POLITECNICO DI MILANO

School of Civil, Environmental and Land Management Engineering
Master of Science in Civil Engineering for Risk Mitigation

Master of Science Thesis

Black spot identification in highways: the case study of sp.14 (Rivoltana)

Supervisor : Prof. Lorenzo Mussone

Co-supervisor : Prof. Roberto Notari

Candidate : Eslsheikh Eslam Ahmed A.A

Matricula : 242320

Person code : 10900788

A.Y. 2024 - 2025

Contents

List Of Tables

List Of Figures

List Of Equations

1 Introduction

Road traffic safety remains a central concern in transportation engineering, particularly on arterial roads that serve both regional mobility and local access functions. These corridors often experience complex traffic dynamics due to mixed vehicle composition, frequent access points, and fluctuating demand patterns. Such conditions increase the difficulty of identifying hazardous locations using traditional crash analysis methods based solely on historical crash counts. The Rivoltana Road (SP 14) represents a typical example of a suburban arterial corridor exposed to these challenges. Located in the eastern metropolitan area of Milan, the Rivoltana connects the city to surrounding municipalities and supports a combination of commuter, commercial, and local traffic. Along its alignment, the road exhibits significant variability in geometric design, traffic flow levels, and operational conditions, resulting in non-uniform safety performance across segments. Conventional approaches to road safety assessment often rely on reactive identification of high crash locations. While effective in highlighting past problems, such methods do not fully capture the probabilistic nature of crash occurrence nor the influence of exposure and traffic saturation. As a result, there is growing interest in predictive and simulation-based frameworks that integrate traffic conditions, roadway characteristics, and historical crash patterns to estimate crash likelihood more comprehensively. This thesis develops an original probabilistic crash prediction framework applied to the Rivoltana Road. The approach is based on segment-level analysis and combines traffic flow estimates, geometric characteristics, and historical crash data to simulate crash occurrence through Monte Carlo techniques. The framework utilizes two complementary methods:

1. **Length–Volume-to-Capacity (VC)-Based Model**: This model incorporates segment length and traffic saturation to predict crash likelihood. It blends exposure-based risk (due to segment length) with congestion risk (due to high volume-to-capacity ratios). A Monte Carlo simulation is used to estimate crash probabilities across different road segments, with a tuning parameter that allows for flexibility in the model's emphasis on geometry versus traffic flow conditions.
2. **Type–Segment–VC Matrix-Based Model**: This method organizes road segments into a three-dimensional matrix based on segment type, VC class, and historical crash frequencies. It simulates crashes by selecting segment types, VC classes, and specific segments within these categories, reflecting the distribution of past crash events. This approach ensures that both the geometry and traffic flow patterns are captured in the prediction model.

The proposed methodology aims to identify road segments that consistently exhibit elevated crash susceptibility rather than relying solely on raw crash frequencies. By focusing on statistical stability and likelihood distributions, the framework supports proactive safety planning and provides a quantitative basis for prioritizing interventions along the Rivoltana Road.

1.1 Objective

The primary objective of this thesis is to develop a comprehensive probabilistic crash prediction framework for the Rivoltana Road (SP 14), a major arterial corridor in the eastern metropolitan area of Milan. The goal is to identify road segments with a high likelihood of crash occurrences by integrating traffic flow estimates, geometric characteristics, and historical crash data. This will enable a more data-driven and predictive approach to road safety analysis, moving beyond traditional reactive methods based solely on past crash records.

The specific objectives of the study are:

1. To apply two complementary probabilistic methods for crash prediction:
    - The Length–VC-Based Model, which estimates crash likelihood by combining segment length and traffic volume-to-capacity (VC) ratios, using Monte Carlo simulations to generate crash probability distributions.
    - The Type–Segment–VC Matrix-Based Model, which organizes road segments into a three-dimensional matrix of segment types and VC classes, and simulates crash occurrences based on historical crash frequencies and traffic patterns.
2. To identify high-risk segments along Rivoltana Road based on crash likelihood estimations, considering both geometric exposure (segment length) and traffic saturation (VC ratios) as key risk factors.
3. To compare the performance of both models in terms of their ability to predict actual crash occurrences along the road, using statistical metrics such as Root Mean Square Error (RMSE) .
4. To provide a predictive tool for proactive safety planning, offering insights that can guide road safety interventions and resource allocation, with the potential for extending the approach to other arterial roads with similar traffic dynamics.

By achieving these objectives, this thesis aims to contribute a novel approach to traffic safety analysis that incorporates both historical data and real-time traffic conditions, ultimately enhancing the accuracy of crash risk predictions and supporting safer road design and management.

1.2 Thesis Outline

Apart from the Introduction, the thesis is organized into six more chapters as follows:

- **Chapter 2: Literature Review**

  This chapter reviews the existing literature on traffic safety analysis and crash prediction modeling. Particular attention is given to the relationship between traffic flow, congestion, road geometry, and crash occurrence. Classical and modern modeling approaches—including volume-to-capacity–based indicators, probabilistic methods, and Monte Carlo simulation techniques—are discussed. The chapter identifies methodological gaps in existing studies and establishes the theoretical foundation upon which the proposed framework is developed.

- **Chapter 3: Study Area and Data Description**

  This chapter describes the study corridors used in the thesis, with a primary focus on the Rivoltana Road and a secondary application to the Paullese Road. The geometric, operational, and traffic characteristics of the roads are presented, along with the rationale for their selection. The chapter also details the data sources, including crash records, traffic flow data, and geometric attributes, and describes the preprocessing procedures such as road segmentation, crash–segment association, and data quality filtering.

- **Chapter 4: Traffic Analysis**

  This chapter focuses on the analysis of traffic conditions along the study corridors. Traffic demand, capacity estimation, and temporal variation in flow are examined for both weekdays and weekends. Volume-to-capacity (VC) ratios are computed and classified to characterize congestion levels at the time of crashes. This chapter establishes the operational traffic context that forms a key input to the crash prediction models developed later in the thesis.

- **Chapter 5: Crash Analysis**

  This chapter presents the crash data analysis, including crash verification, spatial filtering, and descriptive statistics. The distribution of crashes is examined with respect to time, location, road geometry, traffic conditions, and environmental factors. Special attention is given to the spatial consistency between crash locations and road segments, ensuring that only reliably georeferenced crashes are used in the modeling phase. The chapter provides key insights into observed crash patterns along the corridors.

- **Chapter 6: Methodology and Crash Prediction Results**

  This chapter introduces the probabilistic crash prediction framework and details the two modeling approaches developed in this thesis:

Method 1, based on a weighted combination of segment length and volume-to-capacity conditions, and

Method 2, based on a hierarchical Type–VC–Segment probability structure.

For each method, the theoretical formulation, implementation steps, and Monte Carlo simulation process are described. Model results are presented, including predicted crash counts, identification of statistically significant segments, and performance evaluation using metrics such as RMSE and coefficient of variation. Parameter tuning and sensitivity analysis are also discussed.

- **Chapter 7: Model Comparison, Validation, and Conclusions**

  The final chapter provides a comparative assessment of the two crash prediction methods and validates their performance using both training and test datasets. Results obtained for the Rivoltana Road are compared with those from the Paullese Road to evaluate the robustness and transferability of the proposed framework. The strengths and limitations of each method are discussed, and the benefits of their complementary use are highlighted. The chapter concludes with a synthesis of the main findings, practical implications for road safety management, and recommendations for future research, including extensions to crash severity modeling and real-time applications.

2 Literature Review

This chapter reviews the key theories and methods used in traffic safety analysis and crash prediction modeling. It explores the relationship between traffic flow and crash occurrence, as well as the role of road geometry and environmental factors in influencing crash risk. Existing probabilistic and simulation-based methods for crash prediction are also discussed, highlighting the strengths and limitations of each approach.

2.1 Ideal, Base, and Actual Capacity

In traffic engineering, understanding the concepts of ideal, base, and actual capacity is essential for designing roads that are both efficient and safe. These capacities reflect the maximum volume of traffic that a road can handle, but they are affected by different factors like road geometry, traffic conditions, and external influences such as weather. These concepts are key when assessing the risk of congestion and crashes on roads like the Rivoltana Road.

Ideal Capacity

Ideal capacity refers to the theoretical maximum number of vehicles that a road segment can accommodate under perfect operating conditions. It assumes optimal roadway and traffic characteristics, including ideal geometric design, homogeneous traffic composition, and the absence of external disturbances such as adverse weather or incidents (Transportation Research Board, 2016; Greenshields, 1935). Specifically, ideal capacity presumes:

o Ideal road geometry (wide lanes, straight alignments, adequate sight distance)
o Homogeneous traffic (uniform vehicle types and consistent driver behavior)
o No external disruptions (no weather effects, no incidents, no work zones)

• **Significance:**
Ideal capacity serves as a benchmark for roadway design and planning, providing an upper theoretical bound of traffic performance under uninterrupted flow conditions. While it represents a useful reference in geometric design and theoretical modeling, ideal conditions are rarely achieved in practice. Consequently, ideal capacity is primarily employed for analytical and comparative purposes rather than for operational performance assessment (Transportation Research Board, 2016; Brilon et al., 2005).

• **Mathematical Representation:**
The ideal capacity can be derived from the fundamental relationship between speed and density described in classical traffic flow theory. Under the Greenshields linear speed–density assumption, ideal flow may be expressed as:

$$Q_{\text{ideal}} = v_{\text{max}} \times k_{\text{max}}$$

Equation 2-1 : fundamental equation of traffic flow

where:

- Q_{ideal}= ideal capacity (vehicles per hour),
- v_{max}= maximum speed (km/h),
- k_{max}= maximum density (vehicles per km).

Base Capacity

Base capacity represents an adjustment of ideal capacity to reflect typical real-world operating conditions, including standard driver behavior, roadway surface characteristics, and prevailing traffic patterns. Unlike ideal capacity, base capacity incorporates realistic operational influences while still assuming relatively stable flow conditions (Transportation Research Board, 2016; Brilon et al., 2005). It accounts for operational characteristics such as:

o Regular traffic fluctuations (e.g., peak and off-peak variations)
o Typical vehicle composition (passenger cars, heavy vehicles, buses)
o Minor delays or operational disturbances

• **Significance:**
Base capacity provides a more realistic benchmark than ideal capacity and is widely used in roadway design and operational performance evaluation. It reflects expected performance under normal conditions rather than under theoretical optimal assumptions. Although less optimistic than ideal capacity, base capacity still assumes reasonably smooth traffic flow with manageable operational constraints (Transportation Research Board, 2016).

• **Adjustment Factors:**
Several factors influence base capacity, including geometric design and traffic composition characteristics. According to standardized procedures in the Highway Capacity Manual, adjustments typically consider:

o Roadway design elements (lane width, shoulder width, curvature, intersection density)
o Traffic composition (proportion of heavy vehicles and buses)
o Driver behavior and operational variability

These factors reduce theoretical capacity to more realistic base values and account for heterogeneity in traffic streams and geometric constraints (Transportation Research Board, 2016;

Mannering & Bhat, 2014). Consequently, base capacity is generally lower than ideal capacity due to these real-world influences.

Actual Capacity

Actual capacity represents the effective real-world capacity of a roadway under prevailing traffic, geometric, and environmental conditions. Unlike ideal or base capacity, actual capacity reflects real-time operational constraints such as congestion, weather, incidents, and traffic composition (Transportation Research Board, 2016; Brilon et al., 2005).

It incorporates influences such as:

- **Traffic congestion:** High demand levels may trigger flow instability and traffic breakdown, reducing effective discharge rates (Kerner, 2004).
- **Weather conditions:** Adverse weather (rain, fog, snow) reduces speed, increases headways, and lowers roadway capacity (Transportation Research Board, 2016).
- **Construction or incidents:** Lane blockages, work zones, and crashes reduce available lanes and disrupt flow continuity, thereby decreasing effective capacity.

• **Significance:**
Actual capacity is the most relevant measure for operational traffic management and safety assessment. It reflects real-time roadway performance and is directly associated with congestion formation and crash risk. When traffic demand exceeds actual capacity, traffic breakdown may occur, resulting in unstable flow conditions, speed variability, and elevated crash likelihood (Kerner, 2004; Golob et al., 2004).

Thus, actual capacity plays a critical role in identifying segments prone to congestion-related crash clusters.

• **Calculation of Actual Capacity:**
Actual capacity is typically modeled as a multiplicative adjustment of base capacity through operational and geometric correction factors, as recommended in the Highway Capacity Manual:

$$C_{actual} = C_{base} \cdot f_w \cdot f_{HV} \cdot f_g \cdot f_M \cdot f_D$$

Equation 2-2 : Highway Capacity Manual (HCM) Capacity Adjustment Model

where:
C_{actual}= actual capacity (veh/h)

C_{base}= base capacity
f_w= lane width adjustment factor
f_{HV}= heavy vehicle factor
f_g= grade (slope) factor
f_M= merging influence factor
f_D= diverging influence factor

These factors reflect reductions in capacity due to heavy vehicles, geometric constraints, and traffic turbulence near merging and diverging areas (Transportation Research Board, 2016; Brilon et al., 2005). Similar adjustment-based formulations have been adopted in capacity modeling research to represent operational degradation under non-ideal conditions.

• **Impact on Safety:**
Exceeding actual capacity often results in traffic breakdown, increased turbulence, and unstable car-following behavior. Such conditions are associated with higher crash probability, particularly rear-end and merging-related collisions (Golob et al., 2004; Lord & Mannering, 2010).

Therefore, distinguishing between base and actual capacity is essential for safety-oriented traffic modeling, as crash risk is more closely related to effective operational capacity than to theoretical design capacity.

2.2 Road Safety and Crash Analysis

Road safety analysis aims to understand the mechanisms that lead to crash occurrence and severity on road networks. Rather than focusing solely on the total number of crashes, modern traffic safety research emphasizes identifying risk patterns, spatial concentration, and traffic conditions under which crashes are more likely to occur. This perspective forms the foundation for predictive crash modeling and the identification of hazardous road segments The interaction between traffic flow characteristics and crash occurrence is a central topic in road safety research. Traffic conditions influence driver behavior, vehicle interactions, and operational stability, all of which contribute to the likelihood and severity of crashes. Understanding this relationship requires examining the fundamental variables that govern traffic dynamics and how they evolve under different operating regimes.

2.2.1 Fundamental Traffic Flow Variables

Traffic flow is commonly described using three interdependent variables: speed, density, and flow. These variables form the basis of classical traffic flow theory and are typically represented through fundamental diagrams.

