Journal of Ethiopian Association of Civil Engineers (EACE Journal)
Vol.15, Issue 1 (2024)
www.eacecivil.org | ISSN 3005-6446 (Print)

# Identification of Accident Black Spot Locations Using Statistical Modeling: A Case Study of Addis Ababa

*Mitiku Dinsamo*
*School of Civil and Environmental Engineering, Addis Ababa Institute of Technology, Addis Ababa University, Addis Ababa, Ethiopia*
*Senior Pavement and Material Engineer, Email: mitikudinsamo@gmail.com*

## Abstract

The research aimed to identify accident blackspot locations within Addis Ababa's Road network and develop an accident prediction model using statistical modeling techniques. Out of fifty-two road segments analyzed, twenty were identified as black spots, characterized by high incidences of speeding vehicles relative to the proposed speed limits. Specifically, nine black spots were found on the road from Bole to Meskel Square, seven from Mexico to German Square, and four from Sarbet to Sost Kutir Mazoria. The study utilized various explanatory variables to develop the accident prediction model, including adjusted average peak hour volume, 85th percentile speed, availability of U-turns, and the number of access points. The model was derived using a generalized linear model (GLM), specifically the negative binomial (NB) regression analysis, which was deemed appropriate due to over-dispersion in the data. Key findings from the statistical analysis revealed that the 85th percentile speed had a significant positive relationship with accident frequency, with a coefficient of 0.105 and a p-value of 0.000, indicating high statistical significance. This means that for each unit increase in the 85th percentile speed, the expected log count of accident frequency increases by 0.105. Additionally, the availability of U-turns was found to increase accident frequency, while the number of access points had a negative effect, with a coefficient of -0.158, suggesting that more access points lead to fewer accidents. To evaluate the model's goodness of fit, several measures were employed, including Pearson chi-square, deviance, Akaike Information Criterion (AIC), and Bayesian Information Criterion (BIC). The results indicated that the negative binomial model fit the data well, with Pearson chi-square and deviance values close to the degrees of freedom, confirming the model's adequacy. Furthermore, comparative analysis with the Poisson model showed that the negative binomial model provided a better fit, as evidenced by lower AIC and BIC values. The findings underscore the importance of considering high-risk factors like 85th percentile speed and the availability of U-turns in traffic management and road safety strategies. The developed accident prediction model offers a robust tool for identifying and mitigating accident blackspots, ultimately contributing to safer road networks in urban areas.

**Keywords**: Accident Black Spots, Statistical Modeling, Traffic volume, Spot Speed, Road Safety.

## 1. Introduction

Road traffic accidents are a significant public health concern globally, with severe economic and social consequences, particularly in urban areas. Addis Ababa, the capital city of Ethiopia, experiences a high rate of road traffic accidents, leading to injuries, fatalities, and economic losses. Identifying accident black spot locations and implementing effective countermeasures are crucial steps in mitigating this problem and improving road safety.

This study focuses on utilizing statistical modeling techniques to identify accident black spot locations and propose targeted countermeasures in Addis Ababa. By analyzing accident and traffic data, investigating site characteristics, and developing predictive models, this research aims to provide insights into the underlying factors contributing to road accidents. Understanding these factors is essential for designing interventions that address specific challenges and reduce accident rates.

## 2. Problem Statement

Addis Ababa, the capital city of Ethiopia, grapples with a pervasive issue of road traffic accidents, which pose significant threats to public safety, economic stability, and overall societal well-being. Despite various efforts to address this problem, the city continues to experience a high incidence of accidents, resulting in injuries, fatalities, and substantial financial losses. The identification of accident black spot locations, where a disproportionately high number of accidents occur, remains a crucial challenge for urban planners, traffic engineers, and policymakers seeking to enhance road safety.

Existing research on road traffic accidents in Addis Ababa lacks comprehensive analyses that integrate accident data, traffic characteristics, and site-specific factors to identify and prioritize black spot locations effectively. Furthermore, the absence of targeted countermeasures tailored to the unique conditions of these locations exacerbates the problem, leading to continued traffic-related casualties and economic burdens.

The absence of a systematic approach to identify and address accident black spots hinders efforts to mitigate the impact of road traffic accidents in Addis Ababa. Therefore, there is an urgent need for research that employs advanced statistical modeling techniques to analyze accident data, investigate site characteristics, and develop predictive models for identifying high-risk locations accurately.

Therefore, this study aims to fill this gap by providing a comprehensive analysis of accident black spot locations in Addis Ababa and proposing evidence-based counter measures to improve road safety. By identifying the underlying factors contributing to accidents and prioritizing interventions based on risk levels, this research seeks to inform policy decisions and urban planning initiatives aimed at reducing the incidence of road traffic accidents in the city.

## 3. Research Objectives

The general objective of this research was to develop a statistically significant accident prediction model for mid-block sections of the selected principal arterial streets using the prevalent geometric and traffic variables. Additionally, the specific objectives of the research include:

- To analyze accident data from Addis Ababa and identify accident black spot locations with disproportionately high incidence of road traffic accidents.
- To investigate prevalent geometric and traffic characteristics of identified black spot locations to understand

underlying factors contributing to accidents.

- To develop an accident prediction model using generalized linear modeling (GLM) techniques to predict the likelihood of accidents at identified black spot locations.
- To evaluate the goodness of fit of the developed accident prediction model and assess its effectiveness in identifying and predicting accident occurrences.
- To propose evidence-based counter measures tailored to the identified black spot locations to mitigate the risk of road traffic accidents and improve overall road safety in Addis Ababa.

## 4. Methodology

### 4.1. Introduction

The methodology for this study included the following steps:

1. **Study Area Selection:** Three principal arterial streets in Addis Ababa were selected based on criteria including the availability of abundant accident data, the potential frequency of accidents, functional classification, and median-divided road networks.
2. **Data Collection:**
   - **Accident Data:** Gathered from police stations with detailed records of road traffic accidents.
   - **Traffic Data:** Collected through traffic volume studies at 10 survey stations along the selected routes, including spot speed studies and analysis of traffic volume.
3. **Identification of Black Spot Locations:**
   - Utilized a point weightage approach to identify the top 20 black spot locations based on accident frequency.
   - Analyzed traffic volume, average daily traffic, and accident data to determine significant parameters for model selection.
4. **Site Investigation Survey:**
   - Conducted on mid-block sections of the selected arterial streets to gather information on geometric design, traffic patterns, and safety provisions.
5. **Model Development:**
   - Employed generalized linear modeling (GLM) techniques to develop an accident prediction model using identified significant parameters.
   - Defined black spot locations statistically based on accident frequency and engineering judgment.
6. **Model Evaluation:**
   - Assessed the goodness of fit of the developed accident prediction model using appropriate statistical measure.
   - Validated the model's effectiveness in predicting accident occurrences at black spot locations.

This comprehensive methodology integrates data collection, statistical analysis, and on-site investigations to identify, model, and address accident black spot locations in Addis Ababa systematically. By combining quantitative and qualitative approaches, the study aims to provide insights and recommendations for improving road safety in the city. Pictorially, it can be presented in Figure 1.

