Journal of Statistics & Management Systems
ISSN 0972-0510 (Print), ISSN 2169-0014 (Online)
Vol. 29 (2026), No. 1, pp. 1–37
DOI : 10.47974/JSMS-1015

Investigating the trends in traffic count data utilizing global and local spatial autocorrelation

Edmund Baffoe-Twum *
Department of Engineering Management and Technology
University of Tennessee, Chattanooga
Dept. 2402, EMCS 326
615 McCallie Ave.
Chattanooga, TN 37403-2598
U.S.A.

Eric Asa †
Bright Awuku §
Adikie Essegbey ‡
Department of Construction Management and Engineering
North Dakota State University
Dept. # 2475, Box 6050
Fargo
ND 58108
U.S.A.

Abstract

It is unclear how to determine any unevenness in Spatiotemporal variability with only a simple glance at datasets. Real-time annual average daily traffic (AADT) data continues to be the primary operations data needed to perfect the exactness of the performance of transportation systems valuation. Despite the enormous investment in data collection procedures (automated and manual), data collection seems sparse and has varying levels of accuracy. The disparity from unoptimized AADT data collection from low-volume roads has stimulated the end-users to close gaps through modeling. However, it is convenient to use statistical techniques to assess the dataset's local/global trends and normality before optimization processes can conveniently be successful. Therefore, exploratory spatial data analysis (ESDA) is explored to identify statistically significant similarities and disparate within datasets collected from the various locations. In addition, the spatial clustering assessment is completed using Getis-Ord Gi statistics. In contrast, spatial autocorrelation

*E-mail: edmund-baffoe-twum@utc.edu (Corresponding Author)
† E-mail: eric.asa@ndsu.edu
§ E-mail: bright.awuku@ndsu.edu
‡ E-mail: adikie.essegbey@ndsu.edu

is explored with Moran's Index. AADT datasets (2009 to 2016) from low-volume roads in Montana, Minnesota, and Washington is explored. The ESDA results indicate that none of the datasets exhibits global trends of global spatial autocorrelation. Nonetheless, there were some indications of local spatial autocorrelation in the datasets. Additionally, the AADT datasets exhibit spatial heterogeneity. The Getis-Ord Gi statistics indicate that most data points are not statistically significant, and the minor hotspot regions are not statistically significant. Moran's Index reveals perfect clustering in the datasets. Though values were dissimilar (alternating values), it expresses the clustering as perfect with spatial outliers or spatial heterogeneity. As a first step before any model is completed, the process will help resolve the variability and divergence within the dataset.

Subject Classification: *62M10.*

Keywords: *Annual average daily traffic, Exploratory spatial data analysis, Hotspots, Getis-ord gi statistics, Moran's index.*

1. Introduction

Low-volume roads carry up to 400 vehicles per day [4], AASHTO [2] and PennDOT [31] have varying classifications for low-volume roads and annual average daily traffic (AADT). However, the Cornel local roads program (2018) and AASHTO [2] guidelines stipulation include design advice for local and minor collector roads carrying average daily traffic volumes of 2,000 vehicles per day or less. Therefore, the road is high volume at an AADT 2000 and above. Low-volume roads are considered rural/low-volume roads, farm-to-market access roads, roads connecting communities, and roads for logging or mining. They are significant parts of any transportation system [23]. They serve the public in rural areas, help improve the flow of goods and services and promote development. In addition, they help to improve public health, education, land, and resource management [23]. Therefore, real-time annual average daily traffic (AADT) data remain the most sought operations data to perfect the exactness of the performance of transportation systems.

Notwithstanding the considerable investment in data collection procedures (automated and manual), data collection is sparse and has varying levels of accuracy. AADT data collected by various stakeholders, typically in low-volume road jurisdictions, do not utilize optimization procedures to cover all locations adequately. Therefore, making meaningful engineering decisions and allocating the already constrained funds to the various transportation agencies is somewhat skewed. The selection of locations for data collection is based on proximity, easiness of accessibility, and comfort. Eventually, the process may lead to essential omissions, such

as possible clusters and hotspots or spatial autocorrelation describing the presence or absence of spatial variations in a given variable at specific locations. This can be for unknown reasons or a lack of information. However, it is essential to note that there is inherent temporal variability in traffic at a site [34], [1], [36], [11], [12], [15], [27]. Factors like the day of the week, vacation or salary disbursement, or economic parameters are readily available [32]. Besides consistent traffic data, the collection process is cumbersome in terms of resources and time [32]. In addition, they may affect traffic flow. Therefore, alternative sampling and estimation methods are likely to introduce additional variability. To ensure adequate and appropriate decision-making, subject data collection to some extent and analyses must be conducted to determine statistically significant similarities or differences. Superlative corrections based on statistical analyses help eliminate the substantial spatial variability the AADT dataset typically exhibits. A specific variable correlates to a location using Exploratory Spatial Data Analysis (ESDA) and considering the same variable's values in the neighborhood [21]. The methods used for this purpose are Spatial Autocorrelation [21]. ESDA is developed from exploratory data analysis (EDA) and is used in spatial statistics, spatial econometrics, and geostatistics [5].

Various traffic-flow pattern predicting models have been developed, considerably enhancing traffic-flow prediction accuracy. However, as a primary input to drive any predictive model, the data and spatial variability automatically influence predicted AADT values. Therefore, understanding factors influencing traffic functioning and monitoring will improve traffic volume (AADT) predictions, inform investigators of data randomness and requirements and reduce skewness. In addition, this will provide accurate data for total dependence during or at any policy-making stage. However, these variabilities are mostly unexplored; hence decisions are made based on the outputs of these models.

Understanding trends and patterns within a dataset is a prerequisite for good judgment. It helps select an appropriate universal model with no restrictions or allocate resources for data collection at the desired location. For example, Staats [36] asserts that some techniques are specific to a particular area. Thus, they cannot be applied to all locations except when altered.

Although Arem et al. [37] and Han and Song [19] specify different techniques for generating excellent summaries of variability in daily traffic counts, reliance on automation is worth considering. For example, suppose all locations are covered for AADT data collection using

automated procedures to ensure equity; there will be data loss during the downtime of possible breakdowns. The process will eventually contribute to the data imbalance or result in a lack of data during the downtime, thus necessitating estimating data appropriately for unsampled locations or areas where the equipment breakdown is experienced. Closing these gaps with a data optimization statistical approach is essential. Optimizing data collection locations can primarily influence the prediction of AADT and many other variables [8], [9]. Hence data clustering and hotspot studies can establish the dataset quality and distribution.

A typical example is exploratory spatial data analysis (ESDA), a set of techniques capable of describing and visualizing a dataset's spatial distributions [21]. ESDA has been utilized in many fields, such as economics, geology, geography, criminology, medicine, etc. In ESDA, emphasis can be placed on the significance of spatial interactions and geographical location to understand immediate or future development [21]. ESDA can discover spatial correlation (spatial clustering or hot spots) with the potential to identify outliers in a dataset. Additionally, the ESDA output can suggest different spatial systems or spatial heterogeneity. By identifying spatial autocorrelation and spatial heterogeneity at a site, the characteristics of the AADT datasets can be considered and further explored to determine the factors or historical events impacting data collection [21]. The statistical procedures can test the hypothesis and the assumptions made during data collection to ensure effective modeling [28].

This article outlines the outcome of ESDA and hotspot analyses using the AADT dataset. The process examined the statistically significant similarities and disparate trends in the dataset. The main ESDA methods explored are global and local spatial autocorrelation. The global spatial autocorrelation emphasizes the overall trend in the dataset and conveys the degree of clustering [28]. The local spatial autocorrelation detects variability and divergence within the dataset in disparity [28]. The local spatial autocorrelation helps us identify hotspots and coldspots within the data or locations [28]. Data from randomly selected three states, Montana, Minnesota, and Washington, are used in this study. Figure 1 illustrates approximate locations data collection locations in each state. The state of Minnesota varied AADT data collection sites yearly for the eight (8) years considered.

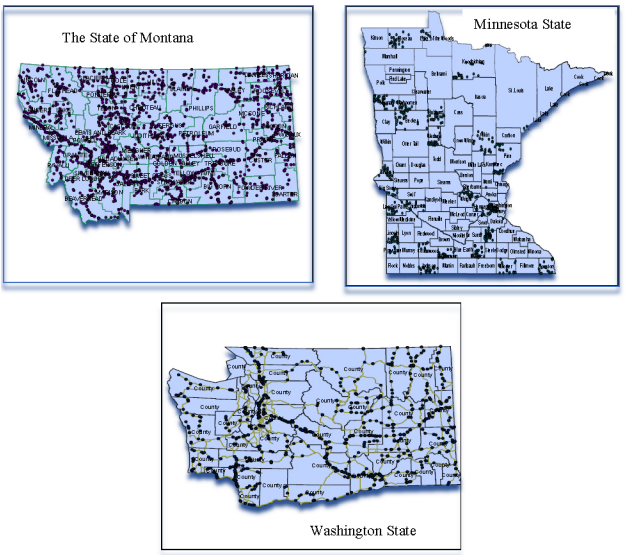

Figure 1

States studied for this research. Dots are AADT collection points.

On the other hand, Montana and Washington AADT data collection locations remained almost the same. However, some additional sites were added, and others were removed each year. This study's main objective is to identify local and global, spatial, and temporal variabilities in the AADT dataset with time. Additionally, the study explains the performance of hotspot analysis and exploratory spatial data analysis as applied to the AADT dataset. Besides, it determines the randomness in quantities not known exactly or unsampled. Therefore, the process can serve as the first step for transportation practitioners to access the randomness in a location dataset and to determine the correlation within the dataset.

2.0 Previous Studies

The need for powerful techniques to study changes in patterns has grown due to the proliferation of social indicator databases [3]. In their publication, Chen, Lu, and Boedihardjo [8] assert that the ever-increasing size of spatial data significantly demands the capability to unearth valuable but unreserved information within the dataset. The interest of the authors was to generate a statistical approach capable of detecting spatial outliers from a dataset [21]. Detecting spatial outliers makes it possible to

uncover data values different from other spatial neighbors. Young et al. [39] discuss the sparse and accuracy issues associated with AADT data collection, even though there have been considerable investments in the process for several years. As a result, Young et al. [39] investigate the need for data, and the fundamental causes of the inaccuracies or limitations using basic statistical models which account for the sampling procedure (probe data). The probe-type data calibrated with continuous counts confirm the potential to meet the accuracy prospects of the end-users of AADT estimates [39]. According to Anselin et al. [3], using a study of clusters and outliers based on data for a child risk scale computed for counties in Virginia, the principles underlying ESDA are explained. Accordingly, Anselin et al. [3] suggest ESDA can leverage the information contained in social indicator databases. The discussions considered "the utility of spatial methods for state agencies in monitoring social indicators at various localities" [3]. Using ESDAs local Moran's Index (MI) and Getis-Ord's Gi* can increase the systematic spectrum allowing for exploring spatial dependence and spatial heterogeneity displayed in voting patterns [24].

