GEOSTATISTICAL INTERPOLATION AND ANALYSES OF WASHINGTON STATE

AADT DATA FROM 2009 – 2016

A Thesis
Submitted to the Graduate Faculty
of the
North Dakota State University
of Agriculture and Applied Science

By

Kunle Meshach Owaniyi

In Partial Fulfillments of the Requirements
for the Degree of
MASTER OF SCIENCE

Major Department:
Construction Management and Engineering

January 2019

Fargo, North Dakota

North Dakota State University
Graduate School

Title

Geostatistical Interpolation and Analyses of Washington State AADT Data from 2009 – 2016

By

Kunle Meshach Owaniyi

The Supervisory Committee certifies that this ***disquisition*** complies with North Dakota State University's regulations and meets the accepted standards for the degree of

MASTER OF SCIENCE

SUPERVISORY COMMITTEE:

Dr. Eric Asa
Chair

Dr. Stephanie S. Day

Dr. Gary R. Smith

Approved:

July 10, 2019
Date

Dr. Jerry Gao
Department Chair

ABSTRACT

Annual Average Daily Traffic (AADT) data in the transportation industry today is an important tool used in various fields such as highway planning, pavement design, traffic safety, transport operations, and policy-making/analyses. Systematic literature review was used to identify the current methods of estimating AADT and ranked. Ordinary linear kriging occurred most. Also, factors that influence the accuracy of AADT estimation methods as identified include geographical location and road type amongst others. In addition, further analysis was carried out to determine the most apposite kriging algorithm for AADT data. Three linear (universal, ordinary, and simple), three nonlinear (disjunctive, probability, and indicator) and bayesian (empirical bayesian) kriging methods were compared. Spherical and exponential models were employed as the experimental variograms to aid the spatial interpolation and cross-validation. Statistical measures of correctness (mean prediction and root-mean-square errors) were used to compare the kriging algorithms. Empirical bayesian with exponential model yielded the best result.

ACKNOWLEDGMENTS

For everything that has a start must surely have an end and to this end, my sincere and humble appreciation goes out to Dr. Eric Asa, my advisor, for his support, advice and help towards the achievement of this degree. Special thanks and love go to my parents; Mr. and late Mrs. E.O. Owaniyi and siblings; Bosede, Busayo, Sanmi, Taiwo, Kehinde and Yinka for their continued prayer, support and love. Sincere appreciation to my committee members: Dr. Stephanie Day and Dr. Gary Smith for their continued help and guidance. To the faculty of the Department of Construction Management and Engineering and to NDSU, I dove my hat and say thank you for giving me the priviledge and the opportunity to fulfil my graduate degree ambition.To all my friends and families, who stood by me in all situations, I remain undoubtedly grateful and forever loyal. Thank you all. To God be the glory.

DEDICATION

This research work is dedicated to God Almighty, the beginning and the end who has been my sustenance and will always be. Also, in loving memory of my mum, late Mrs. Kehinde A. Owaniyi who passed unto glory before I could complete this phase of my life, continue to rest in peace mum.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. INTRODUCTION

1.1. Background

Annual Average Daily Traffic (AADT) data in the transportation industry today is an essential tool which is used in various fields such as highway planning, pavement design, traffic safety, transport operations, analyses, and policy-making. Minnesota Department of transportation defines AADT as the hypothetical estimation of the total number of vehicles using a specific section of roadway (in both directions) on any given day of the year (MnDOT, 2017). The Federal Highway Administration (FHWA) defines AADT as the “total volume of vehicle traffic of a road for a year divided by 365” (FHWA, 2001). This represents the total number of vehicles in a year divided by 365 and this is developed by using factors to adjust for the vehicle type, season (winter, summer, holidays) and which day of the week (weekend days). These all contribute significantly to the variability observed in AADT data.

Transportation planners and policy decision-makers rely heavily on AADT metrics to assess highway performance and guide their future planning and funding decisions. For instance, AADT data is used in the calculation of vehicle miles traveled (VMT), which in turn establishes the basis for distributing highway funds related to maintenance and safety (Staats, 2016). Transportation planning requires the use of accurate traffic data to produce estimates of traffic volume predictions over time and space (Shamo et al, 2015). Thus, annual average daily traffic (AADT) data is an important component of transportation design, operation, policy analysis, and planning. The use of traffic volume forecasting models for the characterization, analysis, and estimation of transportation data has shown to be a beneficial method for reducing high costs, overcoming spatial constraints, and limiting the errors associated with data collection and analysis in transportation planning (Shamo et al, 2015). Furthermore, AADT serves as the framework for estimating other transportation planning factors including crash rate predictions, vehicle emissions, and forecasting future travel demand (Reginald et al, 2016). For these reasons, the state department of transportation (DOT) planners and other affected stakeholders often take great efforts to collect and utilize this data (Staats, 2016).

Thus, AADT values are essential for numerous transportation planning and engineering tasks, such as evaluating the level of service, prioritizing capital investments, and assessing accident exposure rates (TII 2016). It is used as an input to scheme appraisal, environmental models, road planning studies, and pavement design (TII 2016). However, it is not feasible to continuously monitor every street in a community to obtain annual volumes, so AADT typically is approximated using short-duration (usually 48 h) ‘‘coverage counts’’ that are then multiplied by adjustment factors to account for daily and seasonal variations. This may be a source of uncertainty during the estimation of AADT values.

A common method for collecting coverage counts is to install pneumatic-tube counters temporarily at locations across the city and rotate the devices as needed (Lowry, 2014). Several other techniques and tools have been used in the estimation of AADT over the years. These techniques and tools include Machine learning techniques, Origin-Destination centrality-based method, Ordinary Linear regression, Florida Turnpike state model, Geographically Weighted Regression, Travel demand modeling, Kriging interpolation, Elasticity variables, Short Traffic Counts and Artificial Neural Network.

1.2. Problem Statement

The importance of annual average daily traffic data in the US and the world at large today cannot be over-emphasized. Various approaches have been used in predicting AADT in the past. Shamo et al. (2015) used three linear geostatistical interpolation kriging techniques in combination with the variogram models to predict AADT values at unsampled locations. Castro-Neto et al (2009) applied support vector regression (SVR) in predicting AADT. Sharma et al. (1999) applied neural networks (multilayered, feed-forward and back, propagation algorithms) to estimate AADT for 63 sites in the Minnesota state highway network.

Another approach is the use of exponential smoothing. Exponential smoothing (ES) techniques are relatively simple and effective methods for time series forecast for short-term horizons (De Lurgio, 1998). Time series modeling is based on the assumption that the historical values of a variable provide an indication of its value in the future (Box and Jenkins, 1970). Eom et al. (2006) used a spatial regression model to estimate AADT for roads in Wake County, North Carolina. The study used observed traffic counts from 200 of the county's 1,200 monitoring stations to estimate traffic volumes. Wang and Kockelman (2009) used spatial interpolation techniques to characterize and interpolate traffic counts in Texas and California. They came up with the conclusion that kriging performed far better than other options for spatial extrapolation - such as assigning AADT based on a point's nearest sampling site, which yields errors of 80%.

Several shortcomings and challenges were identified from the existing AADT estimation methods. One of the main challenges of accurate measurement of AADT as mentioned by Sababa in his 2016 article is having a complete, precise and reliable traffic data. In his research, he indicated that often the transportation agencies reported the problem of missing hourly volume from the permanent traffic count stations with a percentage of missing traffic data varying between 10% to 60%. Transport Infrastructure Ireland (TII) in its October 2016 publication identified a number of variables that need to be considered as they affect the quality of AADT estimation. These factors include geographical location, road type, the day of the week and seasonality. The use of the same approach for each data type has a risk of generating prediction errors and may result in inaccurate estimate adversely affecting aspects of the transportation design, planning, and policy-making processes (Wang and Kockelman, 2009).

Several AADT estimates have been carried out using the kriging method and researchers have looked at the differences between some of these kriging methods. Asa et al. (2012) compared linear and non-linear kriging methods for characterization and interpolation of soil data. Moyeed and Papirtz (2002) in their research, 'an empirical comparison of kriging methods for non-linear spatial point prediction' also compared linear and non-linear kriging methods. Other researchers have compared the kriging method with other methods, but none has compared linear, non-linear and bayesian methods of kriging which was carried out in this research.

1.3. Research Questions

With the help of this research, the following questions would be addressed

1. What are the methods of AADT estimation and prediction?
2. What are the factors that influence the accuracy of AADT estimation and prediction methods?
3. Which of the kriging methods is best used for AADT estimation and prediction?
4. What differences can be inferred between these kriging methods?

1.4. Research Objectives

The objectives of this study are as follows:

1. Perform a systematic literature review to identify the current approaches used in the estimation of annual average daily traffic data
2. Identify the factors that influence the accuracy of estimation and prediction methods
3. Investigate which of the kriging methods is best for AADT prediction.
4. Postulate a hypothesis to test if the prediction means an error of the kriging methods are the same or not.

1.5. Research Contribution

This research is part of the continual effort at adding to the body of knowledge by testing and comparing the effectiveness of the different kriging methods at estimating and predicting AADT values. This research broadly categorizes the kriging method into three and compare them to test which of the methods has the least prediction error in predicting AADT data. A systematic literature review was also conducted to document AADT research that has been performed to date as well as research gaps.

1.6. Research Organization

This thesis comprises of chapters 1 through 6. Chapter 1 introduces the research background, describes the problem to be solved and sets the research objectives to be achieved.

Chapter 2 provides a comprehensive review of the existing AADT estimation and prediction methods. The main purpose of this review is to research and document the methods that are being used for AADT estimation and to identify the factors that contribute to the accuracy of AADT estimation methods.

Chapter 3 describes the exploratory data analysis of the AADT data from the Washington Department of Transportation. Histogram, normal probability plot and run sequence plot were carried out to analyze the trend and characteristics of the data.

Chapter 4 describes the research methodology used. This consists of data exploration, structural analysis, cross-validation and hypothesis postulation.

Chapter 5 discusses the results, infer differences and evaluates the performance of the prediction tools in the analysis of AADT.

Finally, Chapter 6 draws conclusions, summarizes the main contribution of this research and provides recommendations for further studies.

1.7. Summary

This chapter introduces the research background, describes the problem to be solved and sets the research objectives to be achieved, the research contribution to the knowledge base and the research organization then followed.

2. ESTIMATING ANNUAL AVERAGE DAILY TRAFFIC DATA: A SYSTEMATIC LITERATURE REVIEW

2.1. Introduction

Several methods are being used today to estimate AADT values across the United States and the world in general. Most of these methods are being used by the transportation agencies were developed by Academic Universities and the Department of Transportations (DOT) working at improving the accuracy of AADT estimation. The electronic databases that were searched for this research were the American Society of Civil Engineers [ASCE], Web of Science [WOS], Science Direct [SD], and Engineering Village (EV). An additional database (web-based source) used i.e., Google.

Of the articles retrieved, 52 articles published from 1979 to 2017 met the inclusion criteria and were included in the final review. 8 AADT estimation and prediction method were identified in the 52 articles used. Descriptive and Anderson-Darling statistical methods were used to analyze the resulting articles and the results of the analysis showed the following as some of the commonly used methods of AADT estimation and prediction: Machine learning techniques, Origin-Destination centrality based method, Ordinary Linear regression, Florida Turnpike state model, Geographically Weighted Regression, Travel demand modeling, Kriging interpolation, Elasticity variables, Short Traffic Counts and Artificial Neural Network.

Table 2.1 below shows the AADT estimating methods identified, their proponents and application in the real world.

Table 2.1. Identified AADT Estimation Methods and Applications

AADT ESTIMATION METHODS	PROPONENTS	APPLICATIONS
Florida Turnpike state model	Florida Department of Transportation (2005)	Roads without traffic counts
Geographically Weighted Regression	Zhao and Park (2004)	County roads
Artificial Neural Network	Sharma et al. (2001), Sharma et al. (1999)	Rural roads
Kriging interpolation	Selby and Kockelman (2011), Eom et al. (2006), Shamo et al. (2015), Wang and Kockelman (2009)	All roads in Texas, Non-freeway roads in a county, Roadways with ATR data, All roads in Texas respectively
Travel demand modeling	Wang T. (2012), Wang et al. (2013), Zhang and Hanson (2009)	All roads in Florida, All roads in Florida, Low-class roads respectively
Ordinary Linear regression	Lu et al. (2007), Shen et al (1999)., Zhao and Chung (2001), Lowry and Dixon, Mohammad et al. (1998)	All roads in Florida, Off-system roads in Florida, County roads in Florida, Streets in an urban area, County roads in Indiana respectively
Origin-Destination centrality-based method	Lowry (2014)	Community roads
Support vector regression with data-dependent parameters	Castro-Neto Jeong, and Han (2009)	Both rural and urban roads (25 counties) in Tennessee

Estimating and predicting values in a particular area of interest is challenging and sometimes complicated due to the unavailability of data from the chosen area. Consequently, the AADT estimation process becomes complex when considering these unknown factors which include the day of the week, season and location amongst others. This chapter provides a systematic review of the existing AADT estimation and prediction methods with the aim of identifying different methods of AADT estimation and their different applications or usage at different locations.

2.2. Previous Studies

Traffic forecasting involves the application of computational, intelligent, statistical and mathematical techniques to model and estimate key parameters such as annual average daily traffic (AADT), design hour volumes (DHV), directional design hour volume (DDHV), and other variables that are inputs in transportation planning, design, operations, and policy analysis (Shamo et al. 2015). Many transportation resources, such as the ASSHTO guidelines for traffic data programs (AASHTO, 1992), outline many transportation engineering activities that require estimates of traffic volume demand parameters such as the annual average daily traffic.

Many AADT prediction studies have been published over the years due to the importance of knowing the traffic demand and using it to design and plan transportation operations. Manoel et al. (2009), identified two main categories of AADT prediction studies from literature: current-year and future-year AADT estimation studies. In the current year approach, the AADT for a particular year (usually current-year) is estimated using predictor variables associated with that year. The AADT values for future years are estimated based on the AADT from previous years, and external variables are also sometimes used (Manoel et al. 2009).

Manoel et al. (2009) evaluated the performance of a modified version of the Support Vector Machine for Regression (SVR) technique in forecasting AADT one year into the future without using any external (predictor) variable. The proposed procedure computes the SVR prediction parameters based on the spreading of the training data, consequently, the proposed method used was termed SVR with data-dependent parameters (SVR-DP), which in order to evaluate its performance, SVR-DP was compared to Holt exponential smoothing (Manoel et al. 2009). The modified SVR uses data-dependent parameters in order to reduce computational time and to achieve better predictors. A comparison of the results showed that SVR-DP outperformed the

OLS-regression technique, which is commonly used for future-year AADT forecasting purposes (Manoel et al. 2009).

The SVR-DP also performed better than the Holt's ES, but one can as well argue that both techniques performed in likewise manner (Manoel et al. 2009). The g performance of SVR-DP can be attributed to the remarkable characteristics of SVR and the incorporation of a data-dependent procedure for computing SVR parameters, this inevitably reduces uncertainty related to parameter selection and computation time (Manoel et al. 2009). They concluded that the SVR-DP technique provides an accurate forecasting technique where no external explanatory variable is used. This can be advantageous because the inclusion of external variables might not be feasible.

Sharma et al. (1996) researched the accuracy of AADT estimates using traffic data from 63 ATR sites in Minnesota. These ATR sites used were grouped into five clusters based on their characteristics. Two of the five groups represented the regional routes with low seasonal traffic, one represented the average rural routes, and the rest groups represented routes serving recreational areas. The results of the study showed predicted AADT values to be off by 11% in 95% of the cases with "regional routes serving commuters and business trips" enjoying the smallest AADT estimation errors and heavy-traffic rural routes serving recreational areas suffering the highest errors. They concluded that it is important to assign each site to its correct group; incorrect assignment carries the greatest potential for significant estimation error. They also found that estimation error fell only moderately with count duration, from 16.5% at 24 hours to 13.13% at 72 hours.

Lam and Xu (2000) also analyzed data at 13 locations and found that neural networks consistently performed better than regression analysis, and 8-hour counts (if AADT is estimated from something less than a 24-hour interval) are most appropriate. Jiang et al. (2006) used a weighted combination of past and present counts along 122 highway segments over a 10-year period to estimate AADT. They concluded that accuracy improved when the averaging was applied on a large scale and that the number of short-period traffic counts (SPTCs) could be reduced on many segments.

Eom et al. (2006) used spatial statistics to improve AADT prediction along non-freeway facilities in Wake County, North Carolina. They found that a model which takes both spatial trend and spatial correlation into account provides better predictions for locations where no observed count data existed. Wang and Kockelman (2009) used spatial interpolation techniques to characterize and interpolate AADT data in Texas and California. They concluded that kriging is a promising way to explore spatial relationships across a wide variety of data sets, including, for example, pavement conditions, traffic speeds, population densities, land values, household incomes, and trip generation rates. They further concluded that further refinements, including spatial autocorrelation functions based on network (rather than Euclidean) distances and inclusion of far more explanatory variables, are possible, and will further enhance estimation.

Utmost recently, Shamo et al. (2015) applied the geostatistical procedure of kriging to predict AADT values at unmeasured locations with the aim to reduce the extra cost imposed on the current Highway Performance Monitoring System (HPMS). The modeling techniques they used applies position and spatial relationships to estimate probable values of AADT at unmeasured locations considered. They came with the same conclusion as Chu (1993) that even though there are good iterative graphical programs, the user would do better than a sophisticated fully automatic fitting procedure. This implied that the same kriging method could not be used for the same data type from year to year due to the changing dynamics of AADT attribute (Shamo et al. 2015).

2.3. Research Methodology

This study was performed as a systematic literature review (SLR) to search electronic databases to retrieve relevant literature and papers. Systematic reviews aim to identify, evaluate and summarize the findings of all relevant individual studies, thereby making the available evidence more accessible to decision-makers (Centre for Reviews and Dissemination, 2008). The SLR allowed for an evidence-based approach to identify, select, analyze, and synthesize data for a specific research topic by documenting all the steps. (Cook et al. 1997; Tranfield et al. 2003). The flow chart of the processes used for this research is shown below.

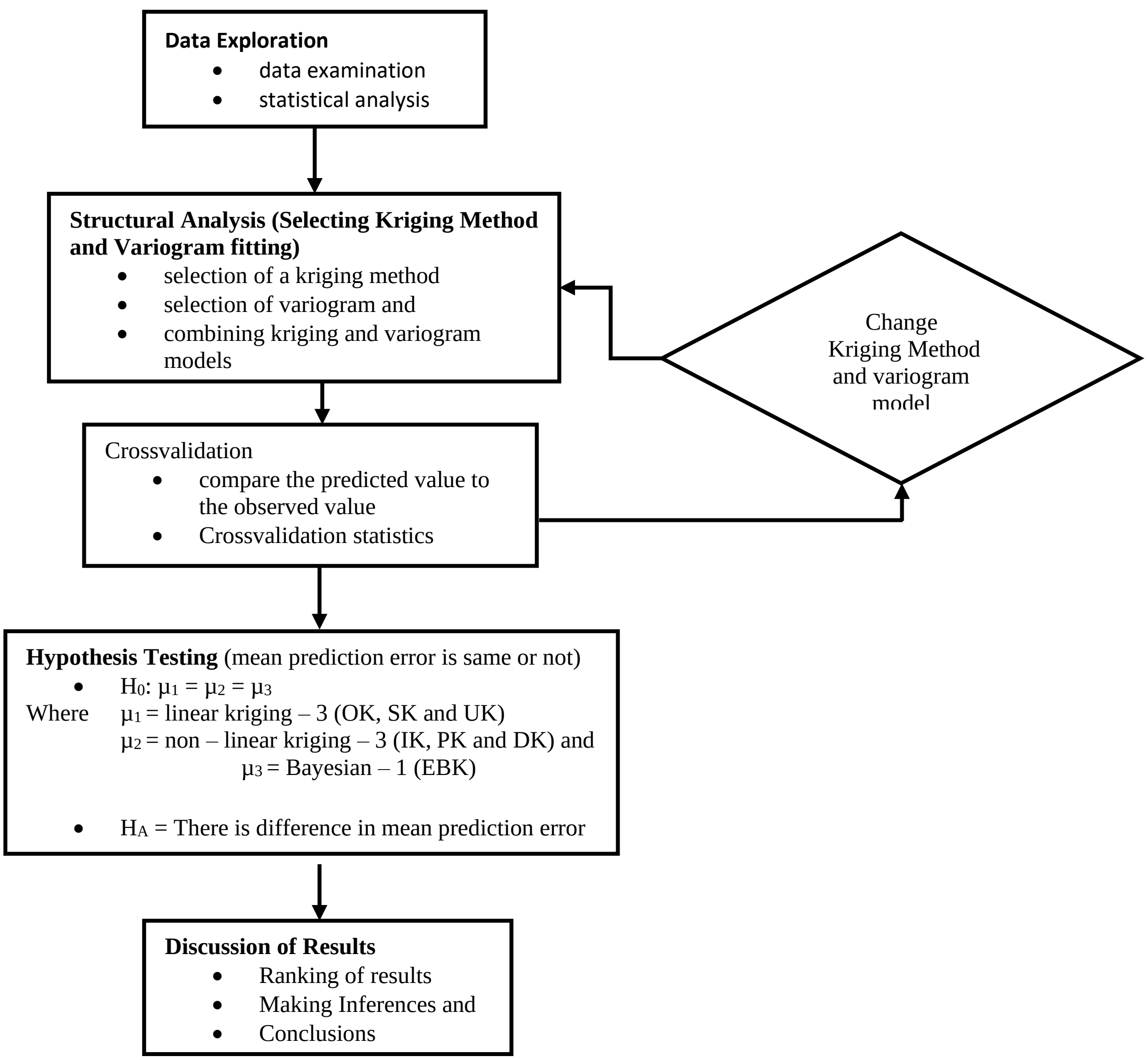

Figure 2.1. Research Method Flowchart

2.3.1. Databases Searched

Five major databases were searched, and resources searched about AADT estimation or prediction were selected. The approach and databases searched are shown in Figure 2.2. The systematic review was performed by searching a combination of databases (such as the American Society of Civil Engineers [ASCE], Web of Science [WOS], Science Direct [SD], and Engineering Village (EV). The additional database was other web-based sources (i.e., Google).

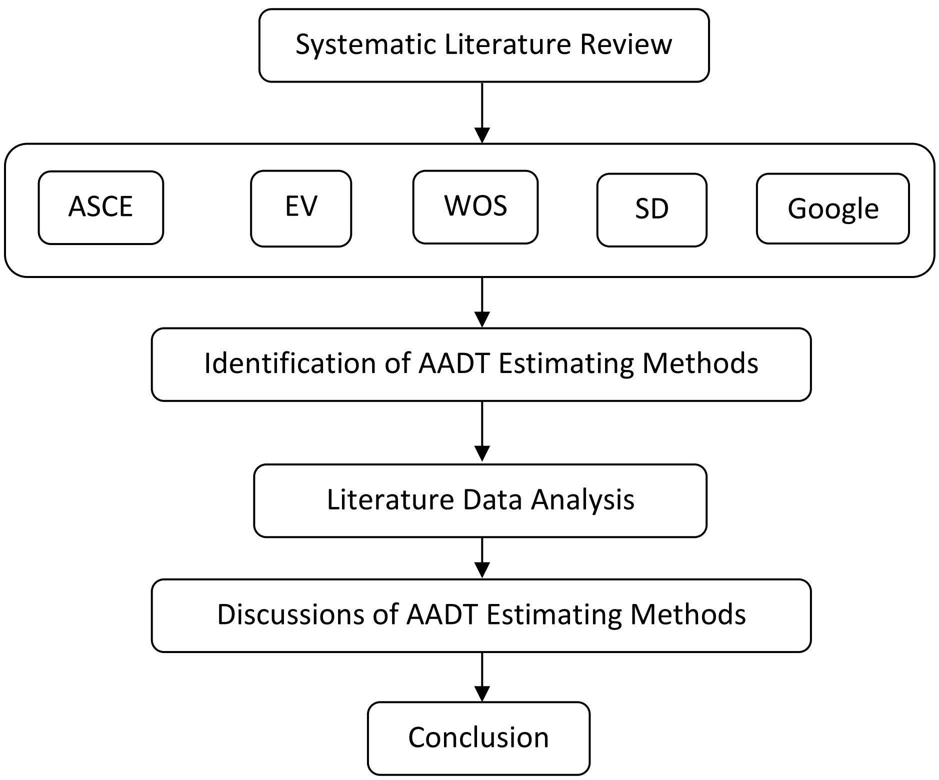

Figure 2.2. The Systematic Literature Review Process

The databases were selected to ensure that a broad range of published literature was retrieved on AADT estimation and prediction. The ASCE, WOS, EV, and SD are major electronic databases involving engineering transportation projects and are highly regarded by the academic community. In order to cross-reference sources, the Google search engine was also included to search for others which might be difficult to find somewhere else and to identify their contributions. The same search sequence was used for all the databases, but the words were slightly modified to suit the format of each particular database in order to avoid missing any important or new information.

To capture all the information, searches were performed directly on the databases. The following keywords were used when the search was conducted: 'annual average data traffic,' 'AADT,' 'annual average data traffic estimation,' 'traffic forecasting,' and 'AADT forecasting.' Other keywords list included 'AADT prediction methods,' 'AADT calculations' and 'estimating AADT.'

The articles were retrieved from different journals and government reports based on the criteria formulated as described in 2.3.2 below. The procedure involved reading the abstracts, and in cases where the information was not available, the entire paper was read. The wide range of databases together with the use of predetermined terms searched was aimed at performing a review to generate a comprehensive list of articles that might be useful for this research. After a document has been retrieved, AADT estimating or prediction methods are used to examine the literature to check its suitability for the review.

2.3.2. Search Inclusion and Exclusion Criteria

The criteria used for inclusion and exclusion of studies in the systematic literature review includes the following. Studies in English were considered from conference papers and peer-reviewed journals (abstracts and full papers), published from 1979 to 2017. The papers included must have focused on AADT estimations and predictions and be available for download. Studies not in English, or not explicitly related to AADT estimations and predictions, or are not related to the review questions were excluded. Also excluded were prefaces, editorials, and poster sessions. Published research which had been peer-reviewed was not independently assessed for study quality and was assumed to be of good quality and coded accordingly. The following is a short description of each of the databases used.

American Society of Civil Engineers (ASCE): According to the website, The American Society of Civil Engineers (ASCE) represents more than 150,000 members of the civil engineering profession in 177 countries. Founded in 1852, ASCE is the nation’s oldest engineering society. It stands at the forefront of a profession that plans, designs, constructs, and operates society’s economic and social engine – the built environment – while protecting and restoring the natural environment. ASCE operates with these goals in mind which are:

- An ever-growing number of people in the civil engineering realm are members of, and engage in, ASCE.
- Civil engineers develop and apply innovative, state-of-the-art practices and technologies.
- All infrastructure is safe, resilient, and sustainable.
- ASCE advances the educational and professional standards for civil engineers.
- The public values civil engineers’ essential role in society.
- ASCE excels in strategic and operational effectiveness.

The database can be accessed through https://ascelibrary.org/journals.

Engineering Village (EV): According to the website, Engineering Village takes engineering research to the next level with a comprehensive database that includes the most authoritative engineering resources available to answer today’s most timely questions—from theoretical to applied, and basic to complex. Academics, government institutions, business researchers and practicing engineers gain an immediate advantage with access to today's most authoritative engineering research, with enhanced user features, that provide deep insight into published engineering work and related disciplines. It offers access to 12 engineering literature and patent databases providing coverage from a wide range of trusted engineering sources. The databases have been carefully selected to provide both breadths as well as the depth of content. The database can be accessed through https://www.elsevier.com/solutions/engineering-village.

Web of Science (WOS): According to the website, the web of science group powers our integrated suite of research intelligence and workflow solutions to help people at every stage of work in a more open, seamless and connected way. The Web of Science Core Collection is the most authoritative global citation index – consistently and continuously curated by an independent team of full-time editors to include only the highest quality and most impactful journals across 254 subject areas. Whether providing pinpoint access to the content one needs, helping one find the best place to publish, or communicating one's research to the world, WOS make sure one has the essential information, analytics, and tools one requires to succeed. The database can be accessed through https://clarivate.com/products/web-of-science/.

ScienceDirect (SD): According to the website, ScienceDirect is built on the widest range of trusted, high-quality, interdisciplinary research. It helps find answers to the most pressing research questions, stay on top of your field and gain in-depth insights into trending research topics as one takes next steps in discovery. This platform gives access to a large database of scientific and medical research. It hosts over 12 million pieces of content from 3,500 academic journals and 34,000. The journals grouped into four main sections include Physical Sciences and Engineering, Life Sciences, Health Sciences, and Social Sciences and Humanities. The database can be accessed through https://www.sciencedirect.com/.

2.3.3. The Literature Selection Process

Figure 2.3 below shows the flow chart of the systematic review which shows the number of papers identified at the various stages of the searches and reviews. From the search and reviews, 601 articles were identified from the 4 databases, 53 abstracts and full-text articles were assessed for eligibility, and 21 abstracts and full papers were considered in the analysis.

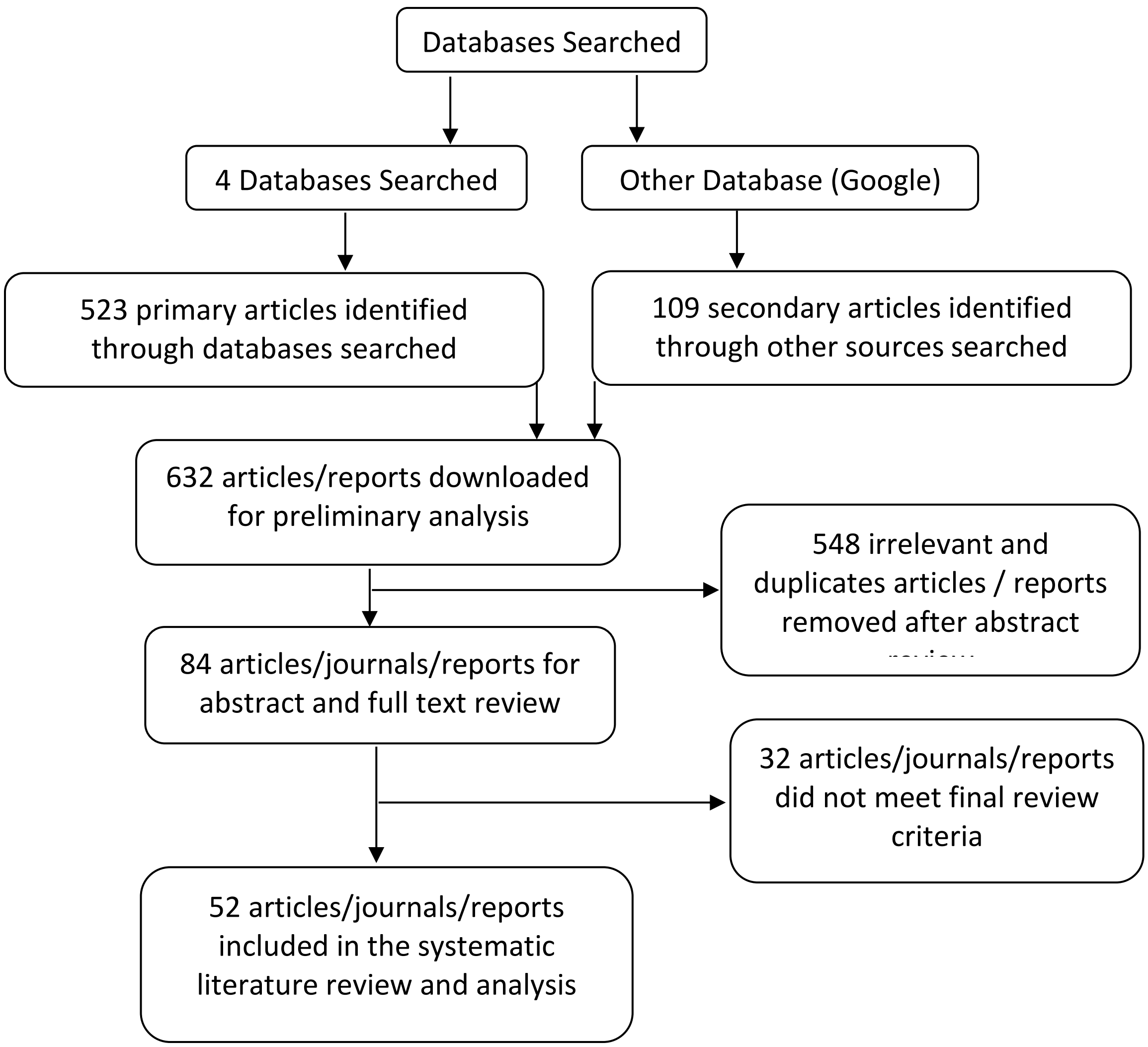

Figure 2.3. The Search, Review and Selection Process of The Systematic Literature Review

A number of papers were not selected because the studies did not specifically discuss AADT estimating or prediction or other related terms to AADT estimating and therefore does not meet the criteria for inclusion. The selected articles were from the 1979 to 2017 period. Table 2.2

below showed the tabulation of papers at the various selection stages from the databases used for this research study. The results showed that the ASCE database had 22 publications which were the highest number, followed by the Google database with a total of 15 publications. SD and EV databases had 9 and 6 respectively. WOS did not produce anything useful for the research after the full review.

Table 2.2. Paper Reviewed at Different Stages of Analysis

Databases	Articles identified	Duplicates and not relevant articles	Articles after duplicate removal	Articles removed after abstract review	Articles for the abstract and full-text review	Articles which did not meet criteria	Articles included in the analysis
ASCE	261	84	177	148	29	7	22
EV	23	8	15	7	8	2	6
SD	212	36	176	162	14	5	9
WOS	12	0	12	10	2	2	0
Other	109	48	61	40	21	6	15

The research papers, journals, and books were selected based on their discussion of AADT estimating or prediction methods. The final papers used for this research study totaled 52 scientific journal articles including dissertation and thesis. Table 2.3 presents an overview of the publications used in this research and the factors each focused on. These publications below are few examples out of the 52 publications that were used for the entirety of the thesis.

Table 2.3. An Overview of the Publications Used for the Research Study

Author(s)	Year	Journal/book title	Type of Study	Remarks/Assumptions
Shashank G; Atul M; and Kara M	2007	Estimates Of AADT: Quantifying the Uncertainty	Variations in AADT estimation errors are investigated across roadway and area types, for both Minnesota and Florida automatic traffic recorder (ATR) sites.	The analytical results of the investigation suggested a variety of recommendations for agencies seeking to reduce and appreciate errors in their AADT estimates. These include sampling in spring and summer months (on weekdays), exercising greater caution with counts on multilane and low-AADT roadways, pursuing appropriate site assignment to ATR groups, and recognizing the effects of distance to the sampling site. With adequate attention, (average) errors in AADT estimates can probably be reduced to the 10 percent level. Nevertheless, this still will have an impact on investment decisions, crash rate calculations, travel demand model validation, and other analyses.
Satish Sharma, Pawan Lingras, Fei Xu, and Peter Kilburn	2001	Application of Neural Networks to estimate AADT on low-volume roads	AADT estimation errors resulting from various durations and frequencies of counts are analyzed by computing average and percentile errors.	The results of the study indicated a clear preference for two 48-h counts as compared to other frequencies (one or three) or durations (24- or 72-h) of counts.

Table 2.3. An Overview of the Publications Used for the Research Study (Continued)

Author(s)	Year	Journal/book title	Type of Study	Remarks/Assumptions
Miguel Figliozzi; Pam Johnson; Christopher Monsere; and Krista Nordback,	2014	Methodology to Characterize Ideal Short-Term Counting Conditions and Improve AADT Estimation Accuracy Using a Regression-Based Correcting Function	The proposed methodology for the analysis of AADT estimation errors using regression models to estimate a correcting function that accounts for weather and activity factors	The results indicated that the proposed methodology is simple and useful for finding ideal short-term counting conditions and improving AADT estimation accuracy.
Y. F. Tang; William H. K. Lam; and Pan L. P. Ng	2003	Comparison of Four Modeling Techniques for Short-Term AADT Forecasting in Hong Kong	The historical data (1994–1998) and available current-year data for 1999 partial daily flows are the input data used for model development. The results of the four models were compared with the real data for validation. The daily flows estimated by the four models were used to calculate the AADT for the current year of 1999	Based on the comparison results, the GML model appears to be the most promising and robust of these four models for extensive applications to provide the short-term traffic forecasting database for the whole territory of Hong Kong.

Table 2.3. An Overview of the Publications Used for the Research Study (Continued)

Author(s)	Year	Journal/book title	Type of Study	Remarks/Assumptions
Mark R. McCord; Prem K. Goel; Zhuojun Jiang; and Patrick Bobbit	2002	Improving AADT and VMT Estimates with High-Resolution Satellite Imagery: Simulated Analytical Results	Developed computer software to simulate AADT and VMT estimation errors with and without the use of satellite data.	Results indicate that adding satellite-based data to ground-based data would decrease AADT and VMT estimation errors and allow for a substantial reduction in ground-based samples for a large range of inputs. Other numerical results show that using satellite-based data could lead to improved estimates while reducing the number of the permanent automatic traffic recorders used to determine temporal adjustment factors in traffic monitoring programs and that there are decreasing marginal benefits from increased satellite supply.
Ehsan Bagheri; Ming Zhong; and James Christie	2015	Improving AADT Estimation Accuracy of Short-Term Traffic Counts Using Pattern Matching and Bayesian Statistics	Two pattern-matching methods and their combination with Bayesian statistics were proposed and tested using permanent traffic counter (PTC) data from Alberta, and their results were compared to the method.	Study results show that, compared to the FHWA method, the proposed methods reduce the 95th percentile of the absolute percent AADT estimation errors (P95) by 0.5 to 31.9 when applied to different testing sites.

Table 2.3. An Overview of the Publications Used for the Research Study (Continued)

Author(s)	Year	Journal/book title	Type of Study	Remarks/Assumptions
Mei Chen; Jingxin Xia; and Alejandro Anaya	2004	Estimating Average Daily Traffic Using ITS Data	In the study, data collected by two regional ITS deployments were screened to weed out erroneous data.	It was observed that, while ITS detector data might not be as accurate as those collected by automatic traffic recorders used in planning applications, it is possible to obtain relatively accurate estimates based on short-term traffic data that has high quality.
Satish C. Sharma; Brij M. Gulati; and Samantha N. Rizak	1996	Statewide Traffic Volume Studies and Precision of AADT Estimates	Investigated in the paper is the statistical precision of annual average daily traffic (AADT) estimates resulting from short period traffic counts (SPTC).	It was found that AADT estimation errors are very sensitive to assignment effectiveness. The study results suggested that highway agencies should put more emphasis on sample site assignments to correct automatic traffic recorder (ATR) groups than on the duration of the count, i.e., whether it be a 24-, 48-, or 72-hr traffic count
Ehsan Bagheri; Ming Zhong; and James Christie	2011	Improving Group Assignment and AADT Estimation Accuracy of Short-term Traffic Counts using Historical Seasonal Patterns	Two pattern matching methods are proposed and tested with the data from a permanent counter on a winter recreational road in Alberta, Canada	It is found that the resulting 95th percentile (P95) AADT estimation errors of the two methods are 10.5% and 8.4% respectively, with contrast to 62.1% from the FHWA method

Table 2.3. An Overview of the Publications Used for the Research Study (Continued)

Author(s)	Year	Journal/book title	Type of Study	Remarks/Assumptions
Benedict Shamo; Eric Asa; and Joseph Membah	2015	Linear Spatial Interpolation and Analysis of Annual Average Daily Traffic Data	Applied three different linear kriging techniques [simple kriging (SK), ordinary kriging (OK), and universal kriging (UK)] and five variogram models (nugget effect, spherical, exponential, Gaussian, and power) to characterize and interpolate the annual average daily traffic of Washington State	Results from the study suggest that using the same combination of kriging and variogram algorithms to characterize and interpolate different AADT datasets (2008, 2009, and 2010) could lead to suboptimal results
Zhuojun Jiang; Mark R. McCord; and Prem K. Goel	2006	Improved AADT Estimation by Combining Information in Image- and Ground-Based Traffic Data	Proposed a weighted combination of both earlier year coverage counts and a current year image containing traffic information for AADT estimation	The accuracy was markedly improved and stable over a large range of important input values. The demonstrated improvements in accuracy and the ease of using this method with existing data are great enough that field testing should now be considered.
Venkata Ramana Duddu; and Srinivas S. Pulugurtha	2013	The principle of Demographic Gravitation to Estimate Annual Average Daily Traffic: Comparison of Statistical and Neural Network Models	The paper focuses on the application of the principle of demographic gravitation to estimate link-level annual average daily traffic (AADT) based on land-use characteristics	The results obtained indicate that statistical and neural network models ensured significantly lower errors when compared to outputs from the traditional four-step method used by regional modelers

2.4. Results

A summary of the results of the review questions is presented.

2.4.1. Methods of AADT Estimation and Prediction

Review question 1 is related to identifying the methods of AADT estimation that have been used in the industry to date. After conducting the literature search and analyzing the results, eight methods were identified as the methods for AADT estimation and prediction. Table 2.4 presents the breakdown and frequency of the identified methods that were used for AADT estimation and prediction. The significance of a factor was weighed by the number of times the factor occurred in the literature. The factors with the largest integer are ranked 1, the second highest as 2, and others. A total of five methods received just one author's opinion, which was ranked as the lowest rank of 4.

The findings from the analysis identified different methods for AADT estimating out of which ordinary linear regression occurred most with kriging interpolation following closely. Four other methods were used which are, Florida turnpike state model, geographically weighted regression, artificial neural network, support vector regression with data-dependent parameters, and origin-destination centrality-based method occurred least.

Table 2.4. Identified AADT Estimation and Prediction Methods (8 Methods)

	Author(s) / Year	Shamo et al (2015)	Lowry (2014)	Castro-Neto et al. (2015)	Shen et al (1999)	Zhao and Chung (2001)	Lowry and Dixon	Mohammad et al (1998)	Wang and Kockelman	Wang et al (2013)	Zhang and Hanson (2009)	Lu et al. (2007)	Selby and Kockelman	Sharma et al. (2001)	Florida DOT (2005)	Wang T. (2012)	Zhao and Park (2004)	Eom et al (2006)	Freq.
	Methods Identified																		
AADT Estimating Methods	Florida Turnpike state model														x				4
	Geographically Weighted Regression																x		4
	Artificial Neural Network													x					4
	Kriging interpolation	x							x				x					x	2
	Travel demand modeling									x	x					x			3
	Ordinary Linear regression				x	x	x	x				x							1
	Origin-Destination centrality-based method		x																4
	Support vector regression with data-dependent parameters			x															4

Table 2.4 showed the 8 identified methods of AADT estimation and prediction that was identified from the 52 publications that were used for this research.

2.4.2. Factors That Influence the Accuracy of AADT Estimation and Prediction Methods

Review question 2 is related to identifying the factors that influence the accuracy of AADT estimation and prediction methods. After conducting the search, a number of factors that influence the accuracy of AADT estimation and prediction were identified and indicated in Table 2.5 below.

Table 2.5. Factors that Influence AADT Accuracy

	Author(s) / Year	Transport Infrastructure Ireland (216)	Sababa (2016)	Paul (2016)	Ohio Department of transportation (2014)	Shashank et al (2007)	Federal Highway Administration (2018)	Venkata and Srinivas (2013)	Miguel et al (2014)	Riccardo et al (2014)	Satish et al (1996)	Eom et al (2006)
	Factors Identified											
Influencers of AADT Estimating and Prediction Methods	Geographical location	x						x		x	x	
	Road Type	x				x	x	x	x	x	x	
	Day of Week	x				x		x				
	Seasonality	x			x				x	x	x	x
	Missing hourly volume		x							x	x	
	Equipment theft			x		x	x	x				
	Equipment damage/vandalism			x								
	Human error			x								

Transport Infrastructure Ireland (TII) in its October 2016 publication identified a few variables that need to be considered which affect the quality of AADT estimation. These include geographical location, road type, day of week and seasonality. One of the main challenges of accurate measurement of AADT as mentioned by Sababa in his 2016 article is having a complete, precise and reliable traffic data. In his research, indicated that often the transportation agencies reported the problem of missing hourly volume from the permanent traffic count stations with a percentage of missing traffic data varying between 10% to 60%. In addition to these factors mentioned earlier, Paul (2016) identified equipment theft, damage, vandalism, and human error as some of the challenges that influenced the accuracy of AADT estimation.

2.5. Discussions

As authorized by Congress (23 U.S.C. 502(h)), every state in the United States submits data from their highway performance monitoring system (HPMS). This is for the purpose of biennial conditions and performance report of the future highway investment needs of the nation (U.S. Dept. of Transportation FHWA 2001). A primary goal of the HPMS traffic data collection effort is to provide a statistically valid estimate of total annual vehicle distance traveled (VDT) (U.S. Dept. of Transportation FHWA 2001). AADT estimation could tend to be tedious and not as straight forward as it seems. It requires some insight into the data which does take time and effort with resources to get it done. The aim of this section is to discuss the methods of AADT estimating that was identified in the literature review and to discuss in depth the kriging methods that are being used to date. These methods are widely used across the transportation field today for AADT estimation and prediction.

2.5.1. Ordinary Linear Regression

Regression analysis may be one of the most popular methods to estimate AADT (Yang et al. 2011). Many papers choose different variables (AADT data) that contribute to AADT and then uses an ordinary linear regression method to estimate or predict the outcome of these variables in connection to AADT. Mohammad et al. (1998) incorporated relevant demographic variables for county roads into a traffic prediction model. Xia et al. (1999) discovered that roadway characteristics such as the number of lanes, functional classification, and type of area can be used to predict AADT values for non-state roads in urbanized areas in Florida. Zhao and Chung (2001) well developed and compared four multiple linear regression models using geographic information system technology. They used four sets of independent variables and these are the roadway characteristics, socioeconomic characteristics, expressway accessibility, and accessibility to regional employment centers. Zhao and Park (2004) used a geographically weighted regression (GWR) method to estimate AADT values. They argued that the GWR models were comparable to ordinary least square models. Kingan and Westhuis (2006) used robust regression methods for AADT forecasting. Once the variables (AADT data) are collected no matter how large. A critical step that must not be missed is to keep significant variables and exclude the non-significant variables in the final model.

