2019

THE USE OF 3-D HIGHWAY DIFFERENTIAL GEOMETRY IN CRASH PREDICTION MODELING

Kiriakos Amiridis
University of Kentucky, kyriakos.amiridis@gmail.com
Digital Object Identifier: https://doi.org/10.13023/etd.2019.219

STUDENT AGREEMENT:

I represent that my thesis or dissertation and abstract are my original work. Proper attribution has been given to all outside sources. I understand that I am solely responsible for obtaining any needed copyright permissions. I have obtained needed written permission statement(s) from the owner(s) of each third-party copyrighted matter to be included in my work, allowing electronic distribution (if such use is not permitted by the fair use doctrine) which will be submitted to UKnowledge as Additional File.

I hereby grant to The University of Kentucky and its agents the irrevocable, non-exclusive, and royalty-free license to archive and make accessible my work in whole or in part in all forms of media, now or hereafter known. I agree that the document mentioned above may be made available immediately for worldwide access unless an embargo applies.

I retain all other ownership rights to the copyright of my work. I also retain the right to use in future works (such as articles or books) all or part of my work. I understand that I am free to register the copyright to my work.

REVIEW, APPROVAL AND ACCEPTANCE

The document mentioned above has been reviewed and accepted by the student's advisor, on behalf of the advisory committee, and by the Director of Graduate Studies (DGS), on behalf of the program; we verify that this is the final, approved version of the student's thesis including all changes required by the advisory committee. The undersigned agree to abide by the statements above.

Kiriakos Amiridis, Student

Dr. Nikiforos Stamatiadis, Major Professor

Dr. Timothy Taylor, Director of Graduate Studies

THE USE OF 3-D HIGHWAY DIFFERENTIAL GEOMETRY IN CRASH PREDICTION MODELING

DISSERTATION

A dissertation submitted in partial fulfillment of the
requirements for the degree of Doctor of Philosophy in the
College of Engineering
at the University of Kentucky

By
Kiriakos Amiridis
Lexington, Kentucky
Director: Dr. Nikiforos Stamatiadis, Professor of Civil Engineering
Lexington, Kentucky
2018

ABSTRACT OF DISSERTATION

THE USE OF 3-D HIGHWAY DIFFERENTIAL GEOMETRY IN CRASH PREDICTION MODELING

The objective of this research is to evaluate and introduce a new methodology regarding rural highway safety. Current practices rely on crash prediction models that utilize specific explanatory variables, whereas the depository of knowledge for past research is the Highway Safety Manual (HSM). Most of the prediction models in the HSM identify the effect of individual geometric elements on crash occurrence and consider their combination in a multiplicative manner, where each effect is multiplied with others to determine their combined influence. The concepts of 3-dimesnional (3-D) representation of the roadway surface have also been explored in the past aiming to model the highway structure and optimize the roadway alignment. The use of differential geometry on utilizing the 3-D roadway surface in order to understand how new metrics can be used to identify and express roadway geometric elements has been recently utilized and indicated that this may be a new approach in representing the combined effects of all geometry features into single variables. This research will further explore this potential and examine the possibility to utilize 3-D differential geometry in representing the roadway surface and utilize its associated metrics to consider the combined effect of roadway features on crashes. It is anticipated that a series of single metrics could be used that would combine horizontal and vertical alignment features and eventually predict roadway crashes in a more robust manner.

It should be also noted that that the main purpose of this research is not to simply suggest predictive crash models, but to prove in a statistically concrete manner that 3-D metrics of differential geometry, e.g. Gaussian Curvature and Mean Curvature can assist in analyzing highway design and safety. Therefore, the value of this research is oriented towards the proof of concept of the link between 3-D geometry in highway design and safety. This thesis presents the steps and rationale of the procedure that is followed in order to complete the proposed research. Finally, the results of the suggested methodology are compared with the ones that would be derived from the, state-of-the-art, Interactive Highway Safety Design Model (IHSDM), which is essentially the software that is currently used and based on the findings of the HSM.

KEYWORDS: 3-D Highway Geometric Design, Differential Geometry, Highway Safety & Crash Prediction Models, Gaussian Curvature, Mean Curvature, Generalized Linear Models.

Kiriakos Amiridis

02/13/2019

Date

THE USE OF 3-D HIGHWAY DIFFERENTIAL GEOMETRY IN CRASH PREDICTION MODELING

By
Kiriakos Amiridis

Nikiforos Stamatiadis
Director of Dissertation

Timothy Taylor
Director of Graduate Studies

02/13/2019
Date

To my beloved grandmother,

Maria Paschalidou

Στην πολυαγαπημένη μου γιαγιά,

Μαρία Πασχαλίδου

ACKNOWLEDGMENTS

The following dissertation, while an individual work, benefited from the insights and direction of several people. First, my Dissertation Chair, Nikiforos Stamatiadis, exemplifies the high quality scholarship to which I aspire. In addition, Dr. Stamatiadis provided timely and instructive comments and evaluation at every stage of the dissertation process, allowing me to complete this project on schedule. Next, I wish to thank the complete Dissertation Committee, and outside reader, respectively: Reginald R. Souleyrette, Gregory D. Erhardt, Arnold Stromberg, and Alexandre Martin. Each individual provided insights that guided and challenged my thinking, substantially improving the finished product.

In addition to the technical and instrumental assistance above, I received equally important assistance from my family and beloved friends; my parents, Giorgos and Kalliopi, and sister, Roxani, provided on-going support throughout the dissertation process and my studies in general. Also, I will always be thankful to my Professor at the National Technical University of Athens, Dr. Basil Psarianos, who conveyed to me the art and secrets of Highway Geometric Design at an early stage and whom I will always consider my Mentor. Furthermore, I wish to thank my high school mathematics professor, Mr. Stephanos Eleftheriou, who revealed to me the beauty and power of mathematics in a way I always keep in mind; I can still hear his words and feel his passion that fuel me with patience and persistence whenever I encounter a research problem that is waiting to be thoroughly explored. I will never forget the first time I met with Mr. Eleftheriou during which he recited to us in class, in ancient Greek, the five axioms of Euclidian Geometry from the prototype book of the “Elements” of Euclid; it was exactly then when I embraced and was fascinated with awe by the field of Geometry. Additionally, I would like to acknowledge all the outstanding teachers that educated and, most importantly, cultured me as well as all the schools I had the luck to attend during my life: 6th Elementary School of Alimos, 2nd Gymnasium of Alimos, 4th Unified Lyceum of Alimos, and the National Technical University of Athens; all of these schools generously and cost-free offered me the utmost merit of humanity: knowledge. Finally, I would like to express my gratitude to the country in which I was raised and which taught me to embrace any difficulty in life that may arise by learning from it and to approach life with endless optimism, Greece.

Kiriakos Amiridis

02/13/2019

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1 INTRODUCTION

Highway safety is a major health issue that requires continued efforts for effectively addressing it and developing sustainable preventive solutions. The roadway environment is a difficult and complex system that people have to deal with and is a major contributor to road traffic injuries and fatalities. In 2015, there were 32,166 fatalities and over 1.7 million injuries (NHTSA n.d.). There is a systematic effort to improve roadway design to address these issues and identify roadway design elements that could contribute to designing roadways having the potential to improve safety by creating an environment that drivers can easily understand. The objective of this research is to contribute to the enhancement of road safety through the development of a 3-dimensional (3-D) model for rural highways that would allow for a more accurate correlation of design elements to their potential crash contribution. The road surface will be modeled as a 3-D surface through differential geometry and B-spline surfaces, leading to a more realistic, complete, and accurate representation of the actual roadway geometry that explicitly or implicitly affects the crash occurrence probability.

Current highway safety research has developed crash prediction models that quantify the impact of single geometric elements on crash occurrence. For example, in the Highway Safety Manual (HSM) there are models that predict the effect of lane width, shoulder width, presence of median, etc. on crashes but each of them defines it singularly without considering their potential interactions (AASHTO 2010). The highway safety community has recognized the need to estimate these interactions and the recent approach to address it is to either estimate the contribution of each geometric element on the crash occurrence alone or use a set of crash modification factors to adjust the estimate for a base condition to the existing features estimating their effect through multiplication of these factors (Washington et al. 2010; Hanno 2004). The number of variable interactions can increase exponentially even when a few are considered, e.g. five variables can produce 27 interactions, resulting in a drastically reduced statistical power of the analysis and a higher probability of not producing statistically significant models. Although there are statistical techniques in theory that may address this issue, practically the problem is still apparent and indicates the need for a much more integrated and coherent modeling approach of the roadway elements.

Another issue that is relevant and underscores the importance of this research effort is the fact that to date highway safety research is based on a 2-dimensional (2-D) approach, i.e., horizontal and vertical alignment, whose principles were initially established in the 1940's and have not drastically changed since then. Over the years, although there have been changes in terms of adjusting minimum values and thresholds of various design elements, the overall methodology regarding highway safety estimation remains intact. Even though the roadway is a 3-D structure, the simplification of its projection in two planes, i.e., horizontal and vertical, has served well in the past when it was adequate to do so. However, given the computational power that is available nowadays one can argue that the geometric design process can be further improved in terms of incorporating a 3-D approach and metrics to express the roadway alignment. Moreover, this approach could be carried forward to safety evaluation and possibly enhance the ability to examine simultaneously the potential contribution of more than a single geometric element on crash occurrence. Today, 3-D geometric interactions are not reported in a quantifiable form in any highway design or safety manual. The coordination between the horizontal and vertical alignment is limited to earthwork estimation or optimization, but not during the design process because the 3-D design incorporation has traditionally been a mathematically and computationally demanding procedure. Therefore, the true extent of the design interactions and implications are not taken into consideration in terms of a holistic 3-D approach (Hassan and Easa 1998a, 2000; Hassan et al. 1996a, 2000; Hanno 2004). Others have noted that improper horizontal and vertical alignment coordination play a crucial role in crash frequency occurrence (Lamm and Smith 1999; Biduka et al. 2002) and could also confuse the driver in terms of selecting an appropriate operating speed (Lamm et al. 1999). Lamm and Choueri (1987) provide a very enlightening description of improper as well as desirable horizontal and vertical combinations that affect the driver's perception and expectation of the roadway. Easa et al. (1999, 2001, 2002) have also highlighted the horizontal and vertical coordination problem and its implications, which could result in sudden fluctuations in operating speed, and underestimation of horizontal element lengths for sight distance calculation. Moreover, the AASHTO Policy of Geometric Design for Highways and Streets (aka Green Book) offers some, qualitative in essence, guidance for horizontal and vertical coordination, indicating the fact that the need for 3-D solutions has been acknowledged from the scientific community as a concept meriting further development and research (AASHTO 2011). Therefore, the development of 3-D models could possibly enhance roadway safety estimation, allowing for the identification, calculation, and incorporation of a plethora of 3-D explanatory variables, well known from differential geometry, into new, more sophisticated and integrated statistical safety models. This integrated approach has the potential to change the entire perspective on roadway safety research and identify the synergistic contribution of roadway design elements on crashes as well as those that are more critical to be addressed. The research to be completed here will examine roadways as surfaces. As such, Differential Geometry will play a prevalent role because it is by definition the field of mathematics that studies the principles of curves and surfaces. Indeed, in the suggested methodology, the roadway surface will be modeled as a 3-D surface in a strict mathematical form.

This research intends to shed more light on the correlation between road safety in rural highways and the effects of combined, 3-D geometric design characteristics. Although crash models can be developed according to the guidelines of the HSM (AASHTO 2010), this research aims to improve the existing models, or even develop new ones, in which the 3-D information will be included. To date, 3-D design elements have not been used as explanatory variables in crash prediction models because of the manner in which roadways are designed, i.e., as a process based on 2-D metrics and elements. Therefore, the novel aspect that this research intends to add to the current literature is the incorporation of 3-D metrics as explanatory variables in crash prediction models. At this point, no specific metrics have been identified but the research proposed here will examine well-known metrics and elements from differential geometry and identify those that have the greatest potential to explain crash occurrence. The use of 3-D metrics could allow for the identification of design errors that are not apparent when one considers and studies in a separate manner the two 2-D alignments, i.e., horizontal alignment and vertical profile. This lack of analytical coordination in 3-D space has often resulted in design errors that had to be addressed after construction. The research proposed here aims to address such errors through the incorporation of 3-D metrics and controls during the design phase and utilizing a 3-D design. It is anticipated that the introduction of 3-D surface metrics will not only address design phase issues, but it could also enhance other applications, e.g. highway safety, and thus could offer the highway engineer a more holistic approach to roadway designers. A discussion of potential metrics is presented in the Methodology section.

It is critical to state that this methodology does not intend to question the validity of safety research findings up to date; it rather suggests a more robust and realistic approach to develop roadway safety models by taking advantage of the 3-D mathematically defined roadway surface that would allow the development of 3-D crash prediction models. In fact, Amiridis and Psarianos (2015a) have demonstrated the similarities and analogies between the 3-D and 2-D traditional approach through an approach they developed named "3-D Differential Road Surface" (3DDRS). In the 3DDRS model, the natural way in which current 2-D guidelines and thresholds could be converted and integrated into 3-D metrics has been demonstrated. Once a 3-D curve has been developed, it can be used to obtain the equivalent 2-D curves that are currently used in the development of alignments. There are two curves that can be used in this manner and those are the 3-D pseudo-geodesic curvature which is equivalent to the horizontal alignment and the 3-D pseudo-normal curvature which is equivalent to the vertical profile. Therefore, the research findings, minimum criteria, and thresholds that are used today will be incorporated in the preliminary models as the starting values, allowing for a direct comparison between the existing predictive models, e.g. HSM regression equations, and the 3-D safety models that will be developed. It is noted that the concepts of pseudo-geodesic and pseudo-normal curvatures were introduced in the highway engineering literature for the first time with the 3DDRS approach and a brief definition and description of these essential concepts are presented in the following chapters.

It is worth mentioning that these 3-D metrics can be calculated only if 3-D curves or surfaces are defined, and therefore the question that may naturally emerge is how can road safety be associated with differential geometry or, in other words, how can the principles of differential geometry be applied on highway design and safety estimation. The idea begins with the simple observation that the roadway surface is indeed a 3-D surface and therefore it should be treated and, most importantly, modeled in this way. Nowadays, the road surface is not viewed as an integrated mathematical structure, but as the byproduct that results from a number of intermediate steps. In a general roadway design process, the centerline is initially defined and then the lane and shoulder widths, combined with their respective cross slopes and superelevation rates, form the roadway surface. Therefore, the roadway surface is not viewed as a separate concept in the design process: it simply occurs. This research aims to change this practice and potentially prove the invaluable advantage of obtaining holistic metrics from an integrated and unified 3-D mathematical surface, compared to metrics that are obtained as pieces of information and do not consider the interactive effects of other variables on them.

2 LITERATURE REVIEW

The scientific community acknowledges that the 3-D coordination of the roadway is essential in highway safety but such an approach has not yet been explored, quantified or implemented in a systematic way (Lamm et al. 1999). The purpose of this thesis is to examine the use of 3-D models in safety prediction and suggest appropriate 3-D geometric metrics that can predict crash frequency. A literature review including prior 3-D attempts and approaches to highway engineering is first presented. However, it must be emphasized that these attempts were mostly related to the 3-D modeling of the roadway and not so much to linking 3-D design and highway safety. Additional efforts that have investigated a more accurate calculation of essential roadway design metrics utilizing a 3-D theory are also presented. Finally, a review of highway safety approaches as they are applied today is also given.

2.1 3-D Highway Design Approach

It can be stated that until World War 2, there were three main concerns when a roadway was constructed: the width, which had to be such in order to accommodate the dimensions of the vehicles, the structural ability of the pavements, which had to satisfy the forces that are imposed from the vehicles, and the grades, which had to be such in order to accommodate drainage runoff issues and not be too steep in order to allow vehicles to actually travel on the roadway (Hanno 2004). It was only after the 40's when geometric design came into play in order to accommodate the increasing speed of vehicles; it is no coincidence that the spiral curve for highway designs was introduced around that time in order to exactly ensure a smooth transition when entering a curve from a tangent. The fact that highways were initially tightly viewed as a military associated asset can be verified by the fact that when President Eisenhower was in office in the USA during the 1953-1961 period, the largest interstate system was designed at that time with military needs in mind. For example, it was advised to design large tangents in length in multiple parts around the USA in order to accommodate the landing or takeoff of airplanes in case needed in a period of war (FHWA n.d.a). Therefore, the current interstate system was not initially designed with a focus on sophisticated geometric design principles. However, at that time and for the scope of needs that the roadway system had to serve, the geometric design that was applied and based on the 2-D traditional design approach, i.e., horizontal and vertical alignment, was adequate.

However, recognizing the deficiencies of the conventional 2-D road design approach a number of researchers tried in the past to address the problem by introducing 3-D modeling approaches. Brauer (1942) was the first who identified the physical properties and meaning of the 3-D road curves and referred to their real differential geometric principles and mathematical functions through the necessity of examining the moving Frenet trihedron. Lorenz (1943) in another effort suggested a cylindrical barrel approach for obtaining the 3-D road axis within a 3-D route planning process. Many years later and in an effort to introduce a 3-D design process Freising (1949) suggested a geometric design system using the curve as the 3-D element for the route planning. Later, Scheck's approach (1973) involved a gradual dynamic optimization of the route plan in the horizontal and vertical alignment plans. Borgmann (1976) examined an interpolation of 3-D fixed points, where the hyperbolic transition curve was used as a flexible 3-D curve resulting from the static properties of continuous slab. Psarianos (1982) carried out extensive research into developing a model representation using the 3-D design elements of a 3-D tangent, a helix and a choroclothoid, which is actually a 3-D spiral curve. In the latter application, the choroclothoid was used as a transition bend between the straight segment and the helix. All of these methodologies were, in fact, 3-D modeling methodologies that proposed techniques of roadway design by incorporating entirely or partially 3-D metrics during the design process. These researchers were the first that tried to view the separate horizontal and vertical alignments into one holistic 3-D approach and essentially encouraged the highway design community and other researchers to start thinking along the same lines. The results of these findings cannot be evaluated or compared to the traditional 2-D approach in the sense of which methodology is better: the final outcome of both the 2-D and 3-D approaches is a roadway. Although one may agree that designing in 3-D is the natural way to design a 3-D infrastructure system as in the case of a roadway, the error incorporated in designing in 2-D compared to 3-D has not been actually quantified, since it is somewhat difficult to convince the highway engineering community to move towards a holistic 3-D design approach without having an absolutely quantifiable justification. However, the 3-D approach can indeed be compared to the 2-D approach when specific applications, such as sight distance, hydroplaning speed, and crash predictions, are calculated or analyzed with both methodologies and then compared to the actual values. In this case and only through the use of specific applications that would eventually produce quantifiable and comparable numerical results one can indeed verify the potential superiority of 3-D methodologies.

More recently, Kühn (2002, 2012) provided an extensive analytical formulation of the 3-D geometric design methodology of a roadway based on fixed, coupling, and dialogue elements. Zuo et al. (2007) also developed a 3-D road calculation methodology based on computer virtual simulation technology in order to solve the 3-D sight distance problem. Hao et al. (2007) integrated visualization in the highway alignment design process in an effort to efficiently address the 3-D road design problem. Makanae (2000) developed a 3-D alignment design system in the virtual space recreated by stereoscopy of aerial photographs. Other 3-D highway design methodologies based on various mathematical functions have been presented by Makanae (2002, 2007), Karri and Jha (2007), Kim and Lovell (2010), Jha et al. (2010), Kühn and Jha (2012), Karri, et al. (2012). All of these methodologies have mostly focused on visualization and optimization techniques basically in terms of the roadway aligment in space in order to eventually improve the highway design process. The visualization techniques are particularly helpful since they allow to design in 3-D space even if 2-D guidelines are used providing invaluable insight of the whole area topography; an aspect that is perhaps the most contributing and crucial factor in highway design. The optimization techniques are very interesting, since their ultimate goal is the automation of the highway design process or at least the suggestion of a satisfactory alignment to start with. Optimization is a rather specialized and unexplored field in highway design, but its ultimate success is questionable given the vast number of competing variables and constraints that come into play as well as the highly non-linear nature of the initial poblem. For example, there are techniques that can optimize the vertical alignment in order to produce the most cost-effective earthwork scheme, but this does not ensure that other especially competing factors such as safety, comfort, operational speed, and appearance among others will be satisfied in an adequate manner as well. Therefore, optimization techniques are very important and further research should be conducted on this field, but it should be viewed as a procedure of suggesting alignment alternatives and not a procedure that can be blindly trusted given also the uniqueness of each project whose parameters cannot be parametrized in a single objective function.

The application that has been mostly analyzed in terms of 3-D calculation is probably roadway sight distance. An interesting aspect of 3-D sight distance calculation is that its results can be compared to the traditional 2-D approach and therefore indicate whether the calculation with a 3-D modeling basis is more accurate. Hassan et al. (1996a) introduced an analytical 3-D model using the finite element method (FEM) whose elements are rectangular (4-node, 6-node, and 8-node) and triangular. Their sight distance model is advanced because it combines horizontal and vertical alignments and accommodates cross slopes and superelevation. Its primary contribution is in the mathematical expression of the roadway geometry, but the sight distance computation is a cumbersome numerical procedure. Nehate et al. (2006) described a methodology to find the available sight distance using Global Positioning System (GPS) data by examining the intersection of line of sight with the elements representing the road surface. Ismail (2007) expanded the 2-D models in Lovell et al. (1999, 2001) to include a 3-D component. This was accomplished using piecewise linear approximations to all of the curvature elements.

In the past, in order to evaluate the actual sight distance in real driving conditions, a number of 3-D models are found in the literature aiming to optimize the available sight distance (Garcia 2004; Zimmermann 2005; Romero and Garcia 2007; Yan et al. 2008; DiVito and Cantisani 2010; Moreno et al. 2010) Kim and Lovell (2010) presented another 3-D sight distance evaluation procedure using thin plate splines where an algorithm is used to determine the maximum available sight distance using computational geometry and thin plate spline interpolation to represent the surface of the road. The available sight distance is measured by finding the shortest line that does not intersect any obstacle. Jha et al. (2001) proposed a 3-D methodology for measuring sight distance, utilizing triangulation methods via an algorithm that was introduced for this purpose consisting of three stages, namely:

road surface development, virtual field of view surface development, and virtual line of sight plane development. However, the process involved multiple software platforms, thus delivering an accurate but non-flexible outcome. All of these sight distance related methodologies highlighted the need to incorporate 3-D methodologies during the design process. However, none of these methodologies quantifies the effect of 3-D metrics on crash occurrence.

Recently, Amiridis and Psarianos (2015a) have developed a 3-D roadway representation methodology developed named “3-D Differential Road Surface” (3DDRS) that allows for utilizing 3-D metrics to define the roadway surface. The proposed method outlined in the 3DDRS approach is a further advancement of the previously mentioned 3-D methods. Its main advantage lies in the fact that the entire roadway is treated as a 3-D mathematical surface, whereas the other methodologies are mostly concerned with the roadway centerline. In addition, the roadway surface is a mathematical 3-D surface in the sense that it is parametrically defined through an interpolation spline function. This is an additional advantage of the 3DDRS methodology because it is now feasible to derive from differential geometry any mathematical calculation of any 3-D metric that requires a mathematically defined surface through a parametric representation. Other 3-D methodologies cannot accomplish the latter because they simply do not parametrize the roadway surface as a, strictly speaking, mathematical surface. Moreover, in the 3DDRS methodology, the resultant 3-D curvature values are strictly controlled in order to belong within a given domain that the user defines based on current highway engineering policies and regulations according to the Green Book (AASHTO 2011). The latter statement is feasible because, in the 3DDRS methodology, once a 3-D curve, i.e. 3-D roadway centerline, is developed, it can be utilized in order to obtain the 2-D curvatures that are currently used in the development of the horizontal alignment and vertical profile. This transition from 3-D to 2-D is achieved via two metrics as described in the 3DDRS: the 3-D pseudo-geodesic curvature that is equivalent to the curvature of the horizontal alignment and the 3-D pseudo-normal curvature that is equivalent to the curvature of the vertical profile.

The 3DDRS methodology has been successfully utilized for a more accurate calculation of the hydroplaning speed directly in 3-D space (Amiridis and Psarianos 2015b). The methodology allows for an easy and automated calculation of geometric parameters of segments, such as segments in which the superelevation rate changes, that could not be calculated before given the available/pre-existing hydroplaning models. Amiridis and Psarianos (2016) also used this methodology to calculate the available sight distance of a roadway directly in 3-D space and in a fully automated and accurate manner. The presented methodology overcomes the conventional 2-D approach of studying the actually 3-D roadway configuration separately and sequentially in its horizontal and vertical alignments. The road surface was simulated as a 3-D B-spline surface with continuous side barriers whose road centerline has been in turn, modeled as a 3-D B-spline curve. Through this approach, the road centerline as well as the right and left edge lines which play a crucial role in the sight distance calculation are mathematically defined both geometrically and analytically through explicit equations.

2.2 Highway Safety Approaches

The purpose of any highway safety model is the accurate prediction of crashes through statistical modeling. Given the fact that the dependent variable is the number of crashes, i.e., discrete counts, specialized regression models and theoretical statistical knowledge is absolutely necessary in order to analyze the data in a proper manner. It is no coincidence that the modeling of crash data is a field in which a number of statisticians conduct a large portion of their research and are fully devoted to. The list of models that have been proposed with their advantages, disadvantages, and implications is extensive and cannot be fully included in this literature review because it would exceed its scope. However, the basic types of statistical processes will be described but in a more informational/tabulated rather than detailed fashion. A very thorough literature review regarding statistical techniques with their associated theoretical background can be found in Lord and Mannering (2010). A concept that should be defined from the outset since it relates to the type of approach to be taken deals with the issue of whether the variance of a dataset is greater than (over-dispersion) or less (under-dispersion) than the mean.

Crashes are positive integers and therefore count models should be utilized in order to model them. The most basic regression model in safety analysis is the Poisson regression. However, the basic assumption of this model is that the variance is equal to the mean and that makes the Poisson model rather inflexible. In crash datasets, the common scenario is the presence of over-dispersion due to the excess of zeros in the dataset. The most basic model that accommodates over-dispersion is the Negative Binomial regression and this is why this regression type is the most commonly used in the literature (Maycock and Hall 1984; Hauer et al. 1988; Miaou 1994; Maher and Summersgill1996; Karlaftis and Tarko 1998; Hirst et al. 2004; Lord and Bonneson 2007; Daniels et al. 2010). Other models that have been tried with varied success include Zero Inflated Negative Binomial regression that attempts to address the presence of a large number of zero crashes. Zero Inflated Models are also an option when dealing with crash data, but this is not applicable here. Zero inflated models, such as Zero Inflated Poisson (ZIP) or Zero Inflated Negative Binomial (ZINB), should only be used when systematic zeros are present. Systematic zeros are zeros that are present not because of a stochastic process, but because of a deterministic one. However, in this research, a crash can occur in any location via the random process imposed by the nature of the problem, and therefore this is why the zero inflated models are not appropriate in this case. In addition, the Gamma Regression Model can be used to account for under-dispersion, but under-dispersion is not the typical situation as noted above. The Bivariate/Multivariate Model can also be utilized for a number of scenarios, such as to handle different types of crashes or to predict the probability, not absolute number, of crash occurrence; it is noted that the Bivariate Model is essentially the Logistic Regression model. A more specialized model is the Random Effects Model that can account for temporal and spatial correlation.

Another model that is widely used nowadays, especially in the HSM, is the Bayesian model that can be applied as both a parametric or non-parametric model. In other words, a distribution may be assumed or not for the data. However, the latter is not the biggest strength of the Bayesian model in crash analysis since the distribution of crashes can be assumed with an adequate enough confidence as Poisson distribution anyway. The biggest strength of the Bayesian model is that the opinion of the analyst/expert plays a role in the statistical analysis process. Nonetheless, it must be emphasized that the HSM does not exactly utilize the Bayesian model, but the Empirical Bayesian model. The Empirical Bayesian approach is something between the Frequentist and Pure Bayesian approach in the sense that researchers first examine the data and then provide their expectations, whereas in the Pure Bayesian case, the researcher is "not permitted" to look at the data; the latter is a critique towards the Empirical Bayesian approach in general.

Issues that could affect the development of crash models include, but are not limited to: time-varying explanatory variables, temporal and spatial correlation, data under-reporting, small sample size, and injury severity and crash type correlation, leading to wrong parameter estimates. In addition, omitted-variable bias and endogenous variables may result in a distorted form of the explanatory variables. The resultant impact of all of these problematic issues is essentially the production of inflated or reduced variable coefficients.

Lord and Mannering (2010) have underscored the complexity of contributing factors incorporated in crash data as well the advanced statistical tools that are necessary to be accounted in order to correctly model crash data. Although there is a vast amount of research pertaining to the optimal statistical model that should be used, these statistical techniques are useful when a new model is intended to be created. Practically, the prediction of crashes are currently carried out with the guidelines of the HSM in which all regression equations are readily applied and no specialized statistical knowledge is essentially needed. The accepted way in which crash predictions are practically performed nowadays is described in the discussion that follows.

2.2.1 Highway Safety Manual

All scientific, observational or empirical research pertaining to highway safety was included for the first time in a manual in 2010 with the publication of the HSM. The core of the HSM approach lies in three main pillars: Safety Performance Functions (SPFs), Crash Modifications Factors (CMFs), and the Empirical Bayes (EB) Statistical Approach, which are briefly described below.

SPFs are defined as “equations used to predict the average number of crashes per year at a location as a function of exposure and, in some cases, roadway or intersection characteristics (e.g., number of lanes, traffic control, or median type)” (HSM 2010). The most common exposure factors that are utilized in highway segments, which are the segment types of interest in this thesis, are the AADT and the segment length. It is expected, in general, that both the AADT and the segment length have a positive relationship with the number of crashes. Equation 1 presents a generalized SPF which is typically developed based on base conditions, i.e., those that are most frequently encountered in the system analyzed. However, besides these basic exposure factors, SPFs can account for site-specific conditions. Historical crashes can be incorporated in the prediction analysis and in this case, the Empirical Bayes model is utilized by calculating a weighted average of observed, i.e., actual, and predicted crashes that were derived from an SPF (Hauer 1997).

$$\text{Predicted Crashes}=\exp[a+\beta \cdot LN(AADT)+LN(\text{Segment Length})] \quad (1)$$

As the FHWA indicates (FHWA, n.d.b), SPFs have three basic functions in the highway safety evaluation process: network screening, countermeasure comparison, and project evaluation.

- *Network Screening*: SPFs can be utilized in order to identify locations in which a safety improvement would be meaningful to consider. These segment locations are indicated by comparing the observed crashes with the predicted segment crashes from other sites with similar roadway characteristics, AADT, and segment length (Part B, HSM).

- *Countermeasure Comparison*: SPFs can be utilized in order to evaluate in a quantifiable manner the impact of treatments and/or countermeasures on specific site locations, which function as base conditions, in terms of crash frequency increase or decrease. This is achieved by first predicting the number of crashes with an SPF for the so-called baseline condition, in which no treatment has been applied, and then applying, i.e., multiplying, a Crash Modification Factor (CMF) to the predicted crashes derived from the baseline SPF. Crash predictions derived from SPFs can be adjusted accordingly with the Empirical Bayes method if historical crashes are available (Part B, HSM).

A CMF is a multiplication factor that essentially indicates the percentage increase or decrease of predicted crashes for a specific element that could be either a geometric design component or a safety treatment in order to calculate the highway safety effect, i.e. improvement or not, of the element or treatment (FHWA, n.d.b). A CMF less than 1.0 has a positive impact on highway safety, whereas a CMF greater than 1.0 has a negative impact on highway safety. For example, a CMF for total crashes for installing centerline rumble strips on rural major collector roads is 0.86, i.e., 14 percent% reduction in crash frequency, whereas a CMF for converting an urban 4-lane cross section to a 5-lane cross section is 1.11, i.e., 11 percent increase in crash frequency (HSM 2010; FHWA, n.d.b).

- *Project Evaluation*: The prevailing process is to utilize the Empirical Bayes method in order to develop CMFs. The basis of the Empirical Bayes method is associated with the SPF combined with historical crashes. The SPFs are essentially calibrated to optimally capture the specific characteristics of a location for a specific time period which is explicitly described in the HSM. Finally, the output of this calibration process is in the form of CMFs (HSM, 2010; Hauer, 1997).

2.3 Summary

The literature review conducted here focused on two elements: 3-D highway geometric design and highway safety modeling. The intent of this thesis is to provide the link between these two fields in a scientific, but also practical manner providing ready-to-use models backed up by a thoroughly described methodology.

Although a number of attempts have been completed starting from the 40’s to develop a 3-D implementation of roadway design, they have not been formally implemented in an explicit quantifiable manner in terms of crash prediction models. The objective of most of these attempts was how to model the roadway, more specifically the roadway centerline, in 3-D space, and not to conduct highway engineering-oriented applications based on this 3-D modeling approach. Extensive research has been conducted acknowledging the fact that the separate coordination of the horizontal and vertical alignment is problematic, but this acknowledgment is not provided in a quantifiable or systematic manner. This is why perhaps, the process of incorporating 3-D metrics in the highway design process has been relatively slow and, in many cases, non-existent. Although interesting 3-D models may have been developed, their potential value cannot be verified because they cannot be compared in a one-to-one and quantifiable manner with current, 2-D procedures because there are no applications that can support their potential advantage. Therefore, the missing part of these 3-D methodological approaches is their practical implementation on applications that can produce quantifiable results by utilizing 3-D metrics as the basis of the analysis, and then comparing these results with the 2-D approach. Only then will 3-D metrics become acceptable and be included in the design process via their explicit integration in manuals and guidelines.

An extensive and very scientifically challenging research, mostly in terms of theoretical and applied statistical methodologies, has been conducted for addressing safety modeling and prediction. All of this knowledge that started compiling from the 70's, mostly in observational and empirical terms at that point, was organized and published in the first edition of the HSM, rather recently, in 2010. The lateness of the HSM as a knowledge compilation underscores the complexity of the crash modeling procedure and mostly the lack of required data in order to quantify the results even though highway safety is a major societal issue in terms of death and serious injure rates. The lack of data is an issue that has been vastly improved in the recent years and will continue to improve in an even more rapid trend. On the other hand, highway safety techniques lack in the incorporation of 3-D metrics as explanatory variables that would potentially assist in a more accurate prediction of crashes. At this point, the state-of-the-art process in crash prediction is the utilization of SPFs. The next step would therefore be the creation of 3-D SPFs in which the desired gap between 3-D design and highway safety would be bridged; this is in fact the objective of this doctoral thesis.

A new 3-D roadway modeling methodology, i.e., 3DDRS, has been developed that can represent the roadway as a 3-D mathematical surface expressing it as a single equation (Amiridis and Psarianos 2015a). A number of applications have been developed based on this prior work, such as sight distance calculation, hydroplaning speed calculation, and incorporation of the cut and fill slopes in the roadway surface itself. The use of this methodology will be investigated and form the basis of this research in order to introduce 3-D metrics whose effect could be quantified in terms of crash frequency occurrence. Another objective of this research is to introduce 3-D metrics in highway design per se in addition to the potential development of statistical models that can predict the number of crashes utilizing the 3-D geometric characteristics of a roadway. The application of the methodology to road safety could significantly strengthen the hypothesis that there is added value in incorporating 3-D metrics in the design process and to show that the highway design community needs to increase its efforts in understanding the 3-D geometric effects of a roadway. The need to consider the roadway design in the 3-D space has been in the forefront of recent research and it may be time to replace the traditional 2-D design approach with a more robust 3-D approach.

3 METHODOLOGY

This chapter describes the main steps of the suggested methodology and clarifies the sequence of tasks to be carried out. The final output of this research attempt will be the suggestion of a crash prediction model/process that utilizes a 3-D highway model in order to ultimately incorporated 3-D geometric metrics derived from differential geometry theory that will function as explanatory variables into regression models. The type of the statistical model and its relevant specifications that will be utilized is discussed in the following sections. The 3-D highway model that will be applied will be the 3DDRS methodology (Amiridis and Psarianos 2015a.).

3.1 3-D Surface Modeling Approach

The 3DDRS methodology to be used in the proposed research should not be compared to the other suggested 3-D approaches in terms of the geometric design or optimization process, but mostly in terms of how it models the roadway surface. In contrast to the 3DDRS, all existing 3-D approaches focus on the development of the road centerline and not on the road surface as a whole. This difference in how the roadway is viewed might turn out to be the most crucial contributing factor in road safety. Additionally, in the proposed methodology not only is the road surface modeled as a 3-D surface, but as noted above this 3-D surface is governed by an explicit mathematical equation: a fact not present in any of the 3-D approaches discussed above. For example, just as a sphere or cone has its own vector form, so does each unique roadway surface that is modeled through this methodology/technique. This allows for various geometric calculations and manipulations that could produce useable metrics for safety estimations. All of the geometric information incorporated in the roadway surface is integrated in a substantially robust mathematical equation and any application, not only for road safety, can be built based on this equation. This equation allows mathematical operations, such as differentiation and integration, to be easily applied on it with considerable computational speed. With this approach, any differential 3-D geometric metric can indeed be calculated based on this interpolation equation. Given the information provided in the literature, it is highly unlikely that the Gaussian curvature, for example, or any other 3-D geometric metric can be calculated based on any of the other 3-D approaches. If a surface is not defined in a strict mathematical formulation as in the 3DDRS methodology, then the Gaussian curvature, and many other significant 3-D metrics that govern the properties of surfaces as indicated by differential geometry theory, cannot be calculated by definition; it is like trying to define a tangent on a non-existent circle. Therefore, the whole goal of this thesis would be clearly unfeasible if any 3-D approaches are used other than the 3DDRS. Hence, the use of the 3DDRS does not imply that it is, in general, better than other 3-D methodologies, but it is the optimal for the specific scope and needs of this research. Since the objective of this research is to address the effects of combined 3-D geometric metrics of a roadway on highway safety, 3-D metrics must be able to be calculated in the first place and therefore the 3DDRS methodology is considered appropriate for utilization here. Current 3-D methods mainly focus on how to optimize the road centerline either purely theoretically or computationally and do not represent realistically the roadway surface itself. In essence, the other models consider that the task is accomplished once the centerline alignment is defined. These methodologies cannot consider an existing roadway and then run applications, e.g. road safety, based on it; this is simply not what they are created for. Finally, the 3DDRS methodology introduced is not useful only for geometric design, but for a large variety of applications because of its flexibility and robustness since all of the geometric information can be derived from a single equation.

The two most critical elements that make the 3-D design process feasible in the 3DDRS approach are the pseudo-geodesic and pseudo-normal curvature. The pseudo-geodesic curvature vector is a generalization of the curvature vector of a curve in the plane and is defined as the arithmetic projection of the 3-D curvature vector of a point to the horizontal plane. In particular, positive values of the pseudo-geodesic curvature correspond to a right turn of the steering wheel as the length, i.e., stationing, of the 3-D road centerline increases, whereas negative values of pseudo-geodesic curvature correspond to a left turn. Equivalently, when the road centerline becomes a straight line, i.e., tangent, then the pseudo-geodesic curvature approaches zero. Equivalently, the pseudo-normal curvature is a generalization of the curvature of the vertical alignment and is defined as the arithmetic projection of the 3-D curvature vector of a point to the vertical plane. Positive values of the pseudo-normal curvature correspond to the curvature of 3-D sag curves, whereas negative values of the curvature correspond to the curvature of 3-D crest curves. This means that the user can impose different limits, explicitly presented in the guidelines of the Green Book, in an algorithm that utilizes a user-friendly manner, depending on the type of vertical curve, i.e., crest or sag.

This approach underscores the correspondence between the values of pseudo-geodesic and pseudo-normal curvature to the well-known horizontal and vertical curvatures of the conventional horizontal and vertical alignment, respectively. The highway designer can then readily associate limiting values based on design policies and guides with the proposed 3-D methodology. Implementing the 3DDRS methodology, the final road design outcome can conform totally to the current, or future, accepted design policies. Therefore, the proposed methodology could be considered as an advancement of current practices and not as an approach that calls them into question. The pseudo-geodesic and pseudo-normal vectors were defined in 3DDRS especially to incorporate the transition from the conventional 2-D approach to the suggested 3-D approach and show that all existing knowledge and guidelines can be incorporated into 3-D design with respective adjustments. Their implementation in this research is original, as this technique is being implemented for the first time.

Given the fact that the roadway is modeled in 3-D space, it can be rationally argued that all further calculations and controls, e.g. sight distance calculation, hydroplaning speed estimation etc., based on the modeling of the roadway surface as a 3-D mathematical surface will be more accurate and precise. In fact, three crucial applications have been directly based on the 3DDRS approach proving that the 3-D realization of the roadway can lead to significant conclusions and provide a general framework based on which essential highway engineering metrics can be evaluated, and, most importantly, calculated with robust algorithms on a universal basis by taking into consideration any possible geometric combinations.

Amiridis and Psarianos (2015b) developed a mathematical methodology that allows a more accurate calculation of the hydroplaning speed directly in space. Currently, the calculation of the water flow path is based on a Digital Terrain Model (DTM) that is constructed based on the horizontal and vertical alignments and cross-section superelevation runoff design. Along this DTM, water flow paths have to be calculated from points on the pavement edge lines along which the water film depth will result based on a rainfall intensity. With the proposed methodology, the water film calculation is implemented in an integrated and fully automated way without having to account the specific location of a point on the edge lines. In addition, with this methodology, all calculations are solely based on the geometry of the surface, which has been modeled as a 3-D B-spline surface.

The advantage of this approach is that the geometric parameters of certain segments, such as segments in which the superelevation rate changes that could not be calculated in an analytical manner with current models, can now be calculated easily, automatically, and with no limits regarding specific conditions, e.g. certain geometric combinations, that must be met in order to yield the application of the methodology feasible. For example, the sight distance equation in the Green Book does not address sight distance on spirals. The 3-D B-spline surface approach allows for the calculation of the sight distance at any point regardless of the type of the curve. In this manner, the 3-D representation allows for the development of universal equations and algorithms that could be used for any type of vertical and horizontal alignment components. Moreover, since the 3-D nature of the road surface is incorporated in this methodology, there is no need of a separate consideration of the horizontal alignment, the vertical alignment and the cross-section design in order to calculate the geometric parameters; an approach necessary in the current conventional models. Finally, the geometric calculations are anticipated to be more reliable since they apply to a 3-D surface, the drainage paths are modeled as geodesic curves, and the calculations can be applied to any type of segment with no restrictions whatsoever.

The fact that the 3DDRS methodology utilizes the concept of geodesic curves, which can only be defined in 3-D surfaces, as it is introduced in Einstein's theory of general relativity makes this research highly unique and pioneering in the field of highway design. In addition, the introduction of the idea of "surface patches", an idea which will also be implemented in this dissertation as discussed later in this chapter, has already been successfully applied in Amiridis and Psarianos (2015b) in order to assess in the hydroplaning speed calculations. The computational part of this approach is attained by expressing the functions in code through the programming language offered by the software Mathematica (Wolfram Research 2018), whereas the whole process is demonstrated as a case study. The same 3-D modeling approach was also utilized in Amiridis and Psarianos (2016) that allowed the calculation of the available sight distance of a roadway directly in 3-D space and in a fully automated and accurate manner. The presented methodology overcomes the conventional 2-D approach of studying the actual 3-D roadway configuration separately and sequentially in its horizontal and vertical alignments. The road surface is simulated as a 3-D B-spline surface with continuous side barriers whose road centerline has been in turn, modeled as a 3-D B-spline curve. Through this approach, the road centerline as well as the right and left edge lines which play a crucial role in the sight distance calculation are mathematically defined both geometrically and analytically through explicit interpolation equations. The sight distance calculation can be made at each point of the road surface because of the integrated information existing in the model through its 3-D character. These calculations are theoretically more reliable, since they are applied on a realistically modeled 3-D roadway surface. It is worth mentioning that this methodology can be applied on existing roadways as well as during the design process by modifying the pseudo-geodesic and/or pseudo-normal curvature of the roadway in order to obtain the desired available 3-D sight distance. Finally, the introduction of the idea of the Point-In-Polygon (PIP) algorithm, an idea which will also be implemented in this dissertation as discussed in this section in order to link the crashes to their respective locations, has already been successfully applied in Amiridis and Psarianos (2016) in order to assess in the 3-D sight distance calculations. A significant extension of this work is found at Amiridis et al. (2016) in which a generic mathematical methodology was developed in order to account the effect of the cross slopes, i.e., cuts and fills, when calculating the 3-D available sight distance.

3.2 General Approach

The first step in the process is to obtain the XYZ coordinates of the roadway centerline and right and left edge lines. In fact, it is relatively easy and straightforward to model a roadway surface nowadays in an accurate manner with all the available technology such as laser scanners, Light Detection and Ranging (LiDAR) technology, and Unmanned Aerial Vehicles (UAV). The second step in this process will be the modeling of the road surface as a 3-D mathematical surface, i.e., 3-D B-spline surface, through an interpolation utilizing the XYZ coordinates of the roadway centerline and respective edge lines. An interpolation surface, just as an interpolation curve, is simply a surface that is forced to pass through specific predefined points, i.e., interpolation points. There are a number of ways, e.g. polynomials, splines etc., in which the interpolation points can be connected with each other. Splines are piecewise polynomials that have significant properties in the points of intersection; for example, the first and second derivatives are equal. The interpolation method that will be utilized in this research is indeed spline interpolation and more specifically B-spline cubic interpolation because of its robust properties. A more detailed discussion regarding B-splines, their properties, and the reason for selecting this specific class of splines is included later in this chapter.

The second step defines the points that will function as interpolation points and these correspond to points on the road centerline and the respective points on the right and left edge lines of the roadway surface for each predefined cross section. At this stage, the road surface will be accurately modeled as a 3-D B-spline surface through an interpolation process which will be applied in the Mathematica platform (Wolfram Research 2018). This 3-D B-spline roadway surface will be further divided into “3-D surface patches”, which are smaller sections of the roadway surface and which be eventually used as the basis of the statistical analysis. In differential geometry, a surface patch is simply a portion of a surface. In this research, various lengths of 3-D surface patches will be examined in order to determine the most appropriate patch length for crash prediction. The 3-D patches that will be created are essentially 3-D quadrilateral surfaces whose two opposite sides are the left and right roadway edge lines and whose two other opposite sides, which are perpendicular to former left and right edge line pair, correspond to two cross sections of the roadway surface whose distance is exactly the constant length of the patch under study. Once the 3-D surface patches are created, crashes can be linked to each patch according to the coordinates of each crash and the PIP computational geometry algorithm, as mentioned above.

The third step involves the collection of the crashes occurred along the roadway to be studied. For this thesis, a 14-year period was utilized consisting of data from 2004 to 2017. Depending on the length of the period, one needs to ensure that there were no interventions or changes along the roadway during the study period to confirm that the roadway geometry remained unchanged throughout the study period. The crash data for the state of Kentucky can be easily downloaded from the website of the Kentucky State Police (Kentucky State Police n.d), which also has very useful filters to personalize the search. Here, the geographic link of the crashes with the specific road segments will be conducted with the software ArcGIS (ESRI 2017), which is a GIS (Geographic Information System) platform.

The fourth step is to define the Annual Average Daily Traffic (AADT) for the roadway for each year, since the AADT will be incorporated in the regression models as an explanatory variable. The AADT should essentially be viewed as an exposure metric since changes in AADT could have an impact on crashes. The inclusion of the AADT increases the predictive power of the models and agrees with current safety prediction practices, i.e., HSM. The AADT data were retrieved from the website of the Kentucky Transportation Cabinet (KYTC n.d.).

The fifth step involves the statistical analysis and evaluation of the potential to predict crashes utilizing the metrics from the 3-D models developed. The explanatory variables must be initially defined. This research is based on the hypothesis that the differential element that will have the most crucial effect on roadway safety will be a 3-D geometric metric that is called Gaussian curvature. Therefore, the statistical model to be developed will be, at least initially, based on the Gaussian curvature itself, which is considered the cornerstone of the study of surfaces. It is so powerful that the field of Geometry as a whole, either Euclidean or non-Euclidean spaces, can be classified according to the sign of the Gaussian curvature (Gray 1998; Lipschutz 1981). For example, the properties of Euclidean

Geometry hold true only in spaces where the Gaussian curvature is zero; the properties of Spherical/Elliptic Geometry govern spaces with positive Gaussian curvature; and the properties of Hyperbolic Geometry apply to spaces with negative Gaussian curvature. This observation reinforces the need of the inclusion of the Gaussian curvature in highway geometric design and, more generally, integrate a holistic differential geometry approach in highway design.