Chen et al., [9] try to establish a balance in sampled AADT, which, they argue, is an essential measurement used in traffic engineering. Chen et al. [9] use the synthetic minority oversampling technique (SMOTE). Besides the SMOTE technique, a generalized linear mixed model (GLMM) is used to estimate AADT, which considers roadway design factors, socio-demographics, and land-use patterns as integrated independent variables. According to Chen et al. [9], the data collected is imbalanced and has biased results from stakeholders not optimizing automatic traffic recorders (ATRs) locations. In the study, Chen et al. [9] use data from Seattle, Washington State, to complete the model. The SMOTE program corrects the imbalanced sample proportions [9]. According to Chen et al. [9], higher population density or more mixed land use results in higher AADT data. Besides, AADT correlates negatively with distance to the nearest arterial road. Therefore, the combined SMOTE and GLMM improve the estimation accuracy of AADT [9]. The authors confirm that an exceptional location condition sometimes contributes to an unbalanced estimation of AADT data values.

Consequently, generating disparities in traffic measurements at the targeted locations eventually impacts the datasets collected's accuracy and stability. Using exploratory spatial data analysis, Gallo and Ertur [18] explore the dynamics of Europe's regional per capita gross domestic product (GDP). Gallo and Ertur's [18] study spans between 1980-1995.

Gallo and Ertur [18] perceive a global and local spatial autocorrelation in per capita GDP during the period under consideration. Furthermore, the persistent spatial disparities between European regions revealed high and low clusters per capita products. In conclusion, the analysis perfects the spatial pattern inquiry of regional growth [18].

In another development, Lian et al. [26] performed a spatial pattern analysis of the JingJinJi metropolitan region's economic growth using exploratory spatial data analysis (ESDA). Lian et al. [26] analyze spatial autocorrelation (global and local spatial disparity), spatial trend, and certified analysis of spatial models. The per capita GDP in the JingJinJi metropolitan region, according to Lian et al. [26], through exploratory spatial data analysis (ESDA), is spatially autocorrelated and dependent on economic growth.

Messner et al. [29], using ESDA, examine two time periods in 78 counties in or around the St. Louis metropolitan area and the distribution of homicides. The periods were 1984 –1988 and 1988 –1993, when homicide was relatively stable and increased homicide, respectively. According to Messner et al. [29], the findings reveal that homicides are distributed nonrandomly, expressing positive spatial autocorrelation. The barriers against the diffusion of homicides are more affluent areas or rural or agricultural areas [29]. The spatial distribution patterns through ESDA provide an empirical foundation for specifying multivariate models, providing formal tests for diffusion processes [29]. The growth and development level of the 76 Turkish regions from 1995 to 2001 is explored with exploratory spatial data analysis by Celebioglu and Dall'erba [7]. Though spatial statistics tools reveal spatial dependence across provinces, choropleth maps developed by Celebioglu and Dall'erba [7] suggest that the Western part of the country is significantly more developed than the East. Celebioglu and Dall'erba [7] again confirm heterogeneity in disseminating local indicators of spatial association statistics. Celebioglu and Dall'erba [7], through ESDA, reveal the distribution of growth across Turkish regions and its relationship with public investments and human capital, which are considered the two development indicators. Long [38], using ESDA, analyzes the production growth rates in Chinese provinces since the late seventies.

Rusche [33] utilizes exploratory spatial data analysis to determine Germany's quality of life's spatial distribution. Rusche [33] identifies statistically significant (dis-)similarities in space to the most attractive place to live and work. Furthermore, Rusche [33] identifies a significant spatial autocorrelation in the quality of life associated with the North-

Mid-South divide. Finally, Rusche [33] again uses exploratory spatial data analysis to enhance the regression specifications to prevent spatial dependencies from the residuals.

Using spatial analysis, Yu et al. [40] identify hazardous road segments or hotspots. Furthermore, Yu et al. [40] confirm the superiority of two spatial analysis methods over four conventional methods. Therefore, spatial analysis methods will present the most precise locations in ongoing research for hotspot delineation. The spatial methods the authors used are kernel density estimation (KDE) and local spatial autocorrelation methods.

Also, Montella [30] compared several methods for identifying crash hotspots. Montella [30] states that erroneously identified hotspots may result in inefficient use of resources for enhancing safety at the locations and may reduce the effectiveness of the safety management process globally. Again, Montella [30] asserted that researchers had compared only a few methods adopted for hotspot safety analysis. Therefore Montella [30] contrasted seven known hotspot identification methods against four robust and informative quantitative evaluation criteria. These methods were Crash frequency (CF), equivalent property damage only (EPDO) crash frequency, crash rate (CR), proportion method (P), empirical Bayes estimate of total-crash frequency (EB), empirical Bayes estimate of severe-crash frequency (EBs), and potential for improvement (PFI). The comparison is made based on the site consistency test, the method consistency test, the total rank differences test, and the total score test. According to Montella [30], the methods' performance is based on hotspots indication efficiency at poorly tenacious safety performance sites, identifying the same hotspots accurately and consistently in the ranking at different times. Montella [30] finally ranks the performance of the identification methods. Coll, Moutari, and Marshall [10] complete a search for an alternative approach to accident analysis and prevention using hotspot identification methods to delineate worst-performing areas or time slots on-road segments. Coll, Moutari, and Marshall [10] contributions help prioritize road safety improvement interventions and augment ongoing research on identifying and ranking road safety composite index estimates for hotspots.

3.0 Methodology

This section uses exploratory spatial data analysis and hotspot analysis to evaluate AADT data. Adopting this approach ensures the necessary understanding and integrity of the data before employing any

modeling techniques to explore the data further. The process is displayed in the flow chart (Figure 2).

3.1 *Description of Dataset*

The data for the studies are acquired from the transportation department of Minnesota, Montana, and Washington. The data were a Microsoft Excel spreadsheet formatted into zipped files. Each state has a compiled and completed set of AADT datasets of annual average daily traffic for several years. The years of interest in this study were from 2009 until 2016. Data for these years had been verified, complete, and reliable for studies. Data extracted were also compiled yearly for each state. Therefore, the dataset helped obtain the information needed for the task.

The data is classified based on the Cornell Local Roads Program (CLRP [13]) suggestion to classify AADT for low-volume roads and incorporate the lower limit for the high AADT value. The literature shows low-volume roads AADT value is set at 400 or fewer vehicles per day. This study uses vehicles per day ranging from greater than zero (0) to a maximum of 2000 vehicles per day. Besides the CLRP, consideration is given to classifications discussed in Sharma et al. [35], PennDOT [31], Apronti et al. [4], and AASHTO [2]. The AADT values are up to 400, 500, 1000, and 2000 vehicles daily.

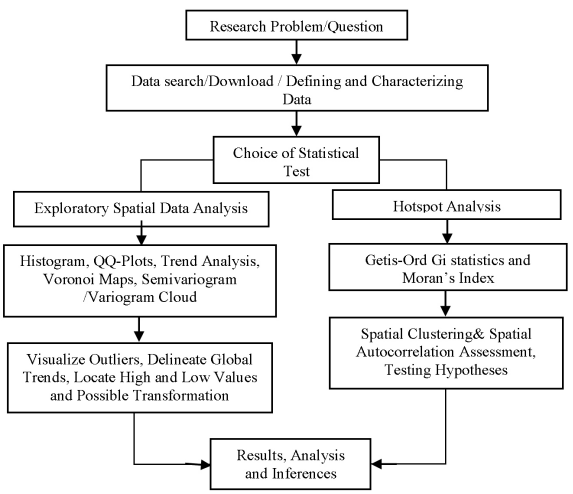

Figure 2

Flow chart for the statistical analysis using ESDA and hotspots

Further screening is also completed to remove data collection stations with zero counts and missing data. The process is completed using conditional clauses in Excel. Finally, each complete data for AADT of up to 400, 500, 1000, and 2000 vehicles per day from 2009 until 2016 were converted, displayed as points, and shapefiles using ArcGIS 10.7 for Desktop [14].

3.2 *Exploratory Spatial Data Analysis*

The cleaned dataset is subjected to exploratory spatial data analysis in ArcGIS 10.7. Exploratory spatial data analysis (ESDA) is a quantitative technique used in examining collected data [15], [18]. ESDA helps in gaining a better understanding of the data trends. As a result, there is a making-of a knowledgeable decision on the possible interpolation [15], [18]. The exploratory spatial data analysis is also applied to determine and make it easy to visualize outliers, delineate global trends in the data, locate areas with high and low values, and possibly transform data [15], [18]. The steps involved include examining the data distribution, determining the general trends, the local and global outlier identification, variations locally, and spatial autocorrelation. Analysis completed on the AADT dataset using exploratory spatial data analysis are histograms, quantile-quantile plots (QQ-Plots), Voronoi maps, trend analysis, and semivariogram/covariance cloud analysis.

The histograms and the quantile-quantile plots (QQ-Plots) determine the normality, skewness, and outliers. The histogram analysis identified outliers as a bar on the extreme left or right or an isolated bar from the rest. It is even more apparent when the bar is more isolated from bars in the histogram [15], [18].

The Voronoi maps (entropy and standard deviation) determine stationarity in the dataset. Voronoi maps (simple Voronoi maps) are handy in determining potential outliers. Voronoi entropy is a quantitative mathematical tool to characterize the orderliness of points distributed on a surface [6]. In an entropy map, the polygons are characterized into five classes based on a natural (smart quantiles) grouping of data values. The entropy assigns a value to a polygon calculated from the polygon and its neighbors [16].

$$\text{Entropy} = -\Sigma\,(\text{pi} * \text{Log pi}) \qquad \text{Equation 1}$$

The difference between the two data points depends on their distance. The relationship assumes a constant variance with trends removed.

Trends analysis tool assesses the trend in data. Thus, the data determines all the possible trends (1st, 2nd, and 3rd order polynomial trends). The Semivariogram/covariance cloud analysis tool determines how a model fits the dataset. It is, again, helpful for detecting local outliers. When a point on the graph has a low value on the x-axis, and a high on the y-axis, the points that make up the pair tend to have different values. The depiction is contrary to data values close together and have comparable values. Thus, it allows for examining the spatial autocorrelation between the measured sample point (based on Tobler's first law). All ESDA steps mentioned were explored, helping determine whether further analysis is required to transform the datasets.

