Development of a Spatial Path-Analysis Method for Spatial Data Analysis

Wiwin Sulistyo[1,3], Subanar[2], and Reza Pulungan[1]

[1]Dept. of Computer Science and Electronics, Faculty of Mathematics and Natural Sciences, Universitas Gadjah Mada, Indonesia
[2]Dept. of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Gadjah Mada, Indonesia
[3]Faculty of Information Technology, Universitas Kristen Satya Wacana, Indonesia
Emails: wiwinsulistyo@staff.uksw.edu, subanar@ugm.ac.id, pulungan@ugm.ac.id

Article Info	ABSTRACT
	Path analysis is a method for identifying and analyzing direct and indirect relationship between independent and dependent variables. This method was developed by Sewal Wright and initially only used correlation analysis results in identifying the variables' relationship. So far, path analysis has been mostly used to deal with variables of non-spatial data type. When analyzing variables that have elements of spatial dependency, path analysis could result in a less precise model. Therefore, it is necessary to build a path analysis model that is able to identify and take into account the effects of spatial dependencies. Spatial autocorrelation and spatial regression methods can be used to enhance path analysis to identify the effects of spatial dependencies. This paper proposes a method derived from path analysis that can process data with spatial elements and furthermore can be used to identify and analyze the spatial effects on the data; we call this method spatial path analysis.

Corresponding Author:
Wiwin Sulistyo
Faculty of Information Technology, Universitas Kristen Satya Wacana, Indonesia
Jl. Diponegoro 52-60, Salatiga, Central Java, 50711, Indonesia
+62 857 4234 7505
Email: wiwinsulistyo@staff.uksw.edu

1. INTRODUCTION

Path analysis is a problem-solving method with statistical approach developed by Sewal Wright in 1921 [1, 2]. It can be used to parse and explain relationship or effects between independent (exogenous) variables and a dependent (endogenous) variable involved in a particular problem. Path analysis is able to deliver results in the form of a diagram that represents the relationship between the variables, either by direct or by indirect correlation, so that the total effect can be calculated with statistical approach in the form of a causal or structural model [1, 3, 4, 2]. The usual approaches for path analysis are correlation and regression [5, 6]. Some studies have made use of path analysis, among others: the identification of the effect of motivation on math score achievement in reformation-based curriculum [7], the measurement of the effect of e-government utilization on the profits of business ventures in Dubai [8], and the identification of the effects of job stress (time stress and anxiety), job satisfaction, and work motivation on Occupational Health and Safety Administration (OSHA) in Texas [9].

It is interesting to study how path analysis can be used to calculate or process spatial data. Spatial data are a type of data taken in a specific location; thus, they may contain attributes and location information, and hence a kind of dependent data [10]. Spatial analysis is a quantitative analysis that involves spatial data, which include a pattern of dots, lines, areas, etc. in the form of maps and coordinates in two or three-dimensional spaces [11]. The spatial effects in data attributes that geographically have relevance must be identified based on the spatial pattern description and relationships [11, 12]. The spatial effects cause the appearance of spatial dependence. Spatial dependency becomes a very important issue in determining the accuracy of the correlation analysis and the regression model built. To produce a proper correlation analysis and regression model we must consider the effects of spatial dependencies in cases that have a spatial tendency, because the regression model will be considered invalid if we ignore the existing

spatial effects. Path analysis, therefore, needs to consider the spatial effects on variables that are calculated, so that the resulting model has a better accuracy.

Two statistical approaches commonly used in spatial analysis are spatial autocorrelation and spatial regression. Given a set S containing n geographical units, *spatial autocorrelation* refers to the relationship between some variables observed in each of the n localities and a measure of geographical proximity defined for all $n(n-1)$ pairs chosen from n [13]. Spatial autocorrelation has been used to analyze the pattern of migratory attack of brown plant hoppers (nilaparvata lugens) in Boyolali, Klaten, Sragen, and Karanganyar districts in Indonesia [14]. Additionally, spatial autocorrelation has also been used to determine the risk factors for outbreaks of Highly Pathogenic Avian Influenza (HPAI) in Bangladesh [15]. *Spatial regression*, on the other hand, is widely used to identify and examine the magnitude of spatial effects between independent variables and a dependent variable [16, 17]. Spatial regression is a method for analyzing the relationship between an outcome of dependent variable Y and one or more independent variables X and allows for spatial dependence in their observations [18]. Spatial regression can be identified by detecting the condition of spatial dependency based on the error (residual) from the regression model that is initially formed [19].

Spatial autocorrelation and spatial regression have been used to identify patterns of relationship and levels of influence between several regions for some particular variables. Spatial autocorrelation, however, is only capable to identify patterns of relationship that occur, while spatial regression produces values of coefficients (β) that describe the magnitude of the direct influence of independent variables on the dependent variable. Several things cannot be done with this approach, namely how to identify the indirect effects of the independent variables on the dependent variable to produce the total effect, and how to build a model that is able to combine the results of spatial autocorrelation and spatial regression. Spatial path analysis is needed to identify the direct and indirect effects so that the total effects between variables, based on the calculation of spatial autocorrelation and spatial regression, that make up the causal (structural) model can be studied.

This study develops a method derived from path analysis to unravel the relationship and effects between variables to identify the spatial effects that occur. Previously this cannot be accomplished by the standard path analysis method. We call this new method *spatial path analysis*. To demonstrate the feasibility of the proposed method, we apply it to analyze the spatial effects in determining rice productivity and harvest area based on several factors, namely area of the region, rainfall, water spring, rice field area, and population. The end result of this study is expected to produce a spatial model that is able to identify and analyze the effects of each independent variable towards the dependent variable by involving the spatial elements therein.

2. PATH ANALYSIS

Path analysis uses correlation and regression analyses to explain the relationship between the observed variables. Many problems in computer science have made use of statistical approaches based on correlation and regression [20, 21]. This analysis computes direct and indirect effects between independent and dependent variables to produce a total effect in the form of structural equations. The stages of path analysis are depicted in Figure 1.