In addition to the Gaussian curvature, there are several other differential elements such as 3-D curvature, torsion, geodesic curvature, normal curvature, and mean curvature that contain rich geometric information and whose potential of functioning as explanatory variables could be examined. All of these metrics can be easily calculated because of the 3-D realization of the roadway surface. There are automated statistical methods such as "forward", "backward", and "stepwise" selection procedures that can assist in this selection procedure according to a specific criterion, e.g. minimization of p-value, AIC (Akaike's Information Criterion), etc. In this case, the AIC will most likely be used because it also indicates the most parsimonious model, meaning that the models can be compared in an objective manner. However, these automated procedures will not be applied blindly; they will only be applied in order to identify which variables seem to be the critical ones in order to prioritize the search of the best model, a procedure that will mainly be conducted manually through a trial and error procedure. Therefore, the final model is intended to include the most appropriate variables by entering and removing variables in a systematic way until it can be reasonably argued that the final model is likely to be best for the given data. As far as the overall statistical model is concerned, the response variable will be the number of crashes and therefore the Negative Binomial Regression is preliminary considered the most appropriate type of regression analysis to be utilized.

Once the suggested statistical model is finalized, the final step is to compare the findings with current crash prediction practices and more specifically with the Interactive Highway Safety Design Model (IHSDM) developed by the Federal Highway Administration (IHSDM n.d.). The IHSDM can provide a crash prediction for roadway segments based on the current HSM procedures and thus could be used to evaluate and directly compare the predictive power of the proposed model with the HSM. This comparison will actually evaluate the findings of this research and indicate whether future research would be pertinent on this specific subject. The purpose of this thesis is to try to enhance the geometric features that come into play in the HSM and not to question it. At this point it should be mentioned that in contrast to the ISHDM model, which requires the data in the conventional 2-D approach, i.e., horizontal and vertical alignment, a procedure very tedious and subjective in nature, the proposed methodology derives all its required geometric data in an absolutely automated and objective manner; this fact by itself demonstrates the power of this methodology and the 3-D approach in general, but this will be further discussed in more detail in the following sections.

After the regression models have been finalized, the predictive power of the models will be examined. To achieve this, a year of the crashes will be used as the training data set. There are 14 years of data in total and therefore this implies that 14 3-D models will be developed each using a 13-year database and the 14th year will be used for the evaluation of the process in each case. A final report will summarize the findings and will actually serve as the basic evaluation criterion regarding the level of contribution of the dissertation to current research results and practices. The final task of the dissertation will focus on the evaluation of the proposed approach by comparing the crash predictions produced from the suggested model with the ones produced from the IHSDM. In addition, the benefit in crash cost prediction will also be calculated in monetary values. . In both cases, i.e., suggested model and IHSDM, the exact same historical crashes are integrated in the prediction process.

3.3 Geometric Data

The data needs for completing this research include roadway segments and their associated crashes. For the roadway segments, the required data are in fact available and have been acquired from the KYTC. At this point, data from three road segments were utilized: KY 420, KY 152, and US 68. This data is available through “mobile mapping technology” through a system that was placed on a vehicle for roadway-based collection. The data that are utilized essentially consist of the GPS data of the vehicle, i.e. latitude and longitude coordinates and altitude, and the superelevation rate data through an inertial navigation system. The GPS data are collected every 5 ft along the roadway path trajectory. Also, it is worth mentioning that this particular technology is vastly used nowadays for a number of applications such as LiDAR point cloud collection and total asset management solution purposes, e.g. bridge and utility record registration.

The acquired data for each segment will have to be manipulated in order to allow for the development of the 3-D models. The Cartesian coordinates, i.e., X, Y, and Z, of the centerline and right and left edge lines must be calculated. The coordinates of the centerline are in the form of geographic coordinates, i.e., latitude and longitude and therefore an appropriate geographic transformation must be applied, which is easily conducted in ArcGIS in an automatic manner. As far as the Cartesian coordinates of the right and left edge lines are concerned, they will be implicitly calculated from the given data: the altitudes of the roadway centerline, as well as the lane width and superelevation rate are calculated along all measured. Therefore, the Cartesian coordinates of the right and left edge lines can be ultimately calculated through a 3-D geometric transformation based on the Cartesian coordinates of the centerline, which have been calculated in the previous step and will function as reference points.

3.4 Crash Data

The crash data that the Kentucky State Police collects, offers numerous categories and filters based on which one can customize the search. It is intended not only to obtain crashes according to categories that are considered critical in addressing the geometric effects on crash occurrence, but also remove potential bias in crash occurrence such as driving under the influence of alcohol, fatigue, distraction from cell phone etc. In addition, the filters intend to separate conditions that although are not directly related to the roadway geometry, e.g. clear weather vs. rain, passenger car vs truck etc., most likely affect the probability of crash.

However, in this dissertation, all crashes will be considered. This decision is based on the fact that the sample size is limited and that the purpose of the statistical model is to prove the correlation between crashes and 3-D geometric metrics and not to suggest an absolutely accurate predictive model. This means that if the model proves the statement above for these datasets, then one could anticipate that accuracy will be improved even more with a more refined crash dataset. Moreover, the HSM software predicts total crashes which means that all categories are included; therefore, the comparison with the results from the HSM software can be conducted in a much more straightforward, pertinent and fair manner.

3.5 AADT Data

One variable that is essential to crash frequency is the AADT, which denotes the average number of vehicles along a roadway segment per day. It is absolutely pertinent to include the AADT as an explanatory variable in the statistical models to be developed since it functions as an exposure factor in crash occurrence and it is also included in the HSM models. Data is collected from traffic counters that the state places on a periodic basis to estimate AADT. These data can then be extrapolated to determine the average AADT to be used in the crash modeling.

3.6 3-D B-Spline Surfaces

There is more than one way to mathematically define a surface; however, here all curves and surfaces are defined according to the "parametric-vector" approach, which is based on vector theory. The essential advantage of the "parametric-vector" approach is simply that it defines curves and surfaces via a vector representation in a rather flexible way. Other approaches require equations to be solved at each point that is intended to be modeled, which drastically reduces the computational speed, and in many cases the system of equations may be unsolvable especially when these equations are in an implicit form (Gray 1998). These equations might even be differential in nature making the production of results practically unfeasible. Moreover, the parameters of the vector representation form usually have a physical meaning and not simply some "dry" mathematical variables. For example, the parameters of the parametric-vector form of the sphere or an ellipsoid are their longitude and latitude angles. The theory behind this approach is not essential in order to understand the procedure, but what is important to understand is the geometric interpretation of the theory as it applies in modeling a roadway surface.

The “parametric-vector” equation of a curve, either 2-D or 3-D, is defined by one variable: *t*. For example, the “parametric-vector” equation of the circular helix, a random 3-D curve, of radius a and slant b is expressed in Equation 2. Equivalently, the “parametric-vector” equation of a surface is expressed by two variables, e.g. u and *v*, defining a whole family of curves. The curves defined by the *u* variable family are called *u*-parametric curves, whereas the curves defined by the *v* variable family are called *v*-parametric curves. For example, the “parametric-vector” equation of the circular helicoid, a random 3-D surface, of radius a and slant b is expressed in Equation 3. Equations 2 and 3 are examples of the approach to be used and are provided to lay the groundwork for a more thorough understanding of the methodology to be used.

$$\vec{h}(t) = (a\ cost, a\ sint, b\ t) \tag{2}$$

$$\vec{h}(u, v) = (a\ v\ cosu, a\ v\ sinu, b\ t) \tag{3}$$

where a and b are constants.

These equations are translated into road surface terms in the following way. The 3-D B-spline surface that has been created is indeed accurately defined by two variables, i.e., *u* and *v*, since it is a 3-D surface. In the case of the roadway surface, one should imagine the *u*-parametric curves as the 3-D road centerline and all other curves parallel to it and imagine the *v*-parametric curves as lines perpendicular to the centerline. In other words, the *u*-parametric curves are the centerline, the right edge line, the left edge line, and all the other parallel curves between them, whereas the *v*-parametric curves are in fact the cross sections of the road surface.

3.6.1 Road Surface

To enable the modeling of the surface of the road as an integrated mathematical surface, an interpolation B-spline surface must be applied. The required interpolation points are the roadway centerline and right and left edge lines of the roadway surface.

To carry out this task, the coordinates corresponding to the left edge line are calculated initially. Afterwards, the coordinates that are calculated are those that correspond to the *u*-parametric curve translated to the right by a fixed number in relation to the left edge line. This process is repeated until the *u* parametric curve corresponds to the right edge line. The step of the discretization of the u-parametric curves is advisable to be applied in such a way in order for the *u* parametric curves to pass successively from the left edge line, the roadway centerline and the right edge line. Eventually, these 3-D coordinates correspond to the control points of the 3-D B-spline surface. In this way, the 3-D XYZ coordinates of each point of the surface of the road can be calculated as a function of the curvilinear coordinates *u* and *v*.

The superelevation of the roadway is calculated through a geometric transformation utilizing the geometric data that are available from the data collection process. Once the superelevation function of the road is calculated, the road surface can easily be modeled through Mathematica (Wolfram Research 2018). Specifically, the road surface is rotated around the road centerline by an angle which essentially corresponds to the superelevation rate at each point along the roadway centerline. This process for the 3-D B-spline surface creation it has been previously validated both visually and numerically (Amiridis and Psarianos 2015a).

In order for the methodology to be usable and therefore effective and applicable it must be implemented in a computational system. All the commands (e.g. mathematical functions, geometric restrictions, and data) should be written in a programming language so that the methodology is applied in a fully automated manner. This has been achieved in this case through the software Mathematica in which given the XYZ data of the roadway centerline and right and left edge lines at a minimum, the required 3-D surface can be automatically created.

3.6.2 3D Surface Patches and Crash Allocation

After the 3-D B-spline is defined, it must be divided into sub-surfaces which, in line with differential geometry terminology, are called surface "patches". Patches are simply curvilinear polygons on a 3-D surface, but in this case these curvilinear polygons are more explicitly and specifically defined as curvilinear rectangles because the *u*- and *v*-parametric curves will be defined in such a way that they will be perpendicular to each other. The division of the road surface into patches has a basic advantage in terms of creating unique small areas of the roadway surface where each crash could be allocated and thus allow for the correlation of the parameters in the model estimation. In addition, the areas where zero crashes have occurred will also be analyzed, i.e., the crash frequency will not be inflated. The latter rationale is in fact the basic logic that spatial statistics are based upon. As a final note, the number of patches will directly define the sample size of the dataset through a one-to-one relationship. An example of surface patches is presented in Figure 3.1.

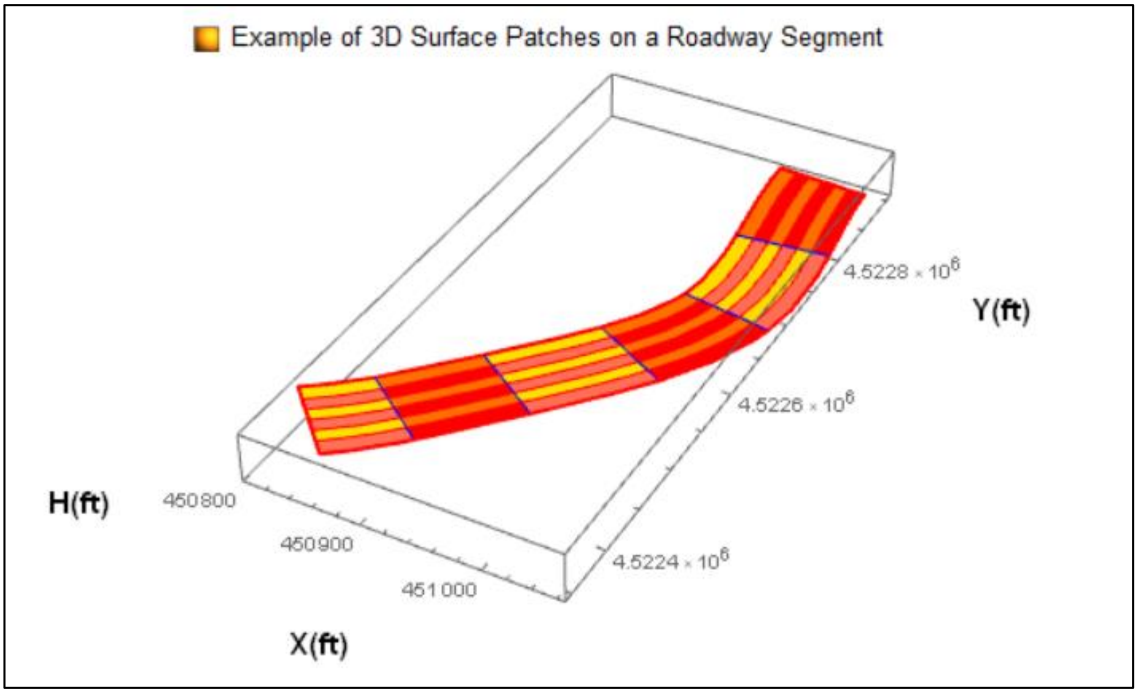

Figure 3.1: Example of 3-D Surface Patches on a Roadway Segment

The 3-D geometric metrics for each patch will be calculated as the arithmetic average of the respective 3-D metrics of nine points: the four vertices which define each particular patch, the four median points between the four patch vertices, and the point in the center of the patch. For example, the Gaussian curvature of a patch will be calculated as the average Gaussian curvatures corresponding to these points. However, for each crash in addition to considering the Gaussian curvature, or any other 3-D metric that will be examined, of the specific patch where the crash occurred, the values of the surrounding patches will also be considered.

When a crash occurs, it can be reasonably argued that not only does the specific geographic point where the crash occurred matter, but the preceding and following geometry of the roadway could also have an effect. It would be therefore somewhat simplistic not to consider the geometric characteristics of the roadway before and after the crash occurrence location. For this reason, the 3-D metrics of the surrounding patches will also be considered as contributing in the final calculation of the 3-D metrics of the each patch with associated weights. The patch where the crash occurred will be called the “principal patch”. A question that arises is how the 3-D metrics will be finally calculated. The weight allocation for the final averaged 3-D metrics will be made as follows: the principal patch will have a weight of 2, whereas the patches exactly before and after the “principal patch” will have a weight of 1. Therefore, the 3-D geometric metrics corresponding to each patch will be calculated based on a weighted average of the principal and surrounding, i.e., preceding and following, patches.

At this point, all patches have a value for each 3-D metric that will be examined in the statistical analysis, i.e. Gaussian curvature, mean curvature, 3-D curvature, torsion, geodesic curvature, normal curvature, etc. The next step will allocate the crashes to their corresponding patches. Initially, the coordinates of the crashes, which are downloaded from the Kentucky State Police website in the form of geographic coordinates, are converted to Cartesian coordinates. The problem of checking whether a point lies inside a polygon is a classical problem in the field of computational geometry and is called the Point-In-Polygon (PIP) problem. The solution to this is achieved with the use of the Jordan

Curve Theorem. The algorithm has been developed and applied in Mathematica in previous research (Amiridis and Psarianos 2016) and it is readily available. By applying the PIP algorithm, all crashes will be linked to their respective patch and after the crashes have been allocated accordingly, the sum of crashes for each patch will be calculated. The sum of crashes for each patch will finally be the response variable in the statistical model development.

3.7 Model Development

The model development lies in the core of this research since its objective is to suggest robust statistical models that can predict the number of crashes in a given roadway segment with specific geometric characteristics using 3-D model metrics. The response variable will be the total number of crashes in each patch. The patch length will affect the sample size for the analysis and a thorough analysis will be undertaken to determine the optimum length for prediction. A count regression model will be developed since the response variable, i.e. crashes, is such. As noted in the previous chapter, the most common count models used in road safety research are the Poisson and Negative Binomial regression models. In crash data there is, almost always, an issue of over-dispersion: the variance is statistically larger than the mean. The issue of over-dispersion often appears when there is a large number of zeros in the dataset, a case that is prevalent when dealing with crash data. The Negative Binomial distribution can account the over-dispersion in a dataset since it has an additional parameter, i.e., dispersion parameter, which is used to model the variance. The Negative Binomial regression can be viewed as an alternative strategy of modeling over-dispersed data that follow a Poisson distribution. The latter observation means that a very important underlying assumption regarding the errors, produced from the Negative Binomial regression, is that they must follow the Poisson distribution. Alternatively, the conditional distribution of the response variable must follow the Poisson distribution, but this assumption is practically impossible to be tested since the probability distributions of the explanatory variables are unknown. In this case, the error/residual approach is the most pertinent method, since the effect of the explanatory variables is "filtered out", in order to assess one of the most essential underlying assumptions of the Negative Binomial Regression model. Although there are a number of other approaches to deal with over-dispersed data, such as the quasi Poisson and Gamma regression models, the Negative Binomial regression will most likely be the final type of regression to be used since it is the most commonly used and accepted model for crash data analysis in the road safety research community as previously noted. In fact, the assessment of the underlying assumptions of the Negative Binomial Regression model, which is an essential task in any statistical analysis, has been extensively studied in prior research aiming to predict the number of crashes at signalized intersections with permitted, protected, and permitted/protected left-turn phasing schemes (Amiridis 2016; Amiridis et al. 2017a).

There are several 3-D metrics that could be tested as explanatory variables such as:

1. Gaussian curvature
2. Mean curvature
3. Length of geodesic curves
4. Metrics of the First Fundamental Form
5. Metrics of the Second Fundamental Form
6. 3-D curvature
7. Torsion
8. Geodesic curvature
9. Normal curvature
10. Pseudo-geodesic curvature
11. Pseudo-normal curvature
12. Darboux vector metric, i.e., vector of angular velocity

These variables are well known metrics and thoroughly described in any typical differential geometry textbook. However, the interesting aspect here is to emphasize their equivalence to roadway geometric elements such as the radius, grade, superelevation etc. Two examples of the analogy/equivalence between 2-D and 3-D metrics have already been discussed above regarding the 2-D analogy of pseudo-geodesic and pseudo-normal curvature (Amiridis and Psarianos 2015a). In addition, some of these 3-D metrics have the potential to allow for incorporating more than one geometric traditional element and this will be explored and presented in this research. For example, the Gaussian curvature can capture the three-way interaction of the horizontal, vertical alignment, and the superelevation in one single value.

One of the goals of this research is to develop models that are surface-based, i.e., not only curve -based. The analysis here is not intended to simply analyze 3-D metrics, but rather to analyze 3-D surface oriented metrics, which can be calculated only on a 3-D surface, and then be used as crash predictors. In this way, all geometric combinations that may be produced by the interaction among the horizontal alignment, vertical alignment, and cross slopes including superelevation rate, will be implicitly expressed by single variables. Considering this concept, the potential 3-D surface variables that could be analyzed are the following:

1. Gaussian curvature
2. Mean curvature
3. Metrics of the First Fundamental Form
4. Metrics of the Second Fundamental Form
5. Geodesic curvature
6. Normal curvature

It is noted that the geodesic and normal curvatures can resemble, in a not strictly accurate but adequate enough manner, the horizontal and vertical curvatures, respectively. However, the intent with this research is to move away from anything that may be linked to the 2-D conventional approach and investigate metrics in which more rich geometric information is hidden. Therefore, the last two variables will not be considered, since they are essentially 2-D variables that are calculated planes that are tangent on 3-D surfaces. In addition, the Gaussian and Mean Curvatures are explicitly defined by the metrics of the First and Second Fundamental Forms where the effect of the metrics of the First and Second Fundamental Forms is “nested” into the Gaussian and Mean Curvature (Lipschutz 1981). Therefore, the potential explanatory variables that will be used are the Gaussian Curvature and the Mean Curvature. An overview, i.e., definition and properties, of the

Gaussian and Mean curvature is briefly presented, without the inclusion of strict mathematical proofs, in Appendix A.

The entire analysis is "patch-based" and therefore a crucial task is the determination of the patch length. In order to address this question, six patch lengths will be tested: 1,500 ft, 1,000 ft, 400ft, 200ft, 100 ft, and 50 ft. Each statistical analysis will be conducted six times and the criterion according to which the "optimal" patch length will be selected is the predictive power of each model. The closer the predicted crashes are to the actual/observed crashes, the higher the predictive power of the model The comparison between predicted and observed crashes is achieved by developing models using 13 years of the crash data and then comparing their prediction to the crashes of the 14^{th} remainder year. However, the final proposed model and the final explanatory variables that will enter the model will be based on all 14 years. As a final note, the scale of the covariates, i.e., whether the explanatory variable should be raised to a power or an exponential transformation would be more pertinent, will be determined according to their respective "grouped" frequency tables. The rational of this approach is discussed in the next chapter in more detail.

3.8 Comparison to Current Safety Estimations

The final task of the dissertation is to compare the suggested prediction models with the crash prediction results that would be produced from the existing models presented in the HSM (AASHTO 2010). In fact, the results will be compared to the predicted crashes from the IHSDM software, which can also utilize the historical crashes that are available from 2004-2017. The comparison will be conducted with training data: for each roadway, one year at a time will be kept outside of the analysis and then the crashes for that year, whose actual/observed crashes are known, will be predicted based on the data of the remaining years. For example, for a given roadway segment, in order to evaluate the predictive accuracy for 2017, the crashes of this year will be removed from the model development that will utilize only the crashes in the 2004-2016 period. Since there are seven roadway segments and 14 years of available crash data, 98 prediction evaluations will be essentially developed. Finally, this procedure allows for comparing the predictive power and accuracy of the proposed modelling approach with the current practices as described in the HSM.

3.9 Approach Summary

This section summarizes the basic methodology and required steps to be followed throughout this research process. Initially, each roadway will be modeled as a 3-D B-spline surface and it will be divided into patches accordingly representing a smaller roadway segment. After the patches have been determined, all corresponding crashes will be selected based on the relevant filters to be used and downloaded from the Kentucky State Police Collision Wizard. Each crash will be allocated to its appropriate patch based on its geographic coordinates. Next, the model development will take place aiming to develop predictive negative binomial crash models in which the number of crashes will be the response variable and 3-D differential geometry metrics will function as the explanatory variables. Once statistical models have been developed, their results will be statistically compared, through real crash data, to those derived from the existing guidelines in the HSM in order to evaluate the reliability, usefulness, and predictive power of the suggested statistical models in this research.

4 CRASH PREDICTION MODEL

This chapter presents the development of the models for crash predictions utilizing the 3-D B-spline roadway surface. To achieve this, three roadway segments are utilized. The following sections present the data used and the analysis undertaken to develop the 3-D SPF models.

4.1 Model Data

4.1.1 Geometric Data

The ArcGIS platform will be used to display the three roadway segments used in order to determine the rural sections for the analysis. The urban (built-up) sections of the roadways selected were visually identified and excluded from the analysis. Specifically, an additional length of 400 ft was excluded from the beginning and end of each urban section in order to filter out the potential “urban effect”. Finally, a 400 ft radius was used to exclude major intersections to obtain only continuous roadway segments. The roadway segments are also shown in Google Earth as .kmz files in Appendix B.

An example of the process undertaken to develop the sections for study is shown here. Figure 4.1, shows the data for KY 420 indicating the rural and urban sections of the roadway. The urban and rural sections of all roadway segments under study are shown in Appendix C. Each of the roadway segments considered was evaluated to determine whether there have been any geometric alterations during the study period. The review identified that there were no changes in the alignments over the 2004-2017 period. This allows for an accurate comparison throughout the entire period of the study. If geometric changes were present, then separate 3-D roadway models would have to be developed in order to capture these changes.

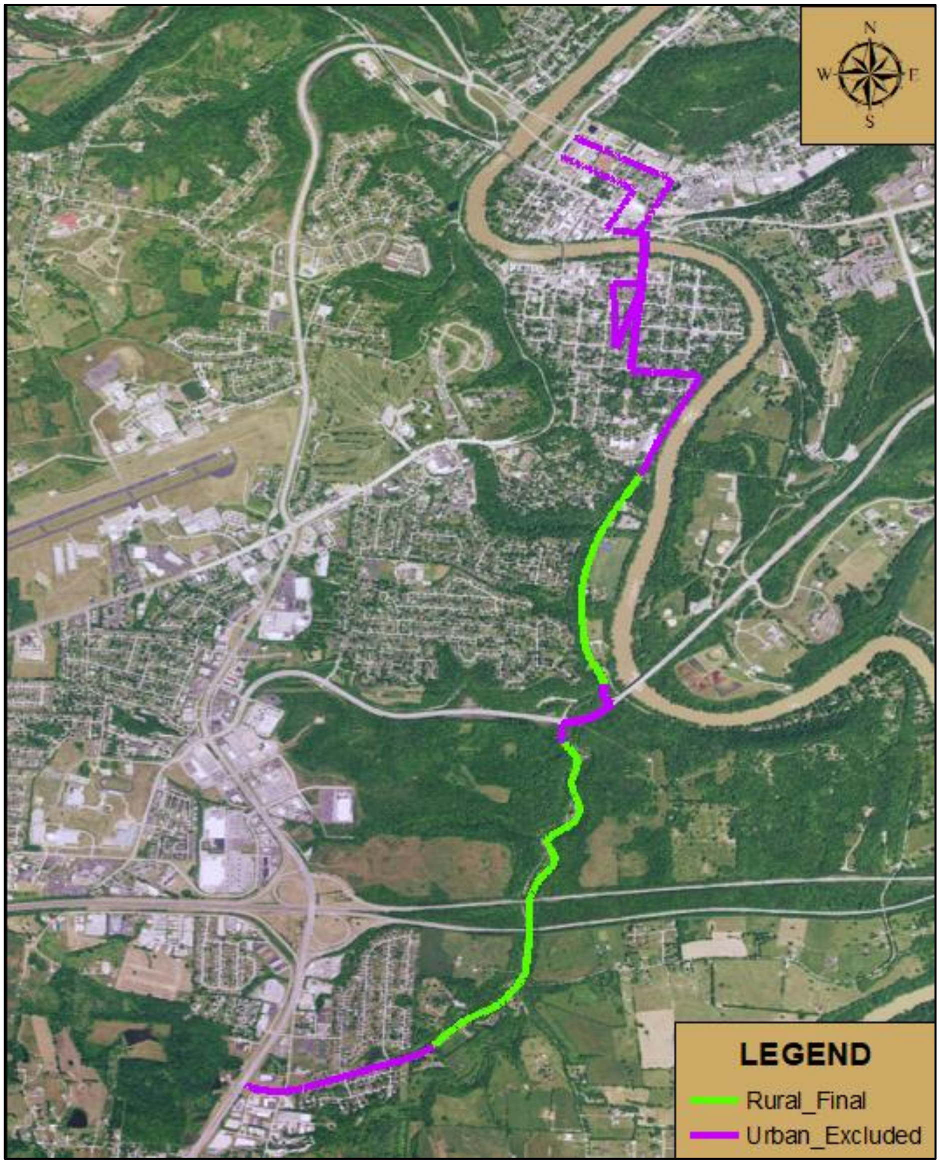

Figure 4.1: Example of Rural/Urban Distinction (KY 420)

The next step involved the development of the 3-D B-spline centerline of the first section (Figure 4.2) and the corresponding 3-D B-spline surface of the roadway segment (Figure 4.3) while considering the superelevation rate of the curves. This process resulted in seven separate roadway segments that were geometrically modeled and statistically analyzed. The 3-D representations of the roadway centerlines and surfaces, as modeled in the Mathematica platform, for all seven segments are also shown in Appendix C. The travel lane width is assumed to be 11 ft according to multiple measurements along the roadway segments, whereas the centerline lengths of these sections are shown in Table 4.1.

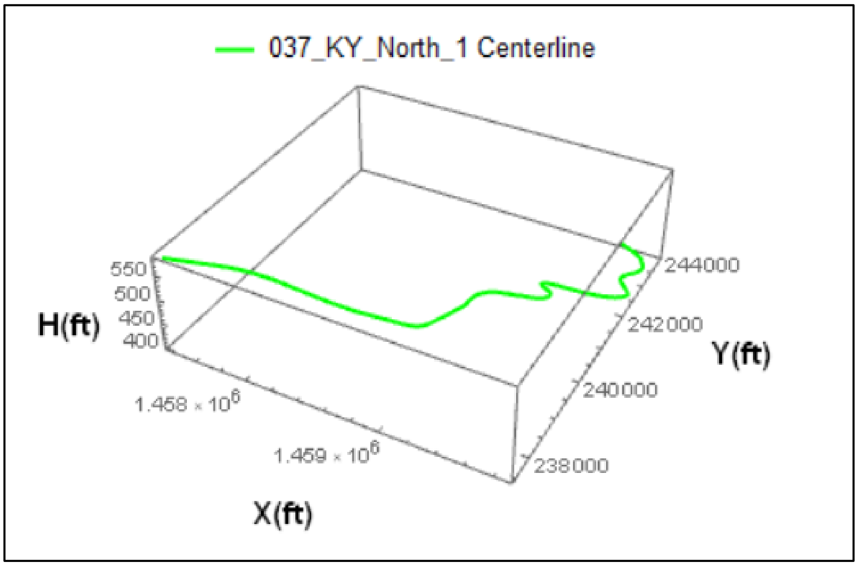

Figure 4.2: 3-D B-Spline Roadway Centerline (KY420-1)

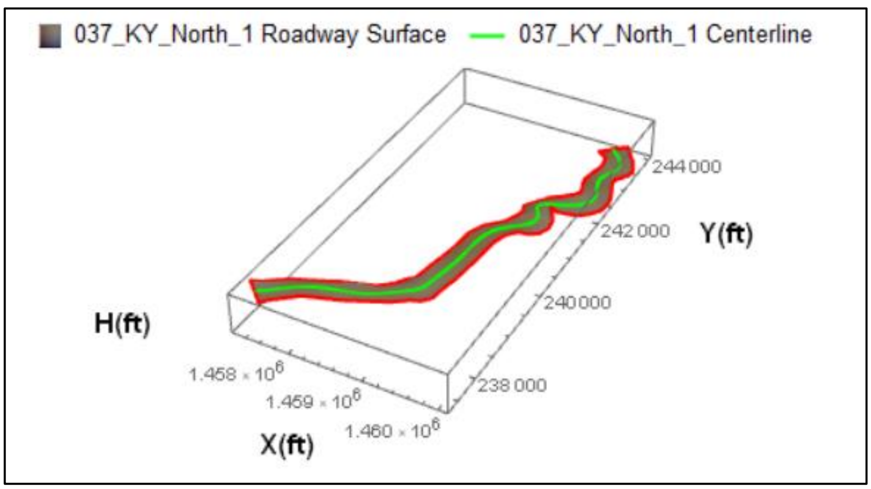

Figure 4.3: 3-D B-Spline Roadway Surface (KY420-1)

Table 4.1: Roadway Segment Lengths

Roadway	Roadway Segment	Length (miles)
KY 420	KY420-1	1.4119
	KY420-2	0.8783
KY 152	KY152-1	8.6922
	KY152-2	2.2392
	KY152-3	3.3862
US 68	US68-1	5.7450
	US68-2	11.5097
Total		**33.8625**

4.1.2 Crash Data

The crash data for each section for the 2004 to 2017 period were retrieved from the Kentucky State Police Wizard. The crashes were plotted with ArcGIS and an example is shown in Figure 4.4.

Figure 4.4: Example of Crash Data Plots (KY 420)

The crash plots for all the other segments are presented in Appendix D. Table 4.2 presents a summary of the crash data by type and other characteristics

Table 4.2: Crash Summary (Percentages)

	Roadway Segment						
	KY420-1	**KY420-2**	**KY152-1**	**KY152-2**	**KY152-3**	**US68-1**	**US68-2**
	Crash Type						
Single Vehicle	73	60	76	61	77	80	75
Rear End	10	8	5	18	2	3	4
Angle	6	16	4	5	2	6	5
Sideswipe Opp. Dir.	6	3	10	12	11	6	8
Head On	4	0	1	0	4	2	4
Sideswipe Same Dir.	1	11	3	4	2	2	2
Other	0	2	1	0	2	1	2
	Number of Motor Vehicles						
Single	73	60	76	56	77	80	74
Two	26	35	23	39	22	20	24
Multi	1	5	1	5	1	0	2
	Crash Severity						
PDO	77.0	83.8	74.5	68.2	67.0	74.7	75.7
Injury	22.7	13.5	22.2	31.8	30.9	24.2	24.0
Fatal	0.3	2.7	3.3	0.0	2.1	1.1	0.3
	Roadway Alignment						
Curve & Grade	46	11	32	23	31	50	45
Curve & Hill Crest	6	0	4	14	6	5	5
Curve & Level	35	14	20	5	33	26	20
Straight & Grade	10	5	20	25	14	8	10
Straight & Hill Crest	1	0	6	25	3	0	3
Straight & Level	2	70	19	8	13	11	17

The data in Table 4.2 indicate that the majority of crashes are “Single Vehicle”. This fact is advantageous for the intended analysis and overall research because this specific crash type is mostly related to the geometric characteristics of a roadway. Table 4.2 also shows that approximately 85 percent of the crashes were related to some combination of the horizontal and vertical alignment. This verifies once again the need, as emphasized also in previous research, of investigating the effects of horizontal and vertical coordination in a more systematic manner to address safety concerns. This further supports the need for considering 3-D solutions and the value of this research proposal. Finally, in terms of severity level, it is worth mentioning that from the total crashes that occurred, 75 percent are property damage only, 24 percent resulted in some kind of injury, whereas only a small percentage (1 percent) resulted in fatalities.

4.1.3 AADT Data Needs

The AADT for each section was obtained from the KYTC. Initially, the corresponding AADT stations for each roadway had to be identified through the interactive map provided by KYTC and the starting longitude and latitude coordinates in order to retrieve the corresponding AADT values. The AADT values were linked to each segment separately; an example of the AADT data is shown in Table 4.3 for a specific station for KY 420. It should be noted that the AADT was not available for all years, i.e. 2004-2017 and that the missing data were estimated by applying piecewise polynomial cubic spline interpolations between known values. The AADT data that were retrieved from all associated stations and for all roadway segments are shown in Appendix E.

Table 4.3: AADT Data for Station ID# 037553; KY 420-1

Year	**AADT**
2004	5051
2005	**5220**
2006	5262
2007	5214
2008	**5110**
2009	4988
2010	4882
2011	**4830**
2012	4962
2013	5147
2014	**5351**
2015	5537
2016	5671
2017	**5716**

Note: Bold numbers are actual AADT counts

In Table 4.4, the AADT stations for each roadway segment are shown.

Table 4.4: Corresponding AADT Stations to Each Roadway Segment

Roadway	Segment	Station ID	Route	Begin MP	End MP
KY 420	1	037553	037_KY_0420_000	0.6	2.145
	2	037A20	037_KY_0420_000	2.145	3.753
KY 152	1	084507	084_KY_0152_000	0	0.232
		084507	084_KY_0152_000	0.232	1.961
		084570	084_KY_0152_000	1.961	6.044
		084569	084_KY_0152_000	6.044	8.605
	2	084A45	084_KY_0152_000	10.976	14.419
	3	084252	084_KY_0152_000	14.419	17.116
US 68	1	084505	084_US_0068_000	0	3.927
		084556	084_US_0068_000	3.927	5.517
	2	084A50	084_US_0068_000	8.391	10.214
		084256	084_US_0068_000	10.214	14.452
		084001	084_US_0068_000	14.452	20.058

As it can be observed from Table 4.4, multiple AADT stations are associated to each roadway segment. The way in which the AADT values were linked to each patch was based on the respective Begin Mile Point (BMP) and End Mile Point (EMP) of each AADT station. For example, two AADT stations are associated with roadway KY 420, i.e., 037553 and 037A20. The BMP and EMP of station 037553 are 0.6 and 2.145 respectively. Therefore, the AADT values from station 037553 correspond to the beginning of the roadway KY 420 until the length of 8,158 ft (2.145-0.6). Now, by definition, the parameter t of any B-spline curve "runs" from 0 to 1. Therefore, in this case, the "breakpoint" of 8,158 ft is converted in terms of the parameter *t*, e.g. 0.42. In addition, the beginning and end points of all the patches of the roadway are known in terms of the parameter *t*. Finally, the midpoint of each patch, i.e., average of the beginning and end *t* variable, is compared to the breakpoint of each AADT station, e.g. 0.42. If the midpoint of the patch is less than 0.42 then the AADT value of the patch is the one retrieved from Sta 037553; if the midpoint of the patch is greater than 0.42 then the AADT value of the patch is the one retrieved from Sta 037A20. This rationale is applied for all patches and AADT stations and this logic describes the way in which all patches are linked to a specific AADT value for each year.

4.2 Statistical Analysis

This section describes the way in which the final crash prediction models are developed. As noted above, six patch lengths, i.e. 1,500 ft, 1,000 ft, 400 ft, 200 ft, 100 ft, and 50 ft were considered. The length of the patch length denoted the way in which the 3-D roadway surface is spatially divided. However, no matter the patch length, in each patch there are four explanatory variables linked to it, namely: Number of crashes that occurred, AADT value, Gaussian Curvature (GC), and Mean Curvature (MC). In addition to the initial values of these variables, transformations were also considered, e.g. $AADT^2$, GC^2 MC^3 in order to identify the optimal scale and combination of these variables. All of these transformations and the justification of the optimal scale are presented in Appendix F. Moreover, statistical interactions of the explanatory variables, e.g. Gaussian*Mean, are also considered. Finally, it should be noted here that the statistical regression model that was utilized is the Negative Binomial Regression because over-dispersion is present in the data and because it was intended to keep the statistics rather simple in order to retain the focus of the thesis on the use of 3-D geometric explanatory variables in highway safety rather than the statistical methods utilized per se. After all, the typical regression model that is utilized for crash prediction modeling is indeed the Negative Binomial Regression.

The analysis to be conducted will serve a dual purpose: 1) demonstrate the proof-of-concept of the proposed 3-D approach; and 2) evaluate the predictive power of the model. These two objectives can be viewed as independent, i.e., failure in demonstrating predictive power of the model does not mean that the proof-of-concept is violated. For example, 3-D metrics may be proven to have a statistically significant effect in crash modeling, but the reason for potential failure in adequately predicting actual crashes may simply rest on the fact that more explanatory variables are required in the model. The proof-of-concept relies on the verification that the 3-D differential geometry metrics of Gaussian and Mean curvature are statistically significant crash predictors. This can be successfully demonstrated if it is proven that the coefficients of the metrics are indeed statistically significant. It should be also noted that depending on whether historical crashes are available, two strategies come into play in order to predict crashes in the most effective way.

4.2.1 Proof-of-Concept

In order to provide the proof-of-concept in the most concrete way, all years, i.e. 2004-2017, and all seven roadways entered the same model which will be called “Integrated Model” (IM) to be distinguished from the models that will be developed for the second objective, i.e., prediction evaluation. The objective of this effort was to establish that it is meaningful to incorporate 3-D highway geometry in crash prediction models. Although the predictive power of the model was not evaluated at this point, this step was crucial because failure to address the statistical significance of the 3-D metrics in crash prediction, would yield any further discussion of prediction power evaluation meaningless. Moreover, the type of the final explanatory variables that will enter this model will function as the basis of the predictive power evaluation of the model. For example, if the variables, AADT, Gaussian[2], and Mean[3] are proven to be the finalists, then these exact variables would be considered in order to evaluate the predictive power of the model; a logic that holds true in most predictive models. For example, even for the variables than come into play in the SPFs in the HSM with a specific transformation, e.g., exp(AADT), does not mean that this particular transformation is the optimal in all cases; it simply means that this transformation is on average adequate.

Although not explicitly stated, a part of the statistical analysis essentially touches the field of spatial statistics since the selection of an acceptable patch length is of utmost importance because it functions as the basis of all further (traditional) statistical analysis. To proceed with this effort, a two-stage simultaneous testing was undertaken that would define the optimal patch length and model to be used. First, for each patch length considered, models with the variables of interest were developed and the most appropriate was selected in terms of statistical significance and the Akaike Information Criterion (AIC) evaluation criterion. Second, these models were then compared to identify the most appropriate patch for analysis and power of prediction evaluation. Since there are six patch lengths tested, six “final models” will eventually be compared to each other.

All of the combinations of the explanatory variables that were utilized until the final model was decided, for all patch length combinations, are presented in Appendix G. The criterion according to which the models were compared was the AIC; the lower the AIC, the more informationally rich the model is. The AIC also functions as an adjusted R-square in the sense that it penalizes the number of variables that enter the model. Furthermore, in order for a model to be further considered as a finalist for additional evaluation, all of the coefficients of the explanatory variables that enter the model must be statistically significant, i.e., p-value<0.05. The demand of p-value<0.05 is associated with the fact that a significance level of 5% is considered; in fact, each p-value, depending on the number of explanatory variables that enter the model, must be less than the predefined "familywise" p-value, which in this case is set to 0.05, according to the Bonferroni, or any other type, correction (Myers et al. 2010). Roughly speaking, this means that if two explanatory variables are considered then the p-value of the coefficient of each variable must be less than approximately 0.025 ($= 1 - \sqrt{0.95}$) assuming that the two explanatory variables are independent in order for the "overall p-value" to be less than 0.05.

More generally, the Bonferroni correction, or any other type of correction, should be applied when explanatory variables are simultaneously inserted in a statistical model. More specifically, the significance level has been assumed to be 5 percent, i.e., there is a 95 percent confidence that the true parameters belong in the constructed confidence interval. However, the significance level of 5% should not be applied to each coefficient, but to the model as a whole; therefore, in order for the significance level of the whole model to be kept at the 5 percent significance level, the p-value of each coefficient should be less than 5 percent. The value of each p-value in order to achieve a "family-wise" error of 5 percent is imposed by the pertinent correction method used, e.g. Bonferroni, Tukey, and by the number of variables; the more variables, the stricter, i.e. less, the p-value must be.

It should be noted here that the AADT, GC and MC are to be used as explanatory variables, i.e. main effects, in the statistical analysis through a multivariate regression analysis. However, even when multiple explanatory variables are intended to enter the model, the analysis should always begin by visualizing the explanatory variables vs. the dependent variable. The Negative Binomial Regression, which is the regression type that will be applied here is a member of the family of Generalized Linear Models (GLMs). Each regression member of the GLMs is associated with a function that is called the "canonical link function", which actually represents the optimal transformation that should be applied to the dependent variable in order to satisfy desirable statistical properties such as unbiased parameters (Hardin and Hilbe 2012). In the case of the Poisson and Negative Binomial Regression, the aforementioned function is indeed the "log-link function". If the logarithmic transformation is not applied to the dependent variable, then the so-called "identity link function" is applied, meaning that the dependent variable is simply the variable "Crashes". Results will be produced even if the log-link function is not applied but the reliability of the results is weakened because the log-link function is the "canonical" link function for the Negative Binomial Regression. This is why the Poisson and Negative Binomial regression models are also often called log-linear models, meaning that the explanatory variables have a linear relationship with the logarithm of the dependent variable. Therefore, in this case the dependent variable will be LN(crashes).

The typical visualization process in order to identify the optimal transformation of each explanatory variables is via scatterplots. The scatterplots for the explanatory variables AADT, GC, and MC are shown in Figures 4.7-4.9 respectively for the 100 ft patch. Typical scatterplots are a great way to identify trends when the dependent variable is continuous, e.g. in the linear regression models. However, in this case the dependent variable, i.e. crashes, is a discrete positive variable; therefore, the creation of typical scatterplots does not produce valuable information. An example of a scatterplot when the dependent variable is discrete is shown in Figure 4.5; in the example, the dependent variable is whether a patient has coronary disease, whereas the explanatory variable is Age.

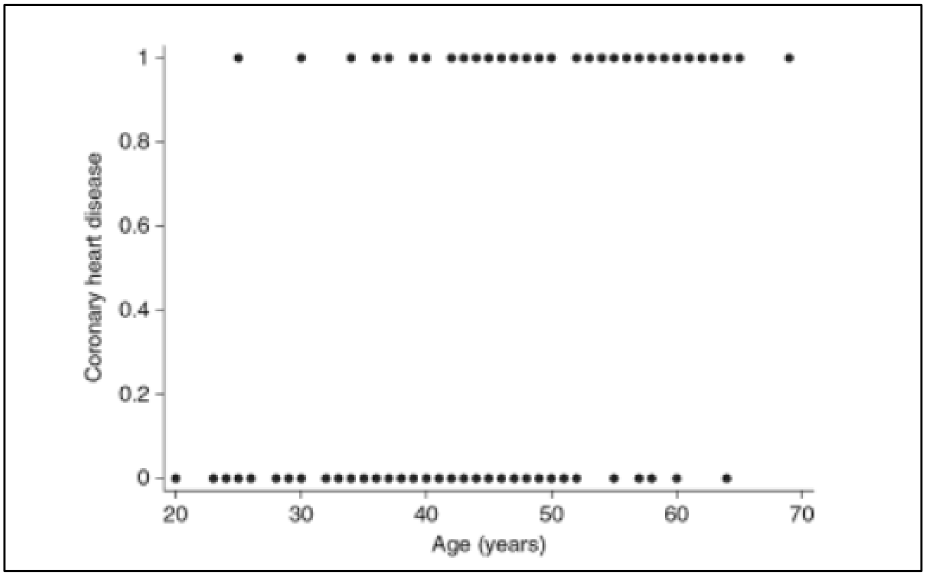

Figure 4.5: Scatterplot of Presence or Absence of Coronary Heart Disease (CHD) by Age for 100 Subjects
Source: (Hosmer et. al., 2013)

According to Figure 4.5, no trend can be conveyed between CHD and Age. This problem can be addressed by the creation of a cumulative frequency distribution (Figure 4.6)

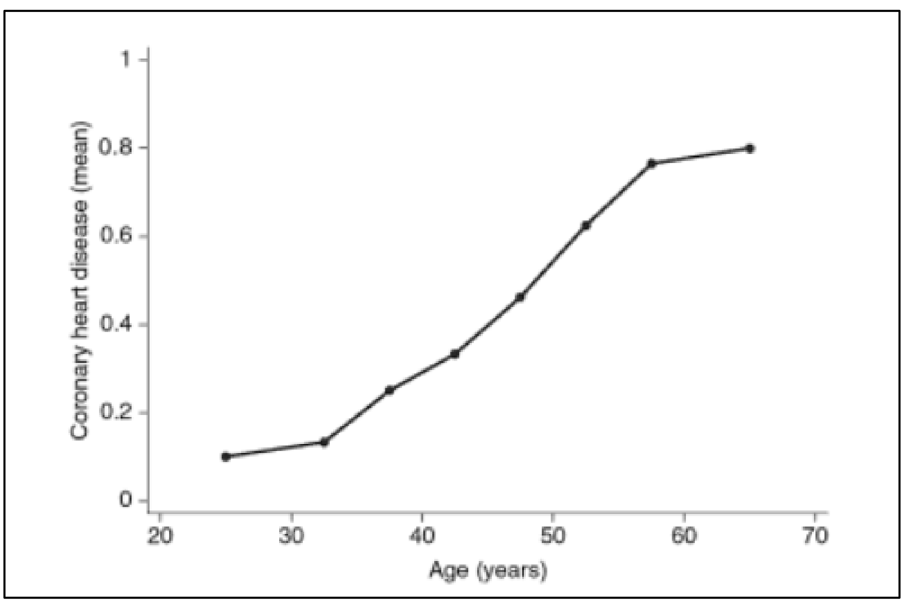

Figure 4.6: Plot of the percentage of subjects with CHD in each Age group
Source: (Hosmer et. al., 2013)

According to the scatterplot presented in Figure 4.6, the determination of a trend between the explanatory and dependent variable is now feasible; in fact, the relationship seems to be linear.

The exact same rational is applied for the needs of this research for all explanatory variables. Each explanatory variable is divided into 10 groups/bins in which 10 percent of the data is present in each bin. Although the selection of 10 bins is arbitrary, it is the most common practice. These bins are then plotted in the scatterplots in Figures 4.7-4.9. This division into bins is only applied for the purposes of the scatterplot creation and no further utilization of these bins is required. It is noted that there are other more sophisticated processes that can be applied to address the scatterplot issue in GLMs in order to decide which transformation of the explanatory variable is pertinent (Hosmer et. al. 2013), but for the needs and scope of this research, the technique described above, i.e. frequency table creation, is adequate.