3.3 *Hotspots Analysis*

Similarly, in ArcGIS 10.7, the hotspot analysis technique of the dataset is explored. The exploration of the dataset helps to authenticate statistically higher predispositions of an area that is clustered spatially. Hotspot analysis is a spatial analysis and uses mapping techniques to identify the clustering of spatial phenomena. A hotspot is depicted as points in a map for a dataset. A future is classified as a statistically significant hot spot if it has a high value and the surrounding features are high [16], [17], [22]. Thus, each event is examined within the context of neighboring features. In this study, the hotspot analysis (HA) is completed on every feature in a dataset. Considered were Getis-Ord Gi statistics and Moran's Index. The Getis-Ord Gi statistic is a tool for assessing spatial clustering of low and high values where the analysis field corresponds to the field's data [16]. The tool has been applied to several investigations, including crime analysis, voting patterns, traffic incident analysis, mineral resources, etc. [16]. The best practice and most reliable results have input features more significant than 30 data points with a fixed distance band to conceptualize spatial relationships [16].

This study's dataset for each AADT considered is more than 30 data points. By visual inspection, the occurrence of spatial clustering in data is indispensable for hotspot analysis. The Gi statistics in the feature dataset is the Z-score. Statistically significant positive Z -scores (larger Z-score) indicate the intensity of clustering of high values (hotspots), whereas the features statistically significantly negative Z-score (smaller Z-score) clustering of low values (coldspot) [16]. The mathematics behind the results is shown below: the Getis - Ord local statistics.

$$G_i^* = \frac{\sum_{J=1}^{n} w_{ij}x_j - \bar{x}\sum_{j=1}^{n} w_{ij}}{S\sqrt{\frac{[n\sum_{J=1}^{n} w_{ij}^2 - (\sum_{j=1}^{n} w_{i^j})^2]}{n-1}}} \qquad \text{Equation 2}$$

$$\bar{X} = \sum_{j=1}^{n} \frac{x_j}{n} \qquad \text{Equation 3}$$

Where x_j is the attribute value for feature j. w_{ij} is the spatial weight between n feature i and j. n is equal to the total number of features and G_i^* statistics is a Z-score, so no further calculations are required. The G_i^* statistics return for each feature in that dataset is a Z-score.

$$S = \sqrt{\frac{\sum_{J=1}^{n} x_j^2}{n} - (x)^2} \qquad \text{Equation 4}$$

http://desktop.arcgis.com/en/arcmap/10.5/tools/spatial-statistics-toolbox/h-how-hot-spot-analysis-getis-ord-gi-spatial-stati.htm

The spatial autocorrelation measured with Moran's I returned values between negative (-1) and positive (+1). Where negative (-1) is an indication of perfect dispersion or perfect clustering of unrelated values, zero (0) indicates a randomly complete arrangement (no autocorrelation or perfect randomness), and a perfect correlation in the north or south divide at positive (+1) value or perfect clustering of comparable values. Moran's I is used in testing hypotheses statistically. The procedure had Moran's I value transformed into z-scores where values greater or lower than 1.96 and -1.96, respectively, indicated autocorrelation at a 5% level of significance [16]. Moran's index calculation is based on the weighted matrix with units i and j. The similarities are calculated with the overall mean and the product of the differences between y_i and y_j.

$$\text{Similarity} = (y_i - y) * (y_J - y), y = \sum n_{i=1} \frac{y_i}{n} \qquad \text{Equation 5}$$

Moran's statistics are generated with the basic form divided by the sample variance (S^2)

$$S^2 = (\Sigma(y_i - y)^2)/n \qquad \text{Equation 6}$$

$$\text{Moran's Index} = \frac{1}{S^2} \frac{\Sigma_i \Sigma_j (y_i - \bar{y})(y_j - \bar{y})}{\Sigma_i \Sigma_j w_{ij}} \qquad \text{Equation 7}$$

The following conditions are to be met when employing the hotspot analysis tool:

- Every feature must have at least one neighbor;
- No feature should have all other features as neighbors;
- Eight neighbors are necessary for each feature when values are skewed.

The output table generated for spatial clusters has P- values, and Z-scores to indicate the feature's high and low values. However, it is worth knowing that a feature may not be statistically essential and identified with a hotspot because of its high value. Thus, a feature identified with the hot spot is the high value when surrounding features are of high value. The summation of a feature with its neighbors is contrasted with the sum of all features correspondingly [16]. A statistically significant Z - scores result from the difference between the sum of neighboring features and the likely local sum when the resulting difference is too substantial to represent a random chance [16].

4.0 Results And Discussion

The collection of techniques to describe and visualize spatial spreading is an exploratory spatial data analysis (ESDA), a subset of exploratory data analysis (EDA). Exploratory data analysis identifies data properties that help in pattern detection in the said data. That summarizes the main characteristics of the data [20]. It also helps in hypothesis formulation from the data and aspects of the models' assessment [20], such as goodness-of-fit in the data. Also, exploratory data analysis emphasizes the description of pattern detection in data. Exploratory data analysis does not explicitly investigate the location component of the dataset [20]. EDA instead deals with the correlation between variables and how they affect each other [20]. ESDA identifies locations or spatial outliers, patterns determining spatial association, clusters, or hotspots suggesting spatial procedures and spatial heterogeneity in different forms. This process's most prominent determinant factor is spatial autocorrelation or spatial association of datasets and locations [20]. The process depends on the Tobler law of spatial proximity: everything is related to everything else; however, near things relate more than distant things. This concept applies to countless phenomena, such as pollution, noise, soil sciences, etc. In a real sense, exploratory spatial data analysis concentrates on spatial and attribute

association or correlates a specific variable to a location, considering the same variable's values in the neighborhood [20]. In exploratory spatial data analysis, several questions relating to the collected or observed dataset were asked. Typically, it is essential to know the data's location and values at the data point and determine its relationship to the observed values. Additionally, it is critical to determine if the dataset is normally distributed.

Therefore, the histogram checks for bell-shaped distribution and outliers in the dataset. A normal quantile-quantile plot (QQ Plot) checks if the data follows the 1:1 line. The normality dataset deviation is transformed using log, box cox, arcsin, and normal score transformation before using the dataset in a model or interpretations requiring normal data distribution. The data must be bell-shaped for normally distributed data, with no outliers, mean to be approximately equal to the median, skewness to zero, and kurtosis to approximately 3. Therefore, the closer the dataset's skewness approximates zero, the better because it may not need a transformation process. Next, Voronoi entropy or standard deviation maps was used to complete a data stationarity check. A dataset's stationarity is measured using the statistical relationship between two data points based on distance. Finally, a nonstationary dataset is transformed to stabilize the variances of the points. The variance, therefore, becomes constant when trends in the data have been removed.

4.1 *Montana*

4.1.1 Exploratory Spatial Data Analysis

From the dataset from Montana state, all categories of AADT values, up to 400, 500, 1000, and 2000 vehicles per day from 2009 to 2015, needed no transformation because the untransformed data skewness indicated a closeness to zero compared to the log-transformed skewness information. However, the dataset's skewness for 2016 preferred a transformation, and visual inspection of the transformed histogram and QQ Plots leaned toward transformation. In addition to the histograms, QQ plots, Voronoi maps, and trend analysis were used to study the dataset's systematic changes across the study area. The data trends were described with the three orders (1^{st}, $2^{nd,}$ or 3^{rd}) of the polynomial. Table 1 shows examples of the histogram results from the Montana AADT dataset. The results cover the years from 2009 to 2016, where AADT categories of 400, 500, 1000, and 2000 vehicles per day are accounted for. The skewness in the dataset ranges between slight and moderate.

Table 1

Summary of skewness and log-transformed skewness – Montana (from histogram)

Year	AADT Count (Montana)	Skewness	Log Transformed Skewness	Apply Transform (Yes/ No)
2009	400	0.3566	-0.8050	No
	500	0.3893	-0.8075	No
	1000	0.5598	-0.7927	No
	2000	0.6819	-0.7720	No
2015	400	0.3424	-0.7616	No
	500	0.3665	-0.7841	No
	1000	0.5794	-0.7764	No
	2000	0.6865	-0.7676	No
2016	400	0.6346	-0.5161	Yes
	500	0.6655	-0.5118	Yes
	1000	0.7859	-0.5159	Yes
	2000	0.8923	-0.5338	Yes

The log-transformed data indicates no need for data transformation. Figure 3 shows the results of 2011_AADT_500 non-transformed and log-transformed graphically for histogram and Normal QQ Plot, respectively. The data count is 1261, with a mean of 213.63, a median of 190, and a kurtosis of 1.9563.

The minimum and maximum AADT values are 10 and 500, respectively. The median and mean are dissimilar. The transformed histogram shows a positive or right skewness; therefore, it departs from the untransformed dataset and is not closely related to a normally distributed histogram. The normal QQ plot of the same year AADT demonstrates a similar trend to the histogram. The skewness value for the untransformed data in 2011 with an AADT of up to 500 is 0.39717. Whereas the log-transformed resulted in a skewness value of -0.77914. The untransformed data can approximate a skewness of zero but not normal distribution in assumption (not symmetrical); however, it is classified as slightly skewed. The value of the transformed data's skewness classifies the output to be moderately skewed, further from the zero value of normality, and therefore not accepted as representation as in data correction. As a result, there is no need for transformation under this assessment condition. The year 2016 data at AADT of 2000 vehicles per day, as indicated in Figure 4, is quite different.