2.1. Path model

Path analysis begins with determining a path model, which describes relationship between independent and dependent variables. A path model determines a causal model between the variables based on a hypothesis that represents the causals that have been formulated in the path analysis. The causal model is constructed based on a theory or the results of previous research concerning relationship of the variables used in the study. Figure 2 shows several examples of path models.

Figure 2(A) shows the relationship among variables X1, X2 and Y, where X1 directly affects X2 and X2 directly affects Y, but X1 only indirectly affects Y. Figure 2(B) shows that both X1 and X2 directly affect Y, but X1 and X2 do not affect each other. Figure 2(C) illustrates that X1 directly affects Y and X1 also indirectly affects Y through X2. Figure 2(D) shows that X1 and X2 directly affect Y and X1 and X2 correlate.

2.2. Path coefficients

Path coefficients are values obtained from carrying out correlation and regression analyses on the independent and dependent variables in the path model [6, 22]. As shown in Figure 2, every variable has relationship which is depicted by arrows to one or more variables. Figure 2(D) also shows that variable X1 correlates with variable X2, while variables X1 and X2 have regressional relationship toward variable Y. Coefficient values produced by correlation or
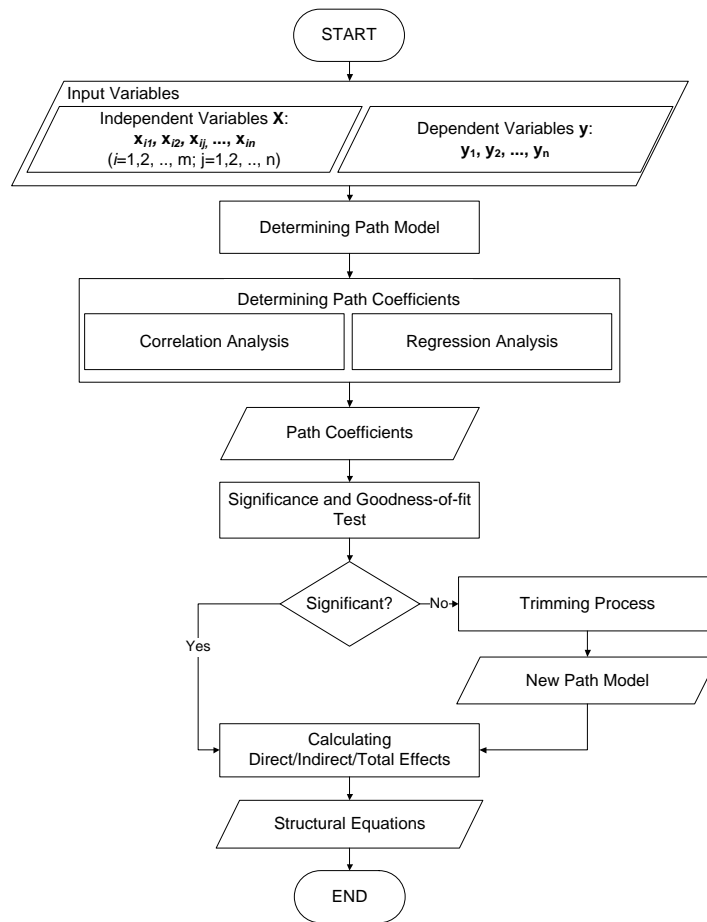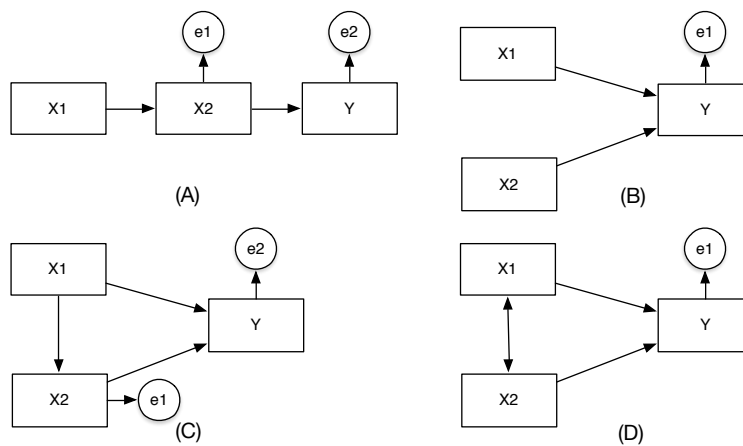
Figure 1. The stages of path analysis

Figure 2. Examples of path models [1]

regression calculation are called *path coefficients*. Therefore, every relationship of several variables in the path model will have a value called a path coefficient.

In this research, some of the path coefficients are obtained by the Pearson correlation equation. Regression analysis, on the other hand, is used to measure the degree of the influence of independent variables toward a dependent variable. One of the methods used to conduct regression analysis is ordinary least squares (OLS) [16, 23, 24, 25]. In this research, some of the path coefficients are obtained by using ordinary least squares.

Furthermore, a significance test is performed to measure the level of significance of each path coefficient value to ensure that the path coefficient formed is significant. This significance testing uses partial test (*e.g.*, t-test) and simultaneous test (*e.g.*, F-test). Beside t-test and F-test, a testing for goodness-of-fit is also performed, which uses coefficient of determination (R^2). Based on the resulting significance level of the path coefficients, a trimming process is carried out to produce a path diagram that consists only of paths that have significant values and to cut away all paths that are not significant.

2.3. Structural equations

The final process of the path analysis is the resulting identification of the effects of independent variables on the dependent variable. The relationship between the variables is described in the form of structural equations. The structural equations are constructed by calculating the direct effects (DE), indirect effects (IE), and the total effect (TE) between the variables [1]. The values of DE, IE, and TE are determined based on the path coefficients.