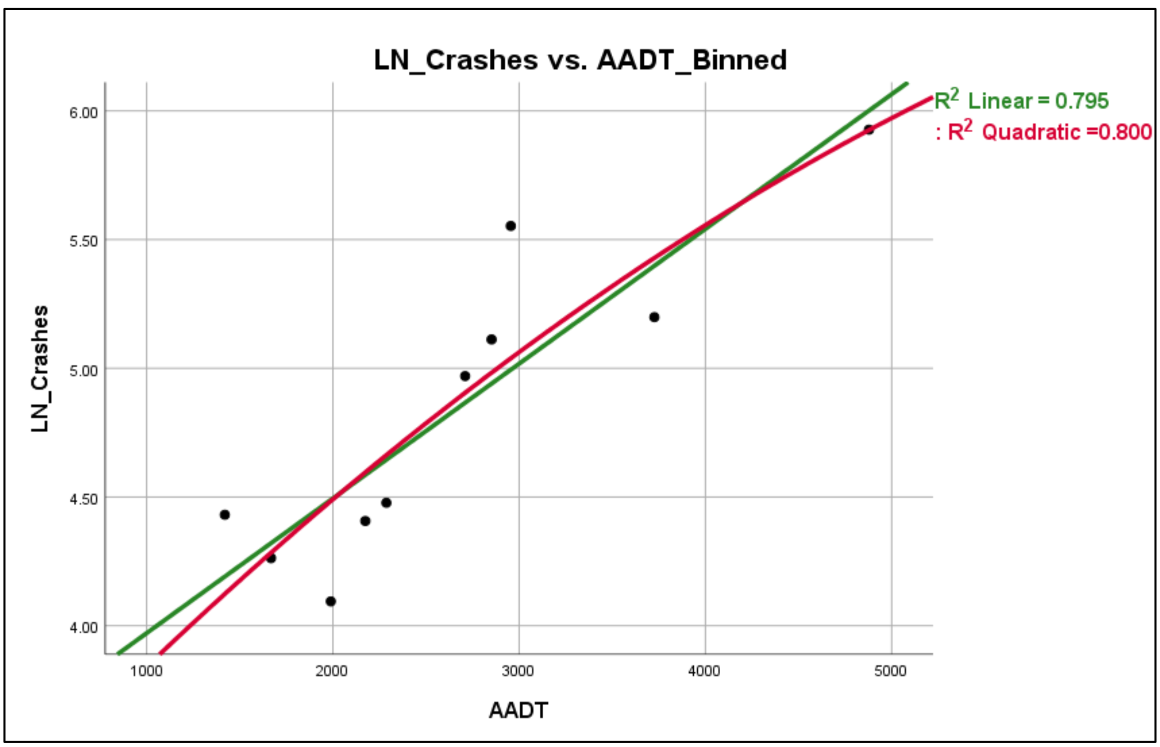

Figure 4.7: Scatterplot LN(Crashes) vs. AADT_Binned

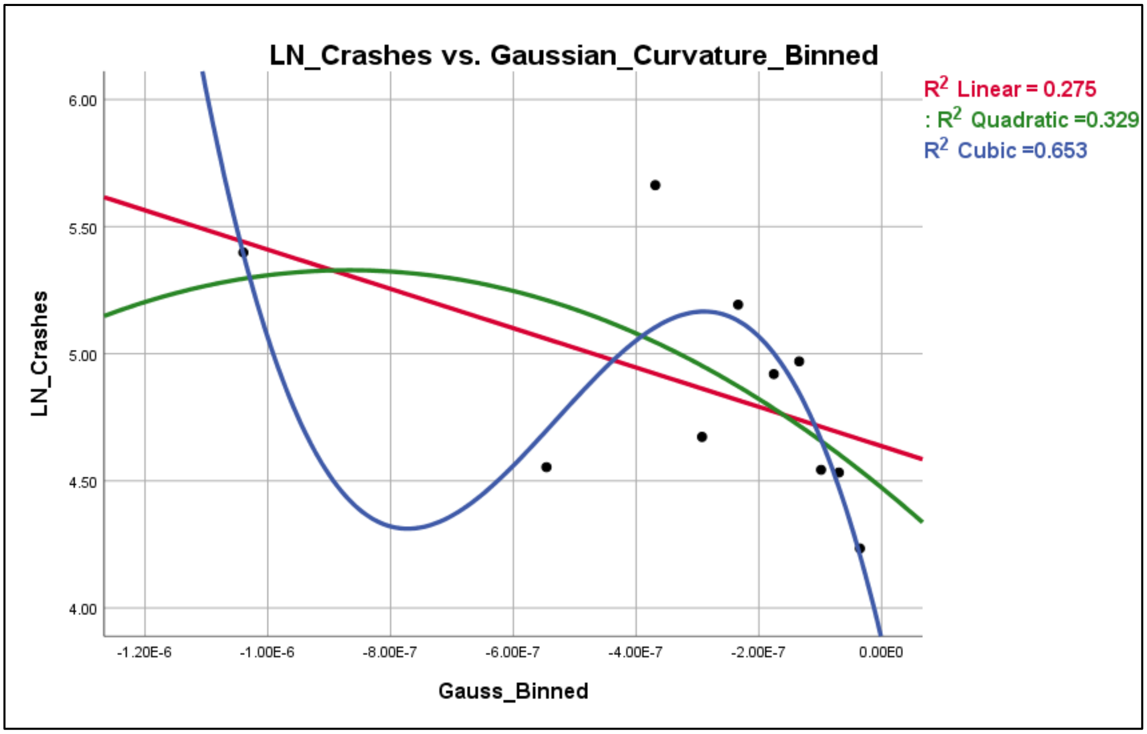

Figure 4.8: Scatterplot LN(Crashes) vs. Average GC- Binned

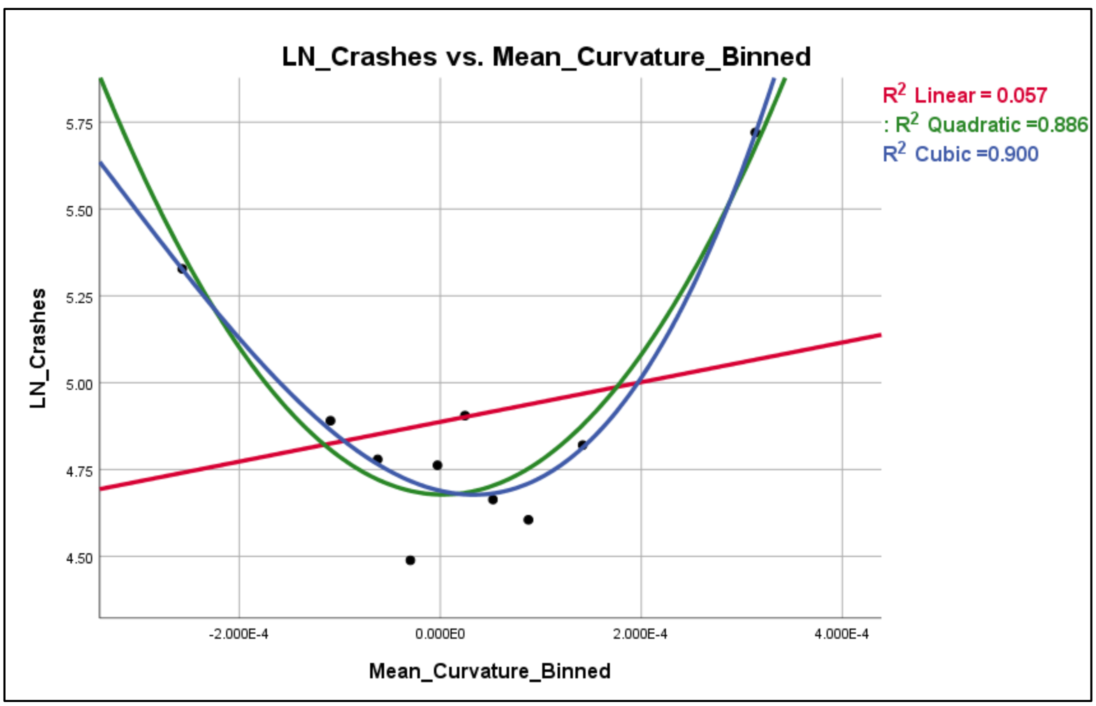

Figure 4.9: Scatterplot LN(Crashes) vs. Average MC _Binned

According to Figure 4.7Figure 4.7, the AADT seems to have a rather linear relationship with LN(crashes), whereas the Gaussian curvature (Figure 4.8) seems to have a cubic relationship with LN(crashes) and the Mean curvature (Figure 4.9) has an essentially quadratic relationship with LN(crashes). However, these observations hold true only when the explanatory variables are plotted one by one against the dependent variable; in other words, there is no guarantee that the nature of these relationships will remain the same when all the explanatory variables enter the model. However, this procedure has revealed that, especially for the Mean and Gaussian curvature, there is some indication that their relationship may not be linear in nature with LN(crashes) and therefore quadratic and cubic transformations may be appropriate for testing.

For each patch, 38 variable combinations were tested until the analysis was finalized. The models considered each variable alone and in a variety of combinations in order to determine the most appropriate and meaningful combination. All these combinations for each patch length are presented in Appendix G. The process for determining whether a model was appropriate was based on an initial determination of whether all the coefficients of the model were statistically significant and accounting for the Bonferroni correction. Then the statistically significant models were compared with the AIC criterion. It is noted that, as a rule of thumb, when two models are compared and their AICs difference is greater than 10, then this difference is “significant”, meaning that the model with the lowest AIC should be kept instead (Hardin and Hilbe 2012). Finally, the assumptions according to which the model is based, e.g., normality of deviance residual distribution, must also be satisfied.

A summary of the variables used in the best models for each patch length are summarized in Table 4.5. The final suggested models as shown in Table 4.5 indicate that the Gaussian Curvature (GC) and Mean Curvature (MC) of 3-D surfaces play a crucial role in crash prediction since they are statistically significant in all models, in which the Bonferroni correction has also been accounted for. In fact, not only are the Gaussian and Mean Curvature statistically significant in all models, but their p-values are also less than 0.001 in all models. The insertion of these two differential geometry metrics it is actually the new aspect that this research introduces to the literature. The use of these metrics can be considered promising because the Gaussian and Mean curvatures are the cornerstones of the study of 3-D mathematical surfaces as a whole in differential geometry. Moreover, the fact that transformed geometric metric, e.g. GC^3 and MC^2, and that the 2-way interaction term GC*MC insert the model, in the 100 ft patch length model, emphasizes the complexity by which roadway geometry affects crash occurrence, a fact that cannot be revealed in such an explicit manner through conventional 2-D geometric metrics. Finally, in terms of computational statistics stability, when a variable is entered in a model with a power, e.g., quadratic, it is beneficial if the "lower power terms" are also included in the model, e.g. linear, for computational reasons. Fortunately, this is the case for both the GC^3 and MC^2 variables since the variables GC^2 and GC, as well as MC are also included in the model with p-values<0.001, meaning that even the Bonferroni correction is amply satisfied.

Table 4.5: Variables Present in Final Best Models for Each Patch Length

Patch length	Explanatory Variables								
	AADT	GC	MC	GC^2	GC^3	MC^2	MC^3	AADT*MC	GC*MC
1500	X	X	X	X	X				
1000	X	X		X	X		X		
400	X	X	X	X	X	X			
200	X	X	X	X	X			X	
100	X	X	X	X	X	X			X
50	X	X	X	X	X	X			X

The criterion used in order to select the most appropriate patch length was based on the overall error prediction which is estimated as the difference between the observed and model-predicted number of crashes. A summary of the predictive ability of each patch length, i.e., the associated error percentage to each, is shown in Table 4.6; it is noted that 1,534 crashes occurred during the 2004-2017 period. Although it may be considered adequate on a practical basis to conclude that the 100 ft patch is the most pertinent patch length for the analysis, an additional statistical metric will also be considered to further validate this assertion, for the comparison among the different patch lengths. The additional statistical measurement used is the Predicted Error Sum of Squares or the so-called PRESS (Caroni and Oikonomou 2017). PRESS is used in order to compare regression models in terms of their ability to predict new values; the model preferred is the one with the smallest value of PRESS (Table 4.6).

Table 4.6: Patch Length Comparison in Terms of Predictive Ability

Patch Length (ft)	Predicted Crashes	Error Percentage of Total Crashes Predicted	PRESS
1,500	1,295	-15.6%	81.94
1,000	1,342	-12.5%	51.63
400	1,582	3.1%	47.40
200	1,546	0.8%	20.90
100	1,532	-0.1%	15.93
50	1,532	-0.1%	15.76

The selected patch length for the final model corresponds to a length of 100 ft because it was observed that this patch length provides the best modeling ability. Even though a smaller patch length leads to an increase in the predictive power of the model, this was true up to a "cut-off" patch length, which in this case was estimated to be 50 ft. In this case, "cut-off" indicates that after a certain point the overall error is not practically improved with the reduction of the patch length.

The results of the model corresponding to the 50 ft patch were identical to the ones derived from the 100 ft patch (Table 4.6). Moreover, a 100 ft patch may be considered more appropriate for transportation related applications because vehicles that have a length over 50 ft such as combination trucks, recreational cars, and buses can be analyzed in a more reliable manner by incorporating a larger surrounding roadway geometry. Therefore, for transportation related consistency, practical effect of overall error reduction as well as computational speed purposes it was decided to utilize the 100 ft patch for the crash modeling process.

The final model corresponding to the 100 ft patch length is summarized in Table 4.7, whereas the regression model is presented in Equation 4. The AIC for the models considered ranged from 11,183 to 11,803. The final model that was kept was indeed the one with the lowest AIC of 11,183 while the second-best model had an AIC of 11,232. It is noted that all of the explanatory variables of the final have a p-value<0.001, a fact that essentially demonstrates the proof-of-concept of this research: 3-D geometric roadway metrics can successfully function as explanatory variables in crash predictive models.

Table 4.7: Coefficient Values of Final Model (E35) for Patch Length = 100 ft

Variable	Coefficient	p-value
(Intercept)	-4.2701	0.000
AADT	0.00037	0.000
GC	-797,670.9567	0.000
GC^2	-62,371,846,845.508	0.000
GC^3	-101,722,389,759,530.720	0.000
MC	347.8188	0.000
MC^2	209,377.3293	0.000
GC*MC	166,612,202.0693	0.000

$$LN(Crashes) = b0 + b1 \cdot AADT + b2 \cdot GC + b3 \cdot GC^2 + b4 \cdot GC^3 + b5 \cdot MC + b6 \cdot MC^2 + b7 \cdot (GC * MC) \quad (4)$$

As noted above, the presence of the Gaussian Curvature and Mean Curvature of 3-D surfaces supports the significance of these variables as crash predictors and the potential interaction with other variables; interactions that can by no means be captured in the 2-D analysis.

At this point, the "Integrated Model" in which all years and roadways are included has been finalized and presented in Equation 4 above. The IM essentially functions as a proof-of-concept for the inclusion of the 3-D metrics in crash prediction models and can, at least theoretically, be used for crash prediction purposes in other roadways. However, the latter is not recommended, especially in areas not in Kentucky in which even driver behavior may be different; hence, more roadways should be analyzed in order to increase the predictive power and representativeness of the model. Nonetheless, this model may be particularly useful when the purpose of an analysis is not the prediction of crashes in absolute numbers, but the comparison of alternatives, e.g. different alignments, in terms of estimating which alternative reduces crash frequency. In addition, it is suggested that the specific coefficient values (Table 4.7) be used for crash prediction purposes only when no historical crash data are available; if crash data are available for a specific roadway segment they should be certainly used in order to incorporate the "special crash pattern" in the adjusted model to be discussed in the next section. Finally, when several years of crash data are available, it is advised that, for crash prediction purposes, the years enter the model as dummy variables. The latter is suggested in order to account for seasonal and time effects. This is further discussed in the next section in which the predictive power of the model is evaluated.

4.2.2 Model Structure and Predictive Ability Evaluation

The ultimate objective of this research is the determination of the predictive ability of the proposed model based on 3-D metrics on safety predictions. The comparison is based on the crash predictions as estimated from the model and the IHSDM. The IHSDM predicts crashes per year for a given roadway through the Empirical Bayes model. To account for the differences that arise throughout the years such as the number of crashes and AADT, IHSDM needs to develop a separate prediction for each year and this approach was considered and applied in the suggested model to obtain an accurate and fair comparison. It is therefore important to consider this in the model developed here and determine how to best approach it. There are two options for incorporating the "year effect" in this analysis: 1) use a separate model for each year developing predictions one year at a time; and 2) insert dummy variables for years to account for the different AADT for each year. The following presents this analysis and the determination of which approach is more appropriate. It should be also noted that there is no concern whether the dummy variables are statistically significant or not at this point; their purpose is to simply increase the predictive ability of the model by accounting for the yearly variation of AADT and random effects in general.

The evaluation will be accomplished by creating training data, i.e., assuming that a certain year is not included in the dataset, running the analysis, and then predicting the crashes of that year and reporting the residuals. For example, the way in which the predictive power of the model will be evaluated for year 2017 is as follows. Suppose that crash data are available for years 2004-2016 and that the intention is to predict the crashes for year 2017. The predictive model will include the explanatory variables of the IM, i.e. AADT, GC, GC^2, GC^3, MC, MC^2, and GC*MC. For the use of the dummy variable approach, in addition to the explanatory variables, a number of dummy variables equal to the number of years of crashes minus 1 is used. In this case, for the 13 years of available date (2004-

2016 period) 12 (=13-1) dummy variables will be used. The crash predictions for year 2017 will be calculated in the following form (Equation 5):

$$LN(13 * Crashes) = b0 + b1 \cdot AADT + b2 \cdot GC + b3 \cdot GC^2 + b4 \cdot GC^3 + b5 \cdot MC + b6 \cdot MC^2 + b7 \cdot (GC * MC) + d1 \cdot D_{2004} + \cdots + d12 \cdot D_{2015} \quad (5)$$

or finally:

$$LN(Crashes) = b0 + b1 \cdot AADT + b2 \cdot GC + b3 \cdot GC^2 + b4 \cdot GC^3 + b5 \cdot MC + b6 \cdot MC^2 + b7 \cdot (GC * MC) + d1 \cdot D_{2004} + \cdots + d12 \cdot D_{2015} - LN(13) \quad (6)$$

The term “-LN(13)” is present in Equation 6 in order to convert the prediction model on a per year basis since the model is based on 13 years of data. In statistical terminology, especially for GLM, this “-LN(13)” term is the so-called offset in the Negative Binomial Regression (Hardin and Hilbe 2012). The crash predictions for any other year will be calculated with the same exact procedure and rationale. The model structure is evaluated using both approaches, with and without dummy variables, and then comparing the predictions to the actual number of crashes. The approach that results in a prediction closer to the actual crashes would be the one to be used.

In Table 4.8 the crashes prediction breakdown per year and roadway segment is presented in which there are three columns for each roadway segment: 1) Actual Crashes (AC), 2) Predicted Crashes Without the Utilization of the Dummy Variables Approach (W/O), and 3) Predicted Crashes Without the Utilization of the Dummy Variables Approach (W/). In addition, Table 4.9 presents the errors/residuals corresponding to the models with and without the dummy variable approach, as well as the corresponding Crash Improvement (CI) that has been achieved with the dummy variable approach.

Table 4.8: Crash Predictions Estimates

	KY420-1			KY420-2			KY152-1			KY152-2			KY152-3			US68-1			US68-2		
Year	**AC**	**W/O**	**W/**	**AC**	**W/O**	**W/**	**AC**	**W/O**	**W/**	**AC**	**W/O**	**W/**	**A**	**W/O**	**W/**	**AC**	**W/O**	**W/**	**AC**	**W/O**	**W/**
2004	29	12	21	3	5	2	12	34	12	2	11	4	10	7	6	9	13	12	34	38	40
2005	15	12	15	0	4	2	8	34	9	4	11	3	8	7	5	6	12	9	30	38	30
2006	20	11	18	2	4	2	9	33	10	2	10	4	4	7	5	17	13	10	29	38	34
2007	21	11	14	3	4	2	3	31	8	1	10	3	3	7	4	6	13	8	30	38	27
2008	26	11	24	3	4	3	13	31	14	2	10	5	7	7	7	12	13	14	50	37	47
2009	33	11	23	3	4	3	8	31	13	6	10	5	5	7	7	17	13	14	38	37	45
2010	34	10	32	3	3	4	25	31	18	7	10	7	7	7	10	22	12	19	51	38	63
2011	30	10	35	0	3	4	21	30	20	8	9	7	8	7	10	26	12	20	72	38	68
2012	48	10	35	7	3	4	17	28	20	5	8	7	5	7	10	16	12	21	71	39	69
2013	28	11	29	2	3	3	10	26	16	10	8	6	4	7	8	13	12	17	68	39	56
2014	5	12	19	0	4	2	13	24	11	1	7	4	10	7	6	7	12	11	53	39	37
2015	12	13	23	3	5	3	17	23	13	5	6	5	7	7	7	18	12	13	42	37	45
2016	12	14	22	3	5	2	21	22	12	8	6	4	13	6	7	9	12	13	33	37	43
2017	12	14	15	4	5	2	7	22	9	5	6	3	6	6	5	12	12	9	24	36	30
Total	**325**	**162**	**327**	**36**	**56**	**36**	**184**	**399**	**185**	**66**	**120**	**66**	**97**	**97**	**97**	**190**	**172**	**190**	**625**	**530**	**634**

Note: AC: Actual Crashes; W/O: Without Dummy Variables; W: With Dummy Variables

Table 4.9: Error Comparison in Crashes

	KY420-1		KY420-2		KY152-1		KY152-2		KY152-3		US68-1		US68-2	
	AC-W/O	AC-W/	AC-W/O	AC-W/	AC-W/O	AC-W/	AC-W/O	AC-W/	AC-W/O	AC-W/	AC-W/O	AC-W/	AC-W/O	AC-W/
2004	17	8	-2	1	-22	0	-9	-2	3	4	-4	-3	-4	-6
2005	3	0	-4	-2	-26	-1	-7	1	1	3	-6	-3	-8	0
2006	9	2	-2	0	-24	-1	-8	-2	-3	-1	4	7	-9	-5
2007	10	7	-1	1	-28	-5	-9	-2	-4	-1	-7	-2	-8	3
2008	15	2	-1	0	-18	-1	-8	-3	0	0	-1	-2	13	3
2009	22	10	-1	0	-23	-5	-4	1	-2	-2	4	3	1	-7
2010	24	2	0	-1	-6	7	-3	0	0	-3	10	3	13	-12
2011	20	-5	-3	-4	-9	1	-1	1	1	-2	14	6	34	4
2012	38	13	4	3	-11	-3	-3	-2	-2	-5	4	-5	32	2
2013	17	-1	-1	-1	-16	-6	2	4	-3	-4	1	-4	29	12
2014	-7	-14	-4	-2	-11	2	-6	-3	3	4	-5	-4	14	16
2015	-1	-11	-2	0	-6	4	-1	0	0	0	6	5	5	-3
2016	-2	-10	-2	1	-1	9	2	4	7	6	-3	-4	-4	-10
2017	-2	-3	-1	2	-15	-2	-1	2	0	1	0	3	-12	-6
Total	**163**	**-2**	**-20**	**0**	**-215**	**-1**	**-54**	**0**	**0**	**0**	**18**	**0**	**95**	**-9**
CI*	**161**		**20**		**214**		**54**		**0**		**18**		**86**	

*CI: Crash Improvement Per Roadway Segment with the Dummy Variables Approach

The summary row in Table 4.9 denotes that the inclusion of dummy variables results in predictions that are closer to the actual crashes than the without using them. Moreover, the insertion of dummy variables is preferred, in general, over the creation of separate models for each year because it is statistically more appropriate: the Bonferroni correction can be applied in a much more robust manner, since the familywise error is explicitly defined, and small sample size issues, which are in general present in crash datasets, are alleviated with the dummy variable approach.

Therefore, at this point it is decided to utilize the dummy variable approach in order to compare the crash predictions of the suggested model with those derived from the IHSDM. The comparison follows in the next section.

The next step involves the evaluation of the assumptions of the model developed, since every regression model is based on some statistical, mostly distribution related, assumptions. This applies in this case as well, and therefore these assumptions must be checked in order to validate the reliability of the model. In practical/applied terms, failure in assessing these assumptions would mean that the coefficients of the model are not reliable, i.e., the coefficients are inflated or deflated compared to the true parameters. Moreover, the defined confidence levels of the coefficients may not hold true, a fact that means that the exported p-values from the models may be highly distorted, which, in turn, means that although the model may be considered statistically significant based on the explanatory variables p-values, it may in fact not be statistically significant since the results may just be artificially in favor of rejecting the null hypotheses.

There are many techniques that have been suggested for the assumption assessment of regression models, but especially in the case of GLMs, this matter remains an open research problem. Therefore, for the scope of this research, the basic assumptions assessment techniques will be checked for which there is a general agreement in terms of their effectiveness and pertinence from the scientific community. More specifically, the assumption assessment will be based on two elements: 1) Residuals Analysis, and 2) Influential Points Identification.

4.2.3 Residuals Analysis

There are two types of residuals that typically come into play in GLMs: Standardized Pearson Residuals and Standardized Deviance Residuals. The terms “standardized” is incorporated in these residuals because the initial Pearson and Deviance residuals have to be standardized in order to account for altering variance among the observations in GLMs and therefore be comparable (Caroni and Oikonomou 2017). Besides these two residual types, other residuals are mentioned in the literature such as Likelihood Residuals (McCullagh and Nelder 1989) and Anscombe Residuals (Hardin and Hilbe 2012). However, there is a general agreement that the standardized deviance residuals are the most pertinent and useful residuals to be utilized. However, the truth of this latter fact also greatly depends on the nature of the application itself.

Although, in theoretical statistics terms, the deviance residuals do not have to follow a normal distribution, the pertinence of the model can be assessed if these residuals indeed follow a normal distribution. Therefore, the demand for normally distributed deviance residuals is imposed by logic rather than theoretical statistics. The normality of these residuals indicates that there is no systematic effect in the data and that the errors/residuals are random; the randomness of observations is a fact highly desired in any regression model. If a systematic effect is present, i.e., the residuals follow a pattern, then this systematic error can be, in general, filtered out with the inclusion of an additional explanatory variable that is meaningful. The Q-Q Plot of the Standardized Deviance Residuals for the developed model is shown in Figure 4.10, whereas the Standardized Deviance Residuals (SDR) for each observation/case is shown in Figure 4.11.

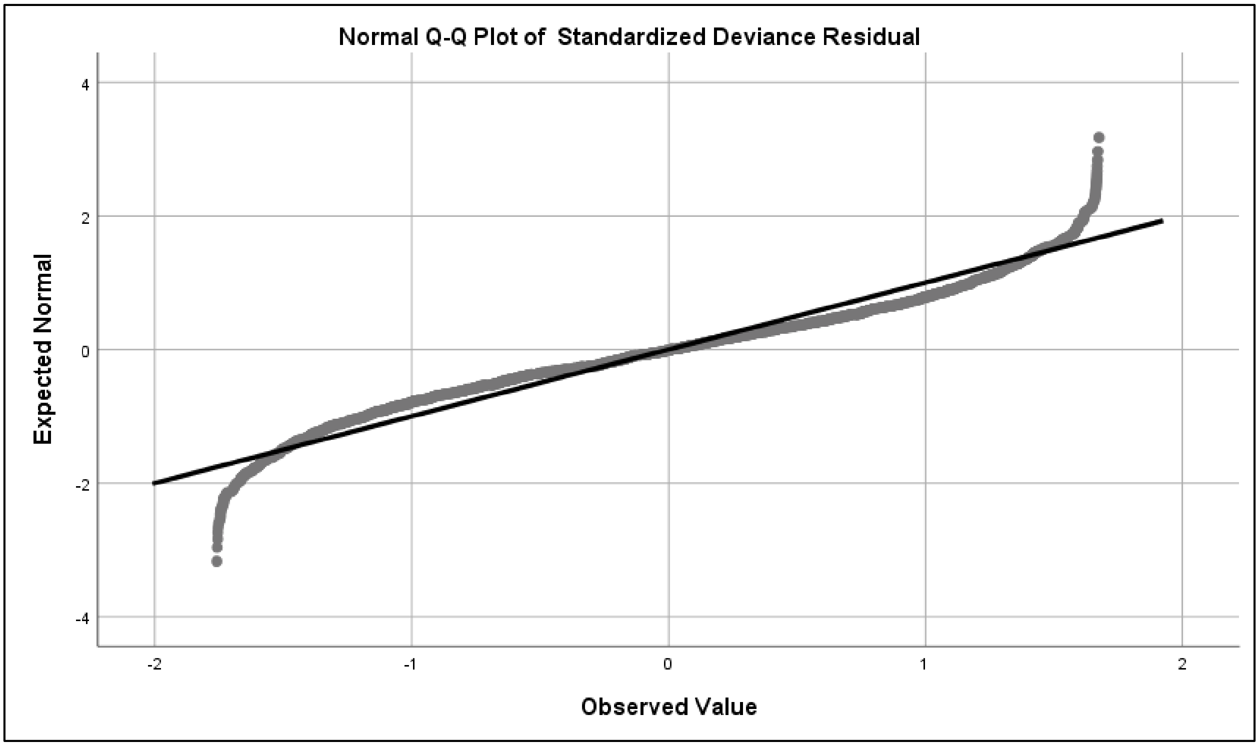

Figure 4.10: Q-Q Plot of Standardized Deviance Residual

According to Figure 4.10, the observations seem to adequately lie on a line. This means that the standardized deviance residuals are also adequately normally distributed. This line represents the theoretical values of the Normal Distribution, meaning that the closer the SDRs are to the line, the more normally distributed they are. It is noted that in the case of GLMs, the normality check of the SDRs, it is not intended to check whether the SDRs are normally distributed, but it is a check in order to implicitly investigate the satisfactory level of fit of the model.

It can be observed that the tails of the SDRs deviate from the theoretical values of the Normal Distribution. However, minor deviation should not be considered a factor that would render the SDRs problematic in terms of failure to view the SDRs as normally distributed. Therefore, it is concluded that the trend of the SDRs are linear, a conclusion that, in turns, implies that the fit of the model can be considered satisfactory at this point.

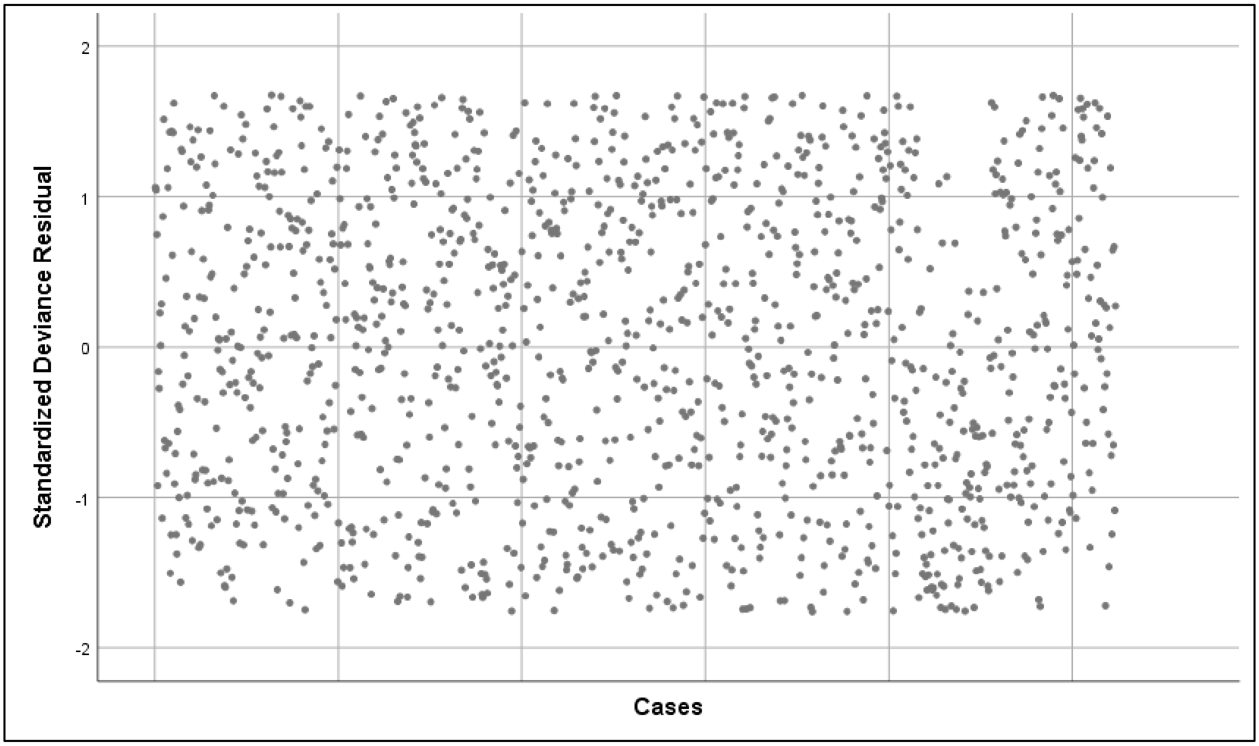

Figure 4.11: Standardized Deviance Residuals Plot

Figure 4.11 investigates the randomness of the observations, which is a factor greatly desired in any regression model. After all, the purpose of a regression model is to exactly capture and absorb all systematic effects/variables and therefore only leave random effects; randomness will always be present in any model and cannot be avoided. The ideal scenario for SDRs vs. Observation scatterplots, i.e. Figure 4.11, is to produce “white noise”; in other words, no pattern should be present. As it can be observed, Figure 4.11 can be characterized by the term “white noise” indicating that the systematic effects have been account for in the model in an effective manner.

4.2.4 Influential Points

A point is characterized as “influential” if its exclusion would have the power to considerably change the coefficient values of the variables included in the model. Briefly speaking, an “outlier” is an observation that has an unusual y, i.e., prediction, value, whereas a “leverage point” is an observation that has an unusual x value. An influential point is essentially an observation that is both an outlier and a leverage point. The most common way to identify influential points is via Cook’s distance; this identification can be implemented either by comparing the values to some absolute cutoff or by simply seeking for observation whose Cook's distance is unusual in terms of the general trend. The latter technique has been proven to be more effective and therefore this approach will be utilized here (Dielmann 2005).

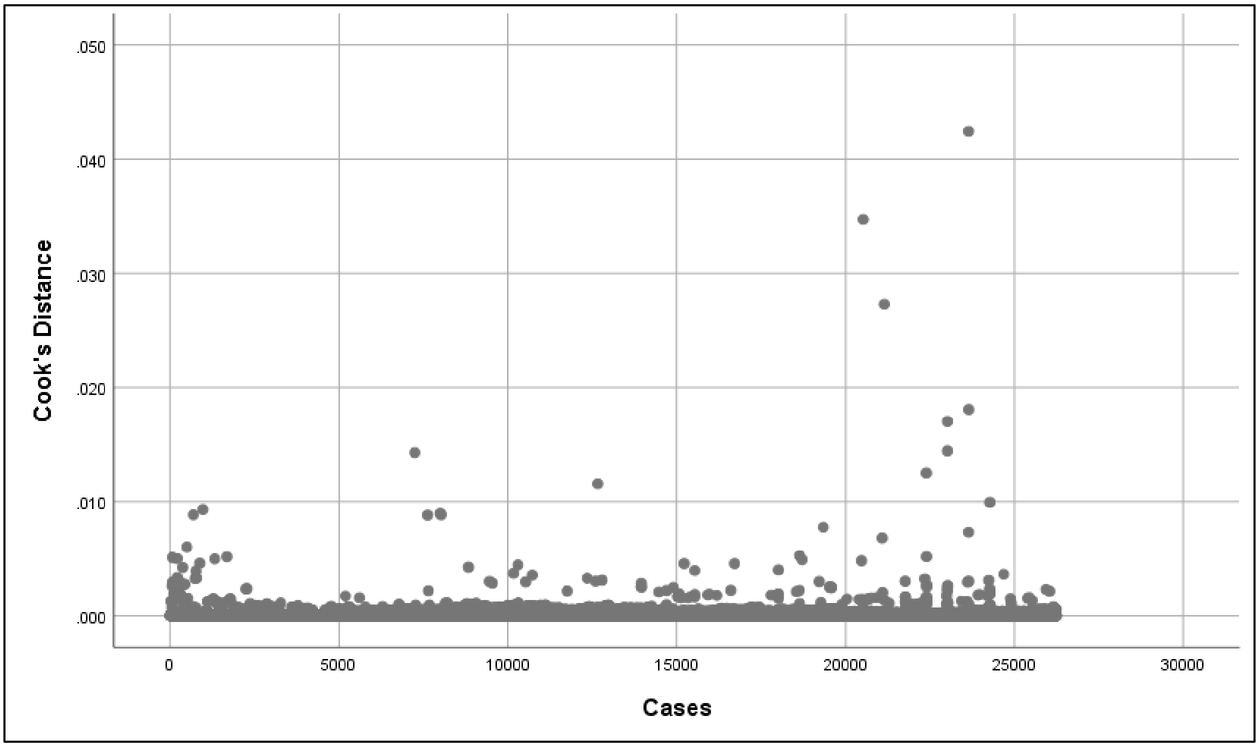

Figure 4.12: Cook's Distance

According to Figure 4.12, there seem to be about 10 points that somehow deviate from the general trend of Cook's distance. However, these minor deviations cannot by any means be considered influential points given the sample size of the dataset. Moreover, it is preferable to not delete observations from the analysis, but even if these observations were deleted, other influential points would appear in the Cook's distance plot. The purpose of the Cook's distance plot is mainly to identify observations with unusual values of Cook's distance and further investigate them; it is true that these observations are in many cases the most interesting observations, containing substantially rich information regarding parameters that may affect the dependent variable or even suggest the inclusion of additional specific meaningful explanatory variables in the model. Finally, a typical absolute cutoff point for Cook's distance is "one" and as indicated in Figure 4.12, the maximum Cook's distance value, i.e., 0.042, is much less than one. Finally, it can be concluded that influential points are not included in the model and no observations are excluded from the analysis; therefore, no further action is necessary to be taken in order to deleted them.

4.3 Comparison to Current Guidelines

This section evaluates the predictive ability of the suggested model with the current safety prediction methodology as utilized in the IHSDM. As described in the Methodology chapter, crash predictions are derived through SPFs, which are the building blocks of the HSM. The equations of the HSM have been incorporated into the IHSDM which makes the calculations automatic. Moreover, the IHSDM can account for historical crash data and essentially adjust the crash prediction results through the Empirical Bayes model. Therefore, the results of the suggested model will be compared to the ones that would be obtained through the IHSDM. The comparison results are based on the prediction models obtained by applying a 100 ft patch for the 3-D roadway surface.

Table 4.10 presents the crashes prediction breakdown per year and roadway segment. There are three columns for each roadway segment: 1) Actual Crashes (AC), 2) Predicted Crashes produced by the IHSDM software (IH), and 3) Predicted Crashes produced by the Suggested Model (SU). In addition, Table 4.11 presents the errors/residuals comparison between the suggested model and the IHSDM, as well as the corresponding Crash Prediction Improvement (CI) that has been achieved with the suggested model. The crash prediction differences will also be presented per mile as well, since this is another unit that the IHSDM utilizes.

Table 4.10: Crash Predictions Estimates Comparison

	KY420-1			KY420-2			KY152-1			KY152-2			KY152-3			US68-1			US68-2		
	AC	IH	SU	AC	IH	SU	AC	IH	SU	AC	IH	SU	AC	IH	SU	AC	IH	SU	AC	IH	SU
2004	29	25	21	3	8	2	12	21	12	2	7	4	10	11	6	9	18	12	34	49	40
2005	15	25	15	0	8	2	8	21	9	4	8	3	8	11	5	6	18	9	30	49	30
2006	20	25	18	2	8	2	9	22	10	2	8	4	4	12	5	17	18	10	29	51	34
2007	21	25	14	3	8	2	3	22	8	1	8	3	3	12	4	6	19	8	30	51	27
2008	26	25	24	3	8	3	13	21	14	2	8	5	7	11	7	12	19	14	50	49	47
2009	33	24	23	3	8	3	8	21	13	6	8	5	5	11	7	17	18	14	38	50	45
2010	34	24	32	3	9	4	25	21	18	7	7	7	7	11	10	22	18	19	51	49	63
2011	30	24	35	0	9	4	21	21	20	8	8	7	8	11	10	26	18	20	72	48	68
2012	48	23	35	7	8	4	17	21	20	5	8	7	5	11	10	16	18	21	71	47	69
2013	28	26	29	2	8	3	10	21	16	10	8	6	4	11	8	13	19	17	68	47	56
2014	5	27	19	0	9	2	13	21	11	1	8	4	10	11	6	7	19	11	53	49	37
2015	12	26	23	3	8	3	17	22	13	5	8	5	7	11	7	18	18	13	42	50	45
2016	12	26	22	3	8	2	21	21	12	8	7	4	13	11	7	9	19	13	33	51	43
2017	12	25	15	4	8	2	7	21	9	5	8	3	6	11	5	12	18	9	24	49	30
Total	**325**	**351**	**327**	**36**	**116**	**36**	**184**	**296**	**185**	**66**	**106**	**66**	**97**	**156**	**97**	**190**	**257**	**190**	**625**	**691**	**634**

Note: AC: Actual Crashes; IH: IHSDM prediction; SU: Suggested model prediction

Table 4.11: Error Comparison in Crashes

	KY420-1		KY420-2		KY152-1		KY152-2		KY152-3		US68-1		US68-2	
	AC-IH	**AC-SU**	**AC-IH**	**AC-SU**	**AC-IH**	**AC-SU**	**AC-IH**	**AC-SU**	**AC-IH**	**AC-SU**	**AC-IH**	**AC-SU**	**AC-IH**	**AC-SU**
2004	4	8	-5	1	-9	0	-5	-2	-1	4	-9	-3	-15	-6
2005	-10	0	-8	-2	-13	-1	-4	1	-3	3	-12	-3	-19	0
2006	-5	2	-6	0	-13	-1	-6	-2	-8	-1	-1	7	-22	-5
2007	-4	7	-5	1	-19	-5	-7	-2	-9	-1	-13	-2	-21	3
2008	1	2	-5	0	-8	-1	-6	-3	-4	0	-7	-2	1	3
2009	9	10	-5	0	-13	-5	-2	1	-6	-2	-1	3	-12	-7
2010	10	2	-6	-1	4	7	0	0	-4	-3	4	3	2	-12
2011	6	-5	-9	-4	0	1	0	1	-3	-2	9	6	24	4
2012	26	13	-1	3	-4	-3	-3	-2	-6	-5	-2	-5	24	2
2013	2	-1	-6	-1	-11	-6	2	4	-7	-4	-6	-4	21	12
2014	-22	-14	-9	-2	-8	2	-7	-3	-1	4	-12	-4	4	16
2015	-14	-11	-5	0	-5	4	-3	0	-4	0	0	5	-8	-3
2016	-14	-10	-5	1	0	9	1	4	2	6	-10	-4	-18	-10
2017	-13	-3	-4	2	-14	-2	-3	2	-5	1	-6	3	-25	-6
Total	**-26**	**-2**	**-80**	**0**	**-112**	**-1**	**-40**	**0**	**-59**	**0**	**-67**	**0**	**-66**	**-9**
CI*	**24**		**80**		**111**		**40**		**59**		**67**		**57**	
CI PM*	**17**		**91**		**13**		**18**		**17**		**12**		**5**	
CI PY*	**2**		**6**		**8**		**3**		**4**		**5**		**4**	
CI PM PY*	**1.2**		**6.5**		**0.9**		**1.3**		**1.2**		**0.8**		**0.4**	

*CI: Crash Improvement Per Roadway Segment with the Suggested Model Compared to the IHSDM

*CI PM: Crash Improvement Per Roadway Segment Per Mile with the Suggested Model Compared to the IHSDM

*CI PY: Crash Improvement Per Roadway Segment Per Year with the Suggested Model Compared to the IHSDM

*CI PM PY: Crash Improvement Per Roadway Segment Per Mile Per Year with the Suggested Model Compared to the IHSDM

Tables 4.10 and 4.11 show that the crash prediction results produced by the suggested model are more comparable to the actual crashes than those obtained from the IHSDM, which is the current crash prediction practice. The best Crash Prediction Improvement, i.e., CI, in absolute numbers for the 14-year period is observed for the roadway segment KY 152-1 (Table 4.11). The aggregated (i.e., sum) error of the suggested model is only -1, whereas the respective error obtained from the IHSDM model is -112. This means that the suggested 3-D geometric safety model has a prediction that is closer to the actual number of crashes compared to the IHSDM model, by 111 crashes. Similarly, the best Crash Prediction Improvement Per Mile Per Year (CI PM PY) is observed for the roadway segment KY 420-2; specifically, the suggested 3-D geometric safety model prediction is closer to the actual crashes compared to the IHSDM model, by 6.5 crashes per year for every mile of highway. These observations and general results could be seen as an indication of the beneficial effects that 3-D geometric metrics can offer to highway safety and can be considered a practical proof-of-concept of this research itself. Another fact that makes this research promising is that the comparison is conducted in a quantifiable manner. However, although the improvement in crash prediction is an important issue on its own, this improvement should be demonstrated not only in crash units, but also in monetary values. Although one may argue, on a philosophical level, that the human life is priceless, this approach does not convey the whole truth as implemented in practice. It is true that even fatalities, injuries, and property damage are incorporated into an optimization scheme in order to reach decisions during the planning/budgeting phase of a project that would eventually have the optimal effect to society as a whole. For example, although it may be observed that the crash occurrence on a particular roadway is problematic, a cost-benefit analysis is typically conducted to determine how to best allocate limited resources to increase their effectiveness. Such improvements are then compared to other competing projects and needs and decisions are reached based on optimizing the available funds for the greater good of the system.

Varying crash predictions can lead to vastly different decisions because they are essentially tightly linked to the cost estimation of a project, new or existing that is considered to be modified. It is therefore imperative that predictions are accurate to avoid assigning incorrect priorities while addressing needs. Although crash costs are just a portion of the total project cost, there are substantial in both economic and societal terms. Crash over prediction may yield a project too expensive, leading to its rejection, whereas crash under prediction will lead to under design problematic issues and inflated crash occurrence.

Considering the potential impact of the overestimation of the IHSDM as compares to the suggested model, one can surmise that when cost-benefit estimates are required for projects those could be grossly miscalculated and thus potentially result in addressing the wrong projects. More accurate cost estimation procedures can greatly benefit both public agencies and private companies during the biding phase of a highway engineering project because their estimations will be in line with reality in a more reliable manner. Moreover, tax payers can feel more confident that their contribution is invested in a better way and in the long run public agencies could potentially design and construct additional infrastructure projects, such as schools and parks, with the same initial budget.

Thus, at this point, the intended proof-of-concept of the research, i.e., 3-D differential geometry metrics have a crash prediction value, has been established and the final model has been compared to the current practices through the use of the IHSDM. The latter comparison verified that the results derived from the suggested model, containing 3-D explanatory variables derived from differential geometry, are closer to the actual/observed crashes, compared to the SPFs of the HSM.

An advantage of the proposed model is that it only requires XYZ data of the roadway instead of the detailed geometric data input of the IHSDM which requires as input the horizontal and vertical alignment information. Therefore, the proposed model is more flexible than the IHSDM. In addition, if the horizontal and vertical alignment plans are not available, then it is difficult, if not impossible, to utilize the IHSDM, since these are essential inputs for the calculation process. The data entry in the IHSDM is also a tedious process and demand a manual entry. The proposed model takes advantage of the automated conversion from the 3-D XYZ data to the horizontal and vertical alignment via the FM-17 software offered by the National Technical University of Athens, Department of Civil

Engineering. The FM-17 offers some semi-automatic tools that assist in the 3-D to 2-D conversion, but the procedure remains subjective and demand manual correction at the end of the process.