- **Speed–Density Relationship:**

As traffic density increases, the available spacing between vehicles decreases, resulting in a progressive reduction in average speed. Under low-density conditions, drivers are generally able to maintain their desired speeds with minimal interaction. However, as density rises, vehicle interactions intensify and operational constraints emerge, forcing drivers to reduce speed and adjust their driving behavior. This inverse relationship between speed and density constitutes a fundamental principle of traffic flow theory and has been extensively documented in the literature (Greenshields, 1935; Transportation Research Board, 2016).

- **Flow–Density Relationship:**
  Traffic flow increases with density up to a critical threshold corresponding to roadway capacity, beyond which additional vehicles lead to a decline in operational efficiency. At low and moderate densities, increases in density result in higher flow rates due to improved space utilization. However, once the critical density is exceeded, vehicle interactions intensify, speed decreases, and traffic flow becomes unstable, ultimately reducing throughput. This parabolic relationship between flow and density represents a fundamental component of classical traffic flow theory (Greenshields, 1935; Transportation Research Board, 2016).

- **Speed–Flow Relationship:**
  At low flow levels, traffic operates under free-flow conditions, where speeds remain relatively high and stable due to minimal vehicle interactions. As flow increases toward roadway capacity, speed gradually declines as a result of intensified interactions, reduced maneuverability, and growing operational constraints. Near capacity, small disturbances may trigger instability and rapid speed reductions. This inverse relationship between speed and flow under congested conditions has been widely documented in empirical and theoretical traffic flow studies (Greenshields, 1935; Brilon et al., 2005; Transportation Research Board, 2016).

These fundamental relationships provide the theoretical foundation for understanding how congestion develops and how traffic conditions transition between stable and unstable states.

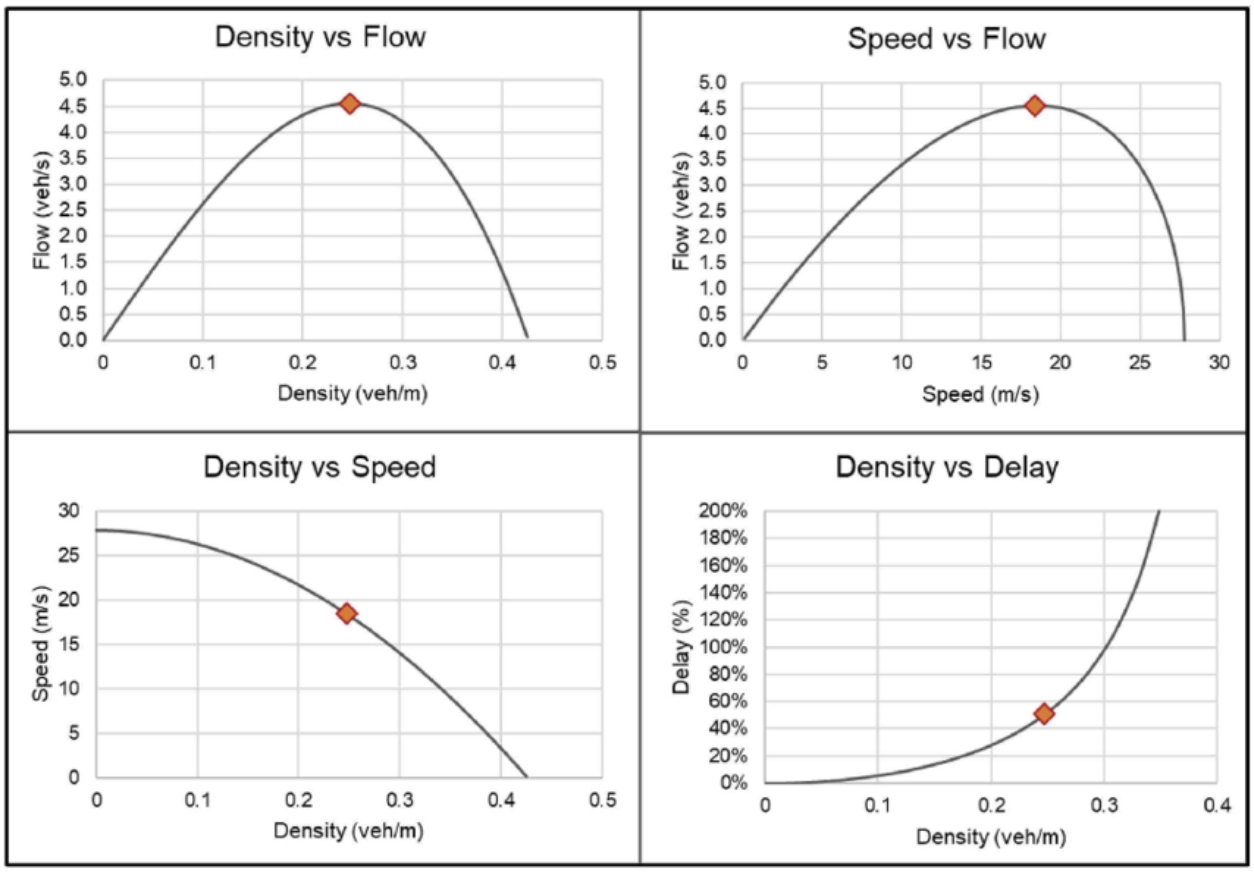


Figure 2-1: Speed–Density-flow relationships (after Greenshields, 1935).

2.2.2 Traffic Volume and Crash Frequency

Numerous studies have demonstrated that crash frequency does not increase linearly with traffic volume. At low traffic volumes, crash frequency generally rises with increasing flow due to greater exposure and a higher number of vehicle interactions. However, as traffic demand approaches roadway capacity, the relationship becomes nonlinear and more complex (Lord & Mannering, 2010; Washington et al., 2010).

Under highly congested conditions, average speeds decrease significantly, which may reduce crash severity despite the persistence of elevated interaction rates. Empirical investigations have identified an inverted parabolic relationship between crash probability and traffic flow, where crash risk tends to peak at intermediate flow levels and decline under severe congestion due to reduced speeds and limited maneuverability (Abdel-Aty & Pande, 2005; Golob et al., 2004; Elvik, 2013).

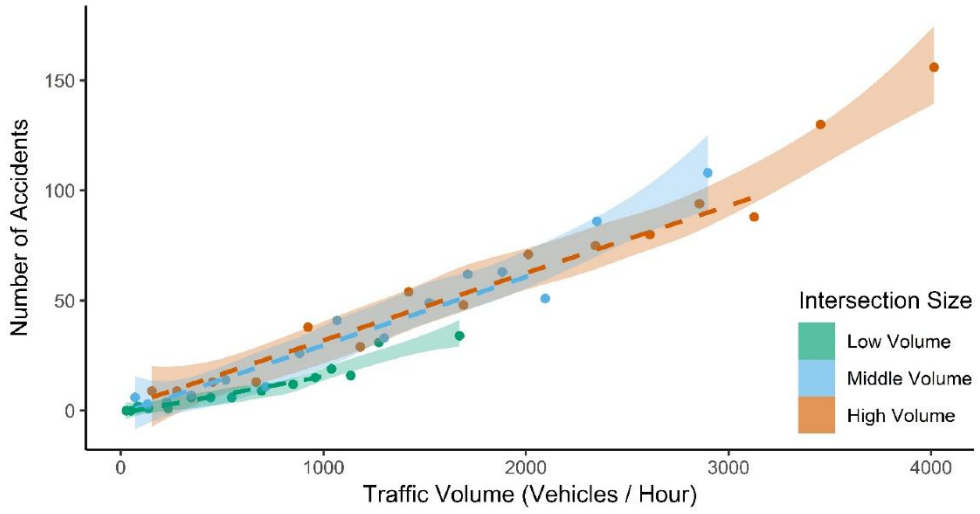


Figure 2-2 : Retallack, A. E., & Ostendorf, B. (2020). Relationship between traffic volume and crash risk.

This nonlinear behavior motivates the classification of traffic conditions into VC-based categories rather than relying solely on average daily traffic values.

2.2.3 Spatial Distribution of Crashes and Segment-Based Analysis

Crashes are not randomly distributed along road networks; rather, they tend to cluster spatially on specific roadway sections due to geometric design characteristics, traffic conditions, access density, and operational factors. This spatial concentration of crashes has been widely documented in traffic safety research and forms the basis for hotspot and segment-level safety analysis (Hauer, 1997; Persaud & Lyon, 2007; Elvik, 2009).

This observation supports the use of segment-based crash analysis, in which roadways are divided into homogeneous segments and crash occurrence is evaluated at the segment level. Such an approach allows for a more accurate association between crash events and roadway characteristics, improving the identification of localized high-risk segments and enabling the integration of traffic, geometric, and operational variables within a consistent analytical framework (Lord & Mannering, 2010; Washington et al., 2010).

Segment-based approaches allow for:
• Better association between crashes and roadway characteristics
• Identification of localized high-risk segments
• Integration of traffic and geometric variables

2.2.4 Traffic Flow States and Crash Risk

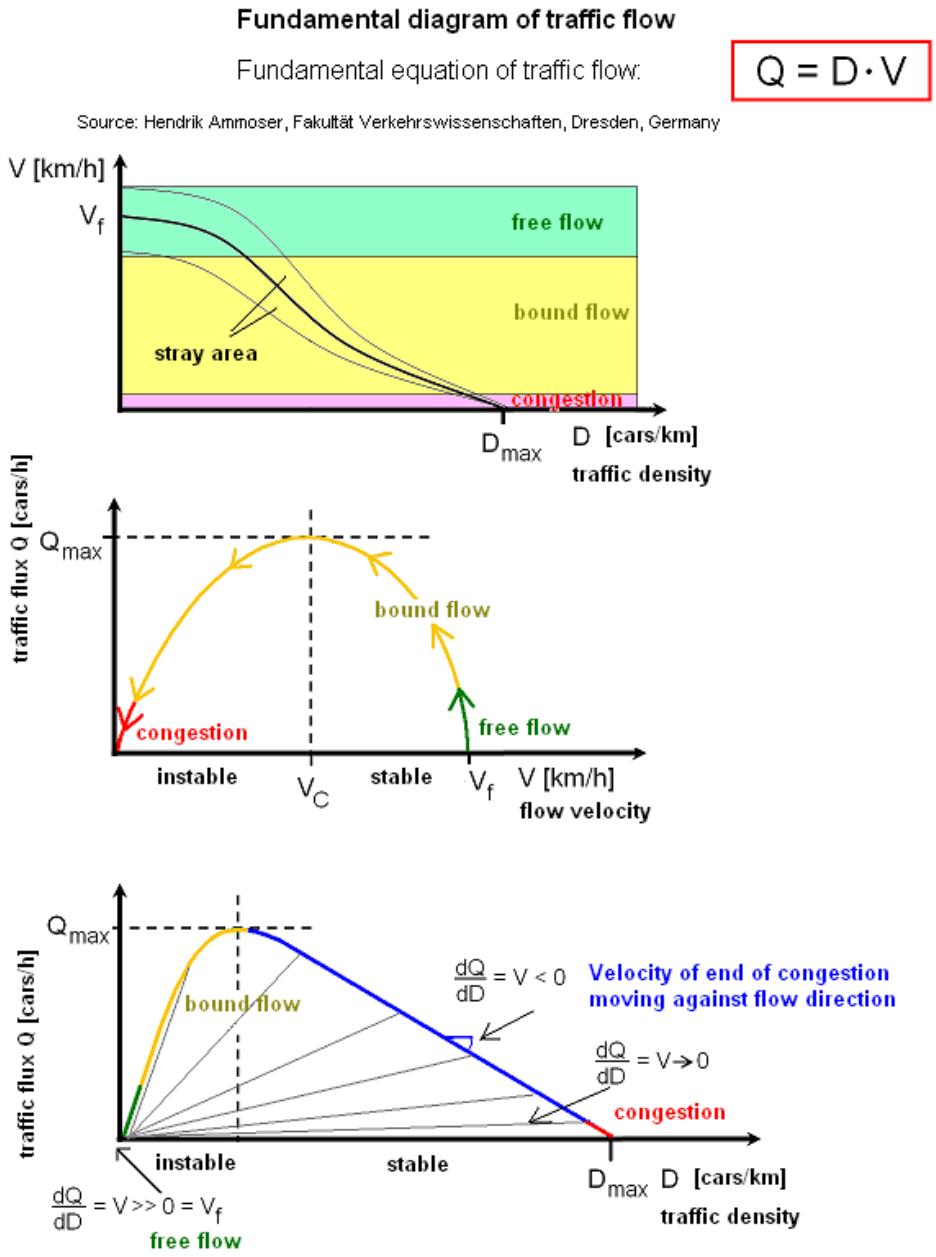
Traffic conditions can be categorized into distinct flow states, commonly described as free-flow, synchronized flow, and congested flow. Each state is associated with specific driving behaviors, vehicle interactions, and safety implications. The three-phase traffic theory and empirical traffic flow studies have extensively documented these operational regimes and their transition dynamics (Kerner, 2004; Transportation Research Board, 2016).

• **Free-flow conditions** are characterized by high speeds and low vehicle interaction. Although crash frequency may be relatively low, the higher operating speeds often contribute to increased crash severity (Elvik, 2009).

• **Synchronized flow** involves moderate speeds and increased vehicle interaction, where lane-changing and car-following behavior intensify. This intermediate regime has frequently been associated with elevated crash probability due to unstable speed adjustments and interaction density (Golob et al., 2004; Abdel-Aty & Pande, 2005).

• **Congested flow** features low speeds and high density. While crash frequency may remain significant due to close spacing, reduced speeds typically limit crash severity (Lord & Mannering, 2010).

Transitions between these flow states are particularly critical from a safety perspective. Sudden changes in speed and spacing increase driver workload, reaction demands, and turbulence in traffic streams, thereby elevating crash likelihood during transitional phases (Kerner, 2004; Brilon et al., 2005).



V_f = "free velocity" - maximum velocity on free lane, selectable by the driver depending on car, skill etc.

V_C = "critical velocity" with maximum traffic flux (about 70...100 km/h)

Figure 2-4 : Kerner, B. S. (2004). The physics of traffic. Springer.

2.2.5 Congestion and Crash Severity

While congestion may contribute to an increase in total crash frequency due to intensified vehicle interactions, it does not necessarily correspond to higher crash severity. Under congested conditions, average travel speeds decline significantly, often resulting in lower-speed collisions that reduce the kinetic energy involved in impact.

As congestion intensifies, crash severity generally decreases because impact energy is directly related to vehicle speed. Lower operating speeds limit the magnitude of forces generated during collisions, thereby reducing the likelihood of severe or fatal injuries (Elvik, 2009; Lord & Mannering, 2010).