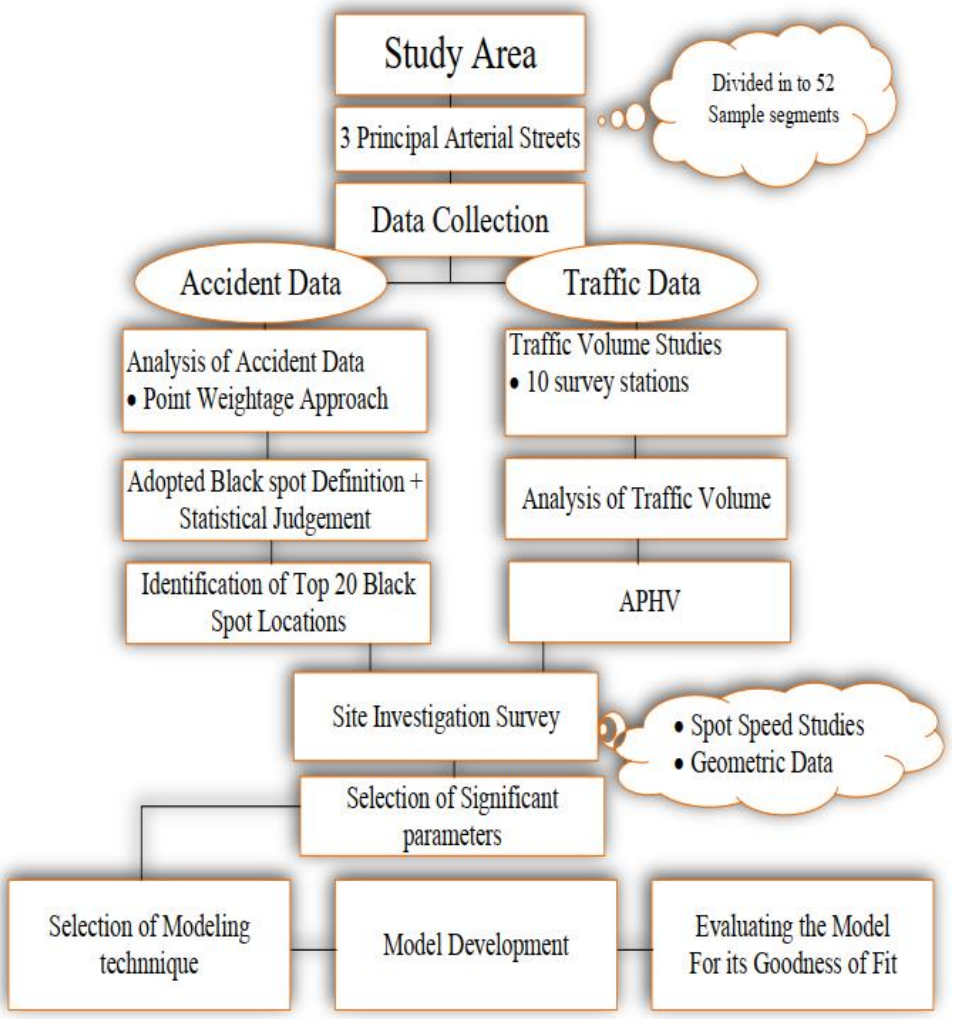



Figure 1 Methodology of the Investigation

### 4.2. Study Area

The selected road networks are detailed in Table 1 with selection criteria previously outlined in the methodology section.

Table 1 The study Road Networks

| No. | Street Name / Label | Street | | Length (km) |
|---|---|---|---|---|
| | | From | To | |
| **1** | Africa Avenue (Bole to Meskel square) | Bole | Meskel square | 4.1 |
| **2** | Mexico Square to German Square | Mexico Square | German Square | 6.0 |
| **3** | Sarbet square to Carl to Sost Kutir Mazoria | Sarbet square | Sost Kutir Mazoria | 3.0 |
| | | | **Total Length (km)** | **13.1** |

### 4.3 Road Network Classifications

Dividing road sections into smaller, specified lengths can significantly enhance the predictive ability and reliability of accident prediction models. This process involves categorizing the roadway network into distinct spots or segments, such as isolated curves, mid-blocks, bridges, and intersections. According to studies by Vistisen (2002) and Overgaard Madsen (2005), it is essential to classify the road network into segments and intersections before identifying accident-prone locations, or "black spots." This preliminary classification ensures that all relevant road characteristics are considered in the analysis.

The Permit Writers Workshop (2002) report emphasizes that spot locations, or short roadway segments, should encompass the area of influence surrounding a hazard. Accident analysis should differentiate between mid-block sections and intersections, considering the unique dynamics of each. For instance, the analysis of intersections should include all accidents within a 250-foot (76-meter) radius from the center, encompassing driveway accidents in this range. This radius should logically account for the stopping sight distance and queue length at intersections (Kwigizile et al., 2013). Conversely, accident analysis within mid-block sections should exclude accidents at intersections and their 76-meter influence zones to properly isolate the impact of mid-block conditions.

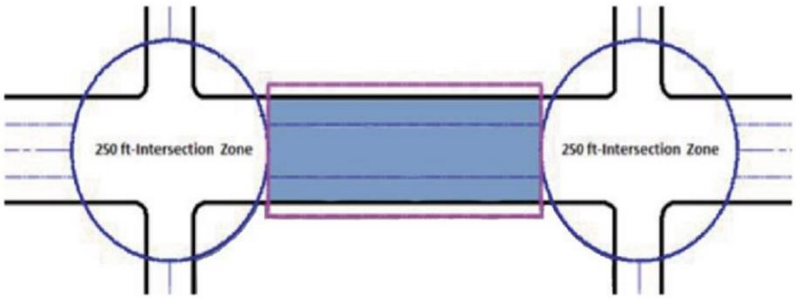



**Figure 2. Influence Zone of an Intersection, Source: Valerian Kwigizile.et.al, (2013)**

Elvik (2007) suggests that dividing road systems into smaller roadway elements can enhance the accuracy of identifying accident-prone areas, or "black spots." However, challenges arise when road segments are divided into non-overlapping sections that may not align with accident patterns, potentially splitting accident peaks across segments. Hauer et al. (2002) address this issue by recommending shorter segment lengths, which helps ensure a more accurate identification of accident concentrations. Best practice guidelines further suggest that the length of roadway segments for black spot analysis should not exceed 0.5 km (Sorensen, 2007).

Following these recommendations, P. Bahat et al. (2013) applied this approach in their study by dividing 3-kilometer road stretches into ten homogeneous sections of 300 meters each for black spot analysis. Typically, segments are defined by a specific length (e.g., 300 feet, 0.1-mile, 1 mile) or the section of a highway between two defined points (Analysis of Accident Statistics, Permit Writers Workshop). Based on these insights,

the current study divides road networks into segments of 200 meters. The classification of mid-blocks into these specified segment lengths was achieved using a combination of site investigation surveys and the Google Earth software package. The classified road networks are summarized in Table 2.

Table 2 Classification of the Study Area

| Network ID | Road Section | Total Length (km) | Study Length (km) | Segment Length (km) | No of Divided Segment | Number of Excluded Intersections |
|---|---|---|---|---|---|---|
| **BM** | Africa Avenue | 4.1 | 3.0 | 0.2 | 20 | 3 |
| **MG** | Mexico Square – German Square | 6.0 | 4.2 | 0.2 | 24 | 3 |
| **SS** | *Sarbet – Sost Kutir Mazoria* | 3.0 | 2.63 | 0.2 | 16 | 2 |
| **Total** | | **13.1** | **9.83** | | **60** | **8** |

### 4.4 Sample Size

The study focuses on a total of 52 road segments across three major road networks in Addis Ababa. The first network, extending from Bole through Welosefer and Dembel to Meskel Square, consists of 20 segments, with three intersections and 17 mid-block segments, including both straight and curved layouts. The second network, from Mexico Square to German Square, includes 24 segments, of which 21 were analyzed while three intersections were excluded from the study. The third network, running from Sarbet through Carl Square to Sost Kutir Mazoria, comprises 16 segments, with two intersections, and 14 segments selected for further analysis. Overall, the research examines a total of 52 road segments.

## 5. Accident Prediction Model

### 5.1 Modelling Parameters

#### 5.1.1 Accident Data

The road traffic accident data utilized in this study were sourced from the respective police departments and the Addis Ababa Police Commission (AAPC). To ensure comprehensive analysis, a three-year period from January 1, 2014, to January 1, 2017, was selected, aligning with established guidelines and previous research findings. The Accident Statistics Guide by the Permit Writers Workshop (2002) and a study by Cheng and Washington (2005) informed this decision, highlighting the adequacy of a three-year timeframe for identifying accident black spots. The dataset included information on severe injury, slight injury, and property damage only (PDO) accidents. Supplementary information from individual accident investigation files was also consulted to enhance data quality, particularly regarding specific accident locations. Overall, the collected data encompassed accident time, date, features, vehicle types, and demographic details of pedestrian victims. A total of 2,497 accidents were recorded across the selected road networks, with 48 fatal accidents and varying levels of severity for injuries, including severe injury, slight injury and PDO incidents.

### 5.1.2 Comprehensive Accident Analysis Using the Point Weightage Approach

This study addresses the need for a comprehensive analysis of road traffic accidents by considering all types, including vehicle-to-vehicle, vehicle-to-pedestrian, and vehicle-to-object crashes, to provide a holistic understanding of the factors influencing road safety. The point weightage method by Mustakim et al. (2008) is employed to classify accidents into four categories: property damage only (PDO), slight injury, severe injury, and fatal accidents, with each category assigned a weight based on severity: fatal accidents (6.0), serious injuries (3.0), slight injuries (0.8), and damage-only incidents (0.2). This method was selected due to its relevance to developing countries, aligning with the study's context, and its focus on emphasizing fatal accidents, thereby capturing the most critical public safety outcomes.

Consequently, the total aggregated frequency of accident was computed by treating each severity using the respective point weights. The general weightage formula is presented in Equation 1 as follows:

$$AN = (6 \times F) + (3 \times S) + (0.8 \times L) + (0.2 \times P) \quad \text{.....(Eqn. 1)}$$

Where F: Fatal accident;
S: Severe Injury;
L: Slight Injury;
P: Property Damage Only

Source: TRL

Many scholars also adopted the above-mentioned point weights due to its comprehensive nature, that it is capable in explaining the accident occurrences by encompassing the four types of accident severity levels. (Mustakim F., 2008, Noorliyana O. et al, 2016). Table 3 provides a summary of accident data across the divided sections, along with the corresponding accident numbers calculated using point weightage over a 3-year period.