As mentioned by Zhao and Chung (2001), although most variables were statistically significant, few added enough explanatory power to be practical and useful. So, it is important to find a criterion to maximize the explanatory power. Thus, it is fundamentally important to do the variable selection in AADT estimation effectively to ensure the efficiency of estimation and accuracy of prediction. Yang et al. in their 2011 publication noted that using the t-test, f-test, or selection of the best model according to the Akaike information criterion (AIC) and bayesian information criterion are some of the commonly used methods that are used to select significant variables. Fan and Li (2001) mentioned that these stepwise deletion procedures may suffer stochastic errors inherited in the multiple stages, and there is no theory on the validity of these multiple selecting steps. Smoothly clipped absolute deviation penalty (SCAD) variable selection was used based on regression models to select the variables that are significant and left out the non-significant variables in their four groups of 19 variables that they collected at the very beginning (Yang et al; 2011).

Once the process of selection and elimination is completed, an estimate of the unknown regression coefficients is then performed. The attractive point of this variable selection for the regression procedure is that it is not only critical for local AADT estimation but also flexible to be applied to other related areas, where the relevant statistics play a key role, e.g., very large-scale intelligent transportation systems and networks (Yang et al. 2011).

2.5.2. Kriging Interpolation

Kriging is a geostatistical interpolation technique that considers both the distance and the degree of variation between known data points when estimating values in unknown areas (ESRI

2018). Kriging is the generic name adopted by geostatisticians for a family of generalized least-square regression algorithms in recognition of the pioneering work of Danie Krige (Goovaerts 1997). A kriged estimate is a weighted linear combination of the known sample values around the point to be estimated (ESRI 2018). This is an advanced geostatistical procedure. It creates a projected surface from a dispersed set of points with z-values (ESRI 2018).

According to ESRI, the direction or distance in between a sample point replicates a spatial linking that can be used to describe variation in the surface as assumed by kriging. A fitted mathematical function is used by kriging tool to a stated number of points, or all points inside a stated area, to determine the output value for each of the location (ESRI 2018).

"Kriging involves a multistep process and this process includes exploratory statistical analysis of data, variogram modeling of data, creating an output surface, and if need be, exploring the variance output surface" (ESRI 2018). Though it was often used in soil science and geology before over the years, its application has spread to various fields including transportation field (ESRI 2018).

Kriging uses kriging weights ($\lambda\alpha$), which are derived from a covariance function (variogram). The variogram is a fitted function used to express the relationship between the known and unknown data points. The variogram approach to developing kriging weights is similar to inverse distance weighting except that in the case of kriging weights, the weights are modeled by the best-fitted variogram (Shamo et al. 2015). One of the major advantages of kriging is that it has the ability to provide estimation/prediction errors. This estimation/prediction error helps in comparing kriging to other methods and it also serves as a basis for stochastic simulation of functions that could represent the relationship between the measured and unmeasured AADT data points. Clark (1979) identified another benefit of kriging which is, its ability to compensate for data clustering by assigning to individual data points within a cluster less weight than isolated data points. This is unlike other predictors like Sichel t estimator, which requires the probability distribution of the samples to be lognormal (Clark 1979).

The difference between kriging and other linear estimation methods is that it is aimed at minimizing the error variance (Shamo et al. 2015). Laslett et al. (1987) compared kriging with other techniques of interpolation and showed that kriging was the only methodology that performed reliably in all circumstances. Kriging has been successfully used in spatial prediction of soil properties (Burgess and Webster 1980), mineral resources, petroleum property evaluation, aquifer interpolation (Doctor 1979), and soil salinity through interpolation of electrical conductivity measurements (Oliver and Webster 1986), meteorology, and forestry.

2.5.2.1. Kriging Methods Used for AADT Estimation

Kriging methods used for AADT estimation is broadly categorized into three. These are:

i. linear kriging methods: This comprises of universal, simple, and ordinary kriging methods
ii. non-linear kriging methods: This comprises of disjunctive, indicator and probability kriging methods and
iii. bayesian kriging method: This is empirical bayesian kriging method.

I. Linear Kriging Methods:

a. **Ordinary Kriging (OK):** Ordinary kriging estimator allows one to account for such local variation of the local mean by limiting the province of stationarity of the mean to the local neighborhood $Z(y\alpha)$ centered on the location y being estimated (Shamo et. al 2015). The assumption here is that the mean is unknown but fixed. Ordinary kriging assumes a linear model form and the equation is given as:

$$Zx(y) = \sum_{\alpha=1}^{n(y)} \lambda\alpha(y)Z(y\alpha) + \left[1 - \sum_{\alpha=1}^{n(y)} \lambda\alpha(y)\right]\mu(y) \quad \text{(Eq. 2.1)}$$

where Z is continuous attribute (AADT); $Z(y)$ is the true value at unsampled location y; $Zx(y)$ is an estimate of value $Z(y)$; $Z(y\alpha)$ is Z datum value at the location $y\alpha$; m is the stationary mean of the random function (RF) $Z(y)$; $\mu(y\alpha)$ is the expected value of random variable (RV) $Z(y)$; and $\lambda\alpha$ is the kriging weights.

The sill, range, and nugget obtained from the variogram used in combination with this estimator is then used to compute the kriging weight (λα) for which the sum is 1 (Shamo et. al 2015). The mean is obtained by requiring the kriging weights sum to 1

$$\sum_{\alpha=1}^{n(y)} \lambda\alpha(y) = 1 \quad \text{(Eq. 2.2)}$$

Hence, the estimator in OK becomes (Shamo et. al 2015)

$$Zx(y) = \sum_{\alpha=1}^{n(y)} \lambda\alpha(y)Z(y\alpha) \quad \text{(Eq. 2.3)}$$

b. **Simple kriging (SK):** Simple kriging estimator considers the mean μ(y) to be known and constant throughout the study range (Shamo et. al 2015). The simple kriging estimator also assumes a linear model form and is given by the equation:

$$Zx(y) = \sum_{\alpha=1}^{n(y)} \lambda\alpha(y)[Z(y\alpha) - \mu] + \mu \quad \text{(Eq. 2.4)}$$

where $\lambda\alpha$ is weights associated with locations $y\alpha$, $Zx(y)$ is an estimate of value $Z(y)$; $Z(y\alpha)$ is Z datum value at location $y\alpha$ and μ is the unknown constant.

c. **Universal kriging (UK):** Universal kriging estimator is applied when the regionalized variable exhibits some form of the trend (Isaak's and Srivastava 1989). The mean varies, and it is unknown. It also assumes a linear model and the equation is given by:

$$Zx(y) = \sum_{\alpha=1}^{n(y)} \lambda\alpha(y)Z(y\alpha) \qquad \text{(Eq. 2.5)}$$

where $\lambda\alpha$ is weights associated with locations $y\alpha$, $Zx(y)$ is an estimate of value $Z(y)$; $Z(y\alpha)$ is Z datum value at location $y\alpha$

II. Nonlinear Kriging Methods:

a. **Indicator Kriging (IK):** Indicator kriging uses the model (ESRI, 2018):

$$I(\mathrm{s}) = \mu + \varepsilon(\mathrm{s}) \qquad \text{(Eq. 2.6)}$$

where μ is an unknown constant, ε(s) is the error(s) and I(s) is a binary variable. The creation of binary data may be with the use of a threshold for continuous data, or 0 or 1 for the observed or count data (ESRI, 2018). Using binary variables, indicator kriging proceeds the same way as ordinary kriging (ESRI, 2018). Probability means is used by indicator kriging to calculate the forecasted values of the unknown points.

b. **Probability kriging (PK):** According to ESRI (2018), probability kriging assumes the model:

$$I(\mathrm{s}) = I(\mathrm{Z(s)} > \mathrm{c_t}) = \mu_1 + \varepsilon_1(\mathrm{s}) \qquad \text{(Eq. 2.7)}$$

$$\mathrm{Z(s)} = \mu_2 + \varepsilon_2(\mathrm{s}) \qquad \text{(Eq. 2.8)}$$

where: μ_1, μ_2 equals unknown constants, I(s) equals a binary variable created via threshold indicator, $I(\mathrm{Z(s)} > \mathrm{c_t})$.

There are now two types of random errors, ε_1(s) and ε_2(s), so there is autocorrelation for each of them and cross-correlation between them (ESRI, 2018). Probability kriging strives to do the same thing as indicator kriging, but it uses cokriging in an attempt to do a better job. (ESRI, 2018).

c. **Disjunctive kriging (DK):** ESRI (2018) on their website showed disjunctive kriging to assume the model:

$$f(Z(s)) = \mu_1 + \varepsilon(s) \quad \text{(Eq. 2.9)}$$

where $f(Z(s))$ is a random function of $Z(s)$ and μ_1 is an unknown constant. DK requires the bivariate normality assumption and approximations to the functions $f_i(Z(s_i))$; these assumptions are difficult to verify, and the solutions are mathematically and computationally complicated (ESRI, 2018).

III. Bayesian Kriging Method:

a. **Empirical Bayesian kriging (EBK):** EBK is a geostatistical interpolation method that programs the most difficult aspects of building a valid kriging model by automatically calculating parameters through a process of sub-setting and simulations. Other kriging methods in geostatistical analysis require the user to manually regulate parameters to receive accurate results, but EBK automatically calculates these parameters (ESRI, 2018). It accounts for the error introduced by estimating by taking into account the underlying semivariogram making it different from other kriging methods and thereby producing a better and more accurate result.

Other kriging methods calculate the semivariogram from known data locations and use this same single semivariogram to make predictions at unknown locations; this process implicitly assumes that the estimated semivariogram is the true semivariogram for the interpolation region. By not taking the uncertainty of semivariogram estimation into account, other kriging methods underestimate the standard errors of prediction (ESRI, 2018).

2.5.3. Travel Demand Modeling

Travel demand modeling utilizes mathematical models to simulate “real world” transportation system and human travel behaviors (Wang 2012). According to the Virginia Department of Transportation (VDOT), the strength of modern travel demand forecasting is the ability to ask critical “what if” questions about proposed plans and policies. Traditionally, the “four-step process” has been used for travel demand analysis and, as its name implies, is composed of four steps: trip generation, trip distribution, mode choice, and trip assignment (Wang 2012).

Trip generation calculates the number of trips made in each traffic zone and it is the unit of geography that is frequently used in travel demand modeling. Trip distribution involves the determination of the distribution of trips among the origin and destination zones. Mode Choice divides the trips between the origin and destination zones according to different modes of travel and finally, trip assignment help allocate the trips to routes by each travel mode (Wang 2012).

2.5.4. Florida Turnpike State Model

According to the Florida Department of Transportation (FDOT), Turnpike State Model (TSM) is a statewide transportation planning model developed by Florida Turnpike Enterprise. TSM provides estimated AADT values generally for all Florida’s major roadways as well as a total number of trips by Traffic Analysis Zones (TAZ). It uses the AADT allocator process which was developed in order to better estimate traffic volumes (AADT) for all streets and roads in Florida. The underlying principle of the Allocator process is to use the results of the TSM statewide transportation model and apply it to all roads and streets in Florida. This process is shown in the picture obtained from the FDOT website

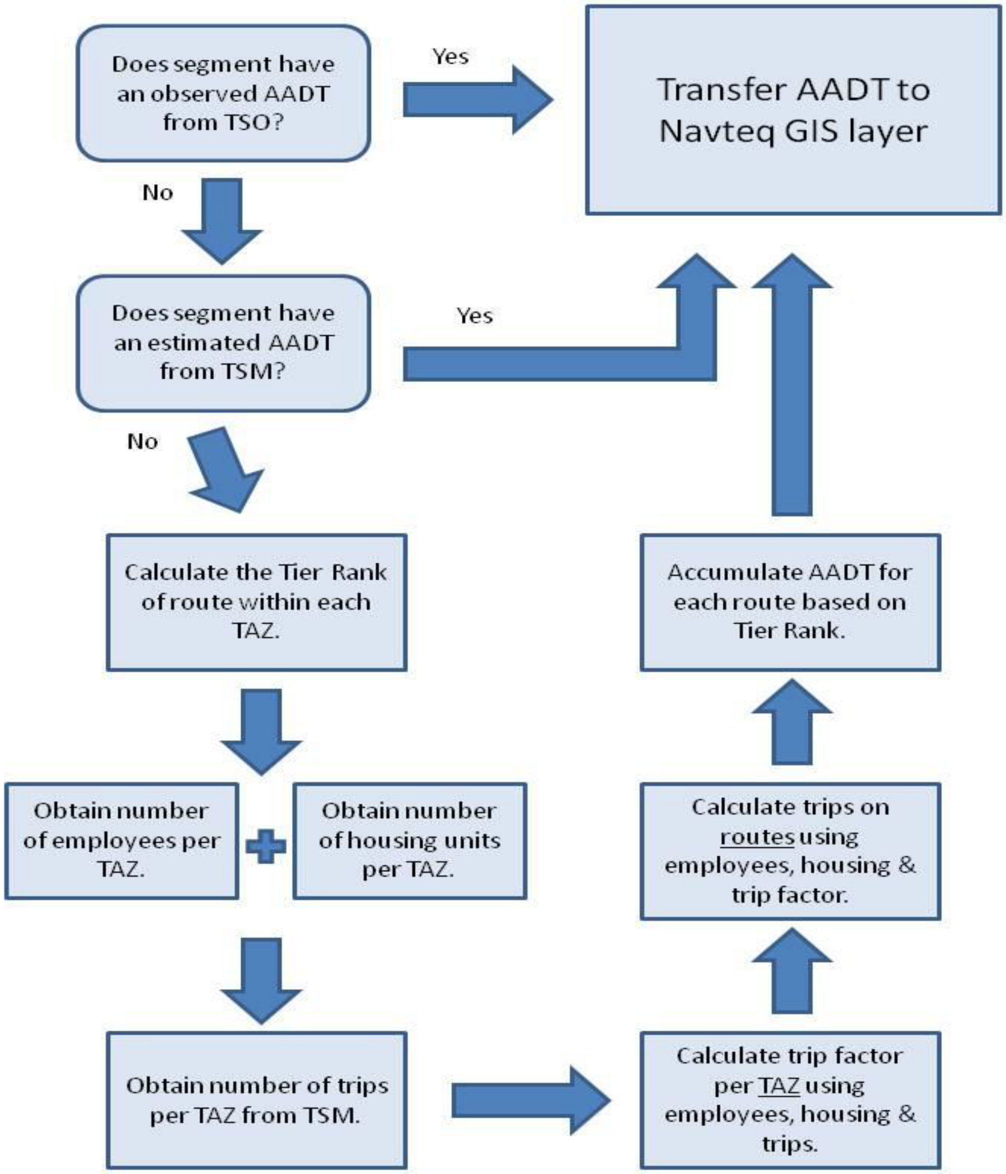

Figure 2.4. TSM Process as Obtained from FDOT Website

2.5.5. Geographically Weighted Regression

Geographically weighted regression (GWR) is a useful exploratory analytical tool that generates a set of location-specific parameter estimates which can be mapped and analyzed to provide information on spatial nonstationarity in relationships between predictors and the outcome variable (Matthews and Yang 2012). GWR models allow model parameters to be estimated locally instead of globally, as in the case of ordinary linear regression (OLR) analysis. (Fang and Nokil 2004). According to Matthews and Yang (2012), GWR is designed to answer the question, “Do relationships vary across space?” It is important to note that the GWR approach does not assume that relationships vary across space but is a means to identify whether or not they do. If the relationships do not vary across space, the global model is an appropriate specification for the data. GWR can also be used to identify interesting locations (areas of variation) for investigation (Akaike 1974).

2.5.6. Artificial Neural Network

An artificial neural network is a computational model composed of simple processing elements called neurons or nodes, which are interconnected by links with weights that perform parallel distributed processing in order to solve the desired problem (Duddu et al 2013). Neural networks have the ability to learn from the environment and to adapt to it in an interactive manner similar to their biological counterparts (Duddu et al 2013). According to Hecht-Nielson (1990) and Lawrence (1993), Neural networks are good at recognizing patterns, generalizing, and predicting trends. They are fast and due to their adaptive nature; they can adapt to changes in the data and learn the characteristics of input signals. Artificial neural networks, also called as neural networks, is a computational model that mimic at least partially the structure and functions of brains and nervous systems of living beings (Cichocki and Unbehauen 1993).

Sharma et al. (1999) carried out a comprehensive comparison between the traditional factor approach and the neural network approach for estimating AADT from 48-h sample counts. They used data from 63 automatic traffic recorders (ATR) sites located on interstate highways and other major roadways in Minnesota. They pointed out that the advantage of the neural network approach would lie in the fact that the classification of ATR sites and sample site assignments to ATR groups would not be required. Origin-Destination centrality-based method

2.5.7. Support Vector Regression with Data-Dependent Parameters

Support vector regression with data-dependent parameters (SVR-DP) has been getting growing attention because of its remarkable characteristics, including a strong theoretical foundation, good generalization performance, the absence of local minima, and sparse representation of solution (Vapnik, 1995). Castro-Neto et al (2009) noted that the implementation of many SVR algorithms requires the computation of adequate SVR parameters, which are crucial to the quality of SVR models developed and the most accurate technique is the use of resampling methods such as cross-validation.

Castro-Neto et al (2009) proposed an SVR methodology that uses data-dependent parameters in order to predict AADT. The methodology uses SVR to predict AADT data in order to enhance prediction accuracy and provides an efficient way of computing SVR parameters. This is achieved by incorporating the equations for computing SVR parameters into the conventional SVR algorithm in order to obtain data-dependent parameters and reduce computational time. The use of data-dependent parameters guarantees that the value of the parameters will give smaller support vectors and a less complex model (Cherkassky & Ma, 2004).

2.5.8. Origin-Destination Centrality-Based Method

This approach is based on the use of centrality. There are various mathematical forms of centrality, all of which seek to quantify the topological importance of each element in a network (i.e. the hierarchy of connectivity for each link and node) (Lowry 2014). The most popular mathematical forms include betweenness centrality, degree centrality, and closeness centrality (Brandes, 2008). Wang et al. (2011) used centrality to analyze street networks, land use intensity, and air transport networks. Over the years, a developed specialized form of centrality called “space syntax” has been shown to exhibit high correlation with pedestrian volumes in small-scale urban environments, like pedestrian plazas (Hillier et al., 1993; Raford and Ragland, 2004).

Large-scale studies have focused on the theoretical implications that space syntax reveals about urban form, spatial cognition, and human movement, with little or no emphasis on the implications for estimating AADT (Lowry 2014). One disadvantage that was discovered in efforts to estimate AADT using space syntax is that the formulation for space syntax does not provide an easy means to incorporate information about origins and destinations (Lowry 2014).

2.6. Limitations and Future Work

The papers that were reviewed were restricted to a selected sample of databases. This might have impacted the research due to the sample size and the selection process. The categories developed in the analysis process were neither mutually exclusive nor definitive. Due to the constraints of scope of work and time, more details cannot be included in this thesis.

2.7. Summary

Annual Average Daily Traffic (AADT) data in the transportation industry today is an important tool which is used in various fields such as highway planning, pavement design, traffic safety, transport operations, and policy. To address the methods used in AADT estimation and prediction as well as the factor that affects the accuracy of these methods, systematic review on estimating methods was performed.

The findings from the analysis identified different methods for AADT estimating out of which ordinary linear regression occurred most and five other methods occurred least. Geographical location, road type, the day of the week and seasonality amongst others were identified as factors that influence the accuracy of AADT estimation and prediction.

3. EXPLORATORY DATA ANALYSIS

3.1. Introduction

This chapter discussed the processes and results of the exploratory data analysis of WSDOT AADT data. Exploratory data analysis provided the vital information from the data and was used to answer part of questions 3 and 4 of the research questions. Graphical methods, tables, and statistical methods were employed in the presentation of the outcomes of this analysis.

The trend of the data, identification of the key locations and attributes for the research was performed using ArcGIS. Histogram, normal probability plot and run sequence plot were carried out to analyze the trend and characteristics of the data.

3.2. Data Acquisition

The data used for this research analysis was obtained from the Washington State Department of Transportation website (WSDOT 2018). The data consist of AADT data from the year 2009 to the year 2016 encompassing all the recorded camera locations in the state. The AADT dataset was downloaded in the form of shapefiles and KML files. The data counts were taken at permanent stations across the state in line with the Highway Performance Monitoring System (HPMS) requirements for continuous data counts (Shamo et al. 2015).

The data consists of different parameters which include the record number for the location, the direction of travel of the vehicles (both ways bound, northbound, southbound, eastbound and westbound), the location of the camera and the data count for each location. Each class of permanent count station consists of highway links (the homogeneous section that has the same features such as AADT and seasonal variation in traffic volume) with similar traffic patterns and characteristics.

Also, the shapefile for state and county boundaries of Washington State were obtained from the U.S. Census Bureau with the road network shapefile obtained from the Washington State Department of Transport website.

3.3. Data Analysis

Data analysis is the process of systematically applying statistical techniques to describe facts, detect patterns, develop explanations, and test hypotheses (Levine and Roos, 2002). It helps in structuring the findings from different sources of data collection. According to Tukey (1977), it provides insight into a large dataset to make meaningful critical decisions in order to avoid human bias from research conclusions with the aid of proper statistical treatments and helps to verify whether the hypothesis is valid, reproducible, and unquestionable. Data analysis consists of several phases including data cleaning, quality analysis, exploratory analysis, and knowledge representation (Tukey, 1977). Several methods are employed in data analysis such as classical analysis, exploratory data analysis, and Bayesian analysis. The methodology used in the present work is the classical approach, which involves data collection, model development (normality, linearity, etc.), analysis, estimation, testing, and conclusions.

Exploratory data analysis (EDA) is used to understand the data. It gives insight into the dataset, discover the underlying structure, extract important variables, test underlying assumptions, and detect outliers and anomalies (Shamo et al. 2015). Examples of tools EDA uses to interpret data includes a histogram, scatterplot, run sequence plot, probability plot and descriptive statistics amongst others. Histogram, Run Sequence Plot and Probability Plot were employed to analyze the AADT dataset for this research.

Run Sequence Plot: According to Chambers (1983), run sequence plots are an easy way to graphically summarize a univariate data set. A common assumption of univariate datasets is that they behave like:

1. random drawings
2. from a fixed distribution
3. with a common location; and
4. with a common scale.

With run sequence plots, shifts in location and scale are typically quite evident. Also, outliers can easily be detected. Run sequence plots are formed by:

- Vertical axis: Response variable Y_i
- Horizontal axis: Index i (i = 1, 2, 3, ...)

For univariate data, the default model by Chambers (1983) is:

$$Y = \text{constant} + \text{error} \qquad \text{(Eq. 3.1)}$$

where the error is assumed to be random, from a fixed distribution, and with constant location and scale. The validity of this model depends on the validity of these assumptions. The run sequence plot is useful for checking for constant location and scale. Even for more complex models, the assumptions on the error term are still often the same. That is, a run sequence plot of the residuals (even from very complex models) is still vital for checking for outliers and for detecting shifts in location and scale (Chambers, 1983).

Histogram: The purpose of a histogram is to graphically summarize the distribution of a univariate dataset (Chambers 1983). The histogram graphically shows the following:

1. center (i.e., the location) of the data;
2. spread (i.e., the scale) of the data;

3. skewness of the data;

4. presence of outliers; and

5. the presence of multiple modes in the data.

These features provide strong indications of the proper distributional model for the data. The most common form of the histogram is obtained by splitting the range of the data into equal-sized bins (called classes). Then for each bin, the number of points from the dataset that fall into each bin is counted. That is

- Vertical axis: Frequency (i.e., counts for each bin)
- Horizontal axis: Response variable

The classes can either be defined arbitrarily by the user or via some systematic rule. A number of theoretically derived rules have been proposed by Scott (Scott 1992).

Normal Probability Plot: The normal probability plot (Chambers et al., 1983) is a graphical technique for assessing whether a data set is approximately normally distributed. The data are plotted against a theoretical normal distribution in such a way that the points should form an approximately straight line. Departures from this straight line indicate departures from normality.

The normal probability plot is formed by:

- Vertical axis: Ordered response values
- Horizontal axis: Normal order statistic medians

The observations are plotted as a function of the corresponding normal order statistic medians which are defined as (Filliben 1975):

$$N_i = G(U_i) \qquad \text{(Eq. 3.2)}$$

where U_i is the uniform order statistic medians (defined below) and G is the percent point function of the normal distribution. The percent point function is the inverse of the cumulative distribution function (the probability that x is less than or equal to some value). That is, given a probability, we want the corresponding x of the cumulative distribution function. Probability plots are used to assess the assumption of a fixed distribution. In particular, most statistical models are of the form (Filliben 1975):

The complete data parameters and characteristics acquired from the WSDOT is shown in Figure 3.1 and Table 3.1 below. The year 2010 raw data downloaded as shown in Figure 3.1 and Table 3.1 showed its data parameters and characteristics after it was imported into the ArcGIS and extracted.

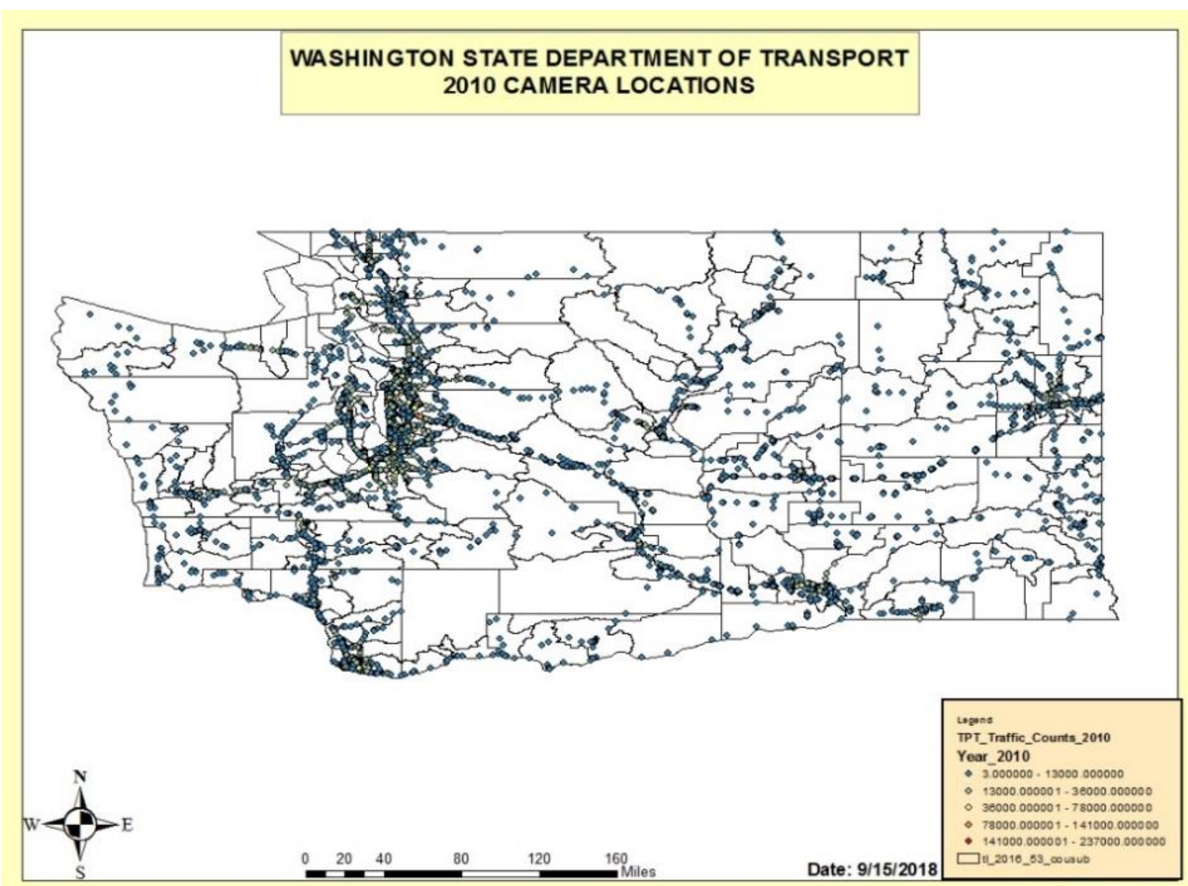

Figure 3.1. The Year 2010 Camera Locations using ArcGIS

Table 3.1. Extracted Data Parameters and Characteristics of The Year 2010 Data

FID	OBJECTID	SRID	ARM	Year_2010	Direction_	Location	LOC_ERROR
0	1	2	0.01	6500	West Bound	After Milepost 0.00 B: BOTHWAYS INTERSECTION SR 529-MAPLE ST, BEGIN ROUTE	NO ERROR
1	2	2	0.12	9000	West Bound	After Milepost 0.11 B: LEFT EXIT TO WALNUT ST	NO ERROR
2	3	290	0	8400	South Bound	At Milepost 0.07 A: DECREASING UNDERCROSSING SR 90, BEGIN ROUTE	NO ERROR
3	4	515	7.75	11000	South Bound	After Milepost 7.70 A: BOTHWAYS INTERSECTION SR 900 E BND	NO ERROR
4	5	515	7.85	12000	South Bound	Before Milepost 7.82 A: BOTHWAYS INTERSECTION SR 900 CO2NDST (COUPLET), END ROUTE	NO ERROR
5	6	522	0	16000	West Bound	At Milepost 0.00 A: DECREASING MISCELLANEOUS FEATURE S1 RAMP AHEAD, BEGIN ROUTE	NO ERROR
6	7	536	0.01	4600	West Bound	After Milepost 0.00 A: BOTHWAYS INTERSECTION SR 20, BEGIN ROUTE	NO ERROR
7	8	543	0	4400	South Bound	At Milepost 0.00 A: BEGIN DECREASING BRIDGE SR 5, BEGIN ROUTE	NO ERROR
8	9	2	14.44	26000	Bothways	Before Milepost 14.37 A: LEFT OFF RAMP SR 522, LEFT ON RAMP SR 522	NO ERROR
9	10	2	13.96	26000	Bothways	After Milepost 13.87 A: BOTHWAYS INTERSECTION 179TH AVE SE	NO ERROR
10	11	2	13.93	22000	Bothways	Before Milepost 13.86 A: LEFT WYE CONNECTION 179TH AVE SE, RIGHT MISCELLANEOUS FEATURE SGN ENT MONROE	NO ERROR
11	12	2	13.04	21000	Bothways	After Milepost 12.95 A: RIGHT INTERSECTION FRYELANDS BLVD SE, LEFT INTERSECTION ROOSEVELT RD	NO ERROR
12	13	2	13.02	23000	Bothways	Before Milepost 12.95 A: RIGHT INTERSECTION FRYELANDS BLVD SE, LEFT INTERSECTION ROOSEVELT RD	NO ERROR
13	14	2	10.17	24000	Bothways	After Milepost 10.08 A: LEFT INTERSECTION WESTWICK RD	NO ERROR
14	15	2	10.15	25000	Bothways	Before Milepost 10.08 A: LEFT INTERSECTION WESTWICK RD	NO ERROR
15	16	2	8.94	25000	Bothways	After Milepost 8.80 A: LEFT OFF RAMP CAMPBELL RD	NO ERROR
16	17	2	8.64	17000	Bothways	At Milepost 8.51 A: BOTHWAYS UNDERCROSSING CAMPBELL RD	NO ERROR
17	18	2	8.02	21000	Bothways	Before Milepost 7.90 A: LEFT ON RAMP CAMPBELL RD	NO ERROR
18	19	2	5.49	21000	Bothways	After Milepost 5.35 A: RIGHT ON RAMP SR 9	NO ERROR
19	20	2	5.17	16000	Bothways	At Milepost 5.04 A: BOTHWAYS UNDERCROSSING SR 9	NO ERROR
20	21	2	4.87	24000	Bothways	Before Milepost 4.75 A: LEFT ON RAMP SR 9	NO ERROR
21	22	2	4	24000	Bothways	After Milepost 3.86 A: RIGHT ON RAMP BICKFORD AVE (OLD SR 2)	NO ERROR
22	23	2	3.97	26000	Bothways	Before Milepost 3.85 A: RIGHT WYE CONNECTION ON RAMP	NO ERROR
23	24	2	3.68	26000	Bothways	After Milepost 3.54 A: RIGHT OFF RAMP BICKFORD AVE (OLD SR 2)	NO ERROR
24	25	2	3.66	31000	Bothways	Before Milepost 3.54 A: RIGHT OFF RAMP BICKFORD AVE (OLD SR 2)	NO ERROR

Table 3.1. Extracted Data Parameters and Characteristics of The Year 2010 Data (Continued)

FID	OBJECTID	SRID	ARM	Year_2010	Direction_	Location	LOC_ERROR
25	26	2	2.89	31000	Bothways	After Milepost 2.75 A: LEFT CENTER OFF RAMP SR 204-HEWITT AVE	NO ERROR
26	27	2	17.07	19000	Bothways	After Milepost 16.98 A: RIGHT INTERSECTION SOFIE RD, LEFT INTERSECTION CALHOUN RD	NO ERROR
27	28	2	15.33	19000	Bothways	After Milepost 15.24 A: LEFT WYE CONNECTION OLD OWEN RD	NO ERROR
28	29	2	15.28	21000	Bothways	Before Milepost 15.21 A: RIGHT WYE CONNECTION MAIN ST	NO ERROR
29	30	2	15.01	28000	Bothways	After Milepost 14.92 A: RIGHT INTERSECTION SR 203-LEWIS ST, LEFT INTERSECTION CHAIN LAKE RD	NO ERROR
30	31	2	14.99	31000	Bothways	Before Milepost 14.92 A: RIGHT INTERSECTION SR 203-LEWIS ST, LEFT INTERSECTION CHAIN LAKE RD	NO ERROR
31	32	2	14.47	37000	Bothways	After Milepost 14.38 A: CENTER INTERSECTION U-TURN ACCESS, LEFT WYE CONNECTION OFF RAMP	NO ERROR
32	33	2	1	73000	Bothways	Before Milepost 0.88 A: LEFT CENTER ON RAMP EBEY ISLAND	NO ERROR
33	34	2	2.21	72000	Bothways	Before Milepost 2.09 A: RIGHT CENTER OFF RAMP SR 204-HEWITT AVE	NO ERROR
34	35	2	2.17	72000	Bothways	After Milepost 2.03 A: LEFT CENTER OFF RAMP EBEY ISLAND	NO ERROR
35	36	2	2.15	71000	Bothways	Before Milepost 2.03 A: LEFT CENTER OFF RAMP EBEY ISLAND	NO ERROR
36	37	2	2.58	28000	Bothways	At Milepost 2.45 A: END DECREASING BRIDGE EBEY SLOUGH	NO ERROR
37	38	2	0.39	73000	Bothways	At Milepost 0.26 A: PTR LOCATION R052	NO ERROR
38	39	2	0.2	21000	Bothways	After Milepost 0.06 A: RIGHT WYE CONNECTION HEWITT AVE (OLD SR 2)	NO ERROR
39	40	2	0.14	23000	Bothways	After Milepost 0.00 A: BEGIN DECREASING BRIDGE W-W RAMP, INCREASING UNDERCROSSING SR 5 SB, DECREASING UNDERCROSSING SR 5 SB	NO ERROR
40	41	2	21.64	19000	Bothways	Before Milepost 21.57 A: RIGHT INTERSECTION FERN BLUFF RD, LEFT INTERSECTION OLD OWEN RD	NO ERROR
41	42	2	21.66	21000	Bothways	After Milepost 21.57 A: RIGHT INTERSECTION FERN BLUFF RD, LEFT INTERSECTION OLD OWEN RD	NO ERROR
42	43	2	22.37	19000	Bothways	Before Milepost 22.30 A: LEFT INTERSECTION 4TH ST	NO ERROR
43	44	2	22.39	17000	Bothways	After Milepost 22.30 A: LEFT INTERSECTION 4TH ST	NO ERROR
44	45	2	22.46	17000	Bothways	After Milepost 22.37 A: RIGHT INTERSECTION JW MANN RD, LEFT INTERSECTION 5TH ST	NO ERROR
45	46	2	23.18	16000	Bothways	Before Milepost 23.11 A: RIGHT EXIT TO CEMETERY RD	NO ERROR
46	47	2	23.32	14000	Bothways	After Milepost 23.23 A: LEFT INTERSECTION SULTAN BASIN RD	NO ERROR
47	48	2	25.27	13000	Bothways	Before Milepost 25.20 A: RIGHT ENTRANCE/EXIT ROADSIDE PARK	NO ERROR
48	49	2	25.78	13000	Bothways	Before Milepost 25.71 A: BOTHWAYS INTERSECTION 363RD AVE SE	NO ERROR
49	50	2	26.28	12000	Bothways	After Milepost 26.19 A: LEFT INTERSECTION KELLOGG LAKE RD	NO ERROR
50	51	2	27.99	12000	Bothways	Before Milepost 27.92 A: BOTHWAYS INTERSECTION 1ST ST	NO ERROR

Table 3.1. Extracted Data Parameters and Characteristics of The Year 2010 Data (Continued)

FID	OBJECTID	SRID	ARM	Year_2010	Direction_	Location	LOC_ERROR
51	52	2	28.01	11000	Bothways	After Milepost 27.92 A: BOTHWAYS INTERSECTION 1ST ST	NO ERROR
52	53	2	29.55	9700	Bothways	Before Milepost 29.48 A: RIGHT INTERSECTION GUNN RD, LEFT INTERSECTION PICKLE FARM RD	NO ERROR
53	54	2	29.57	7900	Bothways	After Milepost 29.48 A: RIGHT INTERSECTION GUNN RD, LEFT INTERSECTION PICKLE FARM RD	NO ERROR
54	55	2	31.29	6500	Bothways	Before Milepost 31.22 A: RIGHT ENTRANCE/EXIT BUSINESS, LEFT INTERSECTION FIR RD	NO ERROR
55	56	2	31.31	6400	Bothways	After Milepost 31.22 A: RIGHT ENTRANCE/EXIT BUSINESS, LEFT INTERSECTION FIR RD	NO ERROR
56	57	2	35.69	6100	Bothways	Before Milepost 35.62 A: LEFT INTERSECTION INDEX-GALENA RD	NO ERROR
57	58	2	35.72	5200	Bothways	After Milepost 35.63 A: LEFT WYE CONNECTION INDEX-GALENA RD	NO ERROR
58	59	2	41.12	5100	Bothways	After Milepost 41.03 A: RIGHT INTERSECTION 634TH PL NE	NO ERROR
59	60	2	41.69	6100	Bothways	Before Milepost 41.62 A: RIGHT INTERSECTION NE 191ST ST	NO ERROR
60	61	2	43.39	5900	Bothways	Before Milepost 43.32 A: LEFT INTERSECTION FS RD #6028	NO ERROR
61	62	2	48.77	5800	Bothways	Before Milepost 48.70 A: RIGHT WYE CONNECTION 5TH ST	NO ERROR
62	63	2	48.81	5500	Bothways	After Milepost 48.72 A: RIGHT WYE CONNECTION 5TH ST	NO ERROR
63	64	2	50.2	4800	Bothways	At Milepost 50.12 A: PTR LOCATION R038	NO ERROR
64	65	2	52.17	4900	Bothways	After Milepost 52.08 A: LEFT INTERSECTION FS RD #6066	NO ERROR
65	66	2	56.77	4900	Bothways	Before Milepost 56.70 A: LEFT INTERSECTION DECEPTION FALLS PARKING	NO ERROR
66	67	2	60.4	4900	Bothways	At Milepost 60.32 A: BEGIN INCREASING BRIDGE TUNNEL CREEK, BEGIN DECREASING BRIDGE TUNNEL CREEK	NO ERROR
67	68	2	66.37	4200	Bothways	Before Milepost 66.24 A: LEFT INTERSECTION YODELIN PL	NO ERROR
68	69	2	66.39	4200	Bothways	After Milepost 66.24 A: LEFT INTERSECTION YODELIN PL	NO ERROR
69	70	2	72.78	4000	Bothways	At Milepost 72.68 A: END BOTHWAYS BRIDGE NASON CREEK	NO ERROR
70	71	2	76.12	4200	Bothways	Before Milepost 76.03 A: LEFT INTERSECTION MERRITT LAKE TR	NO ERROR
71	72	2	78.39	4200	Bothways	After Milepost 78.28 A: RIGHT INTERSECTION WHITE PINE RD	NO ERROR
72	73	2	80.28	4200	Bothways	At Milepost 80.20 A: PTR LOCATION R058	NO ERROR
73	74	2	84.81	4300	Bothways	Before Milepost 84.74 A: LEFT WYE CONNECTION SR 207	NO ERROR
74	75	2	84.85	5100	Bothways	After Milepost 84.76 A: LEFT WYE CONNECTION SR 207	NO ERROR
75	76	2	86.27	5100	Bothways	After Milepost 86.18 A: RIGHT INTERSECTION WINTON RD	NO ERROR
76	77	2	90.51	5100	Bothways	Before Milepost 90.44 A: LEFT ENTRANCE/EXIT TUMWATER CAMPGROUND	NO ERROR
77	78	2	110.22	16000	Bothways	After Milepost 110.13 A: RIGHT INTERSECTION GOODWIN RD, CENTER INTERSECTION MEDIAN XROAD, LEFT INTERSECTION HAY CANYON RD	NO ERROR

Table 3.1. Extracted Data Parameters and Characteristics of The Year 2010 Data (Continued)

FID	OBJECTID	SRID	ARM	Year_2010	Direction_	Location	LOC_ERROR
78	79	2	111.14	16000	Bothways	Before Milepost 111.07 A: RIGHT WYE CONNECTION APLETS WAY	NO ERROR
79	80	2	111.18	16000	Bothways	After Milepost 111.09 A: RIGHT INTERSECTION APLETS WAY, CENTER INTERSECTION MEDIAN XROAD, LEFT INTERSECTION NAHAHUM CANYON RD	NO ERROR
80	81	2	112.03	16000	Bothways	Before Milepost 111.96 A: RIGHT WYE CONNECTION COTLETS WAY	NO ERROR
81	82	2	112.09	21000	Bothways	After Milepost 112.00 A: LEFT WYE CONNECTION NAHAHUM CANYON RD	NO ERROR
82	83	2	112.68	21000	Bothways	After Milepost 112.59 A: RIGHT INTERSECTION OLD MONITOR RD, CENTER INTERSECTION MEDIAN XROAD	NO ERROR
83	84	2	113.18	21000	Bothways	At Milepost 113.10 A: PTR LOCATION P01	NO ERROR
84	85	2	113.28	21000	Bothways	Before Milepost 113.21 A: RIGHT INTERSECTION OLD MONITOR RD, CENTER INTERSECTION MEDIAN XROAD, LEFT INTERSECTION RED APPLE RD	NO ERROR
85	86	2	115.17	21000	Bothways	Before Milepost 115.10 A: RIGHT WYE CONNECTION MAIN ST	NO ERROR
86	87	2	106.43	15000	Bothways	Before Milepost 106.36 A: LEFT INTERSECTION FRONTAGE RD	NO ERROR
87	88	5	95.43	59000	Bothways	After Milepost 95.35 A: LEFT OFF RAMP MAYTOWN RD (OLD SR 121)	NO ERROR
88	89	5	100.8	69000	Bothways	Before Milepost 100.74 A: LEFT ON RAMP TUMWATER BLVD	NO ERROR
89	90	3	2.32	19000	Bothways	After Milepost 2.31 A: RIGHT INTERSECTION MILL ST	NO ERROR
90	91	3	2.24	15000	Bothways	Before Milepost 2.25 A: BOTHWAYS INTERSECTION HARVARD AVE	NO ERROR
91	92	3	1.52	14000	Bothways	After Milepost 1.51 A: RIGHT WYE CONNECTION ARCADIA RD	NO ERROR
92	93	3	1.48	13000	Bothways	Before Milepost 1.49 A: RIGHT INTERSECTION ARCADIA RD, LEFT INTERSECTION ARCADIA AVE	NO ERROR
93	94	3	0.12	13000	Bothways	After Milepost 0.11 A: RIGHT ON RAMP SR 101	NO ERROR
94	95	18	14.23	33000	Bothways	After Milepost 13.69 A: RIGHT ON RAMP SE 256TH ST	NO ERROR
95	96	18	19.08	26000	Bothways	After Milepost 18.57 A: RIGHT ON RAMP 244TH AVE SE	NO ERROR
96	97	3	0	7200	Bothways	At Milepost 0.00 A: BOTHWAYS UNDERCROSSING SR 101, BEGIN ROUTE	NO ERROR
97	98	002CONEWPRT	0.47	5400	South Bound	Before Milepost 334.86 A: LEFT WYE CONNECTION SR 2	NO ERROR
98	99	002CONEWPRT	0.08	5700	South Bound	After Milepost 334.45 A: RIGHT INTERSECTION SR 20, RIGHT INTERSECTION WASHINGTON AVE	NO ERROR
99	100	002CONEWPRT	0.06	7900	Bothways	Before Milepost 334.45 A: RIGHT INTERSECTION SR 20, RIGHT INTERSECTION WASHINGTON AVE	NO ERROR

Table 3.1 above shows the examples of the data acquired from the WSDOT database and compiled in an excel file. The parameters include the FID number, which is the serial number for each location, the direction of travel of the vehicles (Bothways bound, Northbound, Southbound, Eastbound, and Westbound), the location of the camera and the data count for each location. The OBJECTID, SRID and ARM columns are given to be able to use the data for analysis in ArcGIS.

3.3.1. Data Trend Analysis Using Directions

This section offers an insight into the trend of the AADT dataset collected at the different collection points. The year 2009 is considered as the benchmark for this analysis because the data downloaded starts in 2009 and there will not be any available dataset to compare it to. Data from previous years (2008 and less) could not be used because it does not have the location properties that were needed to analyze it in ArcGIS software.

Table 3.2. Data Trend Bothways bound

Years	Total AADT	Percentage Increase in AADT data from the previous year (%)
2009	137732340	0
2010	139156080	0.010337
2011	135926060	-0.02321
2012	135216339	-0.00522
2013	137054410	0.013594
2014	137636640	0.004248
2015	140523570	0.020975
2016	142659550	0.0152

Table 3.2 above provides the summary of the AADT data acquired for each year from 2009 to 2016 and the percentage increase in AADT data for each year for the Bothways bound collection points in Washington State. There is a total increase of 4,927,210 from the year 2009 to the year 2016. Figure 3.2 below represents the trend of AADT count dataset taken at the Bothways bound collection points in Washington State. Table 3.2 also showed a percentage decrease in the years 2011 and 2012 with a decrement of 2.32% and 0.52% respectively from the preceding year. The highest percentage increase in AADT count can be noticed in the year 2015 and 2016 with an increment of 2.10% and 1.52% respectively.

