Cartography M.Sc.

Master thesis

Refining Spatial Autocorrelation Analysis for Dasymetrically Disaggregated Spatial Data

Dennis Punong Dizon

2020

Refining Spatial Autocorrelation Analysis for Dasymetrically Disaggregated Spatial Data

submitted for the academic degree of Master of Science (M.Sc.)
conducted at the Department of Aerospace and Geodesy
Technical University of Munich

Author:	Dennis Punong Dizon
Study course:	Cartography M.Sc.
Supervisor:	Dr.-Ing. Christian Murphy (TUM)
Reviewer:	Dr.rer.nat. Nikolas Prechtel (TUD)

Chair of the Thesis
Assessment Board: Prof. Dr. Liqiu Meng

Date of submission: 09.09.2020

Statement of Authorship

Herewith I declare that I am the sole author of the submitted Master's thesis entitled:

"Refining Spatial Autocorrelation Analysis for Dasymetrically Disaggregated Spatial Data"

I have fully referenced the ideas and work of others, whether published or unpublished. Literal or analogous citations are clearly marked as such.

Munich, 09.09.2020 Dennis Punong Dizon

Acknowledgment

The idea behind this thesis had its early beginnings two years ago when I submitted my scholarship application for the MSc Cartography program. In my motivation letter to the Selection Committee, I expressed my research interest in the area of spatial statistics, and that if given the opportunity to become part of the program, I envision working on a topic that merges spatial statistics and cartography, my two greatest academic passions. Fast forward to two years, and I already see the realization of that vision through the completion of this master's thesis.

To the Selection Committee who granted me the honor and privilege of becoming the first student from the Philippines to become part of the Erasmus Mundus MSc Cartography program;

To the highly competent professors of the Technical University of Munich, the Technical University of Vienna, and the Technical University of Dresden who gave the best advanced education in cartography;

To Juliane Cron and Edyta Bogucka who tirelessly and patiently guided the roster of MSc Cartography students and made sure that we have a smooth learning journey all throughout the program;

To Dr. Doracie Zoleta-Nantes and Dr. Arnisson Andre Ortega, my former professors at the Department of Geography of the University of the Philippines who, without any doubt and reluctance, vouched their strong recommendations for my admission application;

To Prof. Francisco delos Reyes, my former thesis adviser and professor at the School of Statistics of the University of the Philippines who further nurtured my interest and knowledge in spatial statistics;

To Dr. Christian Murphy, Dr. Nicholas Prechtel, and Dr. Liqiu Meng, members of my Thesis Assessment Board, who helped me develop my research ideas into an academic work that I can take pride of;

To all the friends I met in and out of the university during my stay in Munich, Vienna, and Dresden and who turned the struggle of being away from my home country into a fun and truly life-rewarding experience;

And to all my family, friends, and loved ones in the Philippines who incessantly showed and made me feel their support during the past two years of my stay in Europe;

I am forever grateful to all of you for helping me turn this dream into an idea, this idea into a vision, and this vision into reality with the completion of this work.

Table of Contents

Table of Figures

Table of Equations

Table of Tables

Abstract

This study explores a refined method for performing spatial autocorrelation analysis with spatial datasets processed for dasymetric mapping, or what is pertinently termed here as "dasymetrically disaggregated spatial data". Conceptually, the refined method applies a spatial weight assignment method where neighborhood weights are assigned based on the area units' spatial configuration, as well as the parent choroplethic zones from where the area units originate. This method is mathematically carried out by algebraically combining two spatial weight matrices: one coming from the conventional, spatial-neighborhood-based spatial weight matrix; and the other from a custom weight matrix based on similarity of choroplethic origins.

Using Siquijor Island, an island province in the Philippines, as test area for conducting a case study, it was affirmatively shown that the revised spatial weight assignment method can yield substantially distinct spatial autocorrelation analysis results. In particular, two different effects were identified, depending on the type of input spatial variable used from the dasymetrically disaggregated spatial data. When the spatial variable used is in the form of raw counts, the refined method produces a filtering effect, resulting in lowered levels of spatial autocorrelation detected in the analysis output. On the other hand, when the spatial variable is in the form of densities, the method oppositely gives an amplifying effect, where intensified spatial autocorrelation levels are instead observed. These two opposite effects are consistently observed whether spatial autocorrelation analysis is measured at the global level, i.e. in terms of the global Moran's *I*, or the Geary's *C* index; or at the local level, i.e. in terms of the local Moran's *I*, or the Getis-Ord G_i^* index.

Keywords: spatial analysis | spatial autocorrelation | dasymetric mapping | spatial weight matrix | cluster and outlier analysis | hotspot analysis | Moran's | Geary's | Getis-Ord | spatial statistics

1. Introduction

1.1. Motivation and research problem

Dasymetric mapping is a cartographic technique that is originally conceptualized to accurately depict the distribution of population across geography on a map (Petrov, 2012). The technique attempts to deterministically delineate portions within discretely defined area units that are considered populated, and consequently, where the values of a given population related variable is to be redistributed (de Greer, 1926; and Wright, 1936 as cited by Mennis, 2009). To achieve this, the original area units are initially disaggregated into smaller subareas, followed by disproportionately reallocating the value of the area unit's population values among the resulting subareas (Dorling, 1993; Langford & Unwin, 1994; Langford, 2003 as cited by Petrov, 2012). This results in a less distorted depiction of population data across space, a positive feature that choropleth mapping, a more popularly used cartographic technique, lacks due to issues related to the "modifiable area unit problem" and "ecological fallacy" in mapping quantities (Crampton, 2011; Langford & Unwin, 1994). Because of this comparative advantage, dasymetric mapping can be considered as a better alternative method to choropleth mapping in order to accurately depict population-based variables across space.

Figure 1 shows an example of a dasymetric map of population for the island province of Guimaras in the Philippines. Here, population data that is originally recorded at the municipal level is disaggregated into smaller subareas based on the varying landcover across the province's land area.

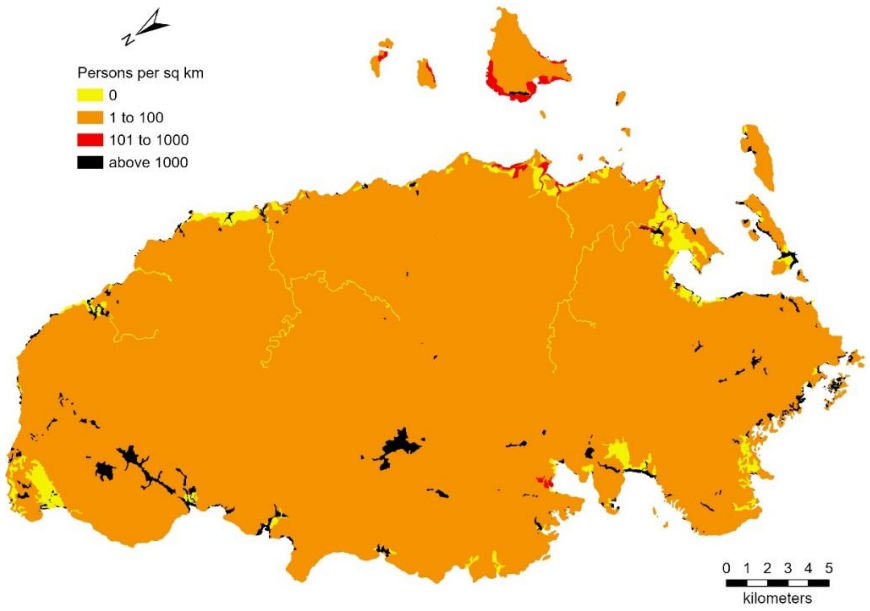


Figure 1. Dasymetric population map of Guimaras Island, Philippines.

Aside from its primary application in cartographic visualization, the underlying spatial data derived from the dasymetric mapping process can also be used as input data for various spatial analysis techniques. One such technique is spatial autocorrelation analysis, a spatiostatistical technique that measures the degree of a quantitative variable's spatial segregation based on the spatial configuration of area units in the data. This analysis technique can be readily performed with any quantitative variable associated to an area-based spatial data, including those that are dasymetrically disaggregated. However, one probable caveat on the reliability of the analysis is its inability to factor in the possible influence of the prior data disaggregation process in the analysis process. In particular, the disaggregated area units are being treated in the analysis as if they are originally disjoint, when on the contrary, these disaggregated units are just a spatially granular form of their formerly aggregated parent units. This poses the possibility of disregarding statistical relationships among area units when performing the analysis in its current form with dasymetrically disaggregated spatial data. Because of this identified gap in the current method, this study explores the possibility of refining the spatial autocorrelation analysis in order to enhance its appropriateness when using spatial data derived from the dasymetric mapping process, or what is formally referred to here as "dasymetrically disaggregated spatial data".

1.2. Research objectives

This study is driven by the following objectives:

1) Identify and set conditions when spatial autocorrelation analysis can possibly be recalibrated when using dasymetrically disaggregated spatial data;

2) Implement computational enhancements in the spatial autocorrelation analysis when using dasymetrically disaggregated spatial data; and

3) Specify differences in the analysis results when running spatial autocorrelation analysis with and without the said computational enhancements.

1.3. Research questions

The research objectives stated above are addressed respectively by answering the following questions:

For research objective 1:

1) *Which applicable spatial autocorrelation measures can possibly utilize dasymetrically disaggregated spatial data?*

2) *Which parameter(s) in the mathematical computation of spatial autocorrelation measures can possibly be modified when using dasymetrically disaggregated spatial data?*

For research objective 2:

3) *What modification(s) in the mathematical computation parameters of spatial autocorrelation measures can be made when using dasymetrically disaggregated spatial data?*

For research objective 3:

4) *How do these modifications differ from each other in terms of results returned in the spatial autocorrelation analysis?*

5) *How do the results of the modified spatial autocorrelation analysis differ from those of the original spatial autocorrelation analysis measures?*

1.4. Significance of the study

As further elaborated in the succeeding chapter, many studies have already demonstrated the use of dasymetric spatial data in analyzing the spatial autocorrelation of population-based variables. In these studies, different measures of spatial autocorrelation are used in their conventional form. However, what this study aims to produce is a reformulated procedure for spatial autocorrelation analysis when the input spatial dataset being used is dasymetrically disaggregated in nature. Furthermore, this study presents a novel approach where the nature of the spatial data (in this case, dasymetric) can prompt changes on how the spatial autocorrelation analysis can be performed.

The results of this study can be of interest to practitioners and professionals involved in analyzing the spatial distribution of populations or any quantitative geographic data. This includes spatial statisticians, demographers, crime analysts, ecologists, epidemiologists, and practitioners of location analytics.

1.5. Hypothesis

In relation to the research questions stated above, it is hypothesized that a substantial difference can be observed when using the modified version of spatial autocorrelation analysis for dasymetrically disaggregated spatial data as proposed in this study. This substantial difference can be quantified in terms of change in the type of detected spatial patterns, the number of occurrences, and the reported statistical significance vis-à-vis when running the conventional form of spatial autocorrelation analysis.

1.6. Scope and limitations

It should be emphasized that though this study highlights the use of dasymetrically disaggregated spatial data, this study does not in any way aim to develop improvements in the existing dasymetric mapping methods nor compare these methods with one another in terms of their accuracy and performance. Commonly used dasymetric mapping methods are identified from the existing body of literature and are adapted in this study's methodology in their original form.

In the aspect of spatial autocorrelation analysis, existing spatiostatistical methods are also reviewed and adapted from the existing literature. Since this study focuses on the refinement of spatial autocorrelation analysis, specific attention is given on retrofitting these existing spatial autocorrelation methods when using dasymetrically disaggregated spatial data as input in the analysis. Moreover, this study does not attempt to compare the spatial autocorrelation methods with one another in terms of their accuracy and performance.

Finally, it must be reiterated that the analytical improvements developed in this study are not directed towards improving the reliability of the dasymetric mapping methods, which in fact covers another area of research that is not tackled here.

1.7. Thesis structure

This manuscript elaborates on the various details of this study through the following chapters:

- Chapter 1 (this section) gives an overview of the research motivation and the problem being addressed; the guiding objectives, questions, scope, and hypotheses of this research; and its foreseen significance in research performed in various related fields.
- Chapter 2 provides a brief discussion of the key concepts revolving around this research, particularly spatial autocorrelation and dasymetrically mapping; as well as the most relevant and recently conducted studies directly related to this study's research problem.
- Chapters 3 and 4 detail the formulated research framework designed for this study based on the objectives that are set and information gathered from the related studies. This section also describes the actual analysis methods and techniques to be undertaken for operationalizing the research framework and applies them using actual data from a selected case study area.

- Chapter 5 details the outputs of the analysis methods and techniques performed using the case study data.
- Chapter 6 gives an interpretation of the results produced and compares if these results are consistent with the hypotheses that were put in place prior to the analysis.
- Chapter 7 synthesizes the results of the study and translates it to a solution statement aligned with the original research problem and set of objectives.
- Chapter 8 lists down the list of bibliographic references cited in this study.
- Chapter 9 contains detailed supplementary information on certain technical aspects of the analysis techniques applied.

2. Review of spatial autocorrelation and dasymetric mapping

2.1. Spatial autocorrelation

The term “spatial autocorrelation” refers to the property or tendency of a spatial variable to exhibit systematic segregation or dispersion over space. A spatial variable is said to be spatially autocorrelated when neighboring locations or nearby areas exhibit similarity in terms of characteristics or magnitudes for a particular attribute or variable of interest (Anselin, 1995). Despite being a naturally observable characteristic in many geographic processes as embodied in Tobler’s First Law of Geography (Tobler, 1979), spatial autocorrelation generally poses complications in many statistical methods, such as in linear regression, where the assumption of independence among observations is a requirement for analytical stability. However, many statistical methods, such as those within the domain of spatial statistics, are developed, not just to accommodate the presence of, but to also measure the degree of spatial autocorrelation present in the data (Dizon, 2018).

Measuring the degree of spatial autocorrelation is equivalent to measuring the presence and the intensity of spatial segregation, i.e. clustering or dispersion, of the variable in question. It can be measured either globally, i.e. to singly characterize the nature of autocorrelation within an entire study area; or locally, i.e. to detect “local pockets of dependence” and describe the spatial autocorrelation of nearby area units (Getis & Ord, 2010). Figure 2 illustrates how global spatial autocorrelation is characterized. In the figure, an increasing global spatial autocorrelation indicates that similar values (visualized as similarly colored hexagonal tiles) tend to bind together spatially over the entire extent of the study area; whereas a decreasing global spatial autocorrelation indicates the opposite tendency of similar values to repel spatially. When values neither tend to spatially bind nor repel, it can be concluded that no spatial autocorrelation is present and there instead is complete spatial randomness. To quantify the level of global spatial autocorrelation, two measures that can be used are the global Moran’s *I* index, and the Geary’s *C* index.

Figure 3 on the other hand illustrates the characterization of local spatial autocorrelation. Here, an increasing local spatial autocorrelation indicates that an area unit is surrounded by either similarly high or similarly low values; whereas a decreasing local spatial autocorrelation indicates that a particular area unit is surrounded by completely dissimilar values. For measuring local spatial autocorrelation, common indices used are the local Moran’s *I* index, and the Getis-Ord G_i^* index.

Figure 2. Visualizing global spatial autocorrelation.

Figure 3. Visualizing local spatial autocorrelation.

One of the measures of global spatial autocorrelation is given by the global Moran's *I* index. The formula (Getis & Ord, 2010) for computing the said index is:

$$I = \left(\frac{N}{\sum_{i=1}^{N}\sum_{j=1}^{N} \mathrm{W}}\right)\left(\frac{\sum_{i=1}^{N}\sum_{j=1}^{N} \mathrm{W} z_i z_j}{\sum_{i=1}^{N} z_i^2}\right)$$

Equation 1. Global Moran's *I* index.

where z_i and z_j are the values (expressed as deviations from the mean) of the spatial variable of interest for every pair of area units (indexed as *i* and *j*); *N* (or *n* for sample-based datasets) is the number of area units in the dataset; and W is a spatial weight matrix, serving as an indicator defining of which pairs of areas *i* and *j* are considered neighbors, and thus will be factored in the index's computation (Esri, Inc., 2018d). The computed global Moran's *I* index can assume a value ranging from +1 to -1 (Esri, Inc., 2018d). A positive global Moran's *I* value indicates spatial clustering at the overall level. A negative global Moran's *I* value on the other hand, indicates spatial dispersion. Finally, a value equal to or closely around 0 suggests complete spatial randomness. The resulting index's statistical significance

can be evaluated by standardizing the values in terms of *z*-scores, and eventually deriving its associated *p*-value (Mitchel, 2005).

Apart from the global Moran's *I* index, another commonly used global spatial autocorrelation measure is the Geary's *C* index (Geary, 1954). The mathematical structure on how it is computed is generally similar with global Moran's *I* index. However, instead of using the product of mean deviations between neighboring area units, Geary's *C* uses the sum of their squared differences (Zhou & Lin, 2008). This can be seen by comparing the Geary's *C* index's formula (Cliff & Ord, 1970) below with the previous one (notations used in the formula are similar with that of the global Moran's *I* index):

$$C = \left(\frac{N-1}{2\sum_{i=1}^{N}\sum_{j=1}^{N}\mathrm{W}}\right)\left(\frac{\sum_{i=1}^{N}\sum_{j=1}^{N}\mathrm{W}(x_i - x_j)}{\sum_{i=1}^{N} z_i^2}\right)$$

Equation 2. Geary's *C* index.

Despite their seemingly slight difference in computation, the ranges in values and their corresponding interpretation differ greatly between the two global spatial autocorrelation measures. In the case of Geary's *C* index, values center around 1 instead of 0, where values less than 1 indicate positive spatial autocorrelation, and values greater than 1 suggest negative spatial autocorrelation (Rodrigues & Tenedório, 2016). Another notable difference between the two global autocorrelation measures is that the Geary's *C* index tends to become more sensitive to local differences among neighboring values than the global Moran's *I* index (Lembo, n.d.; Unwin, 1996). Because of this property, it is possible that the values reported by these two indices specify differing patterns, where one may report positive spatial autocorrelation while the other reports a negative one. Such effects can be addressed by the use of local spatial autocorrelation statistics, which gives more detailed insight on the type of spatial pattern present in each area unit's neighborhood.

The localized counterpart of the global Moran's *I* index, called the local Moran's *I* index, provides a "local indicator of spatial association" (Anselin, 1995), or synonymously, a measure of local spatial autocorrelation. The formula (Anselin, 1995) for computing the said index is:

$$I_i = \left(\frac{z_i}{\sum_{i=1}^{N}(z_j)^2/N - 1}\right)\left(\sum_{j=1}^{N}\mathrm{W}z_j\right)$$

Equation 3. Local Moran's *I* index.

The local Moran's *I* index decomposes global spatial autocorrelation down to the level of the area units, resulting in the measurement of every area unit's contribution to the spatial segregation of the entire variable. In other words, it can identify area "pockets" where localized spatial clusters and/or spatial outliers of the spatial variable exist (Anselin, 1995).

Unlike the resulting values for global Moran's *I* index which are bounded within the range of +1 and -1, the local Moran's *I* index can assume any positive or negative value. A positive value designates an area as a spatial cluster, i.e. an area unit with a value that is similarly high or similarly low with respect to its neighbors (Esri, Inc., 2018b). In contrast, a negative value suggests that the area unit is a spatial outlier, i.e. its value is dissimilar with respect to its neighbors (Esri, Inc., 2018b). Once the type of local spatial autocorrelation is determined (cluster or outlier, or neither) for every area unit, the intensity and significance of local spatial segregation is measured through a permutation-based significance test (Anselin, 1995). One characteristic of the local Moran's *I* index is that it does not readily describe the nature of the spatial cluster or the spatial outlier (Anselin, 1995; Mitchel, 2005). To determine this, the formula is decomposed and the sign of its two components are identified. These components are z_i and $\sum_{j=1}^{N} \mathrm{W}z_j$. When both these components are positive, the area unit is identified as a high-value cluster, termed as a *hotspot* (Anselin, 1995). When both these components however are negative, the area unit is identified as a low-value cluster, or what is called a *cold spot* (Anselin, 1995). Conversely, having differing signs for these two components in an area unit corresponds to that area unit being a spatial outlier, where a negative z_i indicates a low-valued area unit surrounded by high-valued neighbors, and a negative $\sum_{j=1}^{N} \mathrm{W}z_j$ indicates a high-valued area unit surrounded by low-valued neighbors. These two types of outliers are termed as *low-high outlier*, and *high-low outlier*, respectively.

As an alternative to the local Moran's *I* index in determining the type of spatial cluster, i.e. hotspot or cold spot; another measure, called the Getis-Ord G_i^* index can be used. The computational formula (Getis & Ord, 2010) for the index (notations are similar with previous formulae) is given by:

$$G_i^* = \frac{\sum_{j=1}^{N} \mathrm{W}x_j - \frac{\sum_{j=1}^{N} x_j}{N-1}\left(\sum_{j=1}^{N} \mathrm{W}\right)}{\sqrt{\left(\frac{\sum_{j=1}^{N} x_j^2}{N-1} - \frac{\sum_{j=1}^{N} x_j}{N-1}\right)\left(\frac{\left[N\sum_{j=1}^{N} \mathrm{W}^2 - \left(\sum_{j=1}^{N} \mathrm{W}\right)^2\right]}{N-1}\right)}}$$

Equation 4. Getis-Ord G_i^* index.

Positive values of the index indicate the presence of hotspots, while negative values indicate the presence of cold spots (Esri, Inc., 2018c). Unlike the local Moran's *I* index, the Getis-Ord G_i^* index computed for each area unit is readily expressed in terms of z-scores (Mitchel, 2005), thus allowing for a more direct interpretation for statistical significance.

Table 1 summarizes and compares the four spatial autocorrelation measures described above in terms of their type (global or local), the spatial pattern aspect they measure, their respective value ranges, and the corresponding values' interpretation.