3. SPATIAL ANALYSIS

Spatial analysis is applied to obtain a more accurate equation model by fixing the errors of data observation. Spatial analysis considers the spatial effects during data observation. Spatial analysis is used to analyze objects (variables) of the research associated with geographic oriented data. Spatial data contains information of a location (coordinates) and the attributes that exist in it. Spatial autocorrelation and spatial regression are approaches that can be used to analyze spatial connectivity.

3.1. Neighborhood analysis

Neighborhood analysis is the process of identifying two or more regions of spatial interaction that intersect each other. Identification of spatial relationships is expressed in the spatial-weights matrix $\mathbf{W}$. There are several methods that can be used for this analysis, among them distance method and contiguity method. For distance method, several approaches can be used to identify relationship between regions expressed as latitude and longitude, namely power function, negative exponential function and general spatial weights. For contiguity method, on the other hand, queen, castle, or bishop contiguity approaches can be used to identify the environmental relationship [26]. The choice of methods depends on the case study and in this paper, we choose queen contiguity. An area that intersects another area will have a value of 1, while an area that does not intersect will have a value of 0 [27]. The set of intersection values between these areas forms a matrix called spatial-weights matrix $\mathbf{W}$. In practice, some methods standardize this spatial-weights matrix to obtain $\mathbf{W}_s$. The result of neighborhood analysis becomes the basis for the calculation of spatial autocorrelation and spatial regression.

3.2. Spatial autocorrelation

Spatial autocorrelation is an analysis of the correlation between the observed regions (spaces) in the form of spatial patterns (patterns of distance, time and area) [28]. On the determination of spatial units (neighborhood analysis), spatial autocorrelation will identify the pattern of spatial relationship based on a certain variable. There are several statistical approaches commonly used to measure the spatial autocorrelation, such as Moran's I [29], Geary's C [30] and the Getis-Ord statistics [31]. In this study, the approach we use to measure the spatial autocorrelation is Getis-Ord statistics (G_i^*). Getis-Ord uses a statistical approach to measure the spatial relationship/correlation by using a matrix based on the intersection of the regions. Getis-Ord statistical method is also used to measure whether the data concentration value is high or low in a specific area [32].

3.3. Spatial regression

Spatial regression measures the effect of an independent variable towards the dependent variable by identifying the spatial effect. General spatial model (GSM) is a linear regression that incorporates the elements of spatial effect into the model. In general, GSM can be written as [16, 23, 25]:

$$\vec{y} = \rho\mathbf{W}_s\vec{y} + \mathbf{X}\vec{\beta} + \vec{u}, \qquad \vec{u} = \lambda\mathbf{W}_s\vec{u} + \vec{e}, \tag{1}$$

where $\vec{y}$ represents the dependent variable and $\mathbf{X}$ represents the independent variables in the form of a matrix containing explanatory variables ($n \times p$), $\vec{u}$ is the error vector (residual) on the spatial autoregressive model (SAR), $\mathbf{W}_s$ is a spatial-weights matrix of standardized queen contiguity, ρ is a coefficient of the spatially lagged dependent variables

$\mathbf{W}_s\vec{y}$, $\vec{\beta}$ is the regression coefficients, which describe the influence of the independent variables ($\mathbf{X}$) to the dependent variable ($\vec{y}$), λ is the error coefficient value of spatially autocorrelated $\mathbf{W}_s$, and $\vec{e}$ is the error vector (residual) in the equation of the spatial autoregressive error model (SER). To carry out this spatial regression, several stages must be completed, namely Moran's test, Lagrange multiplier (LM) and maximum likelihood (ML).

Moran's test   One approach that can be used to test spatial dependency is Moran's test (Z_I) [27, 16, 33]. Moran's test is used to measure the significance level of the result of Moran's I calculation. Moran's I statistic shows the spatial autocorrelation in residual least squares. The Moran's test uses standardized spatial-weights matrix ($\mathbf{W}_s$) to perform the calculations [27]. This test is conducted based on the residual value ε of OLS. The value of Moran's I statistic will be tested to find out its level of significance. From the result of the Moran's test, the existence of the spatial dependence of the variable can be ascertained.

Lagrange multiplier   The next test for the spatial dependency is to determine the appropriate spatial regression equation. This process is performed by using Lagrange multiplier. Lagrange multiplier test consists of spatial error dependency (LM_{error}) and spatial lag dependency (LM_{lag}). Lagrange multiplier performs calculations based on the least-squares residual (OLS) involving spatial-weights matrix ($\mathbf{W}_s$). Anselin in [16] formulated the equations to test spatial lag (LM_{lag}), while Burridge in [34] constructed the equations to test spatial error model (LM_{error}) [17].

Maximum likelihood   In order to calculate the spatial regression we need to determine the values of ρ and λ in the spatial regression equation. In this study, maximum likelihood method is used to obtain these values [16].

4. PROPOSED APPROACH

Basically, the spatial path analysis we now develop is based firmly on the previous path analysis method. The development of the spatial path analysis emphasizes on the spatial process approach involving statistical methods to identify and analyze the spatial effects on variables. In consequence, changes occur in the calculation of the path coefficients. In spatial path analysis there are additional variables, *i.e.*, maps. A map is a spatial data variable that contains an image and attributes. It is in the form of files consisting of shapefile shape (*.shp), shapefile shape index (*.shx) and shapefile attribute (*.dbf).

The stages of the spatial path analysis are depicted in Figure 3, which are further described as follows:

Stage 1 Specifying the path model on the basis of a hypothesis that represents the formulated causals, so that the independent ($\mathbf{X}$) and dependent ($\vec{y}$) variables will be identified in the path model [5].

Stage 2 After the path model is specified, neighborhood analysis is carried out. In this study, the method used to identify the neighborhood is queen contiguity. A spatial-weights matrix is produced to calculate the spatial autocorrelation and identify the spatial dependence. Unstandardized spatial-weights matrix is used to calculate the spatial autocorrelation (Getis-Ord), whereas the spatial-weights matrix is used for calculating spatial dependency (Moran's test).