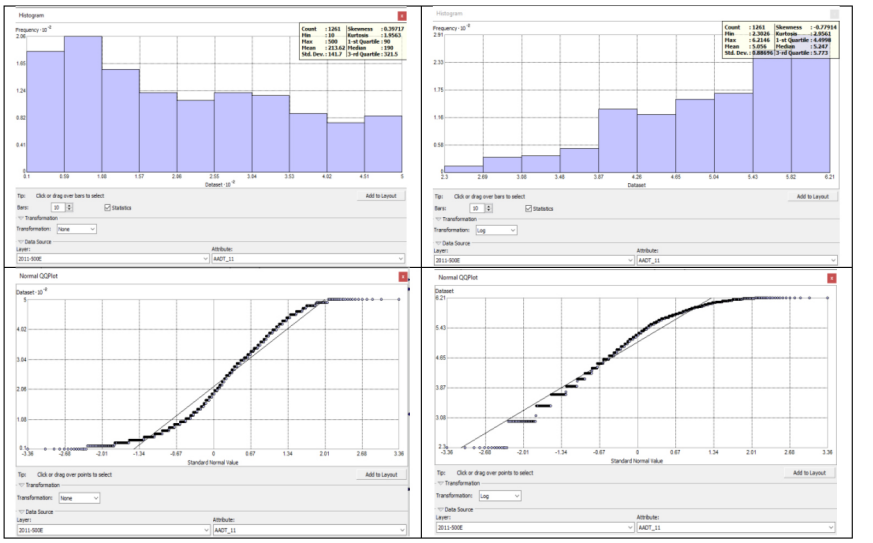

Figure 3

Montana 2011_AADT_500 untransformed (left side) and log-transformed (right side) for histogram and normal QQ Plot

The log-transformed results show a necessitation of data transformation leading to an approximate normality before using other models to generate an appropriate output. Figure 4 shows the difference in the output of the assessment processes. The log-transformed data and graphs (Histogram and QQ plot) confirm the needed transformation.

The data count for Montana 2016 AADT up to 2000 vehicles per day is 3583. It generated statistics with a mean of 600.49, a median of 397, and a kurtosis of 2.6075. The minimum and maximum values are 3 and 2000, respectively. The median and mean are dissimilar. The untransformed and log-transformed skewness resulted in 0.89324 and -0.53377, respectively. The log-transformed negative skewness value is much closer to zero or normality than the untransformed. However, both are categorized as moderately skewed. Therefore, transforming the dataset before using any other model rather than the exploratory spatial data analysis is worth it. The transformed skewness value is at the lower limit of the moderately skewed classification. In contrast, the untransformed is within the upper limit.

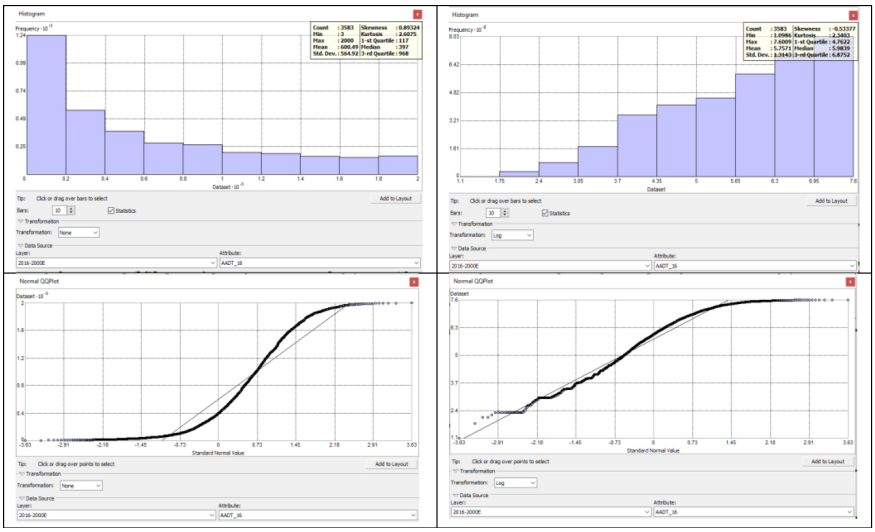

Figure 4

Montana 2016_AADT_2000 untransformed (left side) and log-transformed (right side) for histogram and normal QQ Plot

Besides using the histogram and the QQ-Plots to determine the skewness in the data, the graph depicts data from locations that are unlike nearby values, thus generating the data's non-normality or Gaussian nature. Figure 5 is the entropy Voronoi map, the three polynomial trend analysis orders, and a semivariogram cloud. The left column is the AADT 500 for 2011, whereas the right is the AADT 2000 for Montana's 2016. The entropy map for both did not show true stationarity in the data, though challenging to determine. The variation in the data per the Voronoi map is inconsistent. The map exhibits a moderate variation in the data for 2011 and 2016, respectively. No trends are recognized when all the polynomial orders are applied. The semivariogram cloud map shows variations as there are several combinations of pairs. Typically, as a low x-axis (distance) value against a low y-axis value (semivariance), low x-axis value against a high y-axis value, high x-axis value against a low y-axis value, and high x-axis value against a high y-axis value. Characteristically, the data evaluated exhibit similar outputs. Only the AADT 2000 data is a noted contestant for probable transformation.

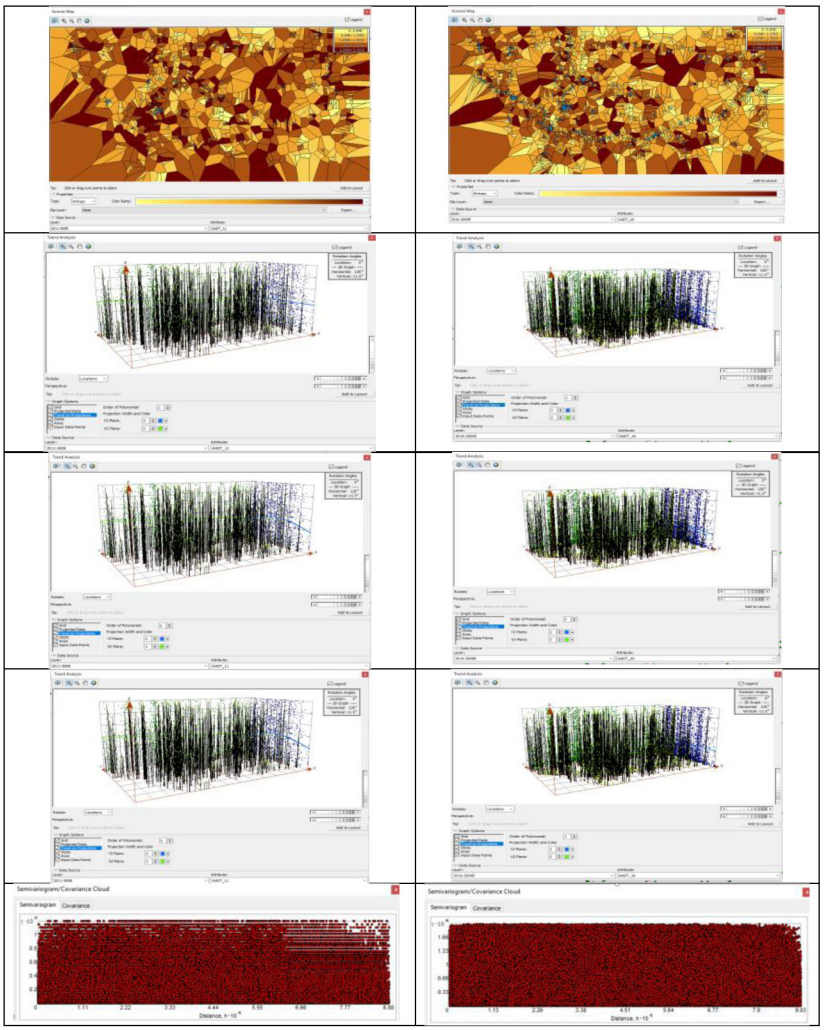

Figure 5

Montana - left column – 2011_500 -AADT and right column 2016_2000-AADT

4.1.2 Hotspots Analysis

The technique adopted is hot spot analysis in the spatial statistical toolset in the ArcGIS toolbox to reduce the subjectivity in analyzing statistically significant clusters in a dataset. In this study, two methods explain the spatial and local/global autocorrelation hotspots in the acquired data. These are Getis-Ord- Gi* statistics and Moran's Index. Each

method tests for randomness besides local/global autocorrelation and spatial autocorrelation. Autocorrelation deals with a variable correlated with itself.

Regarding spatial autocorrelation, the random variable's values are either more or less similar as a function of the distance between them at paired points [25]. As the study deals with a univariate response variable, it correlated itself. However, spatial autocorrelation could be considered bad or good based on the set objectives. The good part is the reliable estimation of unsampled sites at nearby locations. In contrast, the bad part relates to the violation of most of the statistical test assumptions. Furthermore, the observations are not independent; thus, P values are less than 0.05 ($P < 0.05$). Both Getis-Ord- Gi* statistics and Moran's index techniques better explain the reality than the optimized hot spot analysis that considers only sampled locations and leaves out unsampled areas [26].

Moran's Index is used to complete hypothesis testing. The null hypothesis equals no spatial autocorrelation (Moran's I = 0.0, - random pattern), and the alternative hypothesis equals spatial autocorrelation. The null hypothesis is rejected with this condition if the generated Z score test statistics meet the condition: Z > 1.96 or Z < -1.96. Table 2 is a global autocorrelation summary of Moran's Index for Montana AADT up to 400, 500, 1000, and 2000. In summary, for AADT up to 400, all but the year 2016 AADT yielded negative Moran's I and Z-scores values. The negative Moran's Index for 2009 to 2015 indicates perfect clustering in the dataset. Nevertheless, values are dissimilar (alternating) and expressed as perfect dispersion. Also, it indicates a possible presence of spatial outliers or spatial heterogeneity. Moran's index value has no significance to the dispersion or clustering in the data. As global statistics, the Moran's Index says nothing about location. Instead, a Moran's I less than zero implied negative autocorrelation, whereas the reverse indicates positive autocorrelation. A Moran's I equal -1 or +1 indicates a strong negative spatial autocorrection or positive spatial autocorrelation. Thus, the negative and positive spatial autocorrection is used to explain the summary of Moran's I in Table 3 for the type of autocorrelation. None of Moran's I value for the years were equal to zero but were close to zero. The summary could not tell if there is a strong negative spatial autocorrection or positive spatial autocorrelation. However, the null hypothesis of no spatial autocorrelation is not rejected except for 2016 for AADT up to 400 and 500. The null hypothesis of no spatial autocorrelation is rejected for 2009-2016 at AADT up to 1000 and 2000. All the Z test statistics are greater than 1.96 at a 0.05 significant level. Thus, it favors the alternative hypothesis of the existence of spatial

autocorrelation. Also, the negative Z – test statistics indicate somewhat probable clustering in the dataset. From Moran's index summary, AADT up to 400 and 500 followed the same trend for the years reviewed. The summary for AADT up to 1000 and 2000 yielded positive Moran's index and Z – test statistics. The positive Moran's Index indicates the perfect clustering of similar values. However, the clustering could result from either high or low values or a combination of both. It is worth noting that the single value from the summary table for each year represents the whole spatial pattern. Thus, confirming the global scale at which Moran's Index explains clustering. The positive Z-score is an indication of spatial clustering in the data. Because of the global depiction of Moran's Index in the classification of clustering, the Getis-Ord- Gi* statistics helped to visualize the locations of the clustered data with 99%, 95%, and 90 % confidence in the display of hot and cold spots as in Figure 6. Though there were clustered hotspots, cold spots, and not significant locations for the images derived from the dataset, each figure has varying locations for the different classifications of hot, cold, and not significant spots with the associated confidence level. Most of the data points were categorized as not significant from the outputs, with approximately similar data points corresponding to 90%, 95%, and 99% confidence levels for both hot and cold spots.