Measure	Type	What is measured	Range	Interpretation
Global Moran's *I*	Global	Spatial clustering (or dispersion) of a spatial variable for the whole set of area units	-∞ to +∞	Negative values suggest spatial clustering; positive values suggest spatial dispersion
Geary's *C*			0 to +∞	Values below 1 suggest spatial dispersion; value above 1 suggest spatial clustering
Local Moran's *I*	Local	Spatial clustering (or dispersion) of a spatial variable for an area unit's neighborhood	-∞ to +∞	Negative values suggest spatial clustering; positive values suggest spatial dispersion
Getis-Ord G_i^*		Presence and type (hotspot or cold spot) of spatial clustering of a spatial variable for an area unit's neighborhood	-∞ to +∞	Negative values suggest presence of a cold spot; positive values suggest presence of a hotspot

Table 1. Comparison of spatial autocorrelation measures.

2.2. Spatial weight matrices

Aside from deciding which index to use in measuring spatial autocorrelation, an important component that must be defined in spatial autocorrelation analysis is the *spatial weight matrix*. This component plays a crucial role in any spatial autocorrelation analysis as it provides a mathematical model for defining the neighborhood configuration for the area units in the input spatial dataset (Getis & Aldstadt, 2004). A spatial weight matrix (consistently denoted as W in the spatial autocorrelation formulae given above) is formally specified as an *N*-by-*N* matrix of spatial weights, where *N* is the total number of area units being analyzed. A spatial weight is a numerical value quantifying the type or intensity of

neighborhood between all pairs of area units. Equation 5 shows the typical structure of a spatial weight matrix (Anselin, 2018a).

$$\mathrm{W} = \begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1N} \\ w_{21} & w_{22} & \cdots & w_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ w_{\mathrm{N1}} & w_{\mathrm{N2}} & \cdots & w_{NN} \end{bmatrix}$$

Equation 5. Structure of a spatial weight matrix.

Each matrix element w_{ij} corresponds to a spatial weight assigned to a given pair of area units *i* and *j*, where *j* is an area unit in *i*'s vicinity that is potentially part of its neighborhood (Anselin, 2018a). Spatial weights are assigned to each matrix element to exhaustively define whether area unit pairs are neighbors (thus, a positive number), or not (thus, a value of 0). Commonly, the diagonal matrix elements, which represent spatial weights of individual area units with respect to themselves, are automatically assigned with a spatial weight of 0 (Getis & Aldstadt, 2004).

There are various schemes on how spatial weights are assigned to elements in the spatial weight matrix. Many of these spatial weight matrices are based on the geometric configuration of the area units, or the proximity of the area units with respect to each other (Getis & Aldstadt, 2004). For instance, certain schemes only allow binary spatial weight values of 1 or 0 to be assigned, where the assignment of such values are based on the contiguity of area units. Others, on the other hand, assign spatial weights within the continuous range of 0 to 1 based on the proximity of area units. These various weight assignment schemes represent the different ways on how neighborhood relationships can be defined and quantified across the area units being analyzed for spatial autocorrelation. This idea of using different spatial weight configurations in spatial autocorrelation analysis introduces an area of exploration for this study, which is the identification, and eventually, the formulation of an appropriate spatial weight matrix when analyzing dasymetrically aggregated data.

Below are the conventional spatial weight assignment schemes being used in spatial statistics. The comparison of the different schemes is adopted from the discussion of Dizon (2018), where the island of Marinduque, an island province in the Philippines, is used as an example for illustration.

2.2.1. Contiguity scheme

In the contiguity-based spatial weight assignment scheme, area units *i* and *j* are considered neighbors if they share a portion of their boundaries (Getis & Aldstadt, 2004). This translates to spatial weight matrix elements having either a value of 1, if there is boundary sharing between *i* and *j*; or 0, if otherwise. The scheme can be further refined by specifying whether the “queen” type of contiguity, i.e. when area units share at least an edge or a node; or the “rook” contiguity, i.e. when area units must strictly share at least an edges and not just a node or nodes, is considered between the area units (Getis & Aldstadt, 2004; Anselin, 2018a).

In the illustration by Dizon (2018) shown in Figure 4, the six municipalities of Marinduque Island are indexed as area units 1 to 6, matching with the six rows and six columns in the spatial weight matrix. As illustrated, the contiguity-based scheme assigns spatial weight values of 1 to the matrix elements representing boundary-contiguous area unit pairs, while a value of 0 is given to the rest. It must be highlighted that the same spatial weight matrix will be produced regardless whether the rook or the queen type of contiguity is used in this example.

$$\begin{bmatrix} 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 1 & 0 \end{bmatrix}$$

Figure 4. Contiguity-based spatial weight assignment scheme (Dizon, 2018).

2.2.2. Nearest neighbors scheme

In the nearest neighbors spatial weight assignment scheme, neighborhood is established if *j* belongs to the *k* number of neighbors that are nearest to *i* (Getis & Aldstadt, 2004; Anselin, 2018b). Unlike in the contiguity-based scheme which defines neighborhood based on sharing of boundaries, this scheme finds the *k* nearest neighbors of every area unit *i* based on geometric centroidal proximity and by ranking the centroidal distances of all surrounding area units *j* from area unit *i* (Anselin, 2018b). This scheme is

comparatively more appropriate if the phenomenon being analyzed has spatial autocorrelation characteristics that is significantly influenced by areas “outside its contiguous neighborhood” (Dizon, 2018). Furthermore, this neighborhood scheme ensures that all area units have the same number of neighbors regardless of their size and adjacency (Esri, Inc., 2018e).

Figure 5 from Dizon (2018) shows how the nearest neighbor scheme (k = 3) produces the spatial weight matrix for the same area units of Marinduque Island.

$$\begin{bmatrix} 0 & 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 1 & 0 \end{bmatrix}$$

Figure 5. Nearest neighbors-based (k = 3) spatial weight assignment scheme (Dizon, 2018).

2.2.3. Distance scheme

The distance-based spatial weight assignment scheme uniquely defines neighborhood as a function of the area unit’s distances with respect to other area units. Here, the values of the spatial weights are computed based on the distance-decay function, where spatial weights are equal to the inverse of the squared distance between area units i and j (Getis & Aldstadt, 2004). This translates to the formula (Getis & Aldstadt, 2004):

$$w_{ij} = \begin{cases} \frac{1}{d_{ij}^2}, d_{ij} \leq \delta \\ 0, d_{ij} > \delta \end{cases}$$

Equation 6. Function for computing spatial weights in the distance-based scheme.

where d_{ij} is the centroidal distance between area units i and j, and δ is the threshold distance, which marks the maximum limit up to where neighborhood is only established (Getis & Aldstadt, 2004; Anselin, 2018b). This function produces spatial weights that can assume any decimal value between 0 and 1.

The distance-based scheme is demonstrated in Figure 6 from Dizon's (2018) same example of Marinduque Island, with δ = 20 kilometers.

$$\begin{bmatrix} 0 & 0{,}004 & 0{,}010 & 0 & 0{,}003 & 0 \\ 0{,}004 & 0 & 0{,}004 & 0{,}005 & 0{,}002 & 0{,}002 \\ 0{,}010 & 0{,}004 & 0 & 0 & 0 & 0{,}003 \\ 0 & 0{,}005 & 0 & 0 & 0{,}003 & 0{,}007 \\ 0{,}003 & 0{,}002 & 0 & 0{,}003 & 0 & 0{,}007 \\ 0 & 0{,}002 & 0{,}003 & 0{,}007 & 0{,}007 & 0 \end{bmatrix}$$

Figure 6. Distance-based (δ = 20 km) spatial weight assignment scheme (Dizon, 2018).

Table 2 summarizes and compares the three spatial weight assignment schemes described above in terms of their neighborhood configuration, spatial weight values, and their corresponding interpretation.

Scheme	Neighborhood configuration	Weight values	Interpretation
Contiguity	Based on sharing of boundaries	0 or 1	For rook contiguity: Area unit pairings *i* and *j* with a spatial weight of 1 share a boundary segment; otherwise, spatial weight is 0.
			For queen contiguity: Area unit pairings *i* and *j* with a spatial weight of 1 share a boundary segment or node; otherwise, spatial weight is 0.
Nearest neighbors	Based on proximity ranking		Area unit pairings *i* and *j* with a spatial weight of 1 indicate that *j* is part of *i*'s *k* nearest neighbors; otherwise, spatial weight is 0.
Distance	Based on distance (Euclidean)	0 to 1	Area unit pairings *i* and *j* with distance that is less than or equal to the threshold distance δ has a spatial weight equal to the inverse of their squared distance; otherwise, spatial weight is 0.

Table 2. Comparison of spatial weight assignment schemes.

2.3. Dasymetric mapping

Despite being around for more than a century (Petrov, 2012), dasymetric mapping is considered less popular among cartographers, partly due to the higher level of simplicity and availability of choropleth mapping tools in most geographic information systems (GIS) software applications (M. Langford & Unwin, 1994; Crampton, 2011). In the recent decades however, dasymetric mapping has slowly regained its popularity within the mapping community with the advent and continuing progress in the fields of GIS and remote sensing (Mennis, 2009). Many case studies recently exemplified the effectiveness of dasymetric mapping in mapping various spatial phenomena. For instance, Poulsen & Kennedy (2004) showed how dasymetric mapping was applied to estimate the spatial distribution of residential burglary, while Mitsova et al. (2012) used variations of dasymetric mapping in a study related to shore zone conservation. In addition, two separate studies, respectively by Bajat et al. (2011) and Garcia et al. (2016), show how dasymetric mapping was used in disaster risk management and mitigation, specifically on mapping out the population at risk of flood hazard, and of landslides, respectively. An application of dasymetric mapping in the domain of health and medicine is also demonstrated as shown by Cleckner and Allen (2014) in their study involving the mapping of mosquito-borne-disease-vulnerable populations. Lastly, a unique study by Brehme et al. (2015), shows dasymetric mapping being used in marine resource management, which notably deviates from the commonly taken trajectory of applying dasymetric mapping on human populations.

In principle, the dasymetric mapping process can be divided into two steps. First, a set of geographic area units, called “choropleth zones” are disaggregated into smaller component subareas, called “dasymetric (or sub-choropleth) zones”; and second, the value of a quantitative attribute or “any statistical surface data” associated to the choropleth zones, e.g. population, is allocated into the resulting dasymetric zones (Mennis, 2009). The details of each of the two steps are described below:

2.3.1. Disaggregation

In this initial step of the dasymetric mapping process, the disaggregation of the choropleth zones into dasymetric zones is performed by overlaying it with a second dataset called the “ancillary data” (Mennis, 2009). Figure 7 below illustrates this process.

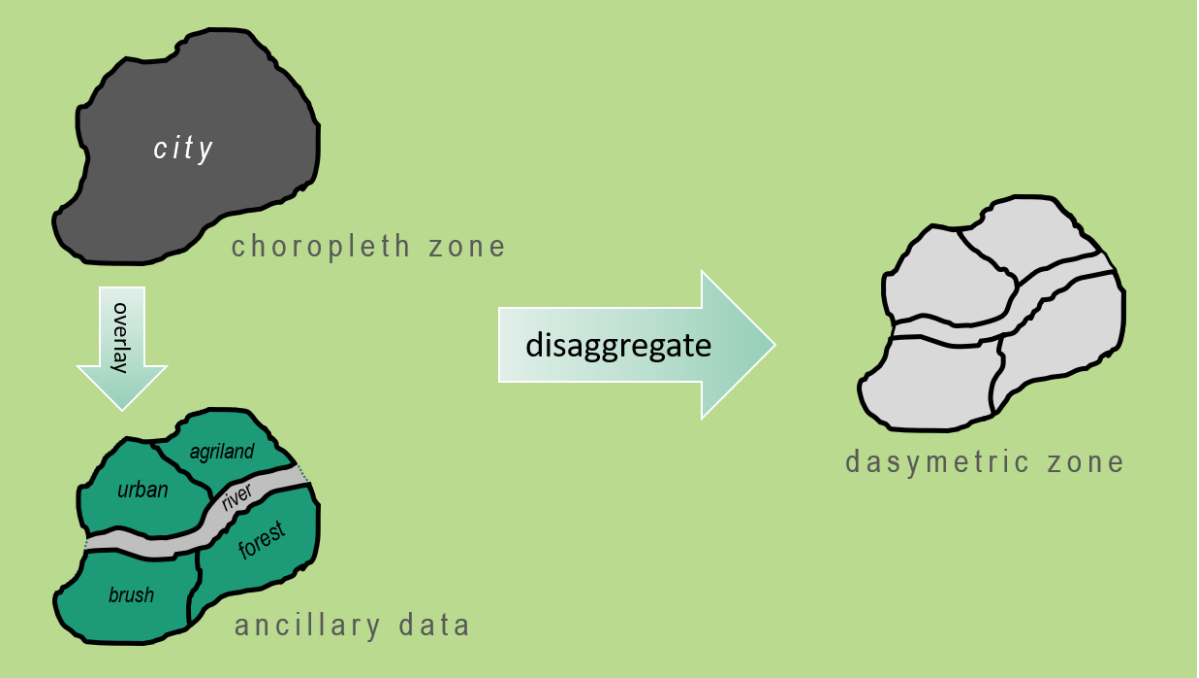


Figure 7. Disaggregation of choropleth zones into dasymetric zones using ancillary data.

Any polygon-based or cell-based spatial dataset can be used as ancillary data so long as it can provide information to the cartographer as to how the attribute can be logically redistributed from the choropleth zones to the dasymetric zones. For instance, landcover data is a widely used ancillary dataset in the dasymetric mapping of human populations since it provides useful information as to how populations can be spatially distributed across an area (Mennis, 2009). However, spatial datasets other than landcover are also used, such as topography, i.e. slope, elevation, curvature (Leyk, Ruther, et al., 2013); soil sealing (Krunić et al., 2015); cadastral boundaries (Maantay et al., 2008; Strode & Mesev, 2013; Jia & Gaughan, 2016); and building clusters (Mans, 2011). Because of its important role in the disaggregation of choropleth zones, the choice and quality of the ancillary dataset is considered essential for the overall accuracy of the dasymetric mapping process (Pavía & Cantarino, 2017).

2.3.2. Allocation

Once the disaggregation of the choropleth zone based on the chosen ancillary dataset is done, the allocation of the quantitative attribute to the dasymetric zones can already be proceed. Though studies on dasymetric mapping have heavily focused on the use of population-related attributes, any quantitative statistical attribute can be used and subjected in the dasymetric allocation process (Mennis, 2009). There are many different methods proposed on how the allocation of the quantitative attribute is done, depending on the geometric algorithm to be applied, and more importantly, the cartographer's

knowledge regarding the “functional relationship” between the quantitative attribute and the ancillary data (Mennis, 2009). Two of the most commonly used methods which are of interest in this study, are the “binary method”, and the “three-class” method (Mennis, 2009).

The binary allocation method is mainly performed by identifying the dasymetric zones that are considered inhabited, thus, where the value of the quantitative attribute is to be exclusively allocated (de Greer, 1926; and Wright, 1936 as cited by Mennis, 2009). An example is shown in Figure 8 to illustrate this method. In the example, the original choropleth zone has a population of 200 and is disaggregated into five dasymetric zones based on landcover. To allocate the choropleth zone’s population value of 200 to the dasymetric zones using the binary method, the dasymetric zones are categorized into either inhabited, or uninhabited areas (labeled as 1 and 0, respectively on the figure). Three out of the five dasymetric zones are considered inhabited, with corresponding landcover classes of agricultural land (agriland), urban, and brush as provided in the ancillary data. The population value of 200 is then finally apportioned among these three inhabited dasymetric zones by basis of the percentage of land area over the whole inhabited area (land areas are specified in the figure beside the ancillary data). The expected result of the allocation using the binary method is shown in the figure, where the urban and brush zones, which individually occupies 3/8 of the total area are each allocated with a population of 75 (3/8 of 200); while the agricultural zone, which occupies 2/8 (or 1/4) of the total land area is allocated with a population of 50 (1/4 of 200).

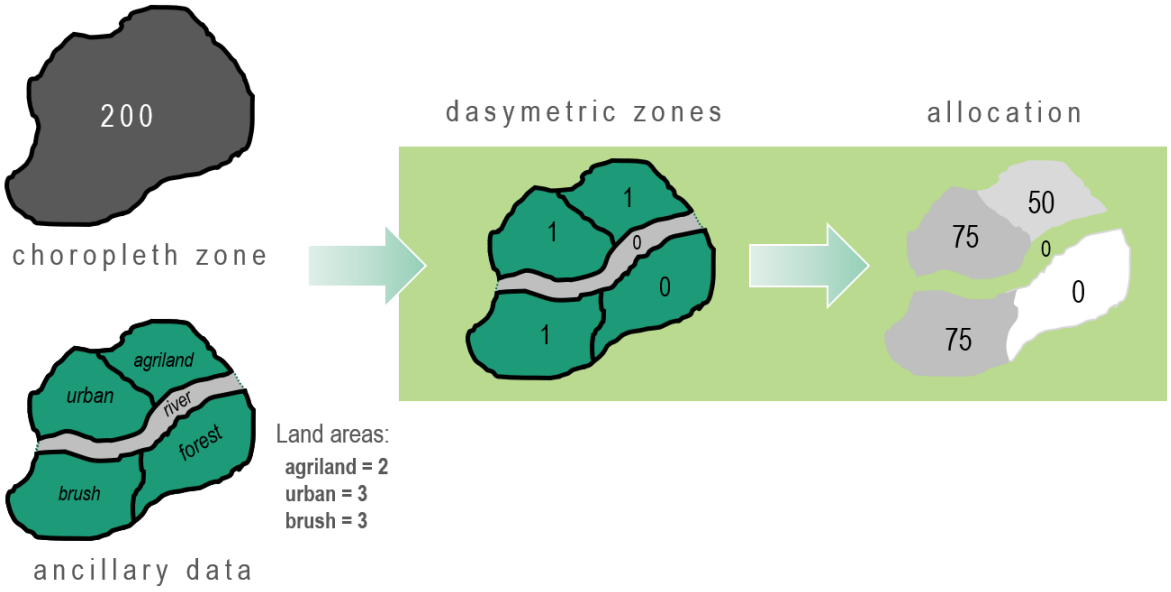


Figure 8. The binary allocation method.

Alternatively, the allocation of values to the dasymetric zones can also be performed using the three-class method. An illustration for the method is shown in Figure 9. In this method, the dasymetric zones defined with the ancillary data are assigned with varying weights depending on the landcover class assigned for the dasymetric zone (Mennis, 2009). The name “three-class” implies that the weight assignment for the allocation is typically done on just three landcover classes, e.g. urban, agriculture, and forest; although any number and set of landcover classes can be used (Mennis, 2009) depending on the cartographer’s assessment. The weight values assigned to the landcover classes is characteristically subjective but must closely reflect the supposed percentage of population present for each landcover class. In the example given in Figure 9 for instance, weights of 0,80 for urban class, 0,10 for brush, 0,05 each for agricultural land and forest, and 0,00 for river are specifically assigned. With the weights assigned per landcover class, the original population value from the choropleth zone is already allocated to the dasymetric zones based on these weights. Furthermore, in cases when multiple dasymetric zones fall in the same landcover class, further allocation is done by again dividing the population value for each landcover proportionately among its corresponding dasymetric zones based on the land area they individually occupy.

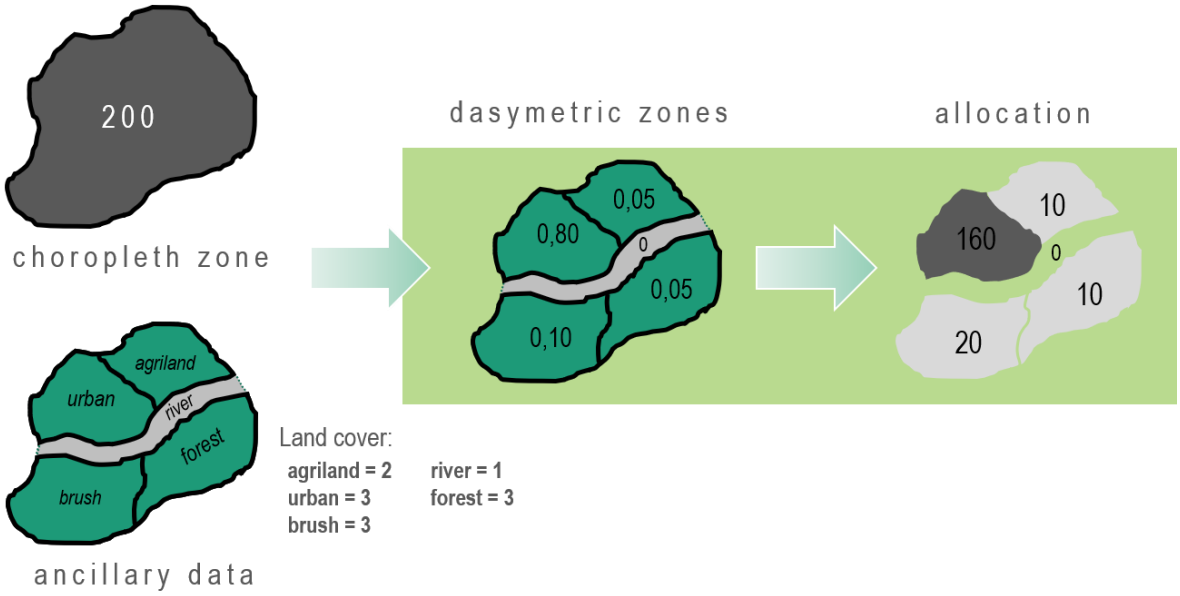


Figure 9. The three-class allocation method.