Table 3 Overall Accidents Number According to Classified Station

| Station ID | Road layout | Fatal | Severe Injury | Slight Injury | Damage to Property | Accident number by Point Weightage/3yrs |
|---|---|---|---|---|---|---|
| **MG 01** | Straight segment | 0 | 0 | 0 | 0 | 0 |
| **MG 02** | Straight segment | 0 | 3 | 2 | 19 | 14 |
| **MG 03** | Straight segment | 0 | 0 | 0 | 1 | 0 |
| **MG 04** | Vertical Curve | 0 | 0 | 0 | 0 | 0 |
| **MG 05** | Vertical Curve | 0 | 1 | 3 | 17 | 9 |
| **MG 06** | Intersection | 0 | 0 | 0 | 0 | 0 |
| **MG 07** | Horizontal Curve | 0 | 10 | 6 | 87 | 52 |
| **MG 08** | Intersection | 0 | 0 | 0 | 0 | 0 |
| **MG 09** | Straight segment | 0 | 0 | 2 | 5 | 3 |
| **MG 10** | Vertical Curve | 0 | 0 | 0 | 0 | 0 |
| **MG 11** | Horizontal Curve | 2 | 13 | 19 | 130 | 92 |
| **MG 12** | Straight segment | 0 | 1 | 0 | 3 | 4 |
| **MG 13** | Horizontal Curve | 0 | 2 | 2 | 13 | 10 |

| Station ID | Road layout | Fatal | Severe Injury | Slight Injury | Damage to Property | Accident number by Point Weightage/3yrs |
|---|---|---|---|---|---|---|
| **MG 14** | Horizontal Curve | 2 | 7 | 16 | 60 | 58 |
| **MG 15** | Intersection | 0 | 0 | 0 | 0 | 0 |
| **MG 16** | Straight segment | 0 | 0 | 0 | 0 | 0 |
| **MG 17** | Horizontal Curve | 6 | 9 | 6 | 38 | 75 |
| **MG 18** | Horizontal Curve | 3 | 12 | 14 | 80 | 81 |
| **MG 19** | Vertical Curve | 1 | 0 | 1 | 3 | 7 |
| **MG 20** | Horizontal Curve | 2 | 10 | 16 | 88 | 72 |
| **MG 21** | Horizontal Curve | 0 | 0 | 0 | 0 | 0 |
| **MG 22** | Straight segment | 1 | 3 | 6 | 43 | 28 |
| **MG 23** | Horizontal Curve | 0 | 0 | 0 | 0 | 0 |
| **MG 24** | Straight segment | 0 | 0 | 0 | 0 | 0 |
| **BM 01** | Straight segment | 0 | 2 | 1 | 57 | 18 |
| **BM 02** | Straight segment | 0 | 0 | 0 | 3 | 1 |
| **BM 03** | Straight segment | 0 | 6 | 1 | 47 | 28 |
| **BM 04** | Horizontal Curve | 0 | 4 | 3 | 56 | 26 |
| **BM 05** | Straight segment | 0 | 0 | 0 | 24 | 5 |
| **BM 06** | Straight segment | 1 | 1 | 0 | 29 | 15 |
| **BM 07** | Straight segment | 5 | 12 | 4 | 142 | 98 |
| **BM 08** | Horizontal Curve | 6 | 10 | 13 | 94 | 95 |
| **BM 09** | Straight segment | 0 | 0 | 0 | 2 | 0 |
| **BM 10** | Straight segment | 0 | 0 | 0 | 0 | 0 |
| **BM 11** | Horizontal Curve | 1 | 3 | 1 | 50 | 26 |
| **BM 12** | Straight segment | 0 | 8 | 5 | 81 | 44 |
| **BM 13** | Straight segment | 2 | 5 | 7 | 169 | 66 |
| **BM 14** | Straight segment | 4 | 10 | 7 | 102 | 80 |
| **BM 15** | Straight segment | 0 | 0 | 0 | 0 | 0 |
| **BM 16** | Straight segment | 0 | 0 | 0 | 0 | 0 |
| **BM 17** | Vertical Curve | 0 | 0 | 0 | 0 | 0 |
| **BM 18** | Straight segment | 1 | 10 | 14 | 204 | 88 |
| **BM 19** | Straight segment | 0 | 0 | 2 | 27 | 7 |
| **BM 20** | Straight segment | 0 | 0 | 0 | 0 | 0 |
| **SS 01** | Straight segment | 0 | 1 | 0 | 18 | 7 |
| **SS 02** | Horizontal Curve | 0 | 0 | 0 | 0 | 0 |
| **SS 03** | Horizontal Curve | 0 | 0 | 0 | 0 | 0 |
| **SS 04** | Straight segment | 0 | 0 | 0 | 1 | 0 |
| **SS 05** | Intersection | 0 | 0 | 0 | 0 | 0 |
| **SS 06** | Straight segment | 0 | 0 | 0 | 0 | 0 |

| Station ID | Road layout | Fatal | Severe Injury | Slight Injury | Damage to Property | Accident number by Point Weightage/3yrs |
|---|---|---|---|---|---|---|
| **SS 07** | Straight segment | 2 | 1 | 2 | 84 | 33 |
| **SS 08** | Straight segment | 1 | 0 | 0 | 6 | 7 |
| **SS 09** | Straight segment | 2 | 8 | 6 | 177 | 76 |
| **SS 10** | Straight segment | 0 | 1 | 2 | 18 | 8 |
| **SS 11** | Straight segment | 0 | 2 | 3 | 16 | 12 |
| **SS 12** | Intersection | 0 | 0 | 0 | 0 | 0 |
| **SS 13** | Straight segment | 0 | 3 | 2 | 60 | 23 |
| **SS 14** | Straight segment | 2 | 4 | 6 | 50 | 39 |
| **SS 15** | Straight segment | 0 | 0 | 0 | 0 | 0 |
| **SS 16** | Straight segment | 0 | 0 | 0 | 0 | 0 |
| **Total** | | **44** | **162** | **172** | **2104** | **1308** |

By employing this point weightage approach, the study ensures that all types of crashes are considered, allowing for a balanced and comprehensive analysis of how different geometric features impact road safety across various accident types. This methodology not only aligns with established research practices but also provides a robust framework for determining which geometric parameters most significantly affect different types of crashes. Consequently, it offers valuable insights for improving road safety measures through targeted interventions.

### 5.1.3 Traffic Data

The analysis in this research relies on two primary traffic parameters: traffic volume, measured as the average peak hour traffic volume (APHV), and the 85th percentile spot speed. In addition, collection of traffic data i.e., traffic volume and spot speed studies were conducted on the selected stations on 2nd, 4th, 9th, 11th,16th and 18th of May, 2017 where the first two consecutive days were spent to conduct the traffic studies along Bole-Meskel square road network. Whilst, the subsequent separate two days were used to perform the required traffic studies along Mexico square – German square, and Sarbet – Sost Kutir Mazoria respectively. Further, the difficulty in managing simultaneous traffic volume counts on the entire survey stations along the selected routes forced the research to make the traffic studies on similar days of different weeks.

### 5.1.4 Traffic Volume

Traffic volume is a critical factor in crash data analysis, significantly influencing crash frequency. Numerous studies, including those by Vogt (1999), Greibe (2003), Pei et al. (2011), Chiou (2013), and Meng and Qu (2012b), have established a strong correlation between traffic volume and crash occurrences. In light of this, our study incorporates traffic volume, measured as Average Hourly Traffic Volume (AHTV), as a key variable to examine accident patterns along designated road networks, guided by empirical evidence and engineering expertise.

To accurately capture peak hour traffic, a manual traffic count is conducted over two days at specific survey stations positioned along selected routes. These stations are

strategically chosen to ensure representative data collection, adhering to best practices from literature and engineering standards. Traffic counts are performed during peak hours over a six-hour period, divided into morning, mid-day, and afternoon peaks, on similar days of different weeks to maintain consistency (Table 4). The data collection is conducted by a team of eight individuals, following recommended count periods and conditions from relevant guidelines (Sanchari Halder et al., 2013; Sharma, 1994; Smith et al., 2002). This approach provides comprehensive insights into traffic dynamics and accident occurrences, supporting informed decision-making in road safety management.