Correspondently, as the AADT level increases, the number of locations consistent with hot, cold, and not significant spots increases. However, the increase may not correlate linearly. Therefore, further analysis may need to be conducted to determine the correlation between AADT levels and the increase in hot, cold, and not significant spots.

Table 2

Moran's Index summary for AADT up to 400, 500, 1000, and 2000 - Montana State

Montana 400 AADT - Global Moran's Index summary								
	2009	**2010**	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**
Maran's Index	-0.011	-0.012	-0.010	-0.011	-0.026	-0.018	-0.016	0.189
Expected Index	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001
Variance	0.002	0.002	0.002	0.002	0.002	0.001	0.001	0.001
Z- Score	-0.251	-0.272	-0.233	-0.234	-0.637	-0.473	-0.428	6.787

Contd...

Montana 500 AADT - Global Moran's Index summary								
	2009	**2010**	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**
Maran's Index	-0.018	-0.023	-0.020	-0.021	-0.023	-0.020	-0.019	0.193
Expected Index	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001
Variance	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
Z- Score	-0.544	-0.657	-0.584	-0.644	-0.728	-0.656	-0.610	7.917
Montana 1000 AADT - Global Moran's Index summary								
	2009	**2010**	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**
Maran's Index	0.081	0.092	0.143	0.106	0.097	0.107	0.156	0.139
Expected Index	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001
Variance	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Z- Score	4.172	4.531	7.110	5.222	4.937	5.754	8. 172	8.358
Montana 2000 AADT - Global Moran's Index summary								
	2009	**2010**	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**
Maran's Index	0.261	0.231	0.235	0.228	0.219	0.241	0.278	0.117
Expected Index	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000
Variance	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Z- Score	21.417	18.386	19.194	18.814	18.490	20.716	24.174	11.303

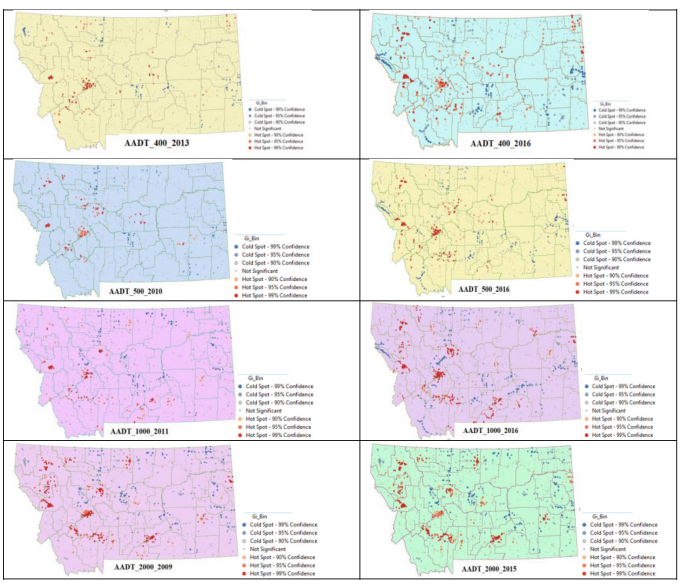

Figure 6

Montana Gi* statistics hotspot analysis output for AADT up to 400, 500, 1000, and 2000

4.2 *Minnesota*

4.2.1 Exploratory Spatial Data Analysis

The skewness data varied for the years reviewed. For example, 2014 to 2016 had skewness ranging between slight and moderate for the actual dataset. In contrast, the skewness from 2009 to 2013 ranged in the highly skewed criteria for the AADT groupings (see summary in Table 3). Therefore, a transformation is needed for the AADT data from 2009 to 2013 to ensure an appropriate, accurate model is obtained.

Table 3

Summary of sample skewness and log-transformed skewness - Minnesota (from histogram)

Year	AADT Count (Minnesota)	Skewness	Log Transformed Skewness	Apply Transform (Yes/NO)
2009	400	1.7175	-0.4965	Yes
	500	2.0045	-0.4213	Yes
	1000	3.0664	-0.0617	Yes
	2000	4.2226	0.2852	Yes
2013	400	1.0925	-0.1274	Yes
	500	1.1572	-0.0941	Yes
	1000	1.4495	-0.0136	Yes
	2000	1.6010	0.0477	Yes
2014	400	0.2198	-0.9850	No
	500	0.2590	-0.9892	No
	1000	0.5525	-0.9210	No
	2000	0.8882	-0.7657	No
2016	400	0.1236	-1.2361	No
	500	0.2484	-1.2176	No
	1000	0.4971	-1.0956	No
	2000	0.6956	-0.9601	No

The histogram and QQ plot in Figure 7 for 2009 and the AADT of up to 2000 vehicles per day represent the original data's transformation. The untransformed data used in the histogram had a count of 560, with a minimum of 5 and a maximum of 1800. The mean is 141. 05, and the median

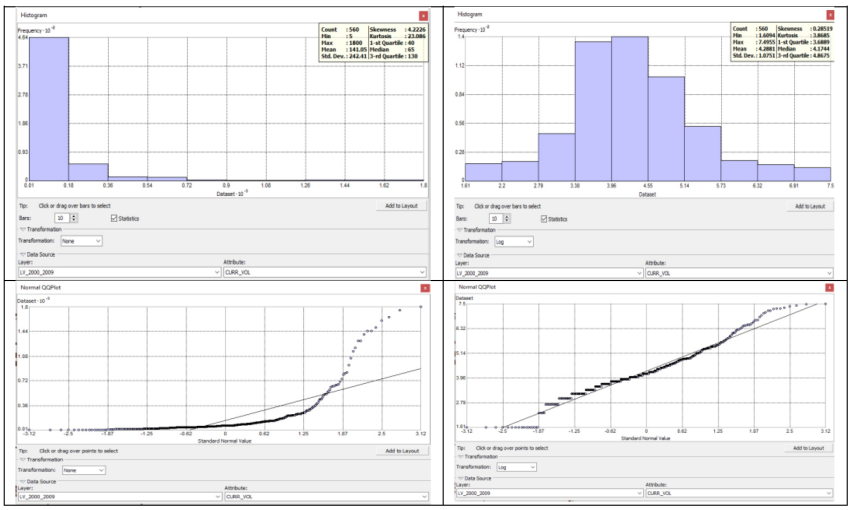

Figure 7

Minnesota 2009_AADT up 2000 untransformed (left side) and log-transformed (right side) for histogram and normal QQ Plot

is 65; both numbers are not similar and, therefore, not symmetrical. The kurtosis is 23.086, and the skewness is 4.2226. However, the log-transformed histogram gave a skewness of 0.28519, approximating symmetry. The result displays a highly positive (right skewness) conversion to a slightly close normally distributed histogram.

Similarly, the inverted S-shaped curve shown on the non-transformed QQ plot is presented in the log-transformed QQ plot to align with the straight line. An indication that transforming the original dataset for 2009 with AADT at up to 2000 vehicles per day may be necessary for further models. Figure 8 presents 2015 at an AADT of up to 1000 vehicles daily. Data transformation is not necessary at any time for other models. The count for the data points is 2695, with a minimum of 5 and a maximum of 1000. The mean is 394.13, and the median is 330. The mean and median are not similar but enough to approximate moderately symmetrical. The kurtosis is 2.1954, whereas the skewness is 0.54169. The skewness is at the lower limit of the moderately skewed data. However, the transformed data has log-transformed skewness resulting in -1.0214. The log-transformed result in highly (-negative) left-skewed data. Thus, it is not necessary to transform the data. This same pattern is shown in the untransformed and log-transformed QQ plot output in Figure 8.

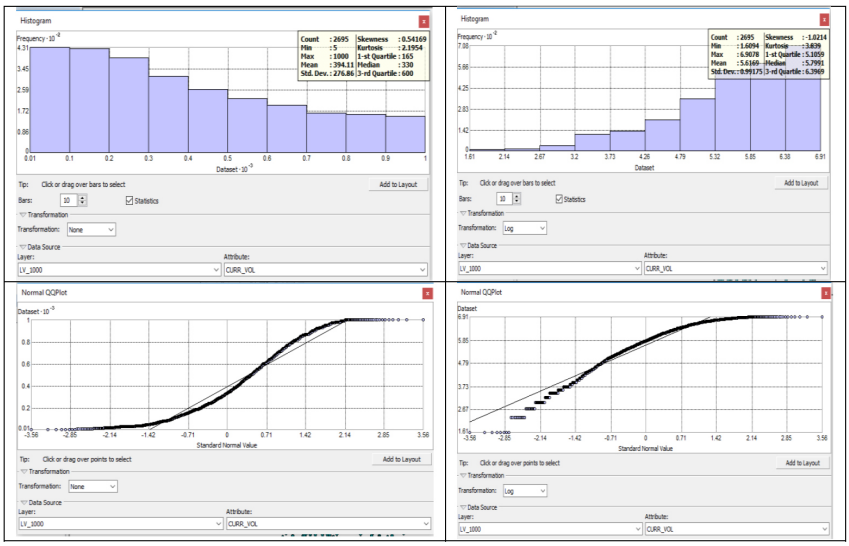

Figure 8

Minnesota 2015_AADT up to 1000 untransformed (left side) and log-transformed (right side) for histogram and normal QQ Plot

The entropy Voronoi map, the three polynomial trend analysis orders, and a semivariogram cloud for AADT of up to 2000 vehicles per day for 2009 and the 2015 AADT of up to 1000 vehicles per day are assessed. The entropy map for 2009 and 2015 did not show stationarity in the data though it is difficult to determine; however, it shows slight to moderate variation. No trends were recognized from the polynomial's 1st, 2nd, and 3rd orders. The semivariogram cloud map showed variations as there were several combinations of pairs.

Typical pairs on the semivariogram cloud are low x-axis (distance) value against low y-axis value (semivariance), low x-axis value against high y-axis value, high x-axis value against low y-axis value, and high x-axis value against high y-axis value. Characteristically, the data evaluated exhibited similar outputs. The AADT up to 2000 data for 2009 is a noted contestant for probable transformation and exhibited some potential outliers.