5 CONCLUSIONS AND FUTURE RESEARCH

Highway design is an engineering principle that combines various aspects and specializations from the engineering spectrum. However, highway design can be also considered a form of "engineering art" in the sense that every project is different and there is not only one solution to each problem. There are many competing parameters embedded in the design process and therefore, engineering judgement is actively present in almost all steps of the design process. The latter fact justifies the need of evaluating several alternative design options before concluding to the final one in order to analyze and approach the problem from different angles. However, even when the final design has been decided, it does not mean that it is the optimal design solution; it simply means that the suggested solution complies in an acceptable manner with the needs, e.g. safety, comfort, and guidelines set forth, environmental, historical and budget constraints, and community needs, imposed by the problem itself.

Highways are large scale 3-D infrastructure systems but unfortunately, they are not treated in that way. The traditional 2-D approach is applied in order to design these 3-D structures. A number of research reports focus on the need of shifting the perceptive of the design analysis from 2-D to 3-D, which would potentially offer a much more holistic approach to roadway design. Although various attempts have been to incorporated or, at least, generate the discussion in a more active manner for the inclusion of 3-D metric in the design process, it can be stated that these attempts have not been accepted as something practical in the sense that they could be readily applied in the near future. In order to convince the scientific community that more research should be conducted towards the 3-D direction, these 3-D suggested models should be compared in a quantifiable manner to the traditional 2-D approaches in order for the potential advantages of the 3-D solutions to be proven in an undisputed manner through practical applications and numerical comparisons. This was indeed the objective of this research: the use of 3-D metrics in crash prediction models and their comparison to results derived from current, 2-D-based practices. The results of this research are very promising since they have demonstrated an improvement in crash predictions compared to the current practice and therefore demand for further research.

The basic 3-D differential geometry metrics that were used were the Gaussian Curvature and the Mean Curvature. These two metrics are very important in differential geometry and especially for the study of 3-D surface properties. Therefore, what has been achieved here is the incorporation of 3-D metrics that essentially govern universal properties of 3-D surfaces, as in this case in which the roadway is treated as a 3-D mathematical surface, in highway safety predictive models. The incorporation of these metrics offers accuracy and flexibility in the suggested 3-D safety models since the roadway is modeled exactly as it is designed and, more importantly, constructed in the field. In addition, the crash predictions can be produced from the suggested model in an absolutely automated procedure, meaning no conversion into the horizontal and vertical alignment is required. Therefore, the evaluation of a project itself or multiple alternatives can be conducted much faster, saving numerous work hours and productivity in general. Another presumably important fact of this research is that the comparison between the suggested model and the IHSDM has shown that the IHSDM tends to overestimate the number of crashes and thus could result in inappropriate cost-benefit evaluations that could distort the project priorities and comparisons. This could help public agencies to better allocate their available funds. Finally, the power of utilizing a 3-D model for interpreting the 3-D roadway surface as the basis of all further analysis, which in this case is crash prediction models development, has been justified in a rather concrete manner; according to the results (see Table 4.11), the incorporation of 3-D analysis substantially improves the prediction of safety models. As far as applicability is concerned, it can be stated that the suggested methodology is rather practical since the only required input data are, at a minimum, the XYZ coordinates of the roadway centerline and edge lines. Given the contemporary surveying technology that is available, e.g. laser scanning, the input data for an existing roadway is relatively easy to be retrieved. For new highway design, the required input data are already available even in the case of the traditional 2-D design approach.

It is emphasized once again that models require further evaluation, since they were developed based on a small sample size of rural roadways. The purpose of this analysis was to demonstrate that the 3-D generated variables have the ability to capture the interactions of the various roadway geometry elements and prove that they could have a better predictive ability than the current practices. The data needs for the proposed model are less demanding than those required for the IHSDM and all current safety prediction practices rely on geometry information for estimating the number of crashes. Finally, it is stressed that the proposed regression equations apply to the specific roadway segments that were examined in this thesis. The models can be further improved with the addition of more roadway segments in order to allow for a more robust statistical evaluation of the coefficients estimated and permit a more accurate and wider-accepted implementation.

Nonetheless, the proposed methodology is easier to implement and less demanding, in terms of data manipulation and subjective decision-making, compared to the IHSDM implementation because segments with homogeneous geometry, e.g., segment constant horizontal radius and vertical grade are not required to be identified, in the suggested methodology, when 3-D data are available. Moreover, there is no need to find a way to convert the 3-D information, i.e., XYZ data of the roadway centerline and edge lines, into a horizontal and vertical alignment; a process that is, in general, subjective. An integrated highway model/system can increase the speed in which roadway related infrastructure is designed and constructed, leading to reduced costs throughout all the design and construction phases of the project. Therefore, this superiority is not necessary to be solely restricted to the comparison of 2-D and 3-D results in terms of design accuracy, precision, or crash prediction but it could be also beneficial in terms of cost reduction such as man-hours that are necessary to conduct the highway design or construction process or take-off estimating accuracy.

Finally, it is advised that crash data be utilized when they are available; if they are not available, then one can use the general regression equation but only to compare between different alignments. In the latter case, the crash predictions will most likely be erroneous in terms of absolute numbers, but the difference in crashes predictions between alternative alignments will be more reliable. This is also the case with the IHSDM tool.

5.1 Future Research Recommendations

Many recommendations for future research can be made since the incorporation of 3-D metric in the highway design process is a rather unexplored field. Some suggestions are presented in this section.

A major concern of this research was the determination of the patch length according to which the statistical analysis was based upon. A future step would be to create a mesh on the roadway surface whose patches would not necessarily have equal lengths, but would alter depending on the special geometric properties of the surface each time. This mesh would look like the meshes that are utilized in the Finite Element Method (FEM). The meshing criteria would be related to the Gaussian and Mean curvature values of the 3-D B-Spline roadway surface. In practical terms, this means that the mesh would be denser in surface areas, i.e., on the roadway surface, that have larger values of Gaussian and Mean Curvatures and sparser in surface areas that are more "flat". It is anticipated that this may increase the predictive ability of the model as well as the computational speed of the procedure. Moreover, it may be easier to identify problematic roadway segments in terms of highway safety due to the more accurate construction of the underlying mesh, which would be purely based on the differential geometry properties of the roadway surface and implemented with computational geometry techniques.

In this research attempt the shoulder width is not included as an explanatory variable because all roadway segments have the same shoulder width and therefore no differentiation is possible in the statistical analysis. The shoulder is not modeled as part of the 3-D B-spline surface because the driver does not, at least typically, travel on the shoulder. However, the intension is to include additional metrics, directly related to highway design, such as shoulder widths and lane widths in the regression models in the future. This task can be accomplished with the inclusion of dummy variables: for example, dummy variables pertaining to shoulder widths of 4 ft, 6 ft, 8 ft, and 10 ft, as well as dummy variables corresponding to lane widths of 10 ft, 11 ft, and 12 ft. However, in order for this to be feasible, sample roadway segments containing all these combinations of shoulder and lane widths, as well any other highway design characteristic desired, should be retrieved and included in the statistical analysis. Nonetheless, at this point CMFs can be incorporated in the suggested 3-D models, creating hybrid 3-D SPFs. For example, since this research was conducted for 11 ft lane width roadways, CMFs can be utilized in order to adjust the crash predictions for different lane widths accordingly; in other words, at this point, 11 ft lane widths function as the baseline conditions, as defined in the HSM terminology, of the analysis. . In general, more explanatory variables of both highway design, e.g. roadside characteristics, and 3-D geometry oriented, e.g., length of geodesic curves, 3-D stopping sight distance, should enter the model. Finally, the rationale, results, and findings of this research can be integrated in the current highway design practices, e.g. IHSDM software, in order to enhance them, by adding/incorporating the 3-D metrics used here into existing SPFs.

The severity type of the crash should be incorporated in the regression models. In this case, small sample size related issues will surely arise, but research should move towards this direction even if it requires waiting some years in order for more crash data to be accumulated. A more detailed breakdown in terms of crash severity type, would eventually allow for a more detailed crash cost estimation since, as one would expect, different crash costs are associated to different crash types (FHWA, n.d.d).

The ultimate objective of this research would be to develop a user friendly and interactive software/tools that would be able to be incorporated in highway design software programs, e.g., Autocad Civil 3D, Microstation, in order to assist in the design process. This tool would be particularly useful for the evaluation and comparison of alternative/competing highway geometric alignments. In addition, this tool would be able to express the evaluation of alternative alignments not only in terms of increase/decrease of crashes, but also in terms of the associated crash cost. Finally, this system could be integrated in the GPS screen of vehicles in order to warn driving when driving on roadway segments in which the crash occurrence probability calls for proportionally more attention.

APPENDIX A: Gaussian & Mean Curvature

A surface S can be defined by the two parameters u, v, in the following format:

$$S(u,v) = (S^1(u,v), S^2(u,v), S^3(u,v))$$

Partial derivatives are developed in terms of *u* and *v*, since they are the two parameters that define a surface. These derivatives are defined below:

$$S_u = \left(\frac{\partial S^1(u,v)}{\partial u}, \frac{\partial S^2(u,v)}{\partial u}, \frac{\partial S^3(u,v)}{\partial u}\right) = \left(S_u^1(u,v), S_u^2(u,v), S_u^3(u,v)\right)$$

$$S_v = \left(\frac{\partial S^1(u,v)}{\partial v}, \frac{\partial S^2(u,v)}{\partial v}, \frac{\partial S^3(u,v)}{\partial v}\right) = \left(S_v^1(u,v), S_v^2(u,v), S_v^3(u,v)\right)$$

$$S_{uu} = \left(\frac{\partial^2 S^1(u,v)}{\partial u^2}, \frac{\partial^2 S^2(u,v)}{\partial u^2}, \frac{\partial^2 S^3(u,v)}{\partial u^2}\right) = \left(S_{uu}^1(u,v), S_{uu}^2(u,v), S_{uu}^3(u,v)\right)$$

$$S_{uv} = \left(\frac{\partial^2 S^1(u,v)}{\partial u\, \partial v}, \frac{\partial^2 S^2(u,v)}{\partial u\, \partial v}, \frac{\partial^2 S^3(u,v)}{\partial u\, \partial v}\right) = \left(S_{uv}^1(u,v), S_{uv}^2(u,v), S_{uv}^3(u,v)\right)$$

$$S_{vv} = \left(\frac{\partial^2 S^1(u,v)}{\partial v^2}, \frac{\partial^2 S^2(u,v)}{\partial v^2}, \frac{\partial^2 S^3(u,v)}{\partial v^2}\right) = \left(S_{vv}^1(u,v), S_{vv}^2(u,v), S_{vv}^3(u,v)\right)$$

The building blocks in order to define the Gaussian and Mean Curvature are the metrics of the so-called First and Second Fundamental Form; each form consists of three metrics. The three metrics, namely: E, F, G, of the First Fundamental Form are defined in Equations 7-9:

$$E = S_u \cdot S_u \tag{7}$$

$$F = S_u \cdot S_v \tag{8}$$

$$G = S_v \cdot S_v \tag{9}$$

It should be noted that the dot symbol " $\cdot$ " indicates the vector inner product.

In order to define the metrics of the Second Fundamental Form, the unit normal vector N must be initially defined (Equation 10):

$$\vec{N} = \frac{S_u \times S_v}{|S_u \times S_v|} \tag{10}$$

where the symbol " $\times$ " indicates the vector cross product.

The three metrics, namely: L, M, N, of the Second Fundamental Form are defined in Equations 11-13:

$$L = S_{uu} \cdot \vec{N} \tag{11}$$

$$M = S_{uv} \cdot \vec{N} \tag{12}$$

$$N = S_{vv} \cdot \vec{N} \tag{13}$$

After the metrics of the First and Second Fundamental Form have been defined, the Gaussian Curvature K and Mean Curvature H can be in turn defined (Equations 14 & 15):

$$K = \frac{L\ N - M^2}{E\ G - F^2} \tag{14}$$

$$H = \frac{G\ L + E\ N - 2\ F\ M}{2\ (E\ G - F^2)} \tag{15}$$

Finally, the natural meaning of the Gaussian and Mean Curvature is briefly described.

At each point on a surface there are infinite possible directions that correspond to it. Loosely speaking, each direction corresponds to a specific surface curvature. However there are two special curvatures that are called principle curvatures and corresponds to the maximum k_1 and minimum k_2 curvature of a point on the surface. The Gaussian and Mean curvature are defined through these principles curvatures in a straightforward manner (Equations 16 and 17):

$$K = k_1 \, k_2 \tag{16}$$

$$H = \frac{k_1 + k_2}{2} \tag{17}$$

Therefore, the Gaussian curvature is essentially the interaction, i.e., product, of the principles curvatures whereas the Mean curvature is the average (this is why it is called Mean) of the principle curvatures.

APPENDIX B: Google Earth Images of Roadway Segments

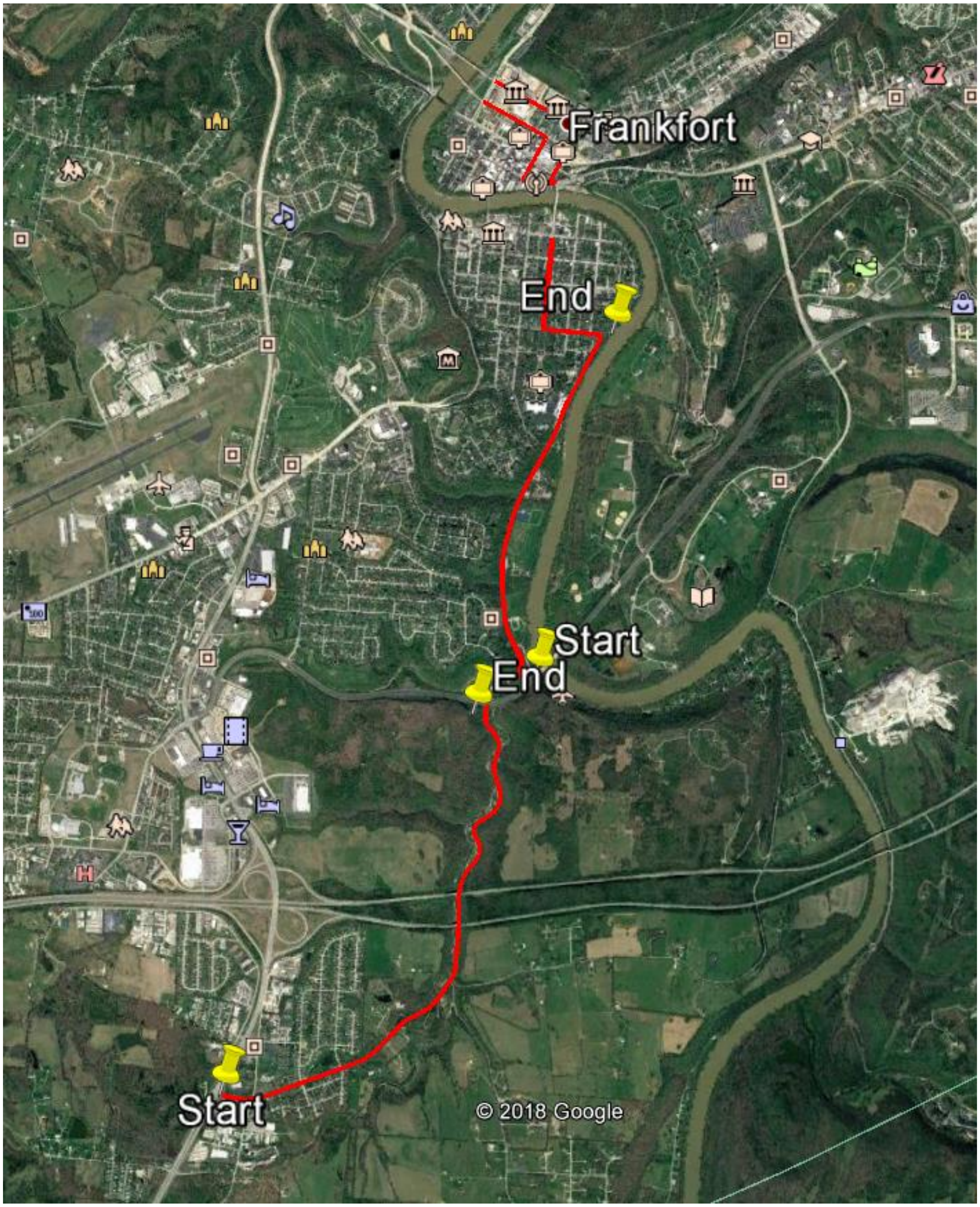

Figure B.1: Google Earth Image of KY 420

Figure B.2: Google Earth Image of KY 152

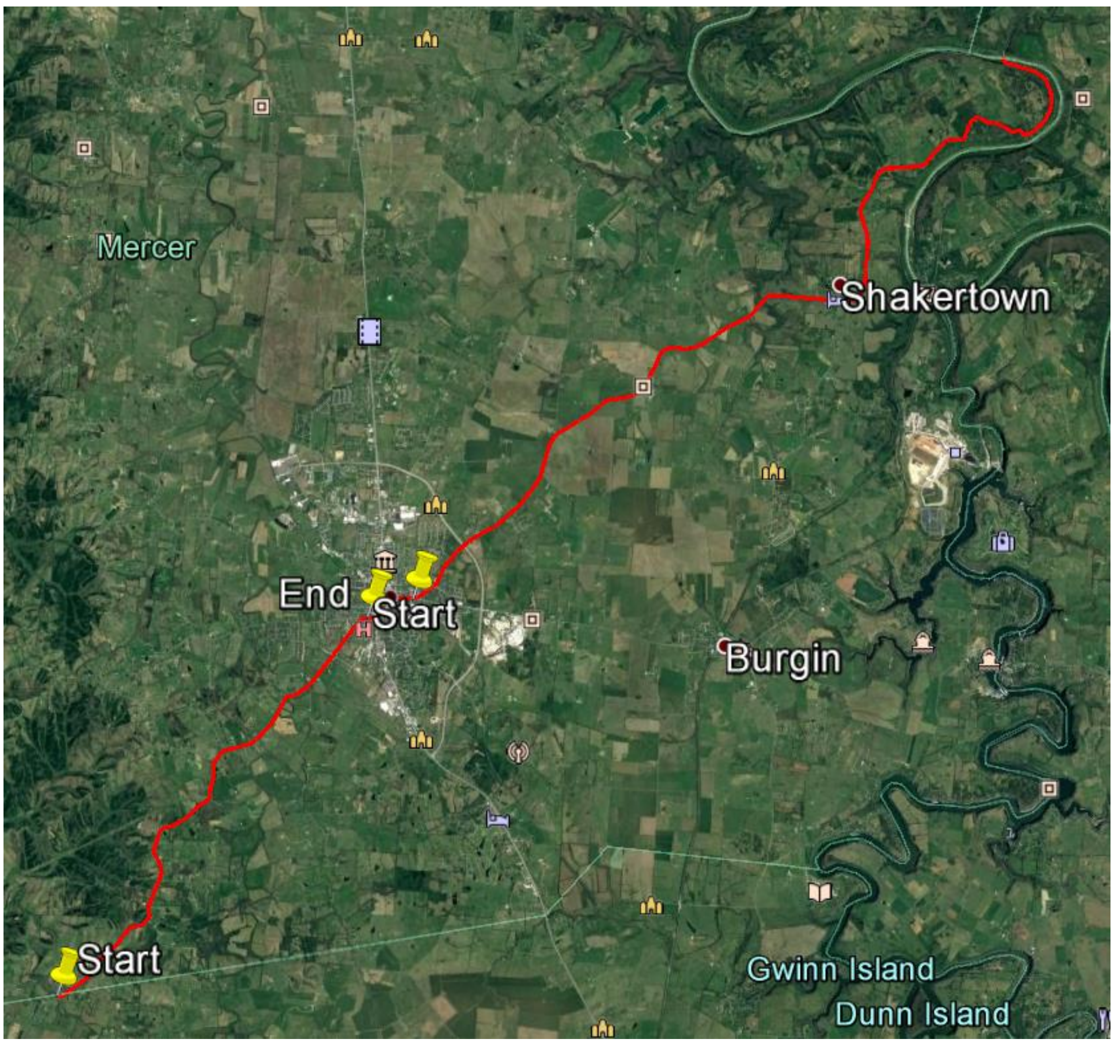

Figure B.3: Google Earth Image of US 68

APPENDIX C: Geometric Roadway Data & Modeling

KY 420 Roadway

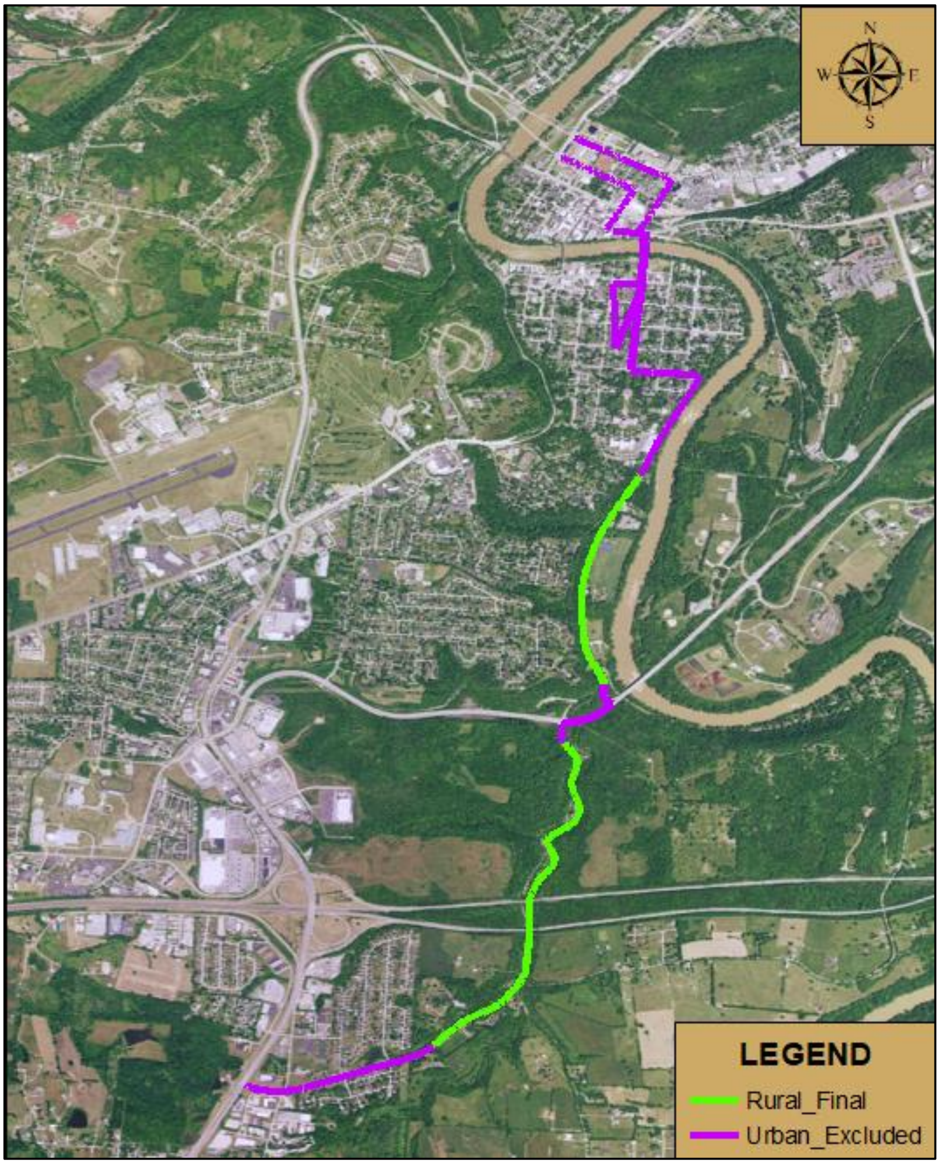

Figure C.1: Rural/Urban Distinction (KY 420)

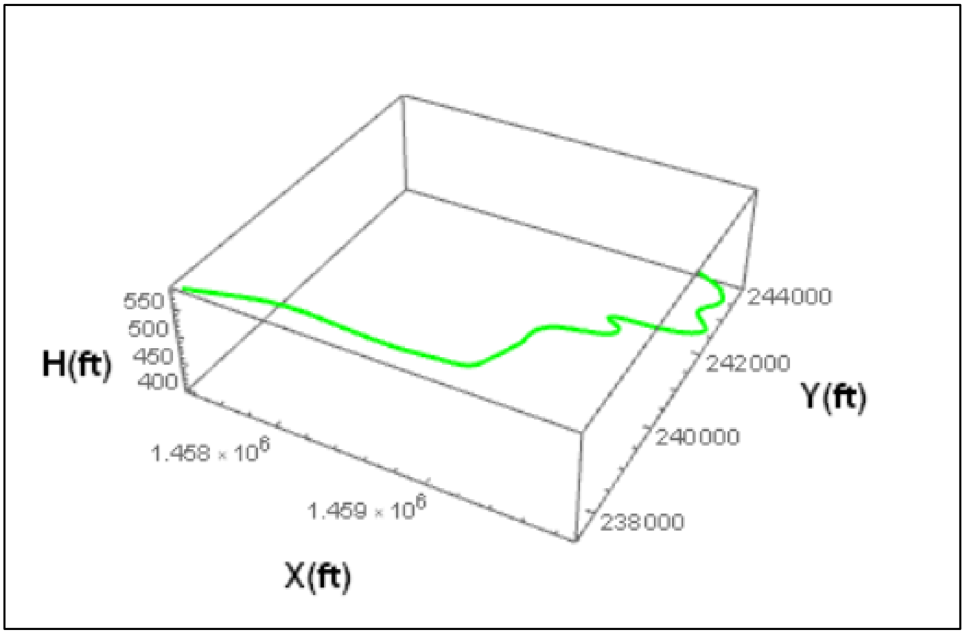

Figure C.2: 3-D B-Spline Road Centerline (KY 420-1)

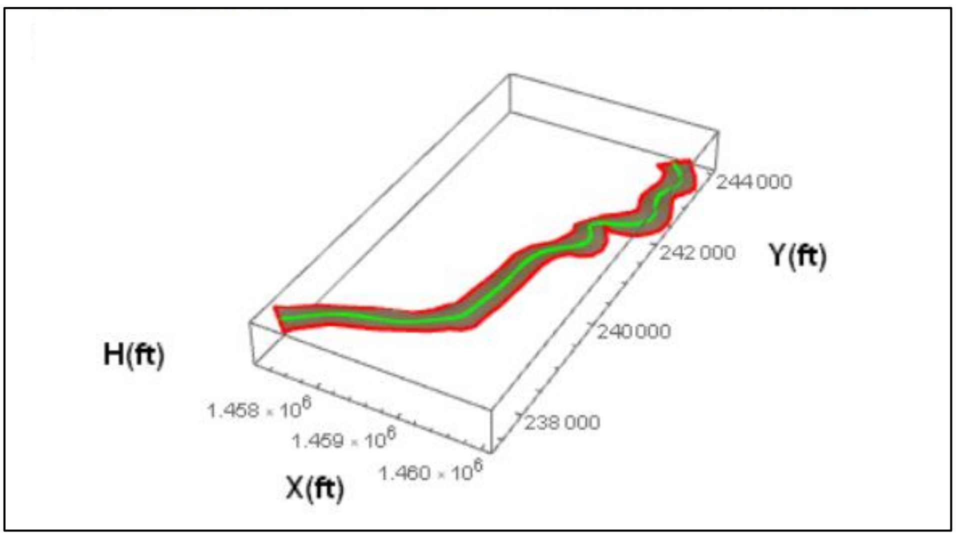

Figure C.3: 3-D B-Spline Road Surface (KY 420-1)

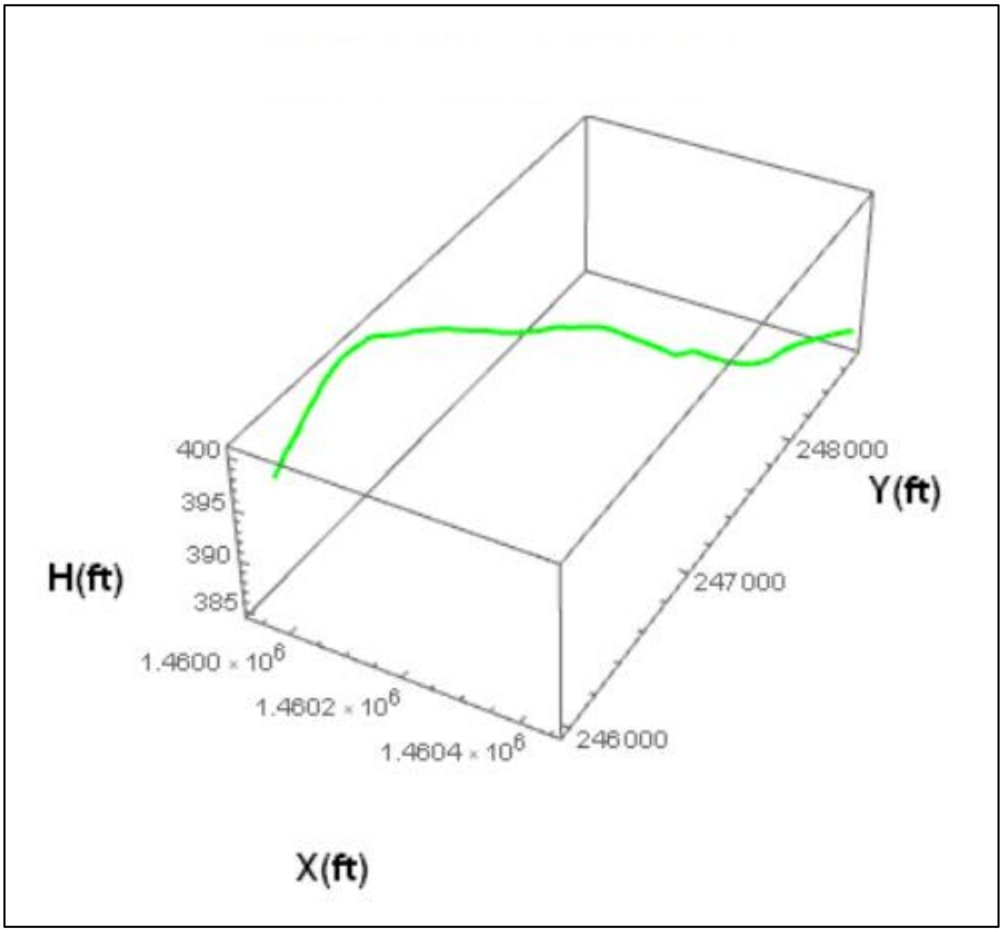

Figure C.4: 3-D B-Spline Road Centerline (KY 420-2)

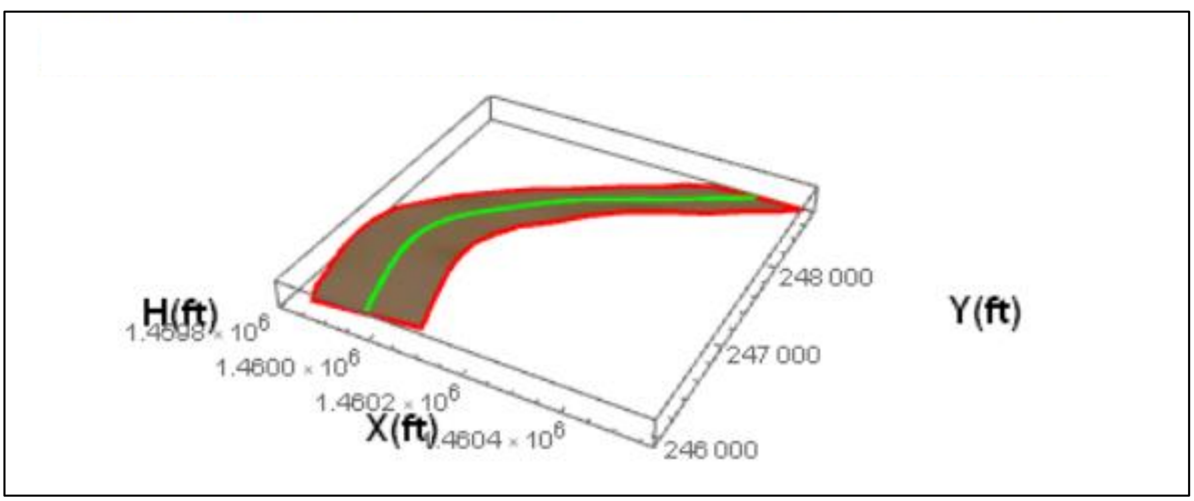

Figure C.5: 3-D B-Spline Road Surface (KY 420-2)

KY 152 Roadway

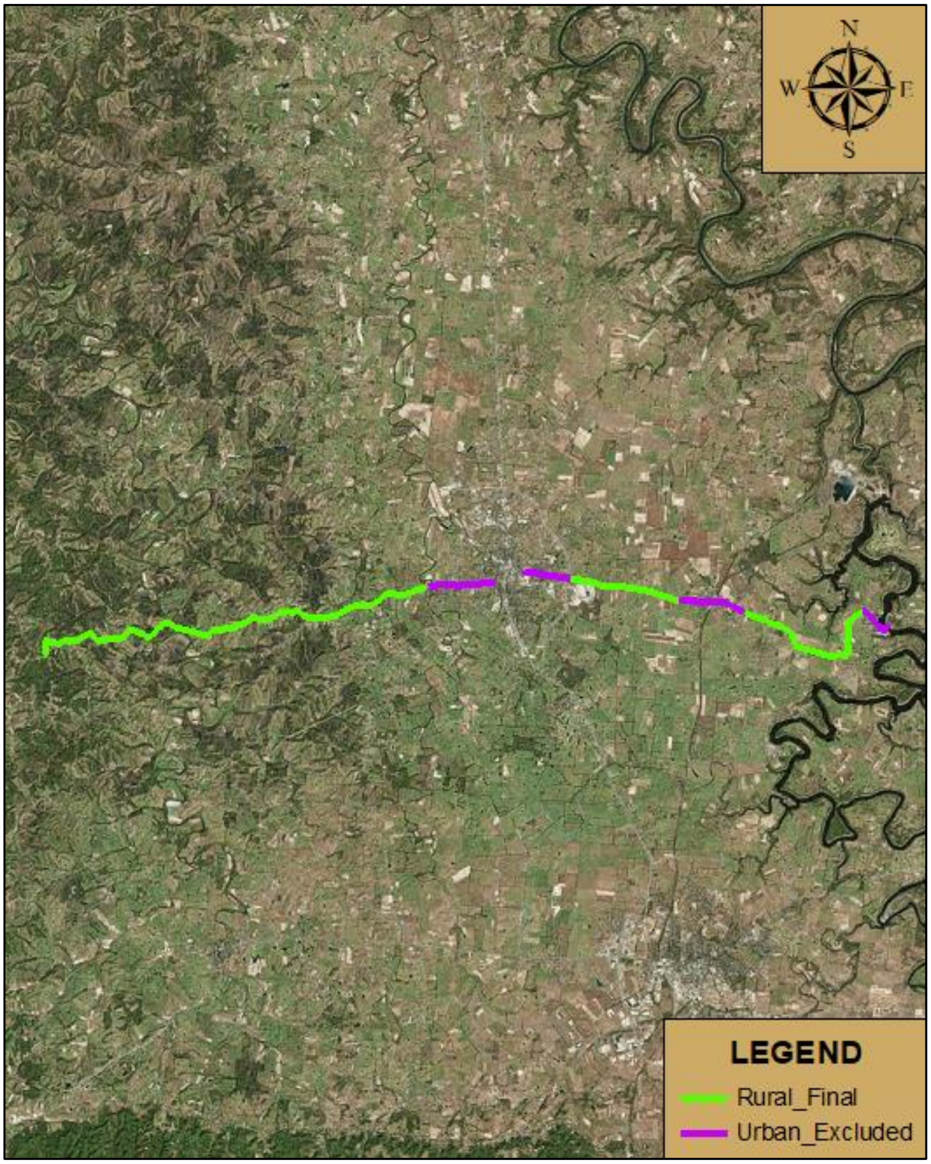

Figure C.6: Rural/Urban Distinction (KY 152)

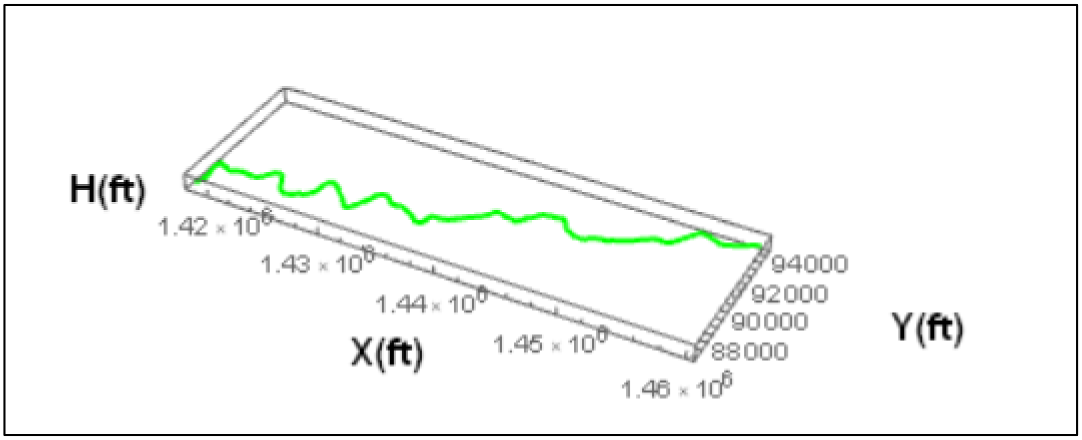

Figure C.7: 3-D B-Spline Road Centerline (KY 152-1)

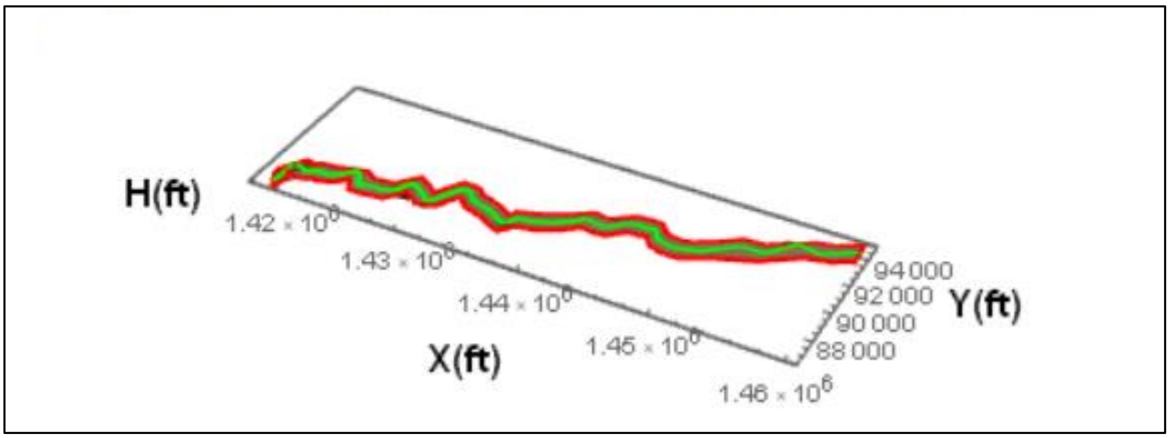

Figure C.8: 3-D B-Spline Road Surface (KY 152-1)

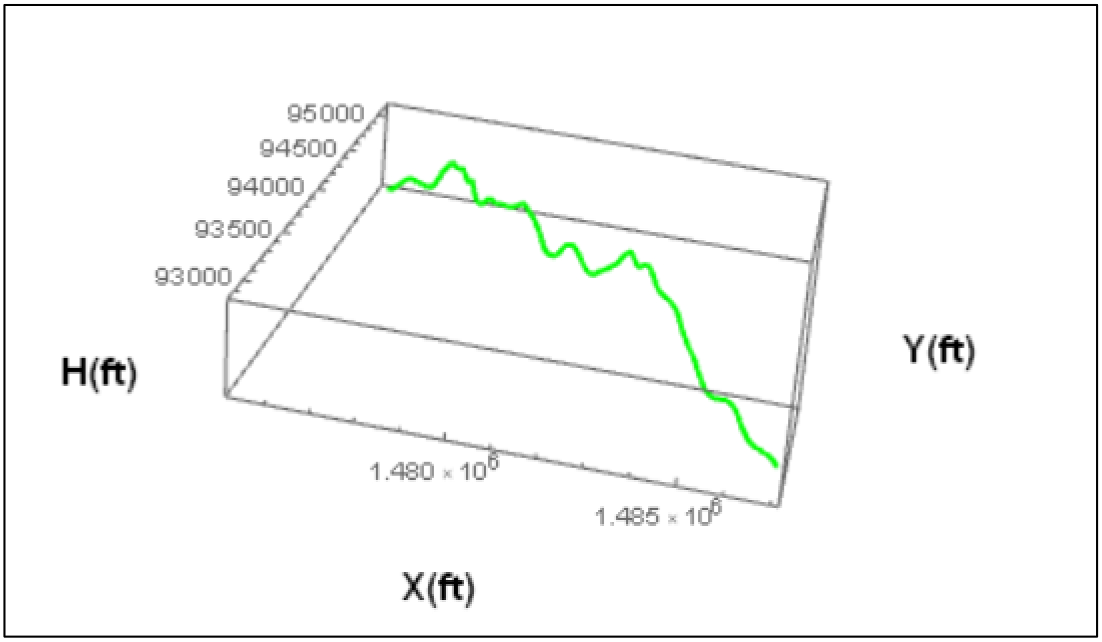

Figure C.9: 3-D B-Spline Road Centerline (KY 152-2)

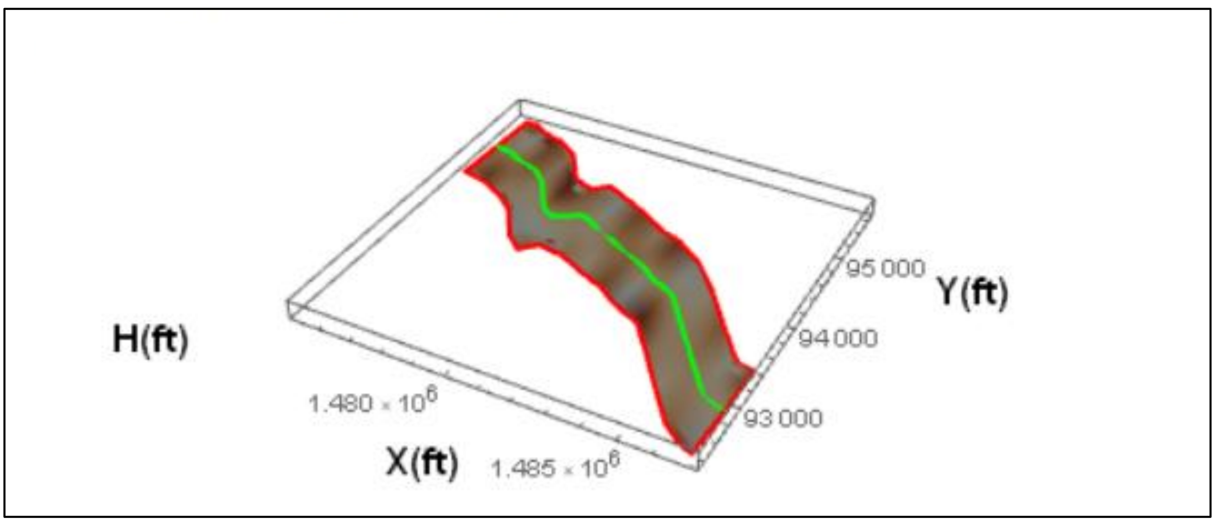

Figure C.10: 3-D B-Spline Road Surface (KY 152-2)

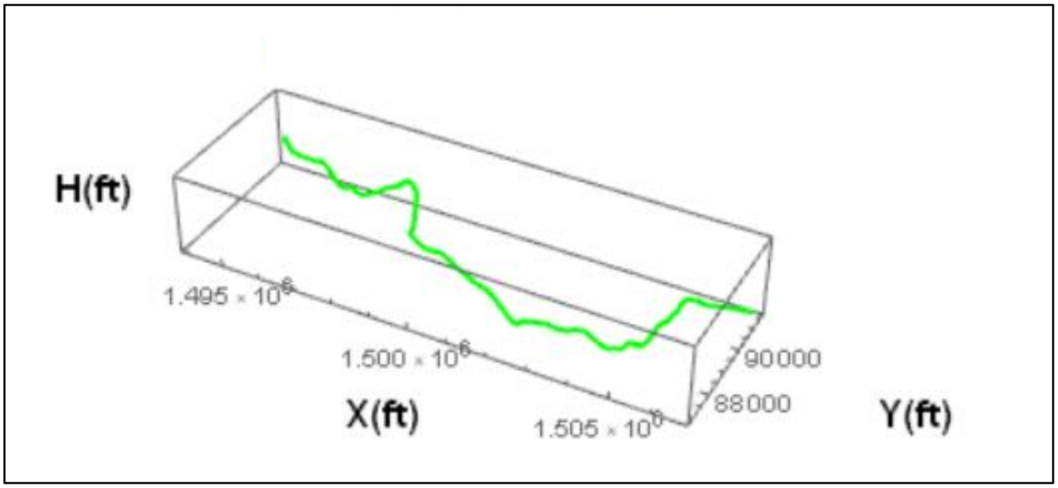

Figure C.11: 3-D B-Spline Road Centerline (KY 152-3)

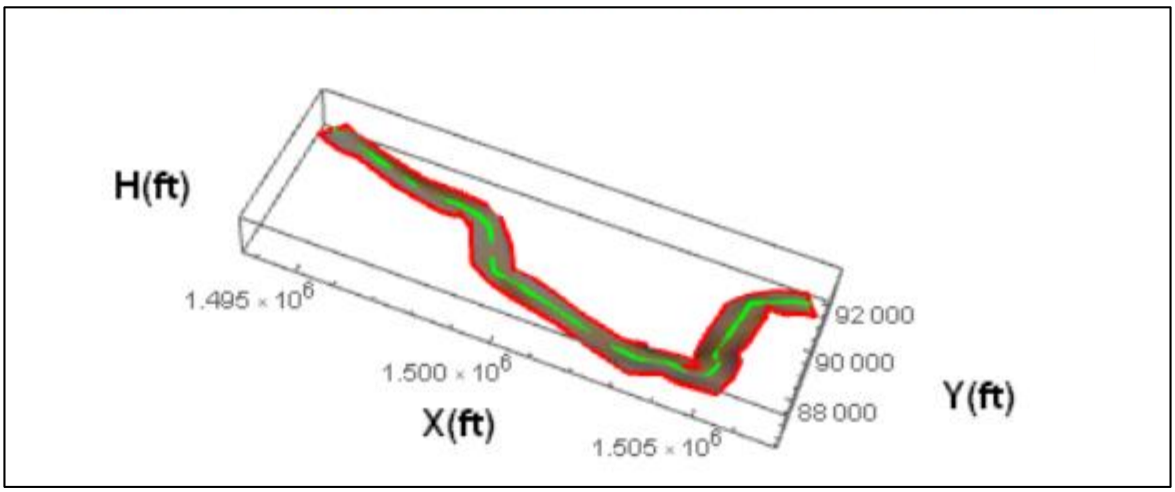

Figure C.12: 3-D B-Spline Road Surface (KY 152-3)