After the disaggregation and allocation of the quantitative attribute, the resulting dasymetrically disaggregate spatial data can be already be visualized cartographically in the form or a dasymetric map; or be subjected to further statistical data processing and analysis. If a dasymetric map is to be prepared,

it is imperative that the attribute be presented in terms of another normalizing attribute. A classic example of this is the population attribute, which can be normalized by land area such that population density instead of the raw population counts becomes the basis for cartographic visualization. On the other hand, if the resulting dasymetrically disaggregated spatial data is to be subjected to statistical data processing such as spatial autocorrelation analysis, it is possible for either the normalized, density-based form (Boo et al., 2015; Choi et al., 2011; Hu et al., 2007; Mosley, 2012; Parenteau & Sawada, 2012; Weeks, 2010); or the raw, absolute form (Eicher & Brewer, 2001; Jia & Gaughan, 2016; Rodrigues & Tenedório, 2016) of the allocated attribute to be used as the analysis input.

The use of dasymetric mapping at various geographic scales has been sufficiently demonstrated in many works provided that the necessary ancillary dataset, and the quantitative attribute for the source data are concurrently available for the chosen area of interest. At the city level for instance, Barrozo et al., (2011) demonstrated this when they mapped out the elderly population in the city of Sao Paulo, Brazil. Exemplifying dasymetric mapping at the county and provincial mapping levels, Bielecka (2005) dasymetrically mapped the population in the Podlasie Voivodship in Poland, and so did Sleeter (2004) for Alameda County, California in the United States. And finally, at the continental level, Batista e Silva et al. (2013) came up with an improved high-resolution population grid map for the entire Europe using the dasymetric method involving continent-wide landcover data.

Aside from cartographers, many scientists have also been making contributions towards further innovate the original dasymetric mapping process to extend its application outside of mapping. These innovations have been infused to tailor many applications, turning dasymetric mapping from a simple cartographic visualization technique into a spatially oriented tool for modeling population distribution (Mennis, 2009). Statistical procedures comprise most of the enhancements applied into the traditional dasymetric mapping technique in the recent years, most of which are focused on small area estimation (Mennis, 2009; Zandbergen & Ignizio, 2010). Langford (2006) for instance, introduced the use of polycategorical- and regional-regression-based models into the method for performing zone-based population estimations, while Mennis & Hultgren (2005) innovated the population estimation method through the inclusion of sampled statistical data into the dasymetric-based estimation process. In addition, Comber et al. (2008) and Kim & Choi (2011) proposed the use of pycnophylactic interpolation; and others by Leyk, Nagle, et al., (2013), Leyk, Ruther, et al., (2013), and Cockx & Canters (2015) incorporated the assumptions of population heterogeneity and non-stationarity into their method.

2.4. Related studies on spatial autocorrelation analysis of dasymetric data

Many studies have affirmed the usability of dasymetrically disaggregated spatial data in the spatial autocorrelation analysis. Most of these studies employ dasymetric mapping in modeling and estimating population-based data. Boo et al. (2015) for instance, performed small area estimation of dog tumor incidents in Switzerland using dasymetrically disaggregated municipal-level data. In addition, Mosley (2012) performed regression modeling of various health indicators in rural Ottawa using dasymetrically disaggregated census tract-level data. Parenteau and Sawada (2012) also used dasymetrically disaggregated data in the land use regression-based modeling of nitrogen dioxide concentrations in Ottawa. While Hu et al. (2007) used dasymetrically disaggregated spatial data to explore relationship between asthma and air pollution in the Pensacola metropolitan region of Florida. These four studies consistently used spatial autocorrelation analysis to assess the spatial autocorrelation of model residuals, a commonly used spatiostatistical method to quantify modeling and estimation accuracy.

Aside from using spatial autocorrelation to quantify modeling and estimation accuracy, other studies also opted for a more direct and descriptive approach to spatial autocorrelation analysis of dasymetrically disaggregated data. Choi et al. (2011) for example, used Moran's global and local *I* indices, and the Getis-Ord G_i^* index to detect the presence of clusters, hotspots, and cold spots in the resulting dasymetrically disaggregated small area estimates of longevity populations in Gangwon-do, South Korea. Similarly, Weeks (2010) employed the same approach in analyzing the spatial distribution of fertility in rural Egypt using Getis-Ord G_i^* index to detect the presence of hotspots and cold spots in fertility and model spatial dependence.

Aside from the above studies where spatial autocorrelation analysis is directly used in its original form, other closely related works focusing on dasymetrically disaggregated data are also seen to potentially give hints to this study regarding how the spatial autocorrelation analysis can be refined. Rodrigues & Tenedório (2016) for instance, looked at the effect of spatial aggregation (instead of spatial disaggregation) on global spatial autocorrelation statistics. Using global Moran's *I* index and the Geary's *C* index as spatial autocorrelation measures, their work showed that the geographic shape, size, and scale of aggregation reflects different effects on the intensity of global spatial autocorrelation. However, a notable finding that is considered of great importance to this study is that the parameter of neighborhood configuration is shown to have a direct effect on the intensity of spatial autocorrelation

detected. This finding gives possible clues as to which parameters can possibly be manipulated when using dasymetrically disaggregated data. In addition, the work of Reynolds (2011), also points to another useful finding for this study, where spatial autocorrelation (measured in terms of level of aggregation as a proxy measure) is found to increase with the number of dasymetrically disaggregated features. The latter also looked at how spatial autocorrelation's relationships with other relevant properties of dasymetric maps, such as the total number of original features involved, and percent of dasymetric over original features. The study, however, did not look at how originally neighboring dasymetric features affect spatial autocorrelation, something which is of great interest for this current research. Factoring in the results produced by the two latter studies, this study is led to focus on the effect of the neighborhood configuration, expressed in terms of spatial weight matrices (Getis & Aldstadt, 2004), as the important parameter that can be manipulated in the spatial autocorrelation analysis involving dasymetrically disaggregated data.

To sum up, the related literature cited above leads to the conclusion that spatial autocorrelation analysis has ubiquitously been applied to dasymetrically disaggregated datasets. However, it can also be drawn that there is a potential of further improving the said analysis technique by recalibrating certain analytical aspects, particularly those related to neighborhood configuration and spatial weight matrices.

3. Implementing the refined spatial autocorrelation analysis

3.1. Conceptual framework

The conceptual framework illustrated in Figure 10 highlights the two central concepts covered in this study: dasymetric mapping and spatial autocorrelation; together with their conjectured relationship.

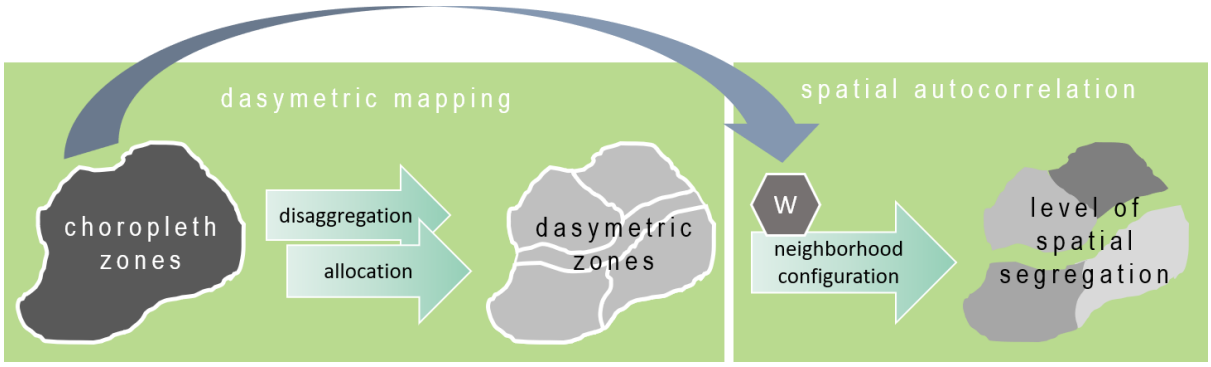


Figure 10. Conceptual relationship of dasymetric mapping and spatial autocorrelation.

As previously described in the preceding chapters, dasymetric mapping entails the process of disaggregating and reallocating a chosen quantitative variable from coarser choropleth zones into finer dasymetric zones, or what this study aptly refers to as “dasymetrically disaggregated (spatial) data”. The resulting data can then be subjected to spatial autocorrelation analysis to characterize the spatial segregation either of the variable itself, or some other data derived from it such as model residuals when the variable is being modeled using linear regression. This is the typical approach that has been copiously applied in many works, particularly the ones cited in the previous section. However, it is postulated here that processing the dasymetric zones directly in this manner disregards certain neighborhood relationships that may exist among dasymetric zones originating from the same parent choropleth zones. The spatial weight matrix (represented by the hexagon labelled W in the figure), is specially highlighted in this study’s conceptual framework because of its potential function in accounting for dasymetrically disaggregated data’s distinct property as a derivative data processed from choropleth zones.

To further illustrate the relationship between the choropleth zones and the spatial weight matrix, Figure 11 shows a hypothetical set of dasymetric zones, shaded into either light green or dark green to categorize them according to their respective parent choropleth zones. The same set of dasymetric zones is used to compare the resulting spatial weight assignment among the three neighborhood

schemes: contiguity-based, nearest neighbors-based, and distance-based. The dasymetric zone labeled *D* is used as the reference area unit, while the dasymetric zones surrounding *D* are labeled with white numbers indicating their assigned spatial weights.

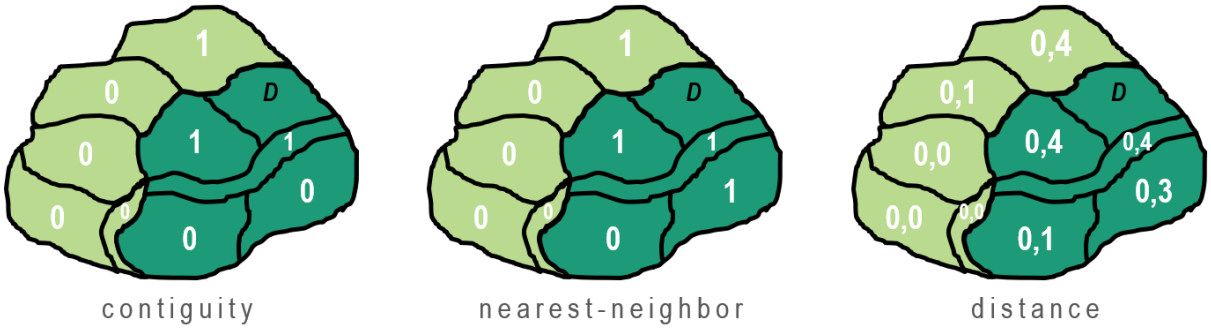


Figure 11. Differences in spatial weight assignment by neighborhood configuration.

In the contiguity- and nearest neighbors-based scheme, some dark green dasymetric zones in the southern area (two in the contiguity scheme, and one in the nearest neighbors scheme) are assigned with a spatial weight of 0, therefore not tagged as neighbors of *D*. On the other hand, one dasymetric zone (the northernmost one) belonging to the light green group is assigned a spatial weight of 1, thus considered as a neighbor of *D* due of its proximity with the latter. A consistent observation can also be seen with the distance-based scheme, where the northernmost dasymetric zone is assigned with a higher spatial weight value of 0,4 than to those located to the south of *D*, which has values of 0,1 and 0,3, respectively. This illustration shows that *D* bears a stronger neighborhood relationship with the nearby dasymetric zone along its northern border than with those located at the southern area, which paradoxically belong to the same parent choropleth zone as does *D*. This neighborhood assignment may be considered flawed since it can be argued that dasymetric zones belonging to the same parent choropleth zones must suitably be more related at the statistical sense than those which are proximate but originating from a different choropleth zone. This argument is further reinforced when we consider the fact that dasymetric mapping strongly works under the assumption that the real underlying spatial distribution of the quantitative attribute is unknown, and that this missing information is merely estimated by the attribute encoded in the choropleth zones (Mennis, 2009). Because of the limiting behavior on the assignment of spatial weights by pure basis of spatial relationship, it is therefore proposed that information regarding the lineage of dasymetric zones with respect to their antecedent choropleth zones be supplied into the spatial weight assignment process. In other words, the proposed spatial weight assignment process when using dasymetrically disaggregated spatial data must simultaneously take

into account the spatial, as well as the original choroplethic configuration of the dasymetric zones. This idea is highlighted in Figure 12 as part of the conceptual framework originally shown in Figure 10.

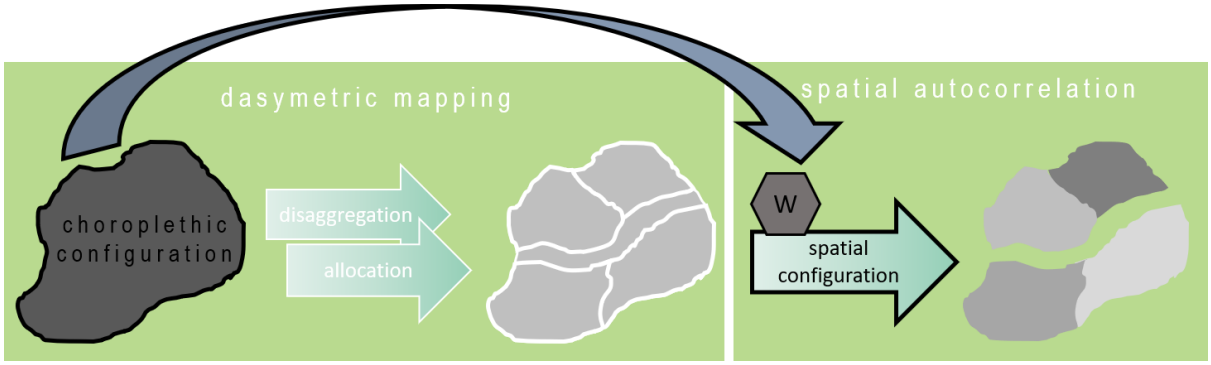


Figure 12. A proposed spatial weight assignment based on choroplethic and spatial configuration.

The revised spatial weight matrix for dasymetrically disaggregated data can be implemented in a hierarchy-based approach, where the spatial configuration and the choroplethic configuration of the dasymetric zones are ranked based on which configuration is to be prioritized. Explicitly, the options for this prioritization can be any of the following:

1) The spatial configuration is prioritized over the choroplethic configuration.

2) The spatial and choropleth configurations are equally prioritized.

3) The choroplethic configuration is prioritized over the spatial configuration.

3.2. Operation

The operation of the conceptual framework is translated into a three-staged workflow. The workflow is shown in Figure 13, and each of the stages are described in further detail below.

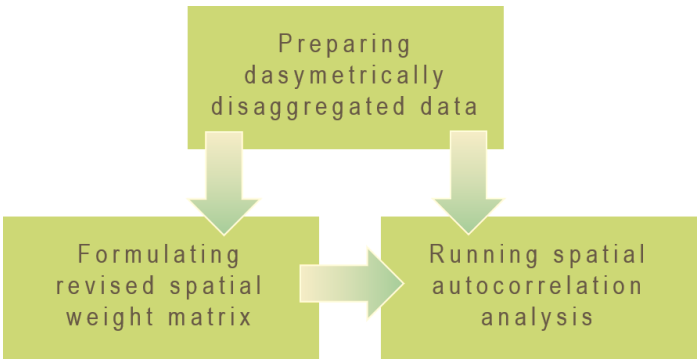


Figure 13. Operation workflow.

3.2.1. Preparing the dasymetrically disaggregated data

In this initial stage of the workflow, the binary and the three-class dasymetric mapping methods are both applied to produce the dasymetrically disaggregated spatial datasets for this study. Vector-based data on administrative geographic units, e.g. city land areas, with population encoded per unit is used as the input choropleth zones. In addition, vector-based data showing areas by landcover type is used as the ancillary dataset. Both datasets are stored as polygon-based vector spatial datasets. Using GIS software, the choropleth zones are to be disaggregated to dasymetric zones by running a spatial intersection between the choropleth zones and the ancillary dataset.

After the disaggregation process, the allocation of the population attribute is applied separately in the two dasymetric mapping methods. For the binary method, the population value is exclusively allocated to the dasymetric zones representing intersections between the city land areas and the urban and agricultural landcover types, leaving all the other landcover areas, such as water, unallocated with any population value (Eicher & Brewer, 2001). On the other hand, for the three-class method, population is allocated disproportionately based on landcover type, such as what was implemented by Holloway et al. (1996, as cited by Eicher & Brewer, 2001), where 70% of a city's population preferentially allocated to urban dasymetric zones, 20% to agricultural and woodland zones, 10% to forested zones, and 0% to zones corresponding to all other landcover types, such as water. Once the population variable is derived, it can then be expressed per dasymetric zone either in terms of raw population counts, or in terms of population densities. At the end of this stage, two sets of dasymetric zones, each respectively having population variables expressed as raw counts and as densities, are readily available for spatial autocorrelation analysis. Comparison can then be made on the output after running the spatial autocorrelation analysis in parallel for both sets of dasymetric zones and population-related variables.

3.2.2. Formulating the revised spatial weight matrix

After preparing the dasymetrically disaggregated datasets, the next stage involves the construction of revised spatial weight matrices which are to be tested and evaluated as part of the spatial autocorrelation analysis. As previously stated in the conceptual framework, this revised spatial weight matrix must simultaneously define the spatial and the choroplethic configurations of the dasymetric zones, while defining the relative prioritization between these two properties. To accomplish this, two spatial weight matrices, each defining the spatial configuration and the choroplethic configuration of the

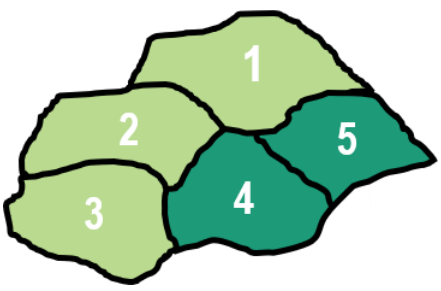
dasymetric zones, are first created separately. This is then followed by fusing the two spatial weight matrices together to form the revised spatial weight matrix incorporating both configurations. Letting W_S be the spatial-configuration-based spatial weight matrix, i.e. any of the conventional spatial weight matrices described in section 2.2; and W_C as the choroplethic-configuration-based spatial weight matrix, the revised spatial weight matrix denoted W_D can be constructed as a linear combination of matrices W_S and W_C, expressed mathematically as:

$$W_D = p_S W_S + p_C W_C$$

Equation 7. Linear combination of spatial weight matrices.

where p_S and p_C are the scalars representing weight values between W_S and W_C. The scalars p_S and p_C define the prioritization between the spatial and the choroplethic configurations, respectively. Their values should sum up to 1 to conveniently express the weights as a proportion value and to facilitate their interpretation and nominalization during analysis. Also, maintaining the weights to sum up to 1 ensures that the range of spatial weight values produced is maintained within the original expected range of 0 to 1.

Figure 14 demonstrates how spatial weight assignment works on a purely choroplethic-based scheme. For such spatial weight matrices denoted W_C, spatial weight assignment is purely based on the similarity of choroplethic origins, such that matrix elements bearing a value of 1 represent dasymetric zones with common parent choroplethic zones, while the rest are given a value of 0. Furthermore, just like how it is done in the conventional spatial weight matrices, the diagonal matrix elements representing spatial weights of individual dasymetric zones with respect to themselves, are consistently assigned with a spatial weight of 0 (Getis & Aldstadt, 2004).



$$\begin{bmatrix} 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

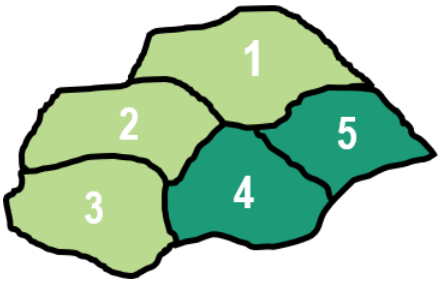
Figure 14. Choroplethic-based spatial weight assignment scheme.

Table 3 shows sample sets of values for p_S and p_C that can be used to perform the linear combination of W_S and W_C to construct W_D, using the possible prioritizations that can be specified between the spatial and the choroplethic configurations previously described in the conceptual framework. Three revised spatial weight matrices, W_{D1}, W_{D2}, and W_{D3}; which correspond to each prioritization option shown in Table 3, are to be constructed. The spatial-configuration-based component, W_S, can be based from any of the three conventional neighborhood schemes, e.g. contiguity scheme, provided that the same neighborhood scheme is consistently used in constructing the three revised spatial weight matrices.

Prioritization	Linear combination
1. The spatial configuration is prioritized over the choroplethic configuration.	$W_{D1} = 0{,}7W_S + 0{,}3W_C$
2. The spatial and choropleth configurations are equally prioritized.	$W_{D2} = 0{,}5W_S + 0{,}5W_C$
3. The choroplethic configuration is prioritized over the spatial configuration.	$W_{D3} = 0{,}3W_S + 0{,}7W_C$

Table 3. Sample linear combinations based on preferred prioritization of configurations.

Figure 15 shows an example of the expected spatial weight matrices for the revised spatial weight matrix W_D. In this example, the linear combination formula for $W_{D1} = 0{,}7W_S + 0{,}3W_C$ is used, where the spatial-configuration-based component W_S is based on the contiguity scheme. As previously noted, the matrix elements contain values falling within the range of 0 to 1, where values 0,3 and 0,7 are assigned for area unit pairings that are considered partially neighbors in both the spatial and the choroplethic aspects.



$$\begin{bmatrix} 0 & 1 & 0{,}3 & 0{,}7 & 0{,}7 \\ 1 & 0 & 1 & 0{,}7 & 0 \\ 0{,}3 & 1 & 0 & 0{,}7 & 0 \\ 0{,}7 & 0{,}7 & 0{,}7 & 0 & 1 \\ 0{,}7 & 0 & 0 & 1 & 0 \end{bmatrix}$$

Figure 15. Revised (choroplethic- and spatial-based) spatial weight assignment scheme.