Stage 3 Performing spatial autocorrelation analysis. At this stage, measurement is performed to determine the strength of the relationship between regions based on the variables used. The method used to analyze the spatial autocorrelation is Getis-Ord statistic. Calculation of the Getis-Ord statistics will generate G_i^* value for each region against other regions based on the variables analyzed. The correlation of G_i^* values for each of these variables is then calculated using Pearson correlation. The correlation coefficients produced will be used as path coefficients that represent the values of the correlation between the analyzed variables.

Stage 4 Performing spatial regression analysis. At this stage, the calculation of regression equation is performed based on the model that has been built to identify causal relationship that occurs. This phase calculates linear regression with ordinary least squares to generate the regression model. Furthermore, identifying the spatial dependency using Moran's test is based on the residual value (ε) of regression model that has been formed. As a result, if there is a spatial dependency, the process of determination of spatial analysis models uses Lagrange multiplier test. If there is no spatial dependency, the coefficients in the regression models that have been produced by OLS will be used as path coefficients on path analysis. Spatial regression models will be determined by the values of ρ and λ generated by calculating LM_{lag} and LM_{error}. The result of calculating LM_{lag} and LM_{error} will determine one of such spatial regression models as SEM, SER, and SGM.

Stage 5 After acquiring spatial regression model, the next stage is determining the values of spatial regression coefficients (ρ and λ) by using maximum likelihood method.
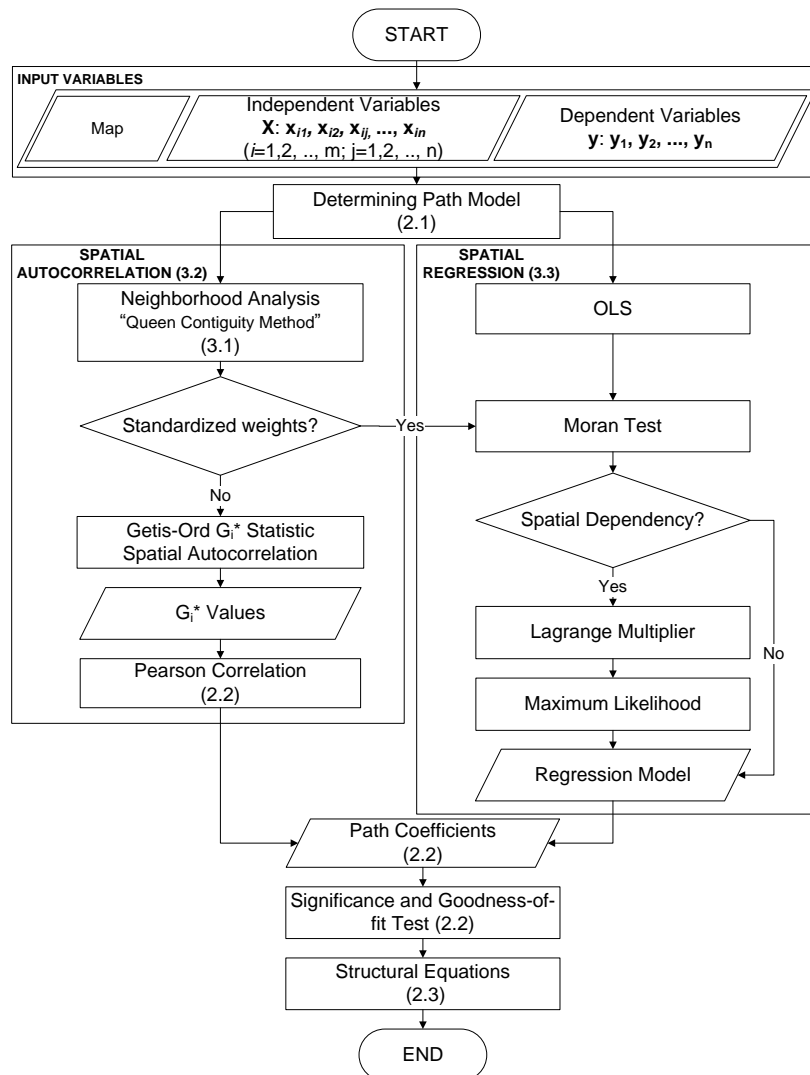
Figure 3. The stages of spatial path analysis

Stage 6 Specifying the path coefficients based on the values of correlation coefficients and regression coefficients. Furthermore, significance tests on the path coefficients are performed using t-test and F-tests as well as a goodness-of-fit test on the regression model based on the values of coefficients.

Stage 7 The next process is forming a structural equation model based on the previous coefficients of path analysis model at Stage 6. At this stage, direct effects and indirect effects are identified to calculate the total effects.

The final result of the spatial path analysis is a structural equation expected to identify the effects of independent variables toward the dependent variable by considering the spatial effects that occur in the model.

5. PROOF OF CONCEPT

Klaten is a regency in Central Java province in Indonesia. Klaten has one of the highest rice productivities in the province. Its total area is 65,556 hectares divided into 26 subdistricts with 39,692 hectares of agricultural land. Geographically, Klaten exhibits very good potentials in agriculture because it has numerous water sources and rainfall of 1,682 mm (year 2015) and most of its area is low land. It is interesting to investigate whether rice productivity and harvest area in Klaten are affected by area, rice field area, rainfall, water source and population.