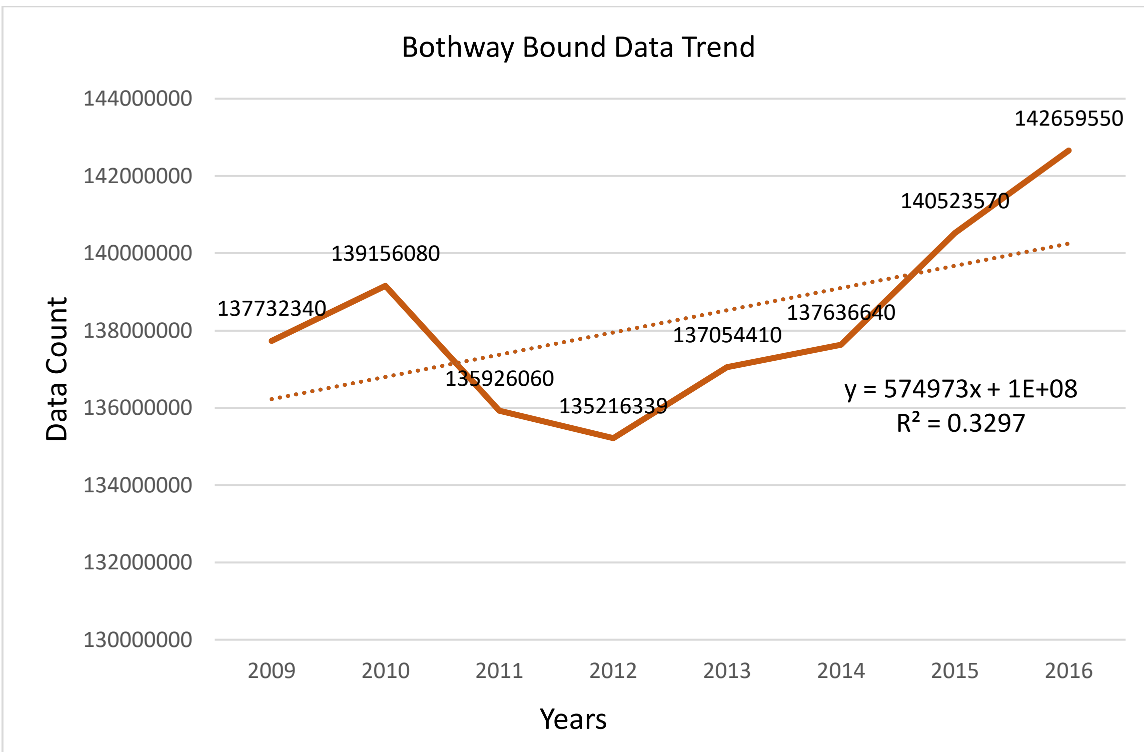

Figure 3.2. The Trend of Bothways AADT Dataset in Washington State 2009 -2016

The data suggest that AADT rose steadily from the year 2009 to 2010 and then suffered a gradual decrease from 2010 to 2012. The year 2013 to 2016 then saw a steady and gradual increase in AADT count. This showed a positive trend in the relationship between the AADT count dataset and the year with an estimated average of 138,238,124 counts from the year 2009 to the year 2016.

The positive trend in Figure 3.2 provided a value of R^2 of 0.3297, with a linear equation (Eq. 3.3) provided below:

$$Y = 574973x + 1E + 0 \qquad \text{(Eq. 3.3)}$$

Table 3.3. Data Trend Northbound

Years	**Total AADT**	**Percentage Increase in AADT data from the previous year (%)**
2009	5423270	0
2010	5797110	0.068933
2011	5960920	0.028257
2012	6116670	0.026129
2013	6290720	0.028455
2014	6341650	0.008096
2015	5849650	-0.07758
2016	5017720	-0.14222

Table 3.3 above provides the summary of the AADT data acquired for each year from 2009 to 2016 and the percentage increase in AADT data for each year for the Northbound collection points in Washington State. There is a total decrease of 405,550 from the year 2009 to the year 2016.

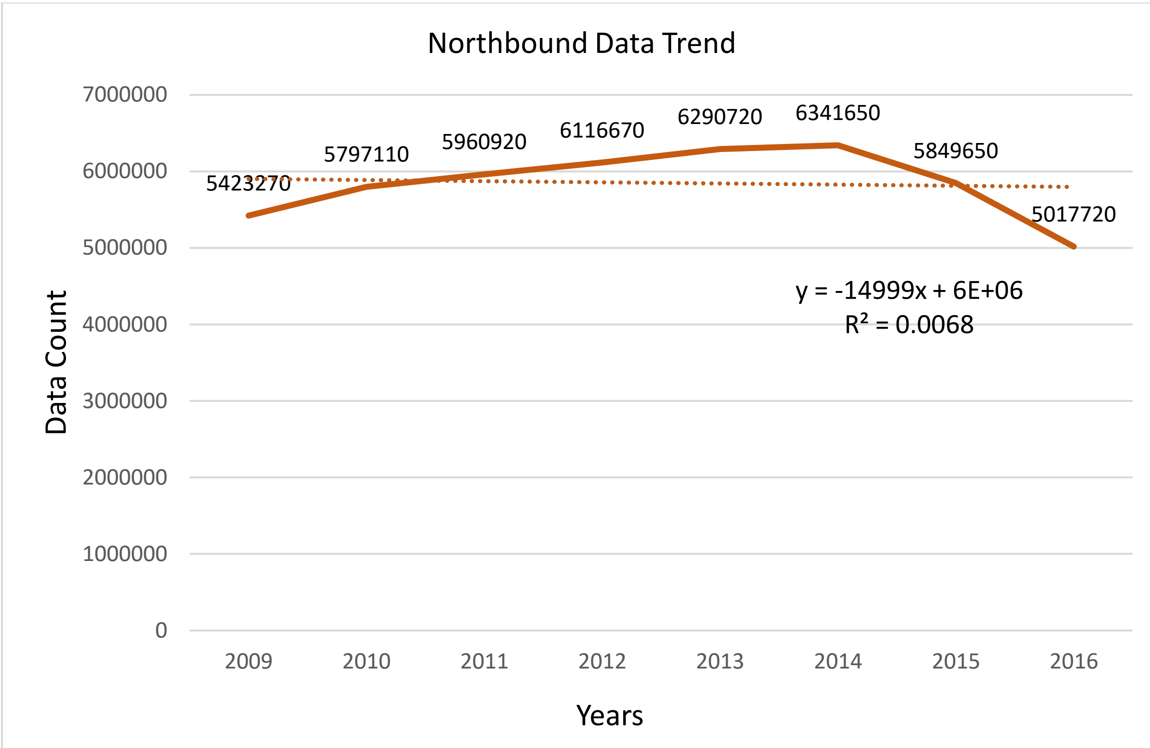

Figure 3.3. The Trend of Northbound AADT Dataset in Washington State 2009 - 2016

Figure 3.3 above represents the trend of AADT dataset count taken at the Northbound collection points in Washington State. The data suggest that AADT rose steadily from the year 2009 to 2014 and then suffered a gradual decrease from 2014 to 2016. This showed a positive trend in the relationship between the AADT count dataset and the year with an estimated average of 5,849,714 counts from the year 2009 to the year 2016.

Table 3.3 above showed a percentage decrease in the years 2015 and 2016 by a decrement of 7.76%, and 14.22% respectively from the preceding year. The highest percentage increase in AADT count can be noticed in the year 2010 and 2013 with an increment of 6.89% and 2.85% respectively. Figure 3.3 positive trend provided a value of R^2 of 0.0068 and a linear equation (Eq. 3.4) provided below:

$$Y = -14999x + 6E + 06 \qquad \text{(Eq. 3.4)}$$

Table 3.4. Data Trend Southbound

Years	Total AADT	Percentage Increase in AADT data from the previous year (%)
2009	5433170	0
2010	5965010	0.097888
2011	5903290	-0.01035
2012	5999670	0.016326
2013	6280060	0.046734
2014	6457500	0.028255
2015	6047770	-0.06345
2016	5115580	-0.15414

Table 3.4 above provides the summary of the AADT data acquired for each year from 2009 to 2016 and the percentage increase in AADT for each year for the Southbound collection points in Washington State. There is a total decrease of 317,590 from the year 2009 to the year 2016. Table 3.4 also showed a percentage decrease in the years 2011, 2015 and 2016 with a decrement of 1.03%, 6.35%, and 15.41% respectively from the preceding year. The highest percentage increase in AADT count can be noticed in the year 2010 and 2014 with an increment of 9.79% and 2.83% respectively.

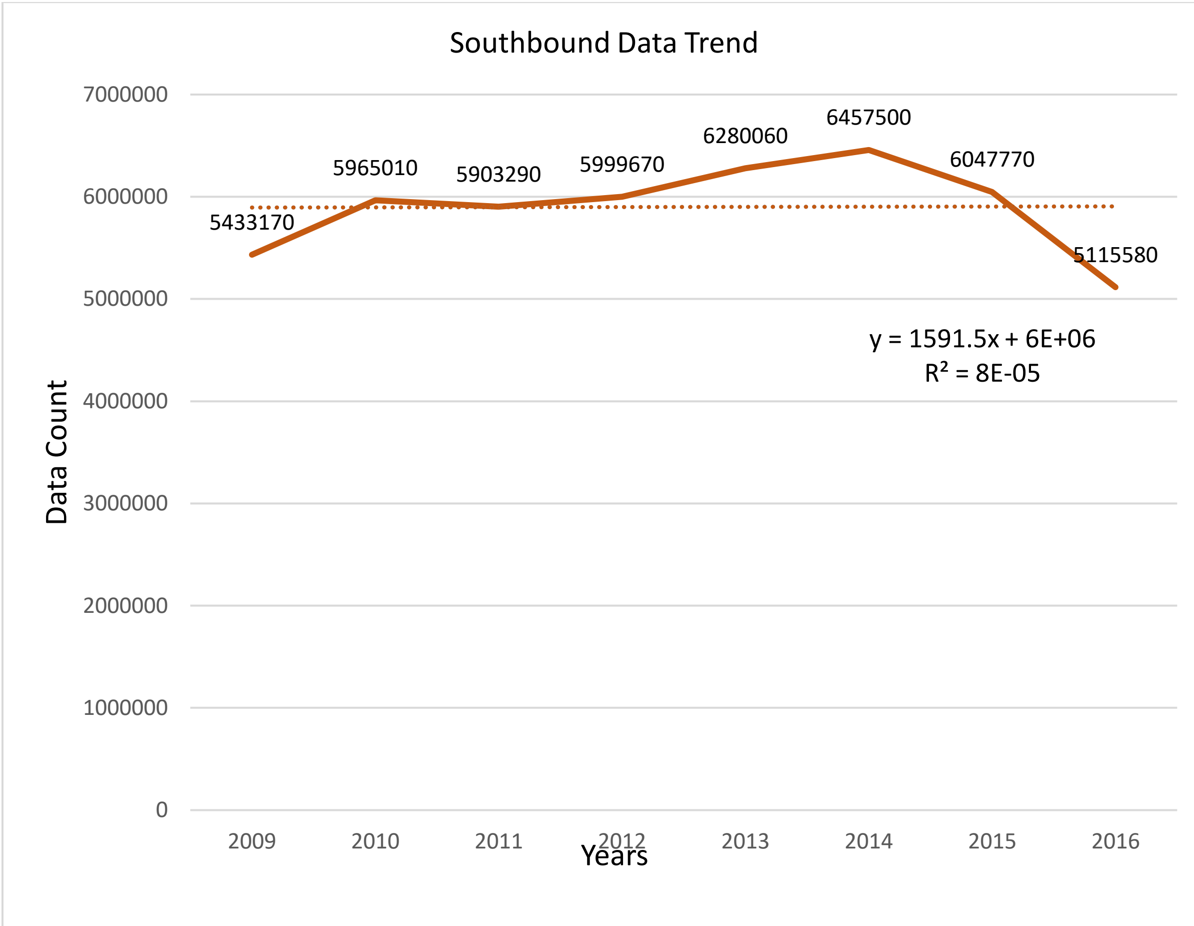

Figure 3.4. The Trend of Southbound AADT Dataset in Washington State 2009 - 2016

Figure 3.4 above represents the trend of AADT count dataset taken at the Southbound collection points in Washington State. The data suggest that AADT rose steadily from the year 2009 to 2014 and then suffered a gradual decrease from 2014 to 2016. This showed a positive trend in the relationship between the AADT count dataset and the year with an estimated average of 5,900,256 counts from the year 2009 to the year 2016. The positive trend in Figure 3.4 resulted in a linear equation (Eq. 3.5) provided below:

$$Y = 1591.5x + 6E + 06 \qquad \text{(Eq. 3.5)}$$

Table 3.5. Data Trend Eastbound

Years	Total AADT	Percentage Increase in AADT data from the previous year (%)
2009	2805210	0
2010	2792940	-0.00437
2011	2574980	-0.07804
2012	2777390	0.078606
2013	2995760	0.078624
2014	3084530	0.029632
2015	2854300	-0.07464
2016	2902230	0.016792

Table 3.5 above provides the summary of the AADT data acquired for each year from 2009 to 2016 and the percentage increase in AADT data for each year for the Eastbound collection points in Washington State. There is a total increase of 97,020 from the year 2009 to the year 2016.

Table 3.5 above showed a percentage decrease in the years 2010 and 2015 with a decrement of 0.44% and 7.46% respectively from the preceding year. The highest percentage increase in AADT can be noticed in the year 2012 and 2013 with an increment of 7.86% and 7.86% respectively.

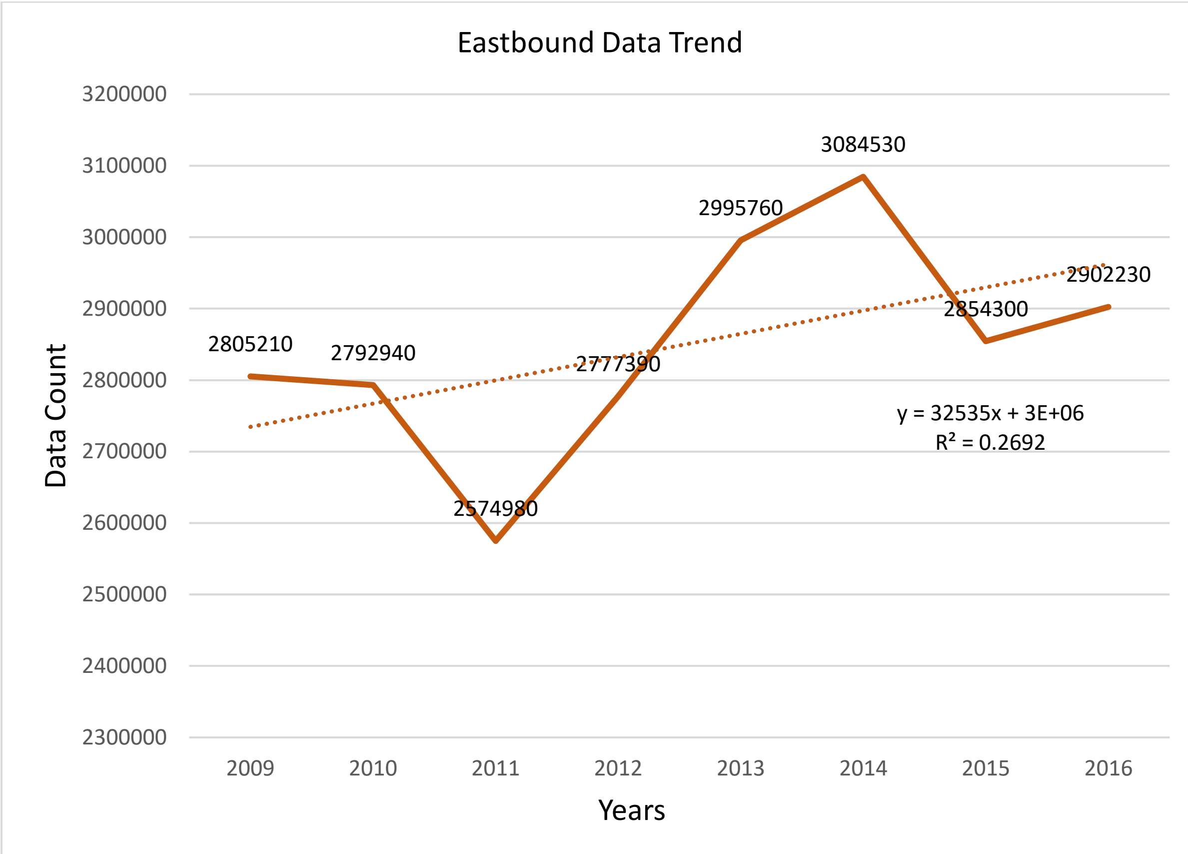

Figure 3.5. The Trend of Eastbound AADT Dataset in Washington State 2009 - 2016

Figure 3.5 above represents the trend of AADT count dataset taken at the Eastbound collection points in Washington State. The data suggest that AADT declined from the year 2009 to 2011. A steady increase is observed between the year 2011 and 2014 and there was a decrease from 2014 to 2015. There was a steady increase in 2016. This showed a positive trend in the relationship between the AADT count dataset and the year with an estimated average of 2,848,418 counts from the year 2009 to the year 2016.

Figure 3.5 positive trend provided a value of R^2 of 0.2692 and a linear equation (Eq. 3.6) provided below:

$$Y = 32535x + 3E + 06 \quad \text{`} \qquad \text{(Eq. 3.6)}$$

Table 3.6. Data Trend Westbound

Years	Total AADT	Percentage Increase in AADT data from the previous year (%)
2009	2844593	0
2010	2910633	0.023216
2011	2712873	-0.06794
2012	2903733	0.070353
2013	3154383	0.08632
2014	3249953	0.030298
2015	3053130	-0.06056
2016	3070490	0.005686

Table 3.6 above provides the summary of the AADT data acquired for each year from 2009 to 2016 and the percentage increase in AADT data for each year for the Westbound collection points in Washington State. There is a total increase of 225,897 from the year 2009 to the year 2016. Table 3.6 showed a percentage decrease in the years 2011 and 2015 with a decrement of 6.79%, and 6.06% respectively from the preceding year. The highest percentage increase in AADT count can be noticed in the year 2012 and 2013 with an increment of 7.04% and 8.63% respectively.

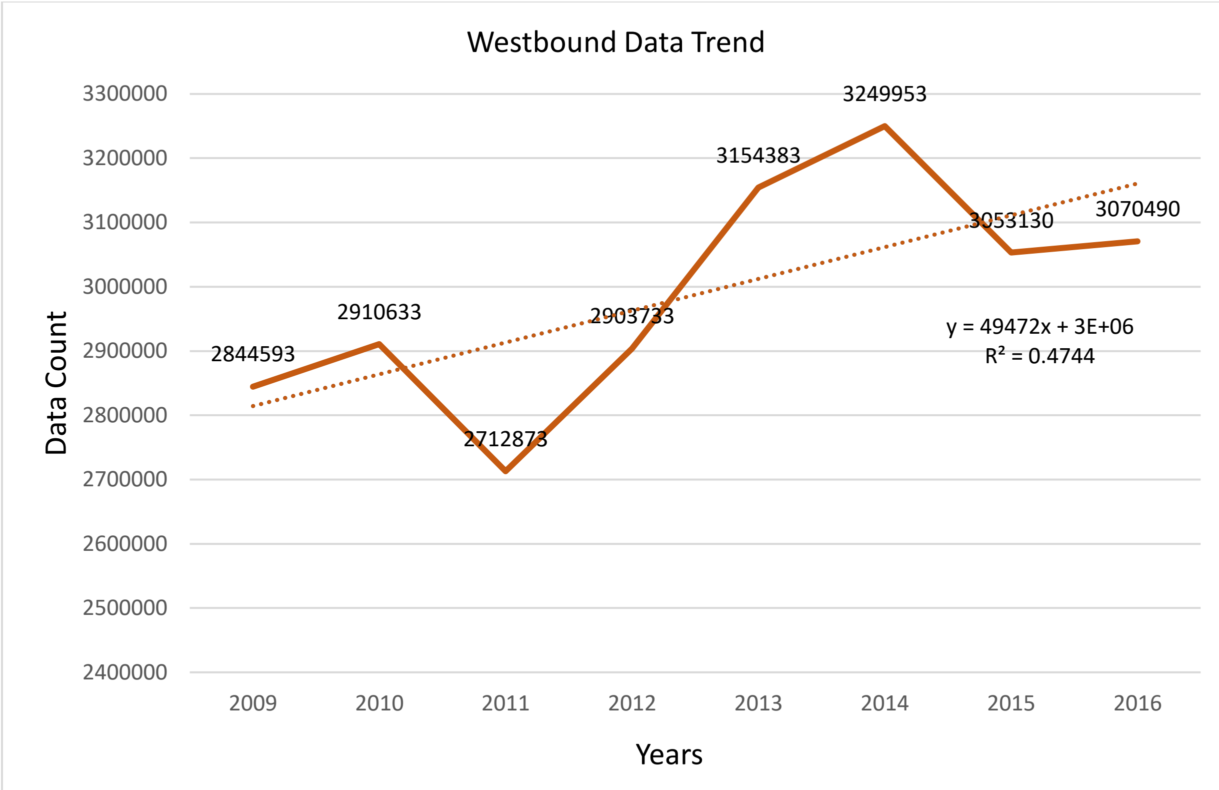

Figure 3.6. The Trend of Westbound AADT in Washington State 2009 -2016

Figure 3.6 above represents the trend of AADT dataset count taken at the Westbound collection points in Washington State. The data suggest that AADT increased from the year 2009 to 2010 and then suffer a steep decrease in the year 2011. A steady increase is observed between the year 2011 and 2014 and there was a decrease from 2014 to 2015. There was a steady increase in 2016. This showed a positive trend in the relationship between the AADT count dataset and the year with an estimated average of 2,987,474 counts from the year 2009 to the year 2016.

Figure 3.6 positive trend provided a value of R^2 of 0.4744, and a linear equation (Eq. 3.7) provided below:

$$Y = 49472x + 3E + 06 \qquad \text{(Eq. 3.7)}$$

3.3.2. Descriptive Analysis Using Yearly Bases

The AADT spatial data used for this section of work consist of data from the year 2009 to 2016 that is made available on the website of the Washington State Department of Transport. The data used is the original data as it was downloaded so that an insight into the raw data about the characteristics and behavioral pattern of the original dataset could be known. Data from previous years (2008 and less) could not be used because it does not have the location properties that were needed to analyze it in ArcGIS software.

Table 3.7. The Year 2009 Dataset Analysis

Variable	N	Mode	Mean	SE Mean	St Dev	Min	Q1	Media n	Q3	Max
Year_2009	7540	12000	20456	389	33739	3	3400	8400	21000	23900 0

Table 3.7 above showed the descriptive data analysis for the year 2009 AADT count dataset. A total number of 154,238,583 counts was recorded with a standard deviation of 33,739. The median for the dataset is 8,400 and a mean of 20,456. The maximum count in the dataset was 239,000 and the minimum was 3 with 12,000 as the mode. The first and third quartile for the data is 3,400 and 21,000 respectively. With the median being 8,400 compared with a mean of 20,456 shows an indication of the data being right-skewed. Figure 3.7 below showed the histogram for the 2009 data.

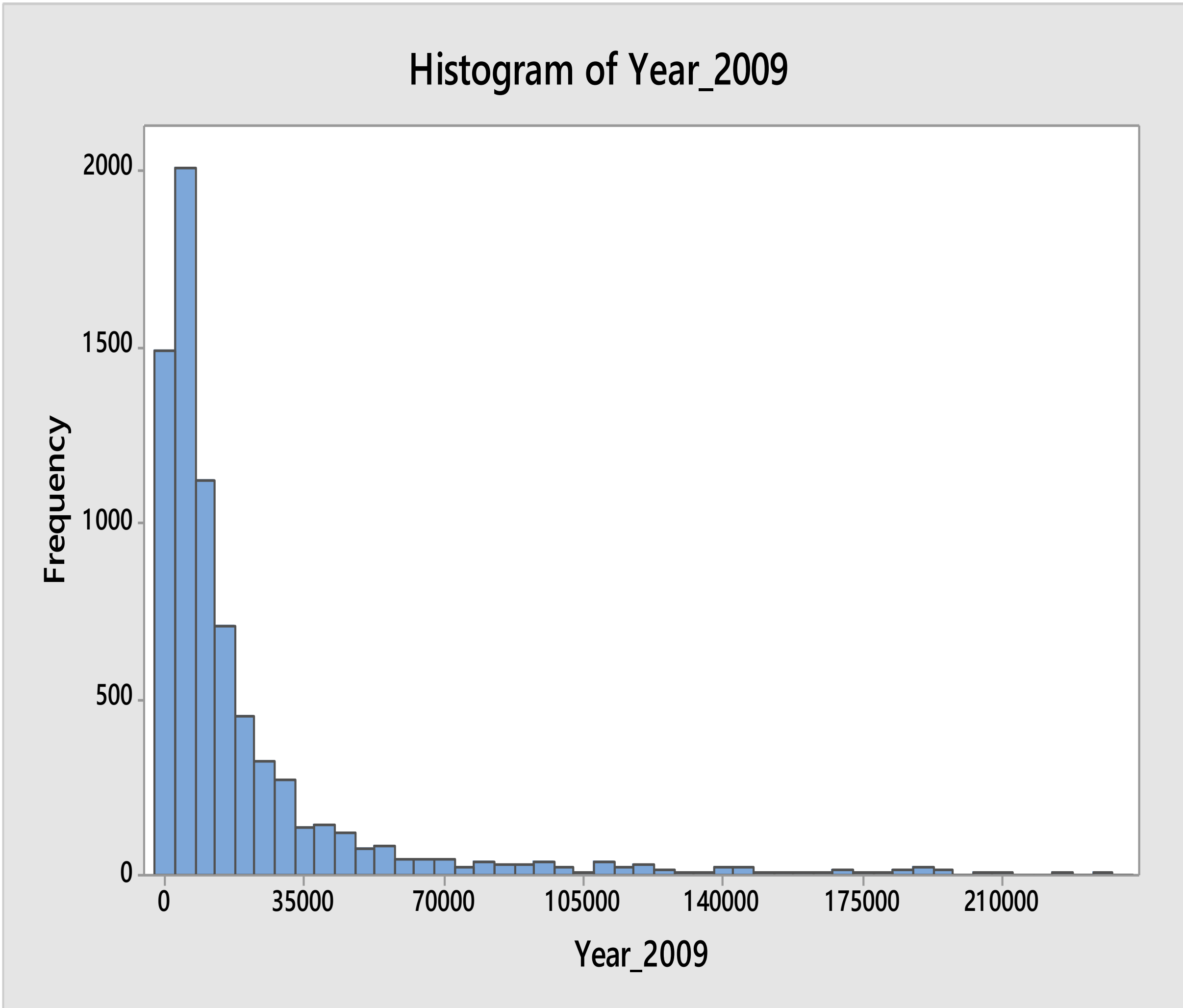

Figure 3.7. Histogram of the Year 2009 AADT

Figure 3.7 above showed that for the year 2009 count dataset, data counts ranging between 0 and 35,000 occurred more in frequency than the rest of counts in the dataset. The higher the data observed increases, the lower the frequency becomes which indicated that the data is positive or right-skewed.

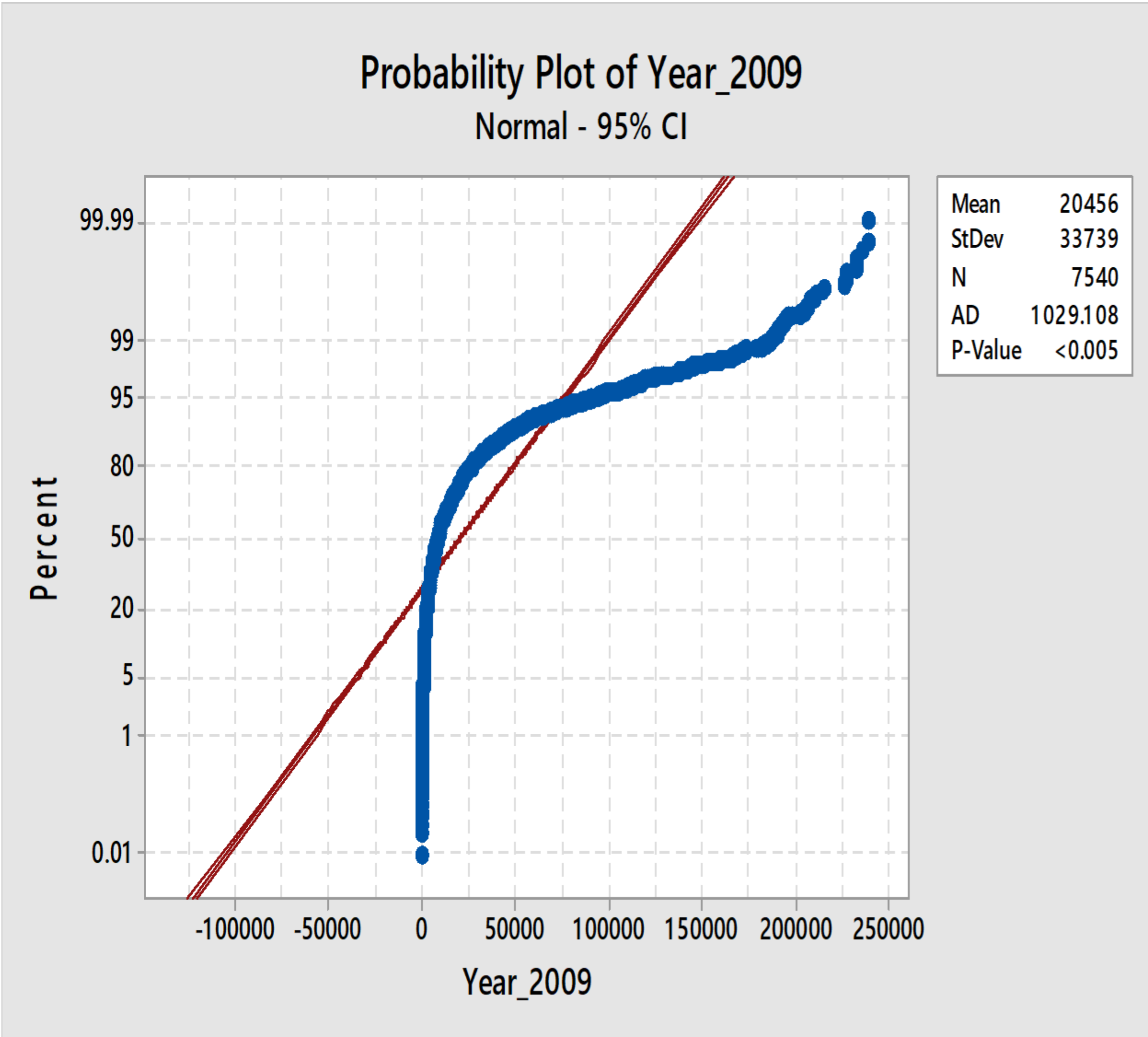

Figure 3.8. Normal Probability Plot for the Year 2009 AADT

Figure 3.8 above showed the distribution of the data count for the year 2009 count dataset. The normal probability plot showed a non-linear pattern indicating that the data does not follow a normal distribution pattern thereby we can reasonably conclude that the dataset is not normally distributed and normal probability plot does not provide an adequate fit for this dataset.

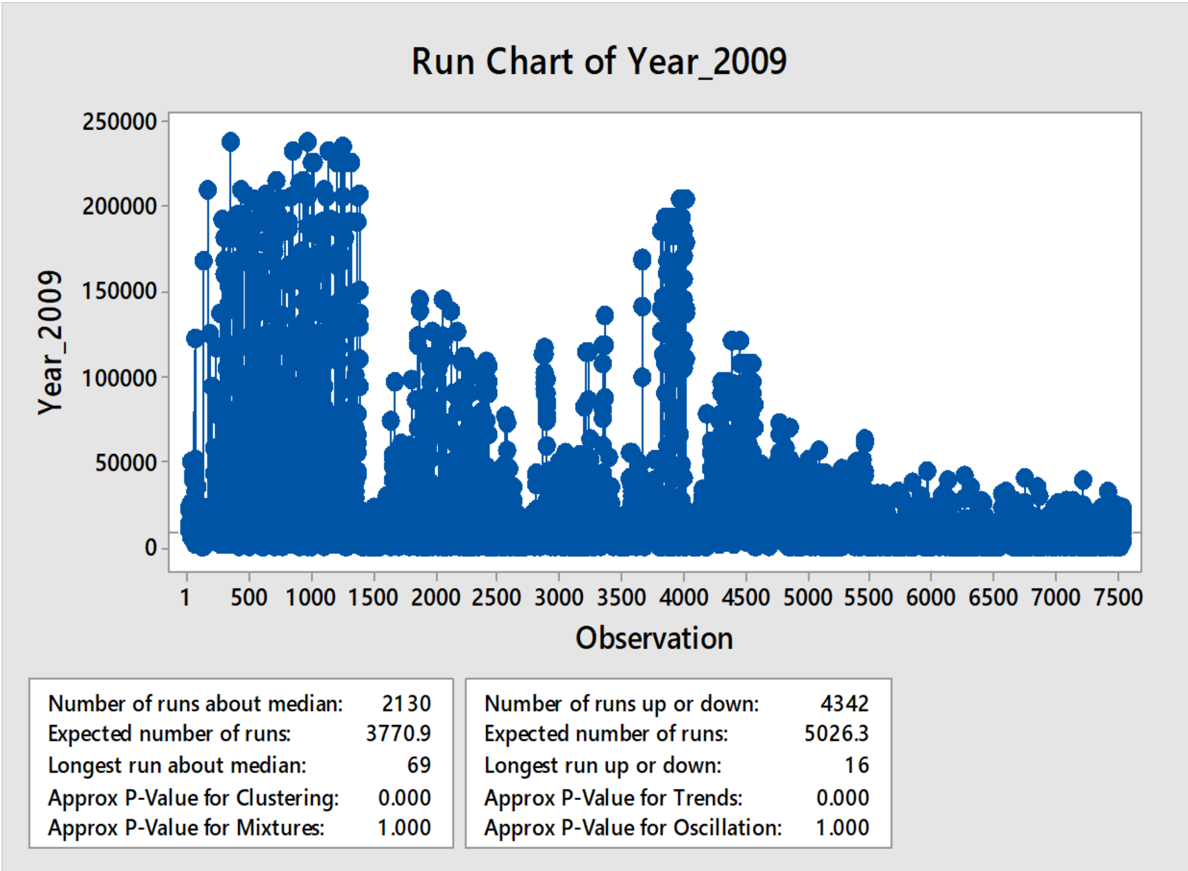

Figure 3.9. Run Sequence Plot for the Year 2009 AADT

The run sequence plot in Figure 3.9 above indicated several significant shifts in different locations in the dataset. This indicated the randomness of the dataset for which the univariate model

$$Y_i = C + E_i \qquad \text{(Eq. 3.8)}$$

is valid.

The camera locations for the year 2009 is shown in Figure 3.10 below using ArcGIS. The camera location points portrayed in altered color for each count division. This reflected the trend with much of the locations on the Westbound side of the state. With the passage of every year, the graphic images exhibit more clusters of the camera points overlaying at various locations.

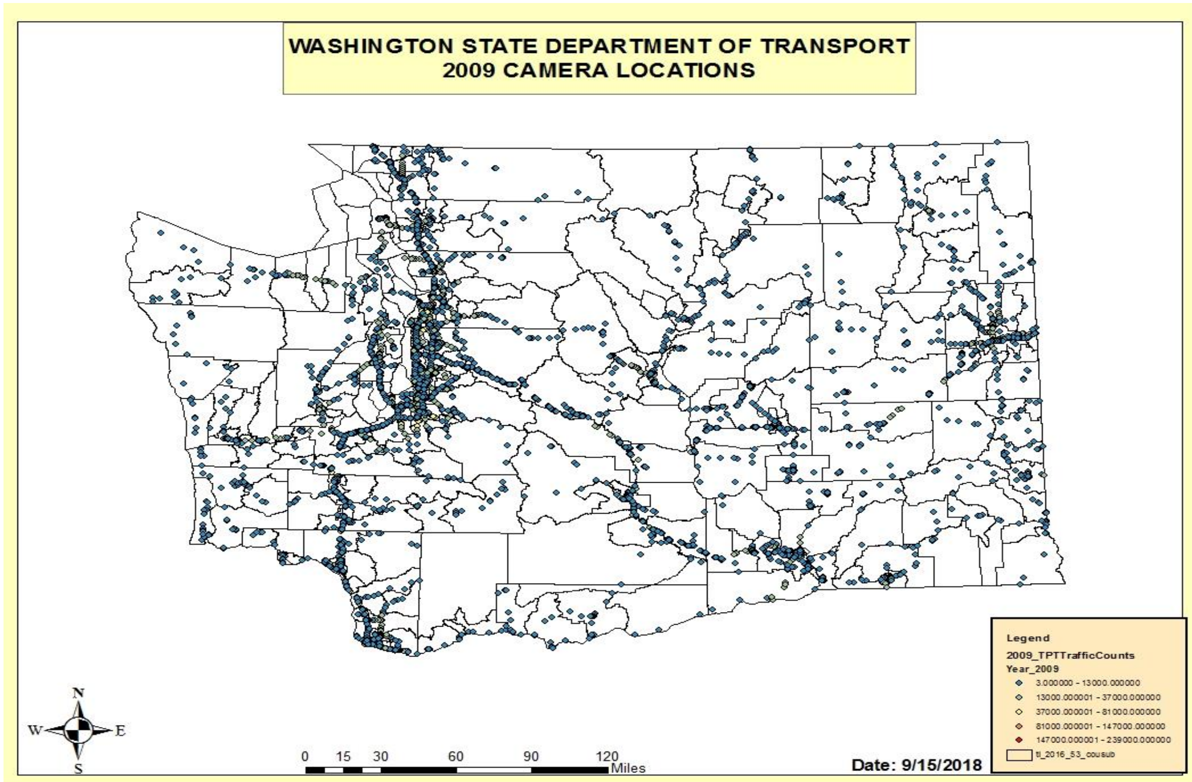

Figure 3.10. 2009 Camera Locations using ArcGIS

Table 3.8. The Year 2010 Dataset Analysis

Variable	N	Mode	Mean	SE Mean	St Dev	Min	Q1	Median	Q3	Max
Year_2010	7789	11000	20108	389	33556	3	3200	8200	20000	237000

Table 3.8 above showed the descriptive data analysis for the year 2010 AADT count dataset. A total number of 156,621,773 counts were recorded with a standard deviation of 33,556. The median for the dataset is 8,200 and a mean of 20,108. The maximum count in the dataset was 237,000 and the minimum was 3 with 11,000 as the mode. The first and third quartile for the data is 3,200 and 20,000 respectively. With the median being 8,200 compared with a mean of 20,108, a plot of the data indicated the data to be right-skewed (Figure 3.11).

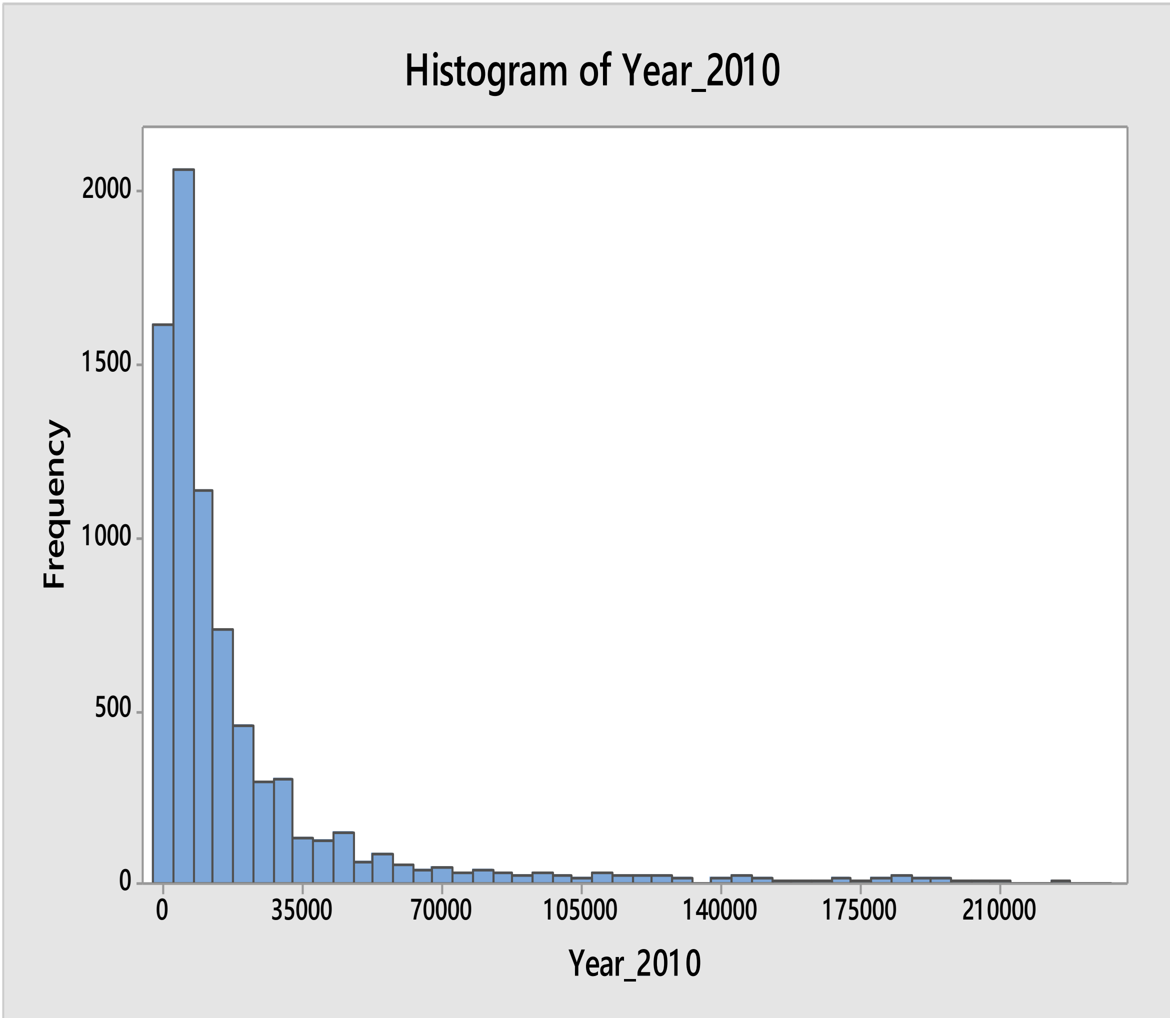

Figure 3.11. Histogram of the Year 2010 AADT

Figure 3.11 above indicated that for the year 2010 count dataset, data counts ranging between 0 and 35,000 occurred more in frequency than the rest of counts in the dataset. The higher the data observed increases, the lower the frequency becomes which indicated that the data is positive or right skewed.

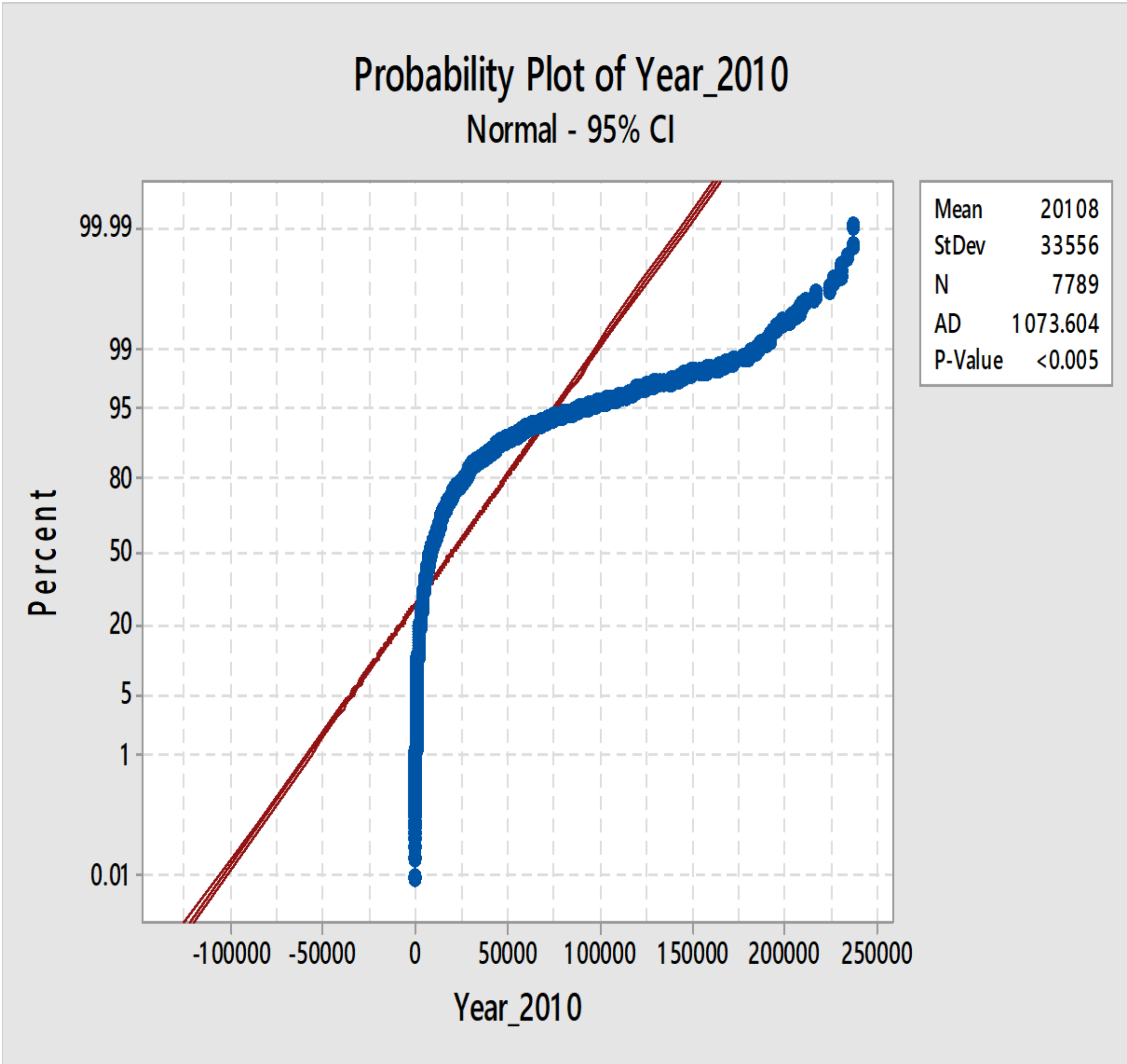

Figure 3.12. Normal Probability Plot for the Year 2010

Figure 3.12 above showed the distribution of the data count for the year 2010 count dataset. The normal probability plot showed a non-linear pattern indicating that the data does not follow a normal distribution pattern thereby we can reasonably conclude that the dataset is not normally distributed and normal probability plot does not provide an adequate fit for this dataset.