4.2.2 Hotspots Analysis

ADDT measurement in Minnesota varied locations where data collection is completed each year. Thus, interpretation varied and did not

follow any trend. Table 4 is a global autocorrelation summary of Moran's Index of Minnesota AADT of up to 400, 500, 1000, and 2000 vehicles per day.

Table 4

Moran's Index Summary for AADT up to 400, 500, 1000, and 2000 - Minnesota

Minnesota 400 AADT - Global Moran's Index summary								
	2009	**2010**	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**
Maran's Index	-0.013	0.67	-0.006	-0.026	-0.015	0.098	-0.017	0.043
Expected Index	-0.002	-0.001	-0.001	-0.002	-0.001	0.0000	0.0000	0.0000
Variance	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Z- Score	-0.51	35.32	-0.664	-1.058	-0.921	7.718	-1.33	4.411
Minnesota 500 AADT - Global Moran's Index summary								
	2009	**2010**	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**
Maran's Index	-0.012	0.951	-0.008	0.057	-0.012	0.106	0.200	0.105
Expected Index	-0.002	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001
Variance	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Z- Score	-0.652	50.04	-0.904	2.535	-0.695	9.212	19.246	9.329
Minnesota 1000 AADT - Global Moran's Index summary								
	2009	**2010**	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**
Maran's Index	-0.001	1.481	0.075	0.097	0.046	0.152	0.145	0.11
Expected Index	-0.002	-0.001	-0.001	-0.002	-0.001	-0.001	-0.001	-0.001
Variance	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Z- Score	0.551	73.635	7.447	4.598	5.091	19.894	15.952	14.855
Minnesota 2000 AADT - Global Moran's Index summary								
	2009	**2010**	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**
Maran's Index	-0.01	1.187	0.197	0.100	0.353	0.141	0.155	0.081
Expected Index	-0.002	-0.001	-0.001	-0.002	-0.001	0.0000	0.0000	0.0000
Variance	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Z- Score	-0.56	58.984	18.4	5.352	22.719	24.764	21.652	20.801

In summary, for AADT up to 400, all but the years 2010, 2014, and 2016 AADT yielded negative Moran's Index and Z-scores values. The Z-test statistics are high, particularly for the year 2010. The same pattern is realized for AADT up to 500, 1000, and 2000 vehicles per day for 2010. Moran's index and Z test statistics for AADT up to 500 are negative for 2009, 2011, and 2013. For AADT up to 1000, the Z- test statistics values

are positive along with Moran's Index except for 2009. That negatively impacted Moran's index value. For AADT up to 2000, only 2009 resulted in negative values for Moran's Index and the Z- test statistic. The negative Moran's Index indicates perfect clustering in the dataset, even if the values are dissimilar (alternating).

Thus, the clustering is expressed as perfect dispersion, possibly indicating spatial outliers or heterogeneity. However, Moran's index value has no significance to the dispersion or even to the clustering in the data. As a global statistic, Moran's Index explained nothing about location. The negative Z – test statistics indicate somewhat probable clustering in the dataset. Positive Moran's Index indicates the perfect clustering of similar values. However, the clustering of the data points could be due to either high or low values or a combination of both. Again, it is worth noting that the single value represents a whole spatial pattern and a specific location, thus confirming the global scale at which Moran's Index explains clustering.

The positive Z- test statistics indicate spatial clustering in the data. The Moran's Index and the Z- test statistics for 2010 for AADT up 400, 500, 1000, and 2000 vehicles per day have Moran's Index approximating to +1 and Z- test statistics greater than +1.96. In addition, the 2010 data exhibited strong positive spatial autocorrelation when considering Moran's Index. Therefore, the alternative hypothesis is accepted as Z-test statistics greater than 1.96 at the 0.05 confidence level

Though the global representation using Moran's Index and a single value in the classification of clustering explained the dataset, the Getis-Ord- Gi* statistics helped visualize the clusters' locations in the data at 99%, 95%, and 90 %, as in Figure 9. There are coincidental hot spots (red dots), cold spots (blue dots), and not significant locations(grey dots) for the images derived from the dataset. Each output map had varying locations for the different classifications of hot, cold, and not significant spots with the associated confidence level. The images likewise show location changes for the yearly measurements for AADT. The minor data points are categorized as not significant, with significant data points corresponding to 90%, 95%, and 99% confidence levels for both hot and cold spots. Moreover, as the AADT category increases, the number of locations consistent with hot, cold, and not significant spots increases. Therefore, the increase in AADT cannot be linearly correlated with the hot, cold, and not significant spots.

Further analysis may need to be conducted to determine the correlation between AADT levels and the possible increase and increase

in the hot, cold, and not significant spots. The hotspots were mostly in the central, eastern, and southeastern parts. The exception is the output for AADT at up to 1000 for 2011 and 2015. Hotspots were observed in the west and southwest of the state. The least of the hotspots in the output were observed in 2009 at AADT up to 400 and in 2012 at AADT up to 2000. The cold spots concentrated primarily on the upper half of the state, yet some were to the southwest. The output for 2016 with AADT up to 2000 showed the western half of the state as cold spots.

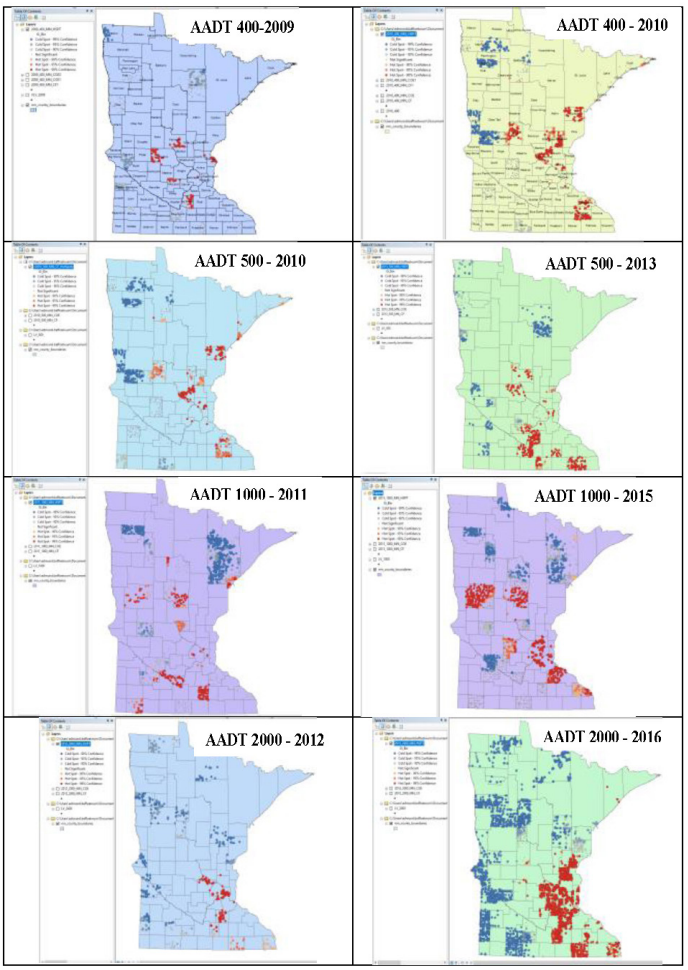

Figure 9

Gi* statistics for hotspots for AADT up to 400, 500, 1000, and 2000 vehicles per day

4.3 *Washington*

4.3.1 Exploratory Spatial Data Analysis

The skewness values for AADT up to 400, 500, 1000, and 2000 for 2009 to 2016 are close to zero (approximates normal distribution). Therefore, none of the log transforms skewness shows a need to transform any datasets under review for the years and AADT groupings (See Table 5). Figures 10 and 11 present the histogram and QQ plot for 2009 and 2016 at AADT of up to 400 and 2000, respectively. For Figure 10, the count is 234, with a minimum of 3 and a maximum of 400. The mean is 233, and the median is 240: the mean and median are approximately similar.

Table 5

Summary of sample skewness and log-transformed skewness (From histogram)

Year	AADT Count (Washington)	Skewness	Log Transformed Skewness	Apply Transform (Yes/NO)
2009	400	-0.1522	-2.0178	No
	500	-0.0617	-1.8211	No
	1000	-0.0720	-1.4618	No
	2000	0.0368	-1.4069	No
2012	400	0.0066	-1.3233	No
	500	0.1402	-1.2507	No
	1000	0.1092	-1.1909	No
	2000	0.0508	-1.3518	No
2014	400	-0.0281	-1.4859	No
	500	0.1148	-1.3697	No
	1000	0.1180	-1.1472	No
	2000	0.0556	-1.2811	No
2016	400	-0.0482	-1.0560	No
	500	0.0711	-1.0128	No
	1000	0.1850	-0.9140	No
	2000	0.0530	-1.1306	No

The skewness is -0.1523, and the kurtosis is 1.9569, whereas the transformed data gives a skewness of -2.0178 and a kurtosis of 11.066. The transformed histogram indicates left-tailed, highly negatively skewed data with outliers. In contrast, the untransformed is slightly skewed with no outliers.

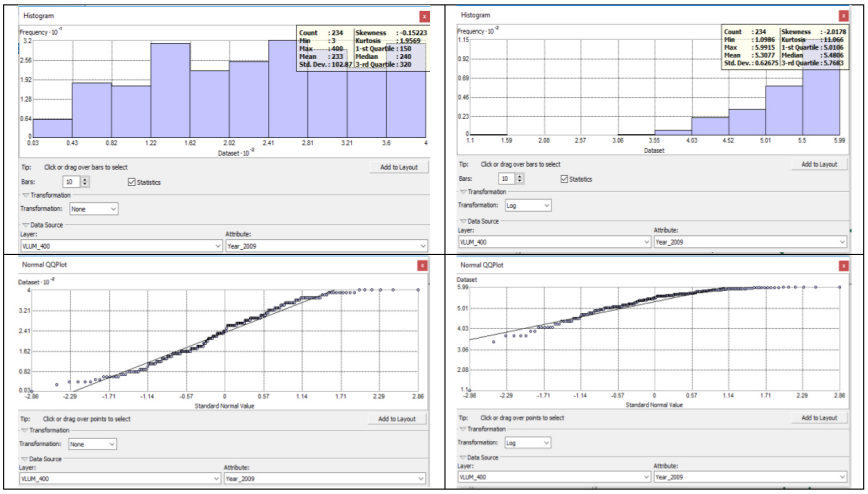

Figure 10

Washington 2009 AADT up to 400 untransformed (left side) and log-transformed (right side) for histogram and normal QQ Plot

Adapting the original dataset's skewness of -0.1523 to -2.0178 may not be necessary for any model. The same can be said from the QQ plot analysis from the same dataset in Figure 10. The outlier in the transformed QQ plot is located at the origin of the plot.