Empirical studies have shown that crashes occurring under free-flow conditions—where speeds are higher—are more likely to produce severe outcomes compared to crashes in congested environments (Abdel-Aty & Pande, 2005; Golob et al., 2004). This inverse relationship between congestion level and crash severity highlights the importance of distinguishing between crash frequency and crash severity when evaluating roadway safety performance.

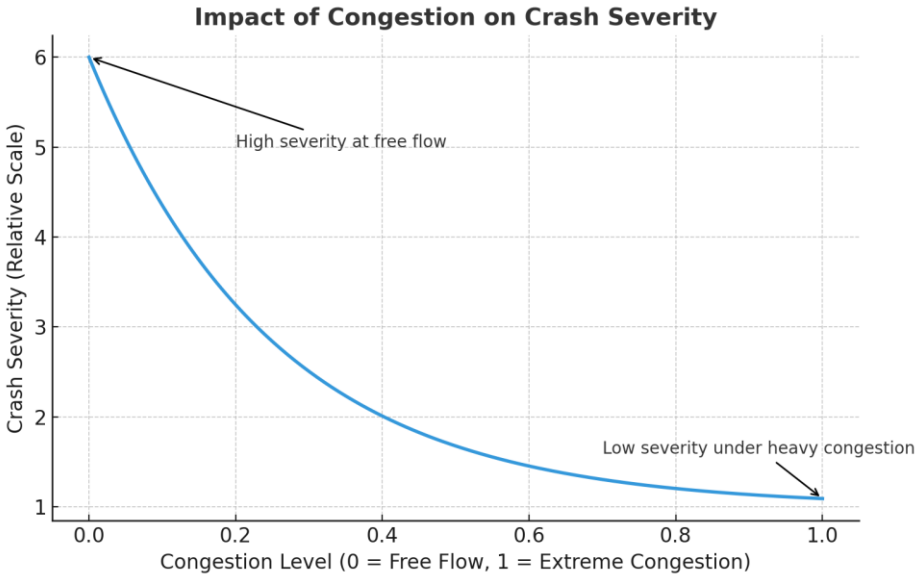


Figure 2-5: Impact of Congestion on Crash Severity (Xu, C., Liu, P., Wang, W., & Li, Z.:2012)

2.2.6 Implications for Predictive Crash Modeling

The reviewed literature highlights three consistent findings:

1. Crash risk is strongly influenced by traffic conditions, not only traffic volume.
2. Road segmentation is essential for meaningful spatial crash analysis.
3. Probabilistic approaches are well-suited for handling uncertainty and variability in crash data.

These insights directly inform the methodological choices adopted in this research, particularly the use of VC-based classification, hourly flow estimation, and probabilistic simulation models for identifying high-risk segments.

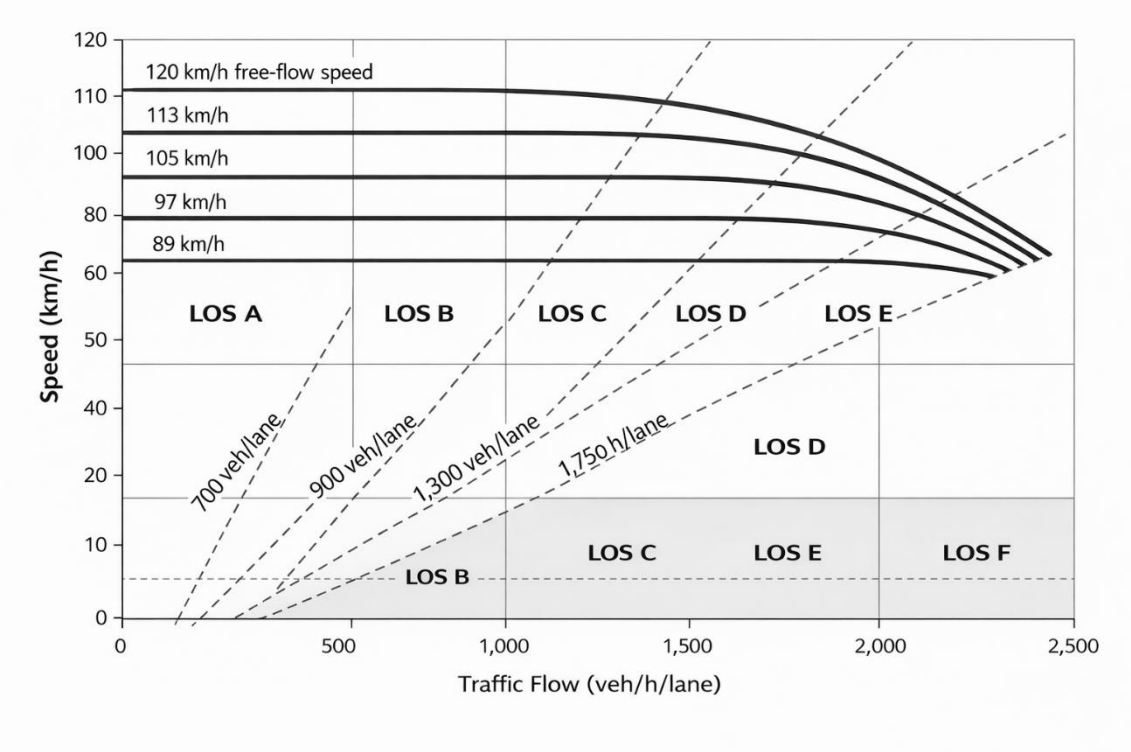
2.3 Traffic Flow, Congestion, and Capacity

Traffic safety analysis cannot be separated from traffic operations. In particular, the concepts of road capacity, congestion, and volume-to-capacity (VC) ratio provide a structured way to interpret traffic conditions beyond simple traffic volume measures. This section reviews these concepts and explains their relevance to crash occurrence and prediction.

2.3.1 Road Capacity and Level of Service

Road capacity is defined as the maximum sustainable hourly flow rate at which vehicles can reasonably be expected to traverse a roadway segment under prevailing geometric, traffic, and environmental conditions. Capacity is typically expressed in vehicles per hour (veh/h) and depends on factors such as lane width, number of lanes, access density, traffic composition, and control conditions. This definition and its operational adjustments are formally established in the Highway Capacity Manual (Transportation Research Board, 2016).

The concept of Level of Service (LOS) is commonly used to qualitatively describe operating conditions, ranging from free-flow conditions (LOS A) to heavily congested conditions (LOS F). LOS provides a descriptive classification of user experience based on performance measures such as speed, density, and delay. However, despite its usefulness in planning and operational assessment, LOS lacks the numerical precision required for quantitative crash modeling, particularly when traffic conditions vary temporally and dynamically. For this reason, continuous indicators such as volume-to-capacity (V/C) ratio are often preferred in safety-oriented analyses (Brilon et al., 2005; Transportation Research Board, 2016).



Urban street class	I	II	III	IV
Range of speeds	70–90 km/h	55–70 km/h	50–55 km/h	40–55 km/h
Typical speed	80 km/h	65 km/h	55 km/h	45 km/h
LOS	**Average travel speed (km/h)**			
A	>72	>59	>50	>41
B	>56–72	>46–59	>39–50	>32–41
C	>40–56	>33–46	>28–39	>23–32
D	>32–40	>26–33	>22–28	>18–23
E	>26–32	>21–26	>17–22	>14–18
F	≤26	≤21	≤17	≤14

LOS: Level of service

Figure 2-6 : Capacity and Level of Service classification based on traffic flow (veh/h)Transportation Research Board (2016). Highway Capacity Manual (HCM). Washington, D.C.

2.3.2 Volume-to-Capacity Ratio (VC) as an Operational Indicator

The Volume-to-Capacity ratio (V/C) is defined as the ratio between traffic demand (volume) and roadway capacity:

$$VC = Q/C$$

where:
Qrepresents traffic flow (veh/h), and
Crepresents roadway capacity (veh/h).

The V/C ratio provides a dimensionless indicator of congestion intensity, allowing traffic conditions to be compared across roadway segments with different geometric and operational characteristics. As a normalized measure, it facilitates safety and performance evaluations independent of absolute capacity values (Transportation Research Board, 2016; Brilon et al., 2005).

Typical interpretations of V/C values are consistent with capacity analysis guidelines and operational stability theory:

• **V/C < 0.25**: Uncongested, free-flow conditions
• **0.25 ≤ V/C < 0.50**: Light to moderate traffic
• **0.50 ≤ V/C < 0.75**: Near-capacity operation with increasing instability
• **V/C ≥ 0.75**: Congested or unstable flow conditions

As V/C approaches or exceeds critical thresholds, traffic flow becomes increasingly unstable, with higher variability in speed and spacing, conditions that are often associated with elevated crash probability (Golob et al., 2004; Lord & Mannering, 2010).

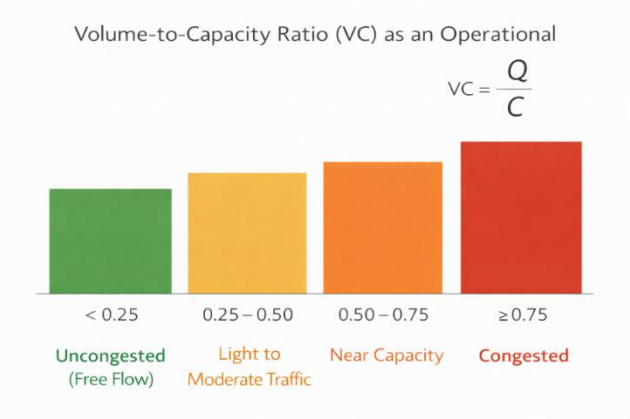


Figure 2-7 volume to capavity ratio (vc)

2.3.3 Temporal Variability of Traffic Demand

Traffic demand varies significantly throughout the day, between weekdays and weekends, and across seasons. Consequently, using average daily traffic (ADT or TGM) alone may mask critical short-term traffic conditions that are directly related to crash occurrence. Several studies have shown that aggregated daily indicators are insufficient to capture dynamic traffic states influencing crash probability (Lee et al., 2003; Abdel-Aty & Pande, 2005).

To address this limitation, traffic demand can be distributed into hourly flow profiles, allowing estimation of traffic flow at the exact time a crash occurs. Hourly flow estimation improves the temporal resolution of crash analysis and supports the calculation of time-specific V/C ratios. The use of temporally disaggregated traffic variables has been widely applied in real-time crash prediction and short-term safety modeling (Oh et al., 2005; Golob et al., 2004).

2.3.4 VC-Based Traffic Classification and Safety Implications

Several studies have demonstrated that crash likelihood and injury severity vary significantly across different V/C levels and operating conditions. Crashes occurring under near-capacity conditions are often associated with increased interaction rates and speed variability, while highly congested conditions may reduce crash severity due to lower operating speeds (Golob et al., 2004; Lord & Mannering, 2010).

By categorizing traffic conditions into discrete V/C classes, it becomes possible to:
- Compare crash behavior across consistent traffic states
- Reduce noise caused by continuous flow variability
- Integrate traffic conditions into probabilistic crash models

The relationship between vehicle impact speed and injury severity has been extensively investigated in traffic safety research. Empirical studies have shown that injury probability follows a nonlinear, sigmoidal pattern as impact speed increases (Elvik, 2009; Rosén & Sander, 2009; Nilsson, 2004).

Relationship Between Impact Speed and Injury Probability

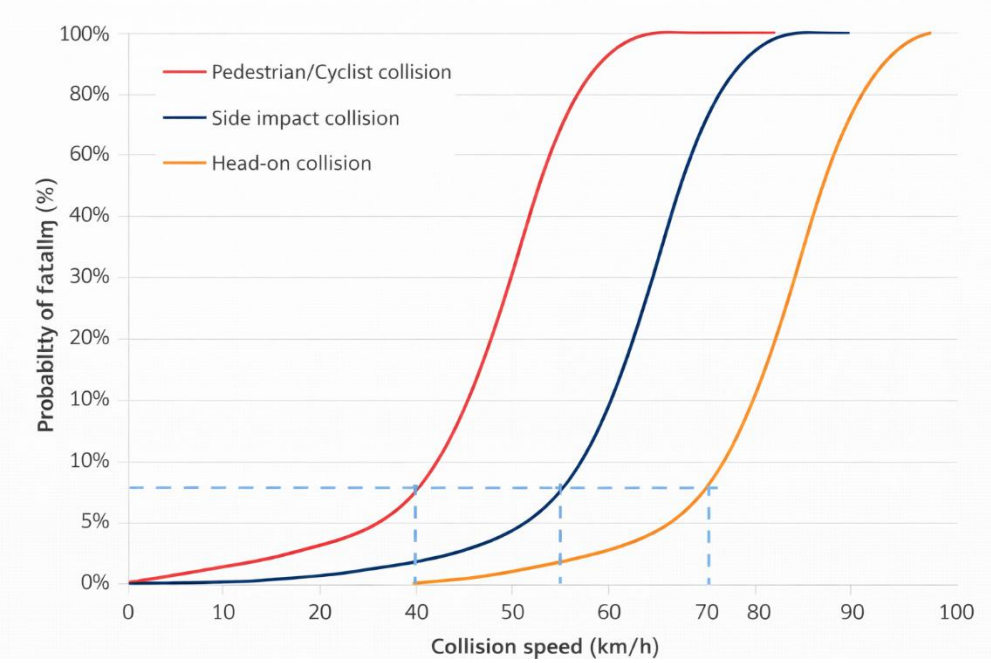


Figure 2-8 : relationship between vehicle impact speed and the probability of severe injury for different crash types

Figure 2-8 illustrates the relationship between vehicle impact speed and the probability of severe injury for different crash types, namely pedestrian/cyclist collisions, side-impact collisions, and head-on collisions. The curves follow a sigmoidal (S-shaped) pattern, highlighting the nonlinear escalation of injury risk as impact speed increases (Rosén & Sander, 2009).

At low impact speeds (below approximately 30–40 km/h), the probability of severe injury remains relatively low across all crash types. In this regime, kinetic energy levels are limited, and vehicle safety systems combined with human tolerance thresholds are generally sufficient to prevent life-threatening injuries (Nilsson, 2004).

As impact speed increases beyond this threshold, injury probability rises sharply. Pedestrian and cyclist collisions exhibit the earliest and steepest increase, reflecting the vulnerability of unprotected road users (Rosén & Sander, 2009). Side-impact collisions display a pronounced increase due to limited lateral structural protection.

Head-on collisions show greater tolerance at moderate speeds due to vehicle frontal energy-absorbing design; however, at high speeds, injury probability approaches unity for all crash types, indicating that safety systems become insufficient to counteract extreme kinetic energy levels (Elvik, 2009).