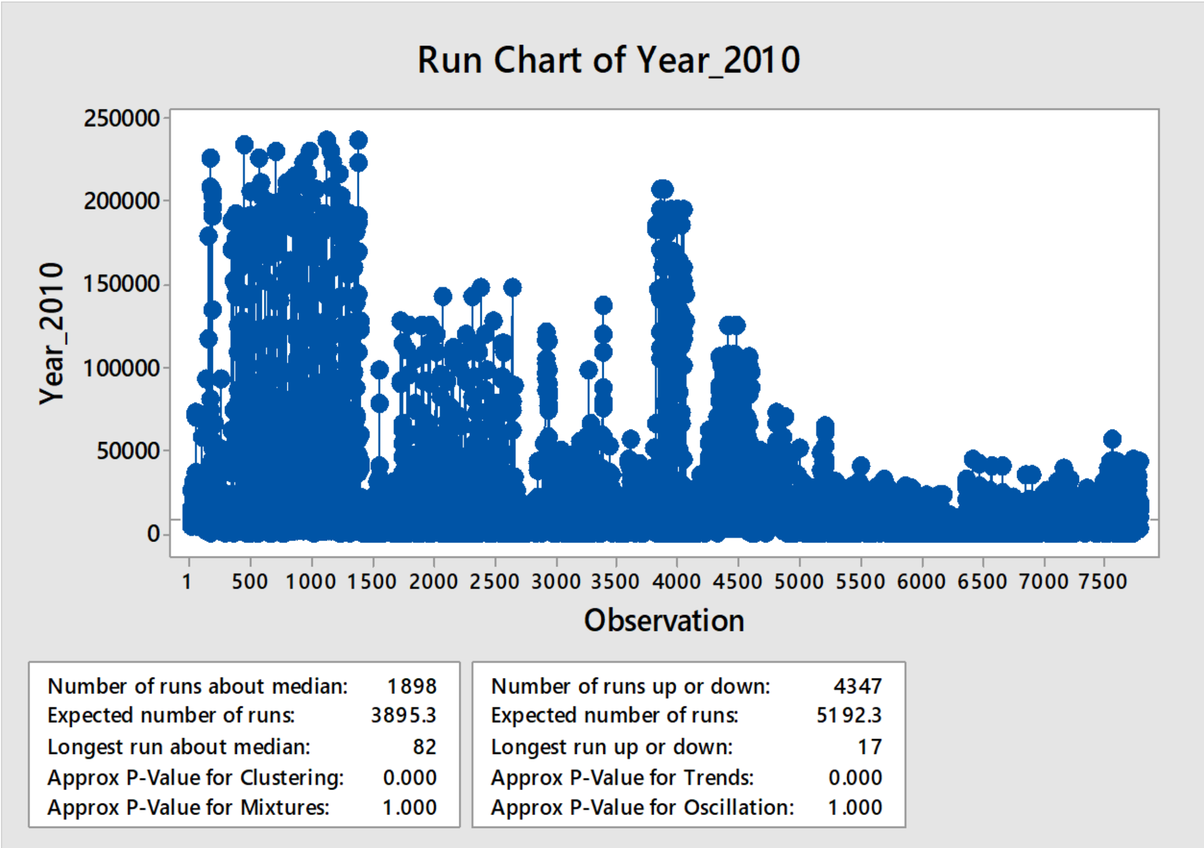

Figure 3.13. Run Sequence Plot for the Year 2010 AADT

The run sequence plot shown in Figure 3.13 above indicated several significant shifts in different locations in the dataset. This indicated the randomness of the dataset for which the univariate model

$$Y_i = C + E_i \quad \text{(Eq. 3.9)}$$

is valid.

The camera locations for the year 2010 is shown in Figure 3.14 below using ArcGIS. The camera location points portrayed in altered color for each count division. Figure 3.14 reflects the trend with much of the locations on the Westbound side of the state. With the passage of every year, the graphic images exhibit more clusters of the camera points overlaying at various locations.

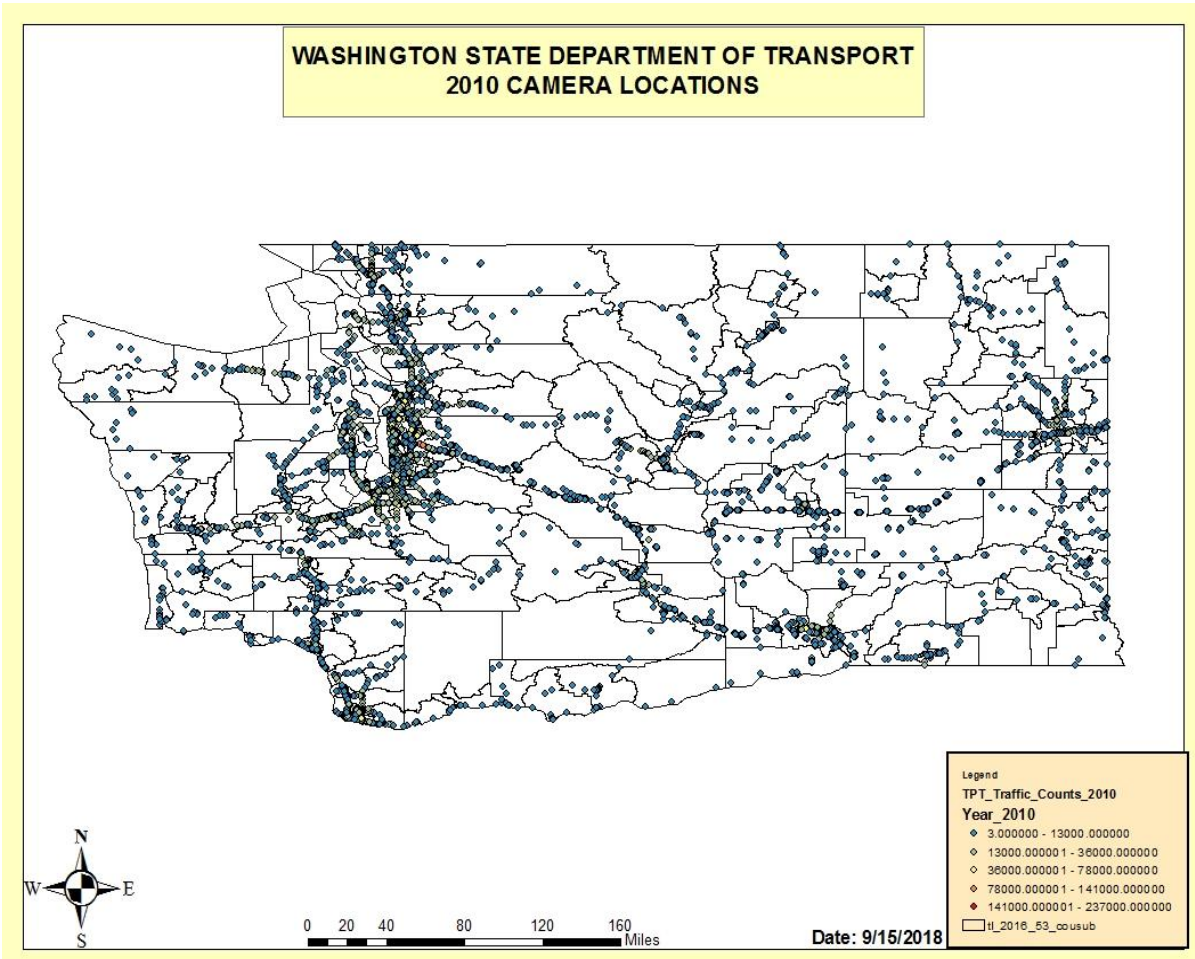

Figure 3.14. 2010 Camera Locations using ArcGIS

Table 3.9. The Year 2011 Dataset Analysis

Variable	N	Mode	Mean	SE Mean	St Dev	Min	Q1	Median	Q3	Max
Year_2011	7645	11000	20023	383	33494	3	3100	8000	20000	229000

Table 3.9 above showed the descriptive data analysis for the year 2011 AADT count dataset. A total number of 153,078,123 counts were recorded with a standard deviation of 33,494. The median for the dataset is 8,000 compared and a mean of 20,023. The maximum count in the dataset was 229,000 and the minimum was 3 with 11,000 as the mode. The first and third quartile

for the data is 3,100 and 20,000 respectively. With the median being 8,000 compared with a mean of 20,023, a plot of the data indicated the data to be right-skewed (Figure 3.15).

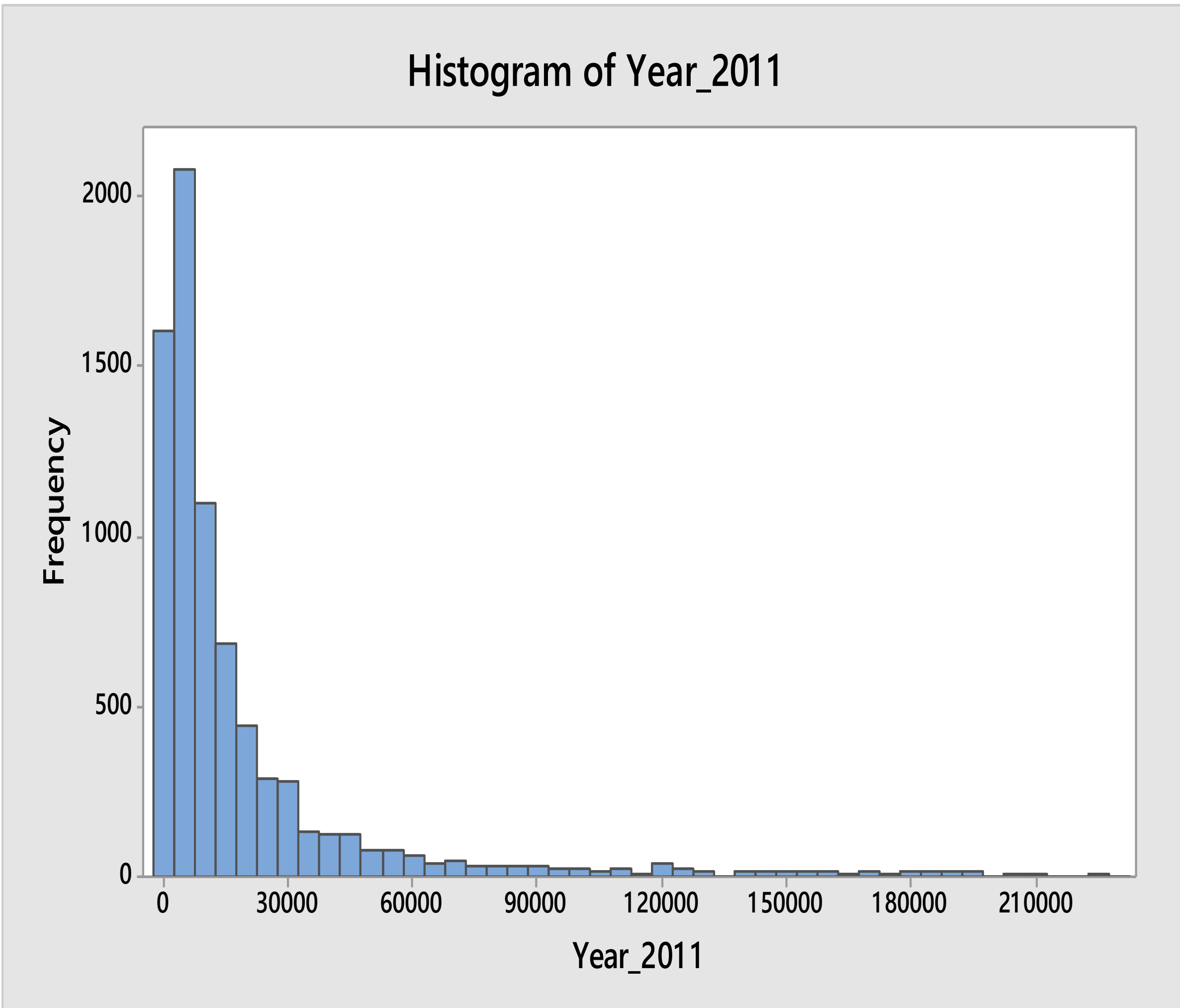

Figure 3.15. Histogram of the Year 2011 AADT

Figure 3.15 above indicated that for the year 2011 count dataset. Data counts ranging between 0 and 35,000 occurred more in frequency than the rest of counts in the dataset. The higher the data observed increases, the lower the frequency becomes which indicated that the data been positive or right skewed.

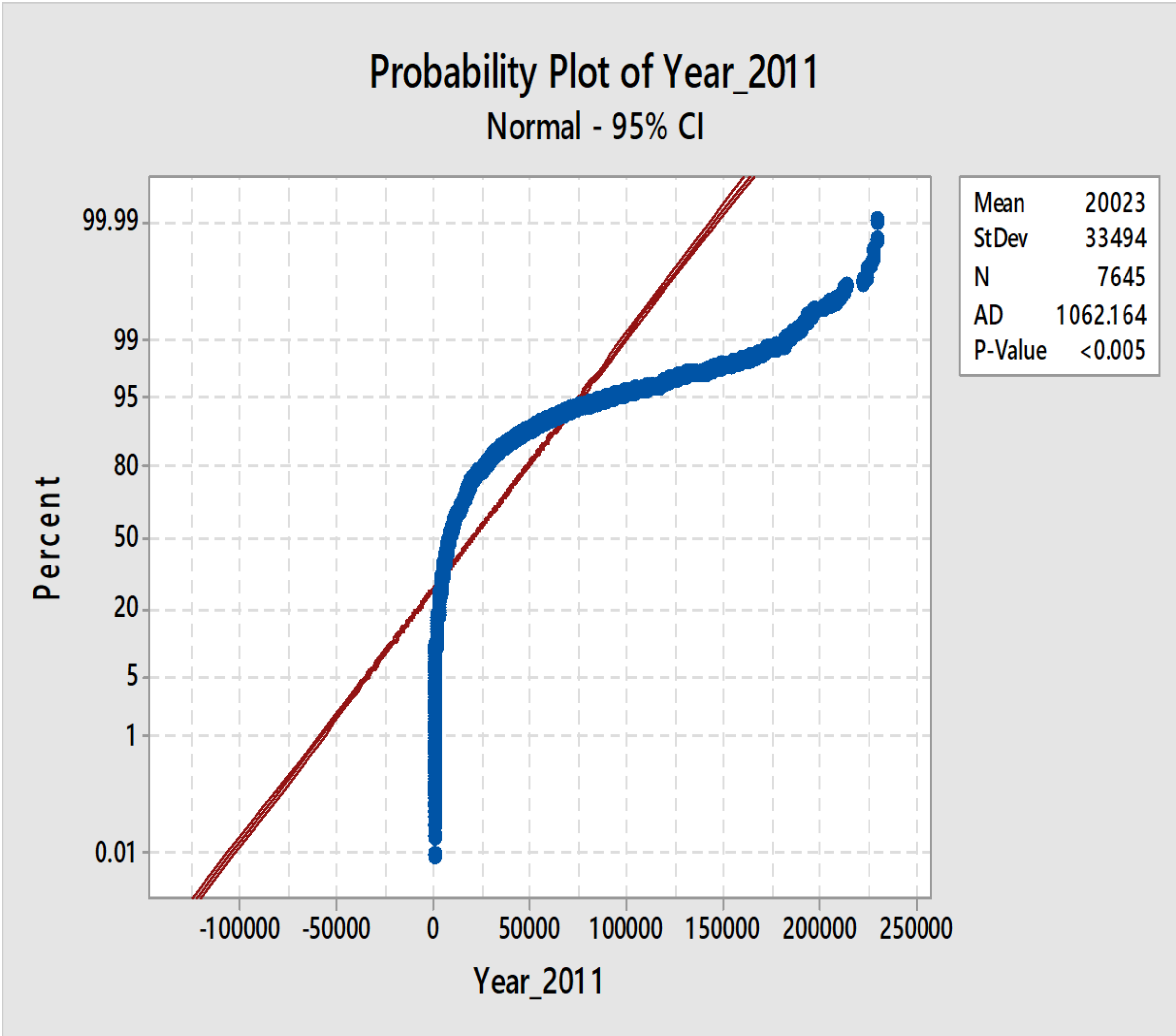

Figure 3.16. Normal Probability Plot for the Year 2011 AADT

Figure 3.16 above showed the distribution of the data count for the year 2011 count dataset. The normal probability plot showed a non-linear pattern indicating that the data does not follow a normal distribution pattern thereby we can reasonably conclude that the dataset is not normally distributed and normal probability plot does not provide an adequate fit for this dataset.

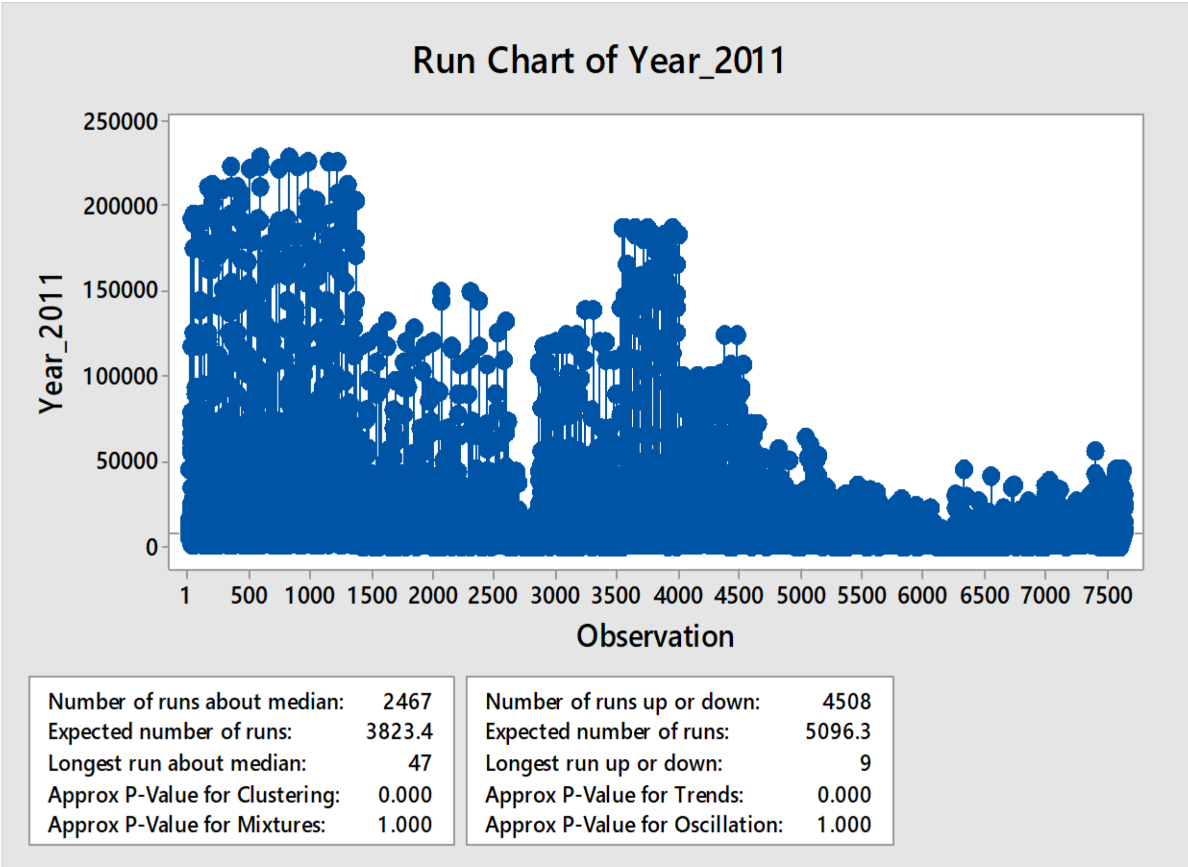

Figure 3.17. Run Sequence Plot for the Year 2011 AADT

The run sequence plot in Figure 3.17 above indicated several significant shifts in different locations in the dataset. This indicated the randomness of the dataset for which the univariate model

$$Y_i = C + E_i \tag{Eq. 3.10}$$

is valid.

The camera locations for the year 2011 is shown in Figure 3.18 below using ArcGIS. The camera location points portrayed in altered color for each count division. Figure 3.18 reflects the trend with much of the locations on the Westbound area of the state. With the passage of every year, the graphic images exhibit more clusters of the camera points overlaying at various locations.

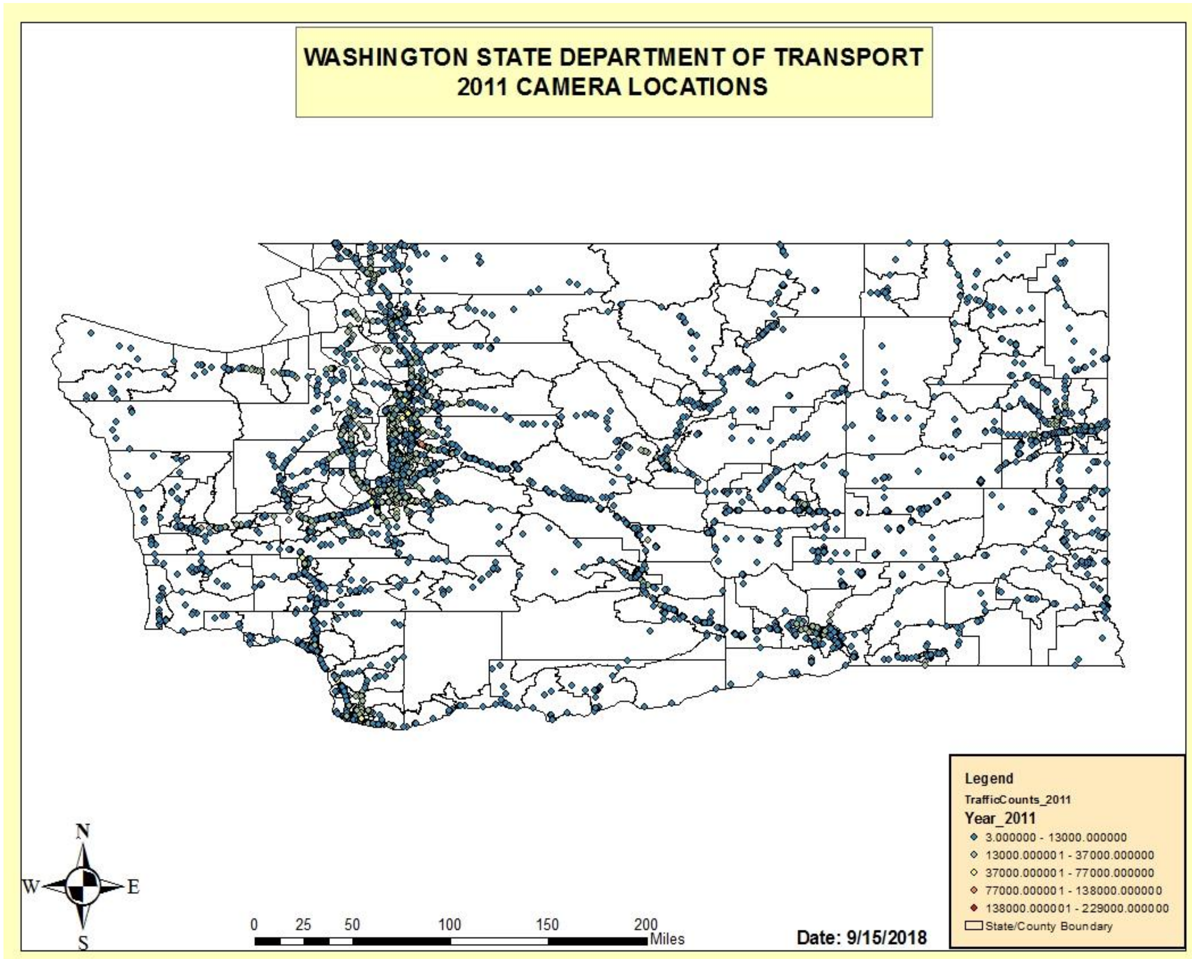

Figure 3.18. 2011 Camera Locations using ArcGIS

Table 3.10. The Year 2012 Dataset Analysis

Variable	N	Mode	Mean	SE Mean	St Dev	Min	Q1	Median	Q3	Max
Year_2012	7693	11000	19890	378	33187	3	3100	8000	20000	228000

Table 3.10 showed descriptive data analysis for the year 2012 AADT count dataset. A total number of 153,013,802 counts were recorded with a standard deviation of 33,187. The median for the dataset is 8,000 compared with a mean of 19,890. The maximum count in the dataset was 228,000 and the minimum was 3 with 11,000 as the mode. The first and third quartile for the data

is 3,100 and 20,000 respectively. With the median being 8,000 compared with a mean of 19,890, a plot of the data indicated the data to be right-skewed (Figure 3.19).

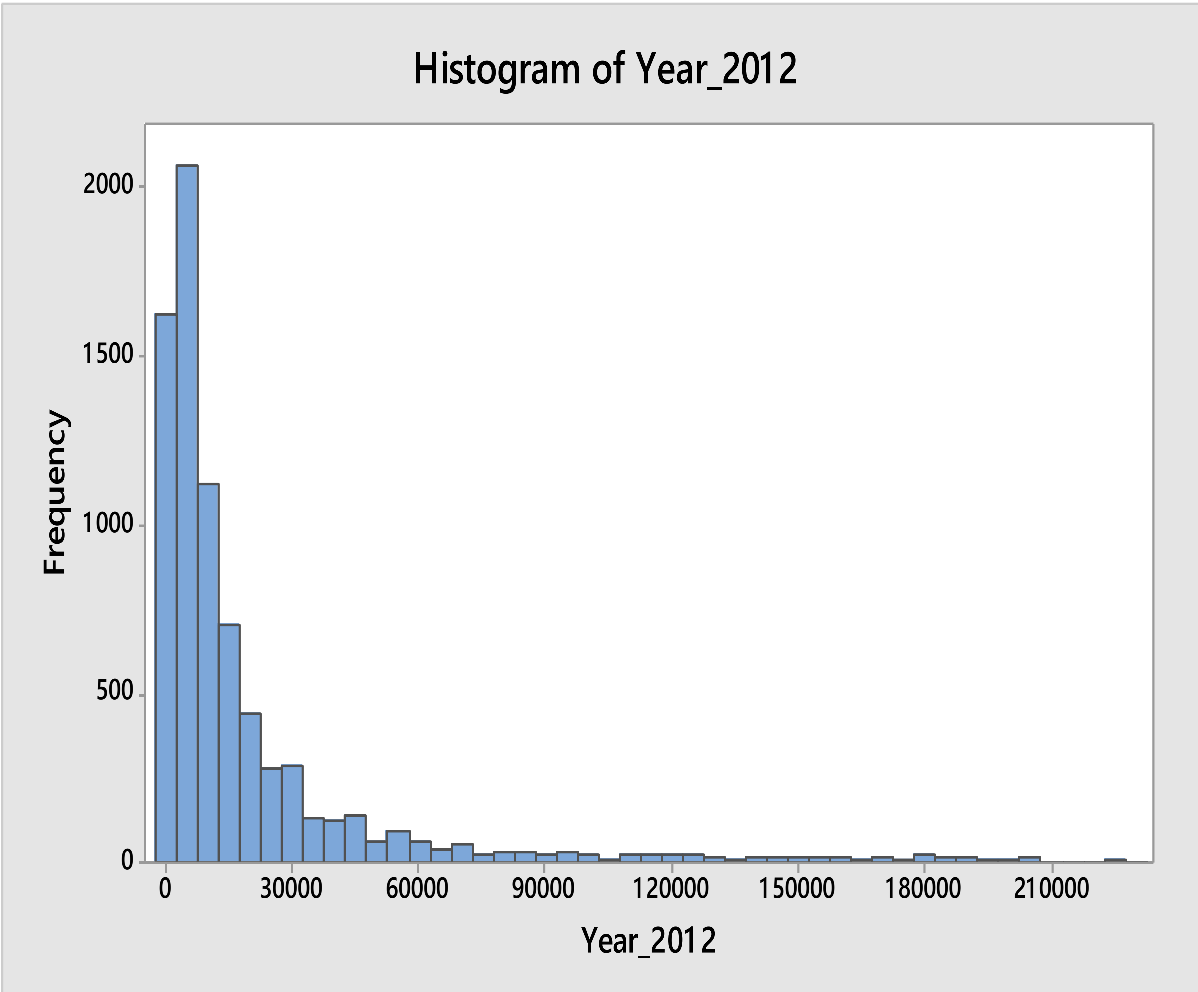

Figure 3.19. Histogram of the Year 2012 AADT

Figure 3.19 above indicated that for the year 2012 count dataset, data counts ranging between 0 and 35,000 occurred more in frequency than the rest of counts in the dataset. The higher the data observed increases, the lower the frequency becomes which indicated that the data is positive or right skewed.

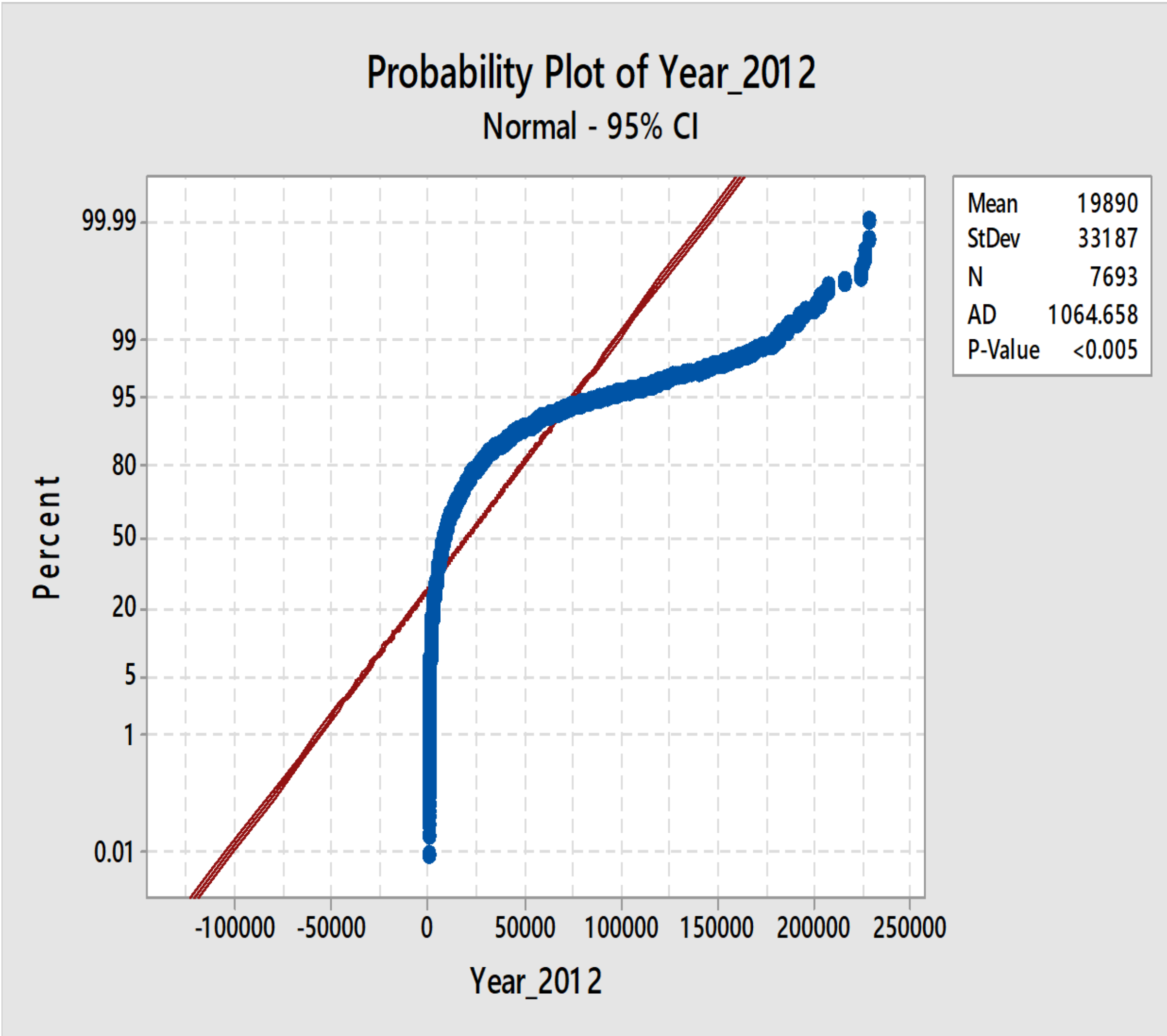

Figure 3.20. Normal Probability Plot for the Year 2012 AADT

Figure 3.20 above showed the distribution of the data count for the year 2012 count dataset. The normal probability plot showed a non-linear pattern indicating that the data does not follow a normal distribution pattern thereby we can reasonably conclude that the dataset is not normally distributed and normal probability plot does not provide an adequate fit for this dataset.

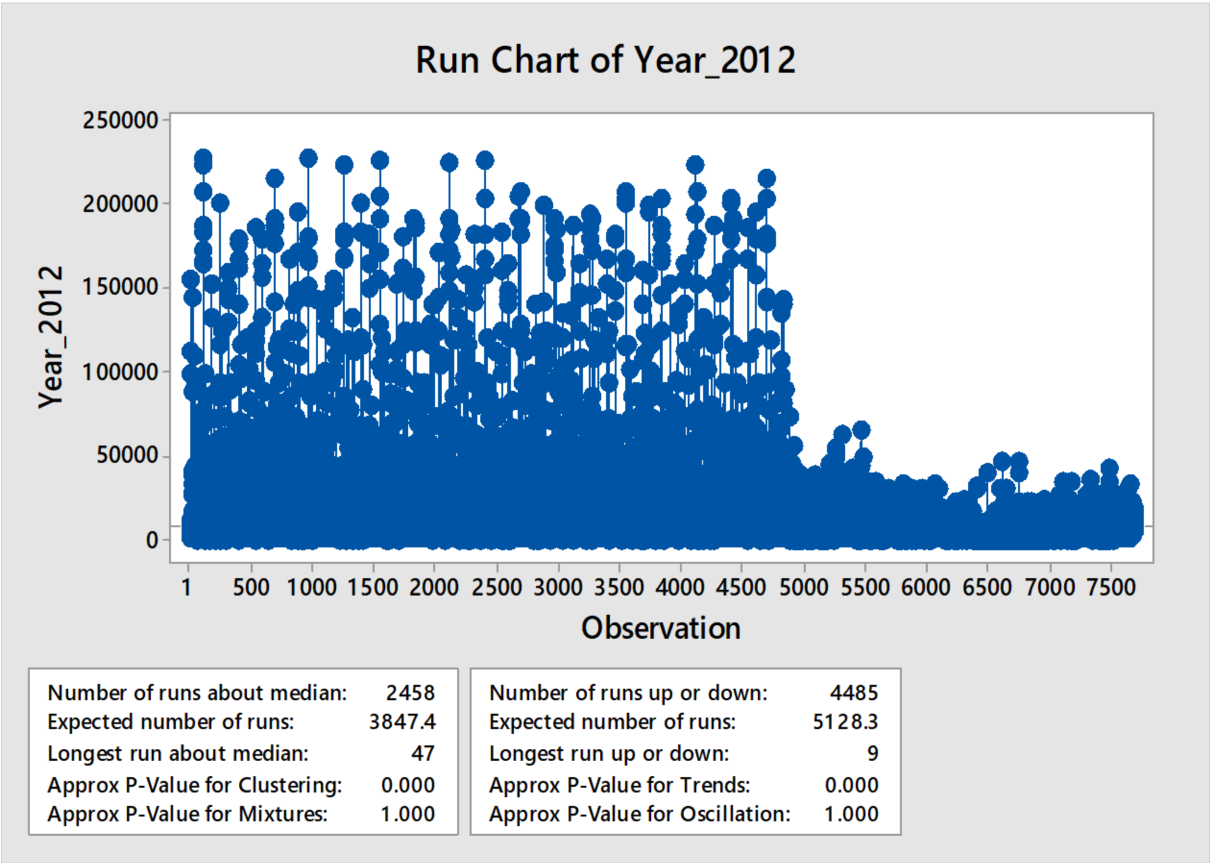

Figure 3.21. Run Sequence Plot for the Year 2012 AADT

The run sequence plot in figure 3.27 above indicated several significant shifts in different locations in the dataset. This indicated the randomness of the dataset for which the univariate model

$$Y_i = C + E_i \qquad \text{(Eq. 3.11)}$$

is valid.

The camera locations for the year 2012 is shown in figure 3.28 below using ArcGIS. The camera location points portrayed in altered color for each count division. Figure 3.28 reflects the trend with much of the locations on the Westbound side of the state. With the passage of every year, the graphic images exhibit more clusters of the camera points overlaying at various locations.

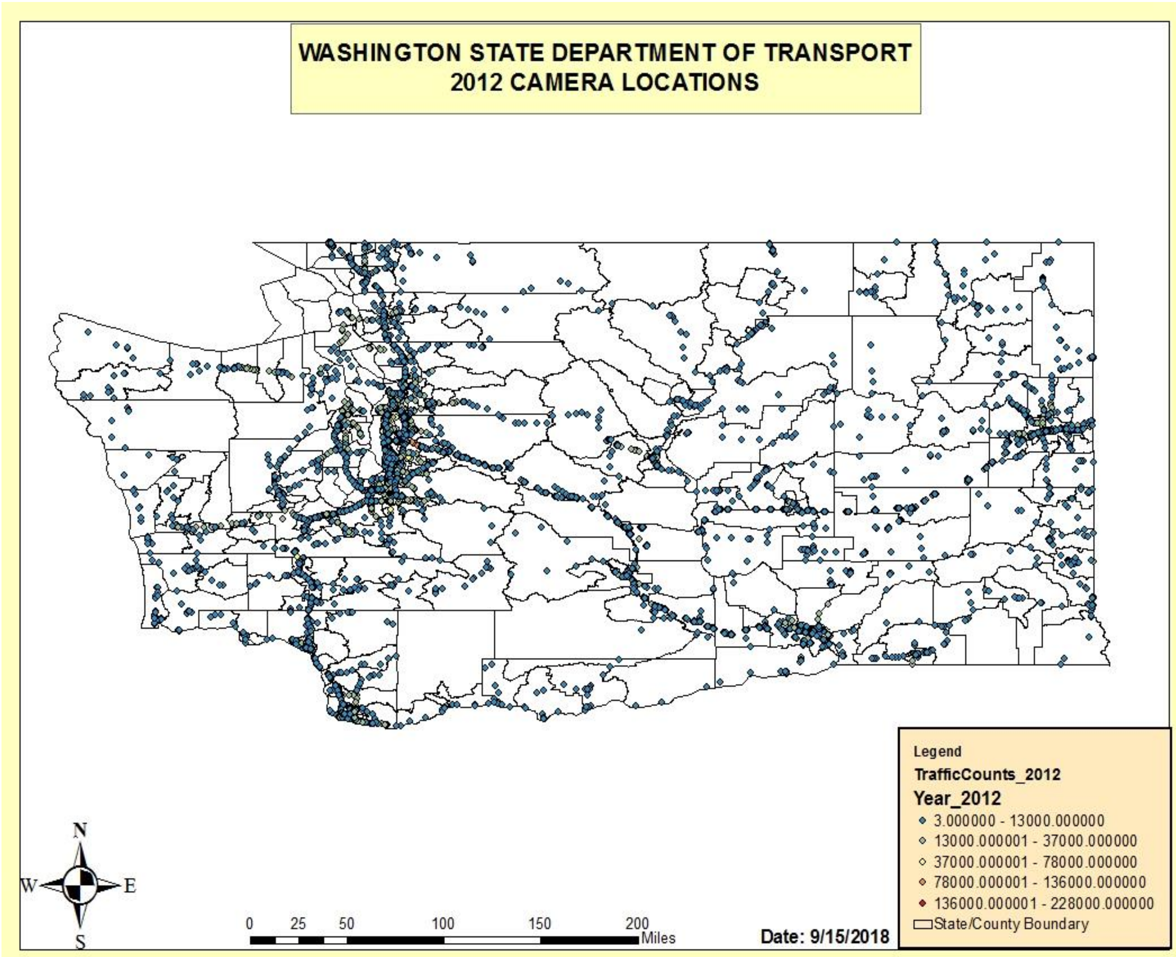

Figure 3.22. 2012 Camera Locations using ArcGIS

Table 3.11. The Year 2013 Dataset Analysis

Variable	N	Mode	Mean	SE Mean	St Dev	Min	Q1	Median	Q3	Max
Year_2013	7788	11000	20022	380	33534	3	3100	8000	20000	229000

Table 3.11 showed descriptive data analysis for the year 2013 AADT count dataset. A total number of 155,775,333 counts were recorded with a standard deviation of 33,534. The median for the dataset is 8,000 compared with a mean of 20,022. The maximum count in the dataset was 229,000 and the minimum was 3 with 11,000 as the mode. The first and third quartile for the data

is 3,100 and 20,000 respectively. With the median being 8,000 compared with a mean of 20,022, a plot of the data indicated the data to be right-skewed (Figure 3.23).

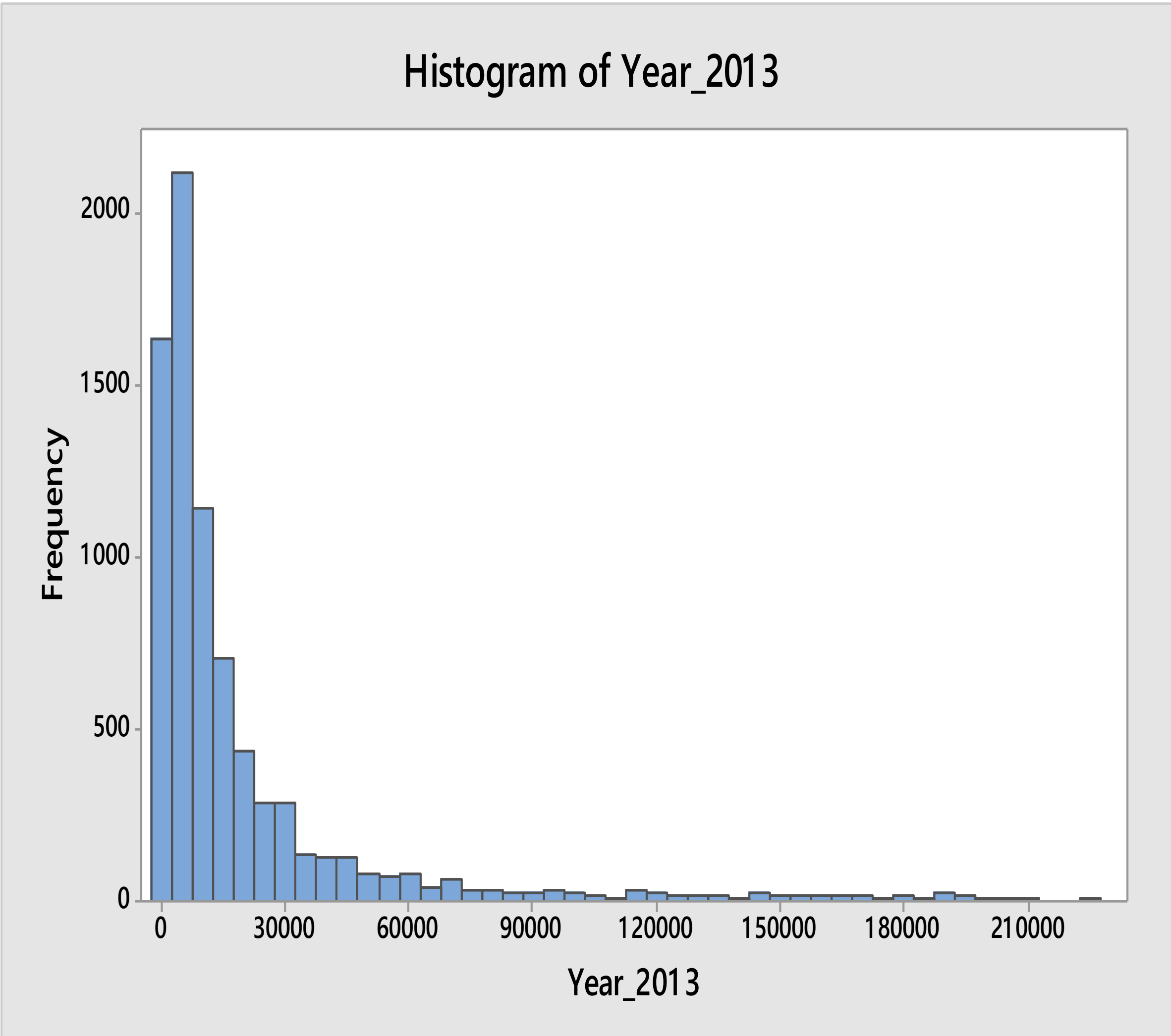

Figure 3.23. Histogram of the Year 2013 AADT

Figure 3.23 above indicated that for the year 2013 count dataset, data counts ranging between 0 and 35,000 occurred more in frequency than the rest of counts in the dataset. The higher the data observed increases, the lower the frequency becomes which indicated that the data is positive or right skewed.