3.2.3. Running the spatial autocorrelation analysis

In this final stage, the revised spatial weight matrices and the population-based variables, i.e. raw counts and densities, derived for both dasymetrically disaggregated datasets are processed together for spatial autocorrelation analysis. As presented in the literature review, both of the raw counts and densities can be utilized separately for spatial autocorrelation analysis. Different spatial autocorrelation measures are also identified from the related literature and these measures are run using different combinations of inputs, i.e. the revised spatial weight matrix, and the dasymetrically disaggregated datasets. The measures which are utilized in the related studies and to be adopted here again are 1) the Moran's local *I* index; 2) the Moran's global *I* index; 3) the Getis-Ord G_i^* index; and 4) the Geary's *C* index. Considering that there are 2 dasymetrically disaggregated datasets, 2 population-based variable inputs (raw counts vs. densities), 3 revised spatial weight matrices, and 4 spatial autocorrelation measures, a total of (2 * 2 * 3 * 4) = 48 possible input combinations is to be processed and later compared. In addition to these 48 input combinations, the spatial autocorrelation analysis stage shall also include using the spatial weight matrix W_S based on the conventional neighborhood configuration chosen for the construction of the revised spatial weight matrices. This shall serve the purpose of validating the spatial autocorrelation results when using the revised spatial weight matrices as opposed to merely using the conventional spatial neighborhood configurations. This validation yields an additional 16 possible input combinations, i.e. 2 dasymetrically disaggregated datasets, 2 population-based variable inputs, 1 conventional spatial weight matrix, and 4 spatial autocorrelation measures). All in all then, this stage of the workflow entails the running of spatial autocorrelation analysis for 64 runs to cover all the possible input combinations.

Table 4 presents a matrix of all the 64 possible input combinations for the spatial autocorrelation analysis to be run. The 64 input combinations are presented as the matrix cells enclosed in the bold black-outlined box, while the set of possible inputs in the spatial autocorrelation analysis are color-coded for each parameter, i.e. green for the dasymetric mapping methods, blue for the spatial weight matrices, and orange for the spatial autocorrelation measures. The resulting indices for all input combinations are to be reported and compared to describe and characterize how spatial autocorrelation differs in response to changes in these three input parameters.

Spatial weight matrices		**Spatial autocorrelation measures**							
		Global				Local			
		Global Moran's *I*		Geary's *C*		Local Moran's *I*		Getis-Ord G_i^*	
Revised	W_{D1}	Binary	3-class	binary	3-class	binary	3-class	Binary	3-class
	W_{D2}	Binary	3-class	binary	3-class	binary	3-class	Binary	3-class
	W_{D3}	Binary	3-class	binary	3-class	binary	3-class	Binary	3-class
Conventional	W_S	Binary	3-class	binary	3-class	binary	3-class	Binary	3-class
Input: Raw counts		**Dasymetric mapping methods**							
Spatial weight matrices		Spatial autocorrelation measures							
		Global				Local			
		Global Moran's *I*		Geary's *C*		Local Moran's *I*		Getis-Ord G_i^*	
Revised	W_{D1}	Binary	3-class	binary	3-class	binary	3-class	Binary	3-class
	W_{D2}	Binary	3-class	binary	3-class	binary	3-class	Binary	3-class
	W_{D3}	Binary	3-class	binary	3-class	binary	3-class	Binary	3-class
Conventional	W_S	Binary	3-class	binary	3-class	binary	3-class	Binary	3-class
Input: Densities		**Dasymetric mapping methods**							

Table 4. Input combinations for the spatial autocorrelation analysis stage.

After assessing the spatial autocorrelation for each of the input combinations listed in Table 4, the spatial autocorrelation results are then compared to ultimately determine if there are substantial differences in spatial autocorrelation analysis when using the proposed spatial weight matrix as hypothesized in this study. The comparison is to be compartmentalized within each spatial autocorrelation measure and within each variable (raw counts and densities). In the context of Table 4, this translates to the comparison being done within each of the 16 columns (8 columns per table) to detect any consistent behavior when shifting from the conventional to the revised spatial weight matrices. When comparing the global spatial autocorrelation statistics, comparison can be made directly on the index's values, or in terms of their standardized units (*z*-scores). In this study, comparison is to be based on z-scores instead of the index's values since two sets of population variables with differing measurement units are analyzed. Also, the *z*-scores are useful in readily deriving the *p*-values in order to assess the attained statistical significance for the analysis. Furthermore, when comparing the local spatial autocorrelation statistics, comparison is to be made based on the number of statistically significant spatial clusters, spatial outliers, hotspots and cold spots.

4. Applying the refined spatial autocorrelation analysis

4.1. Areas of interest

Two areas of interest (AOIs) are set for the actual implementation of the data processing and analysis workflow of this study. One AOI serves as the extent for the *training dataset*, while the other AOI covers the extent for the *test dataset*. The training dataset is used for the computational development and refinement of the data processing and analysis procedure before getting applied to the larger test dataset, which is considered as the main dataset for analysis in this study, and from which the analysis results are reported and interpreted. Both AOIs are located in the country of the Philippines (see Figure 16).

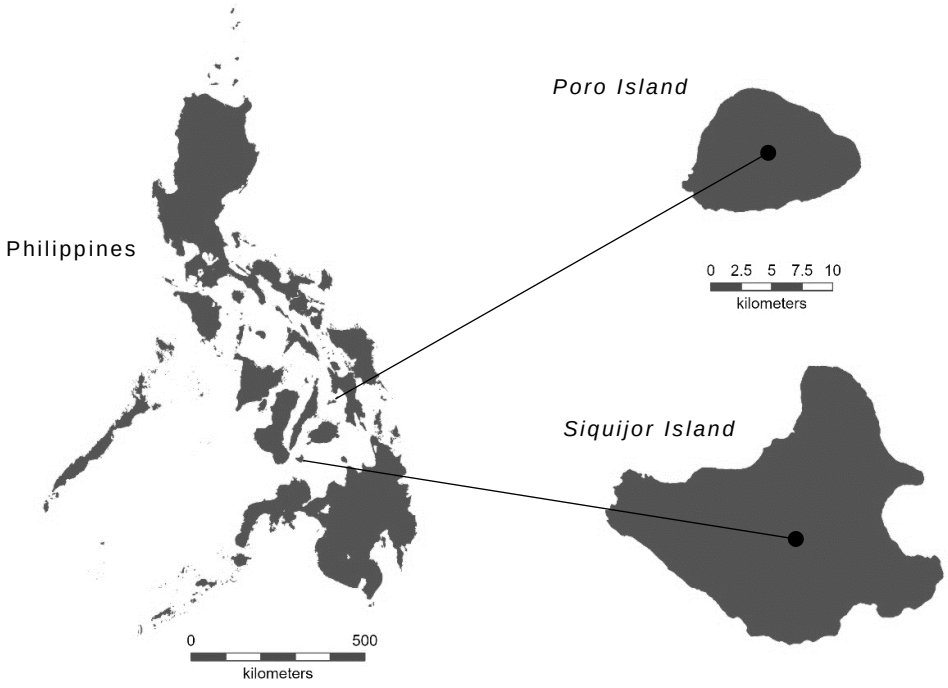


Figure 16. Areas of interest for the case study.

The first AOI is Poro, an island located in the central Philippines with a land area of about 100 square kilometers. Poro Island is chosen as the AOI for the training dataset, primarily because of its compact insular configuration. This compact configuration is favorable for this case study because this ensures that all dasymetric zones have neighbors assigned to them regardless of the spatial configuration used, especially with the contiguity-based scheme, where isolated area units end up having no assigned neighbors. Furthermore, its small land area is seen to facilitate the initial development of the training dasymetric data and spatial weight matrices due to least number of choropleth zones and dasymetric

zones to be processed. For the second AOI, Siquijor, which is an island province also located in the central Philippines and having a land area of about 321 square kilometers, is selected for the test dataset also because of its compact and regular insular configuration as Poro Island. Also, the island province's land area, which is approximately three times that of Poro Island, is another characteristic which is favorable for being chosen as the test dataset AOI.

4.2. Data and analysis parameters

The dasymetrically disaggregated spatial datasets for the case study are prepared using ArcGIS Pro. The spatial datasets used for this are:

1) Municipal land areas for the choropleth zones; and

2) Landcover areas for the ancillary data

Both these datasets are sourced from the National Mapping and Resource Information Authority (NAMRIA), the national mapping agency of the Philippines. Figure 17 shows a map view of these input spatial datasets for Poro Island (left) and Siquijor Island (right).

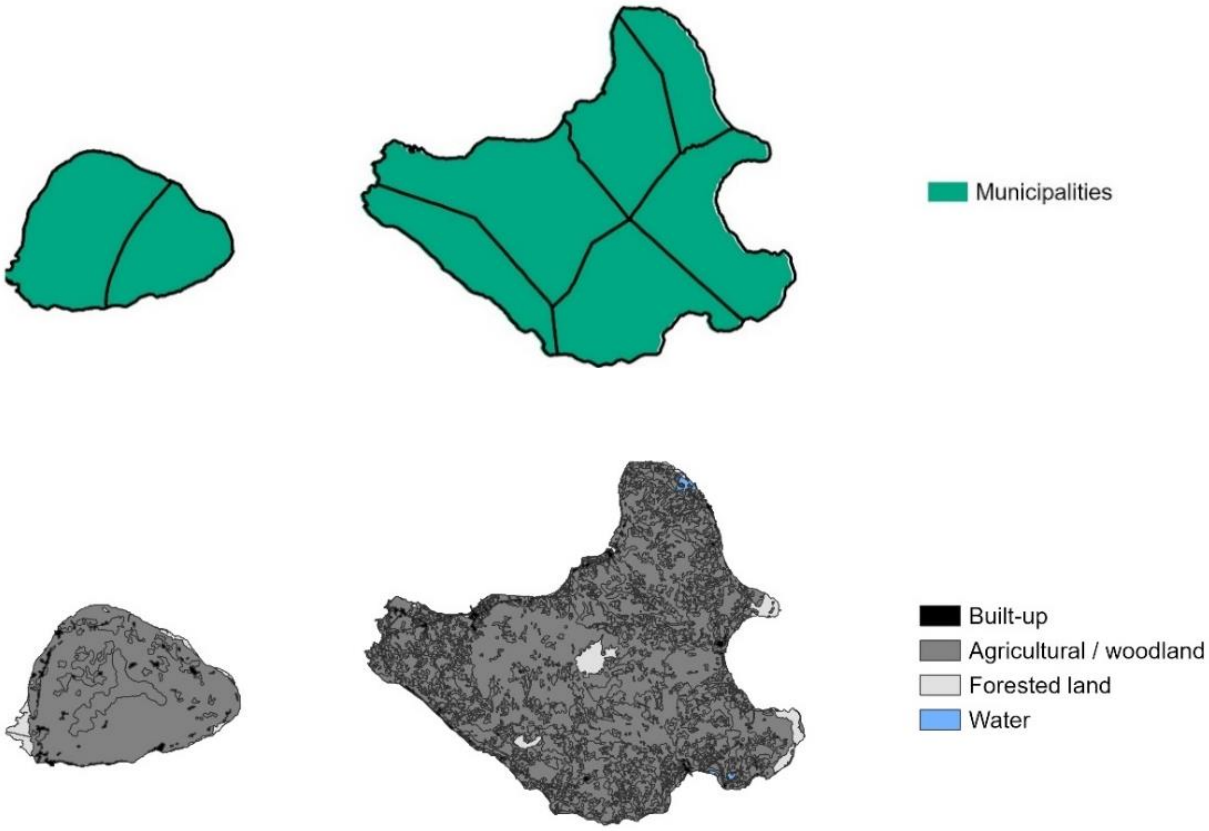


Figure 17. Municipal land areas as dasymetric zones, and landcover as ancillary dataset.

NAMRIA uses a specific set of landcover classification suitable for the Philippines. In order to convert this to what can be used for dasymetric mapping, the landcover classification from the NAMRIA data is reassigned to the reclassification applied by Eicher & Brewer (2001) in their landcover ancillary dataset. Table 5 shows how this reclassification is made for both the binary and the three-class methods.

NAMRIA's landcover classification	**Landcover classes for dasymetric mapping**	
	Binary	**Three-class**
Built-up	Inhabited	Urban area
Annual crop		Agricultural / woodland
Brush / shrubs		
Grassland		
Perennial crop		
Closed forest	Uninhabited	Forested land
Mangrove forest		
Open forest		
Fishpond		Water / barren
Inland water		
Marshland / swamp		
Open / barren		

Table 5. Reclassification of landcover types from the ancillary dataset.

For the three-class method, the weighting used for the allocation of the registered voters per municipality is adapted from Holloway et al. (1996, as cited by Eicher & Brewer, 2001), where 70% is allocated to urban areas, 20% to agricultural and woodland areas, 10% to forested areas, and 0% to water and barren areas.

Adapting the methods of the different case studies cited in the literature review, the case study analyzes the registered voters variable both in terms of raw counts and as densities. By doing this, it can be determined if the revised spatial weight matrices yield the same effect whether the variable used from the dasymetrically disaggregated spatial data is normalized or not. Because of the difference in units and nature of the variables to be compared, comparison of the global spatial autocorrelation measures between the raw counts and the densities is to be based on their standardized values, i.e. in terms of *z*-scores. Local spatial autocorrelation measures on the other hand shall be compared based on the

number of clusters and outliers detected. Statistical significance is to be considered established at a maximum attained *p*-value of 0,05.

The actual attribute data to be used here are the municipalities' registered voters population, which is sourced from the Commission of Elections (Comelec) of the Philippines. Election data for the years 2013 (for Poro Island) and 2016 (for Siquijor Island) are to be used. For the spatial component of the three revised spatial weight matrices – W_{D1}, W_{D2}, and W_{D3}; as well as the conventional spatial weight matrix W_S, the queen contiguity-based scheme is preferentially used because of the compact insular configuration of both AOIs, resulting in all dasymetric zones sharing a portion of its boundary with at least one neighboring zone. Otherwise, if the study areas had non-compact configurations, e.g. elongated, fragmented, or perforated; the two other spatial configurations (nearest-neighbors-based, or distance-based) can be alternately used, since these schemes ensure that non-zero spatial weights are assigned across all dasymetric zones despite them not being totally interconnected.

The construction of the revised and the conventional spatial weight matrices, and the computation of the spatial autocorrelation indices are to be performed using geoprocessing and modeling tools available in ArcGIS Pro. The functionality to directly compute for the Geary's *C* index is not directly available in the said software application. However, since the computational formula for the Geary's *C* index is closely identical to that of the global Moran's *I* index (compare Equation 1 and Equation 2), the script-based geoprocessing tool in ArcGIS Pro for the latter is modified in a separate script in order to be able to compute the Geary's *C* index. The modification applied in the script is guided by the full computational specifications of the two indices as described in Bivand & Wong (2018) and Esri, Inc. (2018a).

5. Results

Using the training dataset, which covers the smaller area of Poro Island, the following models and scripts were successfully completed, and eventually run for the larger test dataset:

1) *Binary Dasymetric Model* – a geoprocessing model that automates the process of creating a dasymetrically disaggregated dataset in ArcGIS Pro through the binary dasymetric method.

2) *Three-Class Dasymetric Model* – a geoprocessing model that automates the process of creating a dasymetrically disaggregated dataset in ArcGIS Pro through the three-class dasymetric method.

3) *Revised Spatial Weight Matrix Model* – a geoprocessing model that automates the process of constructing a revised spatial weight matrix in ArcGIS Pro based on the method described in section 3.2.2.

4) *Spatial Autocorrelation (Geary's C)* – a script-based geoprocessing that computes the Geary's *C* index of an input dataset to determine spatial autocorrelation. This script is modified from the original Global Moran's *I* tool available in ArcGIS Pro.

A copy of the abovementioned models and scripts are available in the media attachment of this thesis manuscript. Also, in the case of the Spatial Autocorrelation (Geary's *C*) script, the computational modifications applied to the global Moran's *I*'s geoprocessing script to enable its computation for the Geary's *C* index is detailed in Appendix 9.1.

The actual analysis results using the above models and scripts, with the Siquijor Island test dataset as input are all detailed below.

5.1. Dasymetric disaggregation

The disaggregation process for the Siquijor Island dataset produced a total of 1 886 dasymetric zones, with allocated values for the estimated number of registered voters ranging from 0 to 10 192. Figure 18 displays the result of the dasymetric disaggregation process of the municipal boundaries as the input choropleth zones, and the allocation of the registered voters count variable for Siquijor Island. The raw count of the registered voters variable in both dasymetric datasets (binary and three-class) is symbolized using the proportional symbolization method. The raw counts variable is illustrated in the

said manner to have an initial visual assessment of the raw counts' spatial segregation, which can then be confirmed later in the spatial autocorrelation analysis.

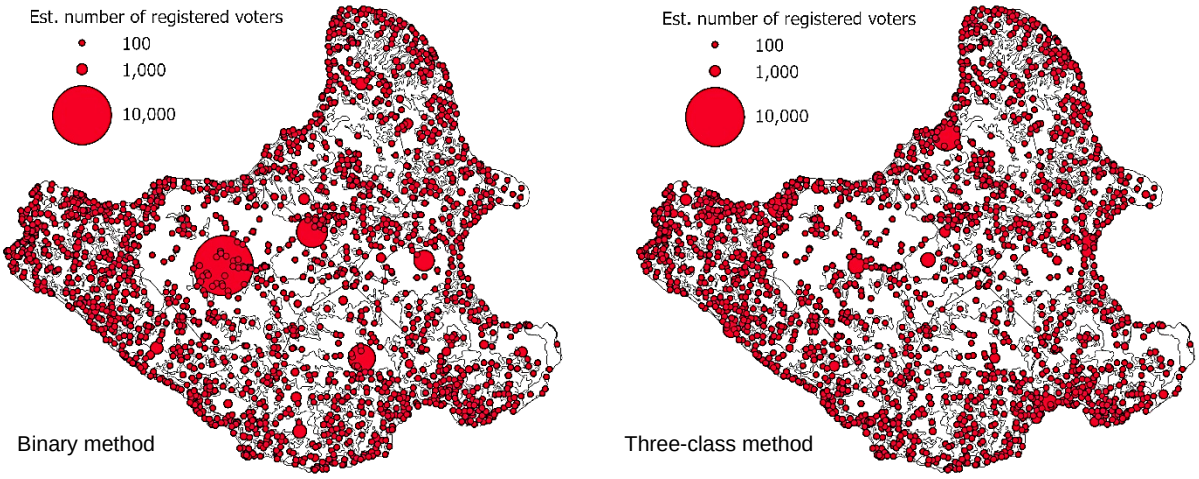


Figure 18. Dasymetric disaggregation of registered voters as raw counts.

The visualization in Figure 18 shows that the resulting raw count variable for the binary method exhibits sharper changes from one neighboring dasymetric zone to another, compared in the three-class method where there is a smoother spatial transition in the dasymetrically disaggregated values. In terms of spatial extent, higher values of registered voters population occupy a larger portion of the study area in the binarily disaggregated dataset than in the three-class dataset. This in return increases the number of adjoining high-valued dasymetric zones in the binary dataset, as well as the number of low-valued dasymetric zones surrounding higher-valued zones. These visually verifiable characteristics point to the possibility of detecting spatial clusters of high values, and spatial outliers of low values in the resulting dasymetrically disaggregated datasets.

On the other hand, Figure 19 displays the disaggregated raw count variable as an actual dasymetric map, where the registered voters variable is displayed in terms of density per square kilometer. The densities in both dasymetric datasets is symbolized using a two-hued color gradient where yellow represents lower density values and blue for the higher density values. In contrast to what is observed in the raw counts variable, there are generally less abrupt changes in the density values of nearby dasymetric zones as shown in the dasymetric map, except in the coastal areas of Siquijor where built-up areas surrounded by less densely populated zones are remarkably highlighted. This is particularly observable in the three-class dasymetric dataset, suggesting that higher levels of spatial segregation can be expected here than in the binary dataset.

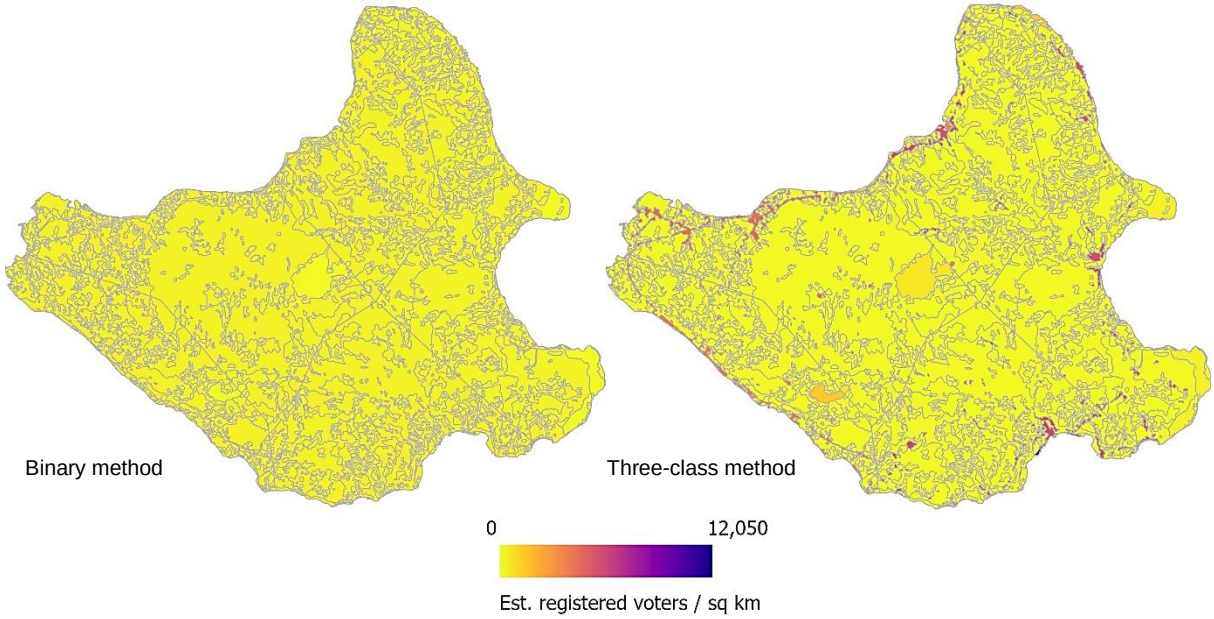


Figure 19. Dasymetric disaggregation of registered voters as densities.