Table 4 Summarized Average Peak Hour Volume of The Entire Study Areas

| **Adjusted average Peak Hour Volume (APHV) along *Bole* – *Meskel* Square** | | | |
|---|---|---|---|
| **Station No** | **Station name** | **APHV** | **Blackspot station ID** |
| **1** | Friendship Bussines Center | 4864 | BM 03, BM 04, BM 07 and BM 08 |
| **2** | Bole Printing Press | 4732 | BM 11, BM 12, BM 13 and BM 14 |
| **3** | Sunshine Laundary | 4563 | BM 18 |
| **Adjusted average Peak Hour Volume (APHV) along Mexico Square – German Square.** | | | |
| **1** | Ethiopian Tobacco Enterprise | 3857 | |
| **2** | International Evangelical church | 3424 | MG 07 |
| **3** | Vatican Embassy | 2982 | MG 11, MG 14, MG 17 & MG 18 |
| **4** | Mekane Eyesus Management School | 3578 | MG 20, MG 22 |
| **Adjusted average Peak Hour Volume (APHV) along *Sarbet* – *Sost Kutir Mazoria*.** | | | |
| **1** | Tomocca Coffee | 2,450 | |
| **2** | Austria Embassy | 2,390 | SS 07, SS 09 |
| **3** | Dimma Cultural Restaurant | 2,382 | SS 13, SS 14 |

### 5.1.5 Spot Speed Studies

Spot speed studies play a crucial role in understanding traffic dynamics at specific locations, offering insights into speed distribution and percentile speeds of vehicles. Among the various percentiles, the eighty-fifth percentile speed emerges as a commonly adopted parameter for developing accident prediction models, as highlighted by previous research (Kalokota and Seneviratne, 1994; Mustakim et al., 2005; Hauer, 2004; Hong et al., 2005). Typically, a sample size of at least 50 to 100 vehicles is recommended for spot speed studies, with data collection methods including the stopwatch, radar meter, or pneumatic road tube method. Given its efficiency and cost-effectiveness, the stopwatch method is often preferred, particularly in studies focused on black spot locations. Data collection and analysis follow established procedures outlined in relevant literature (Smith et al., 2002), ensuring methodological rigor and reliability in speed-related decision-making processes.

### 5.1.6 Geometric Design Data

Similarly, geometric designs of road networks are widely recognized as key contributors to road traffic accidents. Numerous studies by scholars such as Kalokota and Seneviratne (1994), Hong et al. (2005), and Shankar et al. (1995) have emphasized their significance in accident analysis and prediction modeling. Based on these literatures and engineering judgment, six geometric parameters were presumed to be influential factors for the occurrences of frequent traffic accidents along the selected road networks. And these parameters, obtained through site investigation, include

road layout (straight segments or curved features), lane numbers, lane widths, walkway and median widths, U-turn availability, and number of access points. Geometric characteristics of divided segments were gathered using a checklist for comprehensive analysis. Table 5 presents the selected geometric parameters, the reasons for their selection, and their corresponding impact on the road.

Table 5 Selection of Geometric Parameters and their Potential Impact on Road Safety

| Geometric Parameter | Justification for Selection | Potential Impact on Road Safety |
|---|---|---|
| **Road Layout (Straight or Curved Segments)** | Selected based on its direct influence on visibility, speed, and driver behavior, as indicated in previous studies. | Curved segments can reduce visibility and increase accident risk; straight segments may encourage higher speeds. |
| **Number of Lanes** | Considered to understand the impact of traffic volume and capacity on accident rates. | More lanes can lead to higher traffic volume and congestion, potentially increasing accidents. |
| **Lane Width** | Included due to its effect on vehicle maneuverability and lateral clearance. | Narrow lanes may increase the likelihood of side-swipe accidents; wider lanes can reduce this risk. |
| **Walkway and Median Widths** | Chosen to assess the protection and safety of pedestrians and to prevent head-on collisions. | Wider walkways enhance pedestrian safety; median widths can act as a buffer, reducing head-on collisions. |
| **U-Turn Availability** | Included based on its effect on traffic flow and conflict points at intersections or junctions. | U-turns can introduce conflict points and increase the chances of collisions, particularly in high-speed areas. |
| **Number of Access Points (i.e Driveways)** | Selected due to their role in influencing entry/exit points and potential conflict zones. | More access points can increase the likelihood of conflict zones and potential accidents. |

### 5.1.7 Adopted Black Spot Definition and Identification

#### 5.1.7.1 Black Spot Definition

This research identified the relevant black spot locations through comparing the recorded traffic accidents with the threshold value of observed accident numbers where the adopted threshold or limiting value was taken from the recommendation set by three scholars, in which they used to develop a threshold value for the context of developing countries. Hence, this step is aimed to identify the list of high accident frequency locations within the sample of locations established by dividing the road into sections and spot. Such identification was based on the threshold values of observed numbers of injury accidents at every site in the sample in three consecutive years. Moreover, the threshold values of observed accident numbers at every sample in three consecutive years can be determined using the following equation (Nguyen.et.al, 2016)

$$Threshold\ value = Ma[\bar{X}, m] \ldots.(Eqn\ 2)$$

Where, $\bar{X}$ = average value of accident number

m = median value of the sample.

Therefore, any sample with observed number of accidents higher than the threshold value is listed as a high accident frequency location, in technical term it is said to be a blackspot location.

### 5.1.7.2 Blackspot Identification Strategy

After determining the accident patterns of the studied locations, black spot locations were identified according to the definition provided in Section 5.1.7.1. This identification process involved multiplying the four types of accident severity levels (Fatal, Severe, Slight, and Property Damage Only) by their respective point weights and summing these to calculate the total frequency of accidents at each road station. The severity of the accidents, weighted accordingly, was also considered in this analysis, as recommended by Mustakim (2005). Scholars suggest that including accident severity in traffic accident analysis, using a point-weight approach, allows for the inclusion of all accidents in the identification process, considering both the number and severity of accidents (Taylor and Thompson, 1977; Ogaden, 1996; Overgaard and Madsen, 2005).

Subsequently, frequency analysis on the given accident data was executed using SPSS software package, version 20 for determining the threshold value of the sample. Hence, the result of the SPSS analysis in Table 4-15, has shown that the mean (average) value of accident frequency is much greater than its median. Therefore, as per the employed definition, the threshold value was equivalent to the mean value of the sample data, i.e., 37.2.

*Threshold value = Max (37.2,26)*
*Threshold value = 37.2*

Table 6 Frequency Statistics of the Accident Data (SPSS Output)

| Statistics | | |
|---|---|---|
| **Total number of accidents by point weightage / 3years** | | |
| **N** | Valid | 35 |
| | Missing | 0 |
| **Mean** | | 37.2 |
| **Median** | | 26.0 |

Upon establishing the threshold value for identifying accident-prone areas, the total number of accidents at each station was compared against this threshold, resulting in the identification of fourteen segments that exceeded the threshold and were initially classified as black spot locations. However, statistical principles suggest that using smaller sample sizes can introduce bias, potentially compromising the accuracy and generalizability of the accident prediction model. To mitigate this risk and improve the model's robustness and validity, the analysis was extended to include the top twenty locations with the highest accident frequencies, ensuring a more comprehensive representation of black spot locations and enhancing the model's predictive power and applicability in real-world scenarios.

Table 7 Top Twenty Black Spot Locations Along the Study Area

| Station ID (Black spot Stations) | Fatal | Severe Injury | Slight Injury | Damage to Property | Total Accident using Point weightage approach / 3yr |
|---|---|---|---|---|---|
| **BM 03** | 0 | 6 | 1 | 47 | **28** |
| **BM 04** | 0 | 4 | 3 | 56 | **26** |
| **BM 07** | 5 | 12 | 4 | 142 | **98** |
| **BM 08** | 6 | 10 | 13 | 94 | **95** |
| **BM 11** | 1 | 3 | 1 | 50 | **26** |
| **BM 12** | 0 | 8 | 5 | 81 | **44** |
| **BM 13** | 2 | 5 | 7 | 169 | **66** |
| **BM 14** | 4 | 10 | 7 | 102 | **80** |

| Station ID (Black spot Stations) | Fatal | Severe Injury | Slight Injury | Damage to Property | Total Accident using Point weightage approach / 3yr |
|---|---|---|---|---|---|
| **BM 18** | 1 | 10 | 14 | 204 | **88** |
| **MG 07** | 0 | 10 | 6 | 87 | **52** |
| **MG 11** | 2 | 13 | 19 | 130 | **92** |
| **MG 14** | 2 | 7 | 16 | 60 | **58** |
| **MG 17** | 6 | 9 | 6 | 38 | **75** |
| **MG 18** | 3 | 12 | 14 | 80 | **81** |
| **MG 20** | 2 | 10 | 16 | 88 | **72** |
| **MG 22** | 1 | 3 | 6 | 43 | **28** |
| **SS 07** | 2 | 1 | 2 | 84 | **33** |
| **SS 09** | 2 | 8 | 6 | 177 | **76** |
| **SS 13** | 0 | 3 | 2 | 60 | **23** |
| **SS 14** | 2 | 4 | 6 | 50 | **39** |

## 6. Analysis & Result

### 6.1 Choices of Model Explanatory Variables

Many safety researchers believe that more variables in an accident prediction model (APM) enhance its effectiveness. However, the selection of explanatory variables should be based on theory, as observed in previous models (Fridstrom et al., 1995). Theoretical approaches alone have proven impractical, leading to a combined approach integrating theory and engineering judgment (Hauer, 2004; Elvik, 2007). Elvik's recommended approach suggests including variables that have demonstrated significant influence on accident numbers, can be reliably measured, and are independent of other variables in the model (Elvik, 2007). Furthermore, explanatory variables should exhibit generality, ensuring the model's applicability to similar locations with comparable geometric and traffic characteristics. The principle of parsimony guides model development, advocating for simplicity to explain data variability without overfitting (Sawalha, 2003).