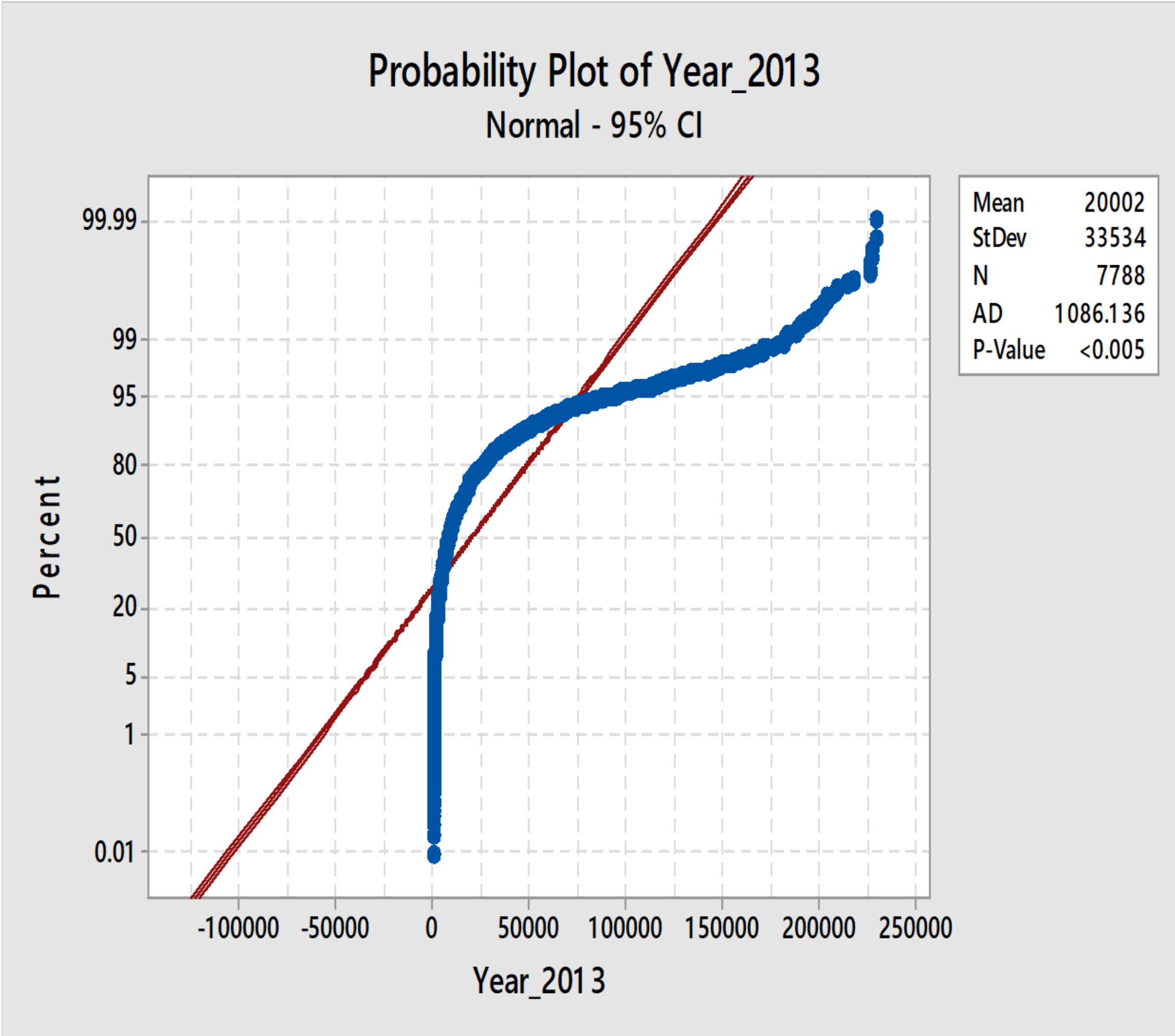

Figure 3.24. Normal Probability Plot for the Year 2013 AADT

Figure 3.24 above showed the distribution of the data count for the year 2013 count dataset. The normal probability plot showed a non-linear pattern indicating that the data does not follow a normal distribution pattern thereby we can reasonably conclude that the dataset is not normally distributed and normal probability plot does not provide an adequate fit for this dataset.

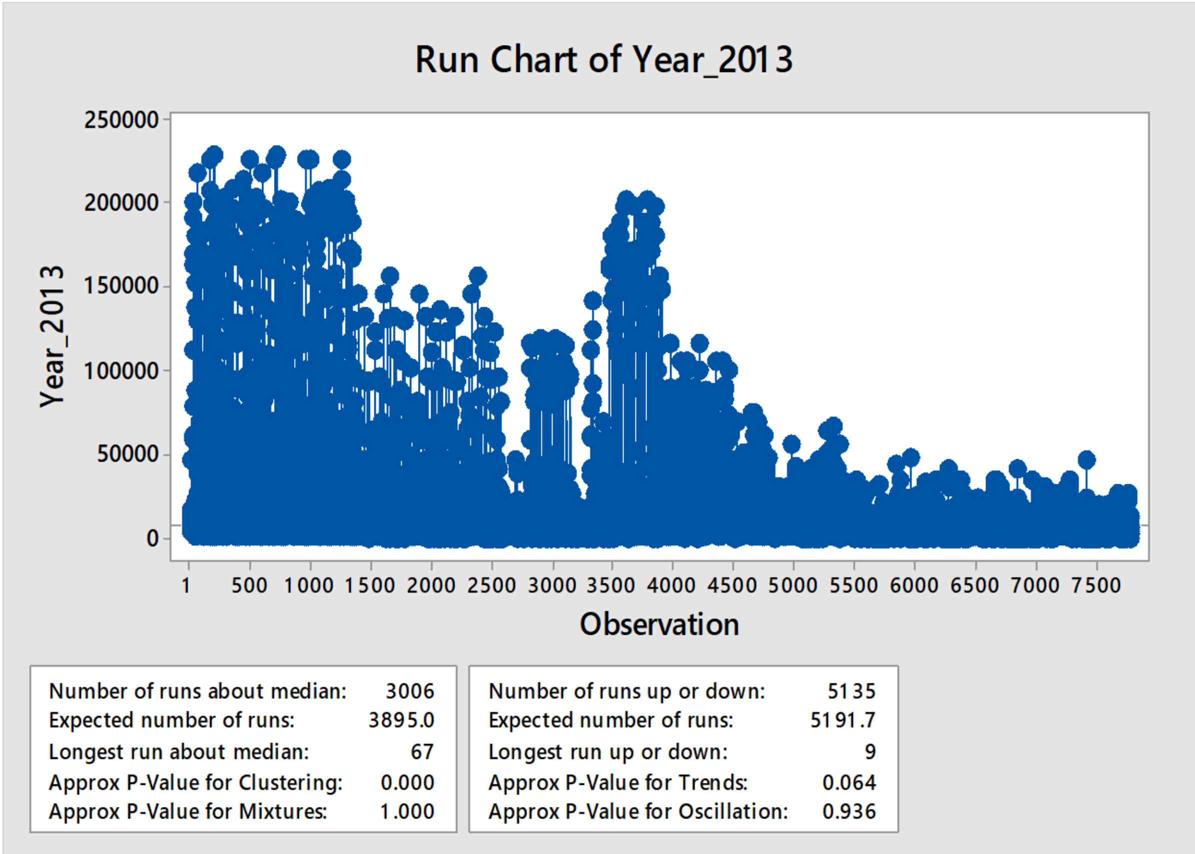

Figure 3.25. Run Sequence Plot for the Year 2013 AADT

The Run sequence plot from Figure 3.25 above indicated several significant shifts in different locations in the dataset. This indicated the randomness of the dataset for which the univariate model

$$Y_i = C + E_i \qquad \text{(Eq. 3.12)}$$

is valid.

The camera locations for the year 2013 is shown in Figure 3.26 below using ArcGIS. The camera location points portrayed in altered color for each count division. Figure 3.326 reflects the trend with much of the locations on the Westbound side of the state. With the passage of every year, the graphic images exhibit more clusters of the camera points overlaying at various locations.

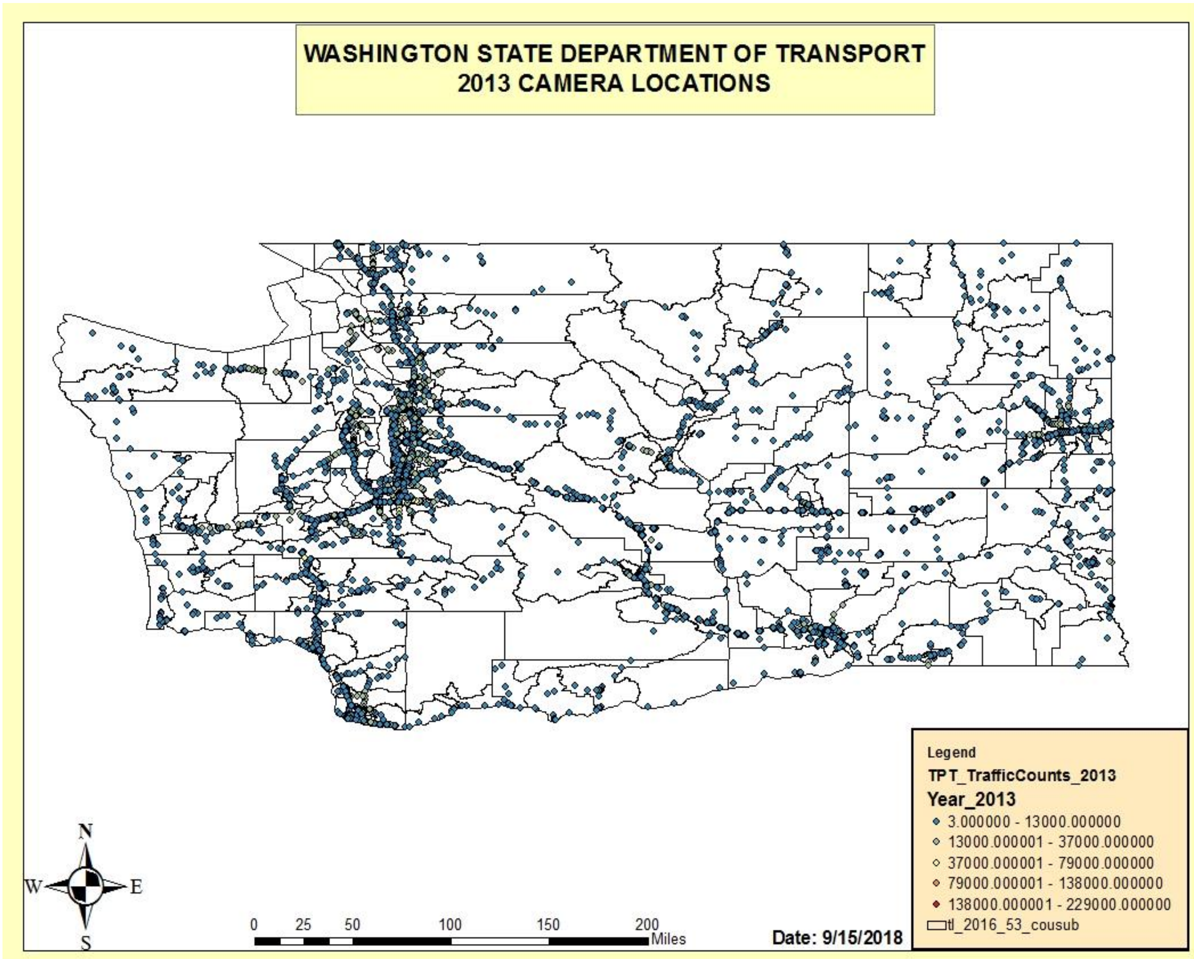

Figure 3.26. 2013 Camera Locations using ArcGIS

Table 3.12. The Year 2014 Dataset Analysis

Variable	N	Mode	Mean	SE Mean	St Dev	Min	Q1	Median	Q3	Max
Year_2014	7692	11000	20381	388	34071	3	3200	8100	20000	232000

Table 3.12 showed descriptive data analysis for the year 2014 AADT count dataset. A total number of 3,084,530 counts were recorded with a standard deviation of 34,071. The median for the dataset is 8,100 compared with a mean of 20,381. The maximum count in the dataset was 232,000 and the minimum was 3 with 11,000 as the mode. The first and third quartile for the data

is 3,200 and 20,000 respectively. With the median being 8,100 compared with a mean of 20, a plot of the data indicated the data to be right-skewed (Figure 3.27).

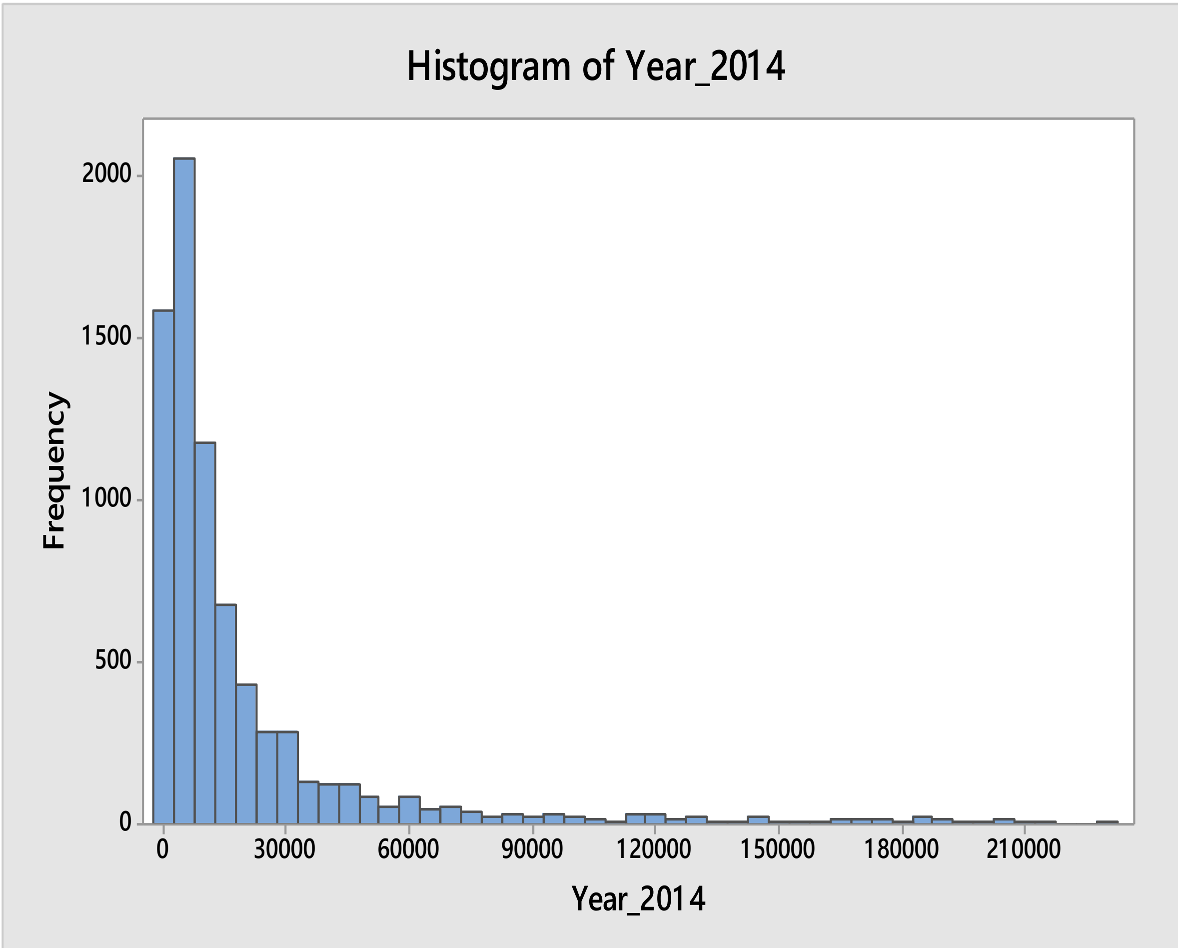

Figure 3.27. Histogram of the Year 2014 AADT

Figure 3.27 above indicated that for the year 2014 count dataset, data counts ranging between 0 and 35,000 occurred more in frequency than the rest of counts in the dataset. The higher the data observed increases, the lower the frequency becomes which indicated that the data is positive or right skewed.

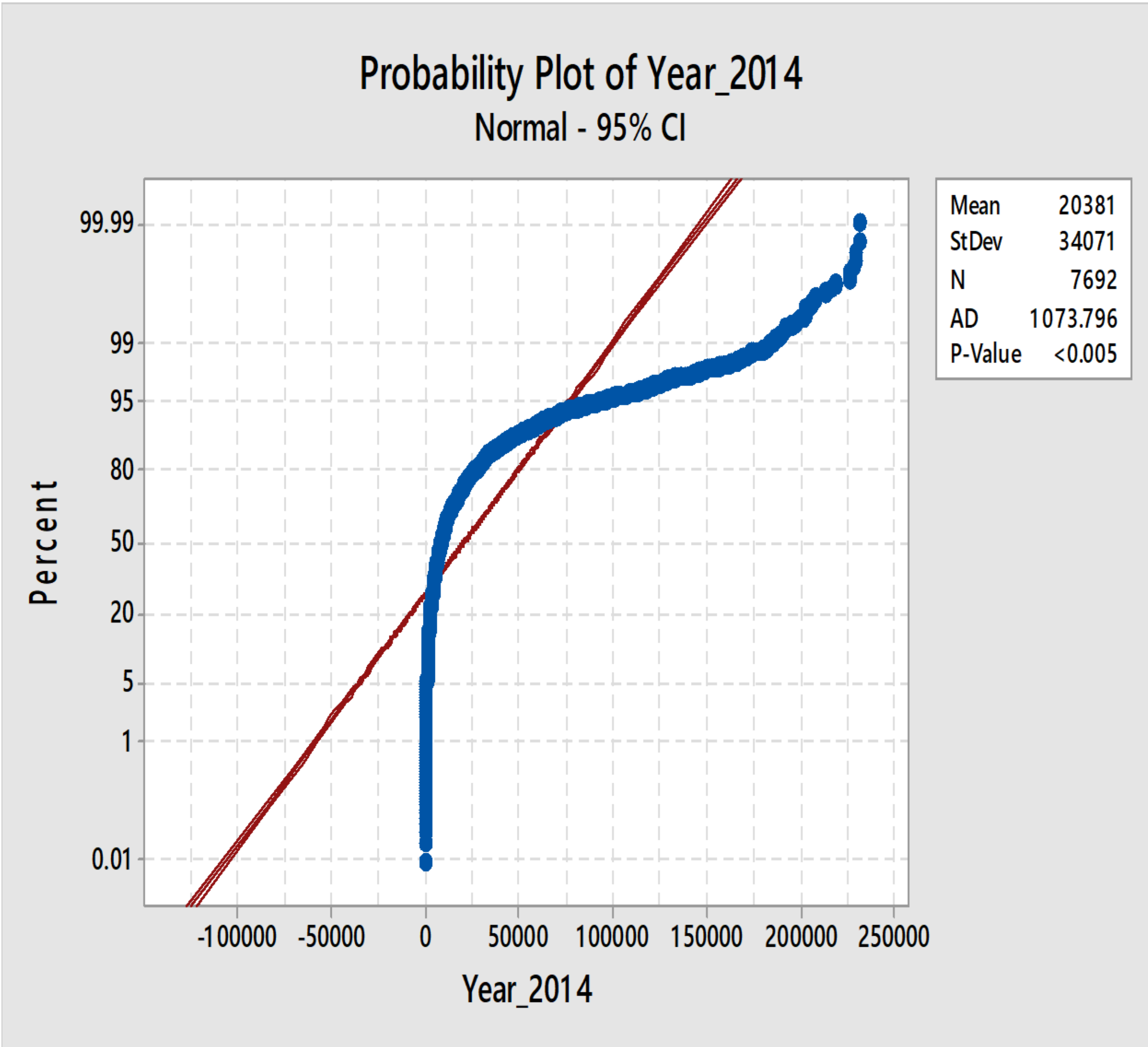

Figure 3.28. Normal Probability Plot for the Year 2014 AADT

Figure 3.28 above showed the distribution of the data count for the year 2014 count dataset. The normal probability plot showed a non-linear pattern indicating that the data does not follow a normal distribution pattern thereby we can reasonably conclude that the dataset is not normally distributed and normal probability plot does not provide an adequate fit for this dataset.

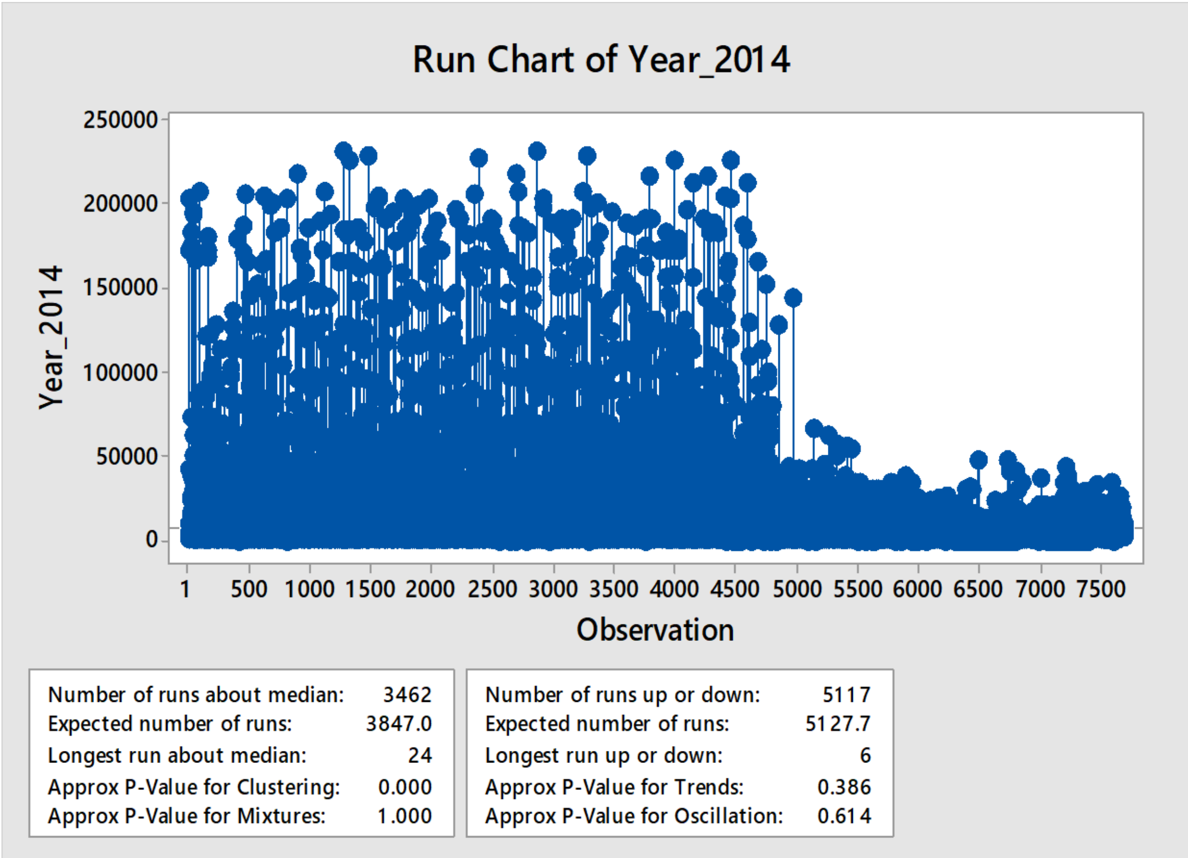

Figure 3.29. Run Sequence Plot for the Year 2014 AADT

The Run sequence plot from Figure 3.29 above indicated several significant shifts in different locations in the dataset. This indicated the randomness of the dataset for which the univariate model

$$Y_i = C + E_i \qquad \text{(Eq. 3.13)}$$

is valid.

The camera locations for the year 2014 is shown in Figure 3.30 below using ArcGIS. The camera location points portrayed in altered color for each count division. Figure 3.30 reflects the trend with much of the locations on the Westbound side of the state. With the passage of every year, the graphic images exhibit more clusters of the camera points overlaying at various locations.

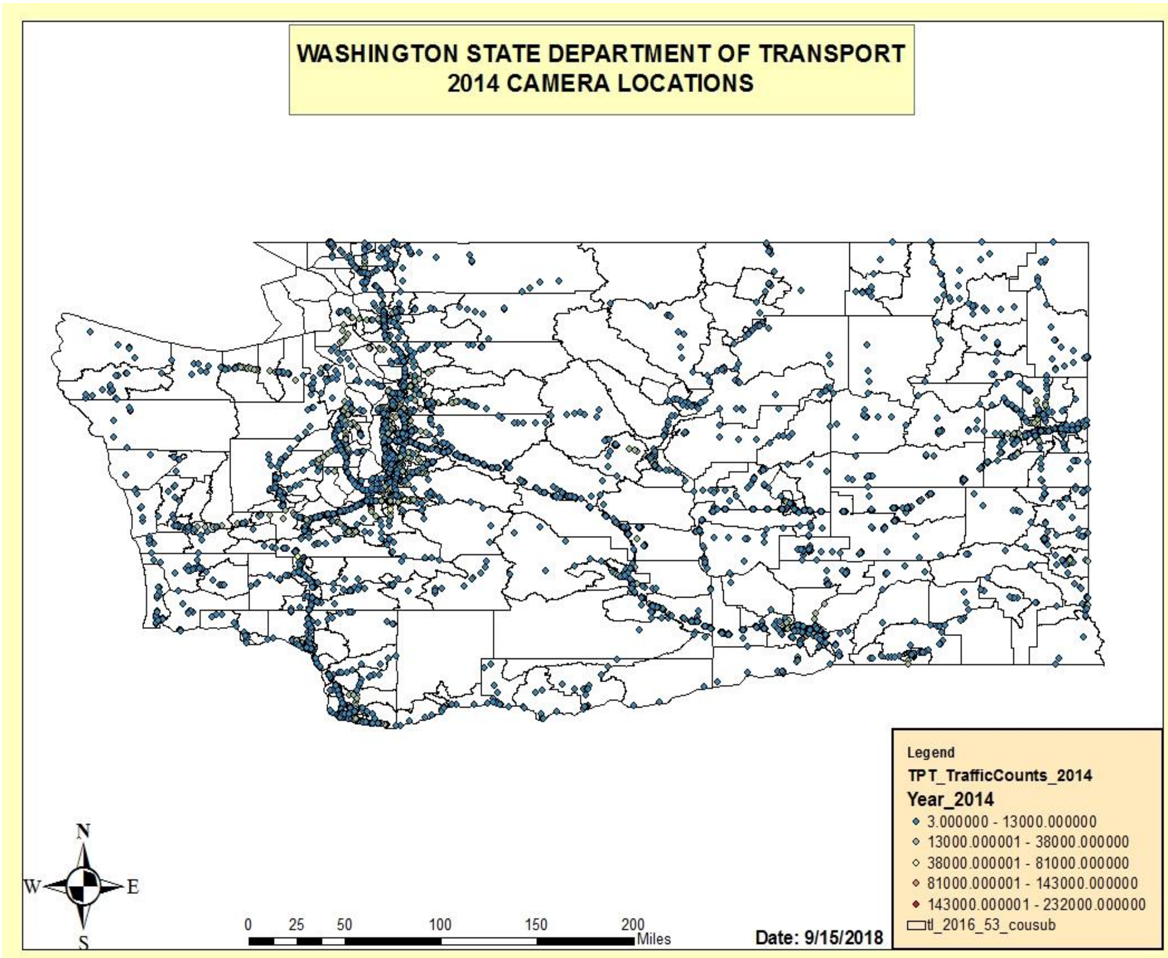

Figure 3.30. 2014 Camera Locations using ArcGIS

Table 3.13. The Year 2015 Dataset Analysis

Variable	N	Mode	Mean	SE Mean	St Dev	Min	Q1	Median	Q3	Max
Year_2015	7333	11000	21591	416	35652	20	3300	8500	22000	242000

Table 3.13 showed descriptive data analysis for the year 2015 AADT count dataset. A total number of 158,328,420 counts were recorded with a standard deviation of 35,652. The median for the dataset is 8,500 compared with a mean of 21,591. The maximum count in the dataset was 242,000 and the minimum was 20 with 11,000 as the mode. The first and third quartile for the data

is 3,300 and 22,000 respectively. With the median being 8,500 compared with a mean of 21,591, a plot of the data indicated the data to be right-skewed (Figure 3.31).

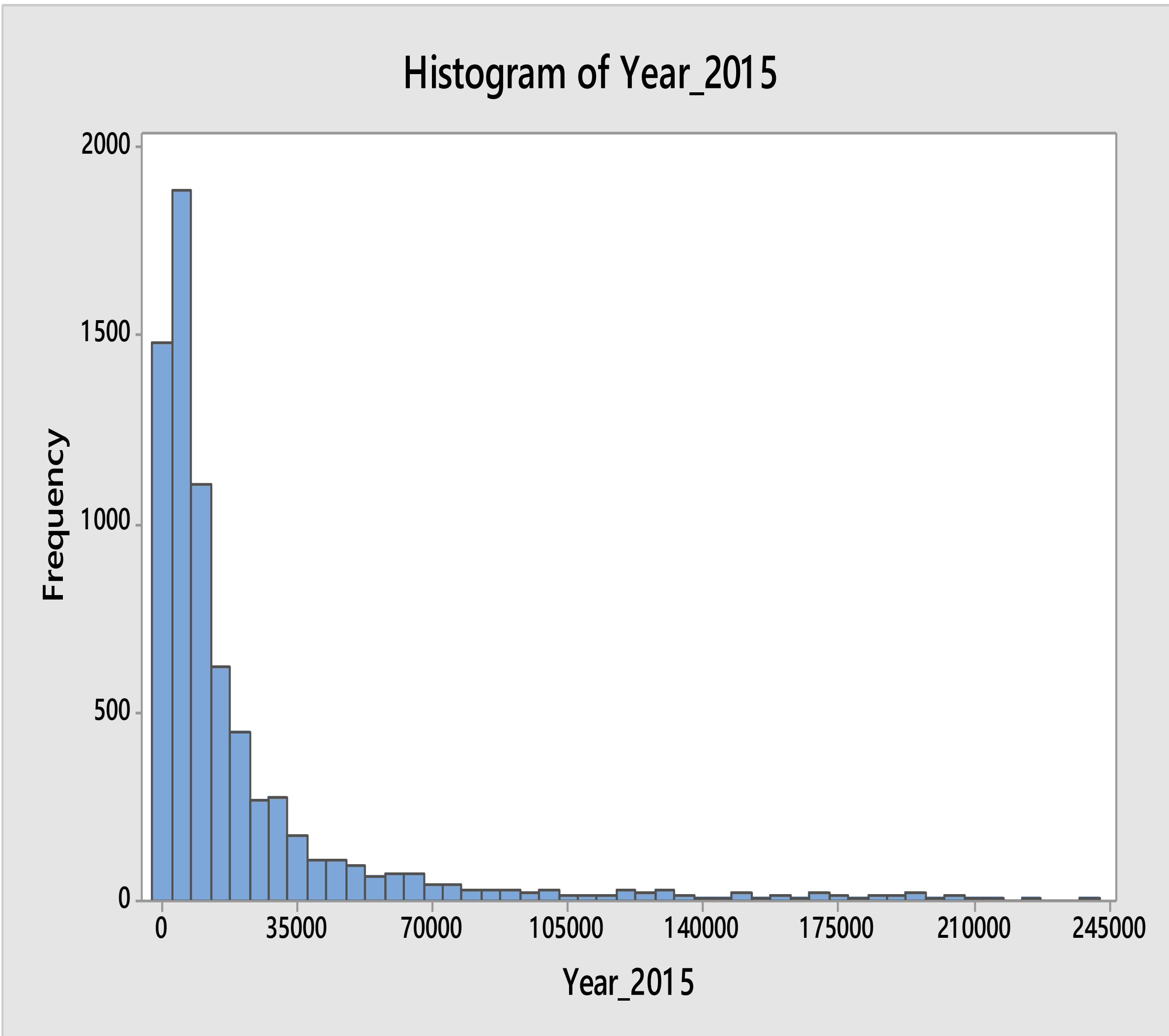

Figure 3.31. Histogram of the Year 2015 AADT

Figure 3.31 above indicated that for the year 2015 count dataset, data counts ranging between 0 and 35,000 occurred more in frequency than the rest of counts in the dataset. The higher the data observed increases, the lower the frequency becomes which indicated that the data is positive or right skewed.

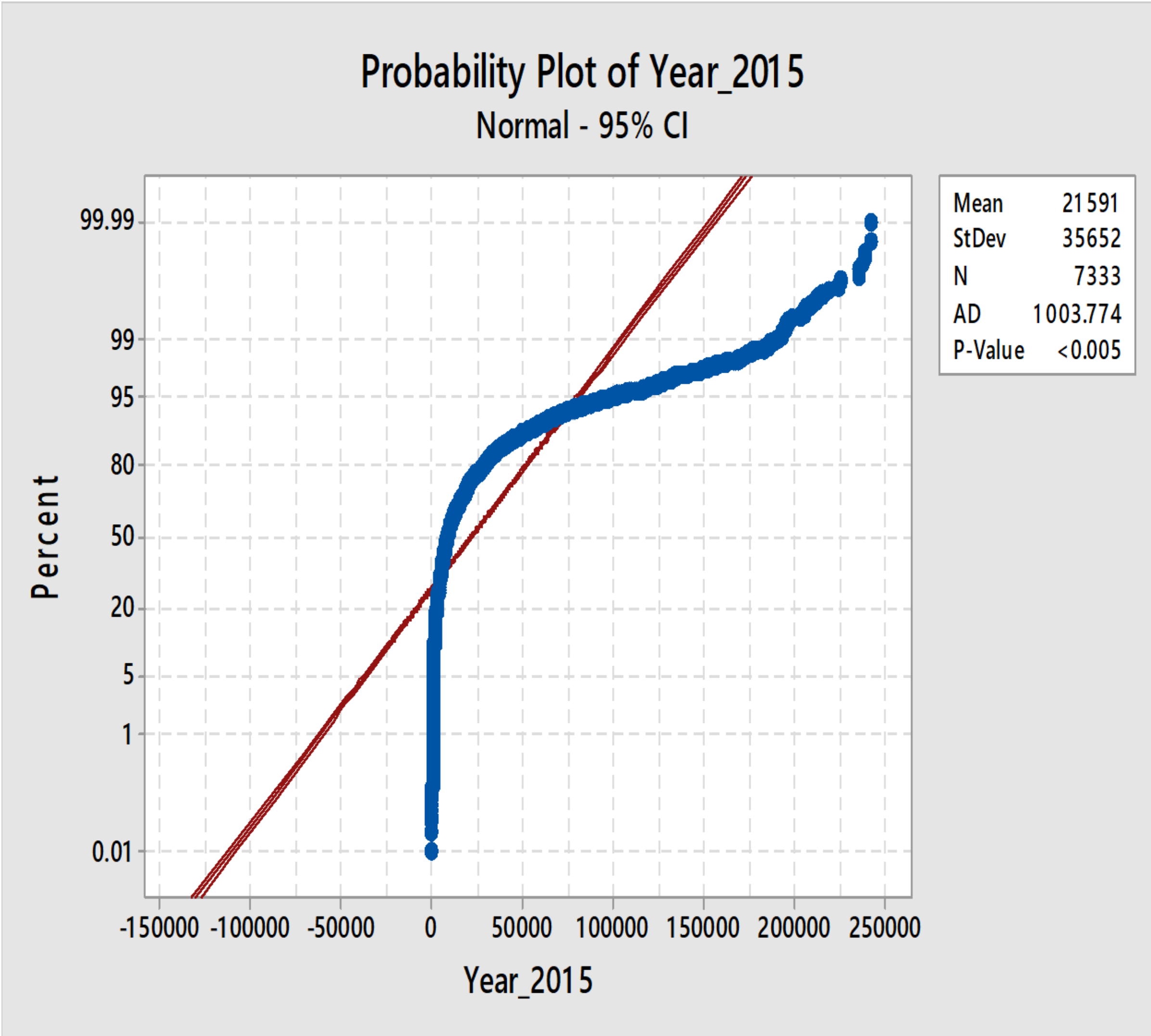

Figure 3.32. Normal Probability Plot for the Year 2015 AADT

Figure 3.32 above showed the distribution of the data count for the year 2015 count dataset. The normal probability plot showed a non-linear pattern indicating that the data does not follow a normal distribution pattern thereby we can reasonably conclude that the dataset is not normally distributed and normal probability plot does not provide an adequate fit for this dataset.

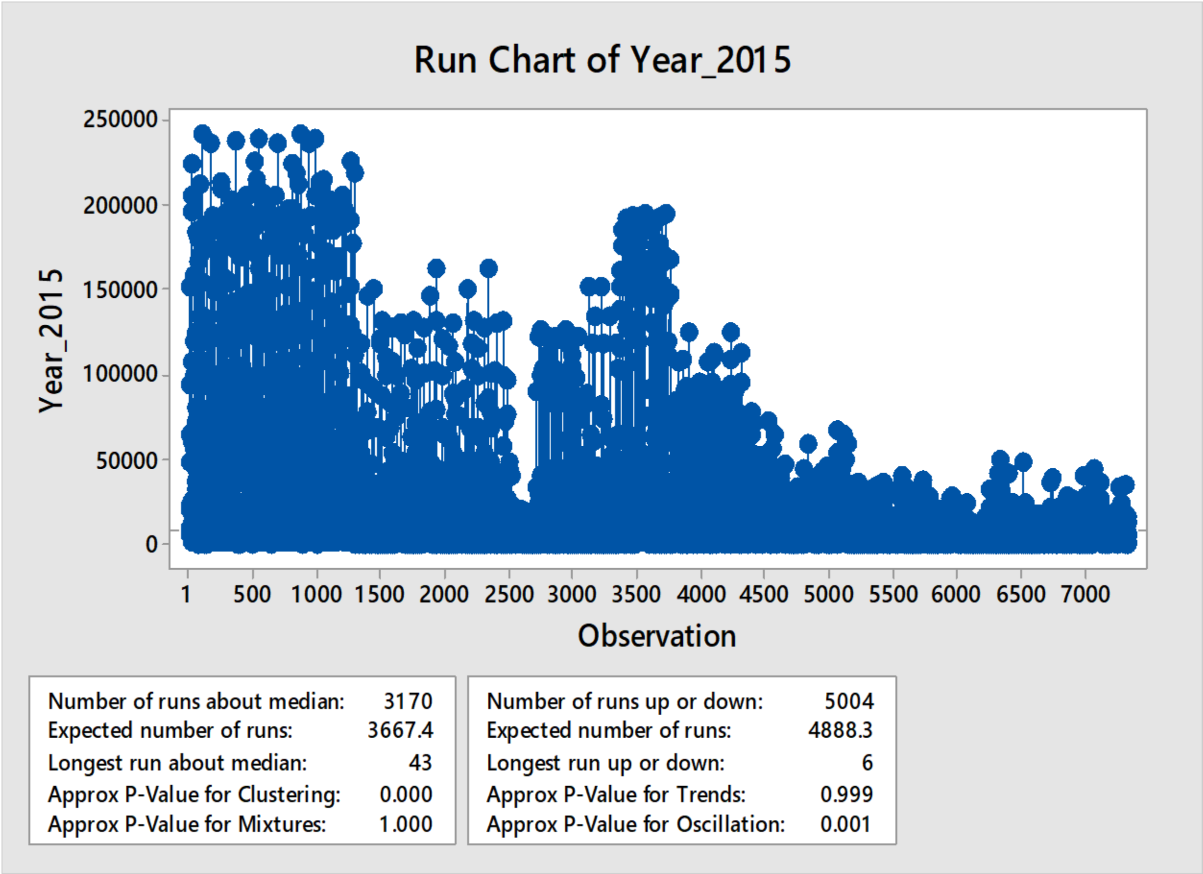

Figure 3.33. Run Sequence Plot for the Year 2015 AADT

The Run sequence plot from Figure 3.33 above indicated several significant shifts in different locations in the dataset. This indicated the randomness of the dataset for which the univariate model

$$Y_i = C + E_i \quad \text{(Eq. 3.14)}$$

is valid.

The camera locations for the year 2015 is shown in Figure 3.34 below using ArcGIS. The camera location points portrayed in altered color for each count division. Figure 3.34 reflects the trend with much of the locations on the Westbound side of the state. With the passage of every year, the graphic images exhibit more clusters of the camera points overlaying at various locations.

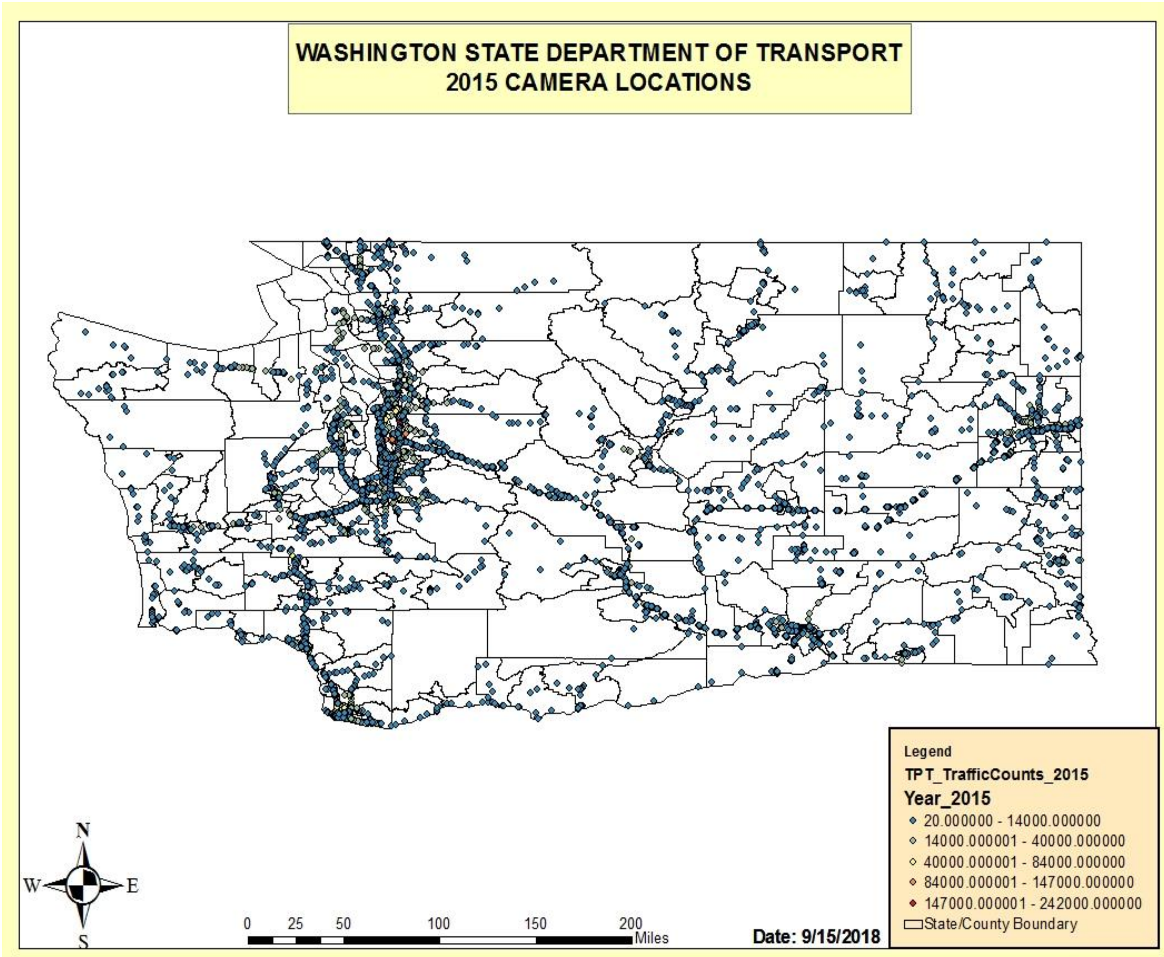

Figure 3.34. 2015 Camera Locations using ArcGIS

Table 3.14. The Year 2016 Dataset Analysis

Variable	N	Mode	Mean	SE Mean	St Dev	Min	Q1	Median	Q3	Max
Year_2016	7101	11000	22358	436	36717	20	3400	8600	23000	245000

Table 3.14 showed descriptive data analysis for the year 2016 AADT count dataset. A total number of 158,765,570 counts was recorded with a standard deviation of 35,652. The median for the dataset is 8,500 compared with a mean of 21,591. The maximum count in the dataset was 242,000 and the minimum was 20 with 11,000 as the mode. The first and third quartile for the data

is 3,300 and 22,000 respectively. With the median being 8,500 compared with a mean of 21,591 a plot of the data indicated the data to be right-skewed (Figure 3.35).

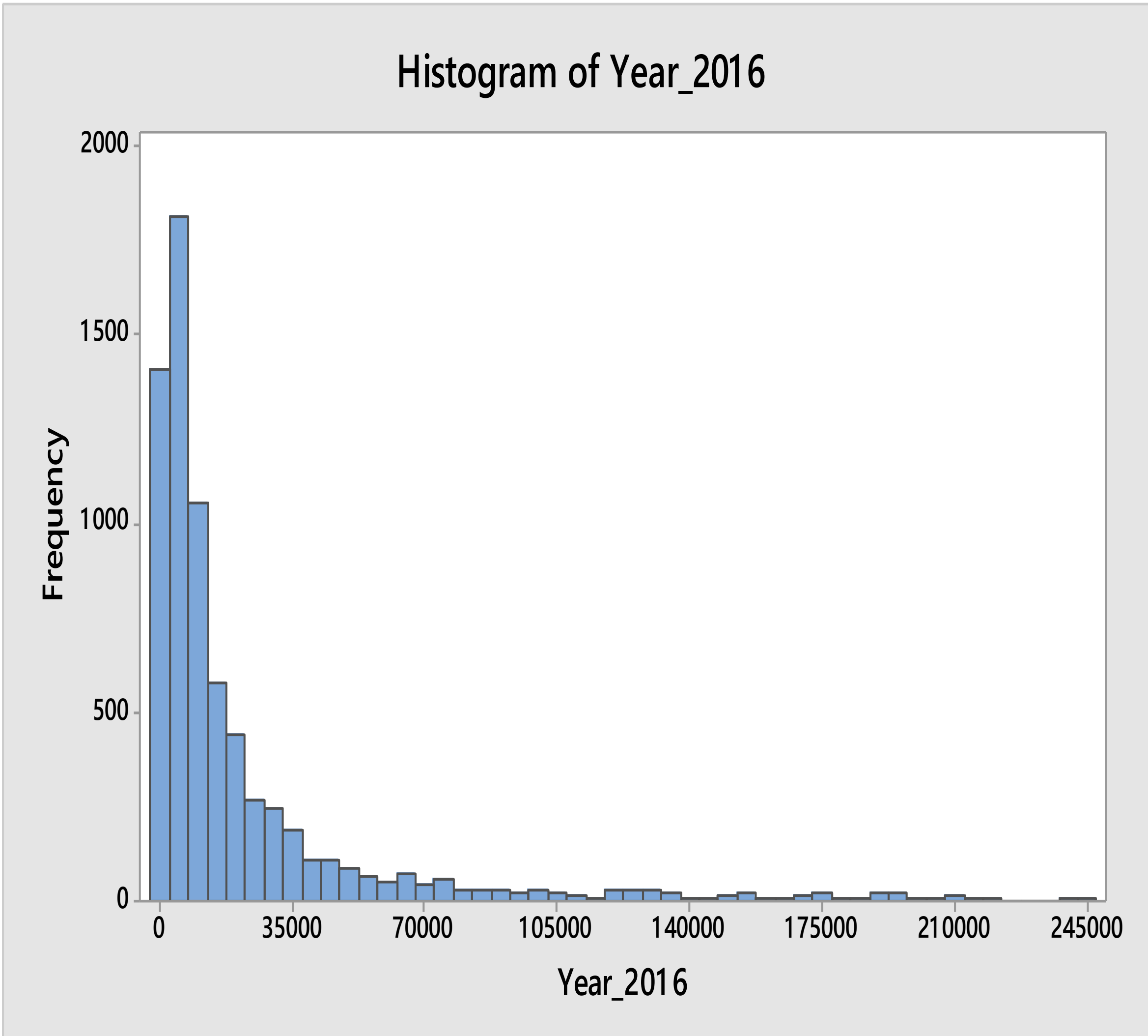

Figure 3.35. Histogram of the Year 2016 AADT

Figure 3.35 above indicated that for the year 2016 count dataset, data counts ranging between 0 and 35,000 occurred more in frequency than the rest of counts in the dataset. The higher the data observed increases, the lower the frequency becomes which indicated that the data is positive or right skewed.