5.2. Spatial weight matrix construction

Since there are a total of 1 885 dasymetric zones derived from the dasymetric disaggregation process, each of the four constructed spatial weight matrices, W_S, W_{D1}, W_{D2}, and W_{D3}, then contains a total of $(1\ 885)^2$ or 3 553 225 matrix elements. This number may be reduced to just 3 551 340 if the 1 885 diagonal matrix elements is not anymore considered due to the de facto zero weight assignment of the area unit pairings where $i = i$. Nonetheless, the original matrix element count is used in determining and reporting the proportion of neighboring area units by neighborhood configuration type in the three spatial weight matrices. These proportions are computed upon construction of the four spatial weight matrix., and Figure 20 shows an Euler diagram (Larsson, 2020) illustrating the comparison of these proportions.

The diagram shows that 82,1% of the 3 553 225 matrix elements represent non-neighboring area unit pairings in neither the spatial nor choroplethic sense. This means that at least 82,1%, or majority of the area pairings will not be considered once the spatial autocorrelation analysis is performed using the four spatial weight matrices. Although this relatively large proportion of non-neighboring area unit pairings is expected for large number of dasymetric zones, this would have been higher if the purely contiguity-based spatial weight matrix, W_S, is merely considered, with the said proportion increasing to 99,7%, or correspondingly, a proportion of just 0,3% for neighboring area unit pairings. However, after incorporating the choroplethic configuration into the W_D spatial weight matrices, 17,6% from the original 99,7% is deducted and reassigned as neighboring area unit pairings. This translates to a sixtyfold

increase in the proportion of neighboring area unit pairings, from just 0,3% (when using just W_S) to 17,9%. In Figure 20, this 17,9% can be seen as composed of three components: 17,6% for the area unit pairings representing choroplethic neighbors, 0,1% for spatial neighbors, and 0,2% for jointly choroplethic and spatial neighbors.

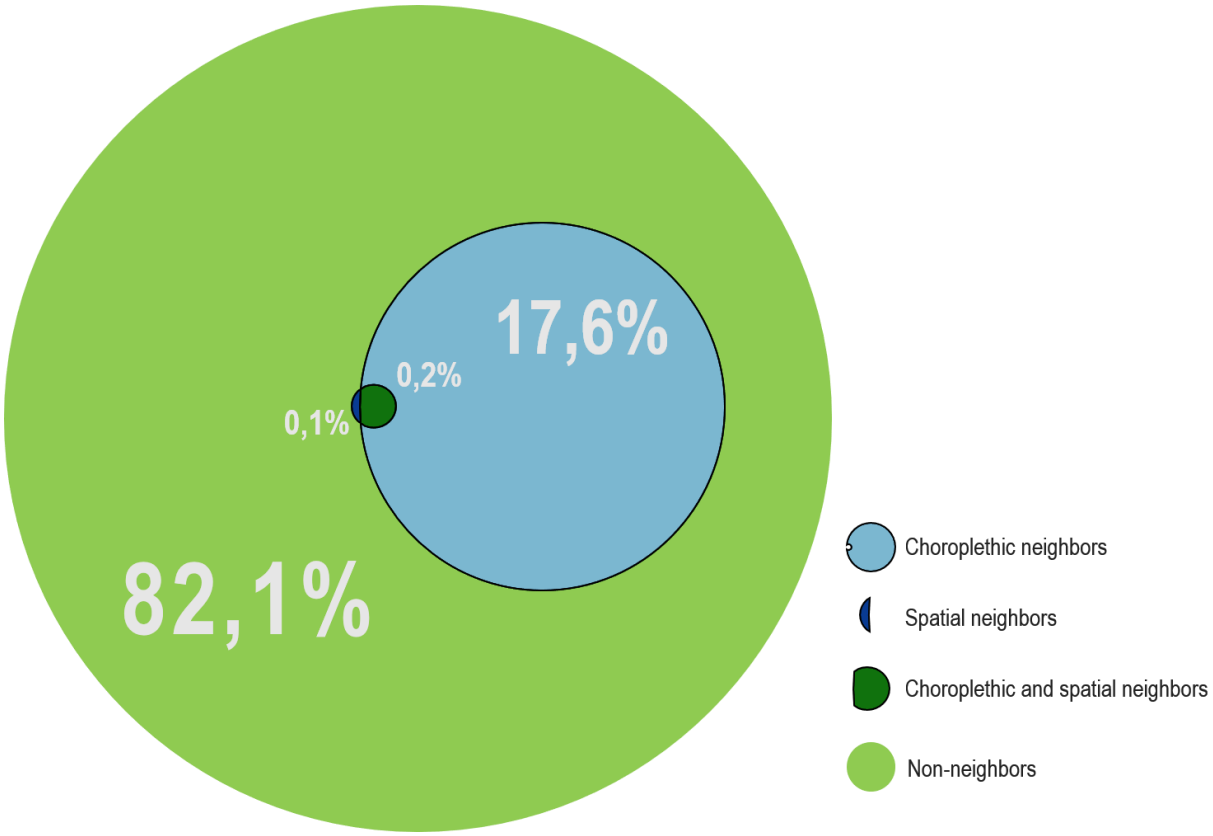


Figure 20. Neighborhood type among area unit pairings.

An alternative visualization in the form of a bar chart in Figure 21 compares these reported differences in the proportion of neighboring and non-neighboring area units between W_S and W_D. From these proportions, it can be expected that the revised spatial weight matrices W_D are to return substantially different results in the spatial autocorrelation analysis compared to when just using W_S. Moreover, differences in the prioritization values for p_s and p_c are expected to also differentiate results among the three revised spatial weight matrices.

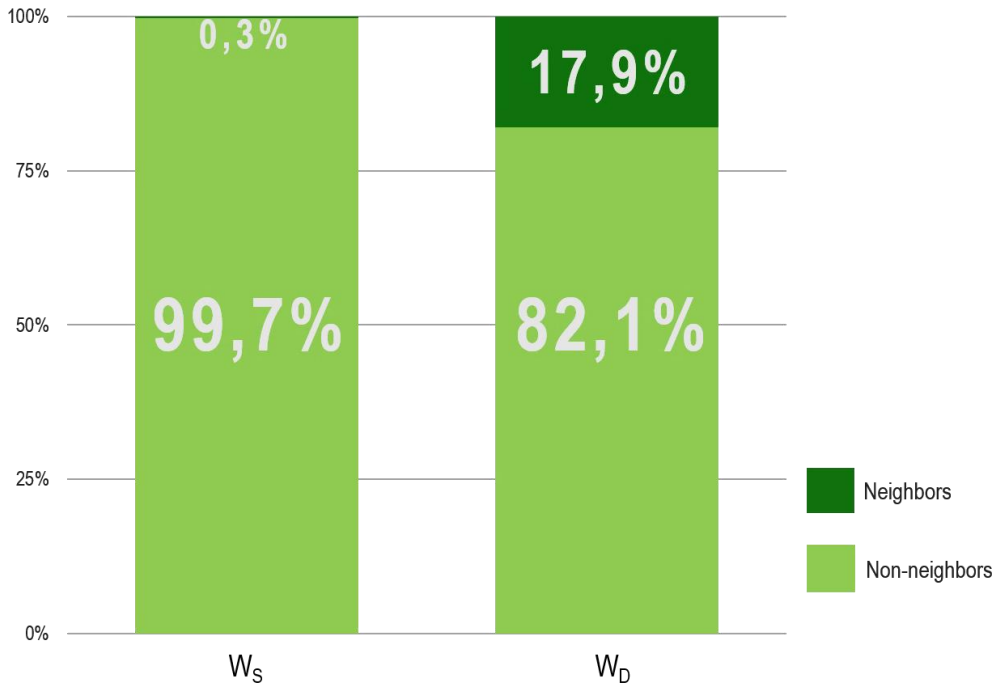


Figure 21. Proportion of neighboring and non-neighboring area unit pairings between W_S and W_D.

5.3. Spatial autocorrelation analysis of raw counts

5.3.1. Global Moran's I

Figure 22 summarizes the computed global Moran's *I* *z*-scores for the raw count of registered voters using the four spatial weight matrices and the two dasymetric datasets as pairwise inputs which are summarized in Table 4. Comparison of the derived values definitively shows a steep decline in spatial autocorrelation when using the W_D spatial weight matrices instead of W_S. Specifically, the global Moran's *I* *z*-score decreases from 24,179 and 3,845 for W_S; to at least -0,598 in all the three revised spatial weight matrices. Negative (albeit small) global Moran's *I* *z*-score values (shown as the bars below the horizontal axis in the figure) are also seen for all the three W_D matrices using the three-class dasymetric dataset and the W_{D3} matrix using the binary dasymetric dataset. These negative values are initially suggestive of an overall spatial dispersion in the data.

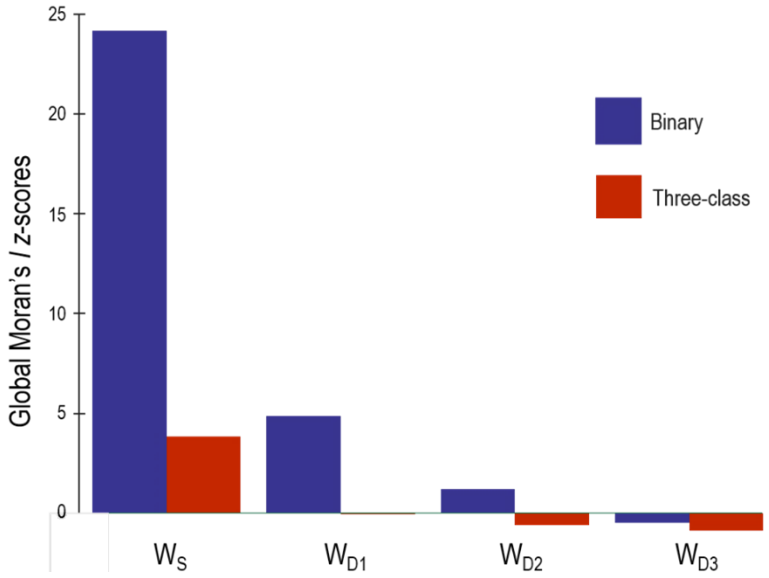


Figure 22. Global Moran's *I* *z*-scores for raw counts by spatial weight matrix.

In terms of statistical significance, *p*-values of ≤0,05 are attained for the global Moran's *I* *z*-scores associated to W_S; and to W_{D1} when using the binary dasymetric dataset. This applies as well to all the negative global Moran's *I* *z*-scores found under the three revised spatial weight matrices. Collectively, these values suggest that as far as the global Moran's *z*-score is concerned, statistically significant overall spatial clustering is only detectable in the dasymetric datasets when using the purely spatial-based weight matrix, and that overall spatial randomness is observed when using all the three revised spatial weight matrices (with exception to W_{D1} when used with the binary dasymetric dataset).

The complete set of computed global Moran's *I* index values and *z*-scores are organized and tabulated per input combination in Table 6 found in the Appendix.

5.3.2. Geary's C

Figure 23 summarizes the computed Geary's *C* *z*-scores using the same set of input combinations. A similar steep decline in *z*-score values is again observed as the computation of the index shifts from using the W_S to the W_D matrices. For instance, in the three-class dasymetric dataset, the highest value for the index is recorded at 30,415 when using W_S, and sharply decreased to at least 1,013 when using the three revised spatial weight matrices. In the aspect of statistical significance, *p*-values of ≤0,05 are again attained for the all input combinations involving W_S; and to W_{D1} when using the three-class dasymetric dataset.

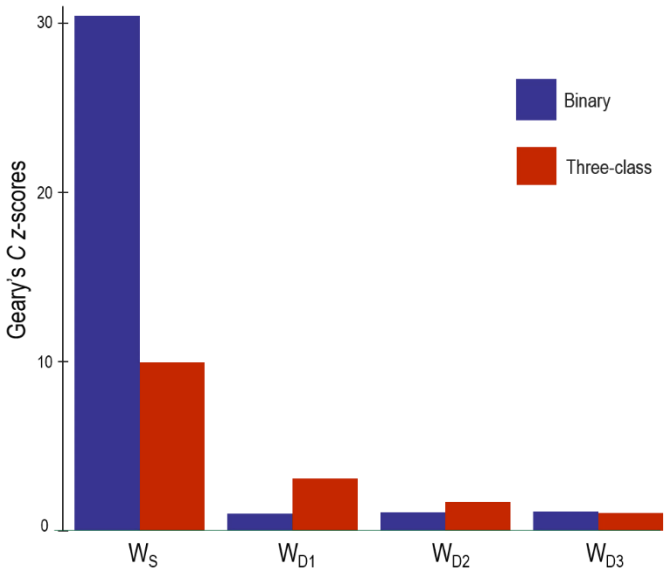


Figure 23. Geary's *C z*-scores for raw counts by spatial weight matrix.

Unlike in the global Moran's index, there is no shifting in signs that occurred in the Geary's *C z*-score, which remains positive across all the input combinations. However, one notable observation that can be made is that the computed Geary's *C z*-scores are all found to be positive. Such set of values signify overall spatial dispersion in both dasymetric datasets, which is contrary to what is concluded from the analysis using global Moran's *I*. As discussed previously in the theoretical background, this contrast in spatial autocorrelation patterns as reported by the two indices can be explained by the latter's greater sensitivity to high differences in values among neighboring dasymetric zones than the former (Lembo, n.d.; Unwin, 1996). This happens when there is strong local negative spatial autocorrelation or local spatial dispersion in the data. This is consistent with what is previously manifested in the map visualization of the dasymetric datasets provided in Figure 18, but this possible presence of local spatial dispersion can only be ascertained in the analysis of local spatial autocorrelation analysis.

The complete set of computed Geary's *C* index values and *z*-scores are organized and tabulated per input combination in Table 7 found in the Appendix.

5.3.3. Local Moran's I

Figure 24 summarizes the number of dasymetric zones identified as statistically significant spatial clusters (left) and outliers (right) in raw counts using the local Moran's *I* index per set of input combinations. Just like the trend already seen in the global spatial autocorrelation results, the graphs likewise show a consistent yet gradual (not steep) decrease in number of localized clusters and outliers

when shifting from the W_S to the three W_D spatial weight matrices. More than half of the dasymetric zones initially identified as clusters and outliers are eliminated upon using the W_{D1} matrix, with the decrease ranging from 57% to 69% among the four input combinations involving W_S. This reduction further progresses when the W_{D3} matrix is used, where only 3% to 23% of the originally identified clusters and outliers are retained, until just 1 spatial cluster remains when using W_{D3}.

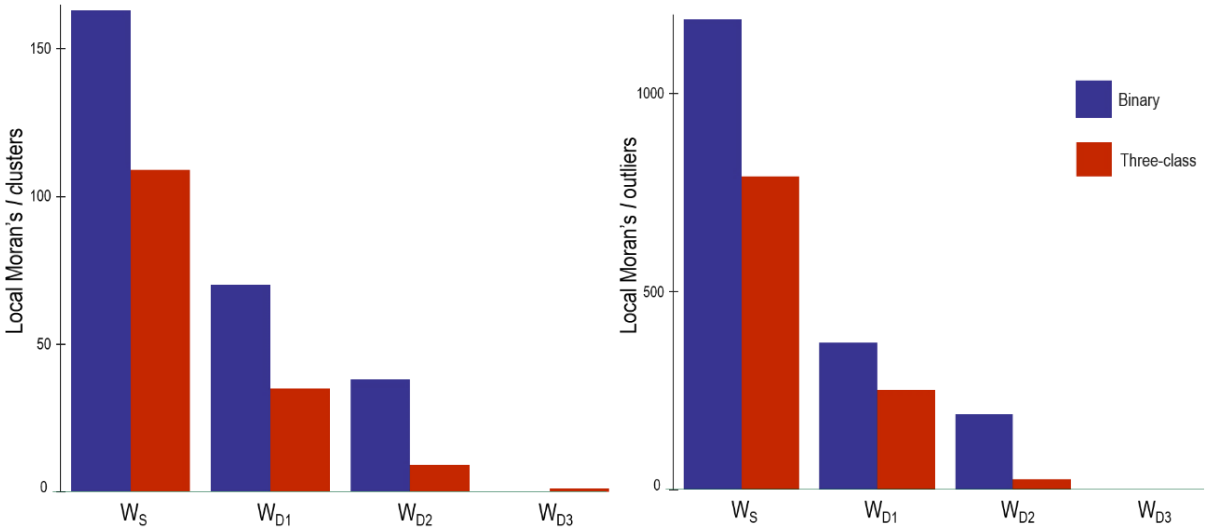


Figure 24. Local Moran's *I* clusters and outliers for raw counts by spatial weight matrix.

The number of spatial clusters and outliers based on the local Moran's *I* values are tabulated per input combination found in Table 8 in the Appendix.

A map view showing the distribution of spatial clusters and outliers is presented in Figure 25 and Figure 26. Figure 25 shows the results for the binary dasymetric dataset while Figure 26 shows the results for the three-class dasymetric dataset. Each figure contains 4 maps, each corresponding to the result produced for each of the four spatial weight matrices. Again, both sets of maps consistently show the declining number of localized clusters and outliers when using the revised spatial weight matrices. However, aside from the declining number, the maps also highlight the area covered by the remaining clusters and outliers, where most of the remaining ones are located along the edges of the choropleth zones (cf. Figure 17, which shows Siquijor Island's municipal boundaries as choropleth zones). The maps also further support the previous remark made in the preceding spatial autocorrelation indices where the binary dasymetric datasets report higher degrees of spatial autocorrelation than its three-class counterpart, this time in terms of the number of localized clusters and outliers detected.

To determine the nature of the spatial outliers, the dasymetric zones tagged as such are further classified into either a low-high outlier, where a low-valued zone is surrounded by high-valued zones; or a high-low outlier, where a high-valued dasymetric zone is surrounded by low-valued zones. In both input dasymetric datasets, all except for 2 of the spatial outliers identified are classified as low-high spatial outliers. These low-high outliers represent 99,8% of all the identified spatial outliers, obfuscating the presence of the mere 2 (or 0,2%) high-low outliers in the study area. This outcome reconciles with the visual pattern of the dasymetric zones shown in Figure 18 where lower-valued zones are enclosed by high-valued zones, as well as with the prior finding in the analysis of Geary's *C z*-scores where local spatial dispersion is present in the data.

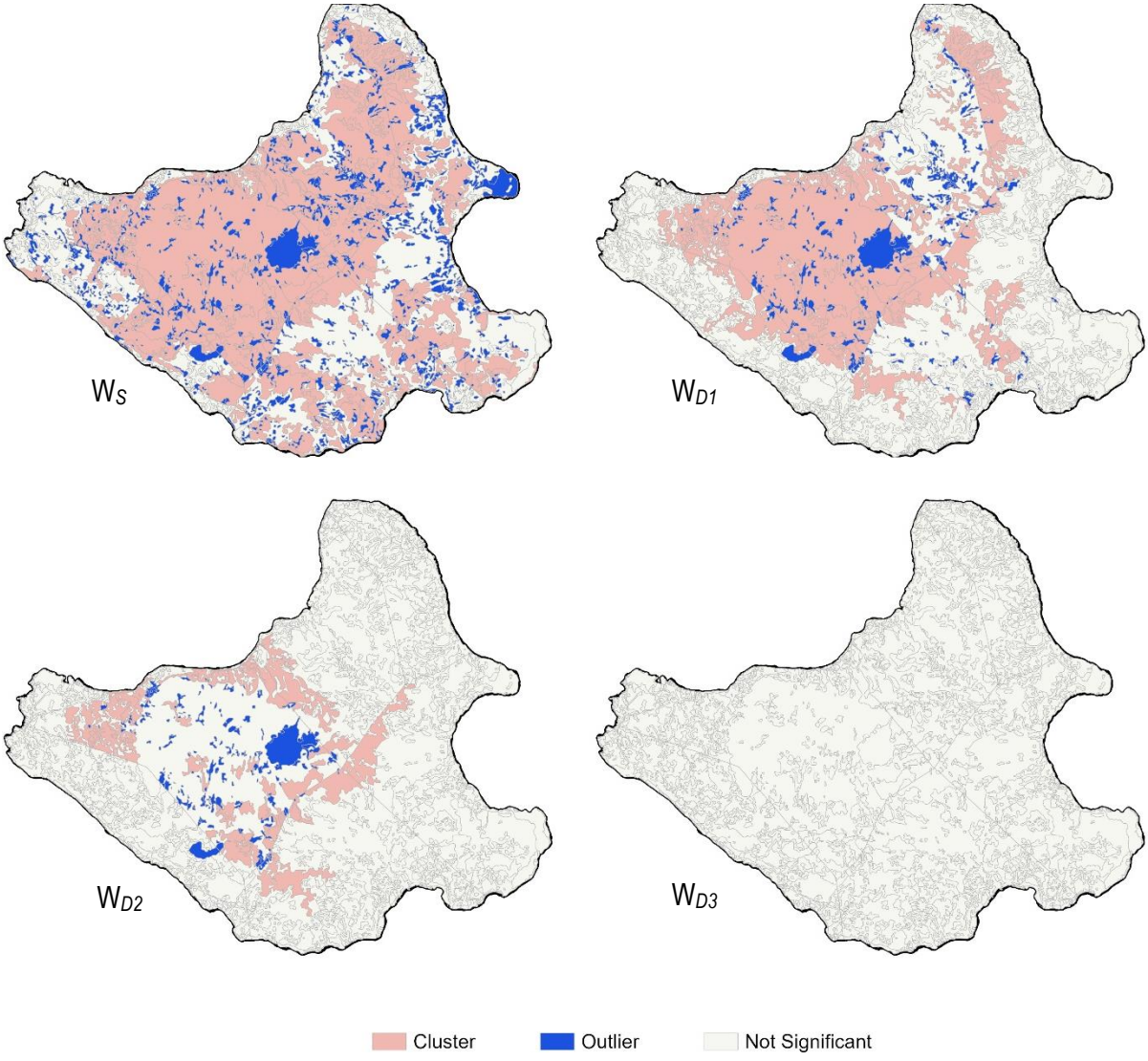


Figure 25. Local Moran's *I* clusters and outliers for raw counts from the binary dasymetric dataset.