### 6.2 Principle of Parsimony of the Model

The selection of explanatory variables for the prediction model relied on two key criteria. Firstly, correlation matrices were employed to assess the strength of relationships between variables, ensuring that highly correlated variables were not included together in the model. Secondly, following Sawalha's recommendation, variables were evaluated based on their ability to reduce the scaled deviance value at a 95% confidence interval. Statistical significance alone was deemed insufficient for variable inclusion, as highlighted by Elvik (2007), who cautioned against relying solely on this criterion due to its potential for bias. Instead, emphasis was placed on variables with statistically significant parameters, as they contribute meaningfully to explaining the variability of accident data and improve model fit (Sawalha, 2003). Following this approach, eight explanatory variables were chosen for analysis, with correlation analysis used to identify and retain variables with minimal correlation. The selected explanatory variables and their corresponding values are forwarded in Table 8.

Table 8 Contingency Table of the Selected Explanatory Variables

| Station ID of Black Spot Locations | Accident number by point weightage No/ 3 yr (AN/3years) | Lane No (LN) | Road Layout (RL) | Average Walkway width (AWW)(m) | Average Median width (AMW)(m) | No of Access Point (NAP) per divided segment | Availability of U-turn (AUT) | Ln of Average Peak Hour Volume (LnAPHV) | 85$^{th}$ % Spot Speed (85$^{th}$ PS) (km/hr) |
|---|---|---|---|---|---|---|---|---|---|
| **BM 03** | 28 | 4 | Straight segment | 4.83 | 2.90 | 1 | No | 8.49 | 50.00 |
| **BM 04** | 26 | 4 | Horizontal curve | 4.90 | 2.90 | 1 | No | 8.49 | 51.40 |
| **BM 07** | 98 | 4 | Straight segment | 4.95 | 2.90 | 3 | No | 8.49 | 63.70 |
| **BM 08** | 95 | 4 | Straight segment | 4.95 | 2.90 | 3 | No | 8.49 | 67.90 |
| **BM 11** | 26 | 4 | Straight segment | 4.70 | 1.50 | 0 | No | 8.46 | 52.00 |
| **BM 12** | 44 | 4 | Straight segment | 4.95 | 2.90 | 1 | No | 8.46 | 56.10 |
| **BM 13** | 66 | 4 | Horizontal curve | 4.70 | 1.50 | 2 | Yes | 8.46 | 59.10 |
| **BM 14** | 80 | 4 | Straight segment | 4.95 | 2.90 | 2 | No | 8.46 | 61.00 |
| **BM 18** | 88 | 4 | Straight segment | 4.95 | 2.90 | 2 | No | 8.43 | 63.00 |
| **MG 07** | 52 | 3 | Horizontal curve | 3.95 | 1.40 | 1 | Yes | 8.14 | 57.70 |
| **MG 11** | 92 | 3 | Horizontal curve | 4.45 | 0.90 | 1 | Yes | 8.00 | 61.00 |
| **MG 14** | 62 | 3 | Horizontal curve | 4.50 | 0.85 | 0 | No | 8.00 | 60.60 |
| **MG 17** | 75 | 3 | Horizontal curve | 3.95 | 0.60 | 0 | Yes | 8.00 | 56.80 |
| **MG 18** | 81 | 3 | Horizontal curve | 3.95 | 4.10 | 0 | No | 8.00 | 62.60 |
| **MG 20** | 72 | 3 | Horizontal curve | 4.25 | 0.90 | 1 | No | 8.18 | 59.90 |
| **MG 22** | 28 | 3 | Straight segment | 4.25 | 0.90 | 2 | Yes | 8.18 | 55.70 |
| **SS 07** | 33 | 3 | Straight segment | 4.45 | 0.90 | 0 | No | 7.78 | 57.10 |
| **SS 09** | 76 | 3 | Straight segment | 4.45 | 0.90 | 2 | Yes | 7.78 | 60.70 |
| **SS 13** | 23 | 3 | Straight segment | 4.35 | 0.90 | 2 | Yes | 7.78 | 48.70 |
| **SS 14** | 39 | 3 | Straight segment | 4.35 | 0.90 | 2 | Yes | 7.78 | 53.00 |

Table 9 The Hypothesized Geometric and Traffic Explanatory Variables

| $N_o$ | Variable Label | Description of Explanatory Variables | Variables Type |
|---|---|---|---|
| | | **Geometric variables** | |
| **1** | **LN** | Lane Number | Categorical; 0: 3 lanes, 1: 4 lanes |
| **2** | **LW** | Lane width | Continuous (m) |
| **3** | **AMW** | Average Median width | Continuous (m) |
| **4** | **AWW** | Average Walkway width | Continuous (m) |
| **5** | **NAP** | Number of access points | Continuous (count) |
| **6** | **AUT** | Availability of U-Turn | Categorical; 0: Absent; 1: Present |
| **7** | **RL** | Road Layout | Categorical; 0: Straight segment; 1: Horizontal curves. |
| | | **Traffic variables** | |
| **1** | **APHV** | Adjusted average peak hour volume | Continuous (veh/hr) |
| **2** | **85$^{th}$ PS** | 85$^{th}$ Percentile Spot speed | Continuous (km/hr) |

Lane width was excluded from consideration due to its consistent nature across the dataset (Table 9). To further refine the selection of explanatory variables, a correlation matrix was constructed using SPSS software.

Two separate matrices were generated using Pearson and Spearman coefficients. The Pearson matrix assessed relationships among continuous variables such as adjusted average peak hour volume (APHV), average walkway width (AWW), average median width (AMW), number of access points (NAP), and 85th percentile spot speed (85th PS). Meanwhile, the Spearman matrix analyzed interactions among categorical variables like lane number, U-turn availability, and road layout. Interpretation of the correlation matrix followed established guidelines, with coefficients falling between -1 and 1. Relevant interpretations of correlation values were referenced from a standard correlation table (Table 10).

Table 10 Standard Correlation Table

| No | Correlation Type | | Relationship Strength |
|---|---|---|---|
| | **Pearson Correlation** | **Spearman Correlation** | |
| **1** | 0.00-0.19 | 0.00-0.19 | Very weak |
| **2** | 0.20-0.39 | 0.20-0.39 | Weak |
| **3** | 0.4-0.59 | 0.4-0.59 | Moderate |
| **4** | 0.6-0.79 | 0.6-0.79 | Strong |
| **5** | 0.8-1.0 | 0.8-1.0 | Very strong |

Independent explanatory variables with low correlation coefficients ($r_s < 0.5$) with other parameters were selected from the correlation matrix for further analysis. The contingency table for all explanatory variables and the two resulting correlation matrices is presented in Tables 11 and 12. This approach ensures that the selected variables are not highly collinear, thereby improving the robustness of the subsequent analysis.