The first step in presenting data trends is the histogram and the QQ plot. The transformed and untransformed data for 2016, at an AADT of up to 1000, is shown in Figure 11. Typically, it exhibits similar output as in Figure 10. The data count is 627, with a minimum of 30 and a maximum of 1000. The untransformed mean and median are 509.49 and 480, respectively. The skewness is 0.1850 for the untransformed data, whereas the transformed is -0.91402, making it unnecessary to transform the dataset. The kurtosis for the untransformed and transformed is 1.7337 and 3.345, separately. The same indication is depicted in the untransformed and transformed QQ plots associated with the data. Outliers are introduced in the data when transformed. Thus, confirming a lack of need to transform the data as depicted in summary Table 5.

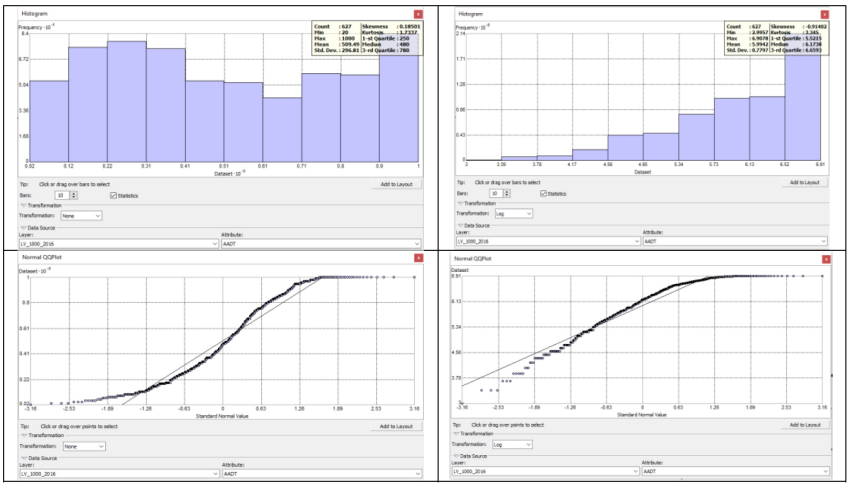

Figure 11

Washington 2016_AADT_1000 untransformed (left side) and log-transformed (right side) for histogram and normal QQ Plot

4.3.2 Hotspots Analysis

ADDT measurements in Washington are consistent at yearly locations for data recording. Table 6 is a global autocorrelation summary of Moran's Index of Washington AADT up to 400, 500, 1000, and 2000. In summary, for AADT up to 400, all years but 2009 AADT yields positive Moran's Index and Z-scores values. The Z-score values were high for all the years except for 2016. The results are similar to Moran's Index and Z-score values for AADT up to 500 and 2000. However, the lowest positive Z-value for AADT, up to 500, was realized in 2013. For AADT up to 2000, the lowest Z-value corresponds to 2010. For AADT 1000, Moran's index and Z scores were positive for all the years. Thus, Moran's index and Z scores were primarily positive for the years and AADT levels under review.

Table 6

Moran's Index Summary for AADT up to 400, 500, 1000, and 2000 for Washington State

Washington State 400 AADT - Global Moran's Index summary								
	2009	**2010**	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**
Maran's Index	-0.142	0.127	0.053	0.107	0.083	0.066	0.103	0.033
Expected Index	-0.006	-0.006	-0.006	-0.006	-0.006	-0.006	-0.006	-0.006

Contd...

Variance	0.004	0.01	0.01	0.01	0.01	0.01	0.011	0.011
Z- Score	-2.073	1.335	0.586	1.112	0.888	0.721	1.035	0.367
Washington State 500 AADT - Global Moran's Index summary								
	2009	**2010**	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**
Maran's Index	-0.131	0.109	0.055	0.063	0.04	0.045	0.138	0.116
Expected Index	-0.005	-0.005	-0.005	-0.005	-0.005	-0.005	-0.005	-0.005
Variance	0.003	0.005	0.006	0.005	0.005	0.006	0.009	0.009
Z- Score	-2.262	1.584	0.798	0.917	0.613	0.645	1.525	1.293
Washington State 1000 AADT - Global Moran's Index summary								
	2009	**2010**	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**
Maran's Index	0.001	0.13	0.168	0.179	0.139	0.13	0.164	0.145
Expected Index	-0.002	-0.002	-0.002	-0.002	-0.002	-0.002	-0.002	-0.002
Variance	0.000	0.000	0.000	0.003	0.003	0.003	0.004	0.003
Z- Score	0.556	2.532	2.859	3.26	2.614	2.386	2.808	2.868
Washington State 2000 AADT - Global Moran's Index summary								
	2009	**2010**	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**
Maran's Index	-0.035	0.101	0.148	0.161	0.176	0.172	0.157	0.164
Expected Index	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001
Variance	0.001	0.001	0.001	0.001	0.001	0.001	0.002	0.002
Z- Score	-1.308	2.739	4.363	4.848	5.439	5.15	3.866	4.049

A negative Moran index indicates perfect clustering in the dataset even if values are dissimilar (alternating). Thus, the clustering is expressed as perfect dispersion. Also, there is a possible indication of spatial outliers or spatial heterogeneity. A negative Moran's Index of less than zero depicts negative autocorrelation. However, it is worth noting that the Moran's index value has no significance to the dispersion or clustering. As a global statistic, the Moran's Index says nothing about location. The negative Z – test statistics indicate somewhat probable clustering in the dataset. Positive Moran's Index indicates the perfect clustering of similar values and positive autocorrelation. However, the clustering may be due to either high or low values. Again, it is worth noting that the single value represents a whole spatial pattern and a specific location, which confirms the global scale at which Moran's Index explains clustering. The positive Z-test statistic is an indication of spatial clustering in the data. Based on hypothesis testing, Z-test statistics greater or less than 1.96 tend to accept the alternative and reject the null hypothesis, thus indicating spatial autocorrelation. It is characterized by the summaries corresponding to

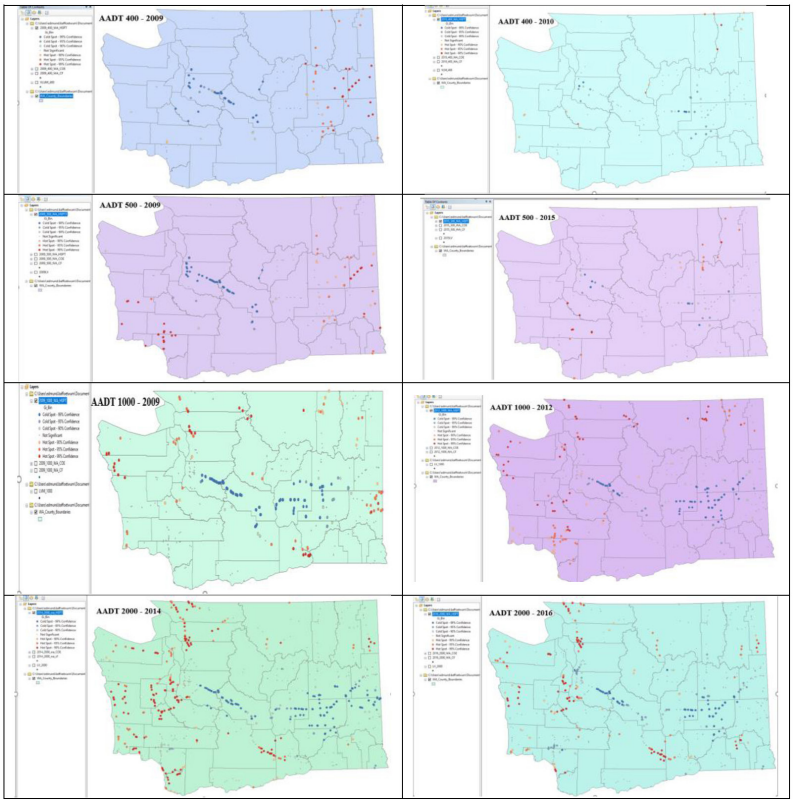

Figure 12

Gi* statistics hotspot analysis of sampled Washington State AADT datasets

2009 for AADTs of up to 400 and 500 and 2010 to 2016 for AADTs of up to 1000 and 2000. Moran's Index's global depiction and a single value in clustering classification are best explained and visualized with the Getis-Ord- Gi* statistics for the clustered data locations with 99%, 95%, and 90 % confidence. Figure 12 has AADT up to 400, AADT up to 500, AADT up to 1000, and AADT up to 2000.

Though there are coincidental hot spots, cold spots, and not significant locations for the output images derived from the dataset, each image has varying locations for the different classifications of hot, cold, and not significant spots with the corresponding confidence level. The output images show no location changes for the yearly measurements of AADT. From the output images, approximately similar data points are classified under not statistically significant classification corresponding to 90%, 95%, and 99% confidence levels for both hot and cold spots. Further

analysis may be needed to determine any correlation between the AADT levels and the possible increase in the hot, cold, and not significant spots.

5.0 Conclusions

To examine Montana, Minnesota, and Washington data, the study utilizes exploratory spatial data analysis (ESDA) and hotspot (Getis-Ord Gi statistics and Moran's Index) techniques. The data analysis reveals that Minnesota's state consistently varied data collection locations yearly. Accordingly, varying conclusions may be drawn on the various locations for this study.

Generally, the conclusions drawn are majorly similar. The ESDA emphasizes the description of pattern detection in data. ESDA is epitomized by identifying locations or spatial outliers, patterns determining spatial association, clusters or hotspots taller bars, and suggesting spatial procedures and spatial heterogeneity in other forms. Similarly, the hotspots emphasize the data spread, showing the high and low values in the datasets. Primarily, the dataset indicates that none of the datasets exhibited global trends that are a global spatial autocorrelation. However, on the other hand, there is some evidence of local spatial autocorrelation and randomness. Also, the AADT values show spatial heterogeneity in the spread. This study shows the general pattern or trends within all the datasets. It gives enough information to further analyses using sophisticated techniques. As the first step to insight into the datasets collected, the output guides decision-makers in these three states in optimizing data recorder stations and data collection coverage.