Overall, the figure emphasizes that crash severity is not linearly related to speed. Even small reductions in impact speed can produce disproportionately large safety benefits. This nonlinear relationship provides strong conceptual justification for incorporating traffic operating conditions—such as the V/C ratio and congestion state—into crash risk modeling frameworks, since traffic flow directly influences prevailing speeds and, consequently, injury outcomes.

2.4 Effect of Road Geometry on Crash Risk

Road geometry plays a critical role in determining crash likelihood, particularly when interacting with traffic volume and environmental conditions. Numerous empirical studies have confirmed that geometric design characteristics significantly influence crash frequency and severity (Hauer, 1997; Elvik, 2009; Lord & Mannering, 2010).

Key geometric factors include:

• **Curvature and Road Alignment:**
Horizontal curves, irregular alignment, limited sight distance, and complex geometric transitions are consistently associated with elevated crash risk. Sharp curves and reduced visibility increase driver workload and error probability, especially under high-speed conditions (AASHTO, 2018; Elvik, 2009). Crash modification factor (CMF) studies have demonstrated higher crash frequencies on curved segments compared to tangent sections.

• **Lane Width and Shoulder Presence:**
Wider lanes and paved shoulders are generally associated with lower crash rates, while narrow lanes increase the likelihood of sideswipe and head-on collisions, particularly at higher traffic volumes. Empirical studies indicate that reductions in lane width correlate with increased crash frequency, especially on arterial and rural highways (HSM, 2010; Lord & Mannering, 2010).

• **Intersection Design and Traffic Flow:**
Intersection type and layout significantly influence crash occurrence. Signalized, unsignalized, and roundabout intersections exhibit distinct safety performance profiles. Locations where high-volume roads intersect with lower-volume roads, or where complex merge and weaving movements occur, tend to experience elevated crash risk (Persaud et al., 2001; Washington et al., 2010).

These geometric characteristics justify the explicit inclusion of segment length, functional classification, and geometric attributes in crash prediction models. When combined with traffic saturation indicators such as the V/C ratio, geometric variables enhance the explanatory power and stability of probabilistic crash models.

2.5 Probabilistic Modeling Approaches in Traffic Safety

Several probabilistic models have been developed to estimate crash risk based on traffic flow and roadway geometry. These models aim to predict crash likelihood by analyzing the interaction between traffic exposure, operational conditions, and geometric characteristics (Lord & Mannering, 2010; Washington et al., 2010). The most widely applied modeling approaches include count-data models, simulation-based techniques, and probabilistic graphical models.

• **Poisson and Negative Binomial Models**

Poisson regression models are commonly used to analyze crash frequency data under the assumption that crashes are rare and independent events following a Poisson distribution. However, crash data often exhibit overdispersion, meaning that the variance exceeds the mean. In such cases, the Negative Binomial model is generally preferred, as it introduces an additional dispersion parameter to account for unobserved heterogeneity (Lord et al., 2005; Mannering & Bhat, 2014). These models form the foundation of modern crash frequency analysis.

• **Monte Carlo Simulations**

Monte Carlo simulation methods involve repeated random sampling of input variables to estimate the distribution of possible crash outcomes. These techniques are particularly effective when dealing with uncertainty and variability inherent in traffic flow, capacity, and exposure measures. By performing multiple simulations, it becomes possible to estimate not only expected crash frequency but also variability and confidence intervals across roadway segments (Rubinstein & Kroese, 2017; Brilon et al., 2005).

Monte Carlo simulation is especially suitable when the objective extends beyond point estimation to include stability assessment and probabilistic variability in predicted crash frequencies.

• **Bayesian Networks**

Bayesian networks have been increasingly applied in traffic safety research to model probabilistic relationships among multiple risk factors and crash occurrence. These graphical models are capable of representing conditional dependencies between geometric, operational, and environmental variables. Bayesian approaches are particularly valuable when dealing with complex, multidimensional datasets and when incorporating expert knowledge into predictive modeling (Sun & Sun, 2006; De Oña et al., 2011).

2.6 Crash Prediction Models Based on Traffic Flow and Road Segmentation

Crash prediction models frequently rely on traffic flow indicators and roadway segmentation to estimate crash risk across a network of road segments. By dividing roadways into homogeneous segments, researchers can develop models that account for spatial variations in traffic conditions, geometric design, and environmental factors. Segment-based modeling has been widely adopted in highway safety research as it improves the association between crashes and roadway characteristics (Hauer, 1997; Lord & Mannering, 2010; Persaud & Lyon, 2007).

• Segment-Level Crash Prediction Models

Segment-level crash prediction models divide a roadway into sections based on geometric features such as curvature, intersection density, lane configuration, and functional classification. Crash risk is estimated for each segment using statistical or probabilistic models, and segment-level predictions may subsequently be aggregated to evaluate overall corridor safety performance. This approach allows for localized identification of high-risk segments and supports targeted safety interventions (Washington et al., 2010; Elvik, 2009).

• Volume-to-Capacity (V/C) Based Models

Volume-to-capacity (V/C) ratios are commonly used to represent congestion intensity and operational stability. In V/C-based crash models, crash risk is estimated by combining roadway geometric characteristics with traffic demand relative to capacity, thereby capturing the influence of traffic saturation on safety outcomes (Transportation Research Board, 2016; Golob et al., 2004).

By integrating operational conditions with structural roadway characteristics, V/C-based models provide an operationally meaningful link between traffic demand, roadway capacity, and safety performance. Such models are particularly suitable for arterial corridors where congestion patterns vary temporally and spatially, as in the case of the Rivoltana (SP14).

2.7 Methodological Gaps and Contributions

Despite the progress made in crash prediction modeling, several gaps remain. Traditional methods often fail to incorporate real-time traffic conditions, heavy vehicle influence, and weather factors in their predictions. Additionally, many models focus on aggregate data, overlooking the spatial and temporal distribution of crashes along a road.

This thesis aims to bridge these gaps by developing a comprehensive probabilistic crash prediction framework for the Rivoltana Road (SP 14). The framework incorporates both segment-level geometry and traffic saturation factors and applies Monte Carlo simulation to estimate crash likelihood across various segments, improving the accuracy of predictions and allowing for targeted safety interventions.

2.8 Summary and Link to Methodology

The concepts reviewed in this section demonstrate that traffic safety is strongly influenced by capacity constraints, congestion dynamics, and temporal variability in traffic demand. The VC ratio emerges as a robust and interpretable indicator that captures these effects in a unified manner.

These findings directly motivate the methodological choices adopted in this research, particularly:

1. The estimation of hourly traffic flows,
2. The classification of traffic conditions using VC levels,
3. The integration of VC classes into probabilistic crash prediction models.

These elements are formally introduced and implemented in the following chapters.

3 Road And Data Preparation

3.1 Study Area : SP.14 (Rivoltana)

The SP14 Rivoltana is a provincial road located in the Lombardy region, serving as a key transportation corridor between the eastern metropolitan area of Milan and the town of Rivolta d'Adda in the province of Cremona. The road extends for approximately 18 km and plays an important role in regional mobility by connecting urban, suburban, and semi-rural areas within the southeastern part of the region.

Along its alignment, the Rivoltana crosses several municipalities, including Segrate, Pioltello, and Melzo, and provides a strategic link between the metropolitan area of Milan and the southern part of Lombardy. Additionally, the road's proximity to Linate Airport adds to its significance, as it supports both local and regional traffic demand related to air travel.

As a result, the road accommodates a combination of commuter traffic, local access movements, and interprovincial travel demand.

The multifunctional role of the SP14 Rivoltana results in consistently high traffic volumes and a heterogeneous road environment. The presence of varying geometric configurations, intersections, and access points leads to complex traffic conditions that may influence safety performance. These characteristics make the Rivoltana Road a suitable and representative case study for segment-level safety analysis and probabilistic crash prediction.

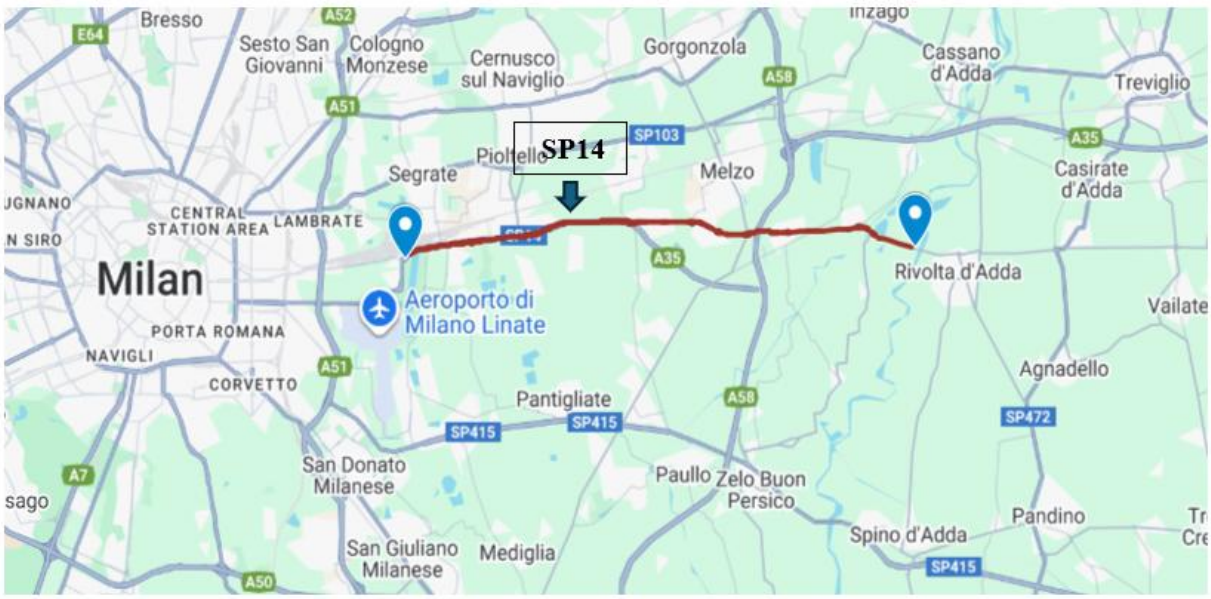


Figure 3-1: sp 14 -Rivoltana Strada visualization on map

Source: Google Maps, https://www.google.com/maps/d/edit?mid=1oBlsZyuzThpS7vrml58fii9TEySzPuc&usp=sharing

3.1.1 : Overall methodological framework for Crash Prediction

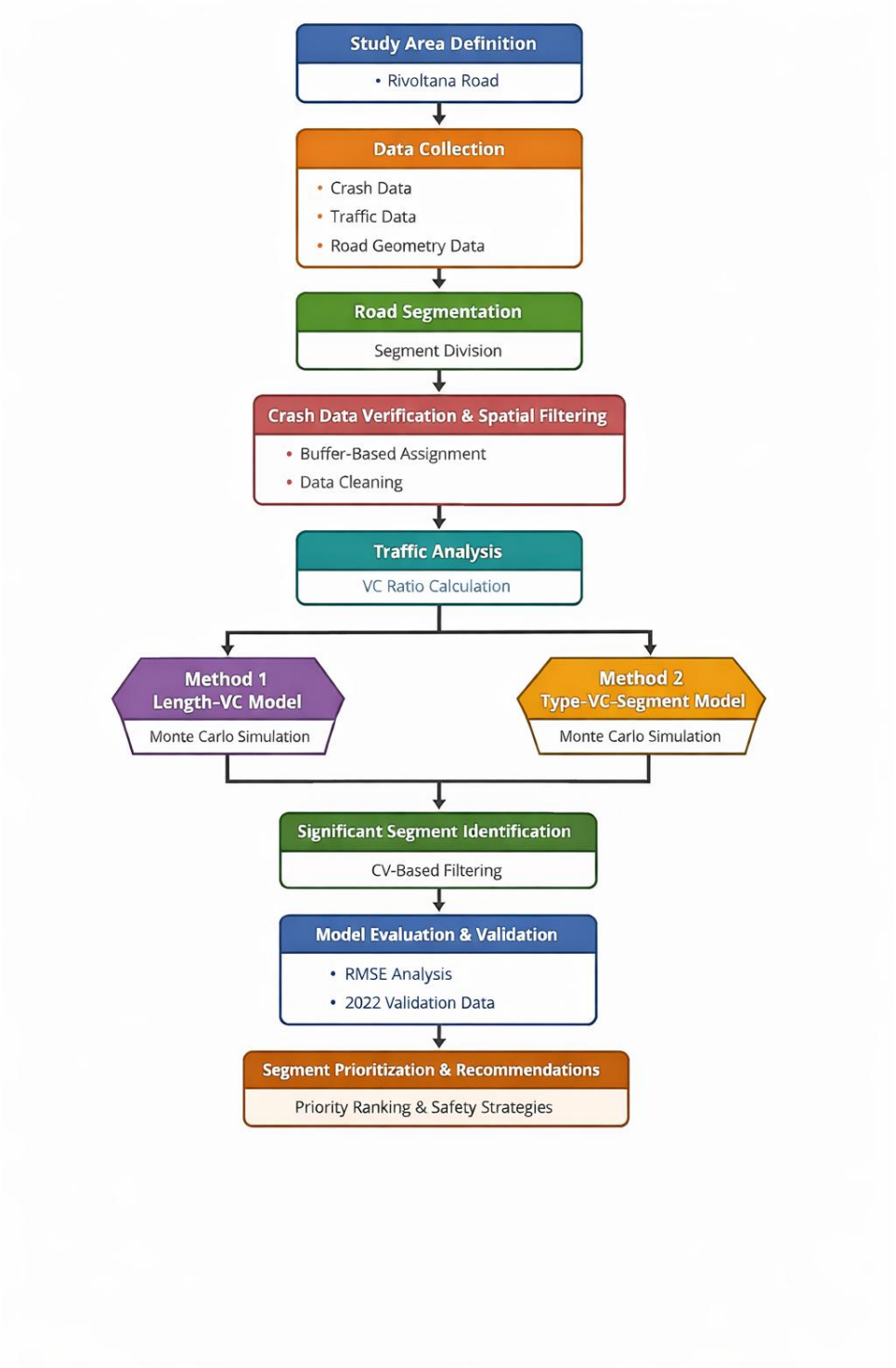


Figure 3-2: Overall methodological framework for Crash Prediction

Traffic Characteristics and Analysis

Traffic conditions along the Rivoltana Road were analyzed using traffic flow data derived from the official traffic monitoring system of Regione Lombardia. The dataset, covering the period from 2018 to 2024, includes Average Daily Traffic (ADT) values and hourly traffic volumes obtained from permanent automatic count stations located along the SP14 corridor.

The traffic dataset was provided for academic research purposes by the course instructor and originates from the Regione Lombardia traffic monitoring database. These data were used to estimate traffic demand and identify peak and off-peak operating conditions, enabling the reconstruction of hourly flow profiles for subsequent safety analysis.