US 68 Roadway

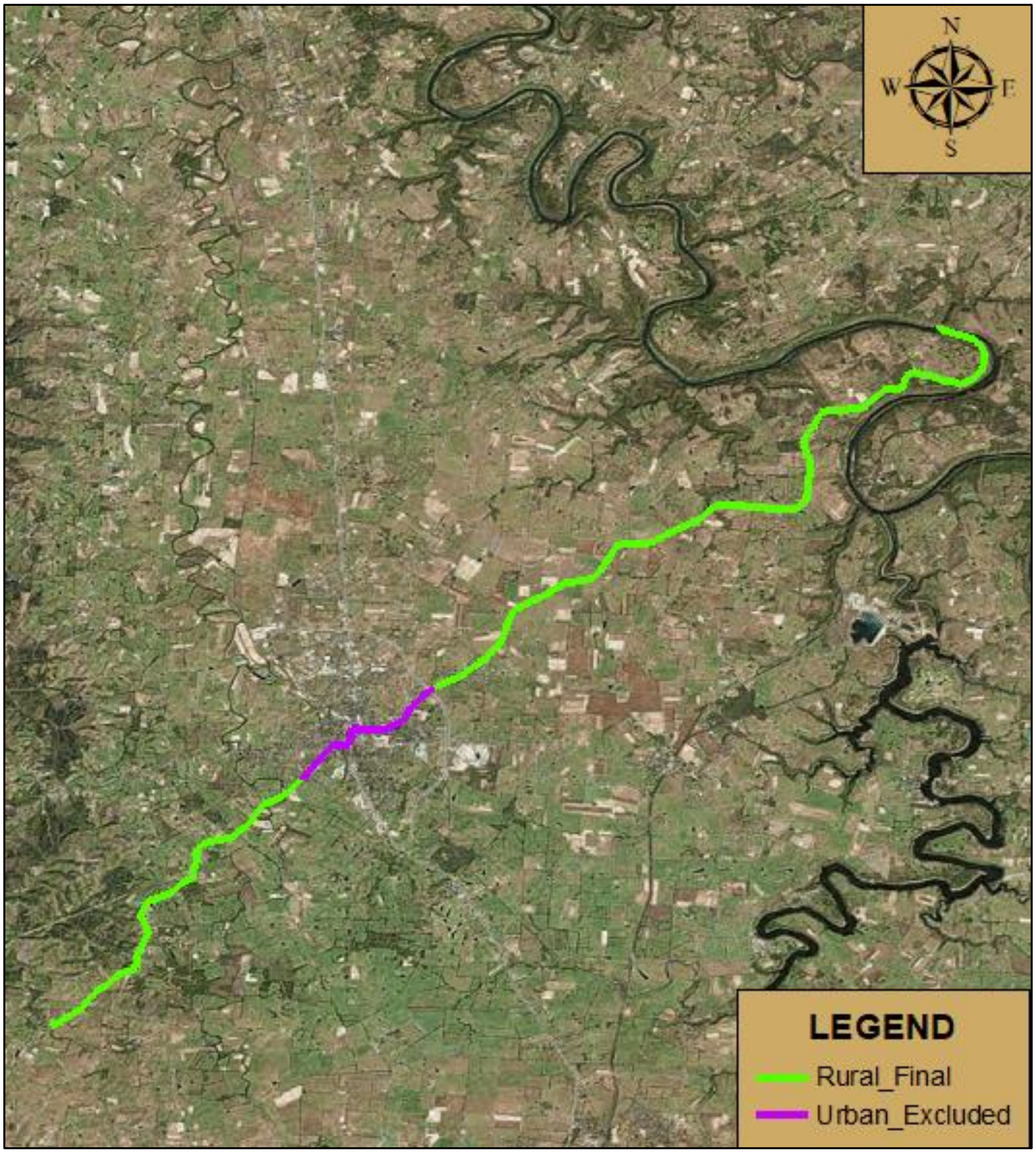

Figure C.13: Rural/Urban Distinction (US 68)

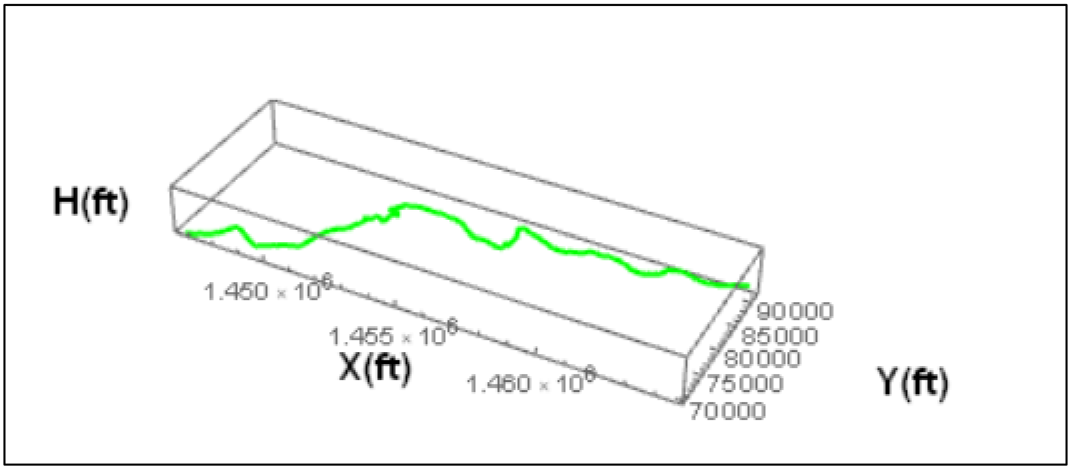

Figure C.14: 3-D B-Spline Road Centerline (US 68-1)

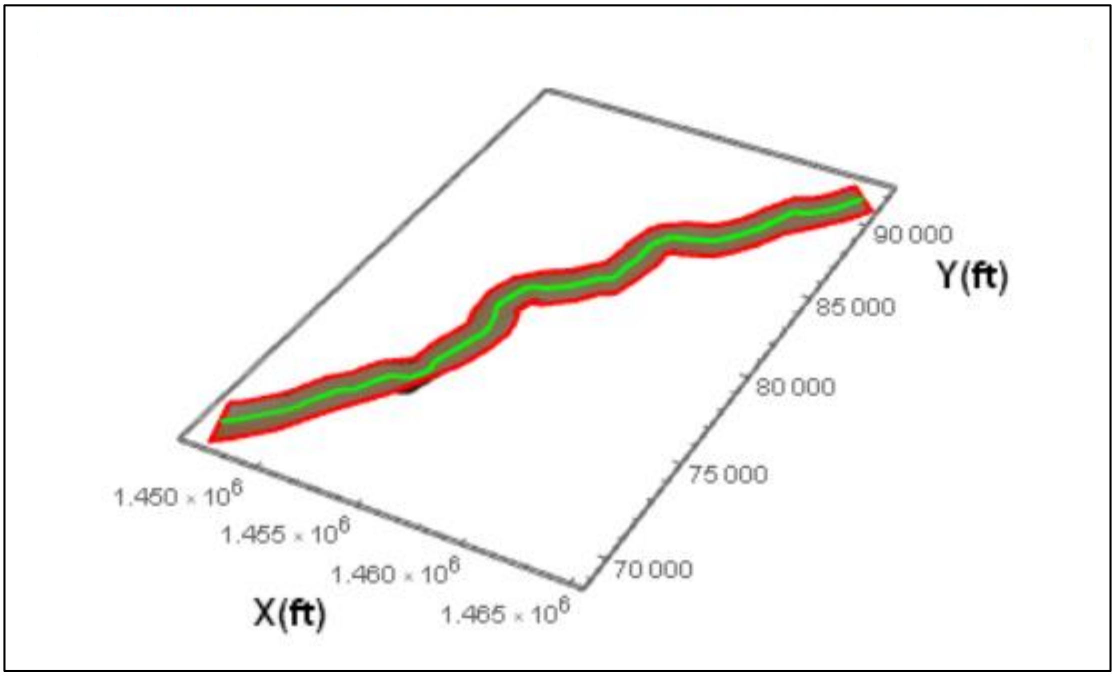

Figure C.15: 3-D B-Spline Road Surface (US 68-2)

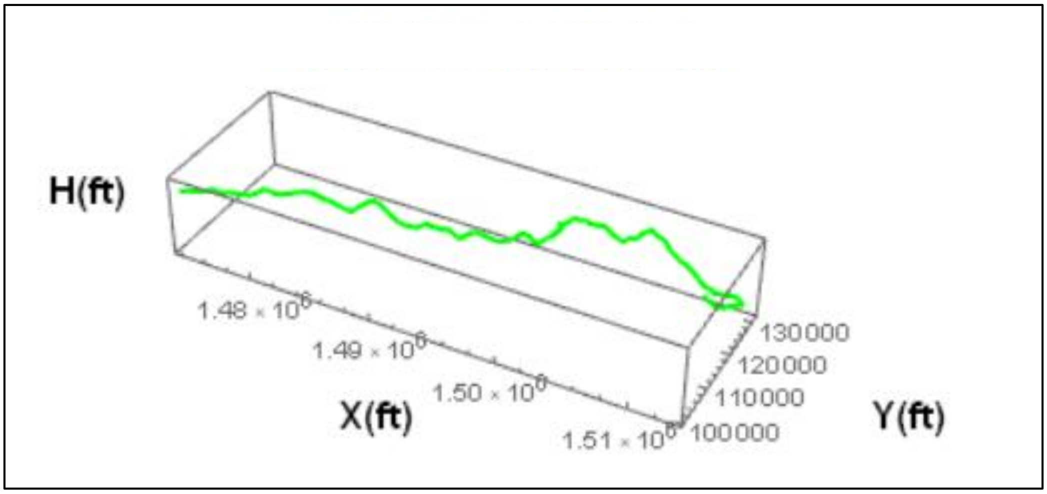

Figure C.16: 3-D B-Spline Road Centerline (US 68-2)

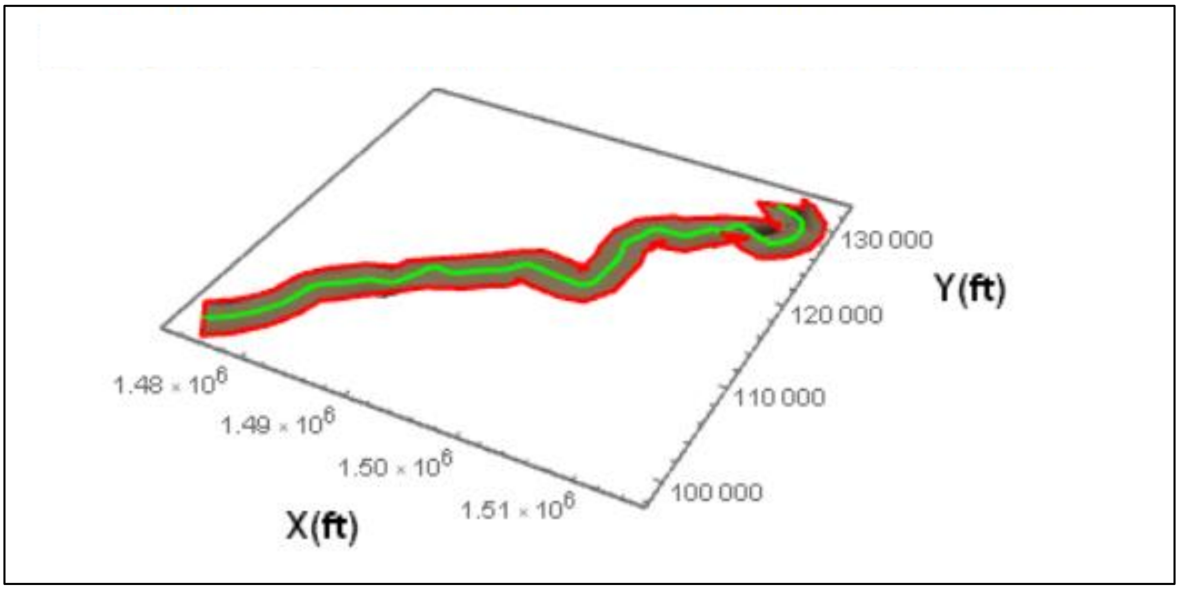

Figure C.17: 3-D B-Spline Road Surface (US 68-2)

APPENDIX D: Crash Data Plots

Figure D.1: Crash Plots, KY 420, 2004 to 2017

Figure D.2: Crash Plots for years 2004 to 2017 (KY 152)

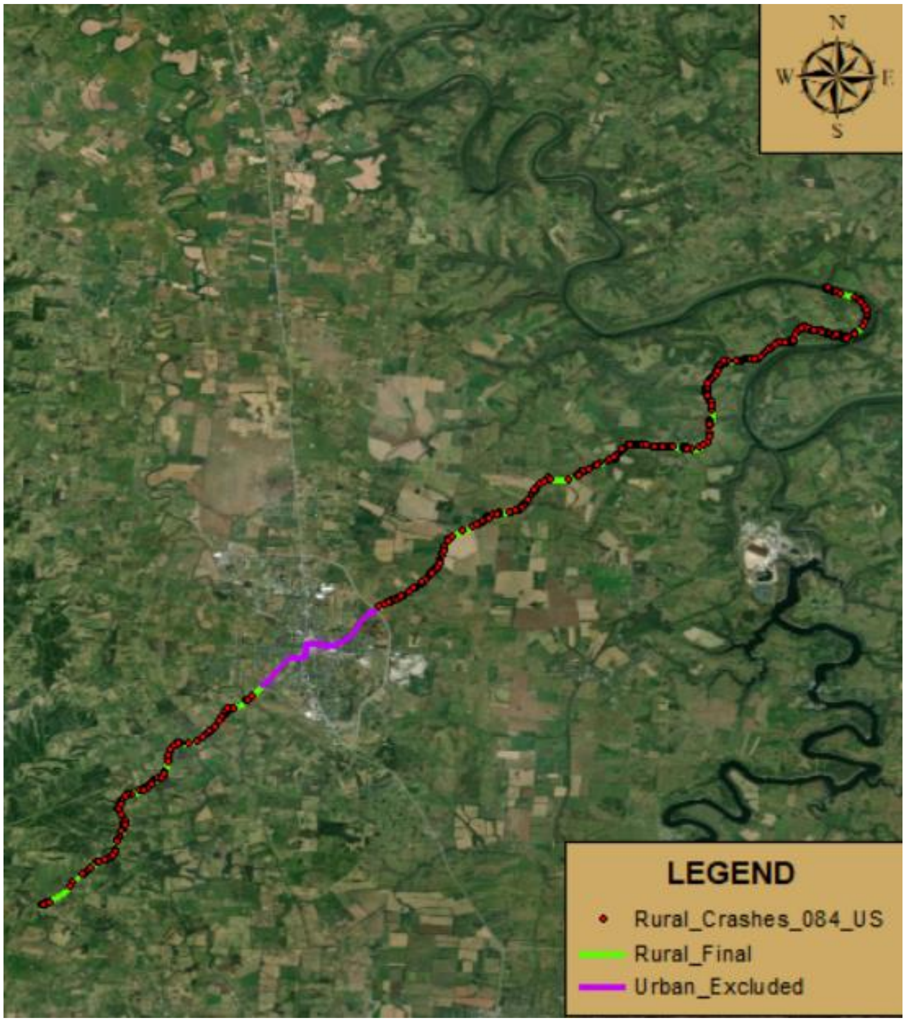

Figure D.3: Crash Plots for years 2004 to 2017 (US 68)

APPENDIX E: AADT Stations and Data

The starting latitude and longitude coordinates to which the interactive map should be zoomed to are presented in Table E.1 for each roadway, whereas the specific AADT data from each station are shown in Figures E.1-E.12. It is worth mentioning that 12 AADT stations come into play in total for the three roadways under study.

Table E.1: Starting Latitude and Longitude Coordinates for All Roadways

	Starting Coordinates	
Roadway	**Latitude**	**Longitude**
KY 420	38.1480066	-84.8972377
KY 152	37.73623318	-85.0163928
US 68	37.6899154	-84.91889505

Year	AADT	Year	AADT	Year	AADT
2018		2008	5110	1998	
2017	5716	2007		1997	
2016		2006		1996	
2015		2005	5220	1995	
2014	5351	2004		1994	4400
2013		2003		1993	
2012		2002	5130	1992	
2011	4830	2001		1991	
2010		2000	4700	1990	
2009		1999		1989	

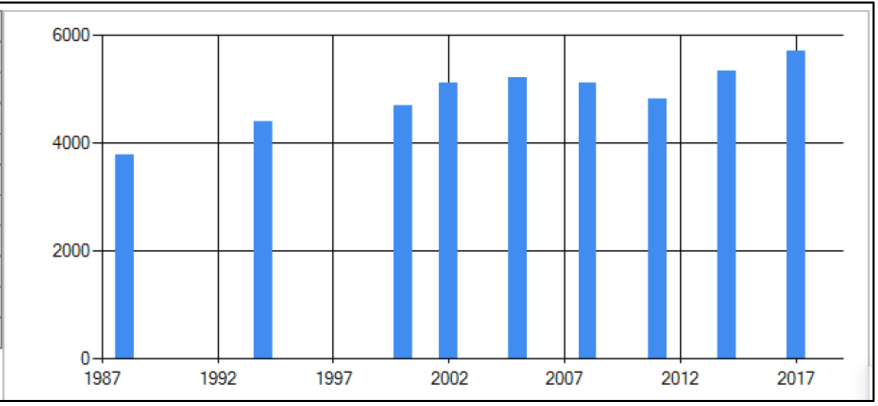

Figure E.1: AADT Data for Station ID# 037553, KY 420-1

Year	AADT	Year	AADT	Year	AADT
2018		2008		1998	4970
2017		2007		1997	
2016		2006	4470	1996	
2015	4696	2005		1995	
2014		2004		1994	
2013		2003		1993	
2012	3717	2002		1992	3800
2011		2001	4810	1991	
2010		2000		1990	
2009	3940	1999		1989	

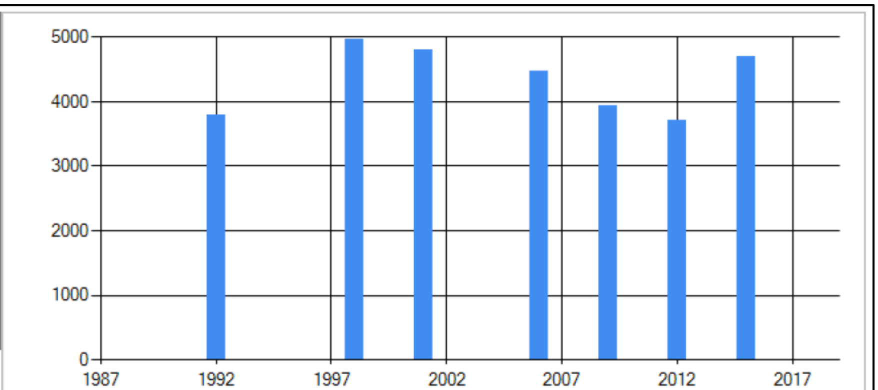

Figure E.2: AADT Data for Station ID# 037A20,KY 420-2

Year	AADT	Year	AADT	Year	AADT
2018		2008	492	1998	
2017	439	2007		1997	
2016		2006		1996	417
2015		2005	640	1995	
2014	478	2004		1994	
2013		2003		1993	
2012		2002	672	1992	
2011	440	2001		1991	
2010		2000		1990	564
2009		1999		1989	

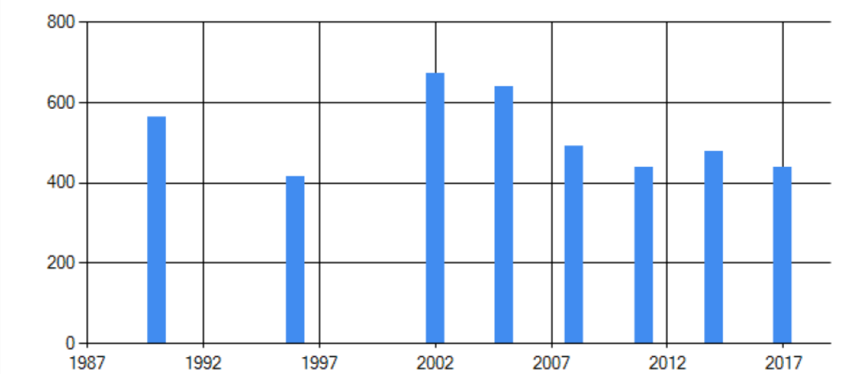

Figure E.3: AADT Data for Station ID# 084507, KY 152-1

Year	AADT	Year	AADT	Year	AADT
2018		2008		1998	
2017		2007		1997	
2016		2006	931	1996	1040
2015		2005		1995	
2014		2004		1994	
2013		2003		1993	
2012	671	2002	845	1992	
2011		2001		1991	
2010		2000		1990	912
2009	539	1999		1989	

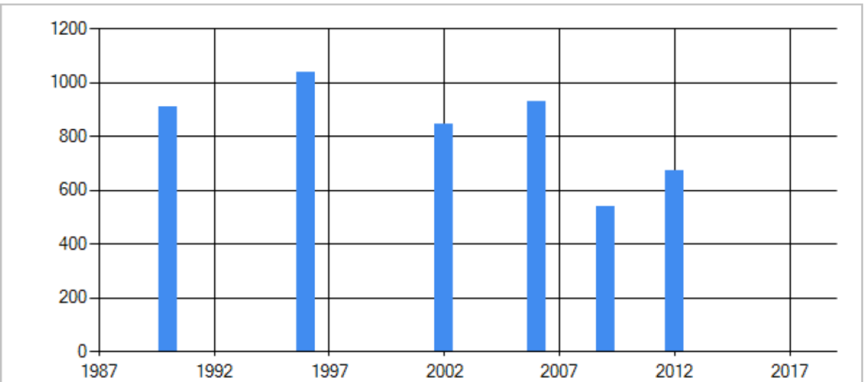

Figure E.4: AADT Data for Station ID# 084570,KY 152-1

Year	AADT	Year	AADT	Year	AADT
2018		2008	2770	1998	
2017	2358	2007		1997	
2016		2006		1996	3020
2015		2005	2950	1995	
2014	2535	2004		1994	
2013		2003		1993	
2012		2002	2970	1992	
2011	2700	2001		1991	
2010		2000		1990	2580
2009		1999		1989	

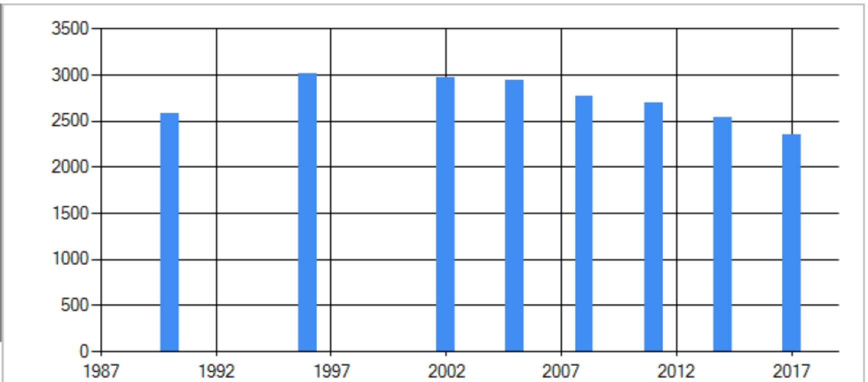

Figure E.5: AADT Data for Station ID# 084569,KY 152-1

Year	AADT	Year	AADT	Year	AADT
2018		2008		1998	
2017		2007	4130	1997	
2016	2865	2006		1996	
2015		2005		1995	4600
2014		2004	4380	1994	
2013	3522	2003		1993	
2012		2002		1992	
2011	4050	2001	3130	1991	
2010		2000		1990	
2009		1999		1989	3490

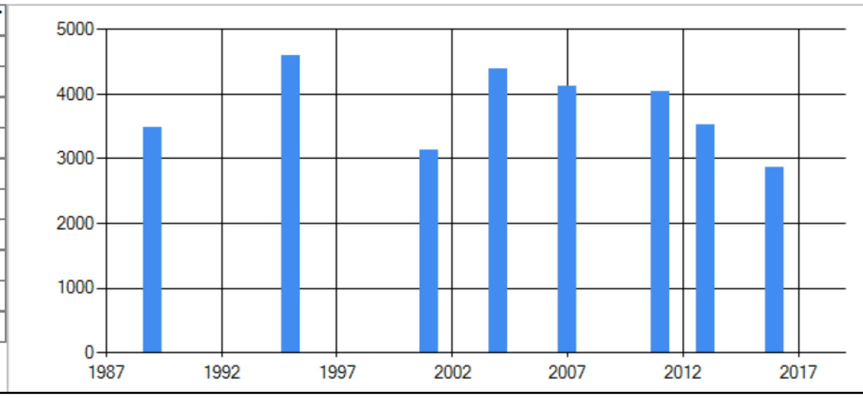

Figure E.6: AADT Data for Station ID# 084A45,KY 152-2

Year	AADT	Year	AADT	Year	AADT
2018		2008	2000	1998	
2017	1575	2007		1997	
2016		2006		1996	1730
2015		2005	1830	1995	
2014	1957	2004		1994	
2013		2003		1993	
2012		2002	1810	1992	
2011	2070	2001		1991	
2010		2000		1990	1610
2009		1999		1989	

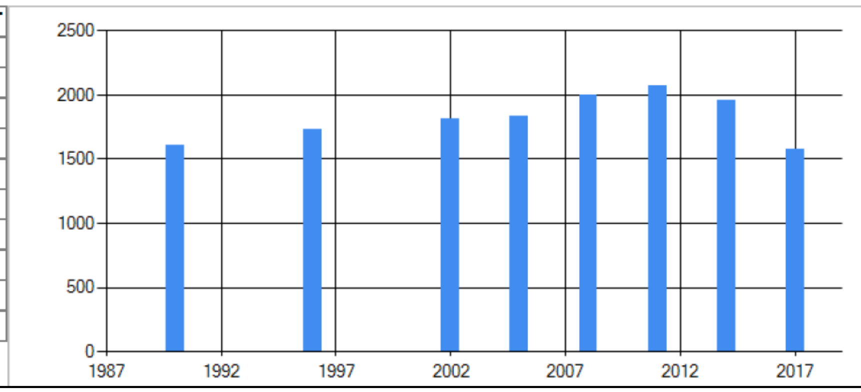

Figure E.7: AADT Data for Station ID# 084252, KY 152-3

Year	AADT	Year	AADT	Year	AADT
2018		2008	1690	1998	
2017	1506	2007		1997	
2016		2006		1996	1150
2015		2005	1520	1995	
2014	1423	2004		1994	
2013		2003		1993	
2012		2002	1840	1992	
2011	1270	2001		1991	
2010		2000		1990	1160
2009		1999		1989	

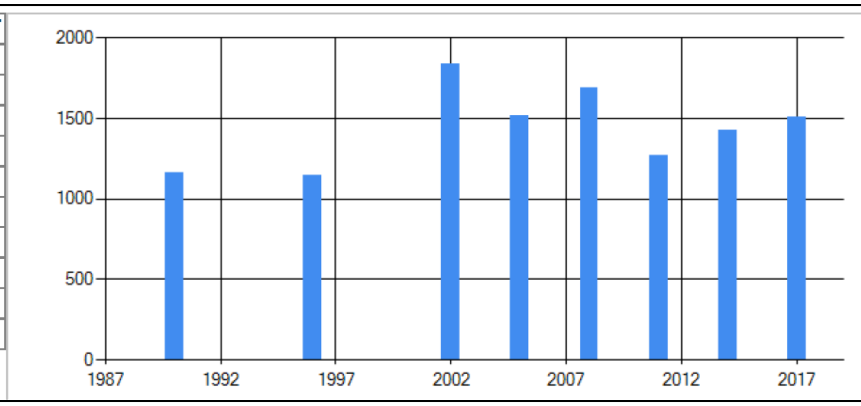

Figure E.8: AADT Data for Station ID# 084505, US 68-1

Year	AADT	Year	AADT	Year	AADT
2018		2008	2680	1998	
2017	2321	2007		1997	
2016		2006		1996	2380
2015		2005	2250	1995	
2014	2236	2004		1994	
2013		2003		1993	
2012	2255	2002	2660	1992	
2011		2001		1991	
2010		2000		1990	1960
2009		1999		1989	

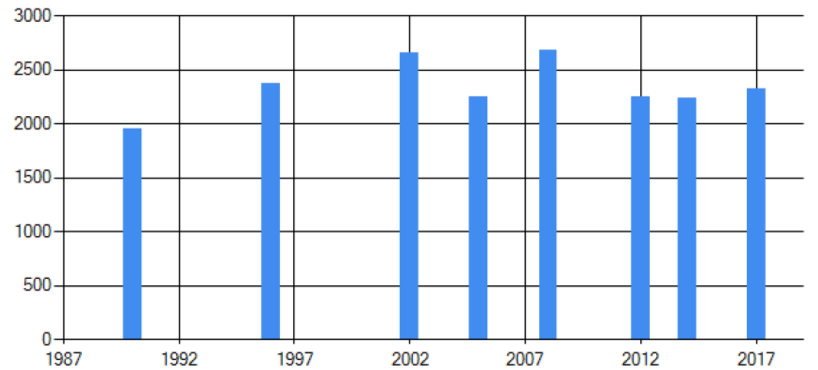

Figure E.9: AADT Data for Station ID# 084556, US 68-1

Year	AADT	Year	AADT	Year	AADT
2018		2008	3080	1998	4030
2017	2969	2007		1997	
2016		2006		1996	
2015		2005	3520	1995	3560
2014	3341	2004		1994	
2013		2003		1993	
2012		2002		1992	3940
2011	2860	2001	3670	1991	
2010		2000		1990	
2009		1999		1989	3450

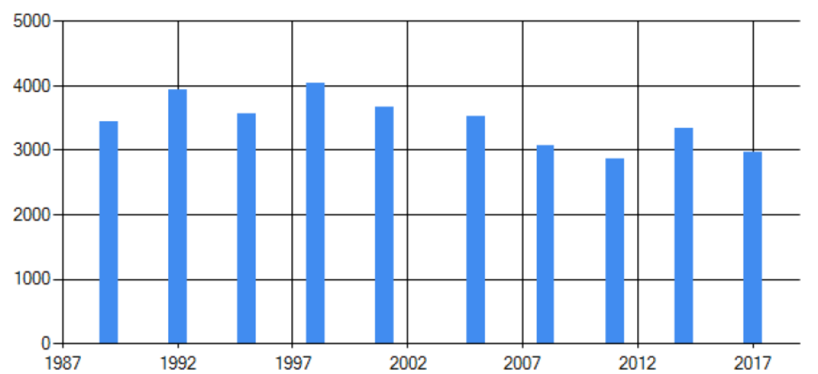

Figure E.10: AADT Data for Station ID# 084A50US 68-2

Year	AADT	Year	AADT	Year	AADT
2018		2008		1998	
2017		2007		1997	1810
2016		2006	2170	1996	
2015	2195	2005		1995	
2014		2004		1994	
2013		2003		1993	
2012	2321	2002	2080	1992	
2011		2001		1991	2090
2010		2000		1990	1920
2009	2180	1999		1989	

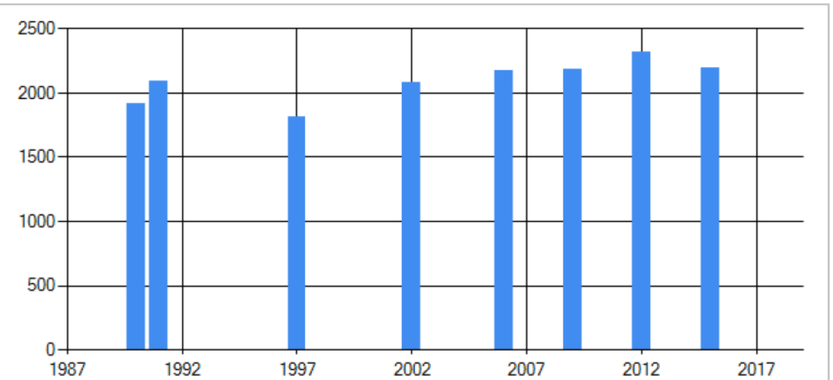

Figure E.11: AADT Data for Station ID# 084256, US 68-2

Year	AADT	Year	AADT	Year	AADT
2018		2008		1998	
2017		2007		1997	
2016		2006	2860	1996	2900
2015	2741	2005		1995	
2014		2004		1994	
2013		2003		1993	2770
2012	2942	2002	3010	1992	
2011		2001		1991	
2010		2000		1990	2340
2009	2850	1999	4620	1989	

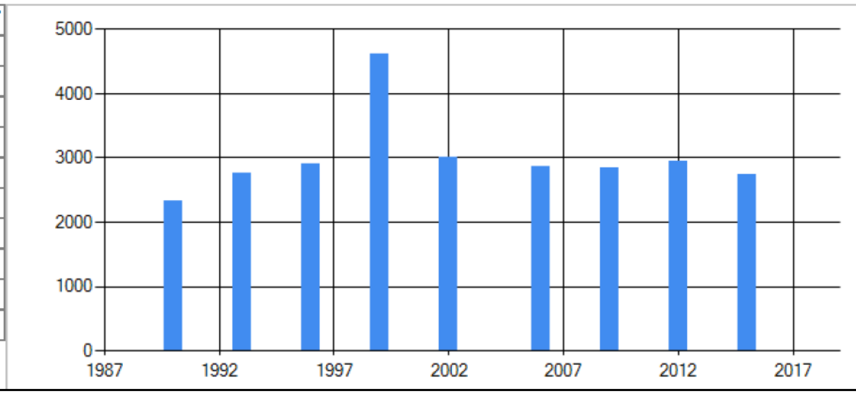

Figure E.12: AADT Data for Station ID# 084001,US 68-2

APPENDIX F: Statistical Analysis for Covariance Scale Determination

Patch Length = 1,500 ft

AADT for Patch Length = 1,500 ft

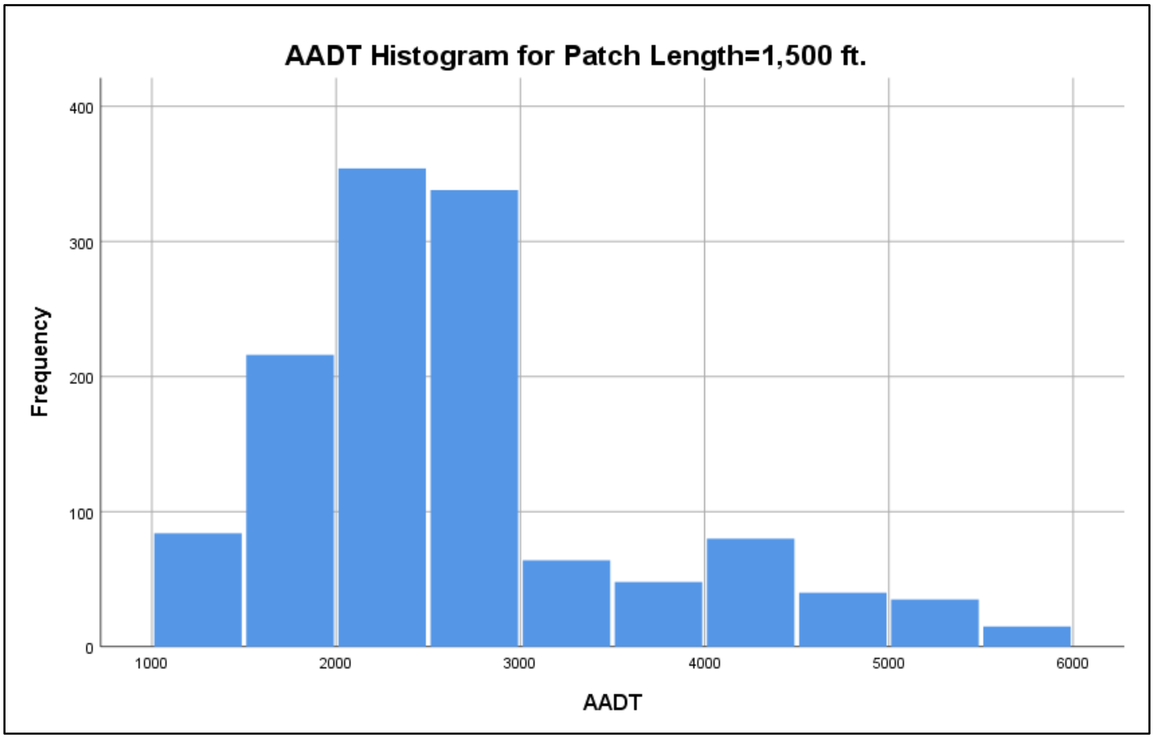

Figure F.1: AADT Histogram for Patch Length=1,500 ft

Table F.1: Descriptive Statistics of AADT for Patch Length=1,500 ft

Variable	**Descriptives**	**Statistic**	**Std. Error**
AADT	N	1,274	
	Mean	2,665	
	95% CI for Mean	[2,610 ÷ 2,719]	
	5% Trimmed Mean	2,595	
	Median	2,365	
	Variance	981,995	
	Std. Deviation	991	
	Minimum	1,270	
	Maximum	5,716	
	Range	4,446	
	Interquartile Range	907	
	Skewness	1.052	0.069
	Kurtosis	0.691	0.137

Table F.2: Mean AADT_Binned vs. LN(Number of Crashes)

Group ID#	Number of Crashes	LN (Number of Crashes)	Mean AADT_Binned
1	84	4.43081680	1,420
2	71	4.26267988	1,668
3	60	4.09434456	1,989
4	82	4.40671925	2,174
5	88	4.47733681	2,287
6	144	4.96981330	2,710
7	166	5.11198779	2,852
8	258	5.55295958	2,955
9	181	5.19849703	3,726
10	375	5.92692603	4,878

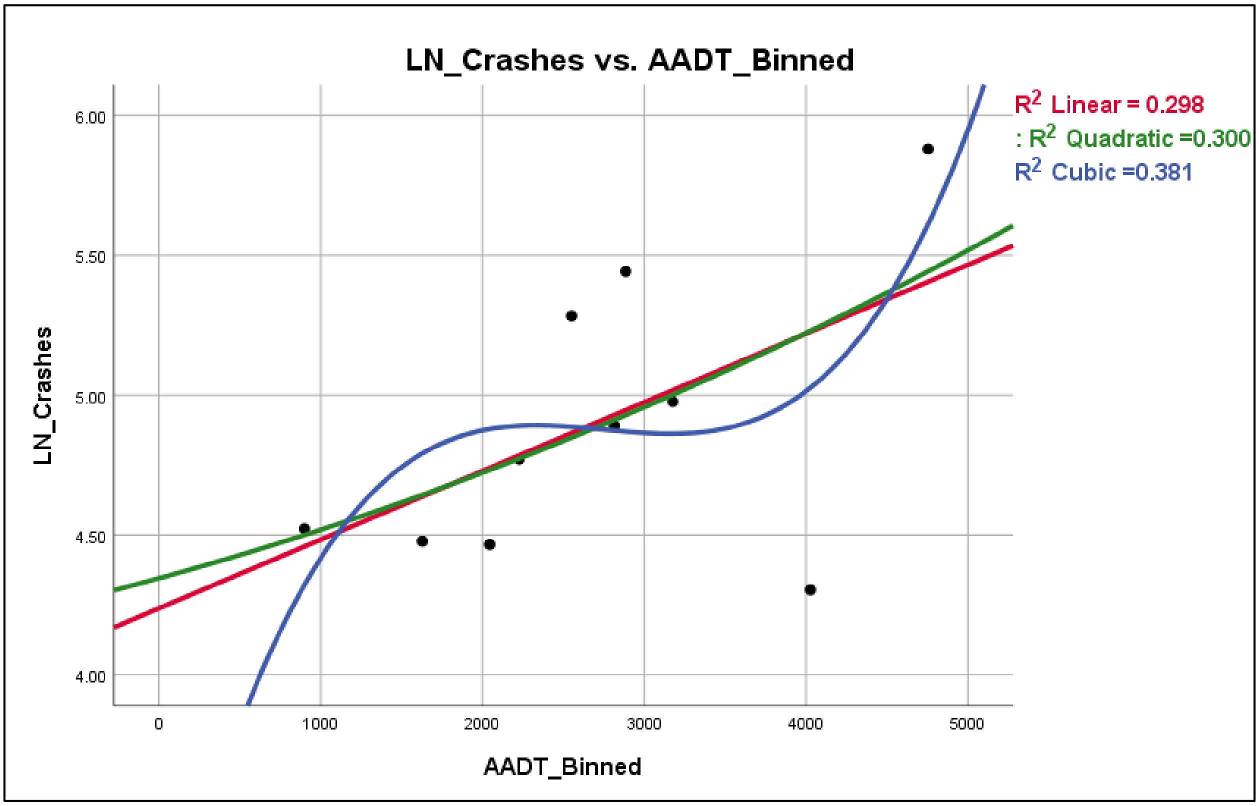

Figure F.2: Scatterplot LN_Crashes vs. AADT_Binned for Covariance Scale Determination

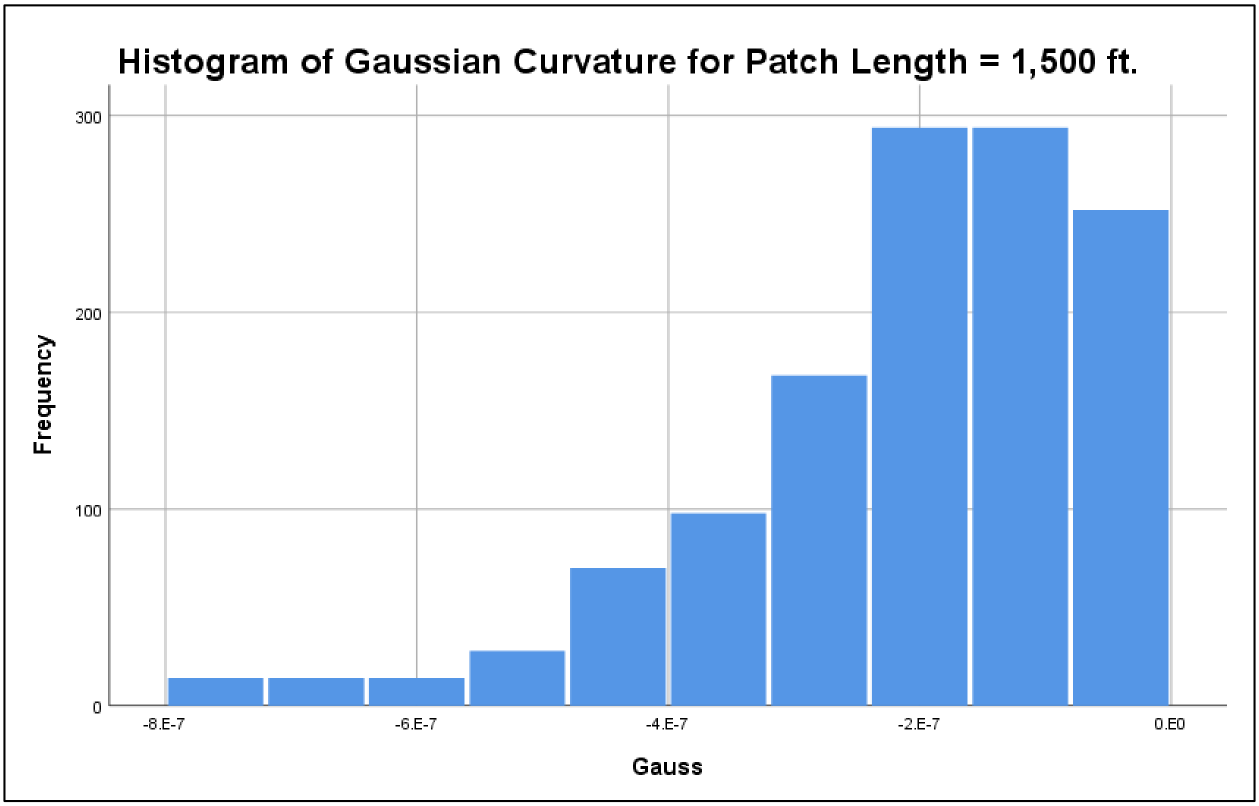

Figure F.3: Gaussian Curvature Histogram for Patch Length=1,500 ft

Table F.3: Descriptive Statistics of Gaussian Curvature for Patch Length=1,500 ft

Variable	Descriptives	Statistic	Std. Error
Gaussian Curvature	N	1,274	
	Mean	2,665	
	95% CI for Mean	[-2.143604E-7 ÷ -1.974639E-7]	
	5% Trimmed Mean	-1.93910583E-7	
	Median	-1.72806000E-7	
	Variance	2.31052338E-14	
	Std. Deviation	1.52004058E-7	
	Minimum	-7.81167000E-7	
	Maximum	-3.59207000E-9	
	Range	7.77574930E-7	
	Interquartile Range	1.94057000E-7	
	Skewness	-1.211	0.069
	Kurtosis	1.749	0.139

Table F.4: Mean Gaussian_Binned vs. LN(Number of Crashes)

Group ID#	Number of Crashes	LN (Number of Crashes)	Mean Gaussian_Binned
1	466	6.14418563	-0.000000531546888888888800
2	117	4.76217393	-0.000000356968444444444450
3	192	5.25749537	-0.000000292121111111111200
4	158	5.06259503	-0.000000224066555555555500
5	95	4.55387689	-0.000000181643111111111120
6	105	4.65396035	-0.000000160482444444444450
7	84	4.43081680	-0.000000128305333333333350
8	118	4.77068462	-0.000000091919155555555570
9	83	4.41884061	-0.000000053352277777777790
10	61	4.11087386	-0.000000017816168749999992

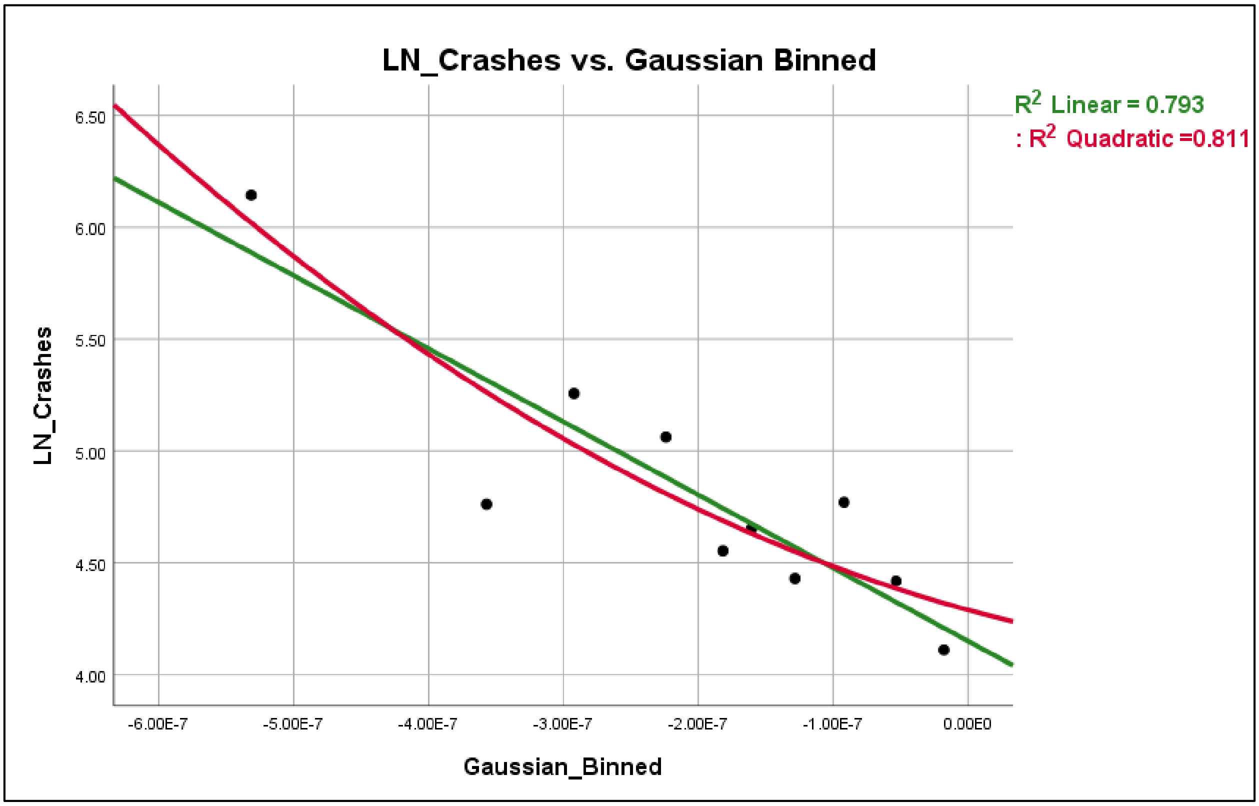

Figure F.4: Scatterplot LN_Crashes vs. Gaussian_Binned for Covariance Scale Determination