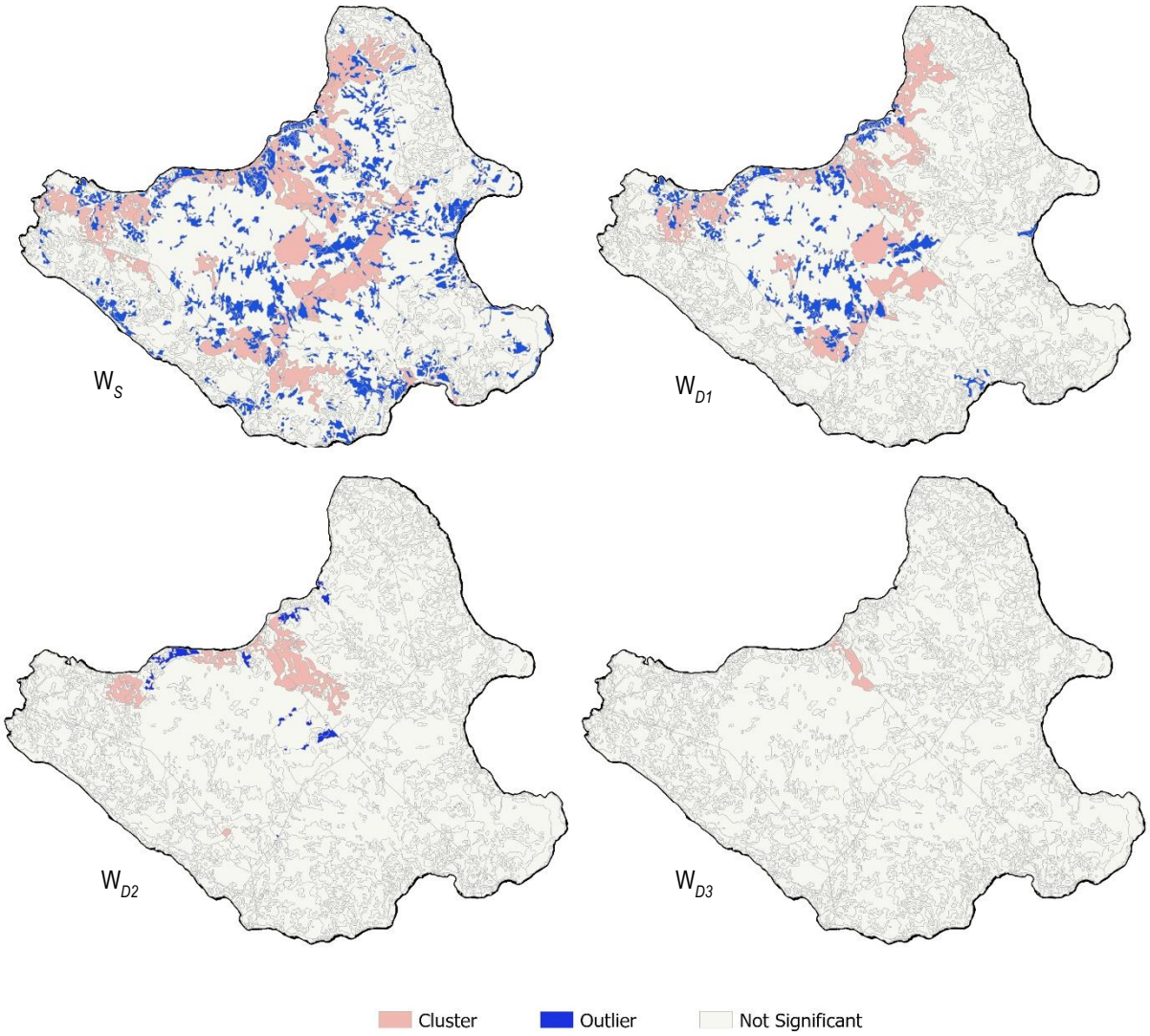


Figure 26. Local Moran's *I* clusters and outliers for raw counts from the three-class dasymetric dataset.

*5.3.4. Getis-Ord G_i**

For the same purpose of determining the types of the clusters detected in the local Moran's *I* analysis, the Getis-Ord G_i* *z*-score is also computed to determine whether the spatial clusters detected are either hotspots, which are clusters of high values; or cold spots, which are of low values. Figure 27 summarizes the number of identified hotspots using the said index per set of input combinations. There are no graphs created for reporting the number of cold spots simply because no cold spots are detected at all in any of the input combinations used. This means that based on the Getis-Ord G_i* index, the observed localized spatial clustering in the study area is purely characterized by clustering of similarly high-valued neighboring zones. The familiar decreasing trend is again seen in the number of hotspots when shifting from the W_S matrix until they dissipate into complete spatial randomness when using W_{D3}.

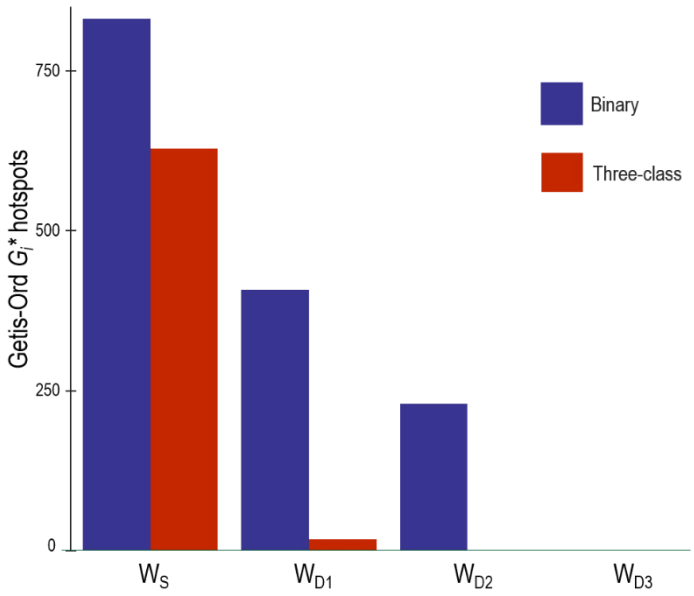


Figure 27. Getis-Ord G_i* hotspots for raw counts by spatial weight matrix.

The number of hotspots and cold spots based on the local Moran's *I* values are tabulated per input combination in Table 9 found in the Appendix.

A multi-map view similar to what is presented previously is also shown in Figure 28 to illustrate the distribution of the hotspots. Although the Getis-Ord G_i* index and local Moran's *I* index are different, both maps show the same spatial pattern of decreasing number of clusters when using the revised spatial weight matrices. The maps also again exhibit higher degrees of spatial autocorrelation for the binary than in the three-class dasymetric dataset, especially along the choropleth zone edges (cf. Figure 17, which shows Siquijor Island's municipal boundaries as choropleth zones).

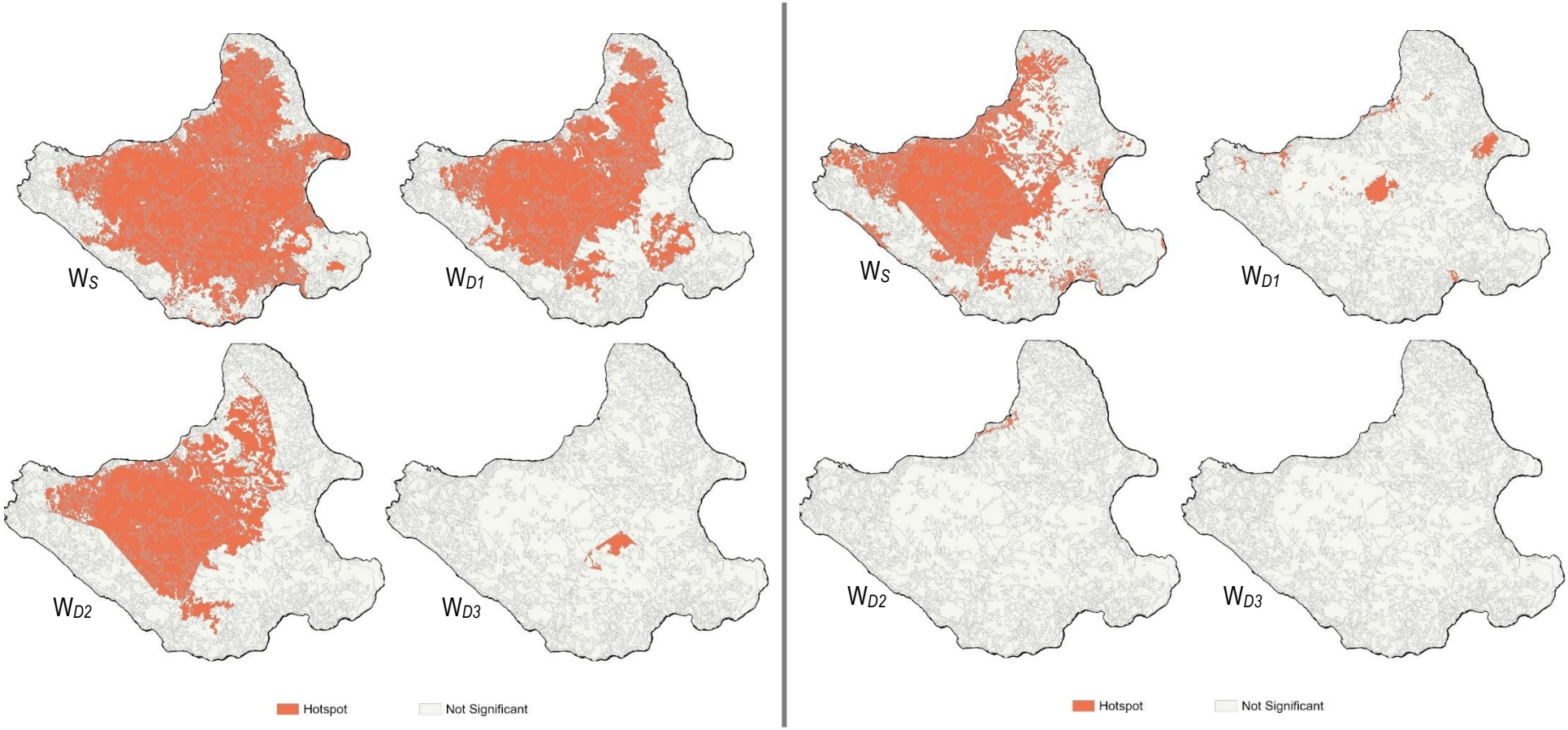


Figure 28. Getis-Ord G_i^* hotspots for raw counts from the binary (left) and three-class (right) dasymetric datasets.

5.4. Spatial autocorrelation analysis of densities

5.4.1. Global Moran's I

Figure 29 summarizes the computed global Moran's *I* *z*-scores for population density of registered voters using the four spatial weight matrices. In contrast to the decreasing trend seen in the spatial autocorrelation analysis of raw counts, the global Moran's *I* *z*-scores for the analysis of densities show a steep and plateauing increase in spatial autocorrelation when shifting from the W_S to the W_D spatial weight matrices. This trend suggests an intensified and sustained overall spatial clustering in the density values.

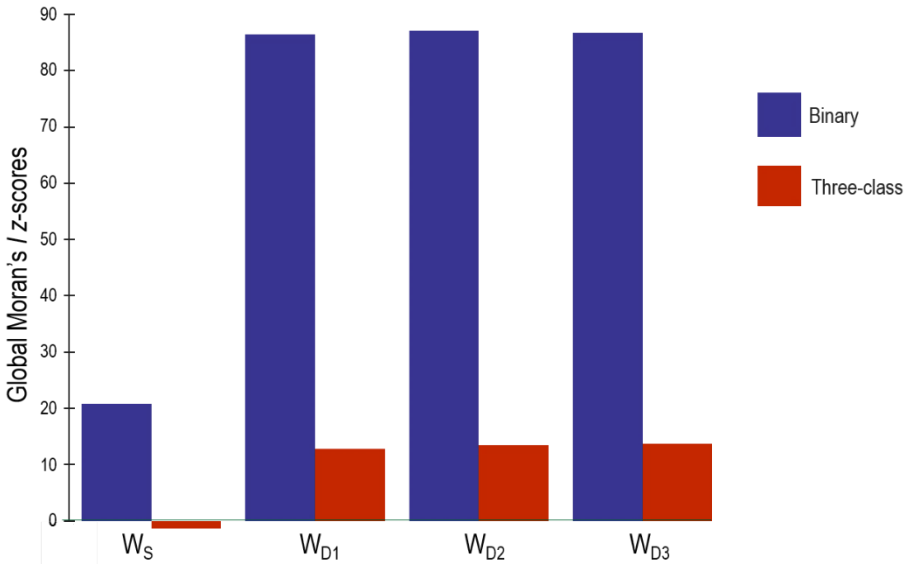


Figure 29. Global Moran's *I* *z*-scores for densities by spatial weight matrix.

The complete set of computed global Moran's *I* values and *z*-scores are organized and tabulated per input combination in Table 10 found in the Appendix.

5.4.2. Geary's C

Figure 30 summarizes the computed Geary's *C* *z*-scores for the same set of input combinations. Unlike for the global Moran's *I* measure, the computed *z*-scores for the Geary's *C* index are all negative. However, the absolute values of the *z*-scores for the said dataset are all increasing and plateauing as well, again indicating a sustained intensification in global spatial clustering when the revised spatial weight matrices are used. This observation is still consistent with the global spatial autocorrelation of the density variable using the global Moran's *I* index. Furthermore, when the two dasymetric datasets

are compared, only the binary dataset exhibited statistically significant results, leaving the three-class results negligible since its corresponding *z*-scores only yield statistically insignificant *p*-values.

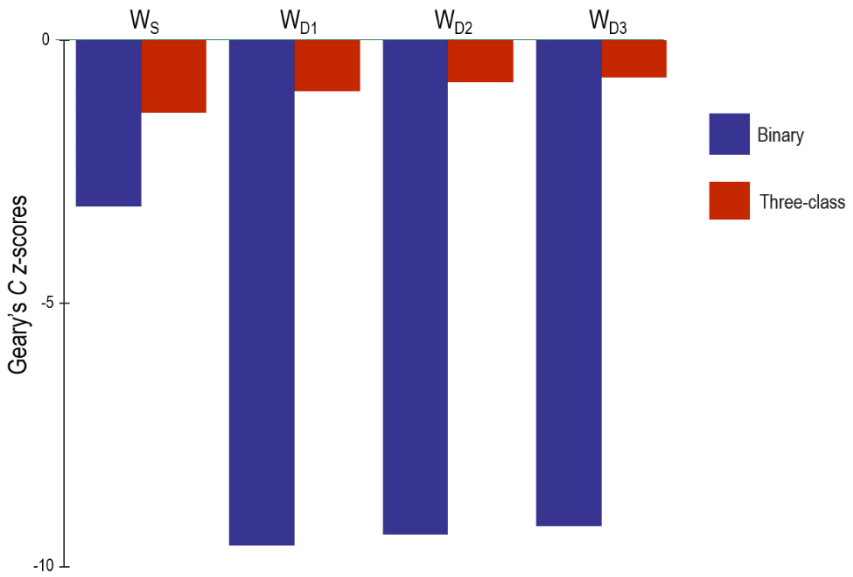


Figure 30. Geary's *C* *z*-scores for densities by spatial weight matrix.

The complete set of computed Geary's *C* values are organized and tabulated per input combination in Table 11 found in the Appendix.

5.4.3. Local Moran's I

Figure 31 shows the number of dasymetric zones identified as statistically significant spatial clusters (left) and outliers (right) in the population densities using the local Moran's *I* index. The legend in Figure 30 still applies to this graph.

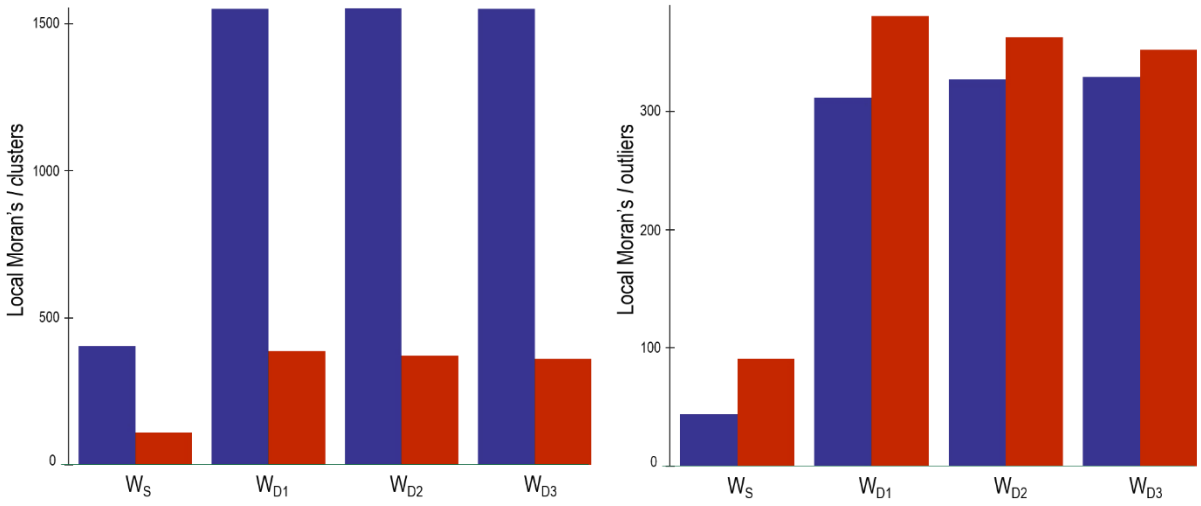


Figure 31. Local Moran's *I* clusters and outliers for densities by spatial weight matrix.

Just like the trend already seen previously in the global spatial autocorrelation, the graphs show a steep and plateauing increase in number of localized clusters and outliers when shifting from the W_S to the three W_D spatial weight matrices. Also, unlike in the raw counts where there is a clear progression in the decrease of values, the changes in the number of spatial clusters and outliers for the densities plateaus for both the binary and the three-class dasymetric datasets when using the revised spatial weight matrices. In terms of the dasymetric dataset type, the difference in counts is more pronounced for the spatial clusters than for the spatial outliers.

The number of spatial clusters and outliers based on the local Moran's *I* values are tabulated per input combination in Table 12 in the Appendix. The map view shown in Figure 32 and Figure 33 again reinforces the increased number of localized clusters and outliers when using the revised spatial weight matrices.

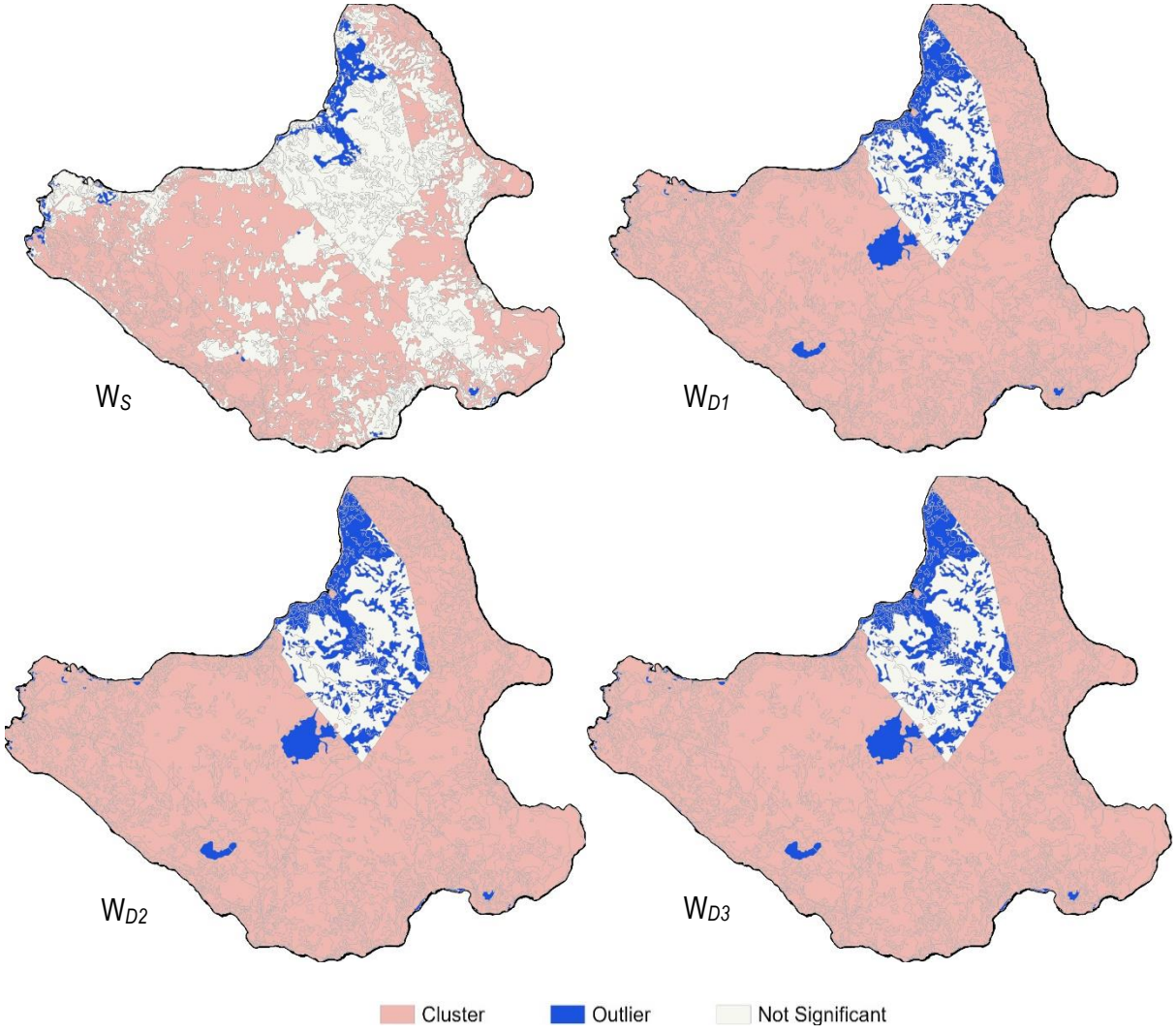


Figure 32. Local Moran's *I* clusters and outliers for densities from the binary dasymetric dataset.