In this section, we will demonstrate how to use spatial path analysis to identify direct and indirect effects on the dependent variables, namely rice productivity and harvest area in Klaten, based on independent variables, namely area, rainfall, water source and population. Issues that will be analyzed include:

1. Is there any effect of the area, rainfall and water source variables toward rice productivity?

2. Is there any indirect effect of the rainfall variable towards rice productivity through water source variable?

3. Is there any effect of the area, rainfall, water source, and rice field area variables toward the harvest area?

4. Is there any indirect effect of the rainfall variable towards harvest area through water source variable?

Based on the problems that have been proposed, a path model is constructed as shown in Figure 4.
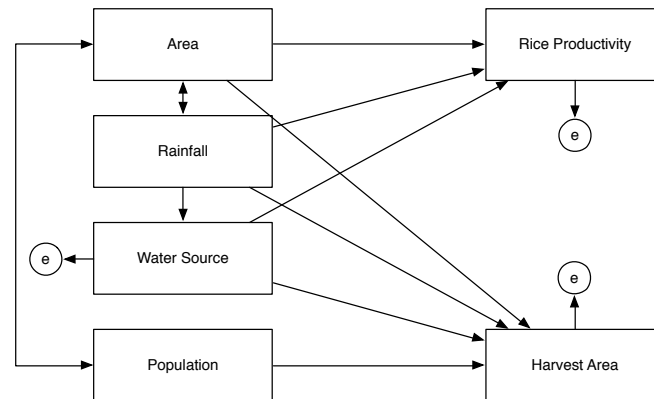

Figure 4. Path model of the case study

In the following, we are going to apply spatial path analysis to identify the inter-variable effects as previously described. This identification process involves the existing spatial effects. Neighborhood analysis produces identification of spatial units based on subdistrict areas as indicated by $polyID$ for each of the 26 subdistricts. Figure 5 shows the map of Klaten and also depicts the result of neighboring identification for each of its subdistricts by using queen contiguity method. This enables us to determine each subdistrict's neighboring subdistricts. The red lines represent that a subdistrict is in a neighborhood with other subdistrict(s). The number on each subdistrict indicates its code $(PolyID)$. The result of this identification is then represented in a spatial-weights matrix for subsequent spatial process.
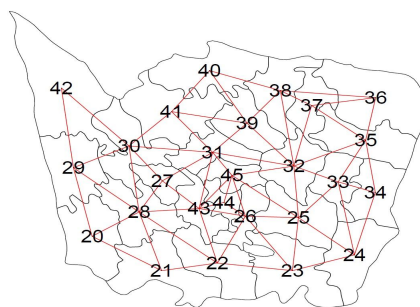

Figure 5. The result of neighborhood identification with queen contiguity approach

Further analysis of spatial correlation is then performed by applying Getis-Ord computation, whose result is depicted in Figure 6. Three variables are analyzed, namely population, rainfall and area. The Getis-Ord spatial autocorrelation aims to identify spatial tendency of each area based on those variables. This process identifies areas with hot-spot status (high values) that exhibit spatial tendency and areas with cold-spot status (low values) that do not exhibit spatial tendency. Figure 6 shows color difference to indicate the difference of the areas' status, where red color represents areas with hot-spot status while blue color represents areas with cold-spot status.

It can be observed that variables rainfall and area are visible in areas with hot-spot status, *i.e.*, subdistricts that are neighboring to each other. Areas with dark red color exhibit hot-spot concentration toward their neighboring areas. Meanwhile, areas with hot-spot status tend to disperse for the population variable.

Beside color that represents status, Figure 6 also depicts values produced from Getis-Ord computation (G_i^*) for each subdistrict. Based on these results, it can be concluded that there exists spatial tendency (hot-spot) in several
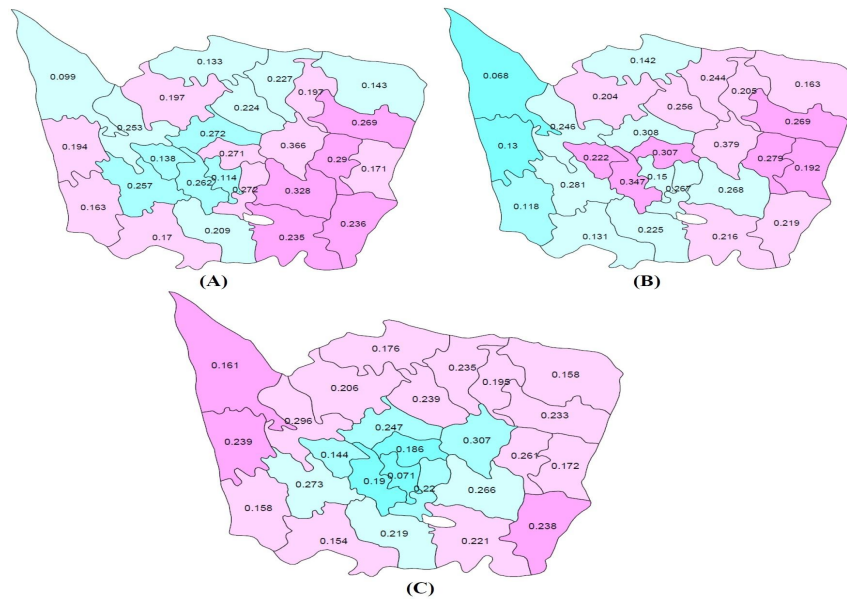
Figure 6. The results of Getis-Ord spatial autocorrelation on: (A) population, (B) rainfall, and (C) area

subdistricts that spatially intersect; and hence they are visually clustered. These conditions exist for rainfall and area variables. On the other hand, for population variable, areas with spatial tendency (hot-spot) are visually dispersed into three groups.

For spatial path-analysis computation, the spatial autocorrelation value (G_i^*) for each subdistrict will be correlated to produce inter-variable correlation values. The correlation results generate the path coefficient values of the variables. Table 1 shows the Pearson correlation analysis result between G_i^* values of the three variables, population, rainfall, and area.

Table 1. The result of Pearson correlation analysis and the corresponding significance

Variables	Population	Rainfall	Area
Population	1.0000	***0.8693	***0.8181
Rainfall	***0.8693	1.0000	***0.6168
Area	***0.8181	***0.6168	1.0000

Significance (p-value): *** = 0.001, ** = 0.01, * = 0.05.