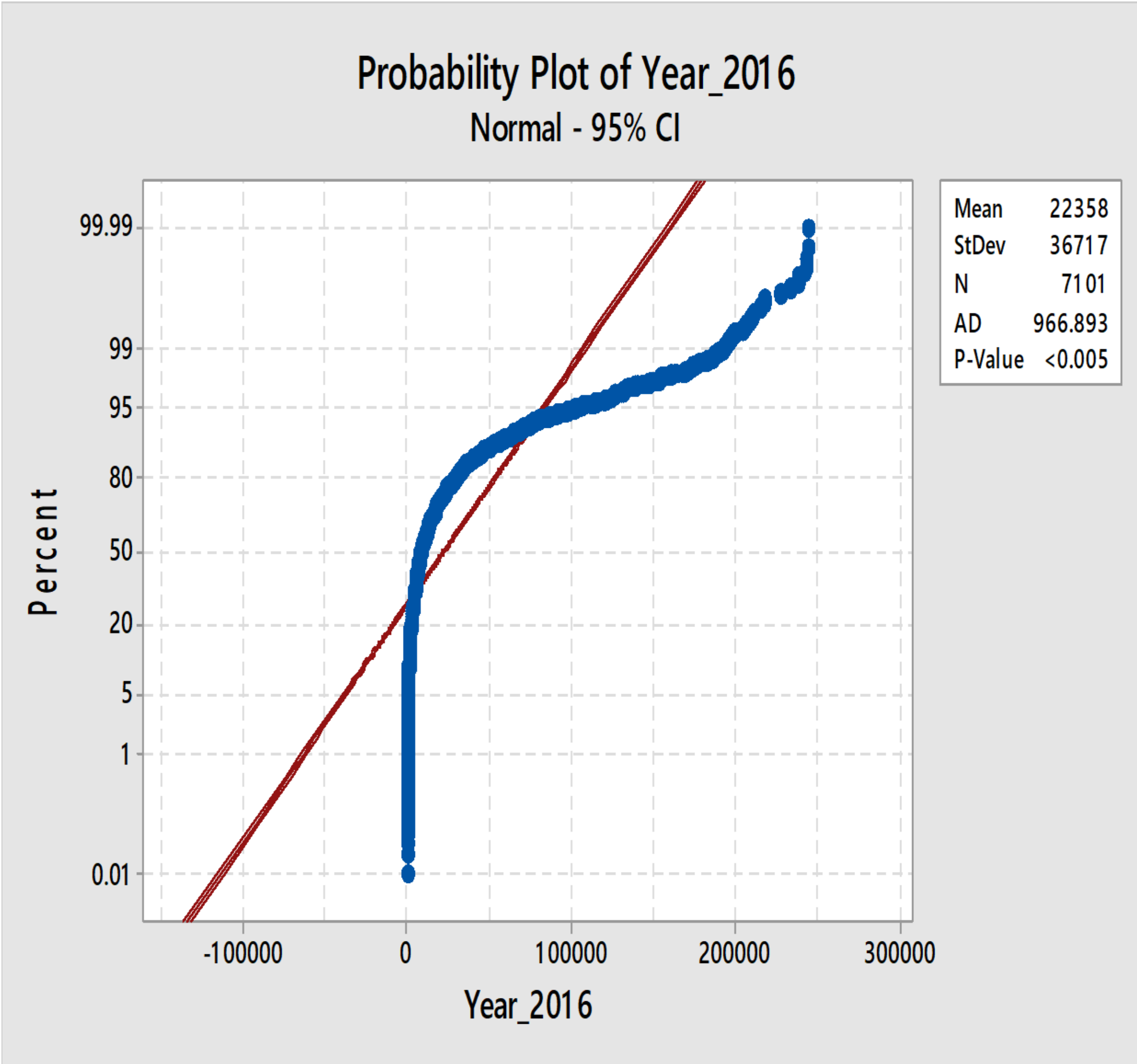

Figure 3.36. Normal Probability Plot for the Year 2016 AADT

Figure 3.36 above showed the distribution of the data count for the year 2016 count dataset. The normal probability plot showed a non-linear pattern indicating that the data does not follow a normal distribution pattern thereby we can reasonably conclude that the dataset is not normally distributed and normal probability plot does not provide an adequate fit for this dataset.

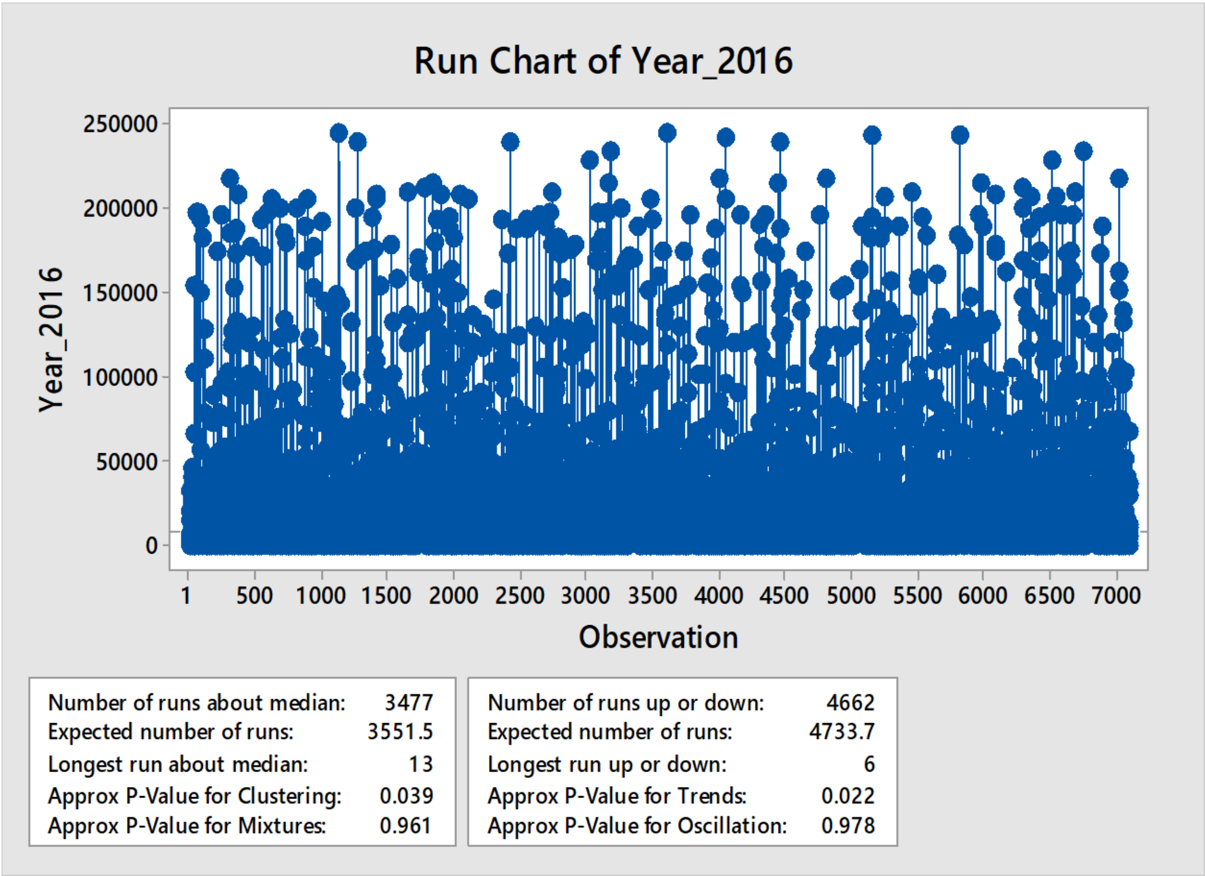

Figure 3.37. Run Sequence Plot for the Year 2016

The Run sequence plot from Figure 3.37 above indicated several significant shifts in different locations in the dataset. This indicated the randomness of the dataset for which the univariate model

$$Y_i = C + E_i \qquad \text{(Eq. 3.15)}$$

is valid.

The camera locations for the year 2016 is shown in Figure 3.38 below using ArcGIS. The camera location points portrayed in altered color for each count division. Figure 3.38 reflects the trend with much of the locations on the Westbound side of the state. With the passage of every year, the graphic images exhibit more clusters of the camera points overlaying at various locations.

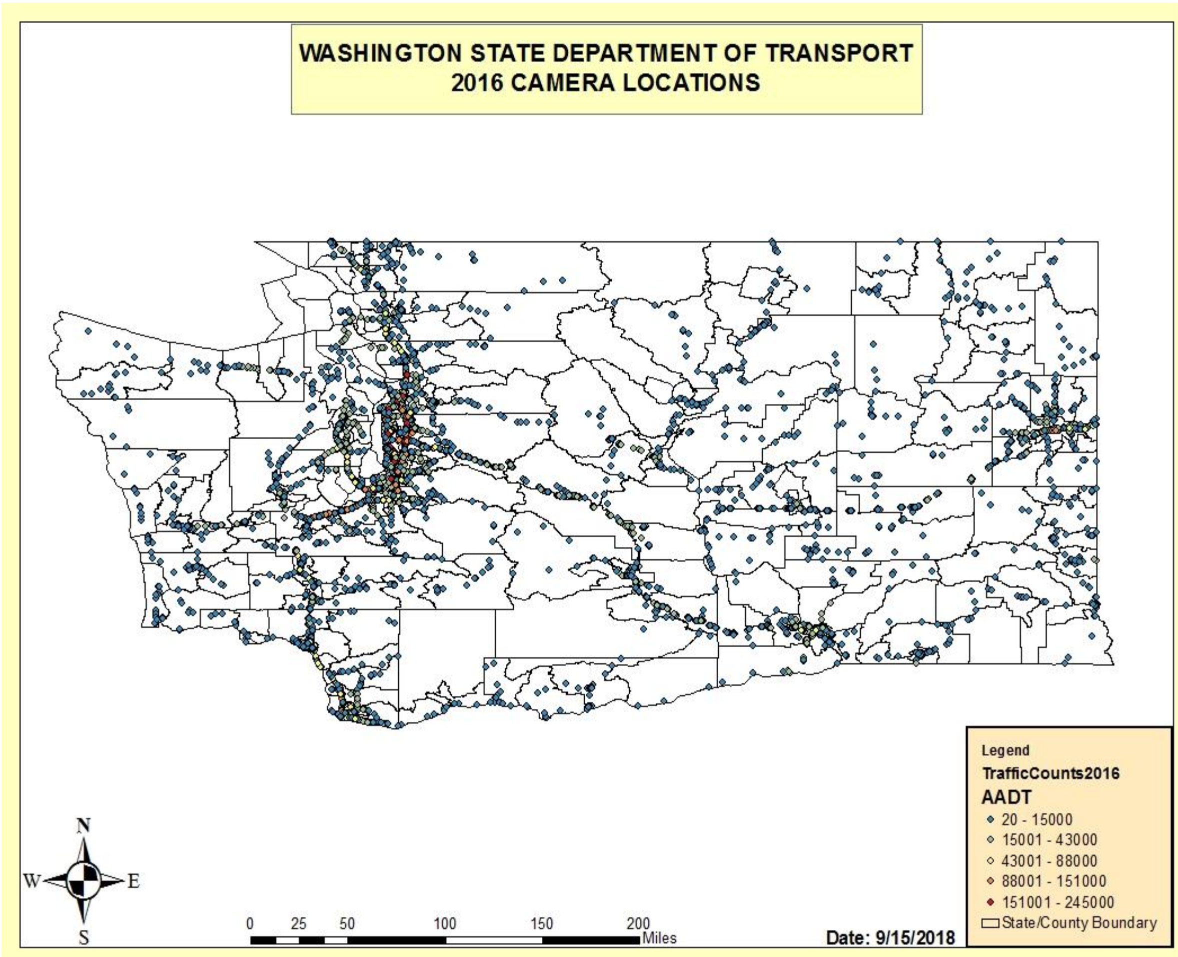

Figure 3.38. 2016 Camera Locations using ArcGIS

3.3.3. Trend Analysis Using Percentage Increment

The traffic recorder data collection sites used for this project has 5 directions. Using the year 2009 has the benchmark year, the percentage increment for these locations for each year as analyzed below. The year 2009 was considered as the benchmark for this analysis because the data downloaded only starts from 2009 and there will not be any available dataset to compare it to. Data from previous years (2008 and less) could not be used because it does not have the location properties that were needed to analyze it in ArcGIS software.

Table 3.15. The Year 2010 Percentage Increment

Locations	Bothway	North	South	East	West	Mean	Median	St. Dev
% Increase in AADT data from previous year (%)	1.03	6.89	9.79	-0.44	2.32	3.92	2.32	4.28

Exploratory data analysis for the locations using percentage increase in AADT from the prior year showed that the average of the percentage increment for the year 2010 is 3.92 with a standard deviation of 4.28. The median for the dataset is 2.32 compared with a mean of 3.92. The highest percentage of the dataset was 9.79 and the lowest is -0.44. With the median being 2.32 compared with a mean of 3.92 showed an indication of the data being right-skewed. Figure 3. below showed the clustered bar chart for this analysis.

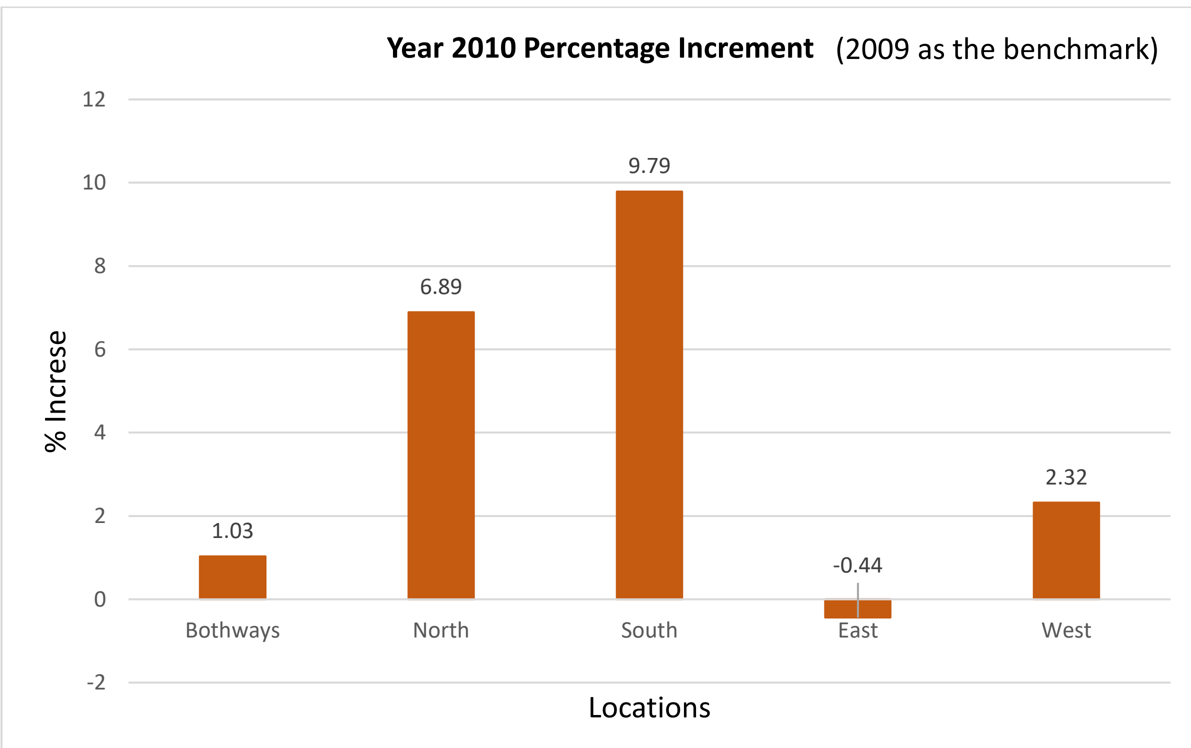

Figure 3.39. The Year 2010 Percentage Increment Clustered Bar Chart

Table 3.16. The Year 2011 Percentage Increment

Locations	Bothway	North	South	East	West	Mean	Median	St. Dev
% Increase in AADT data from previous year (%)	-2.32	2.83	-1.03	-7.8	-6.79	-3.02	-2.32	4.35

Exploratory data analysis for the locations using percentage increase in AADT from the prior year showed that the average of the percentage increment for the year 2011 is -3.02 with a standard deviation of 4.35. The median for the dataset is -2.32 compared with a mean of -3.02. The highest percentage of the dataset was 2.83 and the lowest is -7.80. With the median being -2.32 compared with a mean of –3.02 showed an indication of the data being left-skewed. Figure 3. below showed the clustered bar chart for this analysis.

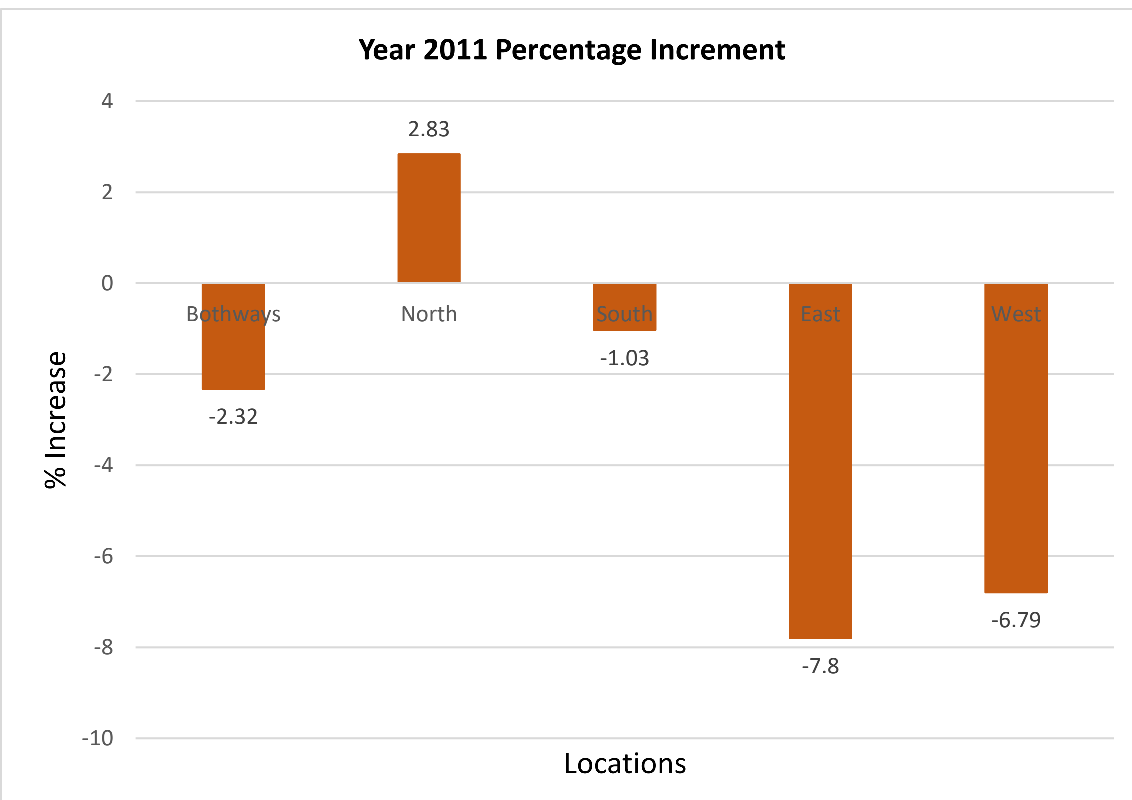

Figure 3.40. The Year 2011 Percentage Increment Clustered Bar Chart

Table 3.17. The Year 2012 Percentage Increment

Locations	Bothway	North	South	East	West	Mean	Median	St. Dev
% Increase in AADT data from previous year (%)	-0.52	2.61	1.63	7.86	7.04	3.72	2.61	3.60

Exploratory data analysis for the locations using percentage increase in AADT from the prior year showed that the average of the percentage increment for the year 2012 is 3.72 with a standard deviation of 3.60. The median for the dataset is 2.61 compared with a mean of 3.72. The highest percentage of the dataset was 7.86 and the lowest is -0.52. With the median being 2.61 compared with a mean of 3.72 showed an indication of the data being right-skewed. Figure 3.41 below showed the clustered bar chart for this analysis.

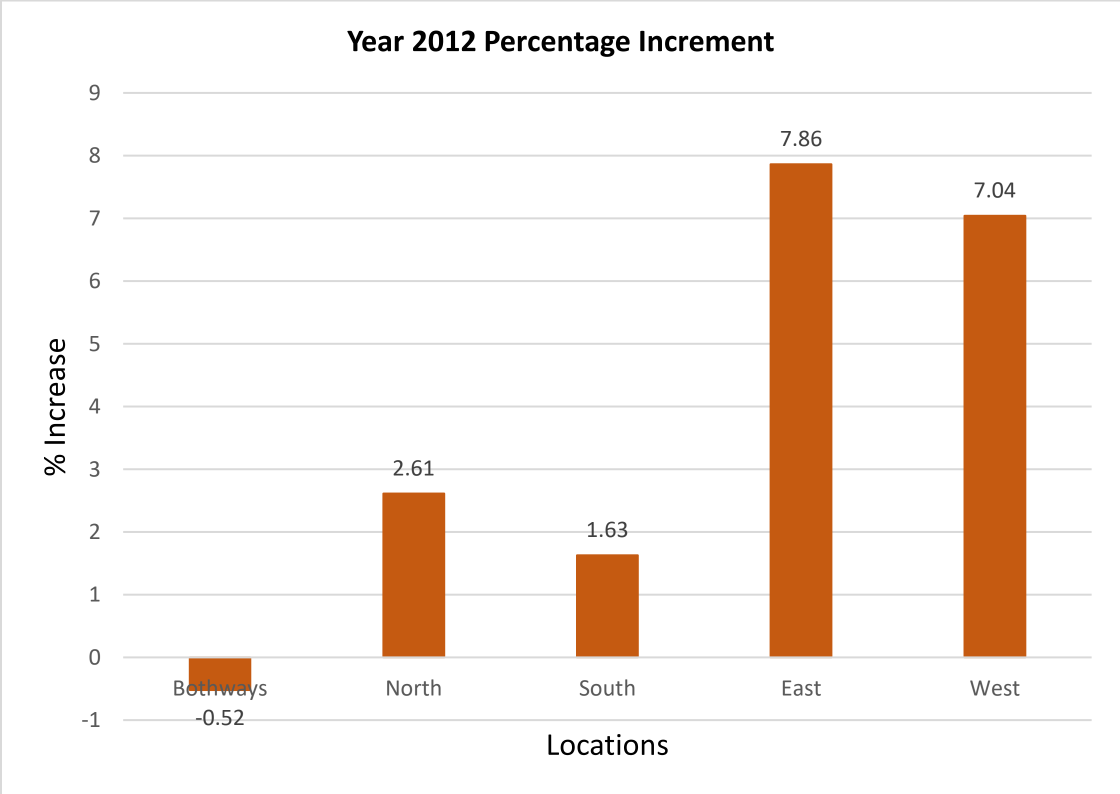

Figure 3.41. The Year 2012 Percentage Increment Clustered Bar Chart

Table 3.18. The Year 2013 Percentage Increment

Locations	Bothway	North	South	East	West	Mean	Median	St. Dev
% Increase in AADT data from previous year (%)	1.36	2.85	4.67	7.86	8.63	5.07	4.67	3.15

Exploratory data analysis for the locations using percentage increase in AADT from the prior year showed that the average of the percentage increment for the year 2013 is 5.07 with a standard deviation of 3.15. The median for the dataset is 4.67 compared with a mean of 5.07. The highest percentage of the dataset was 8.63 and the lowest is 1.36. With the median being 4.67 compared with a mean of 5.07 showed an indication of the data being right-skewed. Figure 3. below showed the clustered bar chart for this analysis.

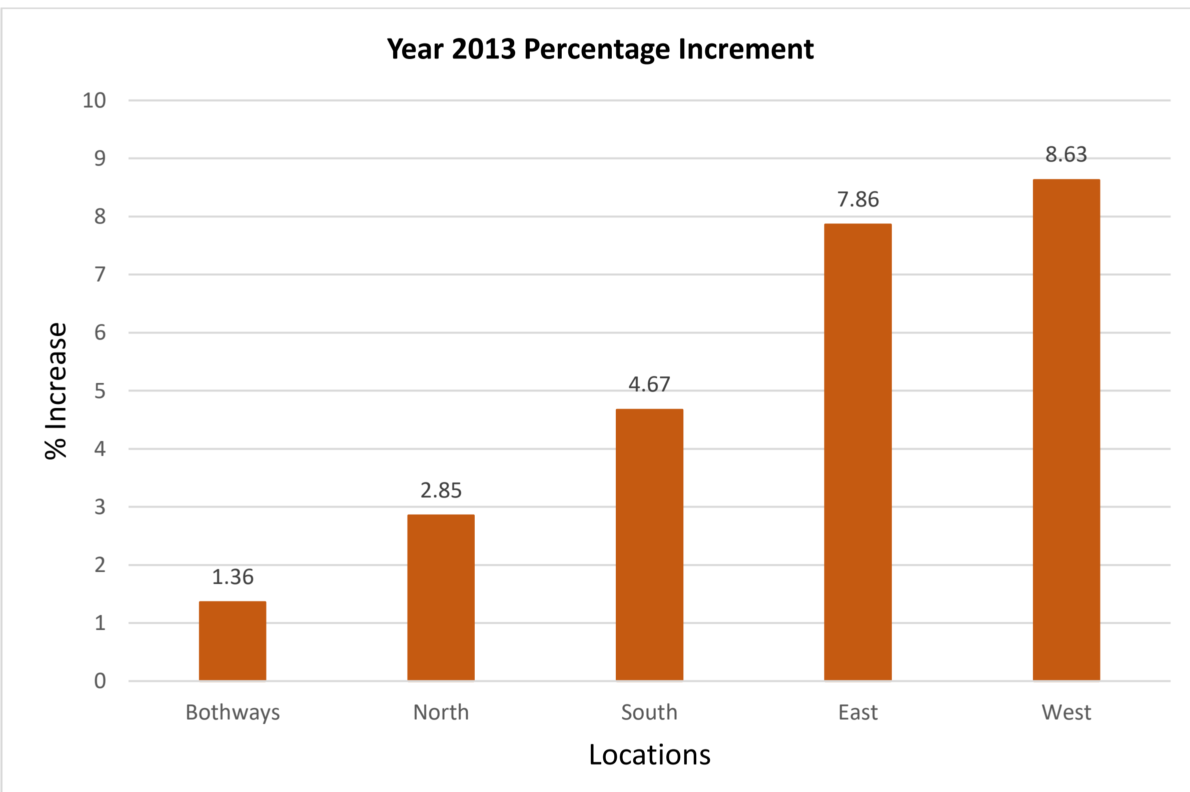

Figure 3.42. The Year 2013 Percentage Increment Clustered Bar Chart

Table 3.19. The Year 2014 percentage increment

Locations	Bothway	North	South	East	West	Mean	Median	St. Dev
% Increase in AADT data from previous year (%)	0.42	0.81	2.83	2.96	3.03	2.01	2.83	1.28

Exploratory data analysis for the locations using percentage increase in AADT from the prior year showed that the average of the percentage increment for the year 2014 is 2.01 with a standard deviation of 1.28. The median for the dataset is 2.83 compared with a mean of 2.01. The highest percentage of the dataset was 3.03 and the lowest is 0.42. With the median being 2.83 compared with a mean of 2.01 showed an indication of the data being left-skewed. Figure 3. below showed the clustered bar chart for this analysis.

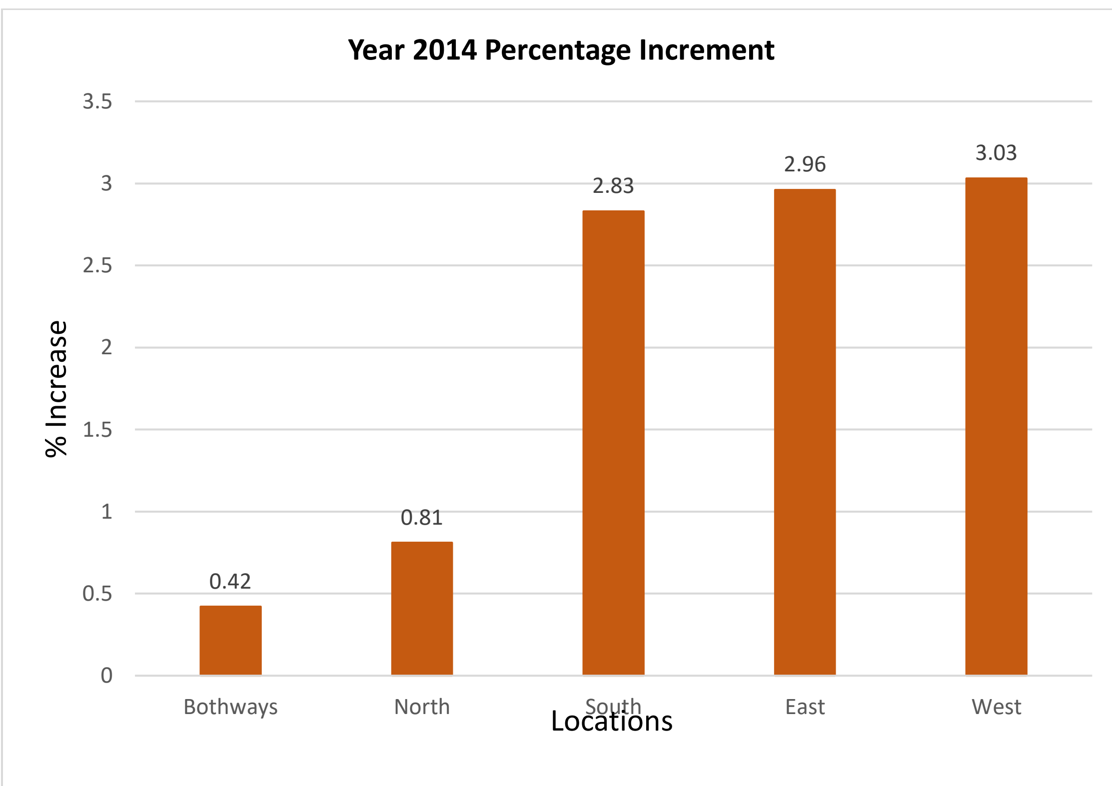

Figure 3.43. The Year 2014 Percentage Increment Clustered Bar Chart

Table 3.20. The Year 2015 Percentage Increment

Locations	Bothway	North	South	East	West	Mean	Median	St. Dev
% Increase in AADT data from previous year (%)	2.1	-7.76	-6.35	-7.46	-6.06	-5.11	-6.35	4.09

Exploratory data analysis for the locations using percentage increase in AADT from the prior year showed that the average of the percentage increment for the year 2015 is -5.11 with a standard deviation of 4.09. The median for the dataset is -6.35 compared with a mean of -5.11. The highest percentage of the dataset was 2.10 and the lowest is -7.76. With the median being -6.35 compared with a mean of -5.11 showed an indication of the data being right-skewed. Figure 3. below showed the clustered bar chart for this analysis.

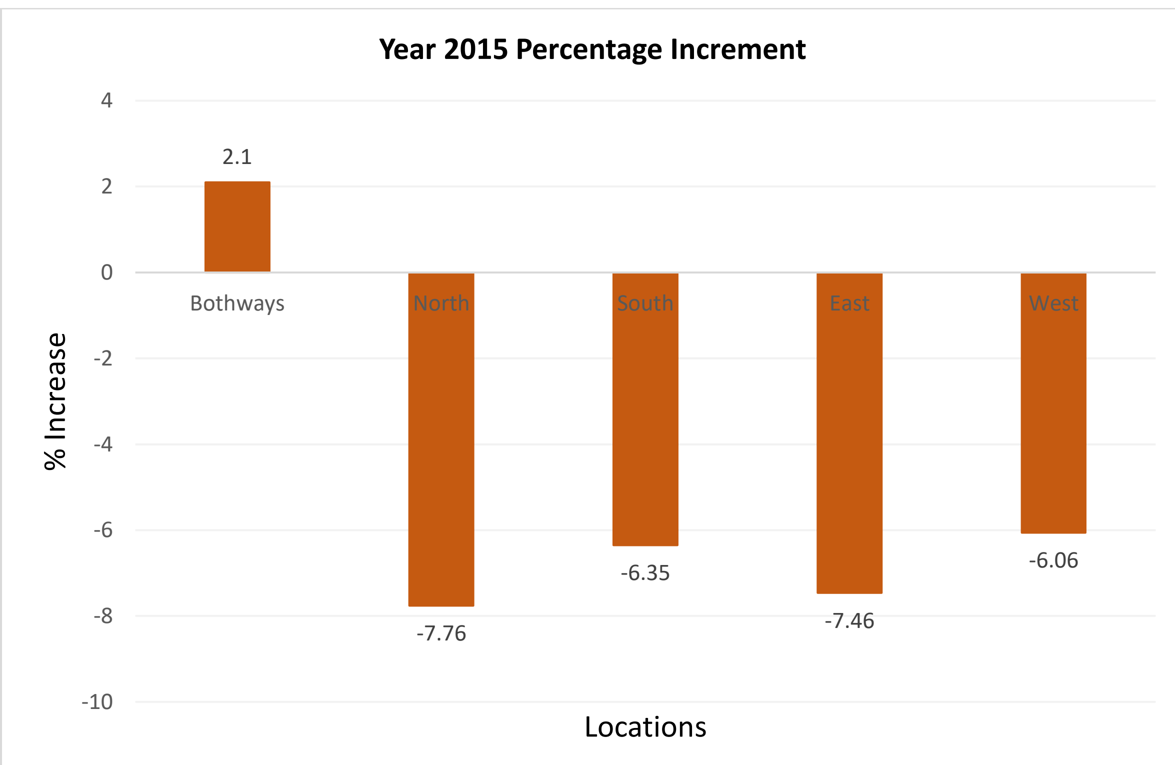

Figure 3.44. The Year 2015 Percentage Increment Clustered Bar Chart

Table 3.21. The Year 2016 Percentage Increment

Locations	Bothway	North	South	East	West	Mean	Median	St. Dev
% Increase in AADT data from previous year (%)	1.52	-14.22	-15.41	1.68	0.57	-5.17	0.57	8.82

Exploratory data analysis for the locations using percentage increase in AADT from the prior year showed that the average of the percentage increment for the year 2016 is -5.17 with a standard deviation of 8.82. The median for the dataset is 0.57 compared with a mean of -5.17. The highest percentage of the dataset was 1.68 and the lowest is -15.41. With the median being 0.57 compared with a mean of -5.17 showed an indication of the data being left-skewed. Figure 3. below showed the clustered bar chart for this analysis.

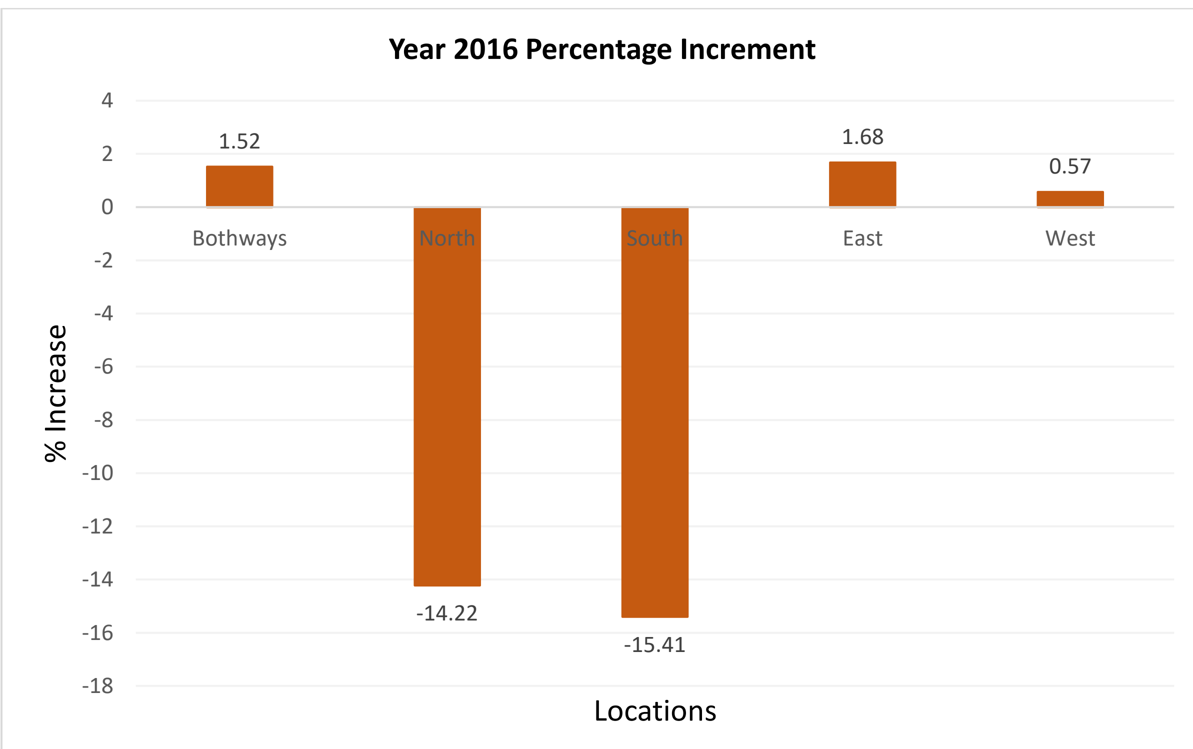

Figure 3.45. The Year 2015 Percentage Increment Clustered Bar Chart

3.4. Summary

This chapter discussed in detail the processes and results of the exploratory data analysis of WSDOT AADT data. Exploratory data analysis provided the vital information from the data and was used to answer part of questions 3 and 4 of the research questions. The result of the analysis showed that the data is positively trended. The direction bound analysis showed the bothways bound collection points has had the highest data count with an increase of 4,927,10 data counts from the year 2009 to the year 2016. The Northbound and Southbound direction data collection points saw a significant decrease of 405,550 and 317,590 respectively in the data count. Eastbound and Westbound direction data collection points have a considerable increase in 97,020 and 225,897 respectively.

The histogram indicated that the dataset is right-skewed. The normal probability plot shows a non-linear pattern thereby we can reasonably conclude that the data does not follow the normal distribution rule indicating its randomness. The run sequence plots indicated several significant shifts in different locations in the dataset, this also indicated the randomness of the dataset.

Furthermore, the descriptive analysis showed that the mode for the data for the year 2009 is 12,000 and 11,000 for the years 2010 to 2016. This indicated that the most common data count collected from the year 2010 to 2016 is about the same.

4. RESEARCH METHODOLOGY

4.1. Introduction

This chapter discussed the research methods, processes, tools, and tests used to address the research questions. The spatial analyst tool used was performed in the ArcGIS. The spatial analyst tool helped create statistical model from a specified point and then interpolates to develop an incessant surface through spatial estimation. This helps to understand the nature and behavioral patterns of the AADT data in a certain area of interest. A hypothesis was postulated and tested using one-way ANOVA to see if all the kriging methods have the same mean prediction error or not.

4.2. Research Approach

The overall framework used to answer questions 3 and 4 of the research questions and objectives is depicted in Figure 4.1. The research approach consists of data exploration, structural analysis, cross-validation, hypothesis testing and discussion of results.

Data exploration involves the examination of the data and statistical analysis. According to Engineering Statistics Handbook (ESH 2003), exploratory data analysis (EDA) is an approach/philosophy for data analysis that employs a variety of techniques (mostly graphical) to: maximize insight into a data set, uncover underlying structure, extract important variables, detect outliers and anomalies, test underlying assumptions, develop parsimonious models, and determine optimal factor settings. It is a philosophy as to how we dissect a data set; what we look for; how we look; and how we interpret (ESH 2012). EDA refers to the critical process of performing initial investigations on data so as to discover patterns, to spot anomalies, to test hypothesis and to check assumptions with the help of summary statistics and graphical representations (Prasad 2018). This aspect of the research was covered in Chapters 2 and 3 of this research.

The rest of the sections are organized as follows: Section 4.2.1 presents the structural analysis of the data. This involves selecting kriging methods and variogram fittings. Section 4.2.2 presents the crossvalidation process. This involves comparing the predicted values to the observed values. The hypothesis testing and the discussion of results come under Chapter 5 of this thesis.

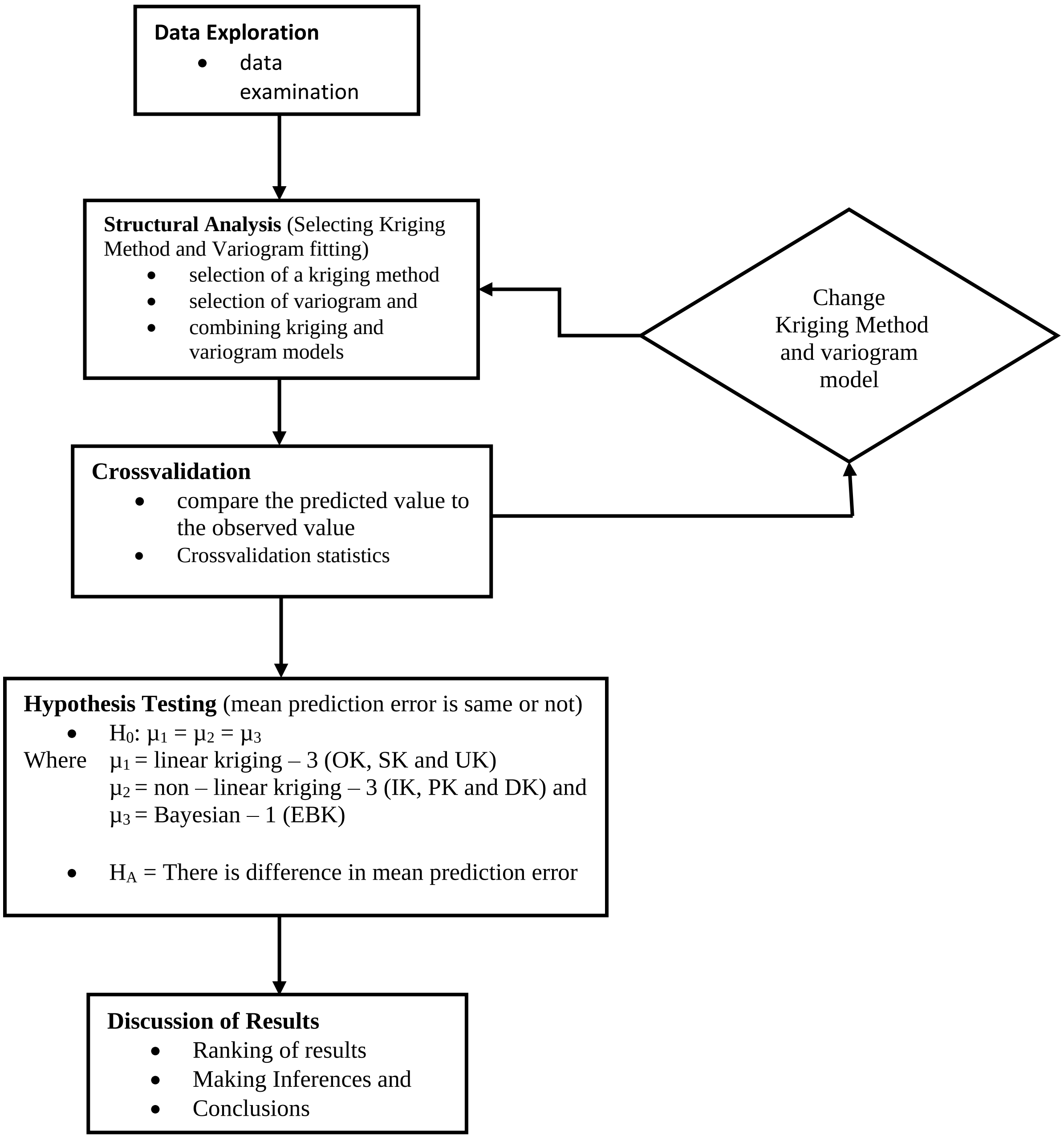

Figure 4.1. Research Method Process

4.2.1. Structural Analysis (Selecting Kriging Method and Variogram Fitting)

The structural analysis involves spatial interpolation process which uses the known values at given locations to estimate a continuous surface or data point (Annaka 2016). This process focuses on geographically raw data correlated by statistical methods (Negreiros et al. 2010). The structural analysis involves the selection of a kriging method and then combining it with a variogram at the same time. In this thesis, structural analysis was conducted by applying the geostatistical procedure of kriging to predict AADT values at unmeasured locations.