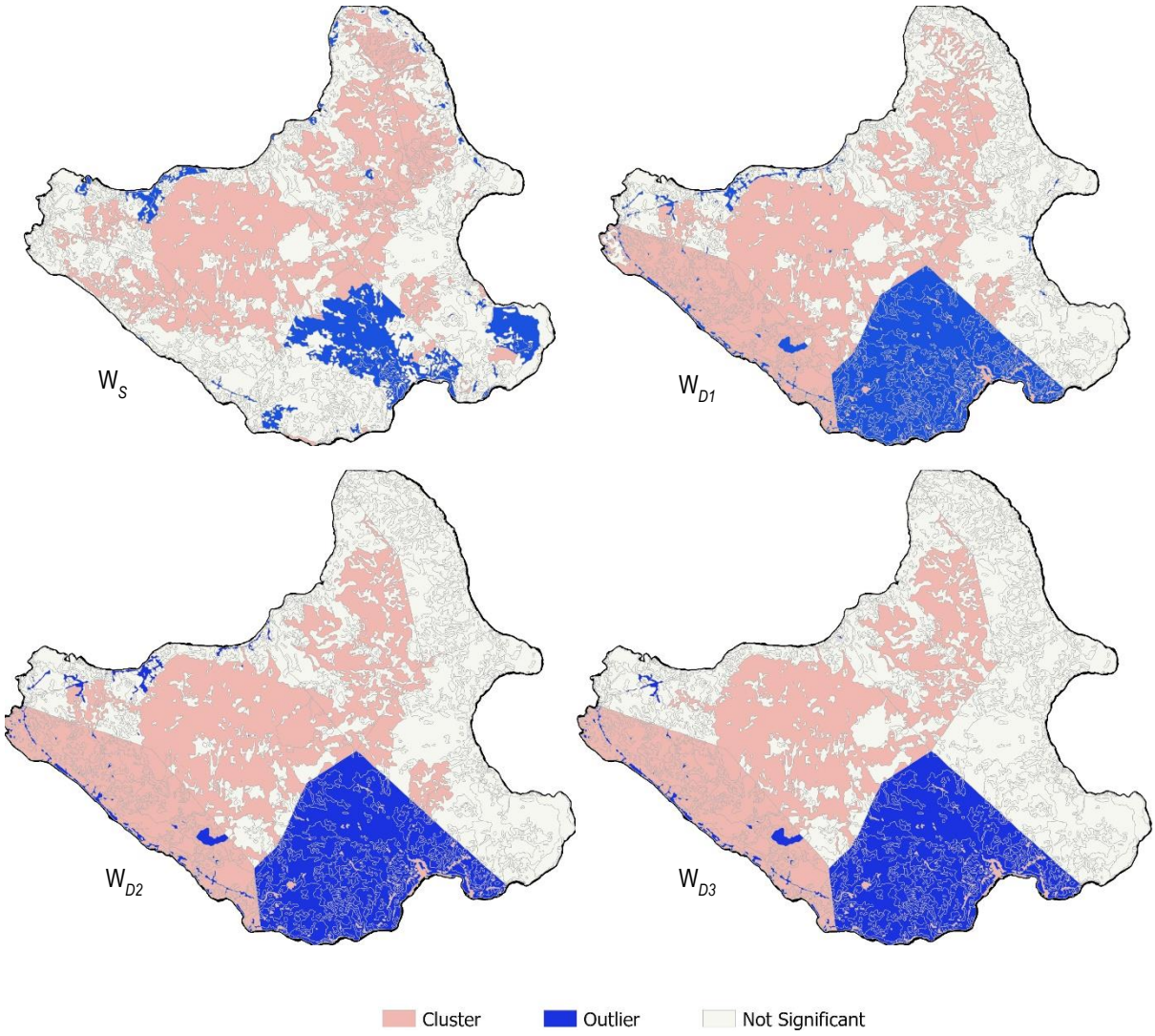


Figure 33. Local Moran's *I* clusters and outliers for raw counts from the three-class dasymetric dataset.

*5.4.4. Getis-Ord G_i**

Figure 34 shows the number of identified hotspots and cold spots per input combination using the Getis-Ord G_i* index. It can be recalled from the spatial autocorrelation analysis of raw counts that no cold spots are detected at all in any of the input combinations used. Here however, a considerable number of cold spots are being detected for the voters' density of the dasymetric zones. More importantly, the increasing trend is again reflected here for the number of hotspots and cold spots detected when shifting from the W_S matrix.

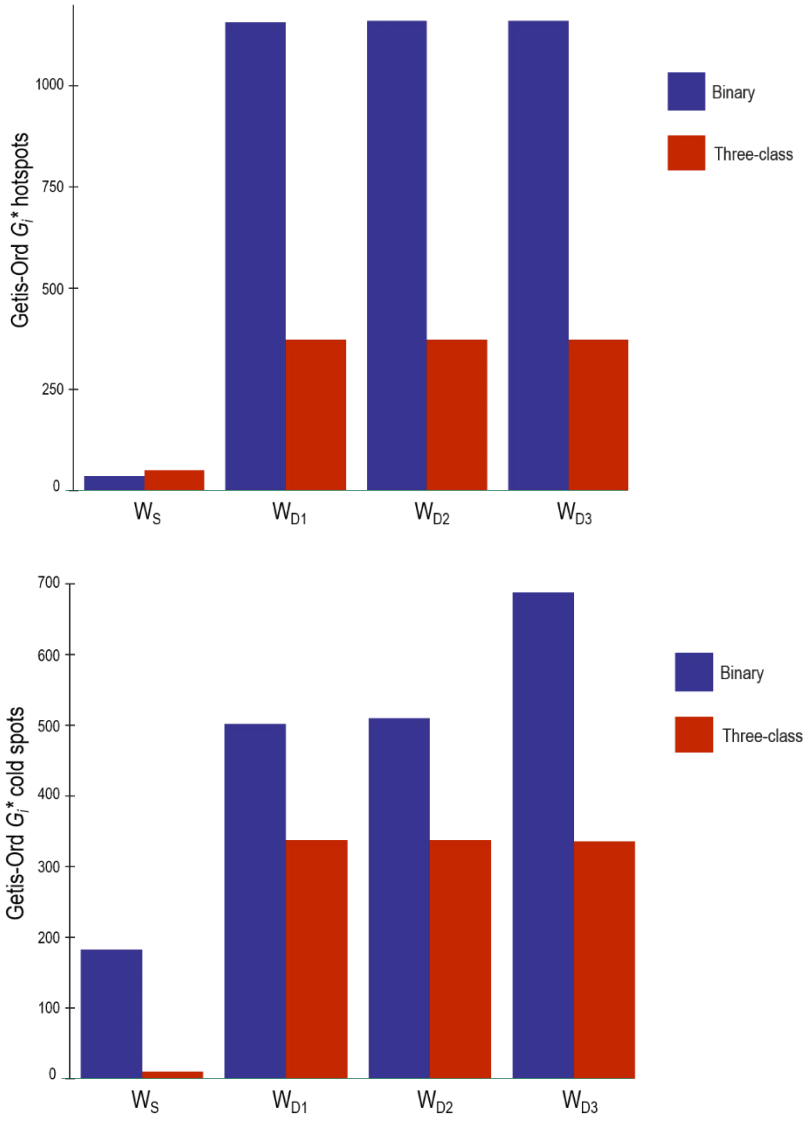


Figure 34. Getis-Ord G_i* hotspots and cold spots for densities by spatial weight matrix.

The number of hotspots and cold spots based on the local Moran's *I* values are tabulated per input combination in Table 13 found in the Appendix.

The map view illustrating the spatial distribution of the hotspots and cold spots is also shown in Figure 35. Although the Getis-Ord G_i* index and local Moran's *I* index are different, both maps show the same spatial pattern of increasing number of clusters when using the revised spatial weight matrices. The maps also again exhibit higher degrees of spatial autocorrelation for the binary than in the three-class dasymetric dataset.

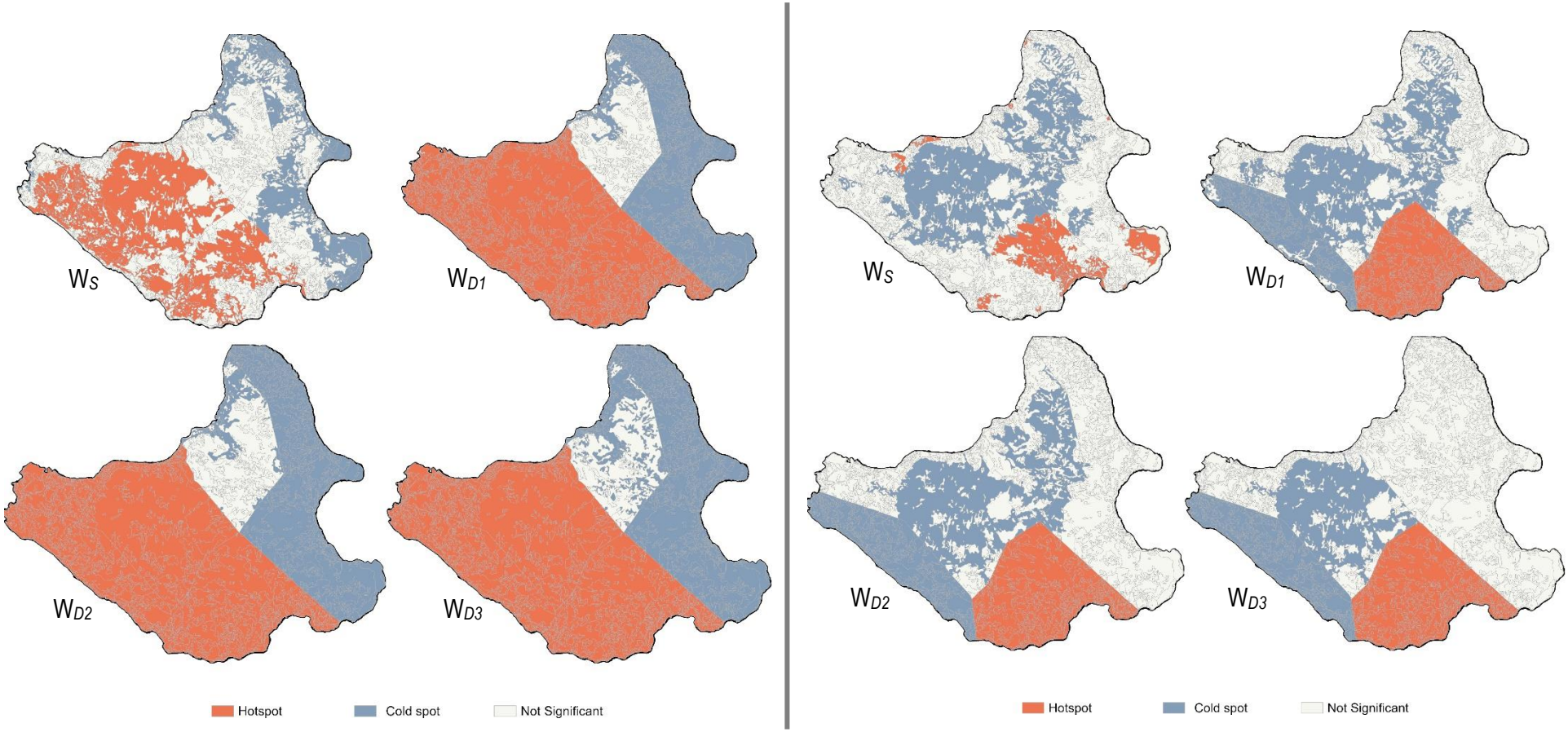


Figure 35. Getis-Ord G_i^* hotspots for densities from the binary (left) and three-class (right) dasymetric datasets.

6. Discussion

Based on the results of the literature review and the case study that was conducted, the research questions initially slated in Chapter 1 are finally addressed and answered below.

1) *Which applicable spatial autocorrelation measures can possibly utilize dasymetrically disaggregated spatial data?*

The related literature points to four commonly used spatial autocorrelation measures in the spatial autocorrelation analysis of dasymetrically disaggregated spatial data. Two of these are for assessing global spatial autocorrelation: global Moran's *I*, and Geary's *C*; and two are for assessing local spatial autocorrelation: local Moran's *I*, and Getis-Ord G_i^*. The decision on which measure to use mainly depends on two factors. The first factor is the extent by which the spatial segregation of the variable must be characterized, whether it is overall (thus, global) or granular (thus, local); and second, which is especially true for the local spatial autocorrelation, is the type of spatial segregation pattern(s) of interest, such as clusters, outliers, hotspots, and cold spots.

2) *Which parameter(s) in the mathematical computation of spatial autocorrelation measures can possibly be modified when using dasymetrically disaggregated spatial data?*

The computational formulae for the four spatial autocorrelation measures cited above commonly share a fundamental parameter which is the spatial weight matrix. Mathematically denoted as W in the formulae, the spatial weight matrix serves the important function of defining subsets in the input area units whose combined values are assessed to gauge whether the data exhibits positive, negative, or zero spatial autocorrelation. Because of this function, the modification of this parameter is seen to influence the spatial autocorrelation analysis of dasymetrically disaggregated spatial data since the input dasymetric zones can properly be organized into neighborhoods based on their spatial configuration and, as introduced in this study, on their choroplethic configuration.

3) *What modification(s) in the mathematical computation parameters of spatial autocorrelation measures can be made when using dasymetrically disaggregated spatial data?*

The spatial weight matrix can be modified and refined through the use of a priority-based spatial weight assignment method, where the spatial weight value of the dasymetric zones is jointly based on their spatial and choroplethic configurations, as well as the relative weighting of these two

configurations. In particular, dasymetric zone pairs are given greater spatial weight values when they are treated as neighbors based simultaneously on their spatial and choroplethic configurations, compared to those which are merely spatial neighbors or sharing common choroplethic origins. Furthermore, aside from the intrinsic configuration of the dasymetric zones, the spatial weight value is also determined by the relative weighting set for the two configurations. In general, the weights can be set by disproportionately increasing (or decreasing) the weight of the spatial configuration over the choroplethic configuration, or vice versa. For the case study, three types of modifications are tested and compared. These are 1) higher priority given to spatial over choroplethic configuration; 2) equal priority given to spatial and choroplethic configurations; and 3) higher priority given to choroplethic over spatial configuration.

4) *How do these modifications differ from each other in terms of results returned in the spatial autocorrelation analysis?*

On the basis of the results derived from the case study, the modifications described above is shown to possibly return contrasting effects in the spatial autocorrelation analysis results, depending on the input spatial variable used from the dasymetrically disaggregated data. When raw counts are analyzed, results show that an increasing priority in favor of the choroplethic configuration corresponds to decreasing levels of detectable spatial clustering or dispersion, reflected both on the spatial autocorrelation index's value, and the number of statistically significant local spatial clusters and outliers. In other words, setting a higher priority for choroplethic configurations results in the dampening of the spatial autocorrelation present in the data towards complete spatial randomness. On the other hand, when densities are analyzed instead, the same increasing priority in the choroplethic configuration contrarily yields increasing levels of detected spatial clustering. This manifests as increasing absolute values in the global spatial autocorrelation indices, and the steep increase and abrupt plateauing in the number of statistically significant local spatial clusters and outliers detected in the data.

5) *How do the results of the modified spatial autocorrelation analysis differ from those of the original spatial autocorrelation analysis measures?*

The modified spatial autocorrelation analysis presented in this study differs from the original method in two aspects. The first aspect is on the identification of neighboring area units among the

dasymetric zones, where an increase in the degree of neighborhood in the data is attained when the modified spatial autocorrelation analysis is implemented. The second and more prominent aspect is on the degree of detected spatial segregation. In comparison to just using the spatial configuration, using the revised spatial matrix is seen to either diminish or intensify the level of spatial clustering or dispersion detected in the variable. In other words, the introduction of the choroplethic configuration in the modified spatial autocorrelation analysis differs from the original method by applying either a filtering effect or an amplifying effect on the spatial autocorrelation analysis of dasymetrically disaggregated spatial data, depending on the type of variable being analyzed.

Certain aspects of the differences noted above remarkably coincides with the findings of Rodrigues & Tenedório (2016), where the original spatial autocorrelation analysis method is used. They found in their study that the level of detected spatial autocorrelation in population, a raw count variable, rises when its less disaggregated geometric configuration is considered, and alternately decreases when its aggregated form is used instead. This finding is consistent with this case study's where it is shown that changes in measured spatial autocorrelation is influenced by whether the aggregated or the disaggregated versions of the spatial datasets is used in spatial autocorrelation analysis. Another aspect in their finding that is deemed consistent with this study is the observed decrease in spatial autocorrelation as the number of neighbors assigned per area unit increases, particularly when using the nearest neighbors spatial scheme. As mentioned previously, one effect of introducing the choroplethic configuration in the spatial weight matrix is the increase in the attained level of neighborhood among the dasymetric zones. This increase in number of neighboring area unit pairs then spontaneously results in decreasing levels of spatial autocorrelation in the case of raw counts, describing a recognizable consistency between the two studies' findings. These consistencies found between Rodrigues & Tenedório's (2016) and this study recapitulate the direct effect of modifying neighborhood configuration on spatial autocorrelation analysis, further bolstering the need of using a choroplethic-configuration-based spatial weight matrix when analyzing dasymetrically disaggregated spatial datasets.

7. Conclusion

This study positively demonstrated that spatial autocorrelation analysis can be further refined when using dasymetrically disaggregated spatial data as input area units, producing substantially different results when compared to the original analysis method. The refinement in the analysis is done by using a revised spatial weight assignment method where the neighborhood configuration of the input area units is dually defined by the spatial configuration, as well as the choroplethic configuration of the dasymetric zones. This is mathematically implemented through a linear combination of two spatial weight matrices, one being the conventional spatial weight matrix based on a spatial neighborhood scheme, and another being a spatial weight matrix based on similarity in choroplethic origins. The refined analysis method is tested through a case study using population data for Siquijor Island's registered voters for the election year 2016. Drawing from the evidences found in the case study results, three conclusions can be made regarding this proposed method for the spatial autocorrelation analysis of dasymetrically disaggregated spatial data.

First, this study confirms that all previously used global and local spatial autocorrelation measures in the analysis of dasymetrically disaggregated data can be refined in the manner described here since they all use spatial weight matrices as a fundamental parameter in their respective computational formulae. This means that the only limiting factor in order to apply the refined spatial autocorrelation analysis proposed here is the type of the input spatial data involved, which must be dasymetrically disaggregated in nature.

Secondly, this study proves that substantial differences can be observed in the analysis output when applying the refined spatial autocorrelation analysis proposed here. Specifically, the revised spatial weight matrix produces two distinct effects depending on the variable in the spatial data being analyzed. When dealing with raw counts, a filtered version of spatial autocorrelation analysis result is produced when using the revised spatial weight matrix, resulting in lowered levels of detected spatial autocorrelation in the output. Conversely, when dealing with densities, an opposite effect is seen, where an intensified version of spatial autocorrelation analysis results is returned, resulting in increased levels of detected spatial autocorrelation levels either in the form of higher levels of global spatial clustering, or greater occurrences of local spatial clusters and outliers.

Finally, since the revised spatial weight matrix is dually composed of the spatial and the choroplethic neighborhood configurations of the dasymetrically disaggregated spatial data, the relative weighting or priority between the spatial and the choroplethic configuration can be variably defined. However, as manifested in the results, it is considered to be optimal for the refined spatial autocorrelation analysis if the relative weights are set in such a way that the dasymetric zones' spatial configuration is either equally, or more prioritized than the dasymetric zones' choroplethic configuration. This relative weighting approach is recommended here in order to avoid dampening the spatial autocorrelation present in the data into complete spatial randomness when analyzing raw counts; or exaggerating the presence of spatial clustering and outliers when analyzing raw counts.

Motivated by the findings produced in this study, follow-up research related to spatial autocorrelation analysis and dasymetric mapping may be undertaken in the areas of spatial statistics, cartography, demography, and other related domains. Case studies similar to those exhibited in this study's literature review can test the usability of the refined spatial autocorrelation analysis proposed here in various real-world applications, such as in crime analysis, epidemiology, and business analytics among others. Moreover, it will also be worthy to determine if the analysis results observed here can be consistently replicated with other dasymetrically disaggregated datasets. In the occasion that such differences are observed, further analysis can then be conducted to determine factors or properties in the datasets that can contribute to such differences.

Aside from the direct application of the technique, studies aimed at further exploring and developing this novel method may also be done. Specific topics include the detailed comparison of the effects of the revised spatial weight matrix on different spatial neighborhood schemes; using the revised spatial weight assignment method in other spatiostatistical techniques, such as spatial crossregression analysis and spatial autoregressive modeling; and the combining of different spatial weight matrices using the linear combination technique demonstrated in this study. Because of the fundamental role of spatial weight matrices in these topics, this study opens up prospects for further inquiry and discussion in the areas of spatial statistics, cartography, demography, as well as other geosciences in general where the statistical analysis of spatial segregation is a topic of interest. ■

8. References

Anselin, L. (1995). Local indicators of spatial association—LISA. *Geographical Analysis*, *27*(2), 93–115.