More specifically, the Pearson correlation between the G_i^* values of the area and rainfall variables is 0.6168 (strong) with statistical significance (p-value) of 0.001 (***), indicating very significant. Similarly, the correlation between area and population variables is 0.8181 (very strong) with statistical significance of 0.001 (***), which indicates very significant. These imply that there are positive relationships between spatial autocorrelation G_i^* values of area and rainfall variables as well as between those of area and population variables. These further suggest that an increase in area variable will also increase rainfall and population variables, although this does not necessarily indicate causal relationships between area variable (explanatory variable) and rainfall and population variables (dependent variables).

Next, the spatial regression estimation is performed; this aims to find the effect of a certain variable on other variables based on the existing spatial effect. As previously described in the computational stage (Figure 3), we run the spatial dependency test based on the residual values of the OLS regression. We use Moran's test for our spatial dependency test. Based on the Moran's test, we find that, in general, all regression residuals (ε_i) exhibit significant spatial autocorrelation; this is shown in Table 2. This result explains that there exist spatial dependence elements in the regression residuals (ε_i).

Based on the test result shown in Table 2, we can determine the appropriate spatial regression model that exhibits significant result for each estimation. We use Lagrange multiplier test for spatial error dependence (LM_{err}) and Lagrange multiplier test for spatial lag dependence (LM_{lag}) to determine our spatial regression model. Based on the significance tests, we find that the appropriate spatial regression model is *spatial lag model*. Table 3 shows the

Table 2. The result of Moran's I test and the corresponding significance

Variables	Rice Productivity	Harvest Area
Population	**0.2419	*0.2394
Rainfall	**0.2668	*0.2694
Area	**0.2187	**0.2180
Water Source	*0.1901	*0.1878
Significance (p-value): *** = 0.001, ** = 0.01, * = 0.05.		

intercept coefficient (β_1), the regression coefficient (β_2), and the spatial coefficient (ρ) of our spatial regression model generated by the spatial lag model estimation.

Table 3. The result of spatial lag model's calculation

Variables	Rice Productivity			Harvest Area			Water Source		
	β_1	β_2	ρ	β_1	β_2	ρ	β_1	β_2	ρ
Population	-	-	-	1397.4000	-0.0012	0.4535	-	-	-
Rainfall	-987.4485	4.6927	0.4652	-190.2105	0.8554	0.4696	3.4457	0.0007	0.2722
Area	6943.8063	0.2715	0.4362	1234.0000	0.0607	0.4382	-	-	-
Water Source	7771.5680	-25.1690	0.4380	1412.8930	-5.0441	0.4411	-	-	-

In the next step, we develop a path model based on the spatial correlation and the regression estimation, which generates a diagram depicted in Figure 7.
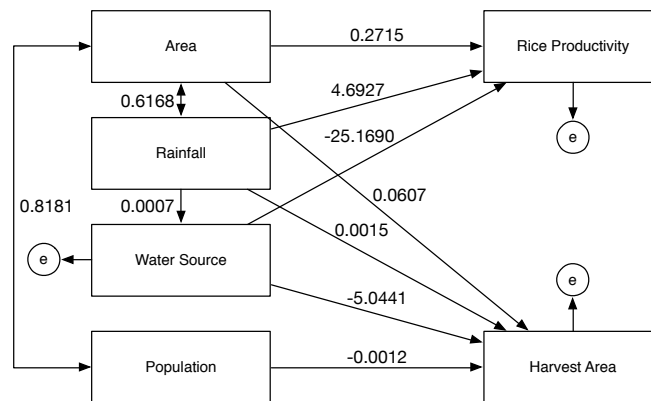

Figure 7. The magnitude of the path coefficients for the variables' relationships

As a result, based on the total effect, we further obtain the following structural equation that represents the relationships between rice productivity (Y_1) variable and harvest area (Y_2) variable with area (X_1), rainfall (X_2), water source (X_3), and population (X_4) variables:

$$Y_1 = 3.1660X_1 + 9.5341X_2 + (-25.1690)X_3, \tag{2}$$

$$Y_2 = 0.1213X_1 + 4.1551X_2 + (-5.0441)X_3 + 0.0484X_4. \tag{3}$$

It then can be concluded that rice productivity is affected by area by 3.1660, by rainfall by 9.5341, and by water source by -25.1690. Meanwhile, harvest area is affected by area by 0.1213, by rainfall by 4.1551, by water source by -5.0441, and by population by 0.0484.

In order to show more closely how all independent variables influence the dependent variable (based on the resulting structural equations), we show the values of all variables side by side in Figure 8 and Figure 9. The x-axis of both charts in the figures represents subdistricts, while the y-axis represents the values of all independent variables (X_1, X_2, X_3, X_4) and dependent variable Y_1 (for Figure 8) and dependent variable Y_2 (for Figure 9).

Based on the results of the first structural equation shown in Figure 8, rainfall (X_2) exhibits the greatest contribution that affects rice productivity (Y_1), implying that higher rainfall increases rice productivity. The second greatest contribution that affects rice productivity comes from area (X_1). As can be seen from the graphical pattern in

the figure, area also supports rice productivity when rainfall drops. For example, this can be indicated at subdistricts 4 and 23. On the contrary, conditions at subdistricts 7, 8, and 25 indicate the cases when low value of area is supported by higher rainfall. Next, although water source (X_3) has greater coefficient value (-25.1690), the fact that it has small value in each area in Klaten makes its contribution to rice productivity less significant. Overall, these results indicate that the most dominant factor in determining rice productivity in Klaten is rainfall, with area also supporting the rainfall's contribution.
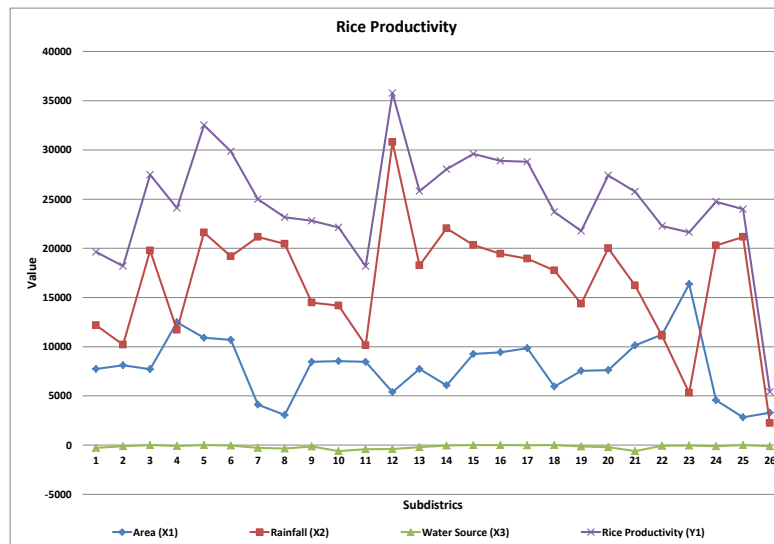