Traffic Volume and Temporal Distribution

Traffic volumes on the Rivoltana Road exhibit pronounced temporal variation. Peak traffic conditions are typically observed during morning and evening commuting hours, reflecting the road's role as a commuter corridor. Off-peak periods are characterized by lower and more stable traffic flows.

The temporal distribution of traffic was analyzed to capture daily and weekly variability. This analysis is essential for understanding the exposure of each segment to traffic demand and for estimating volume-to-capacity (VC) ratios used in crash prediction modeling.

3.2 Data Preparation and Segmentation Methodology

This section describes the data sources used in the study and the procedures adopted for data preparation and road segmentation. Proper preprocessing and segmentation are essential to ensure consistency between traffic, geometric, and crash data and to support reliable crash prediction modeling.

Data Sources

The analysis is based on three primary data sources: roadway geometry data, traffic flow data, and historical crash records.

Road geometry information was obtained from the official databases of Regione Lombardia and the provincial road authority responsible for the SP14 Rivoltana corridor. The dataset includes detailed information on road alignment, lane configuration, intersection layout, and access points along the study corridor.

Traffic flow data were derived from the official traffic monitoring system of Regione Lombardia and cover the period from 2018 to 2024. The dataset includes Average Daily Traffic (ADT) values and hourly traffic volumes recorded at permanent automatic counting stations located along the SP14 corridor. Where continuous measurements were not available for specific segments, traffic volumes were spatially interpolated using standard traffic engineering procedures to ensure consistent coverage across the entire roadway.

Crash data were obtained from official accident records maintained by the competent regional authority and cover the same multi-year observation period (2018–2024). The dataset includes the geographic location of each crash as well as descriptive attributes such as date, time, and crash type. Only crashes occurring on the mainline of the SP14 Rivoltana were considered in order to ensure consistency with the adopted segmentation framework.

Data Cleaning and Preprocessing

Prior to analysis, all datasets underwent a systematic preprocessing phase. Records with missing or inconsistent spatial references were excluded or corrected when possible. Traffic and crash data were checked for temporal consistency to ensure that the observation periods aligned across datasets.

Crash locations were georeferenced and mapped onto the Rivoltana Road alignment. Each crash was then assigned to a specific road segment based on its spatial position. Traffic volumes were normalized to a common reference period to allow for comparison between segments and operating conditions.

3.2.1 Road Segmentation

To capture spatial variability in traffic operations and safety performance, the SP14 Rivoltana was divided into homogeneous roadway segments. The segmentation was performed manually using high-resolution satellite imagery (Google Earth Pro), through detailed visual inspection of the corridor. The objective was to identify locations where significant changes in geometric or operational characteristics occurred, ensuring that each segment maintained internal homogeneity.

Segmentation Criteria

Segments were defined based on roadway attributes known to influence traffic behavior and crash risk. The adopted criteria included:

- Horizontal alignment (distinguishing between straight and curved sections)
- Presence and type of intersections
- Lane configuration and cross-sectional changes

• Access point density
• Traffic control characteristics

Segment boundaries were established at locations where one or more of these attributes changed significantly.

Segmentation Implementation

The segmentation process consisted of systematically reviewing the entire corridor and identifying transition points between different roadway configurations. Boundaries were assigned at changes in the number of lanes, intersection layout, introduction of auxiliary lanes (e.g., acceleration or deceleration lanes), and variations in horizontal alignment.

Each resulting segment was classified according to its dominant geometric and operational characteristics. This structured classification ensures consistency between geometric, traffic, and crash datasets and provides a reliable spatial framework for the probabilistic crash prediction models developed in subsequent chapters.

Illustration of the Segmentation Process for One-Way and Two-Way Road Sections

The following figures illustrate the segmentation approach adopted for the SP14 Rivoltana, highlighting the distinction between two-way divided carriageway sections and one-way traffic sections at intersections.

In two-way divided sections , the roadway consists of two physically separated carriageways, each serving a single traffic direction. In these cases, segmentation was performed independently for each direction of travel, treating each carriageway as a separate analytical entity. This approach allows directional traffic volumes, crash occurrences, and capacity-related parameters to be assigned accurately to each segment. Segment boundaries were defined based on geometric continuity and changes in alignment, and each segment was associated with its own spatial coordinates and traffic attributes.

In contrast, one-way traffic sections, typically located at intersections or complex junctions , were segmented considering the operational characteristics of the traffic flow within the intersection influence area. These segments represent areas where traffic movements are governed by turning maneuvers, lane merging or diverging, and signal or priority control. In such cases, segmentation boundaries were defined to capture changes in traffic behavior and conflict points rather than purely geometric continuity.

The segmentation strategy ensures that both linear roadway segments and intersection-influenced sections are represented consistently within the dataset. By distinguishing between one-way and two-way configurations, the adopted approach improves the accuracy of traffic assignment, crash localization, and capacity estimation, thereby providing a reliable basis for the probabilistic crash prediction models developed in subsequent chapters.

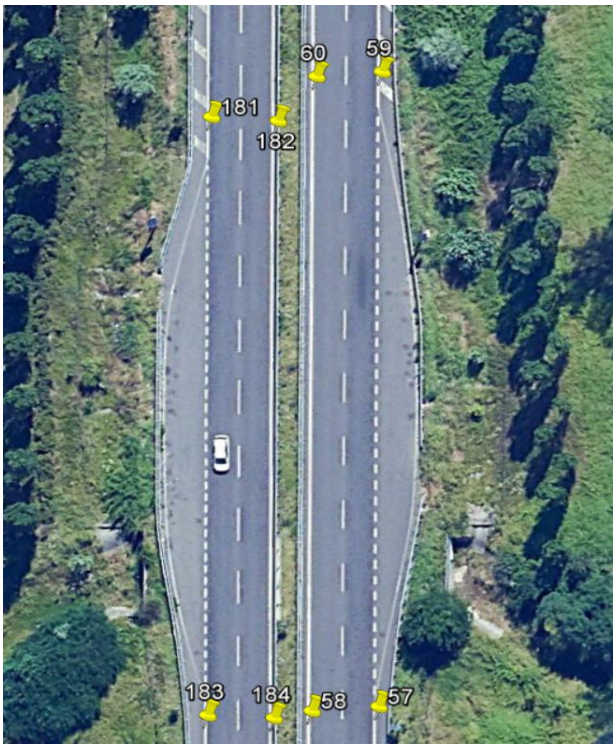


Figure 3-3 : oneway road segmentation

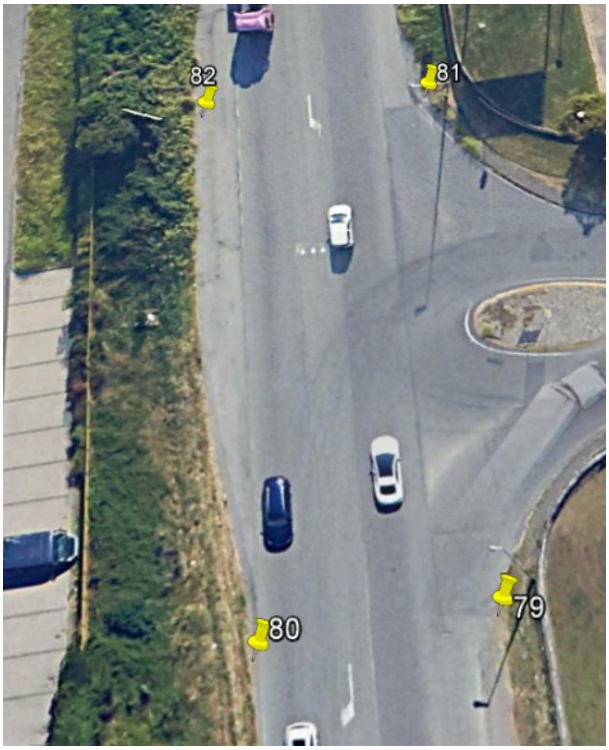


Figure 3-4 : twoways road segmentation

Segment Length and Homogeneity

Segment length was determined as a balance between spatial resolution and statistical robustness. Very short segments may lead to unstable crash rate estimates, while excessively long segments may mask local safety issues. Therefore, segment lengths were selected to preserve homogeneity while ensuring sufficient exposure for meaningful analysis.

Each segment was assigned a unique identifier and associated with corresponding traffic, geometric, and crash attributes. Segment length was explicitly included as a modeling variable, as it directly influences exposure and crash probability.

3.3 Creation of the 'Rivoltana40.mat' Dataset

Following data preprocessing and road segmentation, all segment-level information was consolidated into a MATLAB-compatible dataset named Rivoltana40.mat. This dataset represents the main input for the subsequent analytical stages of the thesis, including the implementation of crash prediction models and the comparative evaluation of methodological performance.

Each road segment was described through a comprehensive set of attributes, resulting in a total of 40 variables per segment. These attributes include spatial, geometric, traffic, and safety-related information. Segment geometry was defined using geographic coordinates, where four points representing the segment boundaries (longitude and latitude) were extracted from Google Earth following the segmentation process. In addition, the dataset includes the identifiers of the points forming each segment, the number of recorded crashes within the segment boundaries, travel direction, number of lanes, posted speed limit, segment type, and all parameters required for capacity estimation.

Adjusted Capacity Estimation

The estimation of adjusted roadway capacity follows the formulation introduced in Equation (2.2), where actual capacity is expressed as a multiplicative adjustment of base capacity through geometric and operational correction factors.

$$C_{\text{actual}} = C_{\text{base}} \times f_w \times f_{HV} \times f_g \times f_M \times f_D$$

where:

- C_{actual}: adjusted (actual) capacity of the roadway segment (veh/h/lane)
- C_{base}: base capacity under ideal conditions assigned to each segment type (veh/h/lane)
- f_w: lane width and lateral clearance adjustment factor
- f_{HV}: heavy vehicle adjustment factor
- f_g: grade adjustment factor accounting for uphill segments
- f_M: merging influence factor, accounting for the effect of on-ramps
- f_D: diverging influence factor, accounting for the effect of off-ramps

Type	segment	base capacity
1	Rectilinear 2 ways	1900
2	Rectilinear 1 way	2000
3	Curve 2 way	1700
4	Curve 1 way	1800
5	Crossroad (give way)	1200
6	Signalized crossroad	760
7	Roundabout	1400
8	Driveway	400
9	Acc/deceleration lane	1500

Table 3-1 : Segment classification used for the segmentation process with their base capacity

Capacity Calculation Procedure

The capacity estimation process was conducted in two main stages. First, an ideal base capacity was assigned to each segment according to its geometric characteristics and segment type. This value represents the capacity under idealized conditions and serves as the reference capacity for subsequent adjustments.

In the second stage, the heavy vehicle adjustment factor was computed using traffic flow data and heavy vehicle percentages available for the study period (2018–2024). Heavy vehicle percentages were not uniformly available along the entire corridor; therefore, a segment-specific estimation approach was adopted.

For each segment, the percentage of heavy vehicles was estimated through a linear interpolation between the two closest traffic count sections for which heavy vehicle data were available. Both inflow and outflow directions were considered in the interpolation process to account for directional traffic variations. This approach allowed the assignment of a unique heavy vehicle percentage to each segment, depending on its spatial location along the corridor.

Heavy Vehicle Adjustment Factor

Once the heavy vehicle percentage was determined for each segment and year, the heavy vehicle adjustment factor was calculated using the following expression:

$$HV = 1/(1 + P_H \cdot (E_H - 1))$$

where:

- P_H: proportion of heavy vehicles expressed as a decimal
- E_H: passenger car equivalent (PCE) for heavy vehicles

In this study, a PCE value of **2** was adopted for heavy vehicles, consistent with the characteristics of the Rivoltana Road segments.

segment	cb/segment type	Nb of Crashes	Directions	N of lanes	LENGTH	lanes width(m)	Speed (Km/h)	slope fg	type	width fw	Base c. Cb/type	fM	fD	cb*fw*fg*fM*fD I.Capacity
1	1500	4	ONEWAY	1	258.80	3.6	40	1	9	1	1500	0.85	0.95	1211.25
2	1800	1	ONEWAY	2	219.61	3.6	70	1	4	1	3600	1	1	3600
3	2000	1	ONEWAY	2	121.43	3.6	70	1	2	1	4000	0.85	1	3400
4	2000	2	ONEWAY	2	279.48	3.6	70	1	2	1	4000	1	0.95	3800
5	2000	1	ONEWAY	2	190.15	3.6	70	1	2	1	4000	0.85	1	3400
6	1500	2	ONEWAY	1	195.61	3.6	70	1	9	1	1500	1	1	1500
7	2000	1	ONEWAY	2	83.86	3.6	70	1	2	1	4000	1	1	4000
8	2000	1	ONEWAY	2	199.18	3.6	70	1	2	1	4000	1	0.95	3800
9	1500	4	ONEWAY	1	513.33	3.6	40	1	9	1	1500	0.85	0.95	1211.25
10	2000	0	ONEWAY	2	464.29	3.6	70	1	2	1	4000	1	1	4000
11	1500	2	ONEWAY	2	339.93	3.6	70	1	9	1	3000	0.85	0.95	2422.5
12	1500	1	ONEWAY	1	616.67	3.6	40	1	9	1	1500	0.85	0.95	1211.25
13	2000	0	ONEWAY	2	582.24	3.6	70	1	2	1	4000	1	1	4000
14	1500	0	ONEWAY	1	352.67	3.6	70	1	9	1	1500	0.85	1	1275
15	2000	0	ONEWAY	2	405.48	3.6	90	1	2	1	4000	1	1	4000
16	2000	0	ONEWAY	2	301.69	3.6	110	1	2	1	4000	1	1	4000
17	2000	0	ONEWAY	2	340.83	3.6	110	1	2	1	4000	1	1	4000
18	2000	0	ONEWAY	2	357.89	3.6	110	1	2	1	4000	1	0.95	3800
19	1500	7	ONEWAY	1	819.24	3.6	40	1	9	1	1500	1	0.95	1425
20	2000	0	ONEWAY	2	872.67	3.6	110	1	2	1	4000	1	1	4000
21	1800	0	ONEWAY	2	400.44	3.6	110	1	4	1	3600	1	1	3600
22	2000	0	ONEWAY	2	884.34	3.6	110	1	2	1	4000	1	1	4000
23	2000	0	ONEWAY	2	225.33	3.6	110	1	2	1	4000	0.85	1	3400

Figure 3-5 : Structure of the Excel dataset containing segment geometry, functional characteristics, and base capacity values used for the creation of the Rivoltana40.mat file.