Data Availability Statements

Some or all data, models, or code that support the findings of this study are available from the corresponding author upon reasonable request. (List items)

Reference

[1] D. Albright, "History of Estimating and Evaluating Annual Traffic Volume Statistics," *Transportation Research Record*, no. 1305, pp. 103–107 (1991).

[2] American Association of State Highway and Transportation Officials (AASHTO), "Guidelines for Geometric Design of Low-Volume Roads," *AASHTO Journal*, 2nd ed. (2019).

[3] L. Anselin, S. Sridharan, and S. Gholston, "Using Exploratory Spatial Data Analysis to Leverage Social Indicator Databases: The Discovery of Interesting Patterns," *Social Indicators Research*, vol. 82, pp. 287–309 (2007). doi: 10.1007/s11205-006-9034-x

[4] D. Apronti, K. Ksaibati, K. Gerow, and J. J. Hepner, "Estimating traffic volume on Wyoming low-volume roads using linear and logistic regression methods," *Journal of Traffic and Transportation Engineering* (English Edition), vol. 3, no. 6, pp. 493–506 (2016).

[5] R. S. Bivand, "Exploratory Spatial Data Analysis," in Handbook of Applied Spatial Analysis, M. Fischer and A. Getis, Eds. Berlin, Heidelberg: Springer (2010). doi: 10.1007/978-3-642-03647-7_13

[6] E. Bormashenko, M. Frenkel, A. Vilk, I. Legchenkova, A. A. Fedorets, N. E. Aktaev, L. A. Dombrovsky, and M. Nosonovsky, "Characterization of Self-Assembled 2D Patterns with Voronoi Entropy," *Entropy*, vol. 20, art. no. 956 (2018). doi: 10.3390/e20120956

[7] F. Celebioglu and S. Dall'erba, "Spatial disparities across the regions of Turkey: an exploratory spatial data analysis," *Annals of Regional Science*, vol. 45, pp. 379–400 (2010). doi: 10.1007/s00168-009-0313-8

[8] F. Chen, C.-T. Lu, and A. P. Boedihardjo, "GLS-SOD: A Generalized Local Statistical Approach for Spatial Outlier Detection," in Proc. ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems (GIS), Washington, DC, USA (2010).

[9] P. Chen, S. Hu, Q. Shen, H. Lin, and C. Xie, "Estimating Traffic Volume for Local Streets with Imbalanced Data," *Transportation Research Record*, vol. 2673, no. 3, pp. 598–610 (2019). doi: 10.1177/0361198119833347

[10] B. Coll, S. Moutari, and A. H. Marshall, "Hotspots identification and ranking for road safety improvement: An alternative approach," *Accident Analysis and Prevention*, vol. 59, pp. 604–617 (2013).

[11] M. Cools, E. Moons, and G. Wets, "Investigating Effect of Holidays on Daily Traffic Counts: Time Series Approach," *Transportation Research Record: Journal of the Transportation Research Board*, no. 2019, pp. 22–31 (2007). doi: 10.3141/2019-04

[12] M. Cools, E. Moons, and G. Wets, "Investigating the Variability in Daily Traffic Counts Through Use of ARIMAX and SARIMAX Models Assessing the Effect of Holidays on Two Site Locations," *Transportation Research Record: Journal of the Transportation Research Board*, no. 2136, pp. 57–66 (2009). doi: 10.3141/2136-07

[13] Cornell Local Roads Program, "What is the Traffic Volume Cut Off Between High-Volume and Low-Volume," *Cornell Local Roads Program*, Apr. 4 (2019). [Online]. Available: https://www.clrp.cornell.edu/q-a/151-low-volume.html

[14] Dirt Artful, "Exploring Spatial Patterns in Your Data Using ArcGIS," (2018). [Online]. Available: https://dirtartful.com/exploring-spatial-patterns-in-your-data-using-arcgis/

[15] J. K. Eom, M. S. Park, T.-Y. Heo, and L. F. Huntsinger, "Improving the Prediction of Annual Average Daily Traffic for Nonfreeway Facilities by Applying a Spatial Statistical Method," *Transportation Research Record: Journal of the Transportation Research Board*, no. 1968 (2006). doi: 10.1177/0361198106196800103

[16] ESRI Support, Exploratory Spatial Data Analysis (2018).

[17] ESRI Support, "Exploratory Spatial Data Analysis (ESDA)," (2018). [Online]. Available: https://desktop.arcgis.com/en/arcmap/latest/extensions/geostatistical-analyst/exploratory-spatial-data-analysis-esda-.htm

[18] J. L. Gallo and C. Ertur, Exploratory Spatial Data Analysis of the Distribution of Regional Per Capita GDP in Europe, 1980–1995, *Research Report, Laboratoire d'Analyse et de Techniques Économiques* (LATEC) (2000). [Online]. Available: https://hal.science/hal-01527222

[19] C. Han and S. Song, "A Review of Some Main Models for Traffic Flow Forecasting," in Proc. *IEEE Intelligent Transportation Systems Conference*, vol. 1, pp. 216–219 (2003).

[20] A. Hassan, "What is Exploratory Spatial Data Analysis (ESDA)," Towards Data Science (2019). [Online]. Available: https://towardsdatascience.com/what-is-exploratory-spatial-data-analysis-esda-335da79026ee

[21] A. Hassan and J. Vijayaraghavan, Geospatial Data Science Quick Start Guide: Effective Techniques for Performing Smarter Geospatial Analysis Using Location Intelligence, Birmingham, UK: Packt Publishing (2019).

[22] ESRI Support, "How Hot Spot Analysis (Getis-Ord Gi*) Accident Analysis and Prevention works," *ArcGIS Desktop* (2018). [Online]. Available: http://desktop.arcgis.com/en/arcmap/10.5/tools/spatial-statistics-toolbox/h-how-hotspot-analysis-getis-ord-gi-spatial-stati.htm

[23] G. Keller and J. Sherar, Low-Volume Roads Engineering: Best Management Practices Field Guide. Produced for the U.S. Agency for International Development (USAID) in cooperation with USDA Forest Service, International Programs & Conservation Management Institute, Virginia Polytechnic Institute and State University (2003).

[24] S. Lee, "Neighborhood Effects," in *International Encyclopedia of Human Geography*, vol. 7, pp. 349–353 (2009). doi: 10.1016/B978-008044910-4.00480-6

[25] P. Legendre, "Spatial Autocorrelation: Trouble or New Paradigm?," *Ecology*, vol. 74, no. 6, pp. 1659–1673 (1993). doi: 10.2307/1939924

[26] J. Lian, X. Li, H. Gong, Y. Wang, and Y. Sun, "The Spatial Pattern Analysis of Economic Growth of JingJinJi Metropolitan Region," presented at the 18th International Conference on Geoinformatics, Beijing, China (2010). doi: 10.1109/GEOINFORMATICS.2010.5567867

[27] H. Liu, M. A. Lee, and J. Khattak, "Updating Annual Average Daily Traffic Estimates at Highway-Rail Grade Crossings with Geographically Weighted Poisson Regression," *Transportation Research Record*, vol. 2673, no. 10, pp. 105–117 (2019). doi: 10.1177/0361198119844976

[28] J. Menzies and J. J. M. Van der Meer, Past Glacial Environments – Geographic Information Systems and Glacial Environments, 2nd ed., Section 14.5.1, Spatial Analysis (2018). doi: 10.1016/C2014-0-04002-6

[29] S. F. Messner, L. Anselin, R. D. Baller, D. F. Hawkins, G. Deane, and S. E. Tolnay, "The Spatial Patterning of County Homicide Rates: An Application of Exploratory Spatial Data Analysis," *Journal of Quantitative Criminology*, vol. 15, pp. 423–450 (1999). doi: 10.1023/A:1007544208712

[30] A. Montella, "A Comparative Analysis of Hotspot Identification Methods," *Accident Analysis and Prevention*, vol. 42, pp. 571–581 (2010).

[31] Pennsylvania Department of Transportation (PennDOT), Traffic Count Policy – Dirt, Gravel, and Low Volume Road Maintenance Program (DGLVRP) (2014). [Online]. Available: https://www.dirtandgravel.psu.edu/sites/default/files/PA%20Program%20Resources/Low%20Volume%20Roads/LVR_Traffic_Count_Policy.pdf

[32] N. T. Ratrout, U. Gazder, and E. S. M. El-Alfy, "Effects of Using Average Annual Daily Traffic (AADT) with Exogenous Factors to Predict Daily Traffic," *Procedia Computer Science*, vol. 32, pp. 325–330 (2014). doi: 10.1016/j.procs.2014.05.431

[33] K. Rusche, Quality of Life in the Regions: An Exploratory Spatial Data Analysis for West German Labor Markets, CAWM Discussion Paper No. 10, pp. 1–25 (2008).

[34] S. C. Sharma, B. M. Gulati, and S. N. Rizak, "Statewide Traffic Volume Studies and Precision of AADT Estimates," *Journal of Transportation Engineering,* vol. 122, no. 6, pp. 430–439 (1996).

[35] S. C. Sharma, P. Lingras, F. Xu, and P. Kilburn, "Application of Neural Networks to Estimate AADT on Low-Volume Roads," *Journal of Transportation Engineering,* vol. 127, no. 5, pp. 426–432 (2001).

[36] W. N. Staats, Estimation of Annual Average Daily Traffic on Local Roads in Kentucky, M.S. thesis, Dept. of Civil Engineering, Univ. of Kentucky, Lexington, KY, USA (2016). [Online]. Available: https://uknowledge.uky.edu/ce_etds/36. doi: 10.13023/ETD.2016.066

[37] B. Van Arem, H. R. Kirby, M. J. M. Van Der Vlist, and J. C. Whittaker, "Recent Advances and Applications in the Field of Short-Term Traffic Forecasting," *International Journal of Forecasting*, vol. 13, no. 1, pp. 1–12 (1997).

[38] G. Ying Long, "Measuring the Spillover Effects: Some Chinese Evidence," *Papers in Regional Science,* vol. 79, pp. 75–89 (2000).

[39] S. E. Young, K. Sadabadi, P. Sekula, Y. Hou, and D. Markow, Estimating Highway Volumes Using Vehicle Probe Data – Proof of Concept, Golden, CO, USA: National Renewable Energy Laboratory (2018). NREL/CP-5400-70938.

[40] H. Yu, P. Liu, J. Chen, and H. Wang, "Comparative Analysis of the Spatial Analysis Methods for Hotspot Identification," *Accident Analysis and Prevention*, vol. 66, pp. 80–88 (2014). doi: 10.1016/j.aap.2014.01.017

Received July, 2022