Table 11 Correlation Matrix Results of Continuous Explanatory Variables

| | | Average Median Width (m) | Average Walkway Width (m) | Number of Access Point | 85th Percentile Spot Speed (Km/hr) | Ln of Average Peak Hour Volume |
|---|---|---|---|---|---|---|
| **Average Median Width (m)** | Pearson Correlation | 1 | 0.494* | 0.179 | 0.271 | 0.632** |
| | Sig. (2-tailed) | | 0.027 | 0.451 | 0.248 | 0.003 |
| | N | 20 | 20 | 20 | 20 | 20 |
| **Average Walkway Width (m)** | Pearson Correlation | 0.494* | 1 | 0.455* | 0.114 | 0.708** |
| | Sig. (2-tailed) | 0.027 | | 0.044 | 0.632 | 0.000 |
| | N | 20 | 20 | 20 | 20 | 20 |
| **Number of Access Point** | Pearson Correlation | 0.179 | 0.455* | 1 | 0.299 | 0.272 |
| | Sig. (2-tailed) | 0.451 | 0.044 | | 0.201 | 0.246 |
| | N | 20 | 20 | 20 | 20 | 20 |
| **85th Percentile Spot Speed (Km/hr)** | Pearson Correlation | 0.271 | 0.114 | 0.299 | 1 | 0.145 |
| | Sig. (2-tailed) | 0.248 | 0.632 | 0.201 | | 0.543 |
| | N | 20 | 20 | 20 | 20 | 20 |
| **Ln of Average Peak Hour Volume** | Pearson Correlation | 0.632** | 0.708** | 0.272 | 0.145 | 1 |
| | Sig. (2-tailed) | 0.003 | 0.000 | 0.246 | 0.543 | |
| | N | 20 | 20 | 20 | 20 | 20 |
| ***. Correlation is significant at the 0.05 level (2-tailed).** | | | | | | |
| ****. Correlation is significant at the 0.01 level (2-tailed).** | | | | | | |

Table 12 Correlation Matrix Result of Categorical Explanatory Variables

| Correlations | | | | | |
|---|---|---|---|---|---|
| | | | **Lane Number** | **Road Layout** | **Availability of U-Turn** |
| **Spearman's rho** | **Lane Number** | Correlation Coefficient | 1.000 | -.328 | -0.533* |
| | | Sig. (2-tailed) | . | 0.158 | 0.015 |
| | | N | 20 | 20 | 20 |
| | **Road Layout** | Correlation Coefficient | -0.328 | 1.000 | 0.167 |
| | | Sig. (2-tailed) | 0.158 | . | 0.482 |
| | | N | 20 | 20 | 20 |
| | **Availability of U-Turn** | Correlation Coefficient | -0.533* | 0.167 | 1.000 |
| | | Sig. (2-tailed) | 0.015 | 0.482 | . |
| | | N | 20 | 20 | 20 |
| ***. Correlation is significant at the 0.05 level (2-tailed).** | | | | | |

The initial stage in developing the model involves constructing a baseline model comprising only exposure variables, as accidents inherently require exposure to occur (Sawalha, 2003). In this study, traffic volume serves as the exposure variable, exhibiting a strong correlation with average walkway width and median width, as indicated by the Bi-variate correlation analysis (Table 11). Consequently, traffic

volume remains a key variable for further analysis, while average walkway width and median width are excluded from consideration. Additionally, the Spearman correlation matrix (Table 12) highlights moderate and weak relationships among categorical parameters such as lane number, availability of U-turn, and road layout.

Lane number significantly influences the other two parameters and is therefore excluded from the analysis. Conversely, the availability of U-turn and road layout exhibit Spearman correlation coefficients below 0.5, warranting their inclusion. Further variable elimination aims to minimize the scaled deviance value, with the availability of U-turn emerging as the most influential variable for inclusion in the final accident prediction model. The significant explanatory variables identified include the 85th Percentile Speed (85th PS), Adjusted Average Peak Hour Volume (APHV), Number of Access Points (NAP), and Availability of U-turn (AUT).

**6.3 Selection of Modeling Method**

Statistical models serve as valuable tools for predicting accident rates by analyzing various factors such as traffic flow, site characteristics, and road geometry. The choice of model depends on data quality and location specificity. In this study, a stochastic model was adopted, considering the random and independent nature of accident occurrences over time. Previous research, notably by Okamoto et al. (1989) and Garber et al. (1990), has demonstrated the suitability of stochastic models for describing traffic crash patterns. Amongst the commonly used count models, Poisson and Negative Binomial distributions are prevalent. However, when the variance of crash frequency exceeds the mean, indicating over-dispersion, Negative Binomial Regression (NBR) is preferred. This approach allows for a statistical parameter, the over-dispersion parameter, to be estimated alongside regression coefficients, offering a more accurate representation of crash data variability. Negative binomial models have been shown to outperform Poisson models in describing the relationship between accident occurrence and road characteristics, as evidenced by previous studies (Chin and Quddus, 2003; Lord and Mannering, 2010). Therefore, based on the observed over-dispersion effect in the data and the consensus in the literature, Negative Binomial Regression (NBR) analysis was employed in this research to develop the accident prediction model, with the dispersion parameter estimated at 1.285.

**6.4 Model Development**

The initial stage of model development involves constructing a basic model comprising exposure variables solely, as outlined in the preceding section. In this study, the exposure variable; traffic volume was retained alongside selected geometric and traffic variables for model development, employing negative binomial regression analysis. The final accident prediction model integrates four key traffic and geometric variables, encompassing one categorical variable, availability of U-turn (AUT), and three continuous variables: number of access points (NAP), adjusted average peak hour traffic volume (APHV), and 85th percentile speed (85th%PS). Following Elvik's framework (2007), explanatory variables in accident prediction models are categorized into those describing exposure to accident risk and those influencing expected accident occurrences per unit of exposure. Accordingly, traffic volume (APHV) serves as the exposure variable, with its natural logarithm utilized as an offset variable in the model. Conversely, the 85th percentile speed, availability of U-turn, and

number of access points are considered risk factors in the analysis conducted using the SPSS software package.

### 6.5 Model Form

According to Mustakim (2005), any Accident Prediction Model (APM) must adhere to two critical conditions. Firstly, its mathematical form should produce logical outcomes, ensuring no negative accident predictions and predicting zero accidents for zero exposure variables, such as road length and traffic volume (AADT). Secondly, the model form must allow for linearization using a known link function to enable generalized linear regression. Mustakim suggests a model form involving the product of exposure measures raised to powers, multiplied by an exponential with other explanatory variables.

Elvik's literature review in 2007 confirms the prevalence of this basic model form in accident prediction models, as used by various researchers (Mountain et al., 1996; Fridstrom, 1999; Gaudry and Lassarre, 2000; Ragnoy, Christensen, and Elvik, 2002; Greibe, 2003). The expected number of accidents, E (λ), is determined by traffic volume (Q) and a set of risk factors (Xi). Traffic volume's impact on accidents is captured by its elasticity (β), representing the percentage change in accidents associated with a 1 percent change in traffic volume. Risk factors' effects are typically modeled exponentially, reflecting the characteristics of Poisson distribution. The log-linear equivalent of the basic form is expressed as:

$$Ln\big(E\ (\lambda)\big) = \alpha + \beta * ln(Q) + \gamma i * X \dots\dots\dots\dots Eqn.3$$

where α, β, and γi represent coefficients of exposure and risk factors, respectively.

Table 13 Coefficients of Variable Estimates (SPSS)

| Parameter Estimates | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Parameter** | **B** | **Std. Error** | **95% Profile Likelihood Confidence Interval** | | **Hypothesis Test** | | | **Exp(B)** | **95% Profile Likelihood Confidence Interval for Exp(B)** | |
| | | | **Lower** | **Upper** | **Wald Chi-Square** | **df** | **Sig.** | | **Lower** | **Upper** |
| **(Intercept)** | -9.754 | .8323 | -11.442 | -8.012 | 137.352 | 1 | .000 | 5.806E-005 | 1.073E-005 | .000 |
| **[AUTurn=0]** | -.511 | .1359 | -.804 | -.234 | 14.117 | 1 | .000 | .600 | .447 | .792 |
| **[AUTurn=1]** | 0[a] | . | . | . | . | . | . | 1 | . | . |
| **NAP** | -.158 | .0696 | -.300 | -.009 | 5.173 | 1 | .023 | .854 | .741 | .991 |
| **85th PS** | .105 | .0150 | .073 | .135 | 48.829 | 1 | .000 | 1.110 | 1.076 | 1.145 |
| **(Scale)** | 1.285[b] | | | | | | | | | |
| **(Negative binomial)** | .042 | .0219 | .013 | .116 | | | | | | |
| ***NB: Dependent Variable: Accident Number by point Weightage / 3 yrs Model: (Intercept), AUTurn, NAP, Eighty fifth PS, offset = LnAPHV*** | | | | | | | | | | |
| ***a. Set to zero because this parameter is redundant.*** | | | | | | | | | | |
| ***b. Computed based on the Pearson chi-square.*** | | | | | | | | | | |

Therefore, the final accident prediction model, developed using the negative binomial approach and expressed in exponential function form, is presented as follows (Eqn. 4):

$$AN/3yr = APHV * e^{(-9.754 - 0.511*(AUT) - 0.158*(NAP) + 0.105*(85th\ PS))} \quad \text{.....Eqn. 4}$$

Where:

APHV = Adjusted average peak hour Volume (Continuous);

AUT = Availability of U-turn (categorical): a value of “0” No and “1” for Yes;

NAP = Number of Access Point (Continuous); and

$85^{th}$ PS = the Eighty Fifth Percentile Speed (Continuous).

The impact of potential explanatory variables is summarized and tabulated in the following table, Table 14.