Mean Curvature for Patch Length = 1,500 ft

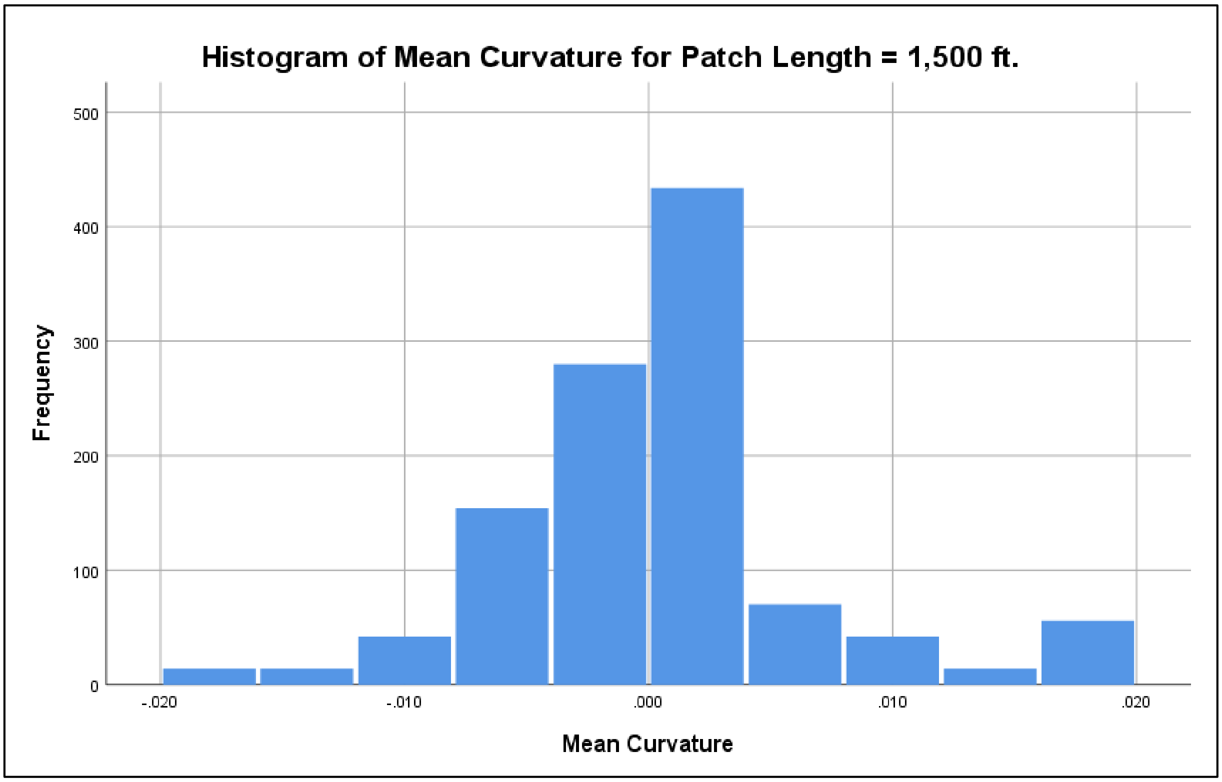

Figure F.5: Mean Curvature Histogram for Patch Length=1,500 ft

Table F.5: Descriptive Statistics of Mean Curvature for Patch Length=1,500 ft

Variable	Descriptives	Statistic	Std. Error
Mean Curvature	N	1,274	
	Mean	2,665	
	95% CI for Mean	[-0.000033462152 ÷ 0.000726867890]	
	5% Trimmed Mean	0.000085686688	
	Median	0.000029762700	
	Variance	0.000042	
	Std. Deviation	0.0064842996496	
	Minimum	-0.0168194980	
	Maximum	0.0196401340	
	Range	0.0364596320	
	Interquartile Range	0.0049812948	
	Skewness	0.733	0.073
	Kurtosis	1.965	0.146

Table F.6: Average Mean_Curvature_Binned vs. LN(Number of Crashes)

Group ID#	Number of Crashes	LN (Number of Crashes)	Average Mean_Curvature_Binned
1	74	4.30406509	-0.010582349625
2	136	4.91265489	-0.005152978000
3	89	4.48863637	-0.002833302875
4	90	4.49980967	-0.000425863638
5	226	5.42053500	0.000000024941
6	271	5.60211882	0.000065651263
7	142	4.95582706	0.000816421125
8	125	4.82831374	0.002343737125
9	52	3.95124372	0.004393670625
10	70	4.24849524	0.014842017750

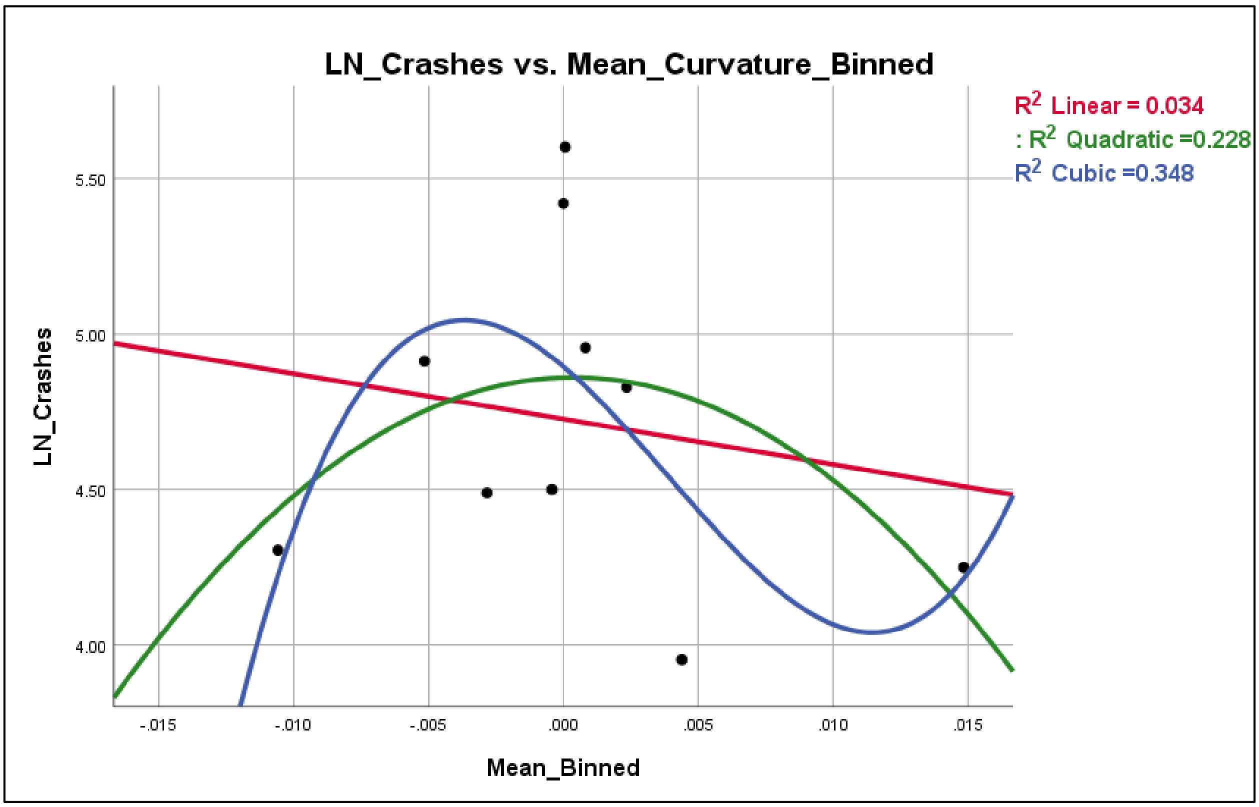

Figure F.6: Scatterplot LN_Crashes vs. Average_Mean_Binned for Covariance Scale Determination

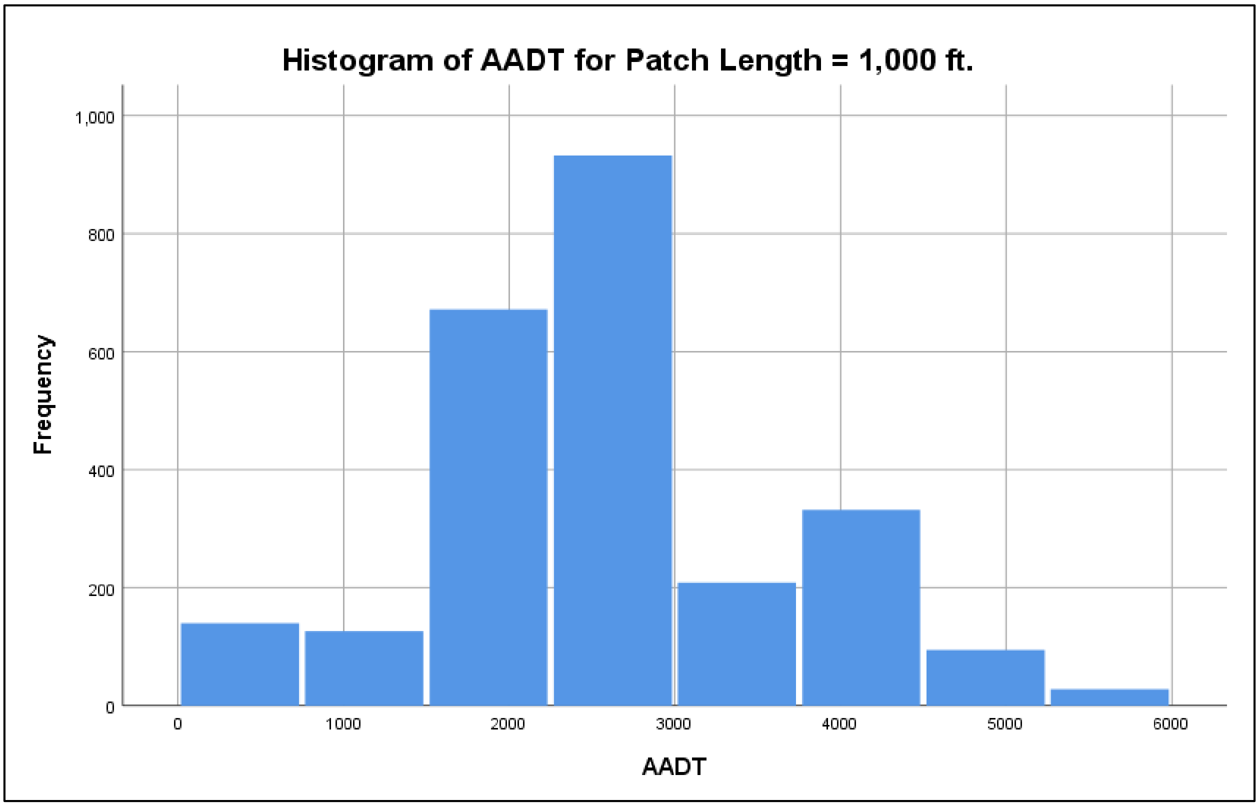

Figure F.7: AADT Histogram for Patch Length=1,000 ft

Table F.7: Descriptive Statistics of AADT for Patch Length=1,000 ft

Variable	**Descriptives**	**Statistic**	**Std. Error**
AADT	N	2,534	
	Mean	2,668	21.296
	95% CI for Mean	[2,626 ÷ 2,710]	
	5% Trimmed Mean	2,655	
	Median	2,741	
	Variance	1,149,224	
	Std. Deviation	1,072.0	
	Minimum	439	
	Maximum	5,716	
	Range	5,277	
	Interquartile Range	983.3	
	Skewness	0.325	0.049
	Kurtosis	0.126	0.097

Table F.8: Mean AADT_Binned vs. LN(Number of Crashes)

Group ID#	Number of Crashes	LN (Number of Crashes)	Mean AADT_Binned
1	93	4.53259949	915
2	89	4.48863637	1,628
3	85	4.44265126	2,032
4	120	4.78749174	2,221
5	196	5.27811466	2,555
6	133	4.89034913	2,815
7	230	5.43807931	2,885
8	146	4.98360662	3,174
9	73	4.29045944	4,024
10	335	5.81413053	4,740

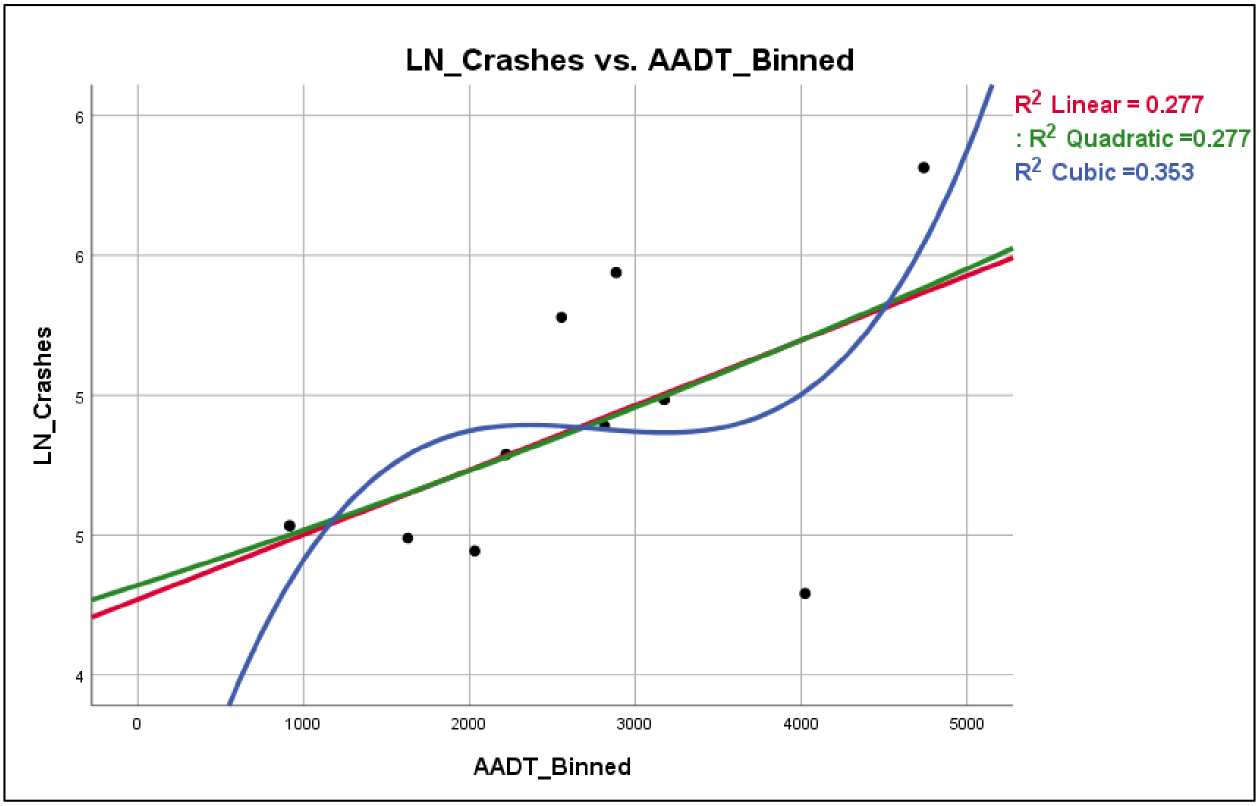

Figure F.8: Scatterplot LN_Crashes vs. Mean_AADT_Binned for Covariance Scale Determination

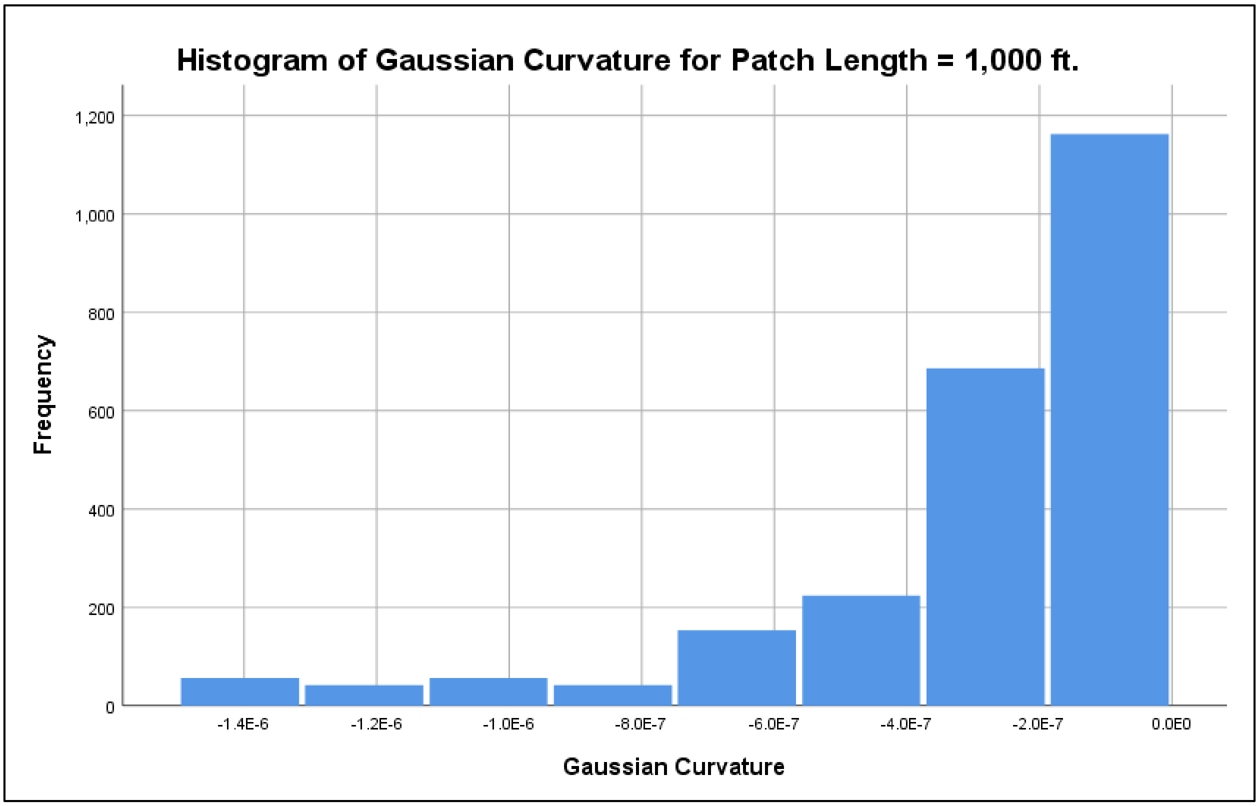

Figure F.9: Gaussian Curvature Histogram for Patch Length=1,000 ft

Table F.9: Descriptive Statistics of Gaussian Curvature for Patch Length=1,000 ft

Variable	Descriptives	Statistic	Std. Error
Gaussian	N	2,534	
	Mean	-0.00000030277559	0.000000006180665
	95% CI for Mean	[-0.00000031490÷ -0.0000002907]	
	5% Trimmed Mean	-0.00000026431226	
	Median	-0.00000020122400	
	Variance	9.2522 E-14	
	Std. Deviation	0.000000304174139	
	Minimum	-0.000001481470	
	Maximum	-0.000000005126	
	Range	0.000001476344	
	IR	0.000000266544	
	Skewness	-1.985	0.050
	Kurtosis	3.948	0.099

Table F.10: Mean Gauss_Binned vs. LN(Number of Crashes)

Group ID#	Number of Crashes	LN (Number of Crashes)	Mean Gauss_Binned
1	221	5.39816270	-0.00000103979039
2	95	4.55387689	-0.00000054602082
3	288	5.66296048	-0.00000036871653
4	107	4.67282883	-0.00000029260972
5	180	5.19295685	-0.00000023383324
6	137	4.91998093	-0.00000017567712
7	144	4.96981330	-0.00000013440056
8	94	4.54329478	-0.00000009873428
9	93	4.53259949	-0.00000006970611
10	69	4.23410650	-0.00000003541577

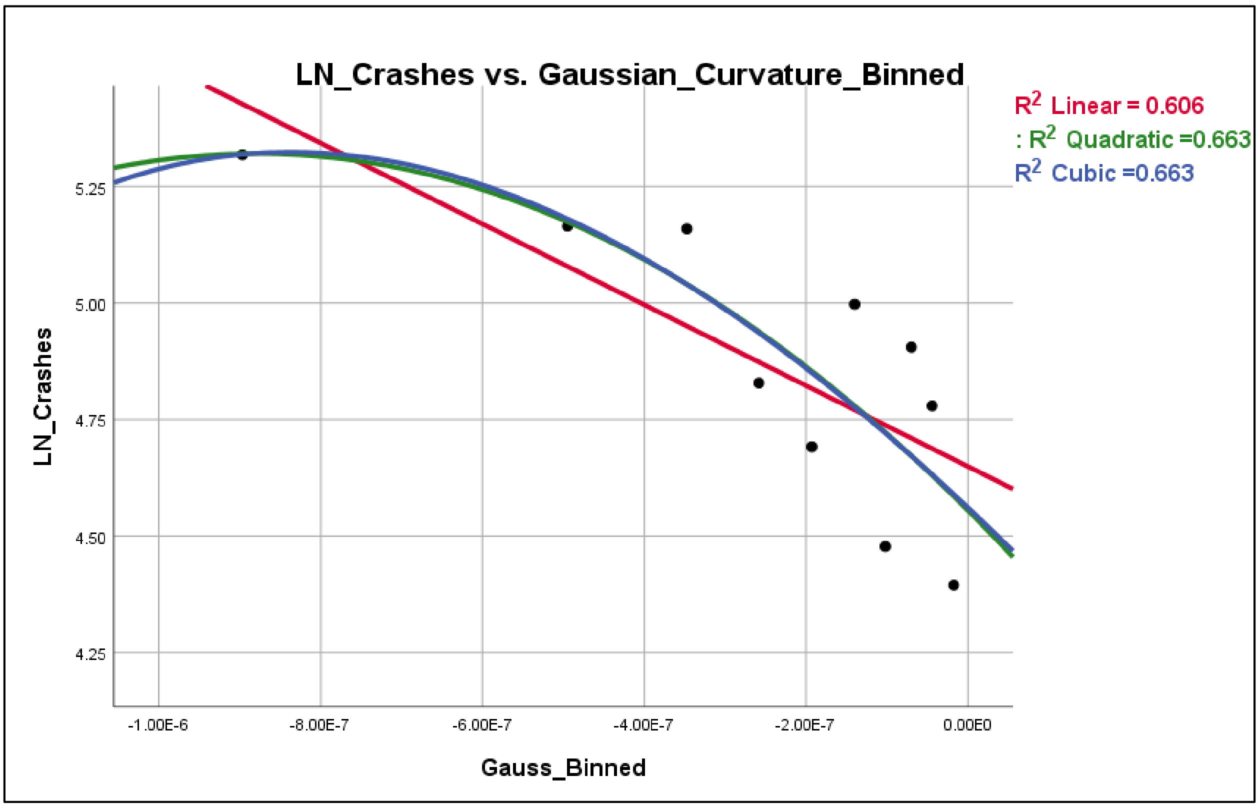

Figure F.10: Scatterplot LN_Crashes vs. Mean_Gauss_Binned for Covariance Scale Determination

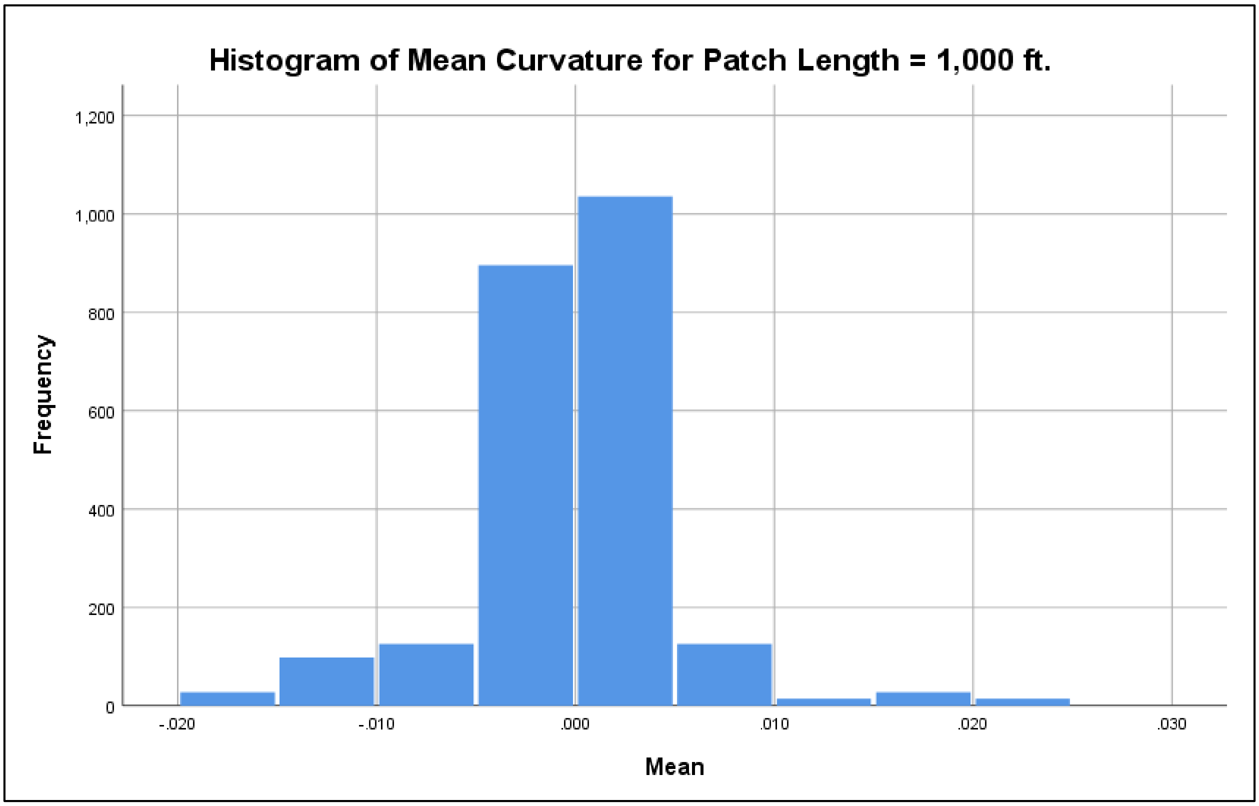

Figure F.11: Mean Curvature Histogram for Patch Length=1,000 ft

Table F.11: Descriptive Statistics of Mean Curvature for Patch Length=1,000 ft

Variable	**Descriptives**	**Statistic**	**Std. Error**
Gaussian	N	2,534	
	Mean	-0.00030684538	0.000106358119
	95% CI for Mean	[-0.00051541020 ÷ -0.00009828056]	
	5% Trimmed Mean	-0.00030654358	
	Median	0.00000812248	
	Variance	0.000027	
	Std. Deviation	0.005173423337	
	Minimum	-0.016932469	
	Maximum	0.024773028	
	Range	0.041705497	
	IR	0.000865688	
	Skewness	0.359	0.050
	Kurtosis	5.529	0.101

Table F.12: Mean Gauss_Binned vs. LN(Number of Crashes)

Group ID#	Number of Crashes	LN (Number of Crashes)	Average Mean_Curvature_Binned
1	206	5.32787617	-0.01086230388
2	94	4.54329478	-0.00322423465
3	165	5.10594547	-0.00073866941
4	246	5.50533154	-0.00014510209
5	133	4.89034913	-0.00002475845
6	107	4.67282883	0.00003989204
7	87	4.46590812	0.00011895493
8	121	4.79579055	0.00036983082
9	157	5.05624581	0.00223817847
10	77	4.34380542	0.00975142119

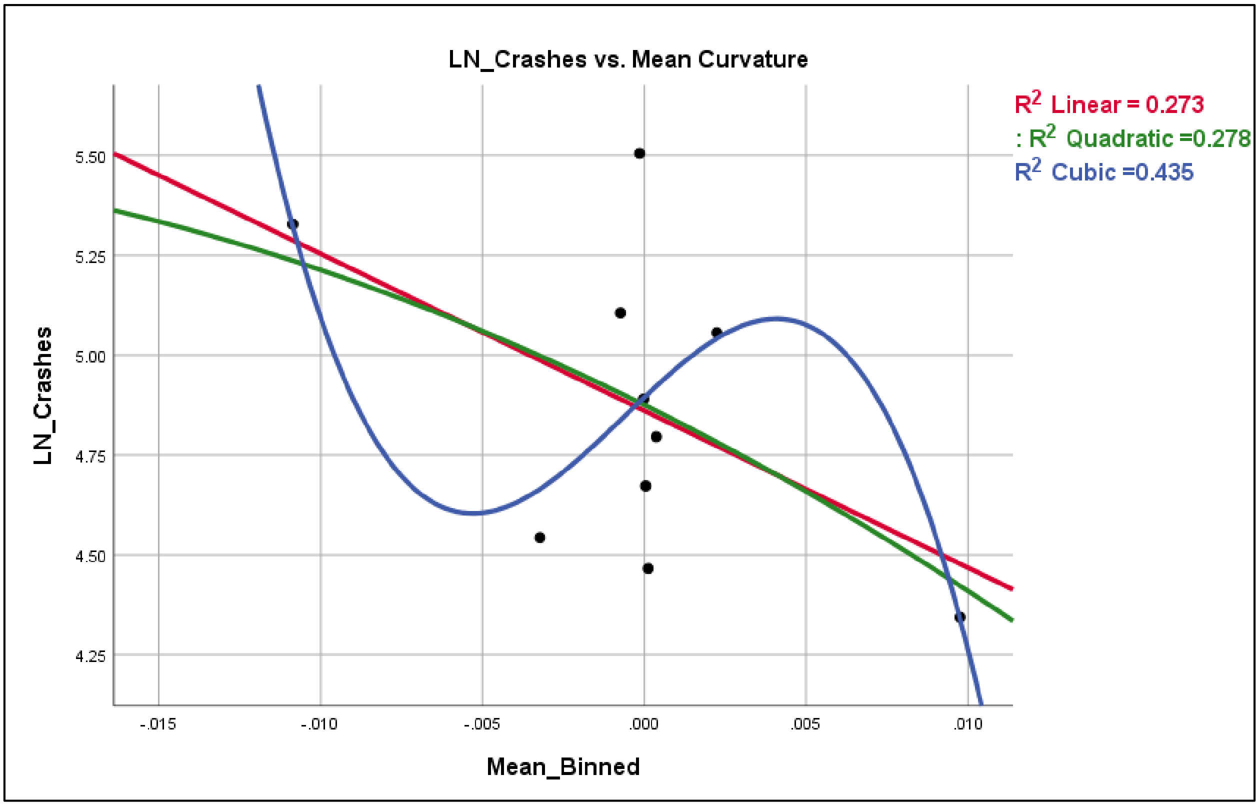

Figure F.12: Scatterplot LN_Crashes vs. Average Mean_Curvature_Binned for Covariance Scale Determination

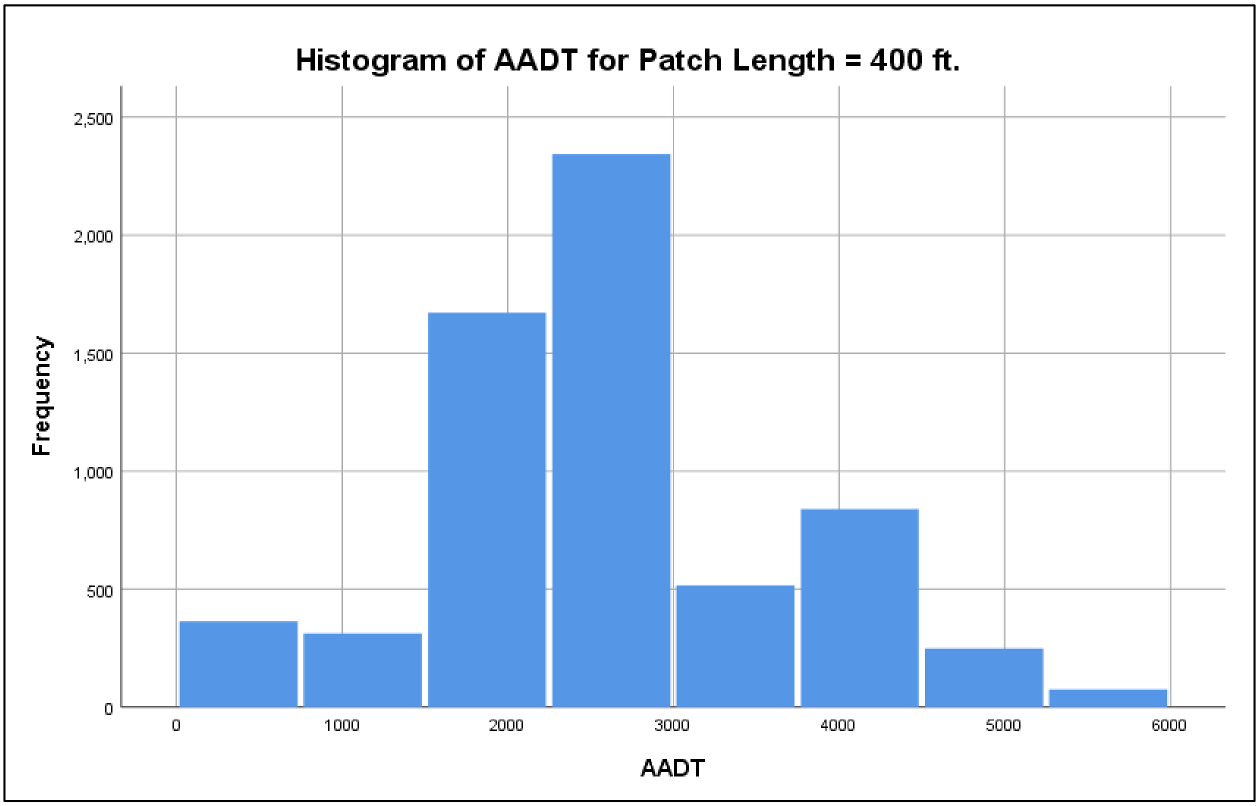

Figure F.13: AADT Histogram for Patch Length=400 ft

Table F.13: Descriptive Statistics of AADT for Patch Length=400 ft

Variable	Descriptives	Statistic	Std. Error
AADT	N	6,370	
	Mean	2,673	13.573
	95% CI for Mean	[2,647 ÷ 2,700]	
	5% Trimmed Mean	2,659	
	Median	2,741	
	Variance	1,173,537.858	
	Std. Deviation	1,084.000	
	Minimum	439	
	Maximum	5,716	
	Range	5,277	
	Interquartile Range	983	
	Skewness	0.329	0.031
	Kurtosis	0.114	0.061

Table F.14: Mean AADT_Binned vs. LN(Number of Crashes)

Group ID#	Number of Crashes	LN (Number of Crashes)	Mean AADT_Binned
1	91	4.51085951	904
2	88	4.47733681	1,628
3	88	4.47733681	2,045
4	117	4.76217393	2,227
5	199	5.29330482	2,551
6	117	4.76217393	2,809
7	248	5.51342875	2,881
8	144	4.96981330	3,175
9	74	4.30406509	4,026
10	356	5.87493073	4,755

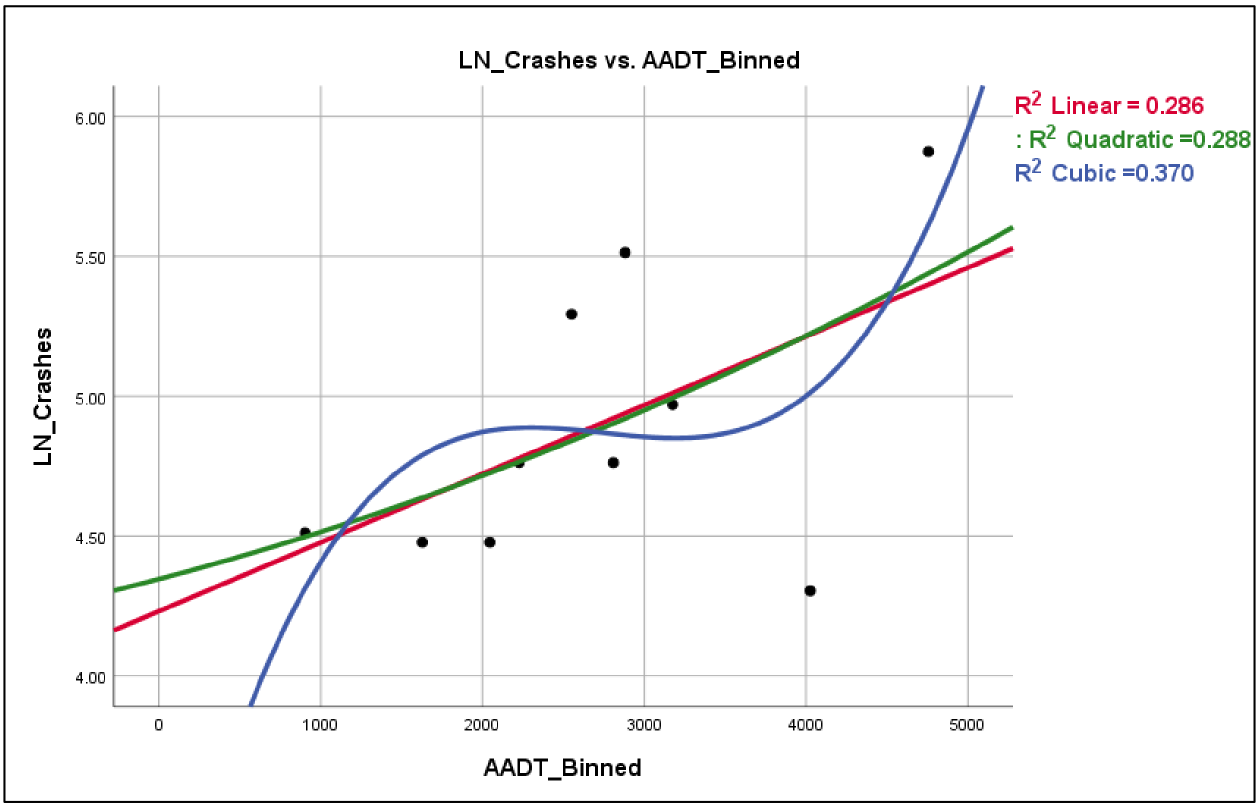

Figure F.14: Scatterplot LN_Crashes vs. Mean_AADT_Binned for Covariance Scale Determination

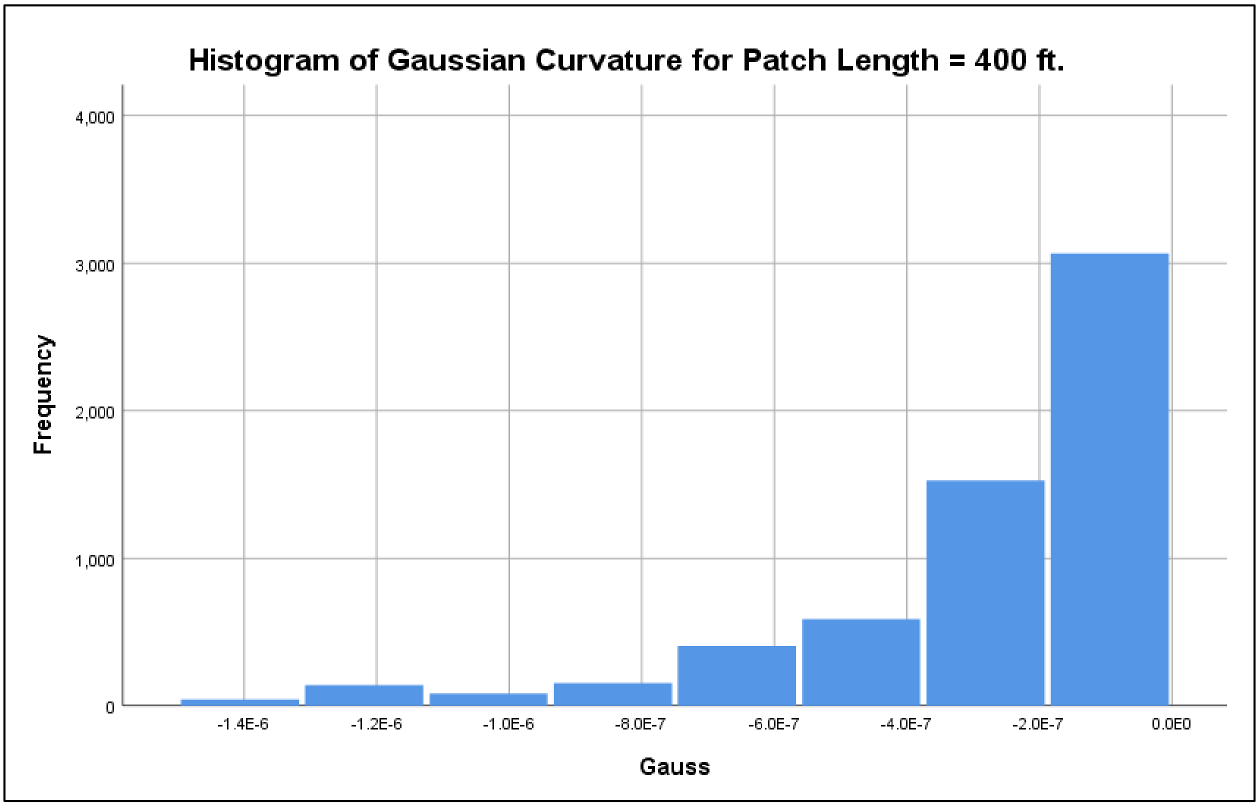

Figure F.15: Gaussian Curvature Histogram for Patch Length=400 ft

Table F.15: Descriptive Statistics of Gaussian Curvature for Patch Length=400 ft

Variable	Descriptives	Statistic	Std. Error
Gaussian	N	6,370	
	Mean	-0.00000027413524	0.000000003623714
	95% CI for Mean	[-0.00000028124÷ -0.00000026703]	
	5% Trimmed Mean	-0.00000024046902	
	Median	-0.00000018110200	
	Variance	7.8867 E-14	
	Std. Deviation	0.000000280832027	
	Minimum	-0.000001411250	
	Maximum	-0.000000000634	
	Range	0.000001410616	
	IR	0.000000287902	
	Skewness	-1.764	0.032
	Kurtosis	3.066	0.063

Table F.16: Mean Gauss_Binned vs. LN(Number of Crashes)

Group ID#	Number of Crashes	LN (Number of Crashes)	Mean Gauss_Binned
1	273	5.60947180	-0.00000094563551
2	160	5.07517382	-0.00000053546207
3	133	4.89034913	-0.00000036520777
4	123	4.81218436	-0.00000027691260
5	139	4.93447393	-0.00000020441570
6	128	4.85203026	-0.00000015786567
7	64	4.15888308	-0.00000011656382
8	101	4.61512052	-0.00000007460986
9	115	4.74493213	-0.00000004227309
10	116	4.75359019	-0.00000001641273

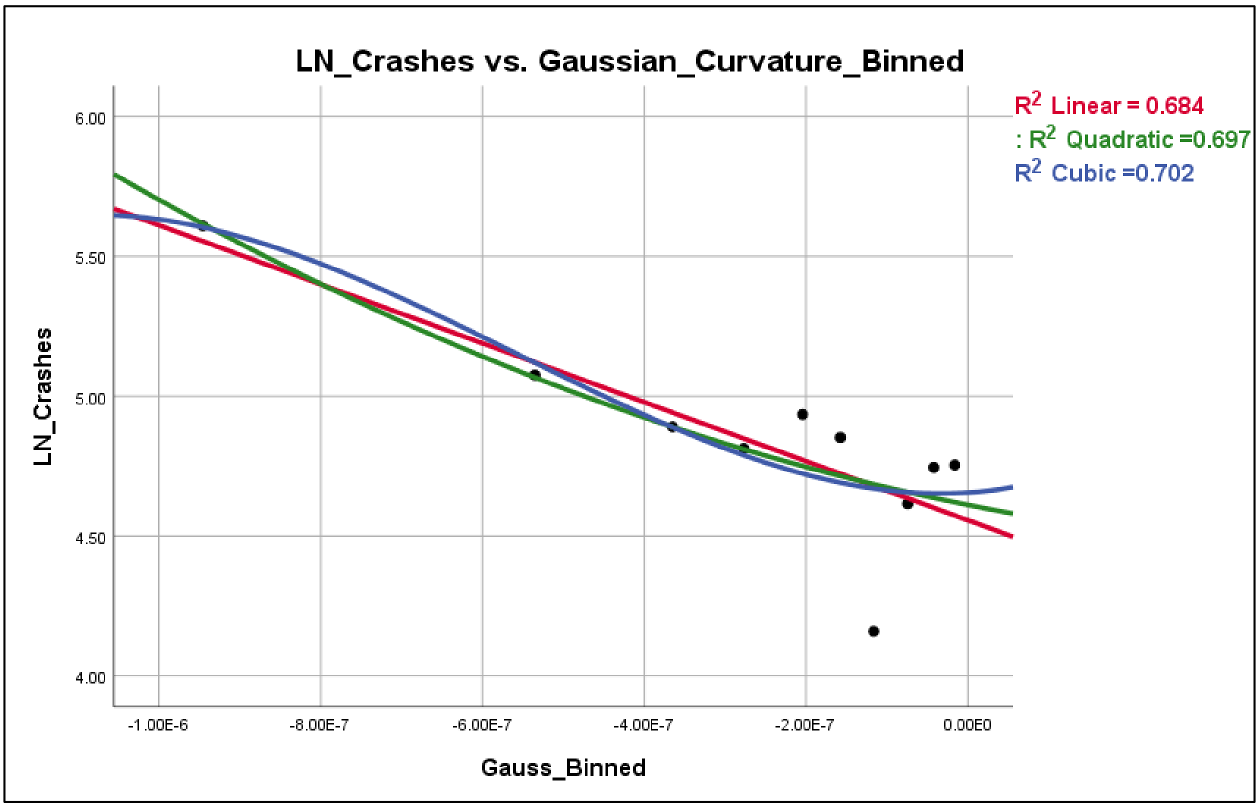

Figure F.16: Scatterplot LN_Crashes vs. Mean_Gauss_Binned for Covariance Scale Determination

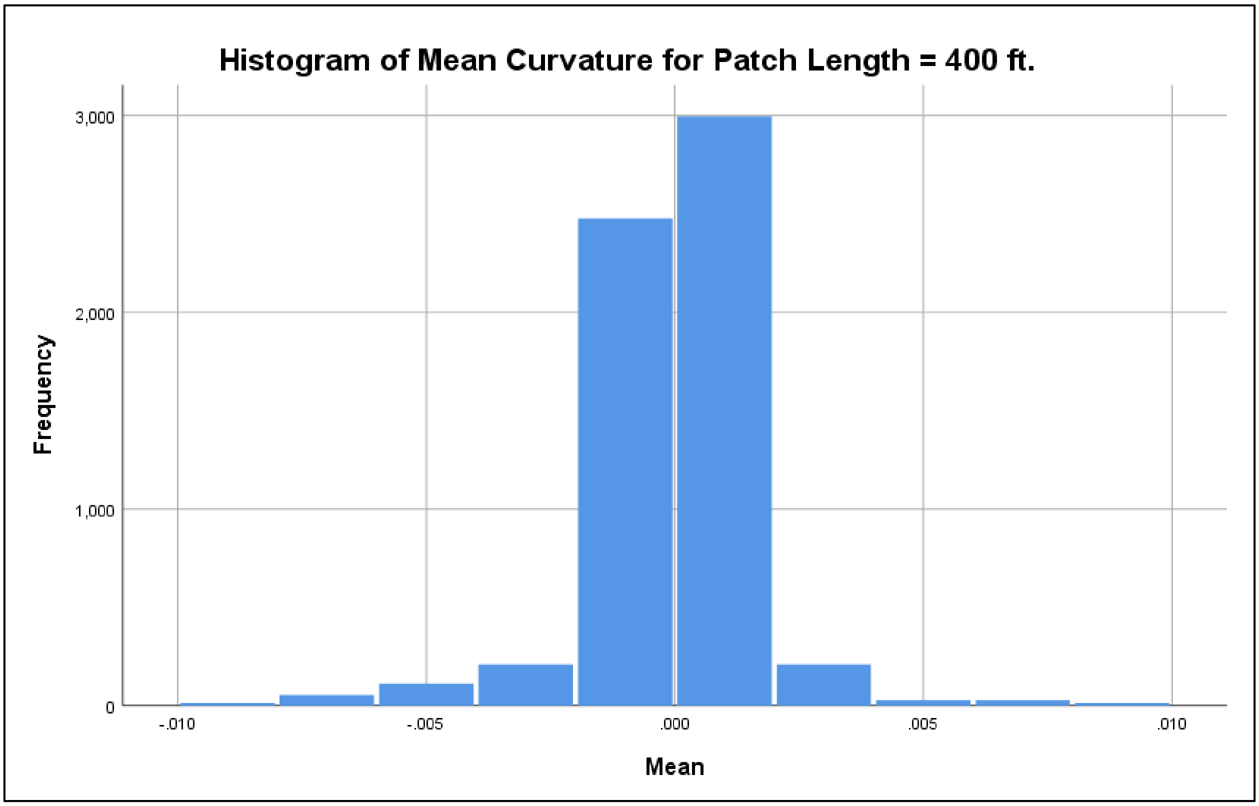

Figure F.17: Mean Curvature Histogram for Patch Length=400 ft

Table F.17: Descriptive Statistics of Mean Curvature for Patch Length=400 ft

Variable	Descriptives	Statistic	Std. Error
Gaussian	N	6,370	
	Mean	-0.0000570858762	0.00001990445335
	95% CI for Mean	[-0.00009610557 ÷ -0.00001806618]	
	5% Trimmed Mean	-0.0000125035190	
	Median	0.0000099411600	
	Variance	0.000002	
	Std. Deviation	0.00156043805370	
	Minimum	-0.00848872100	
	Maximum	0.00932665900	
	Range	0.01781538000	
	IR	0.00034459100	
	Skewness	-0.315	0.031
	Kurtosis	10.871	0.062