2020					2021					2022				
%of HV	fHV	Avg. daily flow		capacity	%of HV	fHV	Avg. daily flow		capacity	%of HV	fHV	Avg. daily flow		capacity
		Weekdays	Weekends				Weekdays	Weekends				Weekdays	Weekends	
4.5833475	0.956175169	9403.3902	7052.5427	1158.2	4.814072	0.95407	10284.61538	8306.804734	1155.6	4.99947	0.952386	9907.692	8002.367	1153.6
4.5833475	0.956175169	22400.0000	18092.3077	3442.2	4.814072	0.95407	20569.23077	16613.60947	3434.7	4.99947	0.952386	19815.38	16004.73	3428.6
4.5833475	0.956175169	22400.0000	18092.3077	3251.0	4.814072	0.95407	20569.23077	16613.60947	3243.8	4.99947	0.952386	19815.38	16004.73	3238.1
4.5833475	0.956175169	22400.0000	18092.3077	3633.5	4.814072	0.95407	20569.23077	16613.60947	3625.5	4.99947	0.952386	19815.38	16004.73	3619.1
4.5833475	0.956175169	18806.7804	15190.0919	3251.0	4.814072	0.95407	17558.28204	14181.68934	3243.8	4.99947	0.952386	16891.49	13643.13	3238.1
4.5833475	0.956175169	9403.3902	7595.0459	1434.3	4.814072	0.95407	8779.141019	7090.844669	1431.1	4.99947	0.952386	8445.746	6821.564	1428.6
4.5833475	0.956175169	18806.7804	15190.0919	3824.7	4.814072	0.95407	17558.28204	14181.68934	3816.3	4.99947	0.952386	16891.49	13643.13	3809.5
4.5833475	0.956175169	18806.7804	15190.0919	3633.5	4.814072	0.95407	17558.28204	14181.68934	3625.5	4.99947	0.952386	16891.49	13643.13	3619.1
4.5833475	0.956175169	9403.3902	7595.0459	1158.2	4.814072	0.95407	8779.141019	7090.844669	1155.6	4.99947	0.952386	8445.746	6821.564	1153.6
4.5833475	0.956175169	18806.7804	15190.0919	3824.7	4.814072	0.95407	17558.28204	14181.68934	3816.3	4.99947	0.952386	16891.49	13643.13	3809.5
4.5833475	0.956175169	18806.7804	15190.0919	2316.3	4.814072	0.95407	17558.28204	14181.68934	2311.2	4.99947	0.952386	16891.49	13643.13	2307.2
4.5833475	0.956175169	9403.3902	7595.0459	1158.2	4.814072	0.95407	8779.141019	7090.844669	1155.6	4.99947	0.952386	8445.746	6821.564	1153.6
4.5833475	0.956175169	18806.7804	15190.0919	3824.7	4.814072	0.95407	17558.28204	14181.68934	3816.3	4.99947	0.952386	16891.49	13643.13	3809.5
4.5833475	0.956175169	9403.3902	7595.0459	1219.1	4.814072	0.95407	8779.141019	7090.844669	1216.4	4.99947	0.952386	8445.746	6821.564	1214.3
4.5833475	0.956175169	18806.7804	15190.0919	3824.7	4.814072	0.95407	17558.28204	14181.68934	3816.3	4.99947	0.952386	16891.49	13643.13	3809.5
4.5833475	0.956175169	18806.7804	15190.0919	3824.7	4.814072	0.95407	17558.28204	14181.68934	3816.3	4.99947	0.952386	16891.49	13643.13	3809.5
4.5833475	0.956175169	18806.7804	15190.0919	3824.7	4.814072	0.95407	17558.28204	14181.68934	3816.3	4.99947	0.952386	16891.49	13643.13	3809.5

Figure 3-6 : Structure of the Excel dataset containing segment-level traffic data (2020–2022) used for the creation of the Rivoltana40.mat file.

Explanation of Traffic and Geometric Input Tables

Figure 3-5 : Segment Geometry, Functional Characteristics, and Base Capacity

This table summarizes the geometric and functional attributes of each road segment, which form the structural basis for capacity estimation and segmentation. The variables include:

- Segment length, number of lanes, lane width, and roadway direction (one-way or two-way);
- Design speed and longitudinal slope;
- Segment type classification, used consistently across the modeling framework;
- Base capacity, derived from standard values and subsequently adjusted;
- Correction factors accounting for heavy vehicles, lane width, and roadway configuration.

These geometric descriptors are used to compute a segment-specific capacity, which is later combined with traffic demand to obtain the VC ratio. The explicit inclusion of geometric parameters ensures that differences in roadway design such as acceleration/deceleration lanes, lane drops, or transitions between cross-sections—are reflected in the safety analysis.

By structuring the road into homogeneous segments, this table enables segment-level crash prediction rather than relying on aggregated corridor-level averages. This segmentation is critical for identifying localized safety issues that may be masked when using coarser spatial units.

Figure 3-6 : Segment-Level Traffic Data by Year (2020-2022)

This table reports the temporal evolution of traffic conditions for each road segment along the Rivoltana Road over the period 2020–2022.

For every segment and year, the dataset includes:

- Average Daily Traffic (ADT / TGM), disaggregated into weekday and weekend values;
- Heavy vehicle percentage (%HV) and the corresponding heavy-vehicle adjustment factor (fHV);
- Segment capacity, computed after accounting for geometric characteristics and traffic composition.

The separation between weekday and weekend traffic is essential, as demand patterns differ substantially between working days and non-working days. This distinction allows the traffic conditions at the time of each crash to be represented more realistically when computing hourly flows and Volume-to-Capacity (VC) ratios.

This table constitutes the primary traffic input for the VC computation and directly supports the hourly flow estimation used in the probabilistic crash models

4 Traffic Analysis

4.1 Introduction

Traffic conditions play a fundamental role in shaping crash occurrence and severity, as they directly influence vehicle interactions, speed variability, and driver behavior. For this reason, a detailed and time-resolved analysis of traffic flow is a necessary prerequisite for any robust crash prediction framework. This chapter focuses on the characterization of traffic demand and capacity along the Rivoltana Road (SP 14), providing the operational basis for the Volume-to-Capacity (VC) analysis adopted in the subsequent modeling phase.

Unlike traditional safety studies that rely on static or annual average traffic values, this work explicitly accounts for temporal variability in traffic conditions. Traffic volumes are analyzed by year, by day type (weekday and weekend), and by hour of the day using representative flow profiles. This approach allows traffic exposure to be aligned with the exact timing of crash events, thereby reducing aggregation bias and improving the interpretability of the resulting safety indicators.

The chapter first describes the available traffic datasets and the procedures used to estimate average daily traffic and segment capacities over multiple years. Particular attention is given to the influence of heavy vehicles and roadway geometry on effective capacity. Subsequently, hourly traffic flows are derived from daily volumes using standardized diurnal profiles, enabling the computation of VC ratios that reflect realistic operating conditions at the time of each crash.

By transforming raw traffic data into segment-specific, time-dependent VC indicators, this chapter establishes the quantitative link between traffic conditions and crash occurrence. The outputs of this analysis serve as direct inputs to the probabilistic crash prediction models developed in Chapter 5, ensuring methodological consistency between traffic analysis and safety modeling.

4.2 Handling of Missing Data

During the data preparation phase, instances of missing or incomplete data were identified across the traffic and crash datasets. Missing data mainly concerned traffic flow measurements and heavy vehicle percentages, particularly in road sections where permanent counting stations were not available or where records were incomplete for specific years.

To ensure consistency and avoid bias in the analysis, a structured approach was adopted for handling missing data. Segments with missing spatial information or unclear geometric definition were excluded from the analysis, as accurate spatial representation is essential for reliable crash assignment and capacity estimation. In contrast, missing traffic-related variables were treated using estimation procedures based on nearby available data.

For traffic volumes and heavy vehicle percentages, missing values were estimated using spatial interpolation between the closest upstream and downstream traffic counting sections. This approach assumes a gradual variation of traffic characteristics along the corridor and allows the assignment of representative values to intermediate segments. When both inflow and outflow measurements were available, they were jointly considered to improve estimation accuracy.

In cases where data were missing for isolated time periods within the study horizon, values were reconstructed using averages from adjacent years with similar traffic conditions. This method preserves temporal consistency while minimizing distortion of long-term traffic trends.

The adopted strategy ensures that missing data do not compromise the continuity of the dataset while maintaining transparency and methodological rigor. All estimated values were used consistently across segments and years and were incorporated into the final Rivoltana40.mat dataset used for subsequent modeling and analysis.

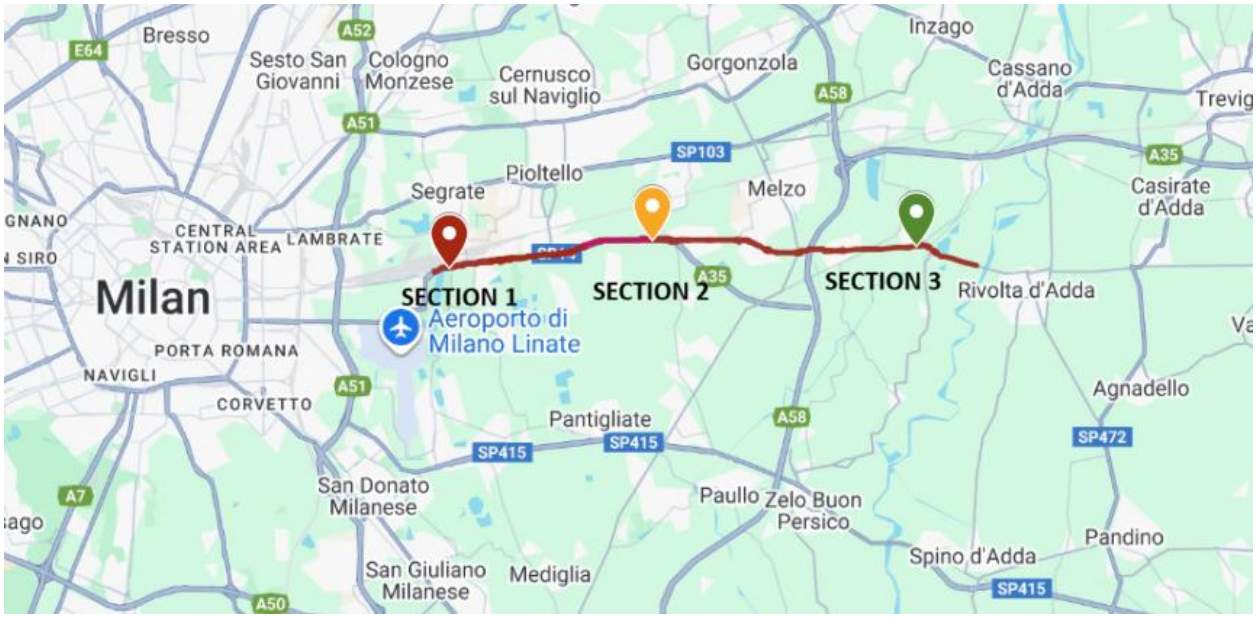


Figure 4-1: section where data was collected along the studied road

4.3 Estimation Process

Traffic flow data were not continuously available for all road sections and all years of analysis. In particular, some sections presented partial monthly records, with missing values for specific months or years. To ensure temporal continuity and consistency across the dataset, an estimation procedure was applied based on the observed seasonal behavior of traffic flows along the corridor.

The estimation approach relies on the assumption that traffic patterns between consecutive months follow similar seasonal trends across different years for the same section and direction of travel. This method allows missing values to be reconstructed using ratios derived from years with complete datasets, while preserving the characteristic monthly variability of traffic demand.

4.3.1 Section 1 (Milano) – Estimation of Missing Traffic Data

4.3.1.1 Stage A: Completing Missing Months Within the Same Year (2023 and 2024)

1. Identify the available data window:
   For the analyzed section, monthly traffic data were available only for the months May–November in the years 2023 and 2024. The months January–April and December were missing.
2. Select a complete reference section:
   Section 2 – Milano direction was selected as the reference section because it provides complete monthly traffic data and shares similar directional demand characteristics with the analyzed section.
3. Compute the mid-year baseline for the analyzed section:
   For each year (2023 and 2024), the average traffic value of the available period April–November was computed. This value represents the mid-year traffic level for that year and section.
4. Extract seasonal relationships from the reference section:
   Using Section 2 (Milano direction), the typical relationship between

   - January–April and May–November, and
   - December and May–November,

   was computed for the same year(s). These relationships describe how traffic in the missing months compares to the available mid-year period.

5. Estimate the missing months for each year:
   The seasonal relationships obtained from the reference section were applied to the mid-year baseline (April–November average) of the analyzed section to estimate:

   - traffic values for January–March, and
   - traffic value for December,

   separately for 2023 and 2024.

6. Rebuild the full monthly series for 2023 and 2024:
   The estimated months (Jan–March and Dec) were combined with the observed months (Apr–Nov) to generate complete monthly datasets for 2023 and 2024.

4.3.1.2 Stage B: Reconstructing Fully Missing Years

7. Define a representative monthly profile from completed years:
   After completing 2023 and 2024, these two years were used to define a representative monthly traffic pattern for the analyzed section (i.e., how traffic is typically distributed across months).
8. Transfer inter-annual traffic variation from the reference section:
   Section 2 (Milano direction) was used to capture how total traffic levels change from year to year along the corridor. These year-to-year changes were used as scaling information to estimate the annual traffic level of the analyzed section in years with missing data.
9. Generate monthly values for the missing years:
   For each missing year, the representative monthly profile was scaled according to the inter-annual variation extracted from the reference section, producing a complete set of monthly traffic values.
10. Integrate final traffic dataset:
    The reconstructed monthly traffic series (completed years + reconstructed years) were included in the final dataset used for capacity estimation, VC classification, and crash prediction modeling.

Tables 4-1 and 4-2 show the available traffic records and the reconstructed monthly values, illustrating the two-stage estimation procedure.