4.2.1.1. Kriging

According to Goovaerts (1997), kriging could be described as the best linear unbiased estimator (BLUE). It is unbiased because the mean error is 0. In kriging, a semivariogram is used to measure the dissimilarity between data points separated by the vector [h] (Goovaerts 1997). Since kriging is a probabilistic approach, it provides a set of possible values with corresponding probabilities of occurrence instead of a single estimated value for the unknown AADT. This stochastic approach reflects our imperfect knowledge of the unsampled value and its distribution (Goovaerts 1997).

Kriging is the generic name adopted by geostatisticians for a family of generalized least-square regression algorithms in recognition of the pioneering work of Danie Krige (Goovaerts 1997). Krige (1951, 1966) developed the kriging method empirically for estimating the quantity of gold in ore bodies in South Africa. According to Shamo et al (2015), kriging uses kriging weights (λ_α), which are derived from a covariance function (variogram). Spatial characterization of a dataset is dependent on fitting the right variogram to the model (Shamo et al. 2015).

One benefit of kriging is its ability to provide estimation errors. The prediction errors in part help in comparing kriging to other methods and it also serves as a basis for stochastic simulation of functions that could represent the relationship between the measured and unmeasured AADT data points (Shamo et al. 2015). This makes kriging the only method that uses spatial statistical theory to optimize interpolation (Clarke 1990).

- **Kriging Methods:** Three classifications of kriging methods are

i. linear kriging method: this comprises of universal, simple and ordinary kriging methods

ii. non-linear kriging methods: this include indicator kriging, probability kriging, and disjunctive kriging methods and

iii. bayesian kriging method: this include empirical bayesian kriging method.

The following kriging estimators were used for the analysis of the AADT data and the prediction error results compared to see which of these methods provided a more accurate result in predicting AADT data for an unknown location.

i. Linear Kriging Methods:

a. **Ordinary Kriging (OK):** Ordinary kriging estimator allows one to account for such local variation of the local mean by limiting the province of stationarity of the mean to the local neighborhood $Z(y\alpha)$ centered on the location y being estimated (Shamo et. al 2015). The assumption here is that the mean is unknown but fixed. Ordinary kriging assumes a linear model form and the equation is given as:

$$Zx(y) = \sum_{\alpha=1}^{n(y)} \lambda\alpha(y)Z(y\alpha) + \left[1 - \sum_{\alpha=1}^{n(y)} \lambda\alpha(y)\right]\mu(y) \qquad \text{(Eq. 4.1)}$$

where Z is continuous attribute (AADT); $Z(y)$ is the true value at unsampled location y; $Zx(y)$ is an estimate of value $Z(y)$; $Z(y\alpha)$ is Z datum value at location $y\alpha$; m is the stationary mean of the random function (RF) $Z(y)$; $\mu(y\alpha)$ is the expected value of random variable (RV) $Z(y)$; and $\lambda\alpha$ is the kriging weights.

The sill, range, and nugget obtained from the variogram used in combination with this estimator is then used to compute the kriging weight (λα) for which the sum is 1 (Shamo et. al 2015). The mean is obtained by requiring the kriging weights sum to 1

$$\sum_{\alpha=1}^{n(y)} \lambda\alpha(y) = 1 \qquad \text{(Eq. 4.2)}$$

Hence, the estimator in OK becomes (Shamo et. al 2015)

$$Zx(y) = \sum_{\alpha=1}^{n(y)} \lambda\alpha(y) Z(y\alpha) \qquad \text{(Eq. 4.3)}$$

b. **Simple kriging (SK):** Simple kriging estimator considers the mean μ(y) to be known and constant throughout the study range (Shamo et. al 2015). The simple kriging estimator also assumes a linear model form and is given by the equation:

$$Zx(y) = \sum_{\alpha=1}^{n(y)} \lambda\alpha(y)[Z(y\alpha) - \mu] + \mu \qquad \text{(Eq. 4.4)}$$

where $\lambda\alpha$ is weights associated with locations $y\alpha$, $Zx(y)$ is an estimate of value $Z(y)$; $Z(y\alpha)$ is Z datum value at location $y\alpha$ and μ is the unknown constant.

c. **Universal kriging (UK):** Universal kriging estimator is applied when the regionalized variable exhibits some form of the trend (Isaak's and Srivastava 1989). The mean varies, and it is unknown. It also assumes a linear model and the equation is given by:

$$Zx(y) = \sum_{\alpha=1}^{n(y)} \lambda\alpha(y) Z(y\alpha) \qquad \text{(Eq. 4.5)}$$

where $\lambda\alpha$ is weights associated with locations $y\alpha$, $Zx(y)$ is an estimate of value $Z(y)$; $Z(y\alpha)$ is Z datum value at location $y\alpha$

ii. Nonlinear Kriging Methods:

d. **Indicator Kriging (IK):** Indicator Kriging uses the model (ESRI, 2018):

$$I(\mathrm{s}) = \mu + \varepsilon(\mathrm{s}) \qquad \text{(Eq. 4.6)}$$

where μ is an unknown constant, $\varepsilon(s)$ is the error(s) and $I(s)$ is a binary variable. The creation of binary data may be with the use of a threshold for continuous data, or 0 or 1 for the observed or count data (ESRI, 2018). Using binary variables, indicator kriging proceeds the same way as ordinary kriging (ESRI, 2018). Probability means is used by indicator kriging to calculate the forecasted values of the unknown points.

e. **Probability kriging (PK):** According to ESRI (2018), probability kriging assumes the model:

$$I(\mathrm{s}) = I(\mathrm{Z(s)} > \mathrm{c_t}) = \mu_1 + \varepsilon_1(\mathrm{s}) \qquad \text{(Eq. 4.7)}$$

$$\mathrm{Z(s)} = \mu_2 + \varepsilon_2(\mathrm{s}) \qquad \text{(Eq. 4.8)}$$

where: μ_1, μ_2 equals unknown constants, $I(s)$ equals a binary variable created via threshold indicator, $I(Z(s) > c_t)$.

There are now two types of random errors, $\varepsilon_1(s)$ and $\varepsilon_2(s)$, so there is autocorrelation for each of them and cross-correlation between them (ESRI, 2018). Probability kriging strives to do the same thing as indicator kriging, but it uses cokriging in an attempt to do a better job. (ESRI, 2018).

f. **Disjunctive kriging (DK):** ESRI (2018) on their website showed disjunctive kriging to assume the model:

$$\mathrm{f(Z(s))} = \mu_1 + \varepsilon(\mathrm{s}) \qquad \text{(Eq. 4.9)}$$

where f(Z(s)) is a random function of Z(s) and µ1 is an unknown constant. DK requires the bivariate normality assumption and approximations to the functions $f_i(Z(s_i))$; these assumptions are difficult to verify, and the solutions are mathematically and computationally complicated (ESRI, 2018).

iii. Bayesian Kriging Method:

b. **Empirical Bayesian kriging (EBK):** EBK is a geostatistical interpolation method that programs the most difficult aspects of building a valid kriging model by automatically calculating parameters through a process of sub-setting and simulations. Other kriging methods in geostatistical analysis require the user to manually regulate parameters to receive accurate results, but EBK automatically calculates these parameters (ESRI, 2018). It accounts for the error introduced by estimating by taking into account the underlying semivariogram making it different from other kriging methods and thereby producing a better and more accurate result.

All kriging estimators are but variants of the basic linear regression estimator $Z^x(y)$ (Shamoet.al 2015). $Z^x(y)$ is defined as

$$Z^x(y) - \mu(y) = \sum_{\alpha=1}^{n(y)} \lambda_{\alpha}(y)[Z(y_{\alpha}) - \mu(y_{\alpha})] \qquad \text{(Eq. 4.10)}$$

where Z is continuous attria bute (AADT); $Z(y)$ is the true value at unsampled location y; $Z^{\,x}(y)$ is an estimate of value $Z(y)$; $Z(y_{\alpha})$ is Z datum value at location y_{α}; m is the stationary mean of the random function (RF) $Z(y)$; $\mu(y_{\alpha})$ is the expected value of random variable (RV) $Z(y)$; and $\lambda a_{\ \alpha}$ = kriging weights.

Shamo et al. 2015, identified the following assumptions that the estimators are modeled under:

1. the unknown sample data, $Z(y)$ and the n sample values, belong to the regionalized variables $Z(y)$ and $Z(y1)$, . . ., $Z(yn)$

2. for any two points $y1$ and $y2$ in the area over which $Z(y)$ is being estimated, the covariance Cov $[Z(y1), Z(y2)]$ of the associated regionalized variables $Z(y1)$ and $Z(y2)$ are known; and
3. the non-negative matrix of covariances between the measured variables (AADT) at the sample points is positive definite.

Other kriging methods calculate the semivariogram from known data locations and use this same single semivariogram to make predictions at unknown locations; this process implicitly assumes that the estimated semivariogram is the true semivariogram for the interpolation region. By not taking the uncertainty of semivariogram estimation into account, other kriging methods underestimate the standard errors of prediction (ESRI, 2018).

4.2.1.2. Variogram Modeling

The variogram is the simplest way to relate uncertainty to distance from an observation (Chiles and Delfiner 1999). The variogram is a fitted function used to express the relationship between the known and unknown data points (Shamo et al. 2015). The variogram approach to developing kriging weights is similar to inverse distance weighting except that in the case of kriging weights, the weights are modeled by the best-fitted variogram (Shamo et al. 2015).

The spherical and exponential variogram models were used in the AADT data analysis.

a) **Spherical Model:** The main characteristics of this model is a gradual decrease in the spatial autocorrelation (i.e. semivariance increment). This gradual decrease continues to a point or distance at which the autocorrelation is beyond zero (ESRI 2018). It is one of the most frequently used models. The equation of the spherical model with semivariance $y(h)$ and range a, as given by Longley et al (2001) and Shamo et al (2015) is

$$y(h) = \left(Sph\frac{h}{a}\right)\left(\begin{matrix} 1.5\frac{h}{a} - 0.5\left(\frac{h}{a}\right)^3 & if\ h\ \leq a \\ 1 & ,\ otherwise \end{matrix}\right) \quad \text{(Eq. 4.11)}$$

Where a is the range and h is the distance.

b) **Exponential Model:** When there is an exponential decrease in spatial autocorrelation with the distance increasing, then this model is more useful (ESRI, 2018). The autocorrelation tends to disappear completely only at an immeasurable distance. This model is also of frequent use in the industry. Previous knowledge of the process of spatial autocorrelation and the spatial autocorrelation of the data itself determines which model will be of use (ESRI, 2018). An **e**xponential model with a practical range a was defined by Shamo et al. (2015) as

$$y(h) = 1 - \exp\left(\frac{-3h}{a^2}\right) \quad \text{(Eq. 4.12)}$$

Where a is the range and h is the distance.

For the basic variograms, practically a sill is reached at a distance of the range (range of influence). In the model, the sill and range of each fitted variogram were determined. The nugget of the fitted variogram was obtained from the point where the variogram cuts the vertical axis (ESRI, 2018). A high nugget is an indication of the variogram modeling the relationship between known and unknown datasets with high variance (ESRI, 2018). A relatively high range value is an indication of the AADT dataset being representative (Shamo et al. 2015).

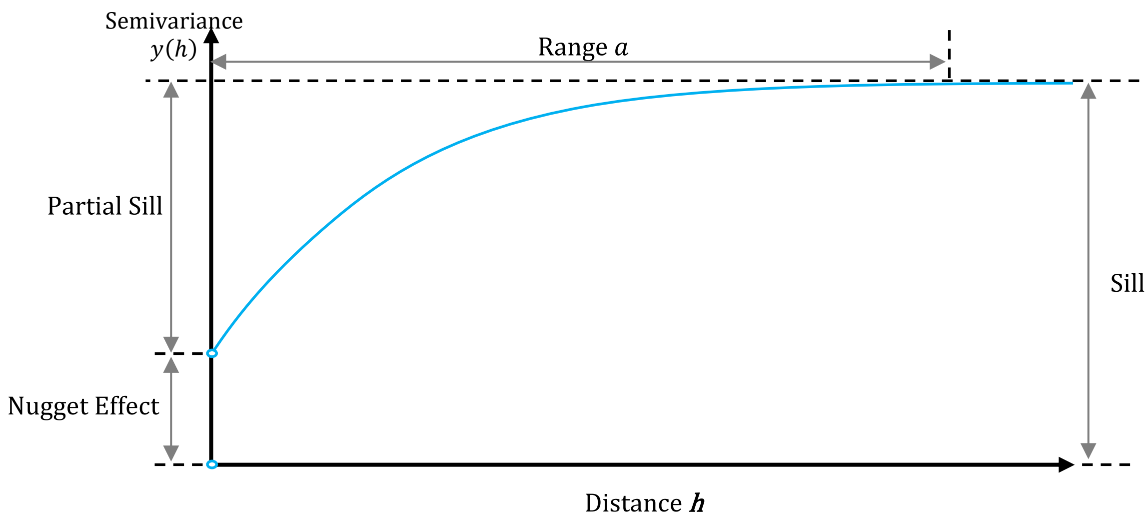

Figure 4.2. Characteristics of the Semi-Variogram Model

- **Sill:** The point or value at which the model touches the range is known as the Sill. (this is the value on the y-axis). Partial sill is gotten when you subtract the nugget effect from the sill (ESRI, 2018).
- **Range:** This is the distance that the model first starts to level or plane out (ESRI, 2018). Range is said to be spatially autocorrelated if points or locations are nearer to each other and vice-versa. (ESRI, 2018).
- **Nugget Effect:** This effect happens when there is either measurement error or spatial variation scale or in some circumstances both (ESRI, 2018). For example, if the semivariogram model crosses the y-axis at 1.5, then the nugget will be 1.5 as well. This calls for a proper knowledge of how the spatial variation scales operate (ESRI, 2018).

4.2.2. Crossvalidation

It's a model validation technique for assessing how the results of statistical analysis (model) will generalize to an independent data set (Georgios 2018). It is mainly used in settings where the goal is a prediction, and one wants to estimate how accurately a predictive model will perform in practice (Georgios 2018). The principle of cross-validation is to estimate $Z(y)$ at each sample point y_{α} from neighboring data $Z(y_{\beta})$, where $\beta \neq \alpha$ and $Z(y_{\alpha})$ is assumed to be unknown. By this, at every sample point y_{α}, a kriging estimate $Z(\alpha)$ and the associated kriging variance σ^2 are estimated. With the true value $Z_{\alpha} = Z(y_{\alpha})$ being known, the kriging error is $E_{\alpha} = Z^{*}(\alpha) - Z_{\alpha}$ and the standardized error is $e_{\alpha} = \frac{}{\sigma}$. If $y(h)$ is the theoretical variogram, E_{α} is a random variable with a mean of zero and a variance of σ^2 while e_{α} is a zero-mean unit variance random variable. The number of validation points is α and the variance at the location y where the AADT prediction is performed is σ^2 (Shamo et. al 2015).

Cross-validation is used to compare the performance of different interpolation models (Davis 1986; Journel 1987; Isaak and Srivastava 1989). Crossvalidation removes one data location then predicts the unknown data using the data at the rest of the locations (ESRI, 2018). In statistics, this step is synonymous to selecting a function of observation, a test statistic, and deriving its probability distribution under the assumed model (Shamo et. al 2015).

The cross-validation tool has the following properties and they are defined on the ESRI (2018) website:

- **Count:** This is the total number of samples used.
- **Mean Error (ME):** This is the averaged difference between the measured and the predicted values and it is shown below as

$$\mathrm{ME} = \frac{1}{\mathrm{n}}\sum_{i=1}^{n}\left\{\hat{\mathrm{Z}}(s_i) - Z(s_i)\right\} \qquad \text{(Eq. 4.13)}$$

where $\hat{\mathrm{Z}}(s_i)$ is the predicted value, $Z(s_i)$ is the observed (known) value and *n* is the number of values in the dataset

- **Root Mean Square Error (RMSE):** This indicates how closely the model predicts the measured values and it is shown below as

$$\mathrm{RMSE} = \sqrt{\frac{1}{\mathrm{n}}\sum_{i=1}^{n}\{\hat{\mathrm{Z}}(s_i) - Z(s_i)\}^2} \qquad \text{(Eq. 4.14)}$$

where $\hat{\mathrm{Z}}(s_i)$ is the predicted value, $Z(s_i)$ is the observed (known) value and *n* is the number of values in the dataset.

- **Average Standard Error (ASE):** This is the average of the prediction standard errors and it is shown below as

$$\text{ASE} = \sqrt{\frac{1}{\text{n}}\sum_{i=1}^{n}\sigma^2(\text{s}_i)} \qquad \text{(Eq. 4.15)}$$

where *n* is the number of values in the dataset and σ^2 is the kriging variance for location s_i.

- **Mean Standardized Error (MSE):** This is the average of the standardized errors. This value should be close to 0and it is shown below as:

$$\text{MSE} = \frac{1}{\text{n}}\sum_{i=1}^{n}\frac{\{\hat{\text{Z}}(\text{s}_i) - Z(\text{s}_i)\}}{\sigma(\text{s}_i)} \qquad \text{(Eq. 4.16)}$$

where $\hat{\text{Z}}(\text{s}_i)$ is the predicted value, $Z(\text{s}_i)$ is the observed (known) value, *n* is the number of values in the dataset and σ is the variance for location s_i.

The mean standardized error should ideally be zero if the interpolation method is unbiased. The calculated mean error, however, is a weak diagnostic for kriging because it is insensitive to inaccuracies in the variogram (Johnston et al., 2001; Webster and Oliver, 2001). The value of ME also depends on the scale of the data and is standardized by dividing by the kriging variance to form the MSE. An accurate model would have an MSE close to zero (ESRI, 2018).

MSE = 0 implies accurate variogram. The model performs good both in the training and the test set.

MSE > 0 implies overestimating. The model performs poorly both in the training and the test set.

MSE < 0 implies underestimating. The model performs poorly both in the training and the test set.

- **Root Mean Square Standardized Error (RMSSE):** This should be close to one if the prediction standard errors are valid. If the root-mean-squared standardized error is greater than one, it is underestimating the variability in the predictions. If the RMSSE is less than one, it is overestimating the variability in the predictions and it is shown below as:

$$\text{RMSSE} = \sqrt{\frac{1}{\text{n}}\sum_{i=1}^{n}\left[\frac{\{\hat{Z}(s_i) - Z(s_i)\}}{\sigma(s_i)}\right]2} \qquad \text{(Eq. 4.17)}$$

where $\hat{Z}(s_i)$ is the predicted value, $Z(s_i)$ is the observed (known) value, *n* is the number of values in the dataset and σ^2 is the kriging variance for location s_i.

If the model for the variogram is accurate, then the RMSE should equal the kriging variance, so the RMSE should equal 1. If the RMSE is greater than 1, then the variability in the predictions is being underestimated, and vice versa. Likewise, if the ASE is greater than the RMSE, then the variability is overestimated, and vice versa (Johnston et al., 2001; Webster and Oliver, 2001).

RMSE = 1 implies accurate variogram. The model performs good both in the training and the test set.

RMSE > 1 implies overestimating. The model performs poorly both in the training and the test set.

RMSE < 1 implies underestimating. The model performs poorly both in the training and the test set.

4.2.3. Operationalization of the Kriging Methods

The itemized steps below showed the step by step process of kriging carried out for the AADT data analysis. Through the process of iteration in steps 2 and 3 (Figure 4.1), a different kriging method is picked and a new variogram is modeled to fit the dataset. The corresponding cross-validated results were then obtained for each kriging method. The different types of variograms were varied for each kriging method and for each case, the best variogram was selected as the best fit for kriging variant on the AADT dataset.

The steps iterated below showed how kriging and semivariogram were performed using ArcGIS:

i. The first thing I did was to select kriging or co-kriging;

ii. The next I carried out was the data input. Here, the data input includes the source of the data and its field. This is useful for data modeling as well as data interpolation;

iii. The next step that ensued is the Semivariogram or covariance modeling this involved various parameters that is used produce the result. This step also produces the nugget and partial sill values. Figure 4.3 below gives an overview of the settings of this step. The semi-variogram was modeled using the exponential model and this gave the best suitable semivariogram for the data.

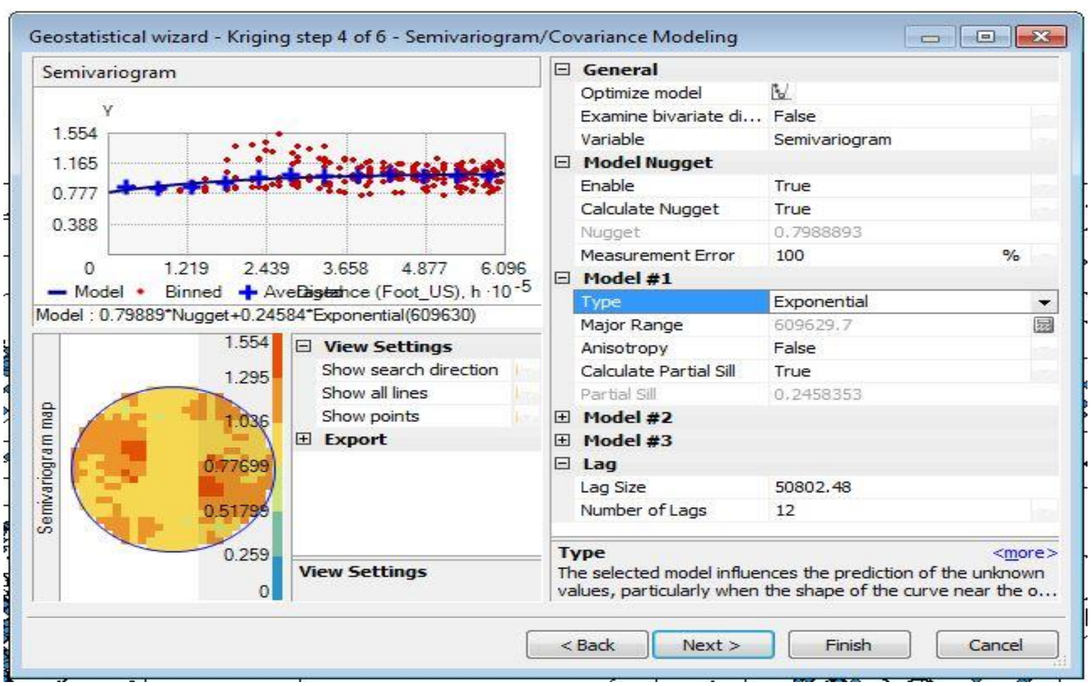

Figure 4.3. The Framework of Semi-Variogram/Covariance Modeling in Geostatistical Analysis

iv. Cross-validation is the last step of the whole process. Here, the tool detailed the mean prediction error statistical values and produced maps that showed areas with a high and low concentration of the dataset.

The steps listed above were then repeated for the rest of the kriging methods and different semivariograms, and the results produced were then compared using the mean prediction errors.

5. DISCUSSIONS OF RESULTS

5.1. Exploratory Spatial Data Analysis Result

The outputs of chapter 3 above (exploratory spatial data analysis) were the trend analysis, histogram, normality probability plot and run sequence plot. The result of the trend analysis of 2009, 2010, 2011, 2012, 2013, 2014, 2015 and 2016 AADT dataset showed that the values were positively trended. Table 5.1 is a summary of the mean, median, mode, and standard deviation of the counts of the years obtained from the analysis result.

Table 5.1. Overview of the Analysis Result

Years	Mean	Median	Mode	St. Dev.
2009	20456	8400	12000	33739
2010	20108	8200	11000	33556
2011	20023	8000	11000	33494
2012	19890	8000	11000	33187
2013	20022	8000	11000	33534
2014	20381	8100	11000	34071
2015	21591	8500	11000	35652
2016	22358	8600	11000	36717

It could be seen from the table above that there was a steady decrease in the mean of counts from the year 2009 to the year 2012. The years 2013 to 2016 saw a steady increase. This is the same for the standard deviation of the counts from the year 2009 through 2016. The median decreased from 2009 to the year 2011 and remained steady till the year 2013 after which it started increasing again to the year 2016. The mode decreased in the year 2010 and remained steady until the year 2016.

Trade, interstate travel, recreational travel, employment dynamics, and other trends could also account for this observation (Shamo et al. 2015). The high variation in AADT was accounted for at some of these locations relative to other locations by selecting a lag size that represents the largest distance amongst the data points. For instance, the lag of 1,318 used in modeling the 2009 AADT was selected after using the spatial analyst tool to conduct a neighborhood analysis on the data points (Shamo et al. 2015).

Another observation in the dataset from the neighborhood analysis was that the dataset was highly clustered. The process of kriging could help with this by assigning to individual data points within a cluster less weight than isolated data points. Invariably, clusters are treated as single points. This observation is consistent with the design of the HPMS coverage system, which allows for the design of new coverages in areas where the randomness in data values are not sufficient.

Histogram plots for the years (2009 - 2016) are lognormally skewed. The implication of the skewness was that the standard deviation of the dataset was directly proportional to the mean AADT. The distribution of the 2009, 2010, 2011, 2012, 2013, 2014, 2015 and 2016 datasets was skewed to the right (i.e., the mean AADT for each year was higher than the median). The covariance is extremely sensitive to outliers (Goovaerts 1997); thereby, reducing the effect of the outliers was prerequisite to achieving good results.

5.2. Kriging Analysis Result

The results of the kriging analysis methods are presented in this section. Comparison and inferences are then made to determine which of the methods has the least prediction error. The two crossvalidation tools used for the comparison and inference is Mean Standardization Error (MSE) and Root Mean Square Standardization Error (RMSSE).

For mean standardization error, the expected target value for the parameters is zero and for root mean square standardization error, the expected target value for the parameters is one. If the values of the parameters are close to the target value, then the model is said to be a good fit and thus acceptable but if the values of the parameters are not close to the target value, then the model is not a good fit and thus unacceptable. This implied the following: if

MSE = 0 implies accurate variogram. The model is good both in the training and the test set.

MSE > 0 implies overestimating. The model performs poorly both in the training and the test set.

MSE < 0 implies underestimating. The model performs poorly both in the training and the test set.

RMSE = 1 implies accurate variogram. The model is good both in the training and the test set.

RMSE > 1 implies overestimating. The model performs poorly both in the training and the test set.

RMSE < 1 implies underestimating. The model performs poorly both in the training and the test set.

Table 5.2. Mean Standardization Error Comparison for Each Year with The Spherical Model

			Mean Standardization Error				
			Target Value = 0				
Year	Model Type	OK	SK	UK	IK	PK	DK
2009	Spherical	-0.00454	0.060108	642.1786613	0.004737	0.001501	-0.00391
2010	Spherical	-0.09962	0.070763	407.7731339	0.014347	0.009071	0.006719
2011	Spherical	-0.13825	0.070536	373.24492	0.011732	0.004674	0.001568
2012	Spherical	-0.12774	0.056056	350.3135346	0.012842	0.009034	0.001325
2013	Spherical	-0.09564	0.053188	368.2149195	0.011705	0.007902	0.001187
2014	Spherical	-0.13446	0.062147	326.0557414	0.012763	0.010853	-0.0023
2015	Spherical	-0.08945	0.066944	459.6733769	0.011618	0.00872	0.004813
2016	Spherical	-0.06974	0.076026	660.4848762	0.012719	0.009641	0.008423

Table 5.2 above details the result of crossvalidation of six of the kriging methods combined with the spherical variogram model. The MSE is the result of the crossvalidation process that was explained in section 4.2.2 of this thesis and the target value is zero. EBK result is not available for

the spherical model because the spherical model does not provide the best visual fit to the empirical semivariances that are used by EBK.

In 2009 and 2014, PK and IK performed best with the spherical variogram. However, in 2010, 2011, 2012, 2013, 2015 and 2016, DK and PK performed best with the spherical variogram. UK has a much higher value compared to the target value thereby not useful in predicting AADT data with the spherical variogram.

Table 5.3. Mean Standardization Error Comparison for Each Year with The Exponential Model

Year	Model Type	OK	SK	UK	IK	PK	DK	EBK
2009	Exponential	-0.00856	0.054546	607.7919537	0.004449	0.001928	0.000374	-0.02178
2010	Exponential	-0.11578	0.071099	427.2266964	0.010287	0.000437	0.012796	-0.04701
2011	Exponential	-0.15611	0.069292	396.2223987	0.00714	0.004467	0.007499	-0.03166
2012	Exponential	-0.14716	0.056249	370.8658539	0.008107	0.002365	0.008112	-0.03669
2013	Exponential	-0.11059	0.050431	385.4671188	0.007604	0.02273	0.007252	-0.04517
2014	Exponential	-0.1609	0.06027	341.1584731	0.009048	0.003544	0.002588	-0.04042
2015	Exponential	-0.10871	0.061838	480.5954159	0.007325	0.004324	0.009874	-0.03995
2016	Exponential	-0.0466	0.071879	584.5150702	0.008478	0.003536	0.012575	-0.06084

Table 5.3 above detailed the result of crossvalidation of the seven kriging methods combined with the exponential variogram model. The MSE is the result of the crossvalidation process that was explained in section 4.2.2 of this thesis and the expected target value was zero. EBK result is available for the exponential model because the exponential model provides a good visual fit to the empirical semivariances that are used by EBK.

In 2009, 2011, 2012, 2014 and 2015, IK, PK, and DK performed best with the exponential variogram. However, in 2010, PK performed best with exponential variogram while 2013, IK and

DK performed best with the variogram. IK and PK performed best with the exponential variogram in the year 2016. UK has a much higher value compared to the target value thereby not useful in predicting AADT data with the exponential variogram.

Tables 5.2 and 5.3 above showed the result of MSE analysis using spherical and exponential variogram models. EBK result is not available for the spherical model because the spherical model does not provide the best visual fit to the empirical semivariances that are used by EBK but the reverse is the case for exponential variogram model has it provides a good visual fit to the empirical semivariances used by EBK. It could be also be inferred that for the MSE using spherical and empirical variogram model, PK was the best result (close to the value of the target than the rest of the other kriging methods).

Table 5.4. RMS Standardization Error Comparison for Each Year with The Spherical Model

			RMS Standardization Error				
			Target Value = 1				
Year	Model Type	OK	SK	UK	IK	PK	DK
2009	Spherical	0.770336	0.852929	7958.774023	0.92883	0.921033	0.985541
2010	Spherical	2.328016	0.830771	4901.444753	0.96163	0.970394	0.972508
2011	Spherical	2.810405	0.826858	4541.525599	0.945223	0.958353	0.98232
2012	Spherical	2.509416	0.86472	4874.466448	0.953585	0.942296	0.999779
2013	Spherical	2.095222	0.875256	5055.033503	0.942304	0.912715	1.004857
2014	Spherical	2.509447	0.842022	4690.140775	0.928743	0.936509	0.991524
2015	Spherical	1.903676	0.819352	5152.625629	0.929471	0.914442	0.959248
2016	Spherical	2.42213	0.791428	7433.361886	0.944403	0.942348	0.934756

Table 5.4 above details the result of crossvalidation of six of the kriging methods combined with the spherical variogram model. The RMSSE is the result of the crossvalidation process that was explained in section 4.2.2 of this thesis and the target value is one. EBK result is not available

for the spherical model because the spherical model does not provide the best visual fit to the empirical semivariances that are used by EBK.

In 2009, 2013, 2014, 2015 and 2016, DK performed best with the spherical variogram. However, in 2010 and 2011, DK, PK, and IK performed best with the spherical variogram while in 2012, DK and IK performed best with the spherical variogram. UK has a much higher value compared to the target value thereby not useful in predicting AADT data with the spherical variogram.

Table 5.5. RMS Standardization Error Comparison for Each Year with The Exponential Model

Year	Model Type	OK	SK	UK	IK	PK	DK	EBK
2009	Exponential	0.787557	0.876301	8117.896995	0.928706	0.931336	0.986774	1.000634
2010	Exponential	2.757813	0.8347	4984.325289	0.947488	0.943208	0.959106	1.075819
2011	Exponential	3.34005	0.83692	4606.534037	0.930796	0.914211	0.976196	1.050659
2012	Exponential	3.049334	0.868773	4933.969393	0.937222	0.953616	0.98479	1.063563
2013	Exponential	2.438474	0.888659	5056.237844	0.927001	1.047878	0.997855	1.076976
2014	Exponential	3.114778	0.85313	4665.230689	0.917721	0.913586	0.991598	1.065393
2015	Exponential	2.38087	0.838719	5231.782127	0.914719	0.902662	0.957626	1.05913
2016	Exponential	1.368427	0.80937	7806.24998	0.930543	0.931682	0.938483	1.097949

Table 5.5 above details the result of crossvalidation of the seven kriging methods combined with the exponential variogram model. The RMSSE is the result of the process that was explained in section 4.2.2 of this thesis and the target value is one. EBK result is available for the exponential model because the exponential model provides a good visual fit to the empirical semivariances that are used by EBK.

From 2009 to 2016, it could be seen that DK and EBK performed best with the exponential variogram. However, PK and IK also seem to perform better with exponential variogram for the years. UK has a much higher value compared to the target value thereby not useful in predicting AADT data with the exponential variogram.

Tables 5.4 and 5.5 above showed the result of RMSSE analysis using spherical and exponential variogram models. It indicated that EBK result is not available for the spherical model because the spherical model does not provide the best visual fit to the empirical semivariances that are used by EBK, but the reverse is the case for exponential variogram model has it provides a good visual fit to the empirical semivariances used by EBK. It could be also be inferred that for the RMSSE using spherical variogram model and, DK came out with the best result close to the value of the target than the rest of the other kriging methods while for RMSSE using empirical variogram model, EBK is a better method to use than the rest of the other methods.

In general, the result of this crossvalidation showed the same kriging methods with variograms cannot work for all the years which confirms the randomness characteristics of that data as seen in the exploratory data analysis section of this thesis. This also confirms that the process of kriging analysis will have to be performed on a year to year basis in order to discover which kriging method will best fit the data.

5.3. Probability Map Result

Probability maps represent the output variance of prediction raster which contains the kriging variance at each output raster cell (ESRI, 2018). Probability maps are used to define areas with the high and low certainty of exceeding a threshold value (Konstantin 2001). The probability maps were produced using the steps that were iterated in section 4.2.3 of this thesis.

The following are the probability maps produced after performing the different kriging interpretations for the Year 2009 AADT shown in Figure 5.1 to

Figure **5.13** below. The maps showed estimated or predicted data counts collected for different raster cells with range values between 3 and 239,000. Figure 5.1 to

Figure **5.13** show the probability of AADT in Washington State that interpolated from the sample database.

This might be useful in order to see what value or range of count is estimated for a particular area of interest. The probability limits in each of the kriging maps differ, but the appearance SK, OK, UK, DK, and EBK are alike in counts division and color variation while IK and PK are alike in counts division and color variation.

The legend below applied to all the maps in terms of the range and color variation

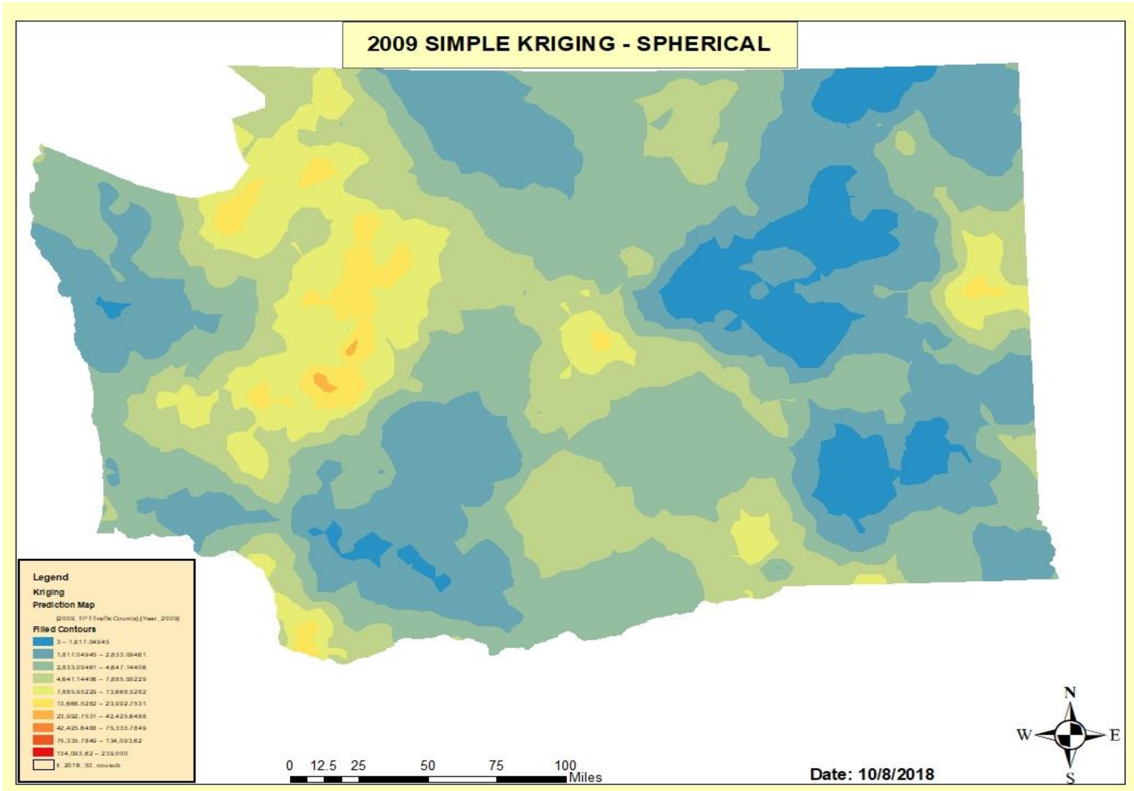

Figure 5.1. SK (Spherical) Map of Washington State AADT Data for the Year 2009 AADT

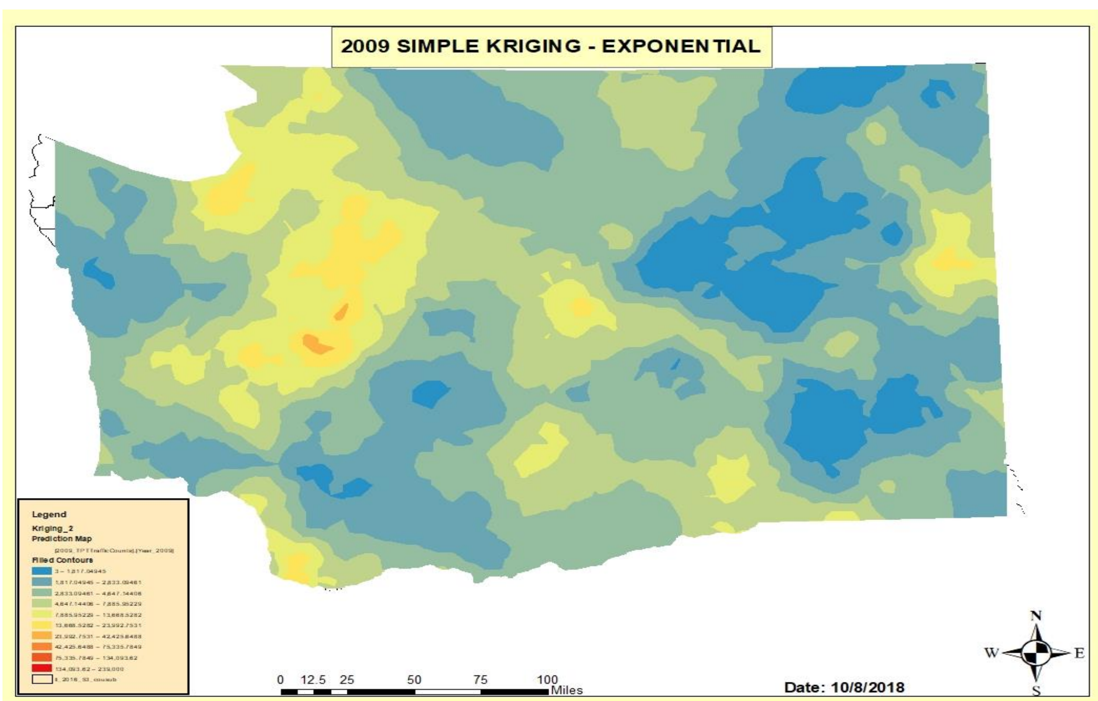

Figure 5.2. SK (Exponential) Map of Washington State AADT Data for the Year 2009 AADT

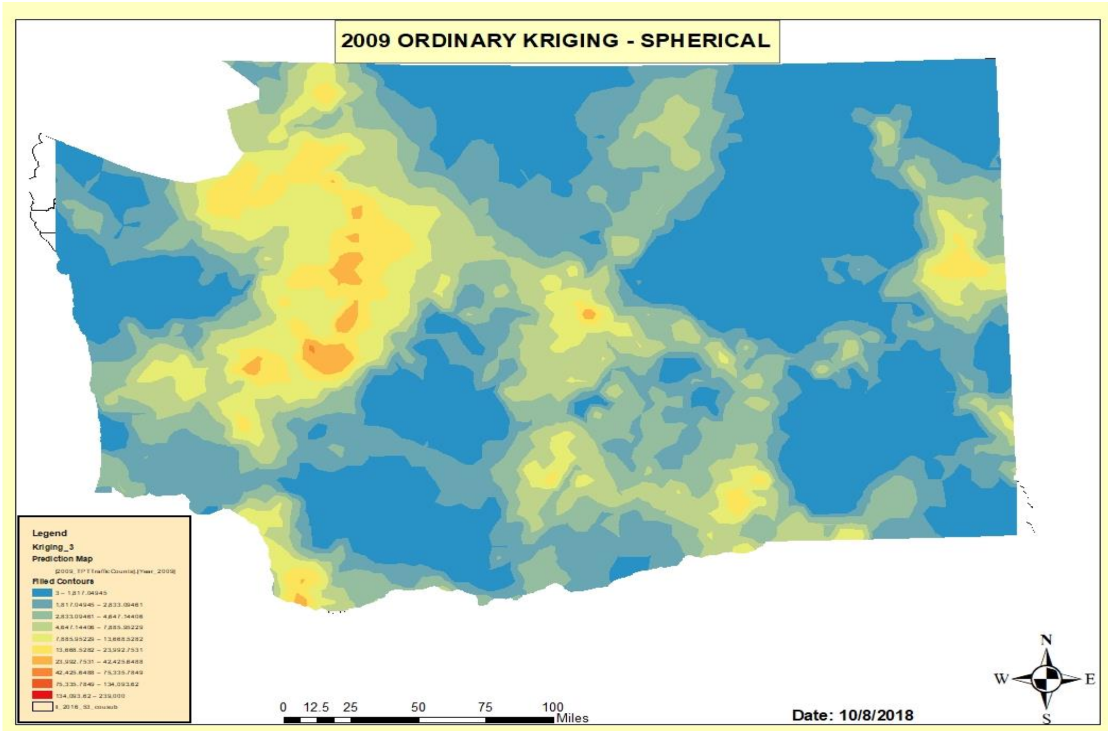

Figure 5.3. OK (Spherical) Map of Washington State AADT Data for the Year 2009 AADT

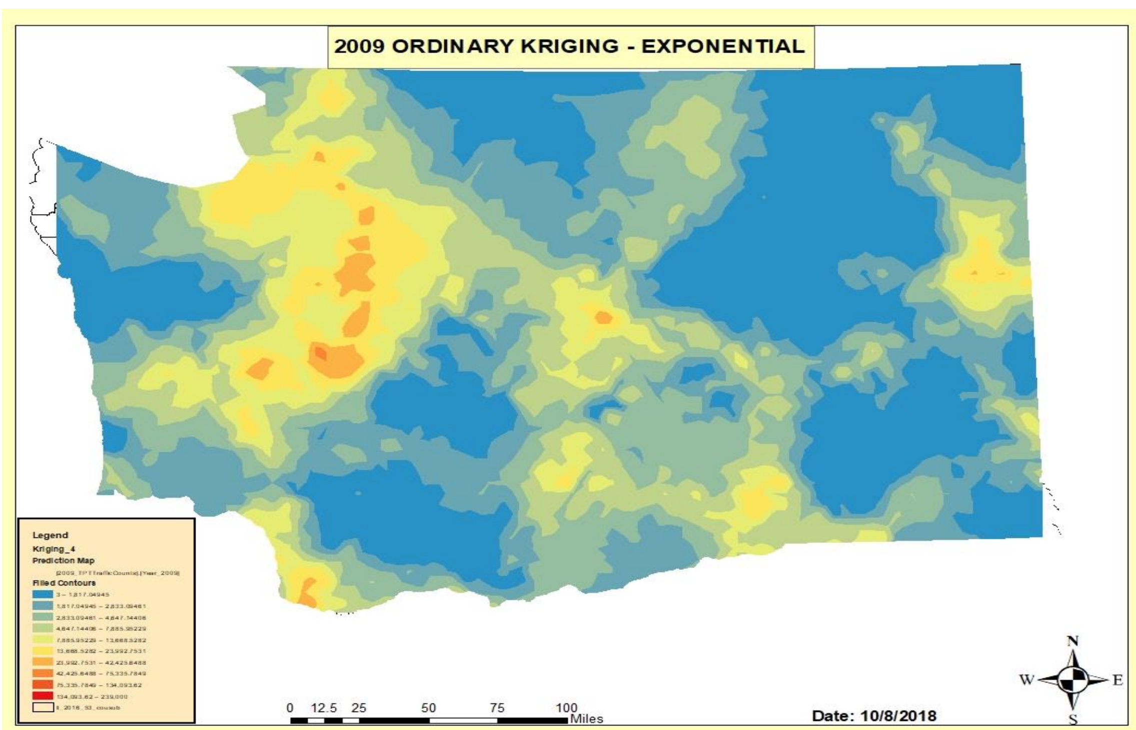

Figure 5.4. OK (Exponential) Map of Washington State AADT Data for the Year 2009 AADT