Table 14 The Significant Explanatory Parameters with Their Respective Impacts

| $N_o$ | Explanatory variables | Assigned Code | Variable type | Significance value | Category | Effect |
|---|---|---|---|---|---|---|
| **1** | Average Peak Hour Volume | APHV | Continuous | 0.000 | Exposure | Positive |
| **2** | $85^{th}$ Percentile Speed | $85^{th}$ PS | Continuous | 0.000 | Risk factor | Positive |
| **3** | Availability of U-Turn | AUT | Categorical | 0.000 | Risk factor | Positive |
| **4** | Number of Access Point | NAP | Continuous | 0.023 | Risk factor | Negative |

### 6.6 Assessment of Goodness of Fit of The Model

#### 6.6.1 Pearson Chi-square test

The Pearson Chi-Square test yielded a value/df of 1.285, which is greater than 0.05, indicating that the model fits the data well. This P-value confirms the adequacy of the model's fit. Table 15 presents the detailed goodness-of-fit results for the Negative Binomial model.

Table 15 Outcome of Goodness of Fit test

| **Goodness of Fit[a]** | | | |
|---|---|---|---|
| | Value | df | Value/df |
| **Deviance** | 20.552 | 15 | 1.370 |
| **Scaled Deviance** | 15.989 | 15 | |
| **Pearson Chi-Square** | 19.280 | 15 | 1.285 |
| **Scaled Pearson Chi-Square** | 15.000 | 15 | |
| **Log Likelihood[b,c]** | -80.335 | | |
| **Adjusted Log Likelihood[d]** | -62.500 | | |
| **Akaike's Information Criterion (AIC)** | 170.671 | | |
| **Finite Sample Corrected AIC (AICC)** | 174.956 | | |
| **Bayesian Information Criterion (BIC)** | 175.649 | | |
| **Consistent AIC (CAIC)** | 180.649 | | |
| ***NB: Dependent Variable: Accident Number by point Weightage / 3 yrs Model: (Intercept), AUTurn, NAP, Eighty fifth PS, offset = LnAPHV*** | | | |
| ***a. Information criteria are in small-is-better form.*** | | | |
| ***b. The full log likelihood function is displayed and used in computing information criteria.*** | | | |
| ***c. The log likelihood is based on a scale parameter fixed at 1.*** | | | |
| ***d. The adjusted log likelihood is based on an estimated scale parameter and is used in the model fitting omnibus test.*** | | | |

#### 6.6.2 Deviance

Essentially, deviance is a measure of the discrepancy between the observed data and the values expected under the model in question. Deviance quantifies the discrepancy between the observed data and

the model's predictions, serving as a pivotal metric for evaluating model fit. When comparing deviance to the degrees of freedom, a value close to the degrees of freedom suggests that the model adequately captures the variability in the data without overfitting. In the above table, Table 7, the value of scaled Pearson chi-square and deviance are almost close to the degrees of freedom, indicating a balance between model complexity and explanatory power.

### 6.6.3 AIC and BIC

In evaluating the fit of statistical models, the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) are essential metrics. Both criteria help to balance model fit and complexity by penalizing models with more parameters, thereby preventing overfitting.

#### *6.6.3.1 AIC*

For this measure, a smaller AIC indicates a better model. Initially, without any variables, the AIC value for the NB model was 193.94. After including the significant parameters, the AIC value decreased to 170.61. This reduction demonstrates that the included parameters effectively represent the traffic accident data, improving the model's fit.

#### *6.6.3.2 BIC*

Similar to the AIC parameter, the BIC parameter also indicated a significantly better fit for the model after including the selected variables. The BIC value decreased from 195.93 before the inclusion of the parameters to 175.65 after their inclusion. This reduction demonstrates that the model with the selected variables provides a superior fit to the data.

### 6.6.4 Omnibus test

As shown in Table 16, it is noted that the Omnibus test value (P-value) for the given model is <0.000, which is below the specified threshold of 0.05. This indicates that the model is statistically significant, meaning that the predictors used in the model provide a significantly better fit than a model with no predictors (intercept-only model).

Table 16 Result of Omnibus Test

| **Omnibus Test[a]** | | |
|---|---|---|
| **Likelihood Ratio Chi-Square** | df | Sig. |
| **22.771** | 3 | <.000 |
| **Dependent Variable: Accident Number by point Weightage / 3 yrs**<br>**Model: (Intercept), AUTurn, NAP, Eighty fifth PS, offset = LnAPHV** | | |
| **a. Compares the fitted model against the intercept-only model.** | | |

### 6.6.5 Residuals

Residuals represent the differences between observed values and the values predicted by the model. Proper examination of residuals helps in understanding the model's accuracy and identifying any potential issues such as outliers or patterns that the model fails to capture.

The assessment of residuals is a critical step in validating the fit of a statistical model. In this study, standardized Pearson residuals were examined to ensure the model's accuracy. Ideally, these residuals should be normally distributed with a mean of zero and a standard deviation of one. The analysis revealed that all standardized Pearson residuals were within the acceptable range,

specifically below an absolute value of 2, indicating no significant deviations.

Furthermore, the residual chart demonstrated that all observations fell within the 95% confidence interval (Figure 3). These findings confirm that the model fits the data well, without substantial outliers or anomalies, supporting the robustness and reliability of the model's predictions. This thorough assessment of residuals underscores the model's validity and its suitability for predicting accident frequencies in the studied road segments.

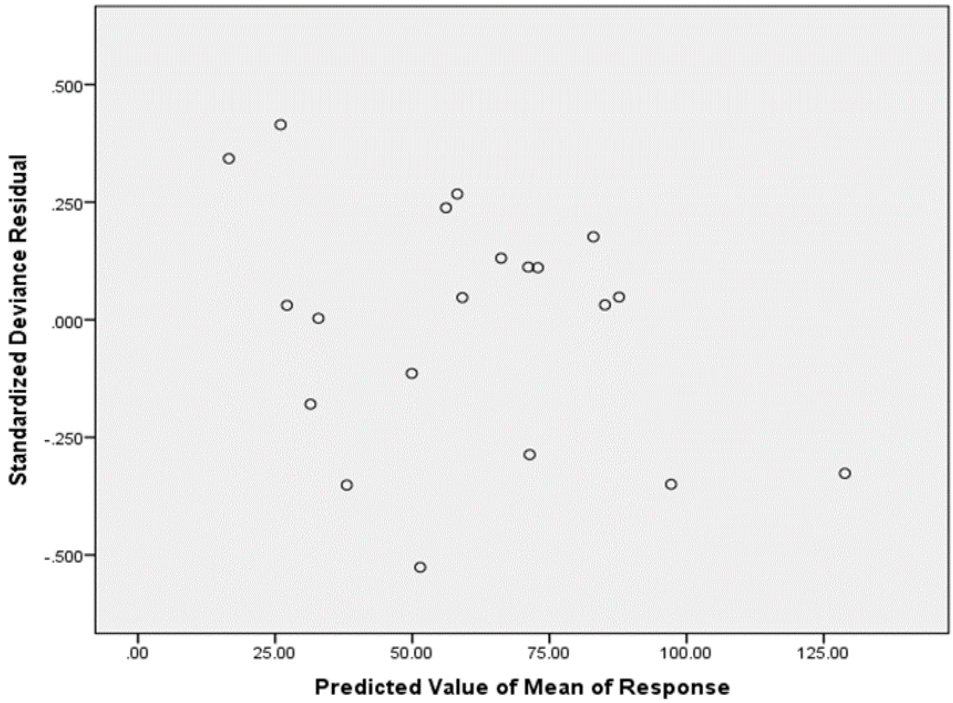



Figure 3 Result of Residual Test

### 6.6.6 Outliers

In negative binomial regression, it is crucial to ensure normality and equal variance of residuals. One method for assessing the impact of outliers is by examining the standardized Pearson residuals. Using SPSS version 20, the computed standardized Pearson residuals for the blackspot locations were all within acceptable limits, specifically below an absolute value of 2. This confirms that the accident data does not include significant outliers. A comprehensive analysis of outliers is detailed in Table 17.