Table F.18: Mean Gauss_Binned vs. LN(Number of Crashes)

Group ID#	Number of Crashes	LN (Number of Crashes)	Average Mean_Curvature_Binned
1	256	5.54517744	-0.0031591563636
2	129	4.85981240	-0.0006234931818
3	61	4.11087386	-0.0001758359773
4	122	4.80402104	-0.0000830496136
5	89	4.48863637	-0.0000178042651
6	203	5.31320598	0.0000376824250
7	163	5.09375020	0.0000982906659
8	169	5.12989871	0.0001916438864
9	97	4.57471098	0.0006045017955
10	122	4.80402104	0.0026171397209

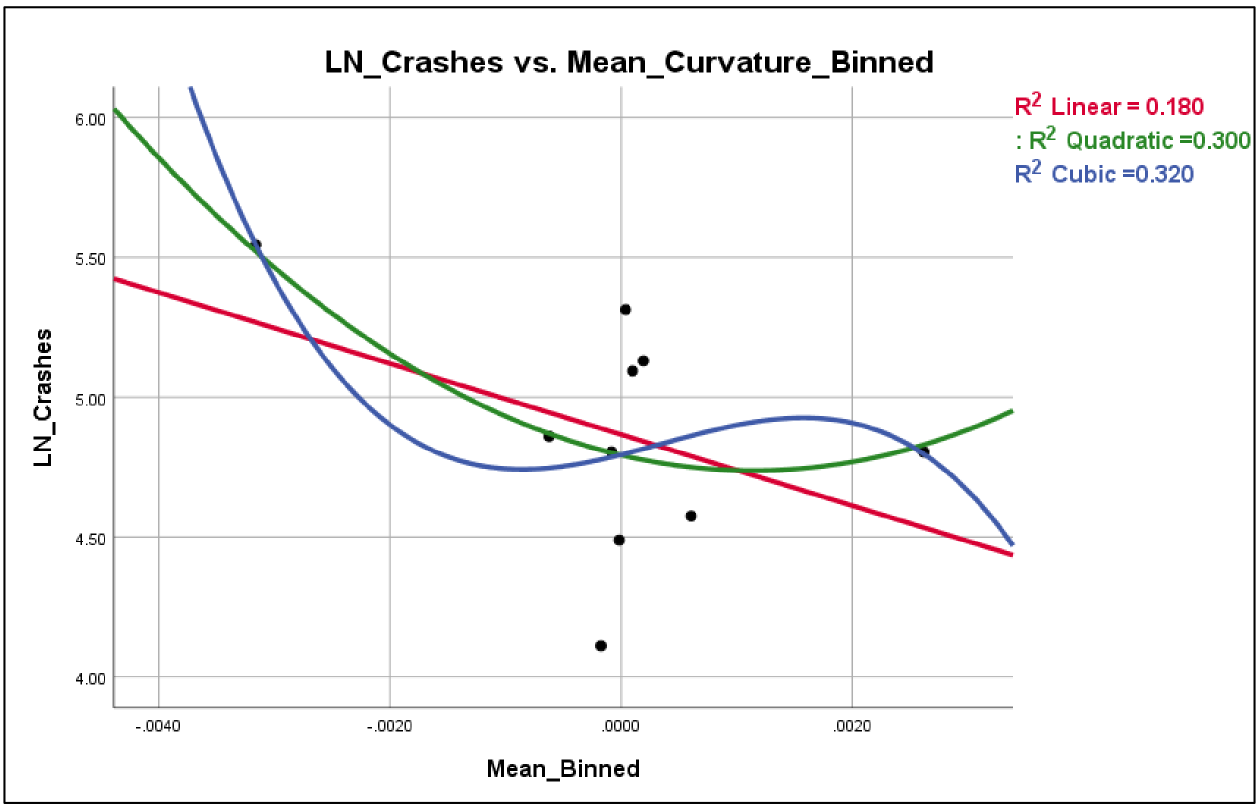

Figure F.18: Scatterplot LN_Crashes vs. Average Mean_Curvature_Binned for Covariance Scale Determination

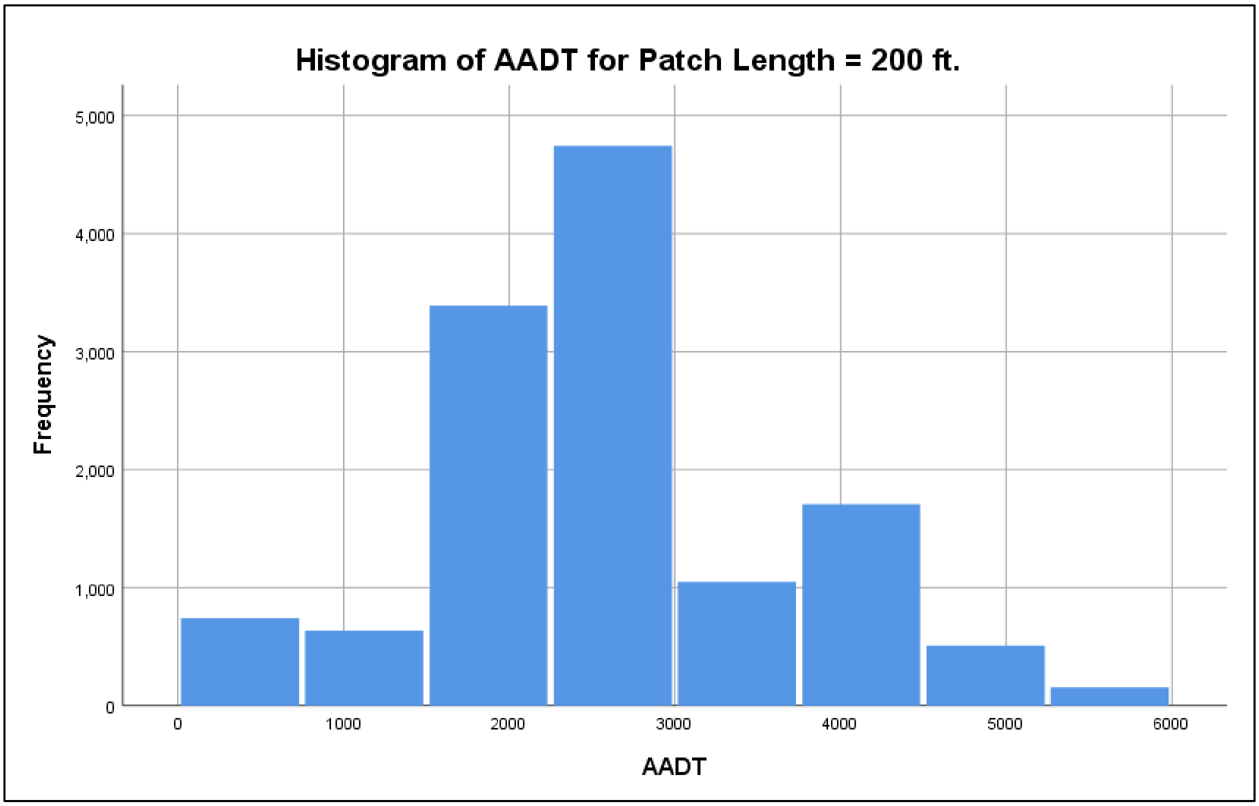

Figure F.19: AADT Histogram for Patch Length=200 ft

Table F.19: Descriptive Statistics of AADT for Patch Length=200 ft

Variable	Descriptives	Statistic	Std. Error
AADT	N	12,936	
	Mean	2,674	9.541
	95% CI for Mean	[2,655 ÷ 2,692]	
	5% Trimmed Mean	2,659	
	Median	2,741	
	Variance	1,177,449.690	
	Std. Deviation	1,085	
	Minimum	439	
	Maximum	5716	
	Range	5277	
	Interquartile Range	983	
	Skewness	0.329	0.022
	Kurtosis	0.109	0.043

Table F.20: Mean AADT_Binned vs. LN(Number of Crashes)

Group ID#	Number of Crashes	LN (Number of Crashes)	Mean AADT_Binned
1	91	4.51085951	904
2	89	4.48863637	1,627
3	88	4.47733681	2,045
4	117	4.76217393	2,226
5	197	5.28320373	2,552
6	134	4.89783980	2,815
7	231	5.44241771	2,885
8	145	4.97673374	3,175
9	74	4.30406509	4,026
10	358	5.88053299	4,757

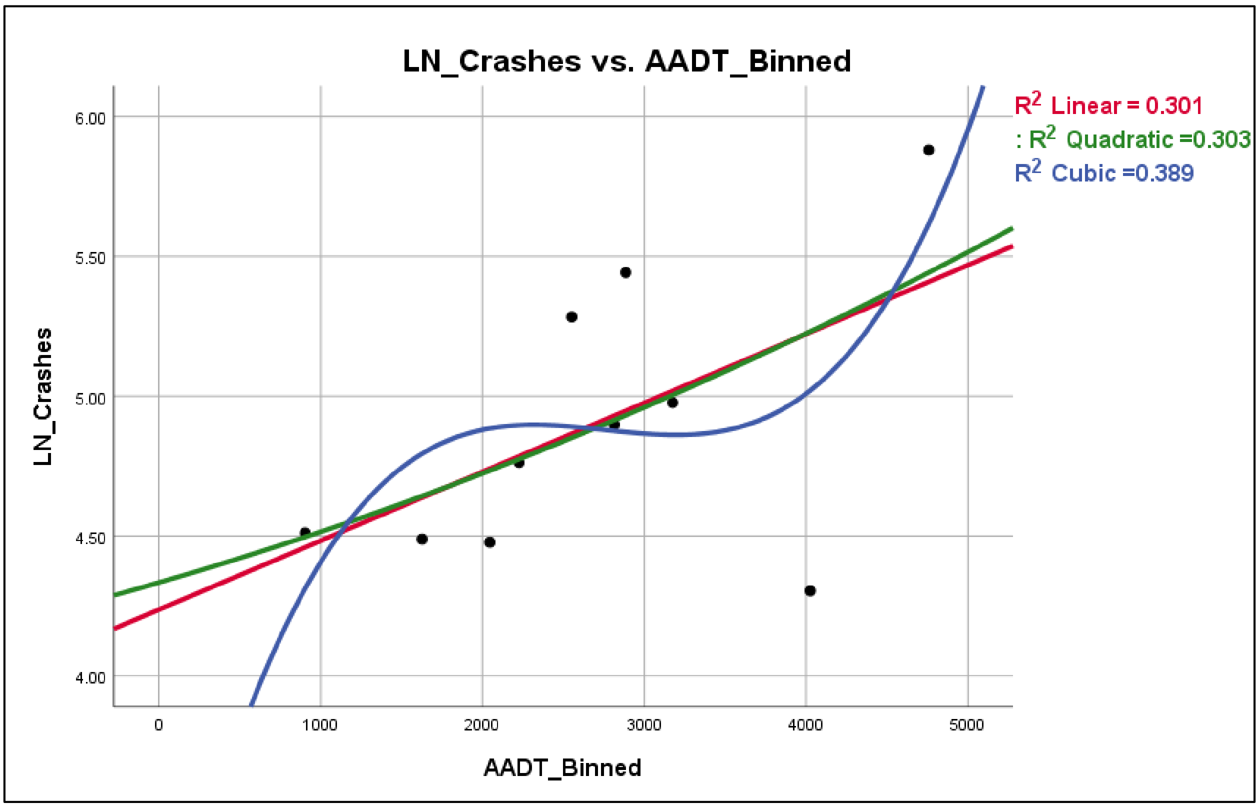

Figure F.20: Scatterplot LN_Crashes vs. Mean_AADT_Binned for Covariance Scale Determination

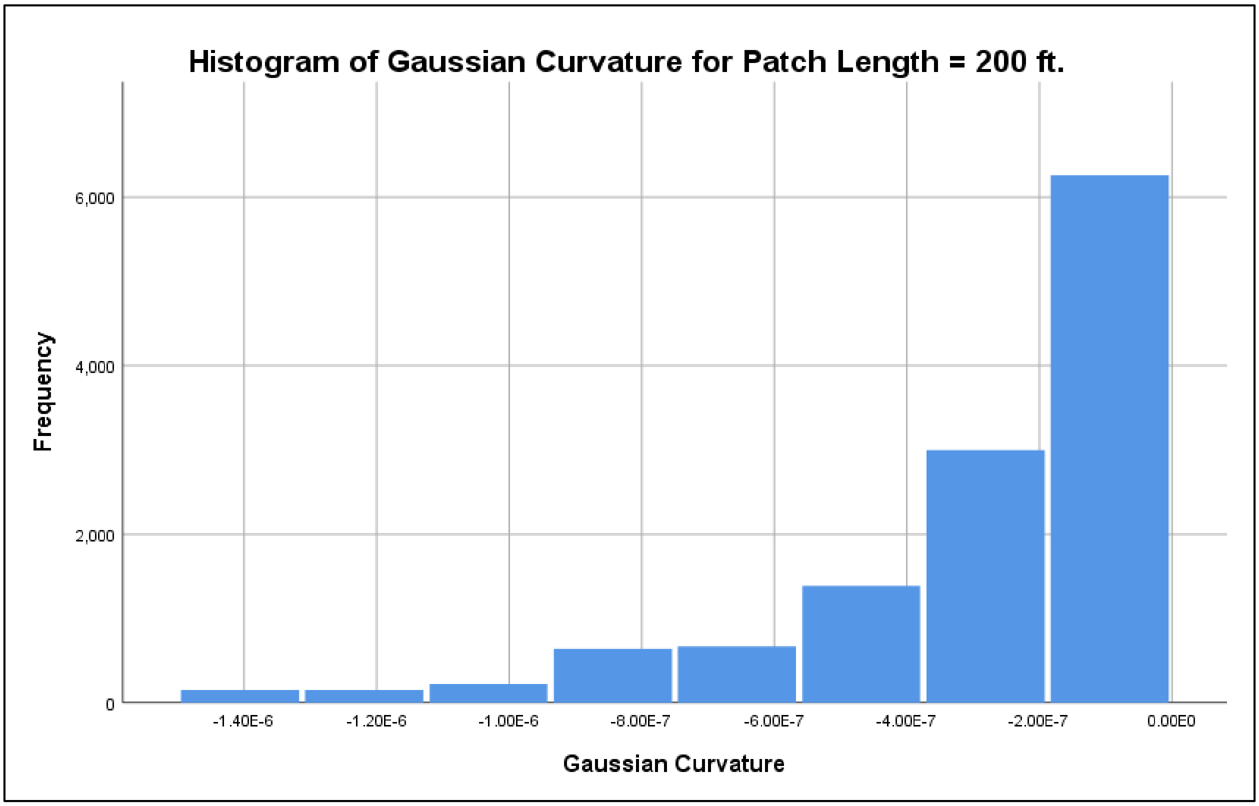

Figure F.21: Gaussian Curvature Histogram for Patch Length=200 ft

Table F.21: Descriptive Statistics of Gaussian Curvature for Patch Length=200 ft

Variable	Descriptives	Statistic	Std. Error
Gaussian	N	12,936	
	Mean	-0.00000028490972	0.000000002588443
	95% CI for Mean	[-0.00000028998÷ -0.00000027984]	
	5% Trimmed Mean	-0.00000025208400	
	Median	-0.00000018741150	
	Variance	8.367 E-14	
	Std. Deviation	0.000000289257738	
	Minimum	-0.000001489390	
	Maximum	-0.000000002114	
	Range	0.000001487276	
	IR	0.000000312415	
	Skewness	-1.729	0.022
	Kurtosis	2.943	0.044

Table F.22: Mean Gauss_Binned vs. LN(Number of Crashes)

Group ID#	Number of Crashes	LN (Number of Crashes)	Mean Gauss_Binned
1	256	5.54517744	-0.00000097265249
2	168	5.12396398	-0.00000056326313
3	153	5.03043792	-0.00000038583212
4	131	4.87519732	-0.00000027637992
5	104	4.64439090	-0.00000021458764
6	88	4.47733681	-0.00000016207796
7	103	4.63472899	-0.00000011778496
8	85	4.44265126	-0.00000008104400
9	124	4.82028157	-0.00000004917910
10	92	4.52178858	-0.00000001994857

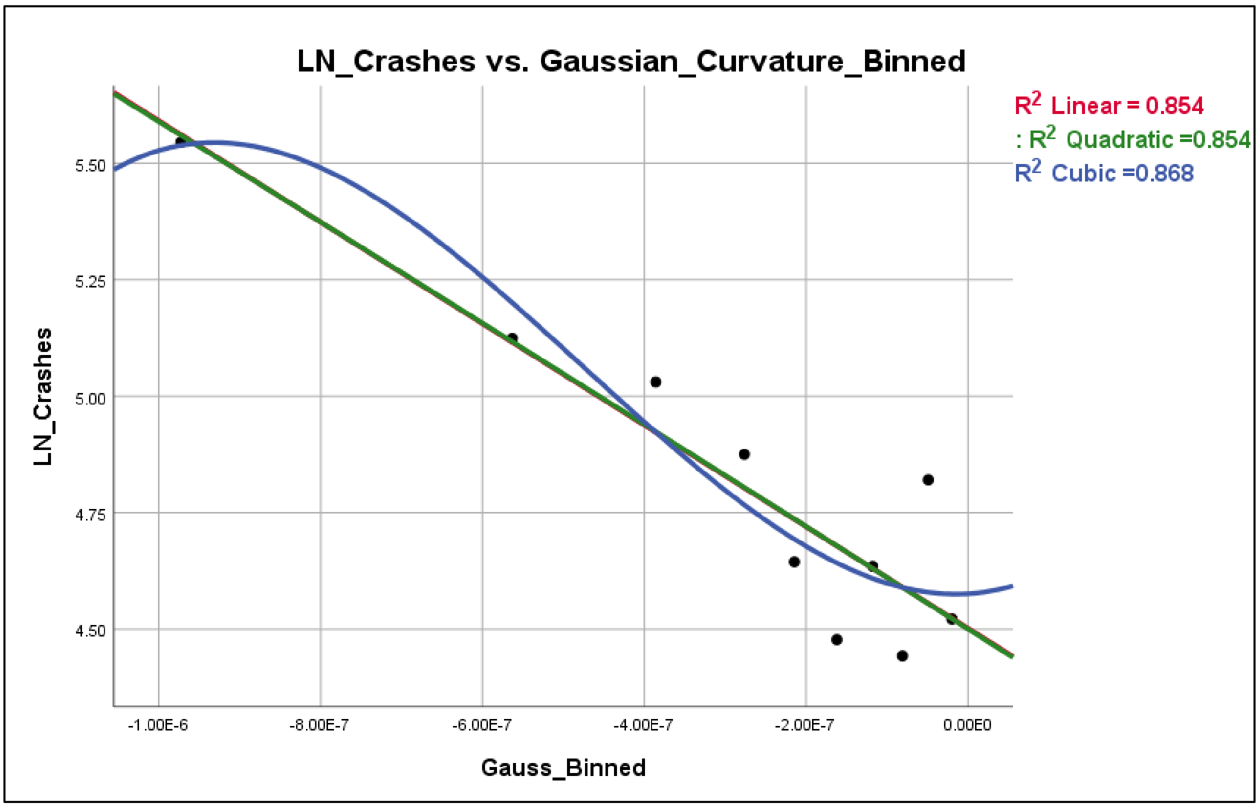

Figure F.22: Scatterplot LN_Crashes vs. Mean_Gauss_Binned for Covariance Scale Determination

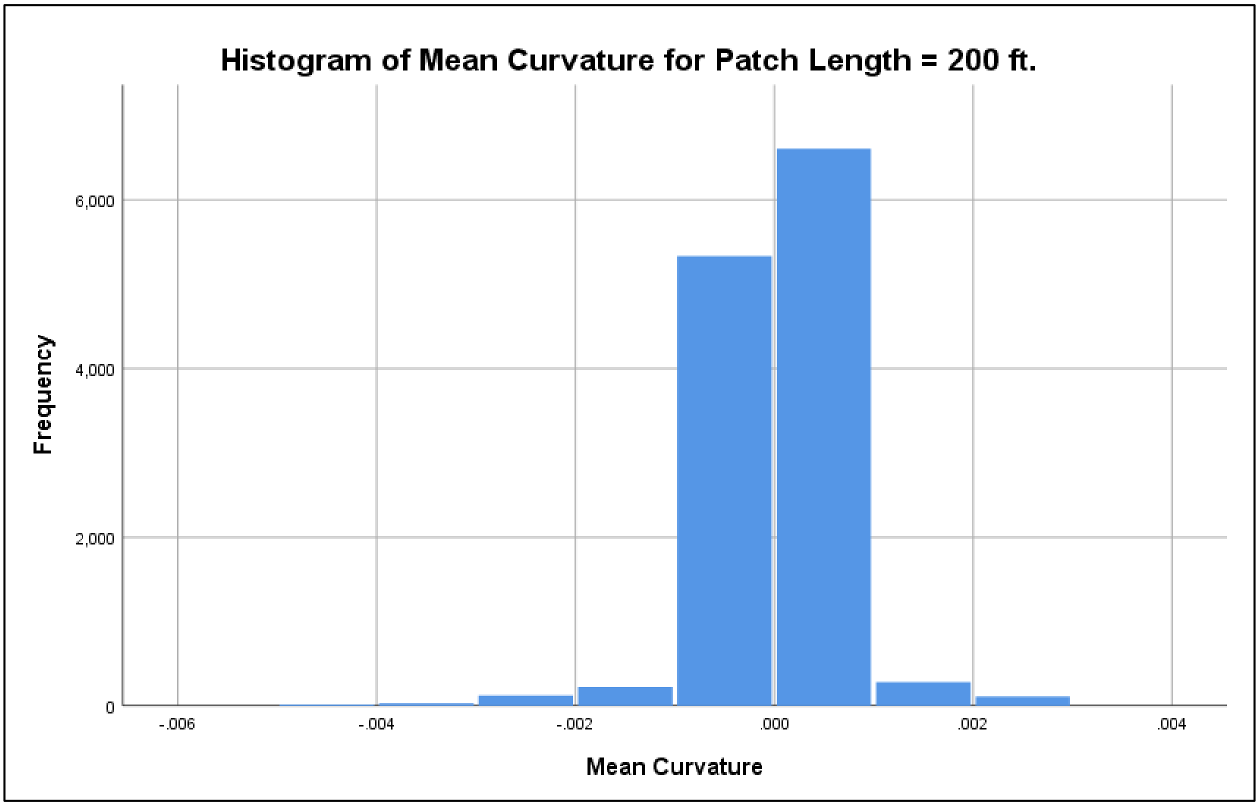

Figure F.23: Mean Curvature Histogram for Patch Length=200 ft

Table F.23: Descriptive Statistics of Mean Curvature for Patch Length=200 ft

Variable	**Descriptives**	**Statistic**	**Std. Error**
Gaussian	N	12,936	
	Mean	0.000017925168	0.0000046855096
	95% CI for Mean	[0.000008740865 ÷0.000027109472]	
	5% Trimmed Mean	0.000025887112	
	Median	0.000019587600	
	Variance	2.7939 E-7	
	Std. Deviation	0.0005285703409	
	Minimum	-0.0042198760	
	Maximum	0.0029446240	
	Range	0.0071645000	
	IR	0.0002185762	
	Skewness	-1.015	0.022
	Kurtosis	16.501	0.043

Table F.24: Mean Gauss_Binned vs. LN(Number of Crashes)

Group ID#	Number of Crashes	LN (Number of Crashes)	Average Mean_Curvature_Binned
1	171	5.14166356	-0.000895598011
2	141	4.94875989	-0.000172234791
3	73	4.29045944	-0.000085944705
4	98	4.58496748	-0.000040076424
5	82	4.40671925	0.000000399880
6	142	4.95582706	0.000037650820
7	138	4.92725369	0.000081904573
8	130	4.86753445	0.000132273418
9	188	5.23644196	0.000245237989
10	200	5.29831737	0.000885169089

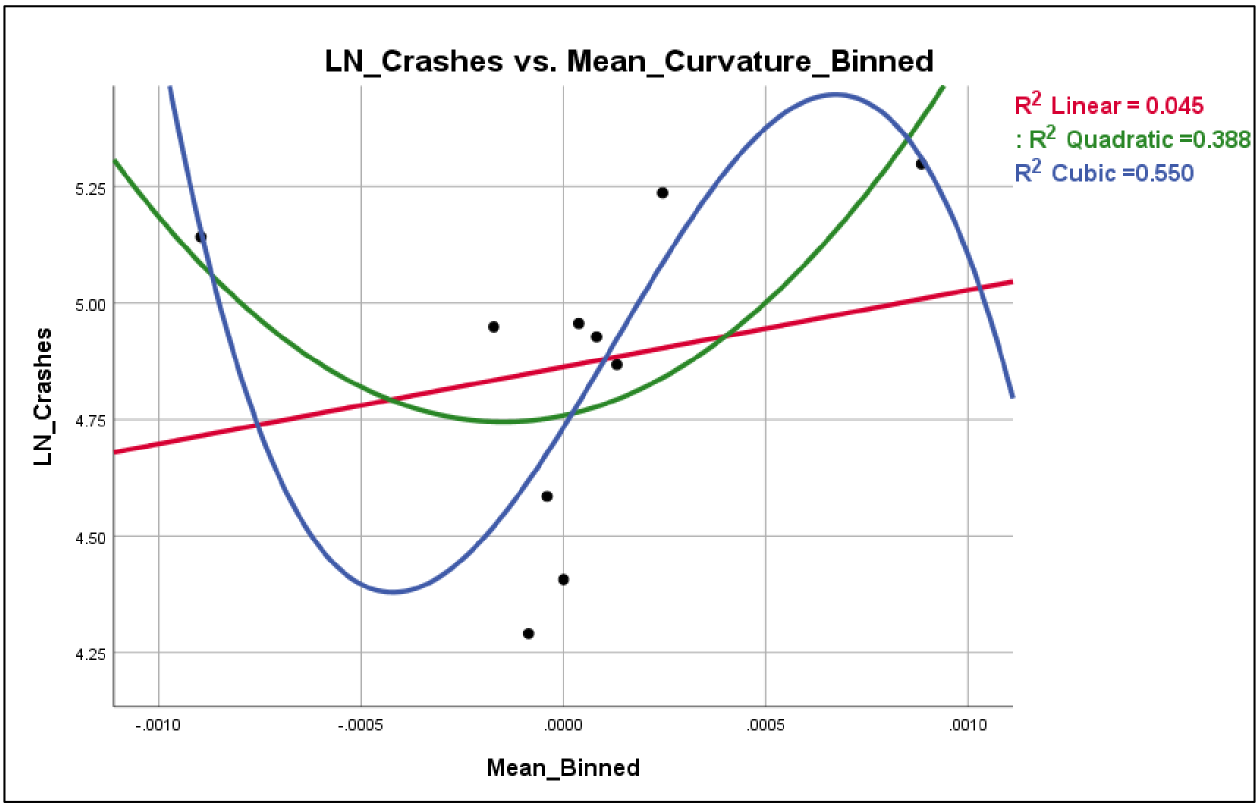

Figure F.24: Scatterplot LN_Crashes vs. Average Mean_Curvature_Binned for Covariance Scale Determination

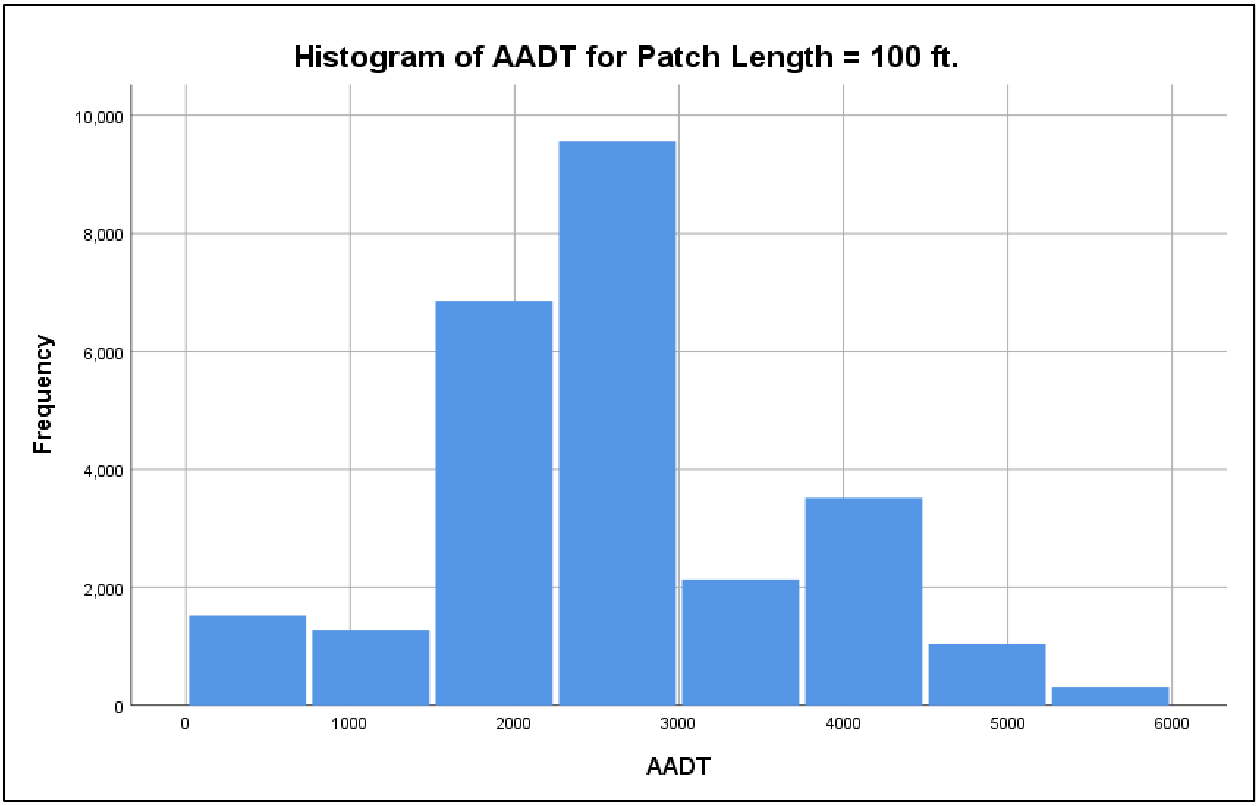

Figure F.25: AADT Histogram for Patch Length=100 ft

Table F.25: Descriptive Statistics of AADT for Patch Length=100 ft

Variable	Descriptives	Statistic	Std. Error
AADT	N	26,236	
	Mean	2,676	6.723
	95% CI for Mean	[2,663 ÷ 2,690]	
	5% Trimmed Mean	2,661	
	Median	2,741	
	Variance	1,185,841.067	
	Std. Deviation	1,088.963	
	Minimum	439	
	Maximum	5,716	
	Range	5,277	
	Interquartile Range	1,041	
	Skewness	0.320	0.015
	Kurtosis	0.083	0.030

Table F.26: Mean AADT_Binned vs. LN(Number of Crashes)

Group ID#	Number of Crashes	LN (Number of Crashes)	Mean AADT_Binned
1	92	4.52178858	900
2	88	4.47733681	1,628
3	87	4.46590812	2,044
4	118	4.77068462	2,227
5	197	5.28320373	2,551
6	133	4.89034913	2,815
7	231	5.44241771	2,885
8	145	4.97673374	3,176
9	74	4.30406509	4,027
10	358	5.88053299	4,753

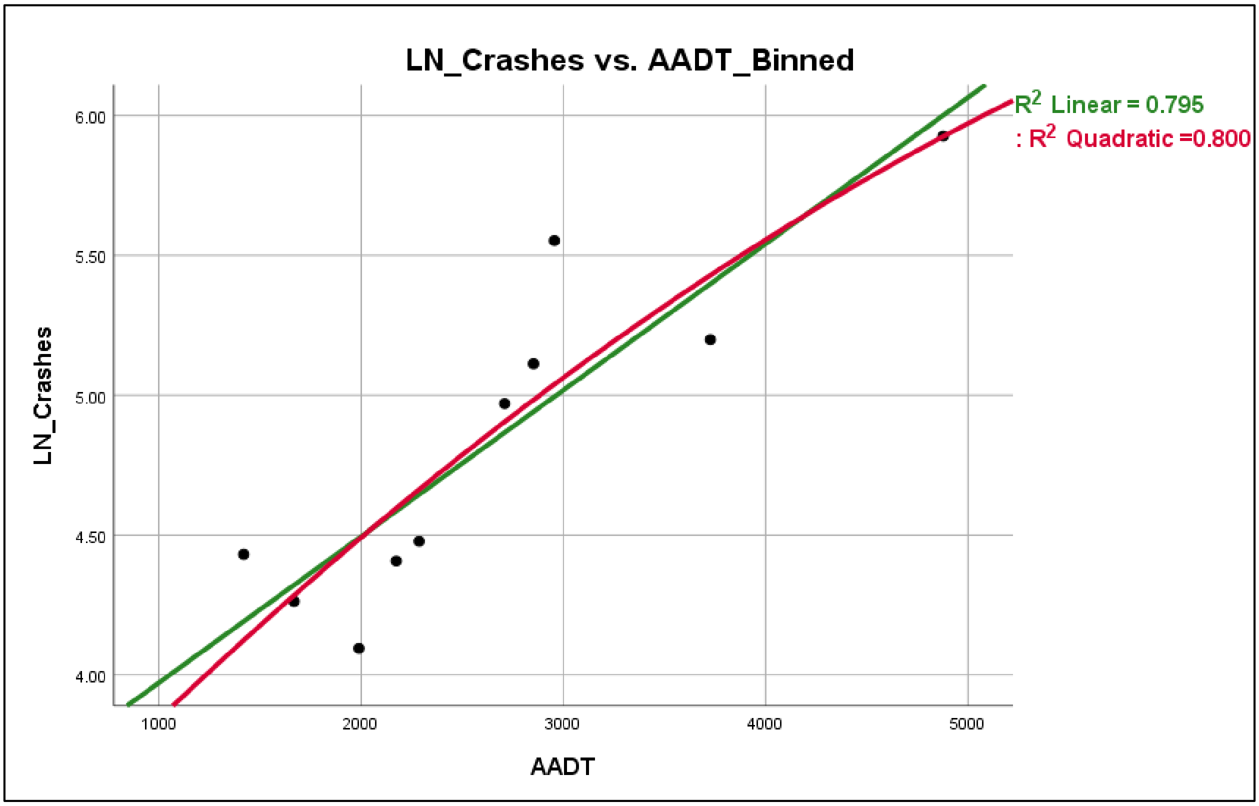

Figure F.26: Scatterplot LN_Crashes vs. Mean_AADT_Binned for Covariance Scale Determination

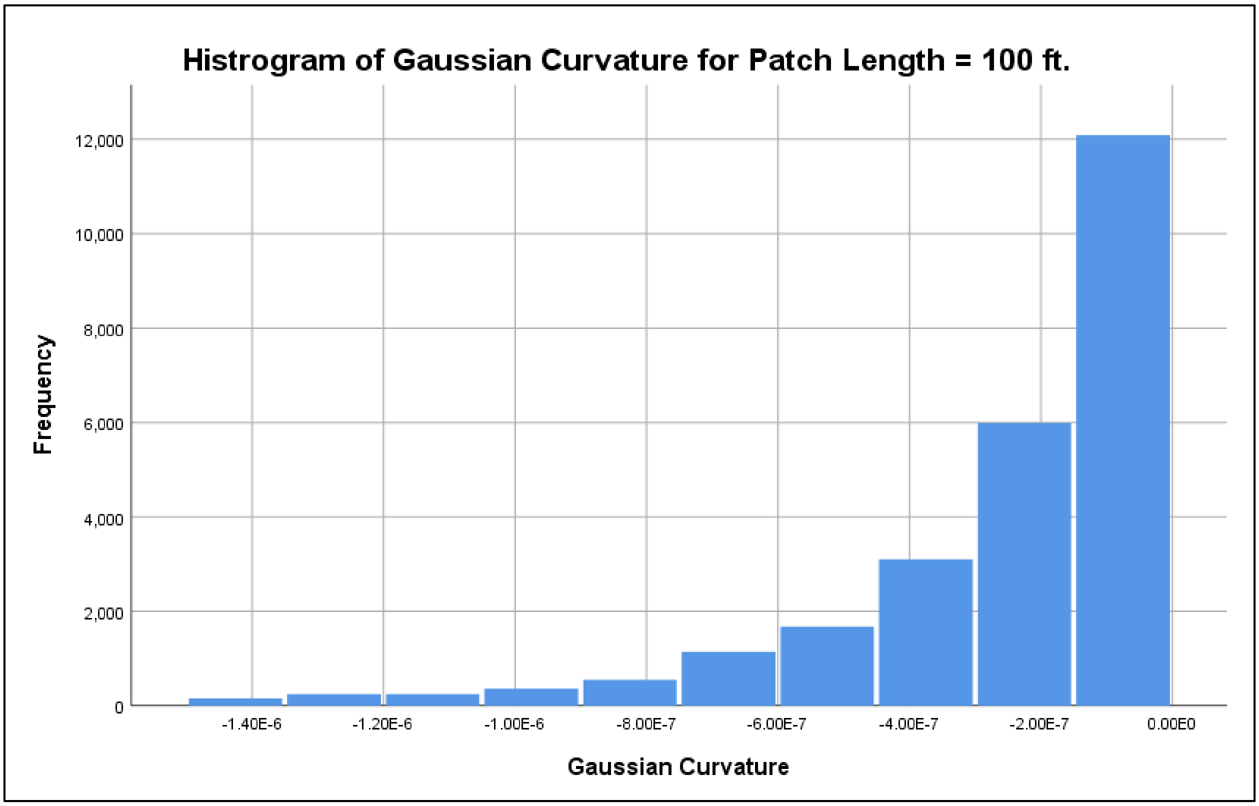

Figure F.27: Gaussian Curvature Histogram for Patch Length=100 ft

Table F.27: Descriptive Statistics of Gaussian Curvature for Patch Length=100 ft

Variable	**Descriptives**	**Statistic**	**Std. Error**
Gaussian	N	26,236	
	Mean	-0.00000025667396	0.000000001674374
	95% CI for Mean	[-0.00000025996÷ -0.00000025339]	
	5% Trimmed Mean	-0.00000022407771	
	Median	-0.00000016332100	
	Variance	7.1709 E-14	
	Std. Deviation	0.000000267784719	
	Minimum	-0.000001499280	
	Maximum	-0.000000000102	
	Range	0.000001499178	
	IR	0.000000280156	
	Skewness	-1.907	0.015
	Kurtosis	3.990	0.031

Table F.28: Mean Gauss_Binned vs. LN(Number of Crashes)

Group ID#	Number of Crashes	LN (Number of Crashes)	Mean Gauss_Binned
1	204	5.31811999	-0.00000089661451
2	175	5.16478597	-0.00000049491986
3	174	5.15905530	-0.00000034748993
4	125	4.82831374	-0.00000025845153
5	109	4.69134788	-0.00000019293687
6	148	4.99721227	-0.00000013994380
7	88	4.47733681	-0.00000010231440
8	135	4.90527478	-0.00000006999132
9	119	4.77912349	-0.00000004442034
10	81	4.39444915	-0.00000001765684

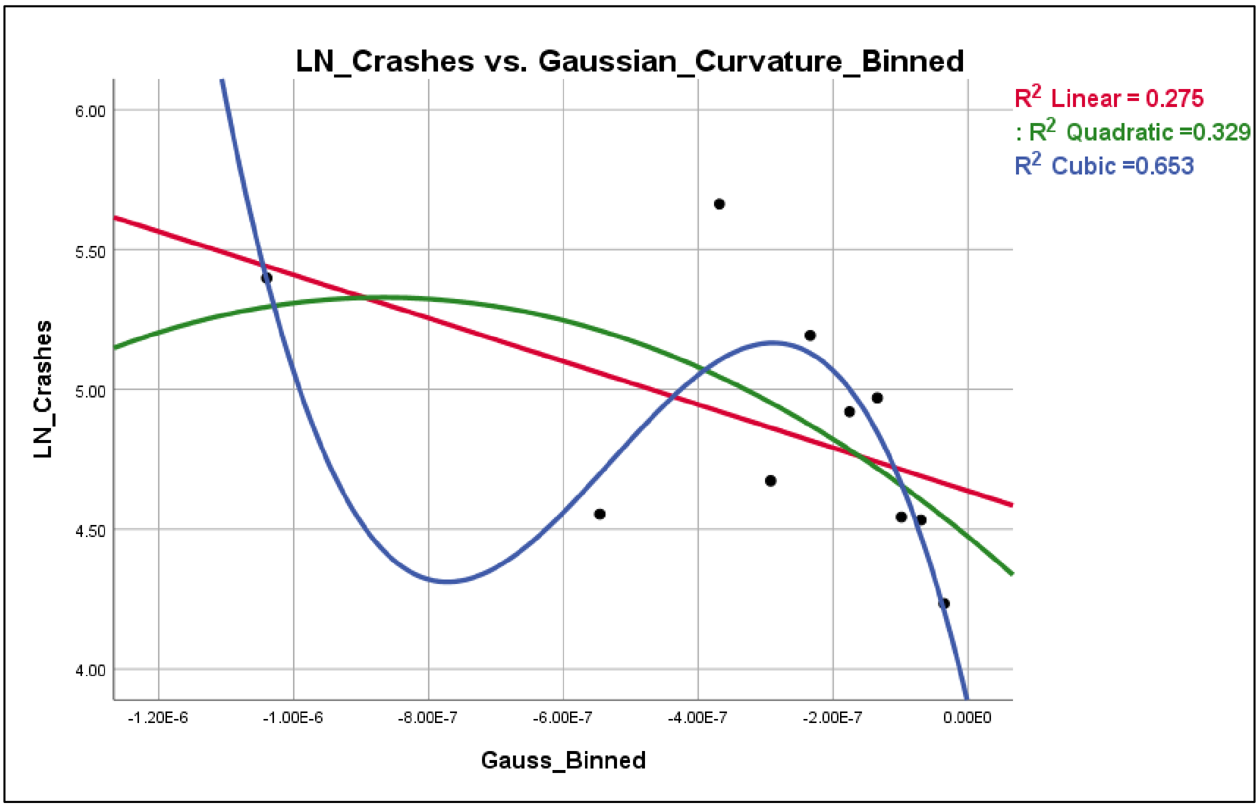

Figure F.28: Scatterplot LN_Crashes vs. Mean_Gauss_Binned for Covariance Scale Determination

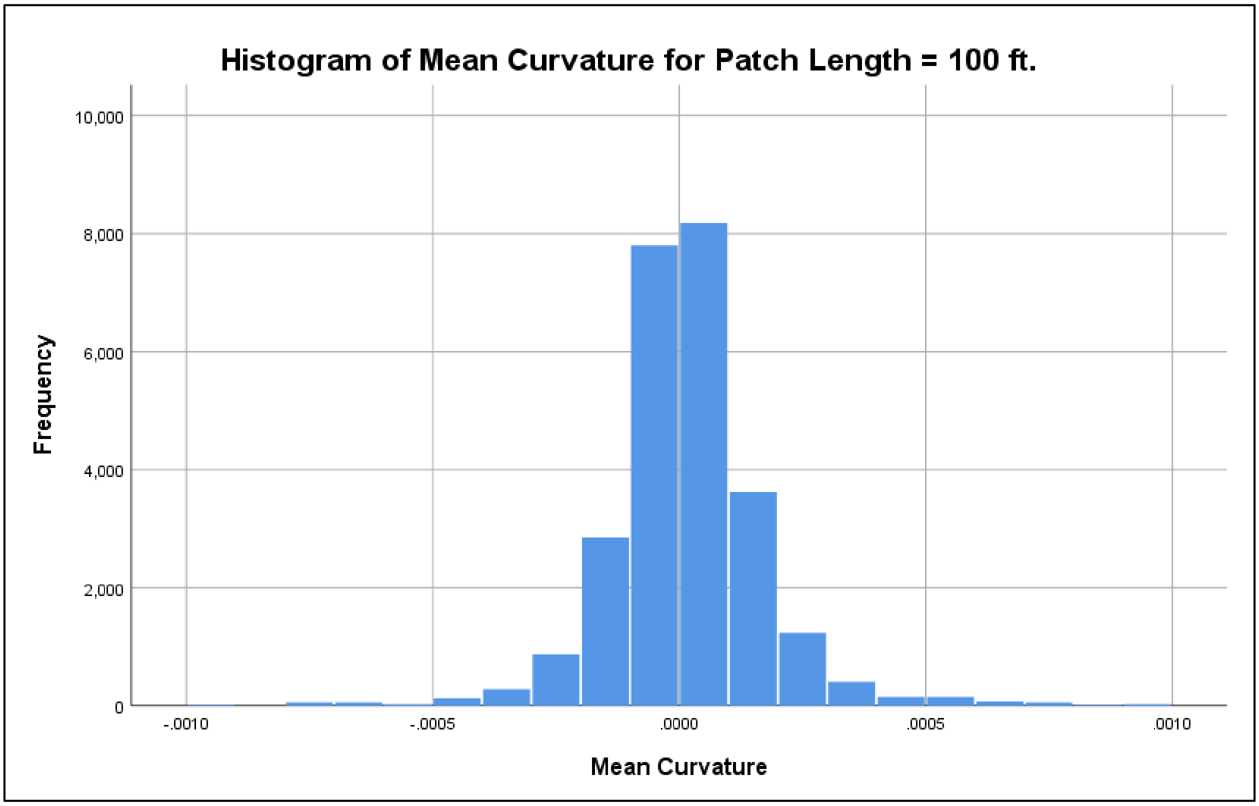

Figure F.29: Mean Curvature Histogram for Patch Length=100 ft

Table F.29: Descriptive Statistics of Mean Curvature for Patch Length=100 ft

Variable	Descriptives	Statistic	Std. Error
Gaussian	N	26,236	
	Mean	0.000015722922	0.0000009823179
	95% CI for Mean	[0.000013797525÷0.000017648319]	
	5% Trimmed Mean	0.000013514588	
	Median	0.000011795600	
	Variance	2.5087 E-8	
	Std. Deviation	0.0001583879164	
	Minimum	-0.0009058790	
	Maximum	0.0009737290	
	Range	0.0018796080	
	IR	0.0001475257	
	Skewness	0.283	0.015
	Kurtosis	5.636	0.030

Table F.30: Mean Gauss_Binned vs. LN(Number of Crashes)

Group ID#	Number of Crashes	LN (Number of Crashes)	Average Mean_Curvature_Binned
1	206	5.32787617	-0.000256756763
2	133	4.89034913	-0.000109274836
3	119	4.77912349	-0.000062272385
4	89	4.48863637	-0.000029809969
5	117	4.76217393	-0.000002843858
6	135	4.90527478	0.000024523926
7	106	4.66343909	0.000052397901
8	100	4.60517019	0.000087679790
9	124	4.82028157	0.000141670839
10	305	5.72031178	0.000313467730