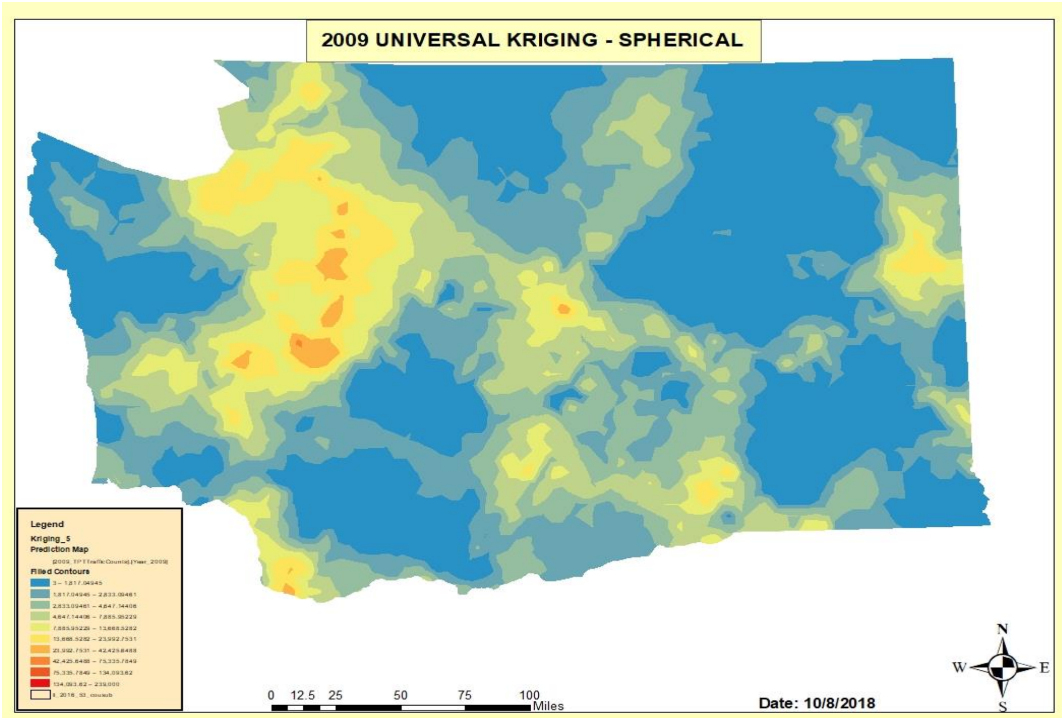

Figure 5.5. UK (Spherical) Map of Washington State AADT Data for the Year 2009 AADT

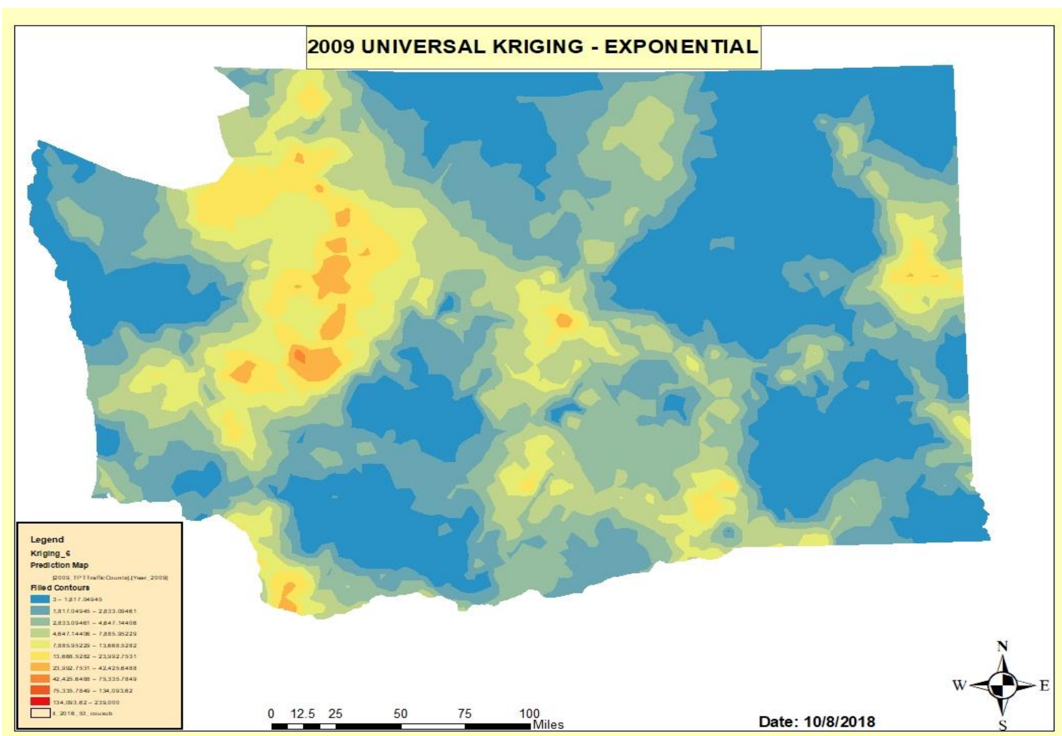

Figure 5.6. UK (Exponential) Map of Washington State AADT Data for the Year 2009 AADT

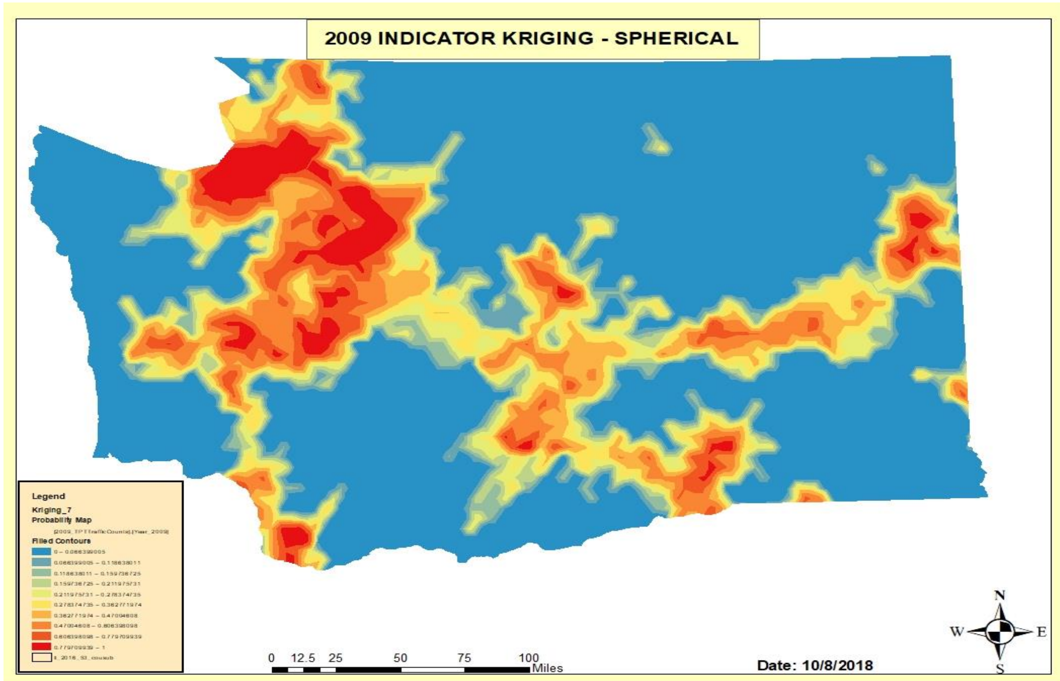

Figure 5.7. IK (Spherical) Map of Washington State AADT Data for the Year 2009 AADT

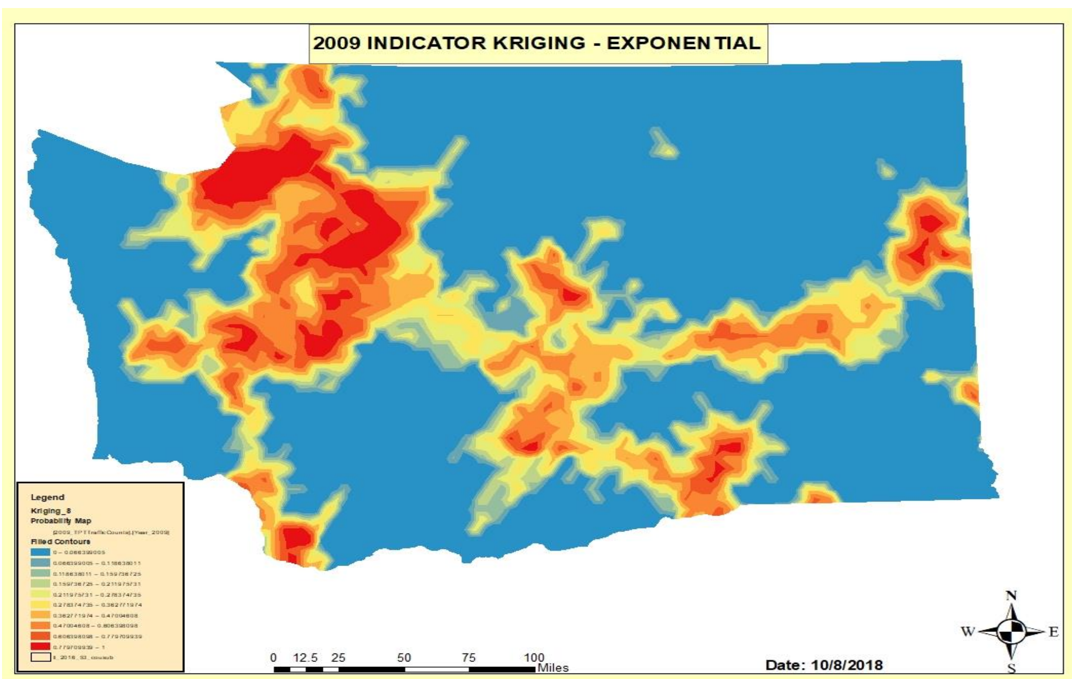

Figure 5.8. IK (Exponential) Map of Washington State AADT Data for the Year 2009 AADT

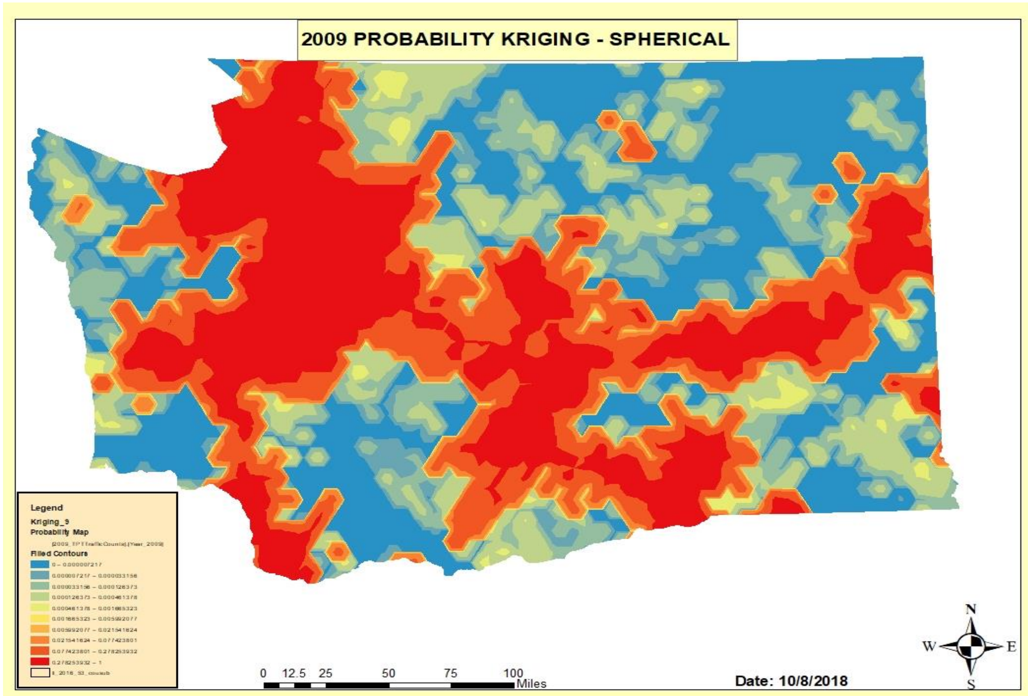

Figure 5.9. PK (Spherical) Map of Washington State AADT Data for the Year 2009 AADT

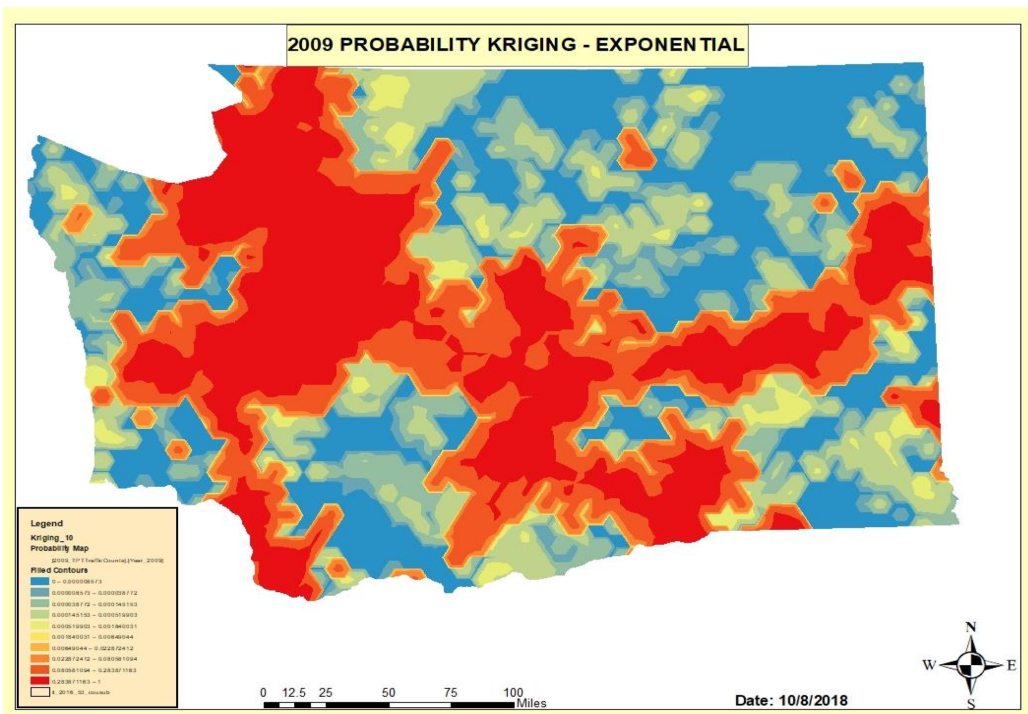

Figure 5.10. PK (Exponential) Map of Washington State AADT Data for the Year 2009 AADT

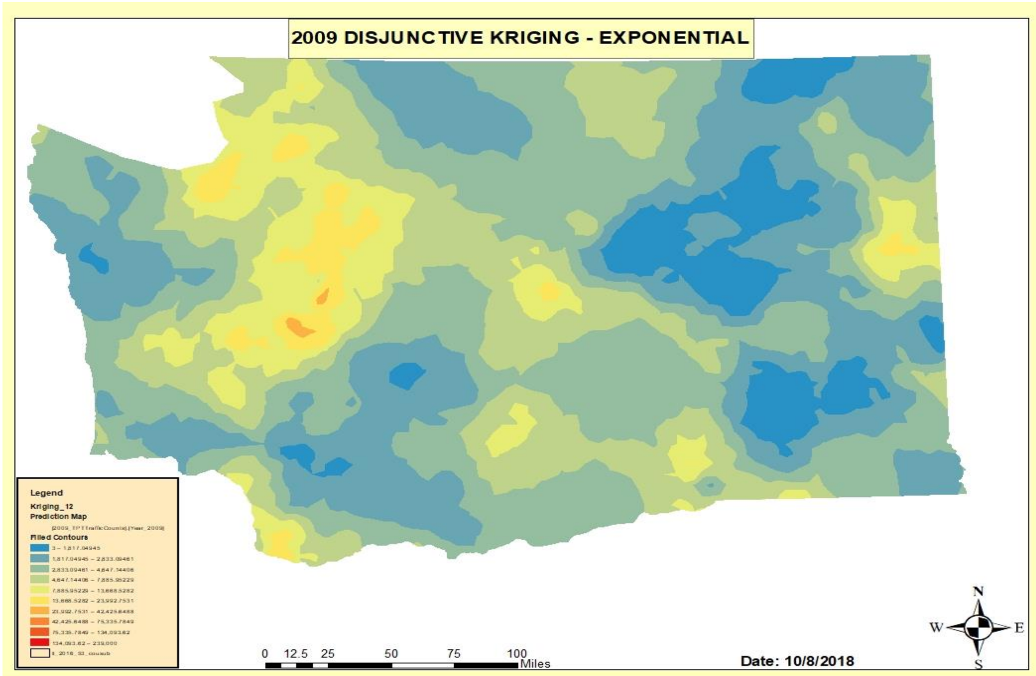

Figure 5.11. DK (Exponential) Map of Washington State AADT Data for the Year 2009 AADT

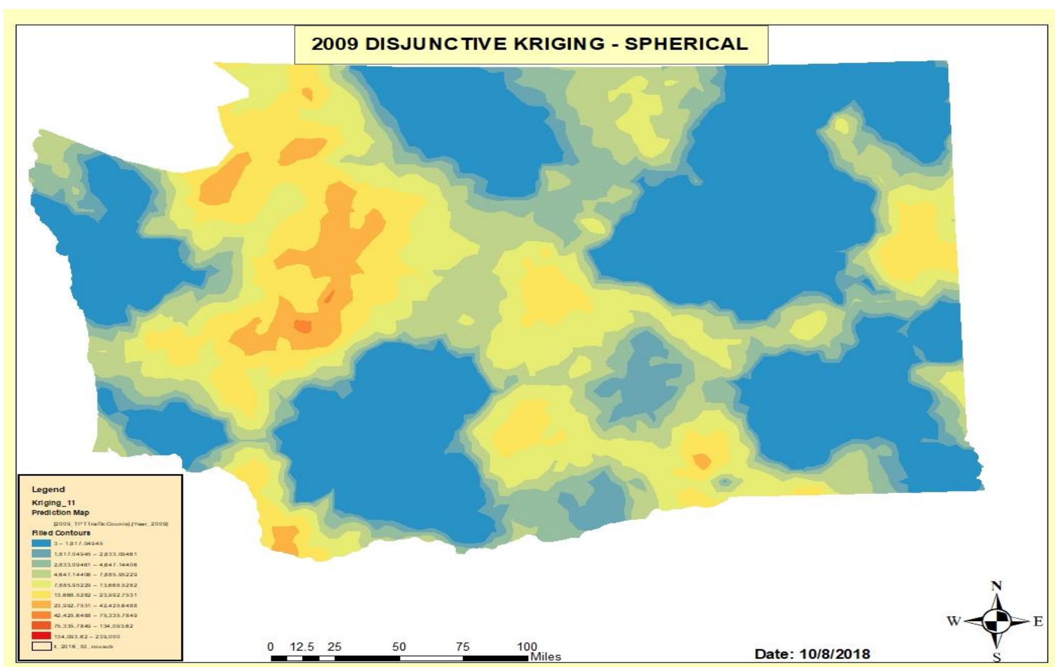

Figure 5.12. DK (Exponential) Map of Washington State AADT Data for the Year 2009 AADT

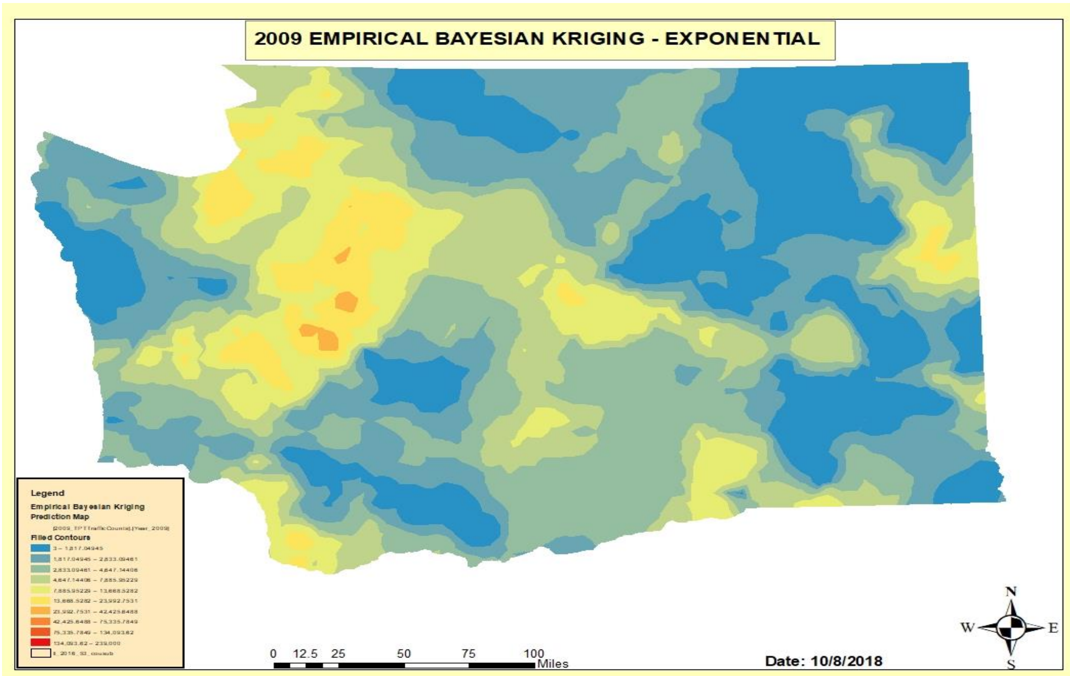

Figure 5.13. EBK (Exponential) Map of Washington State AADT Data for the Year 2009 AADT

5.4. Hypothesis Testing

A hypothesis as defined by Merriam Webster dictionary (2019) could be defined as an assumption or concession made for the sake of an argument. It is a tentative assumption made in order to draw out and test its logical or empirical consequences. This could be an interpretation of a practical situation or condition taken as the ground for action. This thesis made the following tentative assumption in order to test its empirical consequences.

Null hypothesis: mean prediction error is the same for all kriging methods.

Statistically,

H_0: $\mu_1 = \mu_2 = \mu_3$

Where μ_1 = linear kriging methods (Ordinary, Simple and Universal)

μ_2 = Non – linear kriging methods (Indicator, Probability and Disjunctive) and

μ_3 = Bayesian Method (Empirical Bayesian)

H_A = There is a difference in the mean prediction error

The hypothesis was tested with one-way ANOVA using the RMS Standardization error result for the years 2009 to 2016 AADT data set from section 5.2 of this thesis.

Table 5.6 to Table 5.11 below showed the result of the hypothesis test using one – way ANOVA test. The result of the test showed that p-value (0.000) is less than the significance level (0.05), thereby the null hypothesis is rejected, and the alternative hypothesis accepted. This implies that the prediction error of the kriging methods is not all equal. Also, the standard error of the estimate (square root of the mean-squared error) was 0.366140. This value indicates that the model might not meet the model assumption and thereby rejects the null hypothesis. In addition, the confidence interval for the difference between the means of OK is (2.143, 2.666). This range does not include zero, which indicates that the difference is statistically significant. This implies that the prediction errors of the kriging methods are not all equal.

Table 5.6. Hypothesis Testing

	In Reality	
Decision	H_0 is True	H_0 is False
Accept H_0	OK	Type II Error β = probability of Type II Error
Reject H_0	Type I Error α = probability of Type I Error	OK

One-way ANOVA: RMS versus Model

Table 5.7. Method of Hypothesis Testing

Null hypothesis	All means are equal
Alternative hypothesis	Not all means are equal
Significance level	$\alpha = 0.05$
Equal variances were assumed for the analysis.	

Table 5.8. Factor Information for the Hypothesis

Factor	Levels	Values
Model	6	DK, EBK, IK, OK, PK, SK

Table 5.9. Analysis of Variance for the Hypothesis

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Model	5	14.263	2.8527	21.25	0.000
Error	42	5.639	0.1343		
Total	47	19.902			

Table 5.10. Model Summary for the Hypothesis

S	R-sq	R-sq(adj)	R-sq(pred)
0.366410	71.67%	68.29%	62.99%

Table 5.11. Means and Standard Deviation for the Hypothesis

Model	N	Mean	StDev	95% CI
DK	8	0.97405	0.02040	(0.71262, 1.23549)
EBK	8	1.0613	0.0283	(0.7998, 1.3227)
IK	8	0.92927	0.01035	(0.66784, 1.19071)
OK	8	2.405	0.895	(2.143, 2.666)
PK	8	0.9423	0.0458	(0.6808, 1.2037)
SK	8	0.85082	0.02598	(0.58939, 1.11225)

Pooled StDev = 0.366410

Figure 5.14 showed that there is a significant difference between the sample mean of the methods. OK has the highest mean (2.4) and SK has the lowest mean (0.8) since there is the difference in the sample mean, this thereby conform to the fact that the null hypothesis should be rejected, and the alternative hypothesis accepted.

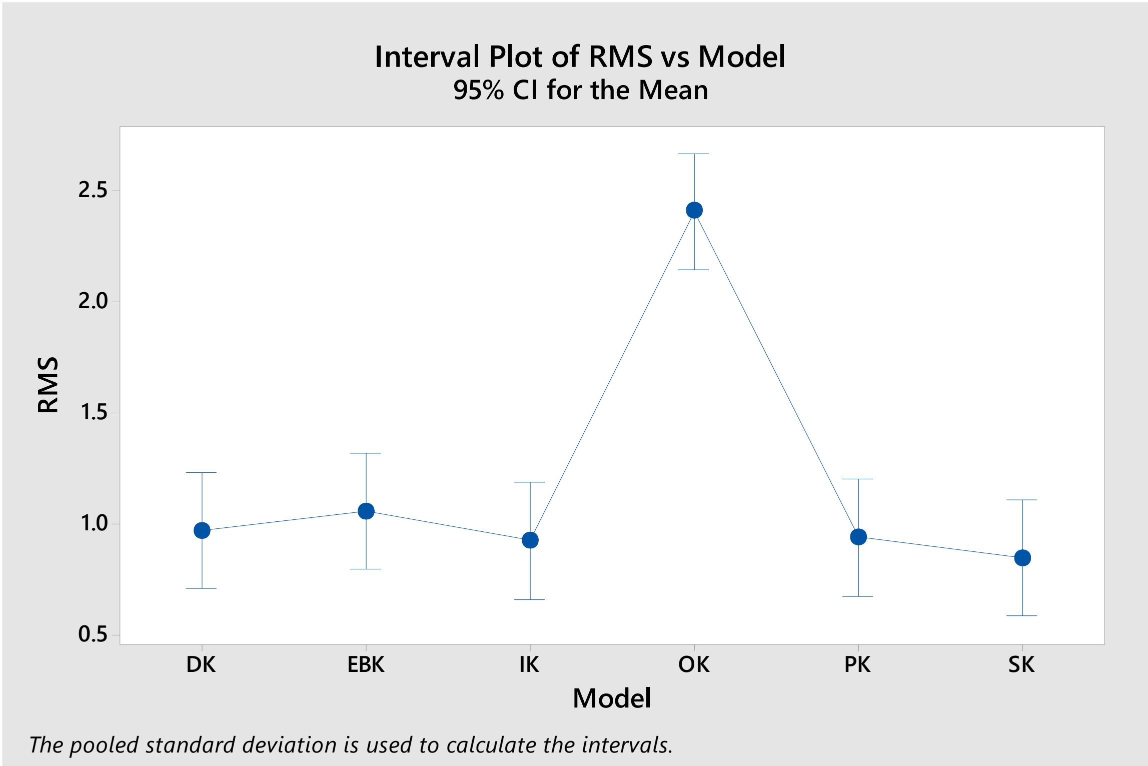

Figure 5.14. Interval Plot of RMS vs Model

5.5. Summary

This chapter discussed the results of the process carried out in order to answer questions 3 and 4 of the research questions and objectives. The exploratory data analysis result showed that the data is positively trended. It was also observed in the dataset from the neighborhood analysis that the dataset was highly clustered mostly towards the western part of the state. This was taken care of by first declustering the data before further analysis.

The crossvalidation result showed that the dataset is random and one kriging method and variogram model cannot be used to model for all the years. Each year had a different method of kriging analysis that best fit to model it. PK worked best in MSE with either spherical or empirical variogram model. DK worked best in RMSSE with spherical model and EBK performed best with an empirical model in RMSSE.

The result of the hypothesis also suggested that there is a significant difference between the mean prediction error of the kriging methods thereby the null hypothesis was rejected, and the alternative hypothesis accepted.

6. CONCLUSION AND RECOMMENDATION

6.1. Conclusion

With the help of this research, the following questions was addressed (1) what are the methods of AADT estimation and prediction; (2) what are the factors that influence the accuracy of AADT estimation and prediction methods; (3) which of the kriging methods is best used for AADT estimation and prediction and (4) what differences can be inferred between these kriging methods.

A systematic literature review was conducted to address research questions 1, 2 and 3. Florida turnpike state model, geographically weighted regression, artificial neural network, kriging interpolation, travel demand modeling, ordinary linear regression, origin-destination centrality-based method and support vector regression with data-dependent parameters were the identified methods of AADT estimation and prediction with ordinary linear regression discovered as the most used method over the years. Also, geographical location, road type, day of the week, seasonality, missing hourly volume, equipment theft, equipment damage/vandalism, and human error were identified as the factors that influence AADT estimation and prediction.

To address research question 4, the research compared linear, non-linear and bayesian methods of kriging using Washington state AADT data count. AADT data from WSDOT was analyzed to study the nature and characteristics of the data. The exploratory spatial data analysis revealed that the data was positively trended. Though some years suffered a decrease at some point, it had little or no effect on the overall positive trend of the data.

The result of the crossvalidation showed that the choice of a kriging method and variogram model cannot be known prior to interpolation. It also showed that probability kriging worked better than other kriging methods when used in mean standardization error (MSE) with either spherical or empirical variogram model while disjunctive kriging worked better than other kriging methods when used in root mean square standardization error (RMSSE) with the spherical model. Empirical bayesian kriging (EBK) performed better than other kriging methods when used with an empirical model in RMSSE. The result of the hypothesis also suggested that there is a significant difference between the mean prediction error of the kriging methods thereby the null hypothesis was rejected, and the alternative hypothesis accepted. This, therefore, can be concluded that the same kriging method cannot be used for the same data type from year to year due to the changing dynamics of AADT attribute.

With EBK having the least error from the result, it may be concluded that EBK will always work better when dealing with empirical variogram model of AADT data.

6.2. Recommendation

The following proposal can put forward for the consideration of future research work:

1. The changing dynamic attribute of AADT makes its analysis more rigorous as the user has to find the best kriging method same data from year to year. In the future, a method that can take care of this changing dynamic attribute should be investigated.
2. The study uses spherical and exponential variograms with different kriging methods. In the future, other variogram models can be used and the researcher can make a comparison between the different variogram models and kriging fittings.
3. The research used a precise data transformation and variogram model to perform the empirical bayesian kriging tool. In the future, the other options under model type can be used to test the empirical bayesian kriging and examine the predicted map.
4. The AADT data analysis used does not include simulation, this can be researched into the nearest future.

REFERENCES

Akobeng, A. K. (2005). Understanding systematic reviews and meta-analysis. *Arch Dis Child* 90: 845–848.

Asa, E., Saafi, M., Membah, J., & Billa, A. (2012). Comparison of Linear and Nonlinear Kriging Methods for Characterization and Interpolation of Soil Data. *Journal of Computing in Civil Engineering*, *26*(1), 11–18. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000118

Bagheri, E., Zhong, M., & Christie, J. (2011). Improving Group Assignment and AADT Estimation Accuracy of Short-Term Traffic Counts Using Historical Seasonal Patterns. In *ICTIS 2011* (pp. 1548–1554). Wuhan, China: American Society of Civil Engineers. https://doi.org/10.1061/41177(415)196

Bagheri, E., Zhong, M., & Christie, J. (2015). Improving AADT Estimation Accuracy of Short-Term Traffic Counts Using Pattern Matching and Bayesian Statistics. *Journal of Transportation Engineering*, *141*(6), A4014001. https://doi.org/10.1061/(ASCE)TE.1943-5436.0000528

Castro-Neto, M., Jeong, Y., Jeong, M. K., & Han, L. D. (2009). AADT prediction using support vector regression with data-dependent parameters. *Expert Systems with Applications, 36*(2), 2979–2986. https://doi.org/10.1016/j.eswa.2008.01.073

Chen, M., Xia, J., & Anaya, A. (2003). Estimating Average Daily Traffic Using ITS Data. In *Towards a Vision for Information Technology in Civil Engineering* (pp. 1–10). Nashville, Tennessee, United States: American Society of Civil Engineers. https://doi.org/10.1061/40704(2003)31

Chiles, J-P., and Delfiner, P. (1999). Geostatistics: *Modeling spatial uncertainty*, Wiley, New York.

Christakos, G. (1984). “On the problem of permissible covariance and variogram models.” *Water Resources. Res*., 20(2), 251–265.

Christensen, O. F., Roberts, G. O., & Sköld, M. (2006). Robust Markov chain Monte Carlo Methods for Spatial Generalized Linear Mixed Models. *Journal of Computational and Graphical Statistics*, *15*(1), 1–17. https://doi.org/10.1198/106186006X100470

Cook, D., C. Mulrow, and R. Haynes. (1997). Systematic reviews: Synthesis of the best evidence for clinical decisions. *Annals of Internal Medicine*, 126: 376-380

Cross-Validation—Help | ArcGIS for Desktop. (n.d.). Retrieved November 1, 2018, from http://desktop.arcgis.com/en/arcmap/10.3/tools/geostatistical-analyst-toolbox/cross-validation.htm

Mohammad D., Sinha K.C., Kucek T., and Scholer C.F. (1998). “Annual average daily traffic prediction model for county roads”, Transportation Research Record, pp.69-77,

De Lurgio, S. A. (1998). Forecasting principles and applications, McGraw-Hill, Irwin, NY.

Denyer D. and Tranfield, D. (2009). Producing a systematic review. Buchanan, David A. (Ed); Bryman, Alan (Ed), The Sage handbook of organizational research methods. Thousand Oaks: Sage Publications Ltd, 671-689.

Duddu, V. R., & Pulugurtha, S. S. (2013). The principle of Demographic Gravitation to Estimate Annual Average Daily Traffic: Comparison of Statistical and Neural Network Models. *Journal of Transportation Engineering*, *139*(6), 585–595. https://doi.org/10.1061/(ASCE)TE.1943-5436.0000537

Eom, J. K., Park, M. S., Heo, T. Y., and Huntsinger, L. F. (2006). “Improving the prediction of annual average daily traffic for non-freeway facilities by applying spatial statistical method." *Transportation Research Record 1968*, Transportation Research Board, Washington, DC.

Federal Highway Administration, *Travel estimation procedures for the local functional system*, 1994. [Online]. Available: http://isddc.dot.gov/OLPFiles/FHWA/013434.pdf

Figliozzi, M., Johnson, P., Monsere, C., & Nordback, K. (2014a). Methodology to Characterize Ideal Short-Term Counting Conditions and Improve AADT Estimation Accuracy Using a Regression-Based Correcting Function. *Journal of Transportation Engineering*, *140*(5), 04014014. https://doi.org/10.1061/(ASCE)TE.1943-5436.0000663

Figliozzi, M., Johnson, P., Monsere, C., & Nordback, K. (2014b). Methodology to Characterize Ideal Short-Term Counting Conditions and Improve AADT Estimation Accuracy Using a Regression-Based Correcting Function. *Journal of Transportation Engineering*, *140*(5), 04014014. https://doi.org/10.1061/(ASCE)TE.1943-5436.0000663

Florida Department of Transportation (2005). 2004 Florida Traffic Information. (CD-ROM), FDOT, Tallahassee, FL.

Fu, M., Kelly, J. A., & Clinch, J. P. (2017). Estimating annual average daily traffic and transport emissions for a national road network: A bottom-up methodology for both nationally-aggregated and spatially-disaggregated results. *Journal of Transport Geography*, *58*, 186–195. https://doi.org/10.1016/j.jtrangeo.2016.12.002

Gastaldi, M., Gecchele, G., & Rossi, R. (2014). Estimation of Annual Average Daily Traffic from one-week traffic counts. A combined ANN-Fuzzy approach.

Gastaldi, M., Rossi, R., Gecchele, G., & Lucia, L. D. (2013). Annual Average Daily Traffic Estimation from Seasonal Traffic Counts. *Procedia - Social and Behavioral Sciences*, *87*, 279–291. https://doi.org/10.1016/j.sbspro.2013.10.610

Georgios D. (2018) Crossvalidation, *Towards Data Science, https://towardsdatascience.com/cross-validation-70289113a072*

Goovaerts, P. (1997). *Geostatistics for natural resources evaluation,* Oxford University Press, New York.

How Kriging works—Help | ArcGIS for Desktop. (n.d.). Retrieved November 1, 2018, from http://desktop.arcgis.com/en/arcmap/10.3/tools/spatial-analyst-toolbox/how-kriging-works.htm

Isaaks, E. H., and Srivastava, R. M. (1989). *Applied Geostatistics*, Oxford University Press, New York.

Jiang, Z., M. R McCord and P. K.Goel. (2006). Improved AADT Estimation by Combining Information in Image- and Ground-Based Traffic Data *ASCE Journal of Transportation Engineering* 132 (7), pp. 523-600.

J. Negreiros, M. Painho, F. Aguilar and M. Aguilar, (2010). Geographical Information Systems Principles of Ordinary Kriging Interpolator. *Journal of Applied Sciences, 10: 852-867.* https://scialert.net/abstract/?doi=jas.2010.852.867

Konstantin K. (2001). Using linear and non-linear kriging interpolators to produce probability maps. *Journal of Environmental Systems Research Institute*, 380 New York Street, Redlands, CA 92373-8100. https://pdfs.semanticscholar.org/9620/5440cdce613d51345b991bd44ebc1d7c7369.pdf

Lam, W.H.K. and J. Xu. (2000). Estimation of AADT from Short-Period Counts in Hong Kong – A Comparison between Neural Network Method and Regression Analysis *Journal of Advanced Transportation* 34 (2), pp. 249-268.

Lowry, M. (2014). Spatial interpolation of traffic counts based on origin-destination centrality. *Journal of Transport Geography, 36*, 98–105. https://doi.org/10.1016/j.jtrangeo.2014.03.007

Lu, J., T. Pan, and P. Liu (2007). *Assignment of Estimated Average Annual Daily Traffic Volumes on All Roads in Florida*. Final Report Prepared for the Florida Department of Transportation, Tallahassee, FL.

McCord, M. R., Goel, P. K., Jiang, Z., & Bobbit, P. (2002). Improving AADT and VMT Estimates with High-Resolution Satellite Imagery: Simulated Analytical Results. In *Applications of Advanced Technologies in Transportation (2002)* (pp. 632–639). Boston Marriot, Cambridge, Massachusetts, United States: American Society of Civil Engineers. https://doi.org/10.1061/40632(245)80

NIST/SEMATECH e-Handbook of Statistical Methods, http://www.itl.nist.gov/div898/handbook/, date.

Paciorek, C. J. (2007). Computational techniques for spatial logistic regression with large datasets. *Computational Statistics & Data Analysis, 51*(8), 3631–3653. https://doi.org/10.1016/j.csda.2006.11.008

Rana A. Moyeed and Andreas Papritz (2002). An empirical comparison of kriging methods for nonlinear spatial point prediction. *International Association for Mathematical Geology, Vol. 34, No. 4, May 2002.* https://link.springer.com/content/pdf/10.1023%2FA%3A1015085810154.pdf

Roadway Data Home - TDA, MnDOT. (n.d.). Retrieved November 1, 2018, from http://www.dot.state.mn.us/roadway/data/index.html

Rossi, R., Gastaldi, M., & Gecchele, G. (2014). Comparison of Clustering Methods for Road Group Identification in FHWA Traffic Monitoring Approach: Effects on AADT Estimates. *Journal of Transportation Engineering*, *140*(7), 04014025. https://doi.org/10.1061/(ASCE)TE.1943-5436.0000676

Selby, B. and Kockelman, K. (2011) "Spatial Prediction of AADT in Unmeasured Locations by Universal Kriging." *Transportation Research Board 90th Annual Meeting, https://trid.trb.org/view/1092064.*

Shamo, B., Asa, E., & Membah, J. (2015). Linear Spatial Interpolation and Analysis of Annual Average Daily Traffic Data. *Journal of Computing in Civil Engineering*, *29*(1), 04014022. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000281

Sharma, S. C., Gulati, B. M., & Rizak, S. N. (1996). Statewide Traffic Volume Studies and Precision of AADT Estimates. *Journal of Transportation Engineering*, *122*(6), 430–439. https://doi.org/10.1061/(ASCE)0733-947X(1996)122:6(430)

Sharma, S. C., Lingras, P., Xu, F., and Liu, G. X. (1999). "Neural networks as an alternative to the traditional approach of annual average daily traffic estimation from traffic counts." *Transportation Research Record 1660*, Transportation Research Board, Washington, DC, 24–31.

Sharma, S., Lingras, P., Xu, F., & Kilburn, P. (2001). Application of Neural Networks to Estimate AADT on Low-Volume Roads. *Journal of Transportation Engineering*, *127*(5), 426–432. https://doi.org/10.1061/(ASCE)0733-947X(2001)127:5(426)

Shen, L. D., F. Zhao, and D. Ospina (1999). *Estimation of Annual Average daily Traffic for Off-System Roads in Florida*. Research Report, Florida Department of Transportation, Tallahassee, FL.

Staats, W. N. (2016). Estimation of Annual Average Daily Traffic on Local Roads in Kentucky. https://doi.org/10.13023/etd.2016.066

Systematic Reviews: CRD's guidance for undertaking reviews in health care. (n.d.), 294.

Tranfield, D., D. Denyer, and P. Smart. 2003. Towards a methodology for developing evidence-informed management knowledge by means of systematic review. *British Journal of Management* 14: 207-222.

Tukey, J. W. (1977). *Exploratory Data Analysis*. Indianapolis: Addison-Wesley Publishing Company.

Understanding disjunctive kriging—Help | ArcGIS for Desktop. (n.d.). Retrieved November 1, 2018, from http://desktop.arcgis.com/en/arcmap/10.3/guide-books/extensions/geostatistical-analyst/understanding-disjunctive-kriging.htm

Understanding how to create surfaces using geostatistical techniques—Help | ArcGIS for Desktop. (n.d.). Retrieved November 1, 2018, from http://desktop.arcgis.com/en/arcmap/10.3/guide-books/extensions/geostatistical-analyst/understanding-how-to-create-surfaces-using-geostatistical-techniques.htm

Understanding indicator kriging—Help | ArcGIS for Desktop. (n.d.). Retrieved November 1, 2018, from http://desktop.arcgis.com/en/arcmap/10.3/guide-books/extensions/geostatistical-analyst/understanding-indicator-kriging.htm

Understanding ordinary kriging—Help | ArcGIS for Desktop. (n.d.). Retrieved November 1, 2018, from http://desktop.arcgis.com/en/arcmap/10.3/guide-books/extensions/geostatistical-analyst/understanding-ordinary-kriging.htm

Understanding probability kriging—Help | ArcGIS for Desktop. (n.d.). Retrieved November 1, 2018, from http://desktop.arcgis.com/en/arcmap/10.3/guide-books/extensions/geostatistical-analyst/understanding-probability-kriging.htm

Understanding simple kriging—Help | ArcGIS for Desktop. (n.d.). Retrieved November 1, 2018, from http://desktop.arcgis.com/en/arcmap/10.3/guide-books/extensions/geostatistical-analyst/understanding-simple-kriging.htm

Understanding universal kriging—Help | ArcGIS for Desktop. (n.d.). Retrieved November 1, 2018, from http://desktop.arcgis.com/en/arcmap/10.3/guide-books/extensions/geostatistical-analyst/understanding-universal-kriging.htm

Using indicator kriging to create a probability map—Help | ArcGIS for Desktop. (n.d.). Retrieved November 1, 2018, from http://desktop.arcgis.com/en/arcmap/10.3/guide-books/extensions/geostatistical-analyst/using-indicator-kriging-to-create-a-probability-map.htm

U.S. Dept. of Transportation Federal Highway Administration (FHWA). (2001). “Traffic monitoring guide.” Rep. FHWA-PL-01-021, Office of Highway Policy Information, Washington, DC.

Wang, T. (2012). Improved Annual Average Daily Traffic (AADT) Estimation for Local Roads using Parcel-Level Travel Demand Modeling, 135.

Wang, X., and Kockelman, K. (2009). “Forecasting network data: Spatial interpolation of traffic counts using Texas data.” *Transportation Research Record 2105*, Transportation Research Board, Washington, DC, 100–108.

Wang, Y., & Kockelman, K. M. (2013). A Poisson-lognormal conditional-autoregressive model for multivariate spatial analysis of pedestrian crash counts across neighborhoods. *Accident Analysis & Prevention*, *60*, 71–84. https://doi.org/10.1016/j.aap.2013.07.030

Washington State Department of Transportation. (n.d.). Retrieved March 5, 2019, from https://www.wsdot.wa.gov/

What is Empirical Bayesian kriging? —Help | ArcGIS for Desktop. (n.d.). Retrieved November 1, 2018, from http://desktop.arcgis.com/en/arcmap/10.3/guide-books/extensions/geostatistical-analyst/what-is-empirical-bayesian-kriging-.htm

Wright, T., Hu, P. S., Young, J., & Lu, A. (1997). *Variability in traffic monitoring data. Final summary report* (No. ORNL/M--6154, 629487). https://doi.org/10.2172/629487

Yang, B., Wang, S.-G., & Bao, Y. (2011). Efficient local AADT estimation via SCAD variable selection based on regression models. In *2011 Chinese Control and Decision Conference (CCDC)* (pp. 1898–1902). Mianyang, China: IEEE. https://doi.org/10.1109/CCDC.2011.5968510

Zhang Z. and Hanson W. “Framework for estimating AADT using coclustering-based collaborative filtering”, *Transportation Research Board Annual Meeting*, 2009.

Zhao F. and Chung S. (2001) “Estimation of annual average daily traffic in a Florida county using GIS and regression”, Paper No. 01-3440, 2001 Transportation Research Board Annual Meeting, Washington, D.C., Jan. 2001. [Online] Available: http://www2.fiu.edu/~zhaof/research/adt-trb2001-Jan2001spacing.pdf

Zhao F. and Park N. (2004). “Using geographically weighted regression models to estimate annual average daily traffic”, *Transportation Research Record*, pp.99-107.

Zhong, M., Bagheri, E., & Christie, J. (2012). Improving Group Assignment and AADT Estimation Accuracy of Short-term Traffic Counts using Historical Seasonal Patterns & Bayesian Statistics. *Procedia - Social and Behavioral Sciences*, *43*, 607–617. https://doi.org/10.1016/j.sbspro.2012.04.134