Anselin, L. (2018a, March 16). *Contiguity-Based Spatial Weights*. GeoDa: An Introduction to Spatial Data Analysis. http://geodacenter.github.io/workbook/4a_contig_weights/lab4a.html

Anselin, L. (2018b, March 16). *Distance-Band Spatial Weights*. GeoDa: An Introduction to Spatial Data Analysis. geodacenter.github.io/workbook/4b_dist_weights/lab4b.html

Bajat, B., Krunić, N., Bojović, M., Kilibarda, M., & Kovačević, Z. (2011). *Population vulnerability assessment in hazard risk management: A dasymetric mapping approach*. 2nd Project Workshop on “Risk Identification and Land-Use Planning for Disaster Mitigation of Landslides and Floods”, Rijeka, Croatia.

Barrozo, L. V., Machado, R. P. P., Luchiari, A., & de Queiroz Filho, A. P. (2011). *Dasymetric mapping of socioeconomic data of the city of Sao Paulo: First approach*. Conference Proceedings. Regional Geographic Conference-International Geographic Union.

Batista e Silva, F., Gallego, J., & Lavalle, C. (2013). A high-resolution population grid map for Europe. *Journal of Maps*, *9*(1), 16–28.

Bielecka, E. (2005). *A dasymetric population density map of Poland*. *48*(22), 9–15.

Bivand, R. S., & Wong, D. W. S. (2018). Comparing implementations of global and local indicators of spatial association. *TEST*, *27*(3), 716–748. https://doi.org/10.1007/s11749-018-0599-x

Boo, G., Fabrikant, S. I., & Leyk, S. (2015). A Novel Approach to Veterinary Spatial Epidemiology: Dasymetric Refinement of the Swiss Dog Tumor Registry Data. *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci.*, *II-3/W5*, 263–269. https://doi.org/10.5194/isprsannals-II-3-W5-263-2015

Brehme, C. E., McCarron, P., & Tetreault, H. (2015). A dasymetric map of Maine lobster trap distribution using local knowledge. *The Professional Geographer*, *67*(1), 98–109.

Choi, D. J., Kim, Y.-S., & Suh, Y.-C. (2011). A Study on the Estimation and Distribution Pattern of Longevity Population in Small Area Using Vector-based Dash Matrix Method. *Korean Journal of Geomatics*, *29*. https://doi.org/10.7848/KSGPC.2011.29.5.479

Cleckner, H., & Allen, T. R. (2014). Dasymetric mapping and spatial modeling of mosquito vector exposure, Chesapeake, Virginia, USA. *ISPRS International Journal of Geo-Information*, *3*(3), 891–913.

Cliff, A. D., & Ord, K. (1970). Spatial autocorrelation: A review of existing and new measures with applications. *Economic Geography*, *46*(sup1), 269–292.

Cockx, K., & Canters, F. (2015). Incorporating spatial non-stationarity to improve dasymetric mapping of population. *Applied Geography*, *63*, 220–230.

Comber, A., Proctor, C., & Anthony, S. (2008). The creation of a national agricultural land use dataset: Combining pycnophylactic interpolation with dasymetric mapping techniques. *Transactions in GIS*, *12*(6), 775–791.

Crampton, J. W. (2011). Rethinking maps and identity: Choropleths, clines, and biopolitics. In *Rethinking Maps* (pp. 44–67). Routledge.

Dizon, D. (2018). *The Spatial Dimension of Electoral Participation in the Philippines* [Unpublished Master's Thesis]. University of the Philippines-Diliman.

Dorling, D. (1993). Map design for census mapping. *The Cartographic Journal*, *30*(2), 167–183. https://doi.org/10.1179/000870493787860175

Eicher, C. L., & Brewer, C. A. (2001). Dasymetric mapping and areal interpolation: Implementation and evaluation. *Cartography and Geographic Information Science*, *28*(2), 125–138.

Esri, Inc. (2018a, February). *Global Moran's I Additional Math*. ArcGIS Pro Tool Reference. https://pro.arcgis.com/en/pro-app/tool-reference/spatial-statistics/h-global-morans-i-additional-math.htm

Esri, Inc. (2018b, February). *How Cluster and Outlier Analysis (Anselin Local Moran's I) works*. ArcGIS Pro Tool Reference. https://pro.arcgis.com/en/pro-app/tool-reference/spatial-statistics/h-how-cluster-and-outlier-analysis-anselin-local-m.htm

Esri, Inc. (2018c, February). *How Hot Spot Analysis (Getis-Ord Gi*) works*. ArcGIS Pro Tool Reference. https://pro.arcgis.com/en/pro-app/tool-reference/spatial-statistics/h-how-hot-spot-analysis-getis-ord-gi-spatial-stati.htm

Esri, Inc. (2018d, February). *How Spatial Autocorrelation (Global Moran's I) works*. ArcGIS Pro Tool Reference. pro.arcgis.com/en/pro-app/tool-reference/spatial-statistics/h-how-spatial-autocorrelation-moran-s-i-spatial-st.htm

Esri, Inc. (2018e, February). *Modeling spatial relationships*. ArcGIS Pro Tool Reference. http://pro.arcgis.com/en/pro-app/tool-reference/spatial-statistics/modeling-spatial-relationships.htm

Garcia, R. A., Oliveira, S. C., & Zêzere, J. L. (2016). Assessing population exposure for landslide risk analysis using dasymetric cartography. *Natural Hazards and Earth System Sciences*, *16*(12), 2769.

Geary, R. C. (1954). The contiguity ratio and statistical mapping. *The Incorporated Statistician*, *5*(3), 115–146.

Getis, A., & Aldstadt, J. (2004). Constructing the spatial weights matrix using a local statistic. *Geographical Analysis*, *36*(2), 90–104.

Getis, A., & Ord, J. K. (2010). The analysis of spatial association by use of distance statistics. In *Perspectives on spatial data analysis* (pp. 127–145). Springer.

Hu, Z., Liebens, J., & Rao, R. (2007). Exploring Relationship Between Asthma and Air Pollution: A Geospatial Methodology Using Dasymetric Mapping, GIS Analysis, and Spatial Statistics. *Proceedings: Geoinformatics 2007*, *6753*. https://doi.org/10.1117/12.761898

Jia, P., & Gaughan, A. E. (2016). Dasymetric modeling: A hybrid approach using land cover and tax parcel data for mapping population in Alachua County, Florida. *Applied Geography*, *66*, 100–108.

Kim, H., & Choi, J. (2011). A hybrid dasymetric mapping for population density surface using remote sensing data. *Journal of the Korean Geographical Society*, *46*(1), 67–80.

Krunić, N., Bajat, B., & Kilibarda, M. (2015). Dasymetric mapping of population distribution in Serbia based on soil sealing degrees layer. In *Surface models for geosciences* (pp. 137–149). Springer.

Langford, M., & Unwin, D. J. (1994). Generating and mapping population density surfaces within a geographical information system. *The Cartographic Journal*, *31*(1), 21–26. https://doi.org/10.1179/000870494787073718

Langford, Mitchel. (2006). Obtaining population estimates in non-census reporting zones: An evaluation of the 3-class dasymetric method. *Computers, Environment and Urban Systems*, *30*(2), 161–180.

Larsson, J. (2020, March 9). *Area-Proportional Euler and Venn Diagrams with Ellipses*. RDocumentation. https://www.rdocumentation.org/packages/eulerr/versions/6.1.0

Lembo, A. (n.d.). *Spatial Autocorrelation*. Retrieved June 17, 2020, from http://faculty.salisbury.edu/~ajlembo/419/lecture10.pdf

Leyk, S., Nagle, N. N., & Buttenfield, B. P. (2013). Maximum entropy dasymetric modeling for demographic small area estimation. *Geographical Analysis*, *45*(3), 285–306.

Leyk, S., Ruther, M., Buttenfield, B. P., Nagle, N. N., & Stum, A. K. (2013). Identifying Residential Land in Rural Areas to Improve Dasymetric Mapping. *Proceedings of the 26th International Cartographic Conference*.

Lial, M. L., Hornsby, J., Schneider, D. I., & Daniels, C. J. (2013). *College Algebra*. Pearson New York, NY.

Maantay, J. A., Maroko, A. R., & Porter-Morgan, H. (2008). Research note—A new method for mapping population and understanding the spatial dynamics of disease in urban areas: Asthma in the Bronx, New York. *Urban Geography*, *29*(7), 724–738.

Mans, G. (2011). *Developing a geo-data frame using dasymetric mapping principles to facilitate data integration*. AfriGEO Conference: Developing Geomatics for Africa.

Mennis, J. (2009). Dasymetric mapping for estimating population in small areas. *Geography Compass*, *3*(2), 727–745.

Mennis, J., & Hultgren, T. (2005). *Dasymetric mapping for disaggregating coarse resolution population data*. 9–16.

Mitchel, A. (2005). The ESRI Guide to GIS analysis, Volume 2: Spartial measurements and statistics. *ESRI Guide to GIS Analysis*.

Mitsova, D., Esnard, A.-M., & Li, Y. (2012). Using enhanced dasymetric mapping techniques to improve the spatial accuracy of sea level rise vulnerability assessments. *Journal of Coastal Conservation*, *16*(3), 355–372.

Mosley, B. (2012). *Estimating Health Outcomes and Determinants in Rural Ottawa: An Integration of Geographical and Statistical Techniques* [Thesis, University of Ottawa]. http://dx.doi.org/10.20381/ruor-6198

Parenteau, M.-P., & Sawada, M. C. (2012). The Role of Spatial Representation in the Development of a LUR Model for Ottawa, Canada. *Air Quality, Atmosphere & Health*, *5*(3), 311–323. https://doi.org/10.1007/s11869-010-0094-3

Pavía, J. M., & Cantarino, I. (2017). Can dasymetric mapping significantly improve population data reallocation in a dense urban area? *Geographical Analysis*, *49*(2), 155–174.

Petrov, A. (2012). One Hundred Years of Dasymetric Mapping: Back to the Origin. *The Cartographic Journal*, *49*(3), 256–264. https://doi.org/10.1179/1743277412Y.0000000001

Poulsen, E., & Kennedy, L. W. (2004). Using dasymetric mapping for spatially aggregated crime data. *Journal of Quantitative Criminology*, *20*(3), 243–262.

Reynolds, D. C. (2011). *Application of Objective Visual Complexity Measures to Binary Dasymetric Maps* [Thesis, University of Wisconsin–Madison]. http://digital.library.wisc.edu/1793/67901

Rodrigues, A. M., & Tenedório, J. A. (2016). Sensitivity Analysis of Spatial Autocorrelation Using Distinct Geometrical Settings: Guidelines for the Quantitative Geographer. *International Journal of Agricultural and Environmental Information Systems (IJAEIS)*, *7*(1), 65–77. https://doi.org/10.4018/IJAEIS.2016010105

Sleeter, R. (2004). *Dasymetric mapping techniques for the San Francisco Bay region, California*. 7–10.

Strode, G., & Mesev, V. (2013). Cadastral-Based Expert Dasymetric System (CEDS) Using Census and Parcel Data. *Proceedings of the 26th International Cartographic Conference*.

Tobler, W. R. (1979). Cellular geography. In *Philosophy in geography* (pp. 379–386). Springer.

Unwin, A. (1996). Geary's contiguity ratio. *Economic and Social Review*, *27*(2), 145–159.

Weeks, J. R. (2010). Spatial Patterns of Fertility in Rural Egypt. In L. Anselin & S. J. Rey (Eds.), *Perspectives on Spatial Data Analysis* (pp. 235–256). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-01976-0_17

Zandbergen, P. A., & Ignizio, D. A. (2010). Comparison of dasymetric mapping techniques for small-area population estimates. *Cartography and Geographic Information Science*, *37*(3), 199–214.

Zhou, X., & Lin, H. (2008). Geary's C. In S. Shekhar & H. Xiong (Eds.), *Encyclopedia of GIS* (pp. 329–330). Springer US. https://doi.org/10.1007/978-0-387-35973-1_446

9. Appendix

9.1. Geary's *C* and Global Moran's *I* computational specifications

The geoprocessing tool in ArcGIS named *Spatial Autocorrelation (Global Moran's I)* is a ready-to-use, Python-script-based tool used to compute Global Moran's *I* indices for a given spatial attribute stored in an area-based spatial dataset. This script-based tool is copied and modified in this study to create a separate script-based tool for computing Geary's *C* indices within ArcGIS.

The script for the *Spatial Autocorrelation (Global Moran's I)* geoprocessing tool can be accessed in the following ArcGIS system directory, together with the other out-of-the-box script-based geoprocessing tools that come with the software.

- For ArcMap:

  C:\Program Files (x86)\ArcGIS\Desktop10.6\ArcToolbox\Scripts

- For ArcGIS Pro:

  C:\Program Files\ArcGIS\Pro\Resources\ArcToolBox\Scripts

The file name of the script is *MoransI.py*.

In order to enable the original *Spatial Autocorrelation (Global Moran's I)* script to compute Geary's *C* indices instead, the following lines of code (lines 309 to 369) in the original script are modified:

```
def processRow(self, orderID, yiDev, nhIDs, nhVals, weights):
        """Processes a features contribution to the Moran's I statistic.

        INPUTS:
        orderID (int): order in corresponding numpy value arrays
        yiVal (float): value for given feature
        nhIDs (array, nn): neighbor order in corresponding numpy value arrays
        nhVals (array, nn): values for neighboring features (1)
        weights (array, nn): weight values for neighboring features (1)

        NOTES:
        (1)  nn is equal to the number of neighboring features
        """

        #### Numerator Calculation ####
        sumW = weights.sum()
        self.s0 += sumW
```

```
        self.numer += NUM.sum(nhVals * weights) * yiDev

        #### Weights Characteristics Update ####
        c = 0
        for neighID in nhIDs:
            ij = (orderID, neighID)
            ji = (neighID, orderID)
            w = weights[c]
            self.s1 += w**2.0
            try:
                self.s1 += 2.0 * w * self.wij.pop(ji)
            except:
                self.wij[ij] = w
            self.rowSum[orderID] += w
            self.colSum[neighID] += w
            c += 1

    def calculate(self):
        """Calculate Moran's I Statistic."""

        s0 = self.s0
        s1 = self.s1
        n = len(self.rowSum) * 1.0
        s2 = NUM.sum( (self.rowSum + self.colSum)**2.0 )
        self.s2 = s2
        self.ei = -1. / (n - 1)
        self.squareExpectedI = self.ei**2
        self.n = n
        scale = n / s0
        s02 = s0 * s0
        n2 = n * n
        n2s1 = n2 * s1
        ns2 = n * s2
        self.gi = (scale * (self.numer/self.denom))
        yDev4Sum = NUM.sum(self.yDev**4) / n
        yDevsqsq = (self.denom / n)**2
        b2 = yDev4Sum / yDevsqsq
        self.b2 = b2
        left = n * ((n2 - (3*n) + 3) * s1 - (n*s2) + 3 * (s02))
        right = b2 * ((n2 - n) * s1 - (2*n*s2) + 6 * (s02))
        denom = (n-1) * (n-2) * (n-3) * s02
        num = (left - right) / denom
        self.expectedSquaredI = num
        self.vi = self.expectedSquaredI - self.squareExpectedI
```

The above lines of code are modified in a separate file copy of the original script as follows:

```
  def processRow(self, orderID, yiVal, nhIDs, nhVals, weights):
          """Processes a features contribution to the Geary's C statistic.

          INPUTS:
          orderID (int): order in corresponding numpy value arrays
          yiVal (float): value for given feature
          nhIDs (array, nn): neighbor order in corresponding numpy value arrays
          nhVals (array, nn): values for neighboring feature (1)
          weights (array, nn): weight values for neighboring features (1)

          NOTES:
          (1)  nn is equal to the number of neighboring features
          """

          #### Numerator Calculation ####
          sumW = weights.sum()
          self.s0 += sumW
          self.numer += NUM.sum(((nhVals - yiVal)**2) * weights)

          #### Weights Charactersitics Update ####
          c = 0
          for neighID in nhIDs:
              ij = (orderID, neighID)
              ji = (neighID, orderID)
              w = weights[c]
              self.s1 += w**2.0
              try:
                  self.s1 += 2.0 * w * self.wij.pop(ji)
              except:
                  self.wij[ij] = w
              self.rowSum[orderID] += w
              self.colSum[neighID] += w
              c += 1

      def calculate(self):
          """Calculate Geary's C Statistic."""

          s0 = self.s0
          s1 = self.s1
          n = len(self.rowSum) * 1.0
          s2 = NUM.sum( (self.rowSum + self.colSum)**2.0 )
          self.s2 = s2
          self.ei = 1.
```

```
        self.squareExpectedI = self.ei**2
        self.n = n
        scale = (n-1) / (2.0 * s0)
        s02 = s0 * s0
        n2 = n * n
        n2s1 = n2 * s1
        ns2 = n * s2
        self.gc = (scale * (self.numer/self.denom))
        yDev4Sum = NUM.sum(self.yDev**4) / n
        yDevsqsq = (self.denom / n)**2
        b2 = yDev4Sum / yDevsqsq
        self.b2 = b2
        left = ((n-1) * s1 * (n2 - (3*n) + 3 - b2 * (n - 1))) / ((n-1) * (n-2) *
(n-3) * s02)
        middle = (n2 - 3 - (b2 * ((n - 1)**2))) / ( n * (n - 2) * (n - 3))
        right = ((n-1) * s2 * (n2 + (3*n) - 6 - b2 * (n2 - n + 2))) / ((4*n) *
(n-2) * (n-3) * s02)
        num = left + middle - right
        self.vi = num
```

The mathematical specifications for the Geary's *C* and global Moran's *I* indices that are used as guide in modifying the script as specified above are provided in Bivand & Wong (2018) and Esri, Inc. (2018a). Aside from the computational modification described above, other modifications to supports proper display of outputs and messaging in the ArcGIS interface are also applied. A copy of the fully modified script is available in the supplementary files provided in this manuscript, together with the associated setup instructions.

9.2. Spatial autocorrelation analysis outputs

Dataset	Binary			Three-class		
Spatial weight matrix	Index	z-score	*p*-value	Index	z-score	*p*-value
W_S	0,292013	24,178506	0,000000	0,056290	3,844595	0,000121
W_{D1}	0,005644	4,878344	0,000001	-0,000499	-0,020487	0,983655
W_{D2}	0,000956	1,208462	0,226870	-0,001429	-0,598079	0,549787
W_{D3}	-0,001100	-0,465083	0,641872	-0,001836	-0,874240	0,381988

Table 6. Computed global Moran's *I* value for raw counts by input combination.

Dataset	Binary			Three-class		
Spatial weight matrix	Index	z-score	*p*-value	Index	z-score	*p*-value
W_S	45,838864	30,414689	0,000000	8,883227	9,951823	0,000000
W_{D1}	1,165298	1,013367	0,310885	1,271933	3,102685	0,001918
W_{D2}	1,177362	1,099669	0,271476	1,147349	1,700296	0,089075
W_{D3}	1,182655	1,130514	0,258260	1,092698	1,067808	0,285607

Table 7. Computed Geary's *C* value for raw counts by input combination.

Dataset	Spatial weight matrix	Clusters	Outliers
Binary	W_S	163	1 188
	W_{D1}	70	371
	W_{D2}	38	190
	W_{D3}	0	0
Three-class	W_S	109	791
	W_{D1}	35	251
	W_{D2}	9	26
	W_{D3}	1	0

Table 8. Number of local Moran's *I* spatial clusters and outliers for raw counts by input combination.

Dataset	Spatial weight matrix	Hotspots	Cold spots
Binary	W_S	831	0
	W_{D1}	408	0
	W_{D2}	229	0
	W_{D3}	1	0
Three-class	W_S	628	0
	W_{D1}	18	0
	W_{D2}	1	0
	W_{D3}	0	0

Table 9. Number of Getis-Ord G_i^* hotspots and cold spots for raw counts by input combination.

Dataset	Binary			Three-class		
Spatial weight matrix	Index	z-score	*p*-value	Index	z-score	*p*-value
W_S	0,544900	-3,152076	0.001621	0,544900	-3,152076	0,001621
W_{D1}	0,846922	-9,588395	0,000000	0,846922	-9,588395	0,000000
W_{D2}	0,851866	-9,385744	0,000000	0,851866	-9,385744	0,000000
W_{D3}	0,854035	-9,232933	0,000000	0,854035	-9,232933	0,000000

Table 10. Computed global Moran's *I* value for densities by input combination.

Dataset	Binary			Three-class		
Spatial weight matrix	Index	z-score	*p*-value	Index	z-score	*p*-value
W_S	0,544900	-3,152076	0.001621	0,806329	-1,376410	0,000000
W_{D1}	2,445148	8,859509	0,000000	1,271938	3,102721	0,001918
W_{D2}	1,734848	4,556142	0,000000	1,147350	1,700313	0,089072
W_{D3}	1,423265	2,619735	0,008800	1,092698	1,067816	0,285604

Table 11. Computed Geary's *C* value for densities by input combination.

Dataset	Spatial weight matrix	Clusters	Outliers
Binary	W_S	403	44
	W_{D1}	1 552	312
	W_{D2}	1 553	327
	W_{D3}	1 550	329
Three-class	W_S	110	91
	W_{D1}	388	381
	W_{D2}	371	363
	W_{D3}	361	352

Table 12. Number of local Moran's *I* spatial clusters and outliers for densities by input combination.

Dataset	Spatial weight matrix	Hotspots	Cold spots
Binary	W_S	36	183
	W_{D1}	1 157	502
	W_{D2}	1 161	510
	W_{D3}	1 161	688
Three-class	W_S	51	10
	W_{D1}	372	338
	W_{D2}	372	338
	W_{D3}	372	336

Table 13. Number of Getis-Ord G_i^* hotspots and cold spots for densities by input combination.