Figure 8. Chart showing how area, rainfall, water source, and population influence rice productivity
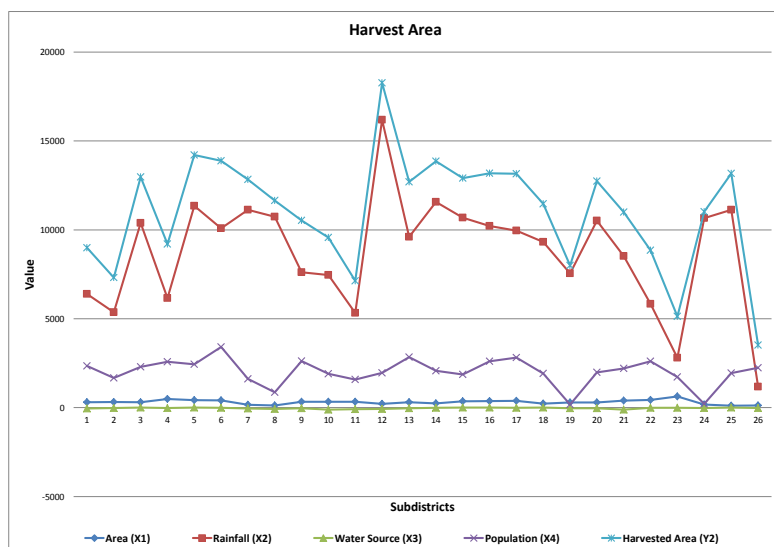

Figure 9. Chart showing how area, rainfall, water source, and population influence harvest area

Figure 9 depicts the results of the second structural equation, namely harvest area (Y_2). Harvest area is heavily affected by rainfall (X_2), with population exhibiting the second largest effect. It can be seen that large population of subdistricts 6, 9, 13, and 22 compensates relatively low rainfall condition to sustain harvest area. Meanwhile, areas with relatively low population are supported by high rainfall level in sustaining their harvest area, such as subdistricts 8 and 24. However, area (X_1) and water source (X_3) exhibit small effects on harvest area. This indicates that subdistricts' areas are not necessarily associated with appropriate environment for rice fields. Similarly, relatively small water source conditions in each area causes water source to have small effect on harvest area. Overall, these

results suggest that the most dominant variable in determining harvest area in Klaten is rainfall, with population also supporting the rainfall's contribution.

Path analysis without spatial analysis   For comparison, we also perform path analysis without considering spatial influences, namely by carrying out the method depicted in Figure 1. Figure 10 shows the resulting path diagram and path coefficients produced by this path analysis. The difference between the results of path analysis and spatial path analysis (as shown in Figure 7) is readily observable. This difference stems mainly from the changes to the variables with the greatest contribution toward rice productivity and harvest area. Other experiments we have conducted also indicate that spatial path analysis produces models that are more consistent than path analysis. This means that spatial dependence is an important factor in path analysis. Proper identification of spatial dependence on the measurement of data results in a more accurate model.
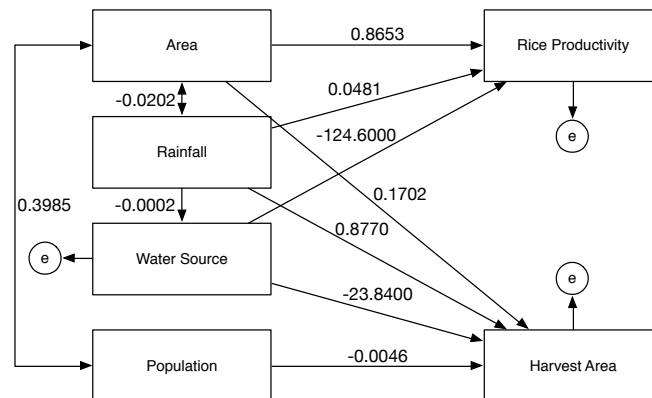

Figure 10. Path coefficients obtained only through path analysis

6. CONCLUSION

       Spatial path analysis is capable of identifying possible spatial effects on variables of interest. The ability of path analysis in disentangling inter-variable relations, both direct and indirect effects, allows one to identify the degree of contribution of each variable toward other variables. Including spatial analysis into path analysis produces structural equations that take existing spatial effects into account. As a consequence, spatial path analysis is supposed to generate better results for cases that involve variables that exhibit spatial tendency.

REFERENCES

[1] R. Schumacker and R. Lomax, *A Beginner's Guide to Structural Equation Modeling*, 3rd ed. Routledge, 2010.
[2] S. W. Leclair, "Path analysis: An informal introduction," *The Personnel and Guidance Journal*, vol. 59, no. 10, pp. 643–646, 1981.
[3] D. L. Streiner, "Finding our way: An introduction to path analysis," *The Canadian Journal of Psychiatry*, vol. 50, no. 2, pp. 115–122, 2005.
[4] R. B. Kline, *Principles and Practice of Structural Equation Modeling*, ser. Methodology in the Social Sciences. Guilford Publications, 2015.
[5] C. Lleras, *Encyclopedia of Social Measurement*. Elsevier Academic Press, 2005, vol. 3, ch. Path analysis, pp. 25–30.
[6] S. Wright, "Path coefficients and path regressions: Alternative or complementary concepts?" *Biometrics*, vol. 16, no. 2, pp. 189–202, 1960.
[7] J. A. Middleton, A. Leavy, and L. Leader, "A path analysis of the relationship among critical motivational variables and achievement in reform-oriented mathematics curriculum," *RMLE Online*, vol. 36, no. 8, pp. 1–10, 2013.
[8] M. A. Badri and K. Alshare, "A path analytic model and measurement of the business value of e-government: An international perspective," *International Journal of Information Management*, vol. 28, no. 6, pp. 524–535, 2008.