SECTION 1 MILANO . DIR.						
	2019	2020	2021	2022	2023	2024
Jan						
Feb						
Mar						
Apr					16429.9679	17087.2
May					16988.6476	17668.2
Jun					17653.8571	18018.7
Jul					16777.9355	18004.5
Aug					11973.9677	12706.7
Sep					17701.8	18409.9
Oct					13704.9677	14253.2
Nov					11082.04	11525.3
Dec						

Table 4-1: available data for sec 1 milan direction

SECTION 1 MILANO . DIR.						
	2019	2020	2021	2022	2023	2024
Jan	13470.12	11225.1	14592.63	13511.69	14727.74	14471.41
Feb	15333.33	12777.77	16611.1	15380.65	16764.91	14054.87
Mar	14429.94	12024.95	15632.43	14474.47	15777.18	16670.16
Apr	15026.99	12522.49	16279.23	15073.37	16429.97	17087.17
May	15537.96	12948.3	16832.79	15585.92	16988.65	17668.19
Jun	16146.36	13455.3	17491.9	16196.2	17653.86	18018.67
Jul	19597.37	16331.14	17964.25	16947.41	16777.94	18004.52
Aug	13986.12	11655.1	12820.61	12094.92	11973.97	12706.68
Sep	20676.48	17230.4	18953.44	17880.61	17701.8	18409.87
Oct	16008.01	13340.01	14674.01	13843.4	13704.97	14253.17
Nov	12944.31	10786.93	11865.62	11193.98	11082.04	11525.32
Dec	9247.765	7706.471	8477.118	10681.17	13031.03	13563.96
AVERAGE	15200.4	12667	15182.93	14405.31	15217.84	15536.17

Table 4-2 Estimated data for sec 1 milan direction

4.3.2 Section 1 (Rivolta Direction) – Estimation of Missing Traffic Data

For Section 1 Rivolta direction, monthly traffic data were not fully available for all months within the study period. As illustrated in the tables, for the year 2023 traffic data were available only for the months June to December, while data for January to May were missing. For the year 2024, traffic data were available for January and for the period June to December, whereas the months February to May were missing.
To reconstruct complete and consistent monthly traffic datasets, a structured estimation procedure was applied. The process was carried out in two main stages: first, missing months were estimated within the same year, and subsequently, once complete yearly profiles were available, the methodology was extended to years with more extensive data gaps.
For the extraction of seasonal traffic patterns, Section 2 – Rivolta direction was selected as the reference section, as it provides complete monthly traffic data and exhibits similar directional and operational characteristics.

4.3.2.1 Stage A: Completion of Missing Months Within the Same Year (2023–2024)

1. Identification of available and missing data:

- 2023: observed traffic values were available for June–December, while January–May were missing.
- 2024: observed traffic values were available for January and June–December, while February–May were missing.

2. Selection of the reference section:
   Section 2 – Rivolta direction was used as the reference section because it provides complete monthly traffic data and reflects the typical seasonal traffic behavior along the Rivoltana corridor in the same travel direction.
3. Computation of a yearly baseline from available months:
   For each year, the average traffic volume of the available months (June–December) was calculated for Section 1. This value represents the mid-to-late-year traffic level for that year and was used as the baseline for estimating missing months.
4. Derivation of seasonal relationships from the reference section:
   Using Section 2 – Rivolta direction, the relationship between traffic volumes in the missing months and those in the available period was identified for the same year. In particular, the relative differences between:

- early-year traffic (January–May) and
- mid-to-late-year traffic (June–December)

were used to describe seasonal behavior.

5. Estimation of missing months for 2023:
   For 2023, the seasonal relationships derived from the reference section were applied to the June–December 2023 baseline of Section 1 to estimate traffic volumes for January–May 2023.
6. Estimation of missing months for 2024 (January preserved):
   For 2024, the observed traffic value for January was retained without modification. The seasonal relationships from the reference section were then applied to the June–December 2024 baseline to estimate traffic volumes for February–May 2024.
7. Reconstruction of complete yearly datasets:
   The estimated values were combined with the observed data, resulting in complete monthly traffic series for:

- 2023: estimated January–May + observed June–December
- 2024: observed January + estimated February–May + observed June–December

4.3.2.2 Stage B: Reconstruction of Years with Larger Data Gaps

8. Definition of a representative seasonal profile:
   Once the datasets for 2023 and 2024 were completed, these years were used to define a representative monthly traffic profile for Section 1 – Rivolta direction, capturing the typical seasonal distribution of traffic for this section.
9. Adjustment for inter-annual variation:
   Inter-annual changes in traffic demand were inferred from Section 2 – Rivolta direction, which provides continuous traffic data over multiple years. These variations were used to scale the representative seasonal profile of Section 1 for years with missing or incomplete data.
10. Generation of monthly traffic values for missing years:
    For each year with missing data, monthly traffic values were reconstructed by combining:

- the seasonal profile derived from the completed years of Section 1, and
- the annual scaling factors obtained from the reference section.

11. Integration into the final dataset:
    All reconstructed monthly traffic values were integrated into the final dataset used for capacity estimation, volume-to-capacity analysis, and crash prediction modeling.

Interpretation of the following tables

It presents the available and reconstructed monthly traffic values for Section 1 – Rivolta direction. The figure highlights the two-stage estimation procedure, showing first the completion of missing months within 2023 and 2024 and subsequently the reconstruction of traffic data for years with more extensive gaps.

SECTION 1 RIVOLTA . DIR.						
	2019	2020	2021	2022	2023	2024
Jan						17160.419
Feb						
Mar						
Apr						
May						
Jun					18526.05	18679.967
Jul					17160.42	18050.161
Aug					11218.58	11523.742
Sep					18254.23	19007.067
Oct					18147.32	18470.533
Nov					17391.94	18087.621
Dec					14829.65	15422.839

Table 4-3Available data for section 1 rivolta di adda Direction

SECTION 1 RIVOLTA. DIR.						
	2019	2020	2021	2022	2023	2024
Jan	15475.7	12896.5	14186.1	14001.1	16988.8	17160.4
Feb	15472.9	12894	14183.4	13461.3	14402.1	14547.6
Mar	20877.9	17398.2	19138.1	20301.5	17130.4	17303.5
Apr	27237.8	20665.2	22731.8	22731.8	16984	17155.6
May	28340.2	25118	27629.8	27629.8	18004.4	18186.3
Jun	28622.7	22091.6	23964	24106.3	18526.1	18680
Jul	22455.9	18713.3	19965	17353.7	17160.4	18050.2
Aug	12101.3	11525.1	12829	11809.6	11218.6	11523.7
Sep	25603.1	21335.9	22643.8	20363.6	18254.2	19007.1
Oct	21692.1	18076.7	19859.9	16625.1	18147.3	18470.5
Nov	12757.6	10631.3	19342.6	17556.2	17391.9	18087.6
Dec	11062.6	9218.8	10140.7	12604.8	14829.7	15422.8
AVERAGE	20141.6	16713.7	18884.5	18212.1	16586.5	16966.3

Table 4-4 : Estimated Data for section 1 Milan Direction

4.3.3 Application of the Estimation Procedure to Section 3

The estimation methodology described above was not limited to Section 1 but was consistently applied to Section 3 in both travel directions in order to obtain complete monthly traffic datasets. As for Section 1, Section 3 presented partial traffic records for several months and years, which required reconstruction to ensure temporal continuity.

For each direction of travel, a reference section with complete monthly data in the same direction was identified, and the same two-stage procedure was followed. First, missing months within partially observed years were estimated using seasonal relationships derived from the corresponding reference section. Subsequently, once complete yearly traffic profiles were obtained, the methodology was extended to reconstruct traffic data for years with larger data gaps by transferring inter-annual trends from the reference section.

By applying the same estimation logic to Section 3 in both directions, a consistent and homogeneous traffic dataset was obtained across all analyzed sections of the Rivoltana Road. This ensures comparability between sections and directions and provides a reliable basis for subsequent capacity calculations, volume-to-capacity analysis, and crash prediction modeling.

SECTION 1 MILANO . DIR.						
	2019	2020	2021	2022	2023	2024
Jan	13470	11225	14593	13512	14728	14471
Feb	15333	12778	16611	15381	16765	14055
Mar	14430	12025	15632	14474	15777	16670
Apr	15027	12522	16279	15073	16430	17087
May	15538	12948	16833	15586	16989	17668
Jun	16146	13455	17492	16196	17654	18019
Jul	19597	16331	17964	16947	16778	18005
Aug	13386	11655	12821	12095	11974	12707
Sep	20676	17230	18353	17881	17702	18410
Oct	16008	13340	14674	13843	13705	14253
Nov	12944	10787	11866	11194	11082	11525
Dec	9247.8	7706.5	8477.1	10681	13031	13564
AVERAGI	15200	12667	15183	14405	15218	15536

SECTION 1 RIVOLTA. DIR.						
	2019	2020	2021	2022	2023	2024
Jan	15476	12896	14186	14001	16989	17160
Feb	15473	12894	14183	13461	14402	14548
Mar	20878	17398	19138	20302	17130	17303
Apr	27238	20665	22732	22732	16984	17156
May	28340	25118	27630	27630	18004	18186
Jun	28623	22092	23964	24106	18526	18680
Jul	22456	18713	19965	17354	17160	18050
Aug	12101	11525	12829	11810	11219	11524
Sep	25603	21336	22644	20364	18254	19007
Oct	21692	18077	19860	16625	18147	18471
Nov	12758	10631	19343	17556	17392	18088
Dec	11063	9218.8	10141	12605	14830	15423
AVERAGI	20142	16714	18885	18212	16587	16966

SECTION 2. MILANO . DIR.						
	2019	2020	2021	2022	2023	2024
Jan	11351	9453	10310	10253	11782	11577
Feb	14314	12423	13547	12781	13412	11244
Mar	13219	11016	12007	13140	12622	13336
Apr	15553	10397	11333	9100.8	12720	13430
May	15889	13425	14634	13526	13782	13962
Jun	15142	11854	15321	12896	12671	13500
Jul	16017	13347	14631	12072	10888	13028
Aug	9786	9515	10189	8796.7	8023.5	8363.1
Sep	17672	14727	15854	13750	12372	14020
Oct	16866	14055	15785	13432	13353	14469
Nov	11418	9515.2	13543	13539	13446	10514
Dec	12876	10730	6407.3	9680.9	11846	12444
AVERAG	14225	11706	12802	11914	12293	12491

SECTION 2 RIVOLTANA . DIR.						
	2016	2020	2021	2022	2023	2024
Jan	11502	9584.3	10543	10406	12626	12236
Feb	15304	12753	14028	13314	14245	12123
Mar	13720	11433	12576	13341	11257	14420
Apr	16190	12283	13511	13511	10095	14296
May	15890	14084	15492	15492	10095	15155
Jun	16151	12466	13522	13603	10454	14345
Jul	16860	14050	14989	13029	12884	13245
Aug	9304.4	9432.8	10500	9665.6	9181.9	8988.1
Sep	18200	15166	16036	14475	12976	14798
Oct	17518	14598	16038	13426	14655	15786
Nov	10246	8538.5	15535	14100	13968	11531
Dec	12656	10546	8672.4	10780	12682	13418
AVERAG	14785	12305	13724	13158	12045	13464

SECTION 3 . MILANO . DIR.						
	2016	2020	2021	2022	2023	2024
Jan	6245.8	5204.8	5673.3	5645.1	6483.2	4630.9
Feb	7001.9	5834.9	6360	6000.2	6236.6	4437.6
Mar	7821.6	6518	7104.7	7774.7	7468.2	5334.5
Apr	3135.9	6147.2	6700.5	5380.9	7520.9	5372.1
May	3014.1	7616.6	8302.1	7673.4	7818.7	5584.8
Jun	3936.8	7825.9	10115	8514.1	8365.5	5975.4
Jul	11032	3193.4	10119	8315	7499.3	5356.6
Aug	6308	6133.4	6567.5	5670.3	5171.9	3694.2
Sep	11876	3696.4	10654	3240	8716.8	6226.3
Oct	7475.6	6229.7	6936.4	5953.4	5918.3	4079.6
Nov	5917.8	4931.5	7019.3	7016.8	6968.9	4977.8
Dec	6606.2	5505.1	3287.3	4966.8	6077.8	4341.3
AVERAG	8207.6	6753.1	7408.2	6845.9	7025.5	5005.9

SECTION 3 RIVOLTANA . DIR.						
	2016	2020	2021	2022	2023	2024
Jan	5573.1	4644.3	5108.7	5042.1	6118	4894.4
Feb	6512.1	5426.8	5969.5	5665.5	6061.5	4849.2
Mar	8787	7322.5	8054.7	8544.4	7209.8	5767.8
Apr	11464	8697.5	9567.2	9567.2	7148.2	5718.5
May	11928	10572	11629	11629	7577.6	6062.1
Jun	11929	9207	9987.3	10047	7721	6176.8
Jul	10315	8596.1	9171.1	7971.6	7882.8	6306.2
Aug	5864.2	5584.9	6216.8	5722.8	5436.4	4349.1
Sep	13980	11650	12364	11119	9367.5	7974
Oct	8196.9	6830.8	7504.6	6282.2	6857.5	5400.1
Nov	5423.2	4519.3	8222.4	7463	7393.2	5914.6
Dec	6153.6	5128	4216.8	5241.4	6166.6	4933.3
AVERAG	9141.2	7594.1	8445.8	8113.9	7220.2	5768.3

Table 4-5 Construction of all traffic flow data for all sections

4.4 Traffic Analysis

The traffic analysis of the SP14 Rivoltana Road is based on comprehensive data from three sections, each of which has two directions of travel: Section 1 (Milano Direction and Rivolta Direction), Section 2 (Milano Direction and Rivolta Direction), and Section 3 (Milano Direction and Rivolta Direction). The data span 2019 to 2024, with monthly Average Daily Traffic (ADT) recorded for each section and direction.

Traffic Flow Analysis Methodology

For each road segment, traffic flow was analyzed by considering the directionality of the flow and the availability of data for each section. The approach is as follows:

1. Complete Segments in a Single Section:
   If a segment is fully contained within a single section, the traffic flow is directly taken from the corresponding section. This ensures accurate representation of traffic flow without the need for estimation or interpolation.
2. Segments Spanning Multiple Sections:
   If a segment lies across or between two sections, interpolation is applied to estimate the traffic flow. The interpolation process considers the traffic flow data from the two neighboring sections and adjusts for any differences in traffic volume and directional movement between the two sections.
3. Interpolation Process:
   The interpolation procedure involves analyzing entrances and exits for each segment to determine how traffic flows across the boundaries of the sections. The number of entrances (where vehicles enter the segment) and exits (where vehicles leave the segment) is counted, and the traffic volume is adjusted based on this distribution. This allows for an accurate estimate of traffic flow on segments that span multiple sections.

Key Observations from Traffic Data

- Higher Traffic in Milano Direction: Traffic flow data consistently show higher volumes in the Milano direction due to the road's role as a primary entry point into Milan.
- Traffic Variability: Traffic volumes exhibit noticeable fluctuations based on time of year, with peak volumes occurring during spring and autumn, while winter and summer months typically show lower traffic flow.
- Impact of Missing Data: For years with incomplete data, the interpolation method ensures that missing months or sections can still be filled with reasonable estimates, maintaining the consistency and accuracy of the traffic analysis.