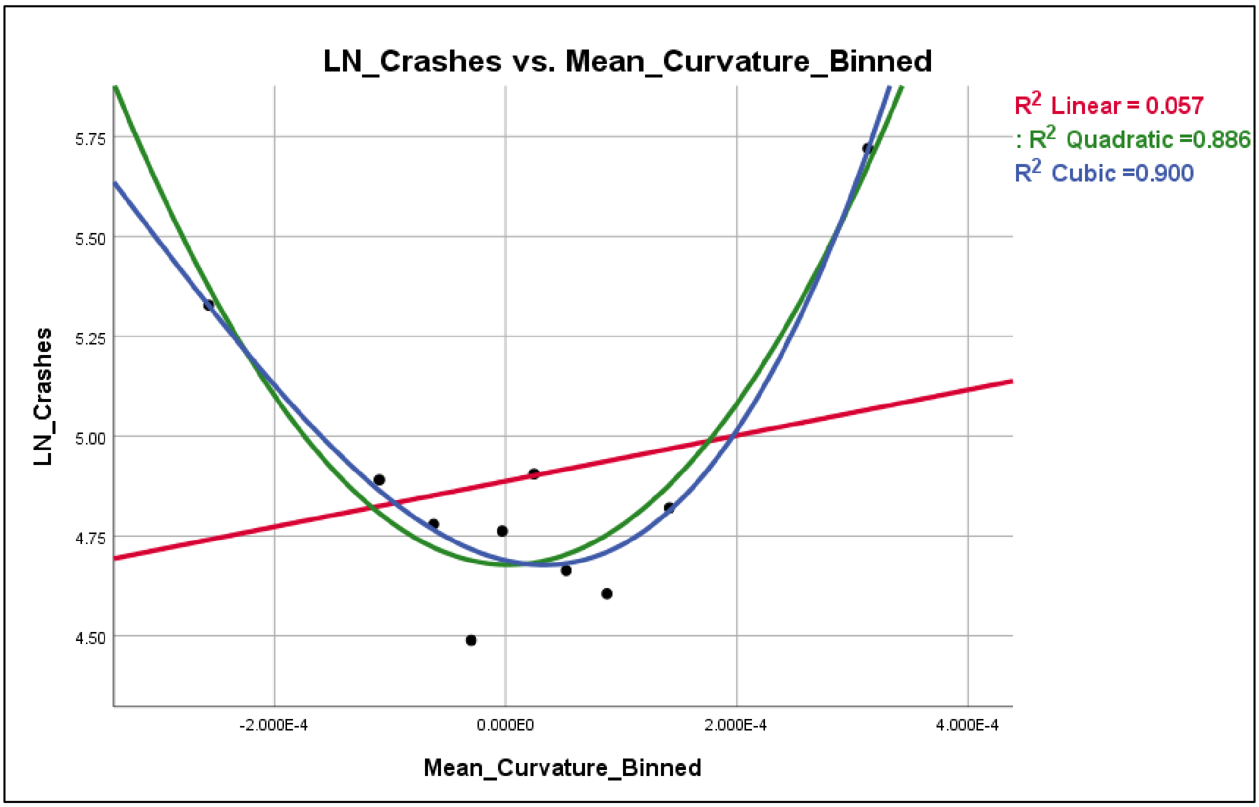

Figure F.30: Scatterplot LN_Crashes vs. Average Mean_Curvature_Binned for Covariance Scale Determination

APPENDIX G: Negative Binomial Regression Analysis

Patch Length = 1,500 ft

Table G.1: Model Testing for Patch Length = 1,500 ft

Model ID#	Number of Explanatory Variables	Variables	p-value	AIC
A1	1	AADT	0.000	3,592
A2	1	GC	0.204	
A3	1	GC^2	0.659	
A4	1	GC^3	0.677	
A5	1	MC	0.542	
A6	1	MC^2	0.682	
A7	1	MC^3	0.677	
A8	2	GC	0.000	3,613
		GC^2	0.000	
A9	2	GC	0.000	3,615
		GC^3	0.000	
A10	2	GC^2	0.000	3,672
		GC^3	0.000	
A11	3	GC	0.000	3,586
		GC^2	0.000	
		GC^3	0.000	
A12	2	MC	0.000	3,717
		MC^2	0.001	
A13	2	MC	0.001	3,717
		MC^3	0.001	
A14	2	MC^2	0.001	3,717
		MC^3	0.001	
A14	3	MC	0.057	
		MC^2	0.051	
		MC^3	0.062	
A15	4	AADT	0.000	3,526
		GC	0.000	
		GC^2	0.000	
		GC^3	0.000	
A16	3	AADT	0.000	3,577
		MC	0.000	
		MC^2	0.000	
A17	6	AADT	0.000	
		GC	0.000	

Model ID#	Number of Explanatory Variables	Variables	p-value	AIC
		GC^2	0.000	
		GC^3	0.000	
		MC	0.006	
		MC^2	0.195	
A18	5	AADT	0.000	3,514
		GC	0.000	
		GC^2	0.000	
		GC^3	0.000	
		MC	0.000	
A19	1	AADT* GC	0.046	3,722
A20	1	AADT*(GC^2)	0.656	
A21	1	AADT*(GC^3)	0.689	
A22	1	AADT* MC	0.549	
A23	1	GC * MC	0.676	
A24	1	(GC^2)* MC	0.677	
A25	1	(GC^3)* MC	0.677	
A26	6	AADT	0.000	3,511
		GC	0.029	
		GC^2	0.004	
		GC^3	0.010	
		MC	0.000	
		AADT* GC	0.020	

Table G.2: Coefficients of Best Model (A18) for Patch Length = 1,500 ft

Variable	Coefficient	Beta	p-value
(Intercept)	-1.6158898868991447	b0	0.000
AADT	0.0003748640404988	b1	0.000
GC	-3,258,298.6756452790	b2	0.000
GC^2	-1,686,373,311,528.25340	b3	0.000
GC^3	-62,891,518,782,965,160.0	b4	0.000
MC	-7.8095296091680980	b5	0.000

$$LN(Crashes) = b0 + b1 \cdot AADT + b2 \cdot GC + b3 \cdot GC^2 + b4 \cdot GC^3 + b5 \cdot MC$$

The equation above corresponds to a 14-year period. In order to convert it to a yearly base, the value LN(14) should be subtracted from the right hand side of the equation:

$$LN(Crashes) = b0 + b1 \cdot AADT + b2 \cdot GC + b3 \cdot GC^2 + b4 \cdot GC^3 + b5 \cdot MC - LN(14)$$

Table G.3: Model Testing for Patch Length = 1,000 ft

Model ID#	Number of Explanatory Variables	Variables	p-value	AIC
B1	1	AADT	0.000	5,103
B2	1	GC	0.008	5,193
B3	1	GC^2	0.064	
B4	1	GC^3	0.071	
B5	1	MC	0.067	
B6	1	MC^2	0.050	5,196
B7	1	MC^3	0.021	5,194
B8	2	GC	0.008	
		GC^2	0.068	
B9	2	GC	0.017	
		GC^3	0.191	
B10	2	GC^2	0.621	
		GC^3	0.784	
B11	3	GC	0.000	5,183
		GC^2	0.000	
		GC^3	0.001	
B12	2	MC	0.575	
		MC^2	0.325	
B13	2	MC	0.691	
		MC^3	0.105	
B14	2	MC^2	0.509	
		MC^3	0.122	
B15	3	MC	0.554	
		MC^2	0.427	
		MC^3	0.123	
B16	4	AADT	0.000	5,087
		GC	0.000	
		GC^2	0.004	
		GC^3	0.010	
B17	2	AADT	0.000	
		MC	0.069	
B18	2	AADT	0.000	5,099
		MC^2	0.017	
B19	2	AADT	0.000	5,099
		MC^3	0.016	
B20	5	AADT	0.000	
		GC	0.000	

Model ID#	Number of Explanatory Variables	Variables	p-value	AIC
		GC^2	0.001	
		GC^3	0.002	
		MC^2	0.091	
B21	5	AADT	0.000	5,077
		GC	0.000	
		GC^2	0.000	
		GC^3	0.000	
		MC^3	0.001	
B22	1	AADT* GC	0.000	5,179
B23	1	AADT*(GC^2)	0.048	5,196
B24	1	AADT*(GC^3)	0.070	
B25	1	AADT* MC^3	0.022	5,194
B26	1	GC * MC^3	0.077	
B27	1	(GC^2)* MC^3	0.086	
B28	1	(GC^3)* MC^3	0.080	
B29	6	AADT	0.000	
		GC	0.004	
		GC^2	0.000	
		GC^3	0.000	
		MC^3	0.002	
		AADT* GC	0.809	
B30	6	AADT	0.000	
		GC	0.000	
		GC^2	0.005	
		GC^3	0.000	
		MC^3	0.000	
		AADT* GC^2	0.163	
B31	6	AADT	0.000	
		GC	0.000	
		GC^2	0.000	
		GC^3	0.000	
		MC^3	0.948	
		AADT* MC^3	0.806	

Table G.4: Coefficients of Best Model (B21) for Patch Length = 1,000 ft

Variable	Coefficient	Beta	p-value
(Intercept)	-1.7903675709256780	b0	0.000
AADT	0.00032774854461651990	b1	0.000
GC	-1,360,470.8069780	b2	0.000
GC^2	-786,107,324,922.35850	b3	0.000
GC^3	-84,173,585,329,077,552.0	b4	0.000
MC^3	2,350.7997806946120	b5	0.001

$$LN(Crashes) = b0 + b1 \cdot AADT + b2 \cdot GC + b3 \cdot GC^2 + b4 \cdot GC^3 + b5 \cdot MC^3$$

The equation above corresponds to a 14-year period. In order to convert it to a yearly base, the value LN(14) should be subtracted from the right hand side of the equation:

$$LN(Crashes) = b0 + b1 \cdot AADT + b2 \cdot GC + b3 \cdot GC^2 + b4 \cdot GC^3 + b5 \cdot MC^3 - LN(14)$$

Table G.5: Model Testing for Patch Length = 400 ft

Model ID#	Number of Explanatory Variables	Variables	p-value	AIC
C1	1	AADT	0.000	7,345
C2	1	GC	0.012	7,475
C3	1	GC^2	0.564	
C4	1	GC^3	0.303	
C5	1	MC	0.000	7,464
C6	1	MC^2	0.002	7,471
C7	1	MC^3	0.002	7,471
C8	2	GC	0.000	7,419
		GC^2	0.000	
C9	2	GC	0.000	7,429
		GC^3	0.000	
C10	2	GC^2	0.001	7,469
		GC^3	0.001	
C11	3	GC	0.000	7,401
		GC^2	0.000	
		GC^3	0.000	
C12	2	MC	0.000	7,454
		MC^2	0.001	
C13	2	MC	0.000	7,458
		MC^3	0.006	
C14	2	MC^2	0.675	
		MC^3	0.817	
C15	3	MC	0.000	7,437
		MC^2	0.000	
		MC^3	0.000	
C16	4	AADT	0.000	7,235
		GC	0.000	
		GC^2	0.000	
		GC^3	0.000	
C17	4	AADT	0.000	7,286
		MC	0.000	
		MC^2	0.000	
		MC^3	0.000	
C18	7	AADT	0.000	7,193
		GC	0.000	
		GC^2	0.000	
		GC^3	0.000	

Model ID#	Number of Explanatory Variables	Variables	p-value	AIC
		MC	0.000	
		MC^2	0.001	
		MC^3	0.029	
C19	6	AADT	0.000	7,196
		GC	0.000	
		GC^2	0.000	
		GC^3	0.000	
		MC	0.000	
		MC^2	0.000	
C20	6	AADT	0.000	7,201
		GC	0.000	
		GC^2	0.000	
		GC^3	0.000	
		MC	0.000	
		MC^3	0.000	
C21	6	AADT	0.000	
		GC	0.000	
		GC^2	0.000	
		GC^3	0.000	
		MC^2	0.523	
		MC^3	0.563	
C22	1	AADT* GC	0.000	7,454
C23	1	AADT*(GC^2)	0.735	
C24	1	AADT*(GC^3)	0.382	
C25	1	AADT* MC	0.000	7,458
C26	1	AADT*(MC^2)	0.002	7,471
C28	1	GC * MC	0.133	
C29	1	GC *(MC^2)	0.007	7,474
C31	1	(GC^2)* MC	0.518	
C32	1	(GC^2)*(MC^2)	0.968	
C34	1	(GC^3)* MC	0.293	
C35	1	(GC^3)*(MC^2)	0.329	
C37	8	AADT	0.000	
		GC	0.000	
		GC^2	0.000	
		GC^3	0.000	
		MC	0.000	
		MC^2	0.002	
		AADT* GC	0.342	
C38	8	AADT	0.000	
		GC	0.000	

Model ID#	Number of Explanatory Variables	Variables	p-value	AIC
		GC^2	0.000	
		GC^3	0.000	
		MC	0.059	
		MC^2	0.005	
		AADT* MC	0.217	

Table G.6: Coefficients of Best Model (C19) for Patch Length = 400 ft

Variable	Coefficient	Beta	p-value
(Intercept)	-2.8918361742282155	b0	0.000
AADT	0.0003887444745069	b1	0.000
GC	-742510.43231766640	b2	0.000
GC^2	-98032335528.572170	b3	0.000
GC^3	-2377047982728732.50	b4	0.000
MC	-65.268028106678460	b5	0.000
MC^2	-264.955793826649260	b6	0.000

$$LN(Crashes) = b0 + b1 \cdot AADT + b2 \cdot GC + b3 \cdot GC^2 + b4 \cdot GC^3 + b5 \cdot MC + b6 \cdot MC^2$$

The equation above corresponds to a 14-year period. In order to convert it to a yearly base, the value LN(14) should be subtracted from the right hand side of the equation:

$$LN(Crashes) = b0 + b1 \cdot AADT + b2 \cdot GC + b3 \cdot GC^2 + b4 \cdot GC^3 + b5 \cdot MC + b6 \cdot MC^2 - LN(14)$$

Patch Length = 200 ft

Table G.7: Model Testing for Patch Length = 200 ft

Model ID#	Number of Explanatory Variables	Variables	p-value	AIC
D1	1	AADT	0.000	9,504
D2	1	GC	0.000	9,671
D3	1	GC^2	0.000	9,711
D4	1	GC^3	0.000	9,721
D5	1	MC	0.000	9,581
D6	1	MC^2	0.000	9,562
D7	1	MC^3	0.000	9,642
D8	2	GC	0.000	9,592
		GC^2	0.000	
D9	2	GC	0.000	9,617
		GC^3	0.000	
D10	2	GC^2	0.000	9,674
		GC^3	0.000	
D11	3	GC	0.000	9,530
		GC^2	0.000	
		GC^3	0.000	
D12	2	MC	0.590	
		MC^2	0.000	
D13	2	MC	0.000	9,575
		MC^3	0.003	
D14	2	MC^2	0.000	9,448
		MC^3	0.000	
D15	3	MC	0.015	9,444
		MC^2	0.000	
		MC^3	0.000	
D16	4	AADT	0.000	9,288
		GC	0.000	
		GC^2	0.000	
		GC^3	0.000	
D17	4	AADT	0.000	9,202
		MC	0.041	
		MC^2	0.000	
		MC^3	0.000	
D18	3	AADT	0.000	9,204
		MC^2	0.000	
		MC^3	0.000	
D19	7	AADT	0.000	

Model ID#	Number of Explanatory Variables	Variables	p-value	AIC
		GC	0.000	
		GC^2	0.000	
		GC^3	0.000	
		MC	0.005	
		MC^2	0.000	
		MC^3	0.160	
D20	6	AADT	0.000	9,072
		GC	0.000	
		GC^2	0.000	
		GC^3	0.000	
		MC	0.014	
		MC^2	0.000	
D21	5	AADT	0.000	9,096
		GC	0.000	
		GC^2	0.000	
		GC^3	0.000	
		MC	0.000	
D22		AADT	0.000	9,077
		GC	0.000	
		GC^2	0.000	
		GC^3	0.000	
		MC^2	0.000	
D23	1	AADT* GC	0.000	9,668
D24	1	AADT*(GC^2)	0.000	9,710
D25	1	AADT*(GC^3)	0.000	9,721
D26	1	AADT* MC	0.000	9,572
D27	1	AADT*(MC^2)	0.000	9,559
D28	1	GC * MC	0.000	9,718
D29	1	GC*(MC^2)	0.000	9,721
D30	1	(GC^2)* MC	0.000	9,723
D31	1	(GC^2)*(MC^2)	0.000	9,726
D32	1	(GC^3)* MC	0.000	9,726
D33	1	(GC^3)*(MC^2)	0.000	9,728
D34	7	AADT	0.000	9,063
		GC	0.000	
		GC^2	0.000	
		GC^3	0.000	
		MC	0.036	
		MC^2	0.006	
		AADT*(MC^2)	0.002	
D35	5	AADT	0.000	9,066

Model ID#	Number of Explanatory Variables	Variables	p-value	AIC
		GC	0.000	
		GC^2	0.000	
		GC^3	0.000	
		MC^2	0.004	
		AADT*(MC^2)	0.001	
D36	6	AADT	0.000	9,053
		GC	0.000	
		GC^2	0.000	
		GC^3	0.000	
		MC	0.000	
		AADT* MC	0.000	
D37	7	AADT	0.000	
		GC	0.179	
		GC^2	0.000	
		GC^3	0.000	
		MC	0.000	
		AADT* MC	0.000	
		AADT* GC	0.279	

Table G.8: Coefficients of Best Model (D36) for Patch Length = 200 ft

Variable	Coefficient	Beta	p-value
(Intercept)	-3.4119717970228263	b0	0.000
AADT	0.0003807129832082	b1	0.000
GC	-103,839.17070852222	b2	0.000
GC^2	-827,114,066.97211050	b3	0.000
GC^3	-1,452,870,025,381.10280	b4	0.000
MC	-285.1232582655780	b5	0.000
AADT* MC	0.1166615708100217	b6	0.000

$$LN(Crashes) = b0 + b1 \cdot AADT + b2 \cdot \mathrm{GC} + b3 \cdot GC^2 + b4 \cdot GC^3 + b5 \cdot MC + b6 \cdot (AADT * MC)$$

The equation above corresponds to a 14-year period. In order to convert it to a yearly base, the value LN(14) should be subtracted from the right hand side of the equation:

$$LN(Crashes) = b0 + b1 \cdot AADT + b2 \cdot GC + b3 \cdot GC^2 + b4 \cdot GC^3 + b5 \cdot MC + b6 \cdot (AADT * MC) - LN(14)$$

Patch Length = 100 ft

Table G.9: Model Testing for Patch Length = 100 ft

Model ID#	Number of Explanatory Variables	Variables	p-value	AIC
E1	1	AADT	0.000	11,554
E2	1	GC	0.008	11,802
E3	1	GC^2	0.020	11,803
E4	1	GC^3	0.020	11,803
E5	1	MC	0.017	11,803
E6	1	MC^2	0.021	11,803
E7	1	MC^3	0.022	11,803
E8	2	GC	0.000	11,609
		GC^2	0.000	
E9	2	GC	0.000	11,629
		GC^3	0.000	
E10	2	GC^2	0.389	
		GC^3	0.385	
E11	3	GC	0.000	11,595
		GC^2	0.000	
		GC^3	0.001	
E12	2	MC	0.000	11,778
		MC^2	0.000	
E13	2	MC	0.000	11,780
		MC^3	0.000	
E14	2	MC^2	0.184	
		MC^3	0.187	
E15	3	MC	0.000	
		MC^2	0.123	
		MC^3	0.812	
E16	4	AADT	0.000	11,329
		GC	0.000	
		GC^2	0.000	
		GC^3	0.000	
E17	3	AADT	0.000	11,527
		MC	0.000	
		MC^2	0.000	
E18	4	AADT	0.000	
		MC	0.000	
		MC^2	0.112	
		MC^3	0.789	
E19	6	AADT	0.000	11,232

Model ID#	Number of Explanatory Variables	Variables	p-value	AIC
		GC	0.000	
		GC^2	0.000	
		GC^3	0.000	
		MC	0.000	
		MC^2	0.000	
E20	5	AADT	0.000	11,304
		GC	0.000	
		GC^2	0.000	
		GC^3	0.000	
		MC	0.000	
E21	5	AADT	0.000	11,242
		GC	0.000	
		GC^2	0.000	
		GC^3	0.000	
		MC^2	0.000	
E22	1	AADT* GC	0.006	11,801
E23	1	AADT*(GC^2)	0.018	11,803
E24	1	AADT*(GC^3)	0.017	11,803
E25	1	AADT* MC	0.015	11,802
E26	1	AADT*(MC^2)	0.019	11,803
E27	1	Gauss* MC	0.021	11,803
E28	1	Gauss*(MC^2)	0.021	11,803
E29	1	(GC^2)* MC	0.020	11,803
E30	1	(GC^2)*(MC^2)	0.021	11,803
E31	1	(GC^3)* MC	0.020	11,803
E32	1	(GC^3)*(MC^2)	0.020	11,803
E33	7	AADT	0.000	
		GC	0.000	
		GC^2	0.000	
		GC^3	0.000	
		MC	0.000	
		MC^2	0.000	
		AADT* GC	0.085	
E34	7	AADT	0.000	
		GC	0.000	
		GC^2	0.000	
		GC^3	0.000	
		MC	0.091	
		MC^2	0.000	
		AADT* MC	0.098	
E35	7	AADT	0.000	11,183

Model ID#	Number of Explanatory Variables	Variables	p-value	AIC
		GC	0.000	
		GC^2	0.000	
		GC^3	0.000	
		MC	0.000	
		MC^2	0.000	
		GC * MC	0.000	
E36	8	AADT	0.000	
		GC	0.000	
		GC^2	0.000	
		GC^3	0.000	
		MC	0.067	
		MC^2	0.000	
		GC * MC	0.025	
		GC * MC^2	0.130	

Table G.10: Coefficients of Best Model (E35) for Patch Length = 100 ft

Variable	Coefficient	Beta	p-value
(Intercept)	-4.2701400273420520	b0	0.000
AADT	0.0003759391060294	b1	0.000
GC	-797,670.95662878790	b2	0.000
GC^2	-62,371,846,845.5081250	b3	0.000
GC^3	-101,722,389,759,530.720	b4	0.000
MC	347.81884723986360	b5	0.000
MC^2	209,377.329346756570	b6	0.000
GC * MC	166,612,202.069389050	b7	0.000

$$LN(Crashes) = b0 + b1 \cdot AADT + b2 \cdot \mathrm{GC} + b3 \cdot GC^2 + b4 \cdot GC^3 + b5 \cdot MC + b6 \cdot MC^2 + b7 \cdot (GC * MC)$$

The equation above corresponds to a 14-year period. In order to convert it to a yearly base, the value LN(14) should be subtracted from the right hand side of the equation:

$$LN(Crashes) = b0 + b1 \cdot AADT + b2 \cdot GC + b3 \cdot GC^2 + b4 \cdot GC^3 + b5 \cdot MC + b6 \cdot MC^2 + b7 \cdot (GC * MC) - LN(14)$$

REFERENCES

AASHTO (2010). *The Highway Safety Manual*, American Association of State Highway Transportation Officials, Washington D.C.

AASHTO (2011). *A Policy on Geometric Design of Highways and Streets, The Green Book*, 6th Edition, Washington D.C.

Amiridis, K. (2014). *3-D Road Alignment By Applying Differential Geometry and B-Spline Interpolation Curves*, Diploma Thesis, National Technical University of Athens, Athens, Greece.

Amiridis, K. and Psarianos, B. (2015a). *Three Dimensional Road Design By Applying Differential Geometry and Conventional Design Approach Criteria*, 94th Annual Transportation Research Board Meeting (TRB).

Amiridis, K. and Psarianos, B. (2015b). *Direct Calculation of Water Film Paths as Geodesic Curves on a Three-Dimensional Road Surface to Address Hydroplaning Phenomena*, 5th International Symposium of Highway Geometric Design (ISHGD), June 2015.

Amiridis, K. (2016). *Safety-Based Guidelines for Left-Turn Phasing Decisions with Negative Binomial Regression*, Master's Thesis, University of Kentucky, Lexington, KY, USA.

Amiridis, K. and Psarianos, B. (2016). *Calculation of the available 3-d sight distance by modeling the roadway as a 3-d B-spline surface*, Advances in Transportation Studies an International Journal, 2016 Special Issue, Vol. 2.

Amiridis, K., Psarianos, B. and Stamatiadis, N. (2016). *Generic Methodology for 3-D Available Sight Distance Calculation*, ASCE International Conference on Transportation & Development (ICTD), June 2016.

Amiridis, K., Stamatiadis, N., Kirk, A. (2017a). *Safety-Based Signalized Intersection Left Turn Phasing Decisions*, Journal of the Transportation Research Board, TRR, 2017.

Amiridis, K., Stamatiadis, N., Kirk, A. (2017b). *Simulated Traffic Conflicts and Crashes for Developing Left Turn Phasing Decisions*, Transportation Research Board, 2017.

Bidulka, S., Sayed, T., Hassan, Y., (2002). *Influence of Vertical Alignment on Horizontal Curve Perception: Phase II Modeling Perceived Radius*, Transportation Research

Record 1796, Transportation Research Board, National Research Council, Washington DC, pp. 24-34.

Borgmann, H. (1976). *Zur Trassierung mit Hilfe von Biege-(Spline-) Linien statisch bestimmt gelagerter Elementarstäbe*, Zeitschrift für Vermessungstechnik 101, 3. [In German]

Brauer, P. (1942). *Zur räumlichen Theorie der Strabe*, Ingenieur Archiv Band XIII. [In German]

Caroni, C., Oikonomou, P. (2017). *Statistical Regression Models*, 2nd Edition. Symeon Editions, Athens, Greece. [In Greek]

Daniels, S., Brijs, T., Nuyts, E., Wets, G. 2010. *Explaining variation in safety performance of roundabouts*. Accident Analysis and Prevention.

Dielmann, T.E. (2005). *Applied Regression Analysis: A Second Course in Business and Economic Statistics*, 4th Edition. Belmont, CA: Brooks/Cole Thomson Learning.

DiVito, M. and Cantisani G. (2010). *D.I.T.S.: A Software for Sight Distance Verification and Optical Defectiveness Recognition*, Proceedings of the 4th International Symposium on Highway Geometric Design, TRB, Valencia, Spain.

Easa, S.M., Gibreel, G.M., El Dimeery, I.A. (1999). *State of the Art of Highway Geometric Design Consistency*, Journal of Transportation Engineering, Vol. 125, No. 4, pp. 305-313.

Easa, S.M., Gibreel, G.M., El Dimeery, I.A. (2001). *Prediction of Operating Speed on Three Dimensional Highway Alignments*, Journal of Transportation Engineering, Vol. 127, No. 1, pp. 21-30.

Easa, S.M., Hassan, Y., Souleyrette, R. (2002). *Three Dimensional Transportation Analysis: Planning and Design*, Journal of Transportation Engineering, Vol. 128, No. 3, pp. 250-258.

ESRI (2017). *ArcGIS Desktop: Release 10.5.1*. Environmental Systems Research Institute, Redlands, CA.

FHWA. Federal Highway Administration (n.d.a). https://www.fhwa.dot.gov/interstate/history. cfm.

FWHA. Federal Highway Administration (n.d.b). http://safety.fhwa.dot.gov/tools/crf/resources/ cmfs/.

FHWA. Federal Highway Administration (n.d.c). https://safety.fhwa.dot.gov/geometric/pubs/mitigationstrategies/chapter3/3_shoulderwidth.cfm

FHWA. Federal Highway Administration (n.d.d). https://safety.fhwa.dot.gov/hsip/ docs/fhwasa17071.pdf

Freising, F. (1949). *Folgerungen aus der Untersuchung des perspektivischen Bildes von Linienelementen der Strabe*, Technische Hochschule Stuttgart, Dissertation. [In German]

Gattis, J.L., Duncan, J. (1995). *Geometric Design for Adequate Operational Preview of Road Ahead*, Transportation Research Record, 1500, Transportation Research Board, National Research Council, Washington DC, pp. 139-145.

Gray, A. (1998). *Modern Differential Geometry of Curves and Surfaces with MATHEMATICA*, CRC Press, New York.

García, A. (2004). *Optimal Vertical Alignment Analysis for Highway design – Discussion*, Journal of Transportation Engineering, Vol. 130, Issue 1, pp. 138.

Hanno, D. (2004). *Effect of the Combination of Horizontal and Vertical Alignments on Road Safety*, M.Sc. Thesis, The University of British Columbia.

Hao, P., S. Zhan-feng, and Teng-Feng, G. (2007). *Integration and Visualization of Highway Alignment Design*, J. Cent. South Univ. (Science and Technology), 35(5), 830-835. [In Chinese]

Hardin J. W., Hilbe J. M. (2012). *Generalized Linear Models and Extensions*, 3rd Edition, Stata Press, College Station, Texas.

Hassan, Y., Easa, S. M. and Abd El Halim, A.O. (1996a). *Analytical Model for Sight Distance analysis on three-dimensional highway alignments*, Transportation Research Record, Vol. 1523.

Hassan, Y., Easa, S.M., Abdelhalim, A.O. (1996b). *Geometric Design Considerations of Combined Horizontal and Vertical Highway Alignments*, Carleton University, Ottawa.

Hassan, Y., Easa, S.M., Abdelhalim, A.O. (1997a). *Design Considerations for Combined Highway Alignments*, Journal of Transportation Engineering, Vol. 123, No. 1, pp. 60-68.

Hassan, Y., Easa, S.M., Abdelhalim, A.O. (1997b). *Modeling Headlight Sight Distance on Three Dimensional Highway Alignments*, Transportation Research Record 1579, Transportation Research Board, National Researh Council, Washington DC, pp. 79-88.

Hassan, Y., Easa, S.M. (1998a). *Design for Sag Vertical Curves in Three Dimensional Alignments*, Journal of Transportation Engineering, Vol. 124, No. 1, pp. 52-58.

Hassan, Y., Easa, S.M. (1998b). *Design Consideration of Sight Distance Red Zones on Crest Curves*, Journal of Transportation Engineering, Vol. 124, No. 4, pp. 343-352.

Hassan, Y., Gibreel, G., Easa, S.M. (2000). *Evaluation of Highway Consistency and Safety: Practical Application*, Journal of Transportation Engineering, Vol. 126, No. 3, pp. 193-201.

Hassan, Y., Easa, S.M. (2000). *Modeling of required Preview Sight Distance*, Journal of Transportation Engineering, Vol. 126, No. 1, pp. 1-20.

Hassan, Y., Easa, S.M. (2003). *Effect of Vertical Alignment on driver perception of horizontal curves*, Journal of Transportation Engineering, Vol. 129, No. 4, pp. 399-407.

Hauer, E., Ng, J.C.N., Lovell, J. (1988). *Estimation of safety at signalized intersections*. Transportation Research Record 1185, 48–61.

Hauer, E. (1997). *Observational before-after studies in road safety*, Pergamon Press, Elsevier Science Ltd., Oxford, England.

Hirst, W.M., Mountain, L.J., Maher, M.J. 2004. *Sources of error in road safety scheme evaluation: a method to deal with outdated accident prediction models*. Accident Analysis and Prevention 36 (5), 717–727.

Hosmer, D. W., Lemeshow, S., Sturdivant, R. X. (2013). *Applied Logistic Regression*, 3rd Edition, John Wiley & Sons, Inc., Hoboken, New Jersey.

IHSDM. Interactive Highway Safety Design Model. https://highways.dot.gov/safety/interactive-highway-safety-design-model/interactive-highway-safety-design-model-ihsdm

Ismail, K. and Sayed, T. (2007). *New algorithm for calculating 3D available sight distance*, Journal of Transportation Engineering, 133(10), pp. 572-581.

Jha, M., Karri, G., and Kuhn, W. (2010). *A New 3-Dimensional Highway Design Methodology for Sight Distance Measurement*, TRB, Washington DC.

Jha, M., Kumar Karri, G.A. and Kuhn W. (2011). *A New 3-Dimensional Highway Design Methodology for Sight Distance Measurement*, The 90th Annual Meeting of the Transportation Research Board, Washington, DC.

Karlaftis, M., Tarko, A. (1998). *Heterogeneity considerations in accident modeling*. Accident Analysis and Prevention 30 (4), 425–433.

Karri, G. and Jha, M.K. (2007). *A New Method for 3–Dimensional Roadway Design Using Visualization Techniques*, Urban Transport XIII (Urban Transport 2007), C.A. Brebbia et al. (eds.), WIT Press, Southampton, U.K.

Karri, K., Maji, A., and Jha, M. (2012). *Optimizing geometric elements of a 3-dimensional alignment in a single stage highway design process*. Washington DC, TRB.

Kentucky State Police. http://crashinformationky.org/KCAP/KYOPS/SearchWizard.aspx

Kim, D. G., and Lovell, D.J. (2010). *A Procedure for 3-D Sight Distance Evaluation using Thin Plate Splines*, proceedings of the 4th International Symposium on Highway Geometric Design, Valencia, Spain, June 2-5, 2010.

Kühn, W. (2002). *Neuartige Modellvorstellungen und Verfahren – Ein Beitrag zur Weiterentwicklung der Entwurfsmethodik für Straßen. Advanced model concepts and methods - A contribution to the development of design methodology for roads*, Technische Universität Dresden, Habilitationsschrift [In German]

Kühn, W. (2012). *3-D Methodology for the Design Process of Safe Rural Highways*, TRB, Washington DC.

Kühn, W. and Jha, M. (2012). *Methodology for checking shortcomings in the three - dimensional alignment*. Washington DC, TRB.

Kentucky Transportation Cabinet. KYTC. https://transportation.ky.gov/Planning/Pages/Traffic-Counts.aspx.

Lamm, R., Choueiri, E. (1987). *Recommendations for Evaluating Horizontal Design Consistency Based on Investigations in the State of New York.* Transportation Research Record 1122, Transportation Research Board, National Research Council, Washington DC, pp. 68-78, Washington DC, TRB.

Lamm, R., Psarianos, B., Mailander, T. (1999). *Highway Design and Traffic Safety Engineering Handbook.* Mc Graw-Hill, New York.

Lamm, R., Smith, B.L. (2004). *Coordination of Horizontal and Vertical Alignment with regard to Highway Esthetics.* Transportation Research Record 1445, Transportation Research Board, National Research Council, Washington DC, pp. 73-85, Washington DC, TRB.

Lipschutz, M. (1981). *Differential Geometry*. Schaum's Outline Series, McGraw-Hill, New York.

Lorenz, H. (1943). *Räumliche Gestaltung von Raumkurven*. Die Strabe Jahrgang 28, 1943. [In German]

Lord, D., Bonneson, J.A. (2007). *Development of accident modification factors for rural frontage road segments in Texas*. Transportation Research Record 2023, 20–27.

Lord, D., Mannering, F. (2010). *The statistical analysis of crash-frequency data: A review ans assessment of methodological alternatives. Transportation Research Part A*, Transportation Research Board, National Research Council, Washington DC.

Lovell, D.J. (1999). *Automated calculation of sight distance from horizontal geometry*, Journal of Transportation Engineering 125(4), pp. 297-304.

Lovell, D.J., Jong, J.-C., and Chang, P.C. (2001). *Improvement to the Sight Distance Algorithm*, Journal of Transportation Engineering, 127(4), 283-288.

Maher, M.J., Summersgill, I. 1996. *A comprehensive methodology for the fitting predictive accident models*. Accident Analysis and Prevention 28 (3), 281–296.

Makanae, K. (2000). *Three-dimensional highway alignment design systems using stereoscopy of aerial photographs and computer graphics*. Proceedings on the Eighth International Conference on: Computing in Civil and Building Engineering, Stanford CA, August 2000.

Makanae, K. (2002). *Functionalization of Highway Alignment for Computer-Based Design*, Proceedings of the Ninth International Conference on Computing in Civil and Building Engineering, April 3-5, 2002, Taipei, Taiwan.

Makanae, K. (2007). *Development of a 3D Highway Modeler using a Driving Game Interface*, Proceedings of the 7th International Conference on Construction Applications of Virtual Reality: October 22-23, 2007.

Maycock, G., Hall, R.D. (1984). *Accidents at 4-Arm Roundabouts*. TRRL Laboratory Report 1120, Transportation and Road Research Laboratory, Crowthorne, UK.

McCullagh, P., Nelder, J. A. (1989). *Generalized Linear Models*. 2nd Edition, Chapman and Hall, London, UK.

Miaou, S.-P. (1994). *The relationship between truck accidents and geometric design of road sections: Poisson versus negative binomial regressions*. Accident Analysis and Prevention 26 (4), 471–482.

Moreno Chou, A., Perez, V., Garcia, A. and Rojas M. (2010). *Optimal 3-D Coordination to Maximize the Available Stopping Sight Distance in Two – Lane Roads*, Proceedings of the 4th International Symposium on Highway Geometric Design, TRB, Valencia Spain.

Myers, J. L., Well, A. D., Lorch, R. F. (2010). *Research Design and Statistical Analysis*, 3rd Edition, New York.

Nehate, G. and M. Rys. (2006). *3D calculation of stopping-sight distance from gps data*, Journal of Transportation Engineering, 132(6), Sep 2006, pp. 691-698.

NHTSA. National Highway Traffic Safety Administration. https://one.nhtsa.gov

Psarianos, B. (1982). *Ein Beitrag zur Entwicklung des räumlichen Trassierungsprozesses von Verkehrswegen und insbesondere von Straben*, Universität Hannover. [In German]

Romero, M.A. and García A. (2007). *Optimal Overlapping of Horizontal and Vertical Curves Maximizing Sight Distance by Genetic Algorithms*, The 86th Annual Meeting of the Transportation Research Board, Washington, DC.

Sayed, T., De Leur, P. (2004). *Predicting the Safety Performance Associated with Highway Design Decisions: A case Study of the Sea to the Sky Highway*, Proceedings of the 2004 Annual Meeting of the Transportation Researh Board, Washington DC, January 2003.

Scheck,H.J. (1973). *Optimierungsberechnungen und Sensitiitätsanalyse als Hilfsmittel bei der Entwurfsbearbeitun von Straben*, Strabenbau und Strabenverkehrstechnik, H. 153. [In German]

Washington, S.P., Karlaftis, M.G., Mannering, F.L., 2010. *Statistical and Econometric Methods for Transportation Data Analysis*, second ed. Chapman Hall/CRC, Boca Raton, FL.

Wolfram Research, Inc., (2018). Mathematica, Version 11.3, Champaign, IL.

Yan, X., Radwan, E., Zhang, F. and Parker J.C. (2008). *Evaluation of Dynamic Passing Sight Distance Problem Using a Finite - Element Model*, Journal of Transportation Engineering, Vol. 134, No.6, pp. 225-235.

Zimmermann, M. (2005). *Increased Safety Resulting from Quantitative Evaluation of Sight Distances and Visibility Conditions of Two-Lane Rural Roads*, Proceedings of the 3rd International Symposium on Highway Geometric Design, TRB, Chicago, USA.

Zuo, W., L. Jian-Pei, and Teng-Feng, G. (2007). *Calculating Method and Appraising Technique of Highway 3-D View Sight Distance*, Journal of Chang'an University (Natural Science Edition), 27(6), 44-48. [In Chinese]

VITA

KIRIAKOS AMIRIDIS

EDUCATION

University of Kentucky: (August 2016 – May 2019 Expected)

Ph.D. in Civil Engineering

Specialization in Highway Geometric Design & Crash Regression Models

PhD Dissertation: «The Use of 3-D Highway Differential Geometry in Crash Prediction Modeling»

Advisor: Dr. Nikiforos Stamatiadis

University of Kentucky: (August 2015 – May 2017)

M.Sc. in Civil Engineering

Specialization in Transportation Engineering

GPA: 4.00/4.00

Rank: #1 out of 60

Master Thesis: «Safety-Based Guidelines for Left-Turn Phasing Decisions with Negative Binomial Regression»

Advisor: Dr. Nikiforos Stamatiadis

University of Kentucky: (August 2015 – May 2017)

Graduate Certificate in Applied Statistics

GPA: 4.00/4.00 (Excellent)

National Technical University of Athens: Athens, Greece (October 2014–May 2015 & October 2017–May 2018)

Department of Applied Mathematics & Physical Sciences

M.Sc. in Applied Mathematics

Major in Numerical Methods for Partial Differential Equations

National Technical University of Athens: Athens, Greece (September 2007- June 2014)

<u>Diploma in Rural & Surveying Engineering</u> – 5 Year Curriculum

Major in Transportation Engineering

Minor in Water Resources

GPA: 8.0/10.0

Rank: #8 out of 100

<u>Diploma Thesis</u>: «3-D Road Design By Applying Differential Geometry and B-Spline Interpolation Curves» Grade: 10.00/10.00 (Excellent)

Advisor: Dr. Basil Psarianos

Apolytirion - 4th Unified Lyceum of Alimos, Alimos, Greece (June 2007)
GPA: 19.0/20.0 (Excellent)

PROFESSIONAL EXPERIENCE

- Langan Engineering & Environmental Services, Manhattan, NY. Staff Engineer Level III; Design of Transportation Infrastructure Projects, Traffic Data Analytics, Traffic Planning & Operations (January 2019 – Present).
- M&J Electrical Contractors, Astoria, NY. Construction Estimator; Quantity Take Offs & Cost Estimating (September 2018 – December 2018).
- Graduate Research Assistant under Professor Nikiforos Stamatiadis at the University of Kentucky (August 2015 December 2018).
- Graduate Research Assistant under Professor Nikiforos Stamatiadis; Project MRI2: Integrated Simulation and Safety for South-eastern Transportation Center, Knoxville, Tennessee (January 2016 – August 2016).
- Teaching Assistant at the University of Kentucky at CE 531 Geometric Design of Roadways (Fall 2017 & Fall 2016).
- Teaching Assistant at the University of Kentucky at CE 331 Transportation Engineering (Fall 2015).

- VTC Services LTD (May 2012 – August 2012); Land Surveying for Metallic Structures and Structures from Reinforced Concrete.
- National Technical University of Athens (June 2011 – July 2011); Land Surveying Works for the Inclusion of a Section of the Nisyros Island in the National Cadastre.
- Summer Internship at the National Bank of Greece (May 2008 – October 2008).

AWARDS & DISTINCTIONS

- Winner of the Southeastern Transportation Center (STC) Outstanding Student of the Year Award in 2017 out of 6 States (Kentucky, Tennessee, Alabama, North Carolina, South Carolina, Florida) across 12 Universities.
- 2nd Best Undergraduate Thesis Award out of all nine Engineering Departments of the National Technical University of Athens for the year 2014 ("Thomaidio Award of Best Undergraduate Thesis").
- Best Undergraduate Thesis Award from the Department of Rural and Surveying Engineering at the National Technical University of Athens for the year 2014.
- Younger Member in the Street and Highway Operations Committee of the Transportation & Development Institute (T&DI) (August 2016 – Present).
- Awarded a fellowship to obtain a Master's in Civil Engineering Program at the University of Kentucky.
- Traffic Bowl Contestant representing the University of Kentucky at the Annual Meeting of Southern District Institute of Transportation Engineers (SDITE) (2016), Nashville, Tennessee.
- Awards of Excellence (over 92.5%) for exceptional student performance and Award of Highest Grade in Class throughout all the classes of Gymnasium (Junior High) and Lyceum (High School) (2002, 2003, 2004, 2005, 2006, 2007).
- Selected from all over Greece to attend the 7th Summer School of Astronomy (August 2006).
- Distinction in the Qualifying Competition "Euclid" of the Greek Mathematical Society for the International Mathematical Olympiad (February 2005).

PUBLICATIONS

Amiridis, K., Stamatiadis, N., Kirk, A., *Left-Turn Phasing Decisions Utilizing Simulated Traffic Conflicts and Historical Crashes*, Advances in Transportation Studies an International Journal (ATS), 2018 Special Issue, Vol. 1., pp 71-82.

Amiridis, K., Stamatiadis, N., Kirk, A., *Safety-Based Signalized Intersection Left-Turn Phasing Decisions*, Journal of the Transportation Research Board (TRR), 2017.

Amiridis, K., Psarianos, B., *Calculation of the available 3-d sight distance by modeling the roadway as a 3-d B-spline surface*, Advances in Transportation Studies an International Journal (ATS), 2016 Special Issue, Vol. 2.

Amiridis, K., Psarianos, B., *Application of Differential Geometry to Solve the Highway Alignment Location Problem*, Scientific Journal of the Orenburg State University (Vestnik OGU), 10(171), 262-270, 2014 Oct.

Amiridis, K., Psarianos, B., *Three Dimensional Road Design By Applying Differential Geometry and Conventional Design Approach Criteria*, Mathematical Design & Technical Aesthetics, 2015, Vol. 3, No1, pp 46-75.

Stamatiadis, N., Kirk, A., Hedges, A., Sallee, T., **Amiridis, K.**, *Left-Turn Guidance*, Kentucky Transportation Center, 2017.

CONFERENCE PROCEEDINGS

Amiridis, K., Stamatiadis, N., Kirk, A. *Development of Left-Turn Phasing Decisions Combining Simulated Traffic Conflicts and Historical Crashes*, 97th Annual Transportation Research Board Meeting (TRB), 2018

Amiridis, K., Stamatiadis, N., Kirk, A. *Safety-Based Decisions for Left-Turn Phasing*, 6th International Road Safety and Simulation Conference (RSS), 2017.

Stamatiadis, N., **Amiridis, K.,** Kirk, A. *Left-Turn Phasing Simulation-Based Decisions*, 6th International Road Safety and Simulation Conference (RSS), 2017.

Stamatiadis, N., Sturgill, R., **Amiridis, K.**, Taylor T. *Estimating Constructability Review Benefits for Highway Projects*, Lean and Computing in Construction Congress – Joint Conference on Computing in Construction, July 2017.

Stamatiadis, N., Sturgill, R., **Amiridis, K.** *Benefits from Constructability Reviews*, World Conference on Transportation Research (WCTR), July 2016.

Amiridis, K., Psarianos, B. *A Direct and Accurate Sight Distance Calculation by Simulating the Road as a Three-Dimensional B-Spline Surface*, 5th International Road Safety and Simulation Conference (RSS), 2015.

Amiridis, K., Psarianos, B. *Direct Calculation of Water Film Paths as Geodesic Curves on a three-Dimensional Road Surface to address Hydroplaning Phenomena*, 5th International Symposium of Highway Geometric Design (ISHGD), June 2015.

Amiridis, K., Psarianos, B. *Three Dimensional Road Design By Applying Differential Geometry and Conventional Design Approach Criteria*, 94th Annual Transportation Research Board Meeting (TRB), 2015.

Amiridis, K., Stamatiadis, N., Kirk, A. *Safety-Based Signalized Intersection Left-Turn Phasing Decisions*, 96th Annual Transportation Research Board Meeting (TRB), 2017.

Amiridis, K., Psarianos, B., Stamatiadis, N. *Generic Methodology for 3-D Available Sight Distance Calculation*, ASCE International Conference on Transportation & Development (ICTD), June 2016.

PROFESSIONAL AFFILIATIONS

- Chartered Member of the Technical Chamber of Greece (October, 2018 Candidate).
- Younger Member in the Street and Highway Operations Committee of the Transportation & Development Institute (T&DI) (August 2016 – Present)
- Member of the American Society of Civil Engineers (ASCE)
- Member of Chi Epsilon Civil Engineering Honor Society, Kentucky Chapter
- Member of the Institute of Transportation Engineers (ITE), UK Student Chapter and International
- Member of Engineers Without Borders, Kentucky Chapter
- Friend of the AFB10 Geometric Design Committee of the Transportation Research Board (TRB)

VOLUNTEERING WORK

Engineers Without Borders (August 2016). Design of a cacao processing plant, where cacao beans will be fermented and dried, community of Coaque, Pedernales, Ecuador. The required information and data were gathered in the assessment trip in August of 2016.

Urban Planning Committee of the Municipality of Alimos, Alimos, Greece (September 2014 – August 2015). Involved in monitoring roadway drainage, pavement design, traffic signals and signs, markings, and lighting.