[9] P. K. Nair, "A path analysis of relationships among job stress, job satisfaction, motivation to transfer, and transfer of learning: Perceptions of occupational safety and health administration outreach trainers," Ph.D. dissertation, Texas A&M University, 2007.

[10] N. Cressie, *Statistics for Spatial Data*, ser. Wiley series in probability and mathematical statistics: Applied probability and statistics.   Wiley, 1993.

[11] S. Openshaw, "Developing appropriate spatial analysis methods for GIS," *Geographical Information Systems: Principles and Applications*, vol. 1, pp. 389–402, 1991.

[12] D. S. A. Gunadi, I. N. S. Jaya, and B. Tjahjono, "Spatial modeling in landslide susceptibility," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 5, no. 1, pp. 139–146, 2017.

[13] A. Getis, "A history of the concept of spatial autocorrelation: A geographers perspective," *Geographical Analysis*, vol. 40, no. 3, pp. 297–309, 2008.

[14] S. Y. J. Prasetyo, N. Widyawati, K. D. Hartomo, and B. H. Simanjuntak, "Geographic information system for detecting spatial connectivity brown planthopper endemic areas using a combination of triple exponential smoothing - Getis Ord," *Computer and Information Science*, vol. 7, no. 4, pp. 21–29, 2014.

[15] L. Loth, M. Gilbert, M. G. Osmani, A. M. Kalam, and X. Xiao, "Risk factors and clusters of highly pathogenic avian influenza (H5N1) outbreaks in Bangladesh," *Preventive Veterinary Medicine*, vol. 96, no. 12, pp. 104–113, 2010.

[16] L. Anselin, *Spatial Econometrics: Methods and Models*, 4th ed.   Springer Science & Business Media, 1988.

[17] B. H. Baltagi, S. H. Song, and W. Koh, "Testing panel data regression models with spatial error correlation," *Journal of Econometrics*, vol. 117, no. 1, pp. 123–150, 2003.

[18] J. P. LeSage, "An introduction to spatial econometrics," *Revue D'économie Industrielle*, vol. 3e, no. 123, pp. 19–44, 2008.

[19] Y. Xu and E. Kennedy, "An introduction to spatial analysis in social science research," *The Quantitavie Methods for Psichology*, vol. 11, no. 1, pp. 22–31, 2015.

[20] A. Khamis, Y. Xu, and A. Mohamed, "Comparative study in determining features extraction for islanding detection using data mining technique: Correlation and coefficient analysis," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 7, no. 3, p. 1112 1124, 2017.

[21] G. Adhani, A. Buono, and A. Faqih, "Optimization of support vector regression using genetic algorithm and particle swarm optimization for rainfall prediction in dry season," *TELKOMNIKA Indonesian Journal of Electrical Engineering*, vol. 12, no. 11, pp. 7912–7919, 2014.

[22] J. C. Westland, *Modern Path Analysis & Structural Equation Models*.   CreateSpace Independent Publishing Platform, 2012.

[23] J. LeSage and R. Pace, *Introduction to Spatial Econometrics*, ser. Statistics: A Series of Textbooks and Monographs.   CRC Press, 2009.

[24] W. H. Greene, *Econometric Analysis*.   Pearson Education, Inc., 2011.

[25] C. F. Dormann, J. M. McPherson, M. B. Araujo, R. Bivand, J. Bolliger, G. Carl, R. G. Davies, A. Hirzel, W. Jetz, W. D. Kissling, I. Kuehn, R. Ohlemueller, P. R. Peres-Neto, B. Reineking, B. Schroeder, F. M. Schurr, and R. Wilson, "Methods to account for spatial autocorrelation in the analysis of species distributional data: a review," *Ecography*, vol. 30, pp. 609–628, 2007.

[26] A. Fotheringham and P. Rogerson, *The SAGE Handbook of Spatial Analysis*.   SAGE Publications, 2008.

[27] J. P. LeSage, *The Theory and Practice of Spatial Econometrics*.   University of Toledo, Ohio, 1999.

[28] S. Y. J. Prasetyo, Subanar, E. Winarko, and B. S. Daryono, "Endemic outbreaks of brown planthopper (nilaparvata lugens stal.) in Indonesia using exploratory spatial data analysis," *International Journal of Computer Science Issues*, vol. 9, no. 1, pp. 162–171, 2012.

[29] L. Anselin and R. J. G. M. Florax, *New Directions in Spatial Econometrics*.   Springer Berlin Heidelberg, 1995, ch. New directions in spatial econometrics: Introduction, pp. 3–18.

[30] S. Shekhar and H. Xiong, *Encyclopedia of GIS*.   Springer Science & Business Media, 2007.

[31] A. Getis and J. K. Ord, "The analysis of spatial association by use of distance statistics," *Geographical analysis*, vol. 24, no. 3, pp. 189–206, 1992.

[32] ——, *Spatial Analysis: Modelling in a GIS Environment*, 1st ed.   Wiley, 1996, ch. Local spatial statistics: an overview, pp. 261–278.

[33] L. Anselin and S. Rey, "Properties of tests for spatial dependence in linear regression models," *Geographical Analysis*, vol. 23, no. 2, pp. 112–131, 1991.

[34] P. Burridge, "On the Cliff-Ord test for spatial correlation," *Journal of the Royal Statistical Society Series B-Statistical Methodology*, vol. 42, no. 1, pp. 107–108, 1980.