Table 17 Output of standardize Pearson Residual

| Standardized Pearson Residual | | | | | |
|---|---|---|---|---|---|
| | | **Frequency** | **Percent** | **Valid Percent** | **Cumulative Percent** |
| **Valid** | -1.777 | 1 | 5.0 | 5.0 | 5.0 |
| | -1.362 | 1 | 5.0 | 5.0 | 10.0 |
| | -1.336 | 1 | 5.0 | 5.0 | 15.0 |
| | -1.210 | 1 | 5.0 | 5.0 | 20.0 |
| | -1.105 | 1 | 5.0 | 5.0 | 25.0 |
| | -.632 | 1 | 5.0 | 5.0 | 30.0 |

| Standardized Pearson Residual | | | | | |
|---|---|---|---|---|---|
| | | **Frequency** | **Percent** | **Valid Percent** | **Cumulative Percent** |
| | -.445 | 1 | 5.0 | 5.0 | 35.0 |
| | .012 | 1 | 5.0 | 5.0 | 40.0 |
| | .113 | 1 | 5.0 | 5.0 | 45.0 |
| | .139 | 1 | 5.0 | 5.0 | 50.0 |
| | .200 | 1 | 5.0 | 5.0 | 55.0 |
| | .214 | 1 | 5.0 | 5.0 | 60.0 |
| | .491 | 1 | 5.0 | 5.0 | 65.0 |
| | .496 | 1 | 5.0 | 5.0 | 70.0 |
| | .577 | 1 | 5.0 | 5.0 | 75.0 |
| | .809 | 1 | 5.0 | 5.0 | 80.0 |
| | 1.071 | 1 | 5.0 | 5.0 | 85.0 |
| | 1.216 | 1 | 5.0 | 5.0 | 90.0 |
| | 1.233 | 1 | 5.0 | 5.0 | 95.0 |
| | 1.712 | 1 | 5.0 | 5.0 | 100.0 |
| | Total | 20 | 100.0 | 100.0 | |

## 7. Conclusion & Recommendation

### 7.1 Conculsion

As identified in the previous section, twenty out of fifty-two road segments were found to be black spots. The research indicated that these black spots experienced high vehicle speeds exceeding the proposed speed limit. Specifically, nine black spots were identified between Bole and Meskel Square, seven between Mexico and German Square, and four between Sarbet and Sost Kutir Mazoria.

The study aimed to develop an accident prediction model using prevalent traffic and geometric variables. The significant parameters identified were adjusted average peak hour volume, 85th percentile speed, availability of U-turns, and number of access points. The model was derived using a generalized linear model (GLM), particularly negative binomial (NB) regression analysis

Analysis using SPSS revealed a positive relationship between the coefficients of the 85th percentile speed and accident frequency, signifying those higher speeds correlate with increased accident likelihood. Additionally, the absence of U-turns along road segments corresponded to a decrease in accident frequency.

The statistically significant coefficient of 0.105 for the 85th percentile speed suggests that a one-unit increase in speed leads to a 0.105 increase in accident frequency. Similarly, the coefficient of 0.158 for the number of access points indicates that each additional point decreases accident frequency by 0.158 per three years.

### 7.2 Recommendation

The Addis Ababa City Road Authority (AACRA) and the National Ministry of Transport should prioritize allocating sufficient funds and addressing black spot locations based on proposed counter measures to reduce traffic accidents influenced by prevalent traffic and geometric parameters. Governmental organizations must also develop effective maintenance strategies for road signs and markings throughout the city's road network.

Moreover, road safety agencies including AATPMO (Addis Ababa Transport Project Management Office), AARTMA (Addis Ababa Road Traffic Management Agency), and other national institutions, should conduct comprehensive studies to identify

black spots and implement cost-effective countermeasures.

Furthermore, the current accident database system used by police stations is outdated and lacks accuracy. A well-structured, computerized database system should be established to provide detailed and accessible data for insurance companies, hospitals, transport agencies, and research institutions. Researchers should not waste resources seeking input data; instead, relevant stakeholders such as the Ministry of Transport, ERA, AACRA, TPMO, and RTMA should continuously assess the road network and provide necessary data like traffic volume and geometric design to facilitate research.

**Future Research Gap**

This research highlights four critical areas for improvement in future studies: research scope, explanatory variables, data collection methodology, and modeling type selection. The current study, focused on three principal arterial roads in Addis Ababa, suggests expanding the scope to include a broader range of roads and functional classifications for more comprehensive insights. Additionally, the use of presumed traffic and geometric variables indicates a need for incorporating a wider array of variables to enhance the accuracy of accident pattern descriptions. The data collection was limited to short-duration counts on two weekdays and used methods that may introduce errors. Future studies would benefit from longer traffic counts and more precise speed measurement techniques, such as radar or pneumatic tubes. Lastly, while this study employed negative binomial regression to address over-dispersion, exploring other modeling methods could yield even more robust results.

**References**

Akaike, H., 1998. Information theory and an extension of the maximum likelihood principle. In: H. Akaike, ed. Selected Papers of Hirotugu Akaike. New York: Springer, pp. 199-213.

Bandyopadhyaya, R. and Mitra, S., 2011. Comparative Analysis of Hotspot Identification Methods in the Presence of Limited Information. In: 3rd International Conference on Road Safety and Simulation Transportation Research Board, Indianapolis, USA.

Berhanu, G., 2000. Effects of road and traffic factors on road safety in Ethiopia. IST RAPPORT 2000: 11.

Berhanu, G., 2004. Models relating traffic safety with road environment and traffic flows on arterial roads in Addis Ababa. Accident Analysis & Prevention, 36(5), pp.697-704.

Bobade, S.U., Patil, J.R. and Sorate, R.R., 2015, February. Identification of Accidental Black spots on National Highways and Expressways. In: National Conference “ACGT, 2015, pp. 13-14.

Cheng, W. and Washington, S.P., 2005. Experimental evaluation of hotspot identification methods. Accident Analysis & Prevention, 37(5), pp.870-881.

Cui, H., Zhao, X., and Sun, Z., 2022. Identifying accident black spots based on the accident spacing distribution. Journal of Traffic and Transportation Engineering (English Edition), 9(6), pp.1017-1026.

Dereli, M.A., and Erdogan, S., 2017. A new model for determining traffic accident black spots using GIS-aided spatial statistical methods. Transportation Research Part A: Policy and Practice, 103, pp.106-117.

Elvik, R. and Mysen, A., 1999. Incomplete accident reporting: meta-analysis of studies made in 13 countries. Transportation

Research Record: Journal of the Transportation Research Board, (1665), pp.133-140.

Elvik, R., 2008. Comparative analysis of techniques for identifying locations of hazardous roads. Transportation Research Record: Journal of the Transportation Research Board, (2083), pp.72-75.

Elvik, R., 2008. A survey of operational definitions of hazardous road locations in some European countries. Accident Analysis & Prevention, 40(6), pp.1830-1835.

Geremew, G., 2024. Prioritization and identification of accident black spots and analyzing road traffic accidents in the Amhara Region: A case study from Dembecha to the Injibara Highway segment. Accident Analysis & Prevention, 36(5), pp.697-704.

Geurts, K., Wets, G., Brijs, T. and Vanhoof, K., 2004. Identification and ranking of black spots: Sensitivity analysis. Transportation Research Record: Journal of Transportation Research Board, (1897), pp.34-42.

Greibe, P., 2003. Accident prediction models for urban roads. Accident Analysis & Prevention, 35(2), pp.273-285.

Hauer, E. and Persaud, B.N., 1984. Problem of identifying hazardous locations using accident data (No. HS-033 641).

Hong, D., Kim, J., Kim, W., Lee, Y. and Yang, H.C., 2005. Development of traffic accident prediction models by traffic and road characteristics in urban areas. In: Proceedings of the Eastern Asia Society for Transportation Studies, Vol. 5, pp. 2046-2061.

Lloyd-Smith, J.O., 2007. Maximum likelihood estimation of the negative binomial dispersion parameter for highly overdispersed data, with applications to infectious diseases. PLoS One, 2(2), p.e180.

Malaysia, H.O., 2008. Blackspot Study and Accident Prediction Model Using Multiple Liner Regression. Construction in Developing Countries, p.121.

Manan, W. and Nadiah, W., 2011. Accident prediction model at un-signalized intersections using multiple regression method. Doctoral dissertation, Universiti Tun Hussein Onn Malaysia.

Miaou, S.P. and Lum, H., 1993. Modeling vehicle accidents and highway geometric design relationships. Accident Analysis & Prevention, 25(6), pp.689-709.

Murthy, N.S.S.R. and Srinivasa, R.R., 2015. Development of model for road accidents based on intersection parameters using regression models. International Journal of Scientific and Research Publications, 5(1), p.1.

Mustakim, F., 2005. Treating Hazardous Locations at Federal Route 50. Doctoral dissertation, Universiti Teknologi Malaysia.

Nguyen, H.H., Taneerananon, P. and Koren, C., 2015. Identifying Black Spots Based on Safety Potential - A Suitable Approach to Accident Reduction in Developing Countries. Journal of the Eastern Asia Society for Transportation Studies, 11, pp.2241-2257.

Niveditha, V., Ramesh, A. and Kumar, M., 2015. Development of models for crash prediction and collision estimation - a case study for Hyderabad City. International Journal of Transportation Engineering, 3(2), pp.143-150.

Reddy, B.S., Vardhan Reddy, L.V., and Sreenivasa Reddy, G., 2017. Statistical analysis and treatment of accident black spots: A case study of Nandyal Mandal. IOP Conference Series: Materials Science and Engineering, 225(1), pp.1017-1026.