This section will now proceed to analyze traffic trends, estimate capacity, and explore the relationship between traffic conditions and crash likelihood across these segments. The use of interpolation ensures a continuous, reliable dataset for the entire Rivoltana Road, even when data gaps occur.

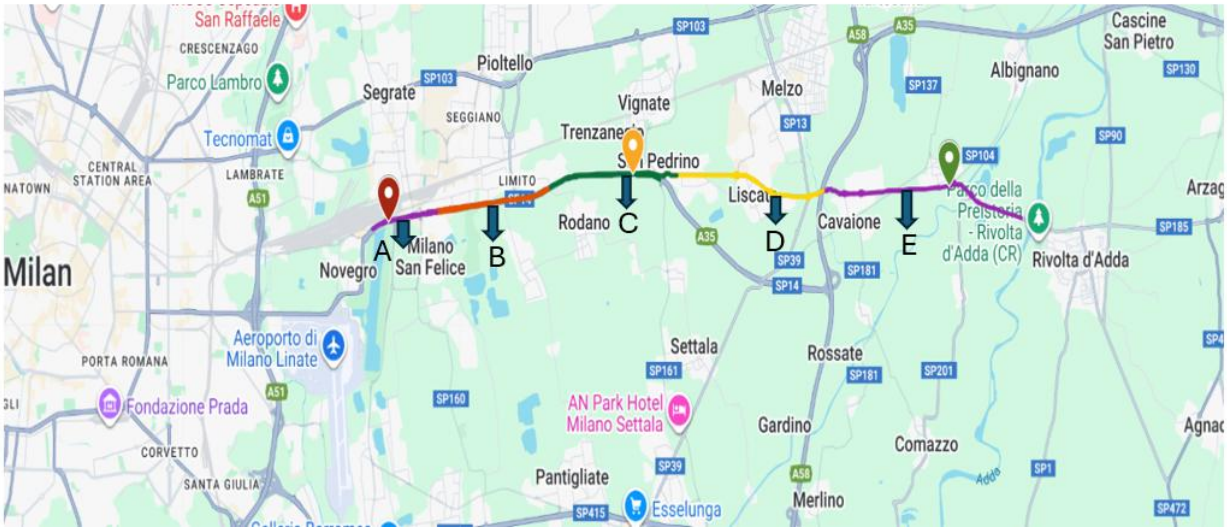


Figure 4-2 overview of divided sections and segments

Traffic Trend Analysis for Each Section and Segment Overview of Divided Sections

The Rivoltana Road (SP14) is divided into three distinct sections, each with two directions of travel:

- Section 1 (Milano Direction and Rivolta Direction)
- Section 2 (Milano Direction and Rivolta Direction)
- Section 3 (Milano Direction and Rivolta Direction)

These sections are represented by color-coded segments in the map, which helps in understanding the varying traffic patterns and trends along the road. The color-coding allows us to associate specific traffic behavior with geographic locations, such as high-traffic urban areas and lower-traffic suburban or rural segments.

Each section will be analyzed based on:

1. Traffic volume trends (monthly, seasonal variation)
2. Peak traffic conditions (morning and evening peaks)
3. Changes in traffic flow due to road characteristics (intersections, entry/exit points, access density)

4.4.1 Traffic Flow Analysis – Section 1

Section 1 represents the westernmost part of the SP14 Rivoltana, closest to the Milan metropolitan area. Due to its strategic role, this section experiences significant traffic demand in both directions. Traffic flow analysis is conducted separately for the Milano direction (1M) and the Rivolta direction (1R) to capture directional differences and temporal trends.

Section 1 – Milano Direction (1M)

The Milano direction in Section 1 shows consistently high traffic volumes throughout the analyzed period (2020–2024), reflecting strong commuter demand toward the metropolitan area.

Annual Traffic Trend

- The average daily traffic (ADT) ranges approximately from:
  - ~12,700 veh/day in 2020
  - to ~15,500 veh/day in 2024
- A clear increase is observed after 2020, indicating recovery and growth following the pandemic period.
- The highest annual average is recorded in 2024, confirming a return to and exceedance of pre-pandemic traffic levels.

Monthly and Seasonal Pattern

- Peak traffic months consistently occur between June and September, with:
  - July and September often exceeding 18,000–20,000 veh/day.
- Lower traffic volumes are observed during:
  - December (minimum values, around 7,700–10,700 veh/day),
  - January–February, reflecting reduced travel activity.
- The seasonal pattern is stable across all years, suggesting a strong and recurring commuter-based demand.

Key Observations

- The Milano direction exhibits:
  - pronounced summer peaks,
  - strong weekday commuter influence,
  - and high exposure levels relevant for safety analysis.
- This direction is expected to operate closer to capacity during peak months, increasing the likelihood of congestion-related crashes.

Section 1 – Rivolta Direction (1R)

The Rivolta direction shows higher traffic volumes than the Milano direction, indicating asymmetric travel demand along Section 1.

Annual Traffic Trend

- The average daily traffic in the Rivolta direction ranges approximately from:
  - ~16,700 veh/day in 2020
  - to ~17,600 veh/day in 2024
- Compared to the Milano direction, the Rivolta direction consistently records higher annual averages.
- The increase from 2020 onward is smoother, indicating a more stable demand profile.

Monthly and Seasonal Pattern

- Peak values are observed between April and October, with:
  - several months exceeding 20,000 veh/day,
  - and peak summer values approaching 28,000 veh/day (notably May–July).
- Winter months (December–February) still show relatively high volumes compared to the Milano direction, rarely dropping below 9,000–11,000 veh/day.

Key Observations

- The Rivolta direction is characterized by:
  - higher baseline traffic,
  - less pronounced winter reductions,
  - and stronger continuity across the year.
- This suggests a mix of commuter, regional, and through traffic, making this direction particularly relevant for capacity and safety evaluations.

Analysis of Monthly Traffic Flow in Section 1 (Year 2024)

Figure 4-3 presents the monthly average daily traffic for Section 1 in 2024, comparing the Milano direction and the Rivolta direction. The graph highlights clear directional asymmetry and seasonal variability in traffic demand.

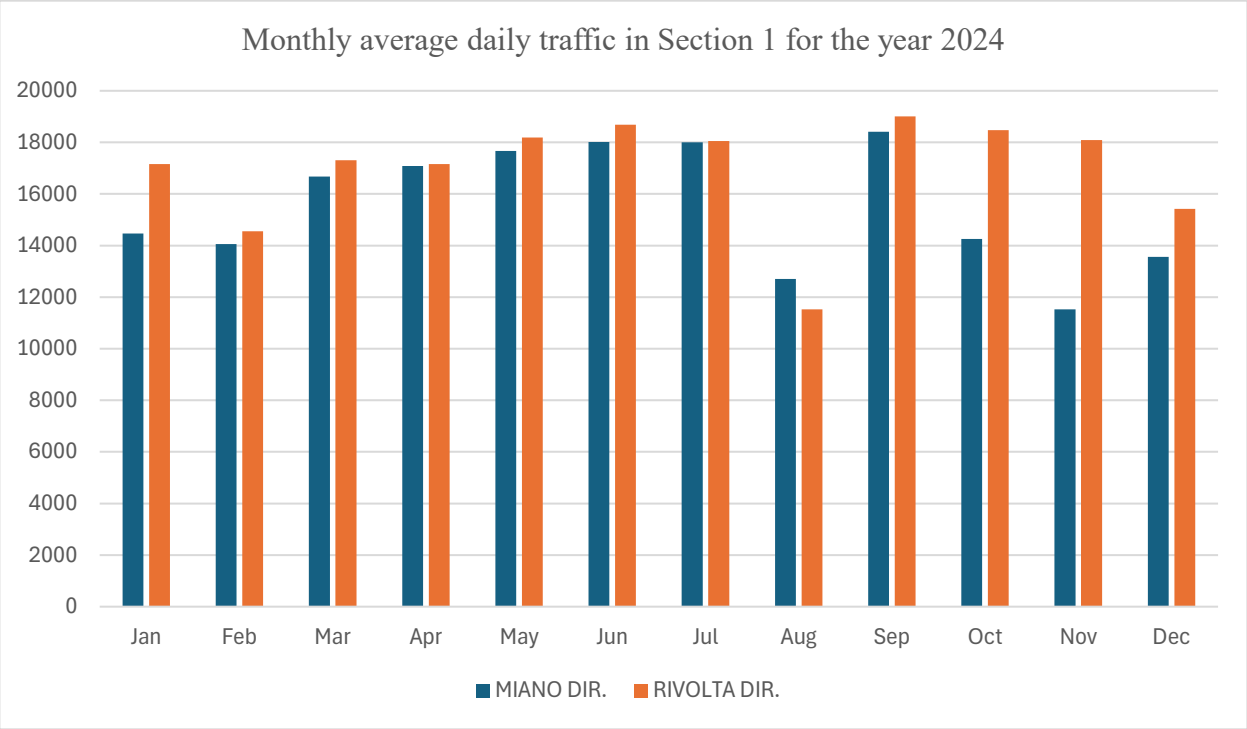


Figure 4-3 : Monthly average daily traffic in Section 1 for the year 2024

Directional Comparison

Across almost all months, the Rivolta direction consistently carries higher traffic volumes than the Milano direction. The difference between directions ranges approximately from 500 to more than 6,000 vehicles per day, with the largest discrepancies observed in October and November. This confirms that the Rivolta direction experiences a higher baseline demand, likely due to a combination of commuter, regional, and through traffic.

An exception is observed in August, where the Milano direction slightly exceeds the Rivolta direction. This inversion may be associated with summer travel behavior, holiday movements, and reduced commuter demand in the Rivolta direction during this period.

Seasonal Traffic Patterns

Both directions exhibit a clear seasonal pattern:

- Winter months (January–February and December) show relatively lower traffic volumes, particularly in the Milano direction.
- Spring and early summer (March–June) are characterized by a steady increase in traffic in both directions, reaching peak values in June.
- August presents a marked traffic reduction in both directions, more pronounced in the Rivolta direction.
- Autumn months (September–October) show a renewed increase in traffic, especially in the Rivolta direction, which reaches its annual maximum in September.

This seasonal behavior reflects typical commuter and recreational travel dynamics along the Rivoltana corridor.

Implications for Traffic and Safety Analysis

The consistently higher traffic volumes observed in the Rivolta direction imply:

- increased exposure to potential conflicts,
- higher demand relative to capacity during peak months,
- and a greater relevance for congestion-related crash risk.

Conversely, the Milano direction shows stronger seasonal variability, suggesting periods of lower demand that may influence crash patterns differently

Traffic Flow Analysis for Section 1 – Year 2019

The following Figure illustrates the monthly average daily traffic in Section 1 for the year 2019, representing pre-COVID traffic conditions. The figure shows traffic volumes for both the Milano direction and the Rivolta direction, providing a baseline for comparison with subsequent years.

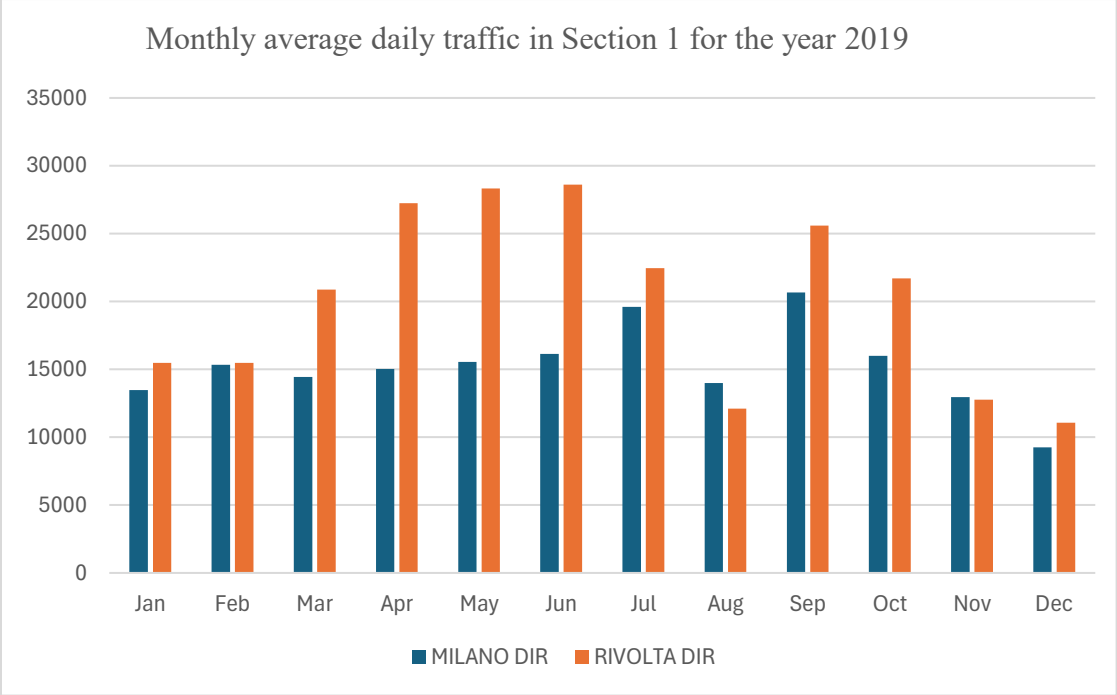


Figure 4-4 : Monthly average daily traffic in Section 1 for the year 2019

Directional Traffic Characteristics in 2019

In 2019, the Rivolta direction consistently carried higher traffic volumes than the Milano direction across almost all months. Peak values in the Rivolta direction occur between April and June, reaching values close to 28,000–29,000 veh/day, while the Milano direction peaks at approximately 20,000–21,000 veh/day during the same period.

The Milano direction shows a more moderate increase throughout the year, with relatively stable traffic between January and June, followed by a decline during the summer months. In contrast, the Rivolta direction exhibits stronger seasonal variation, with a sharp increase in spring and early summer and a pronounced reduction in August, likely related to holiday travel patterns.

Seasonal Pattern in 2019

Both directions exhibit a clear and typical seasonal trend:

- Winter months (January–February and December) show lower traffic volumes.
- Spring (March–June) corresponds to the highest traffic demand, particularly in the Rivolta direction.
- August presents a substantial drop in traffic in both directions, more pronounced in the Rivolta direction.
- Autumn months (September–October) show a secondary increase in traffic volumes before declining again toward the end of the year.

This seasonal behavior reflects normal pre-pandemic commuting and regional travel dynamics along the Rivoltana corridor.

Comparison Between 2019 and 2024

A comparison between 2019 and 2024 highlights important changes in traffic demand over time and provides insight into the impact of the COVID-19 pandemic and subsequent recovery.

Pre-COVID vs Post-Recovery Traffic Levels

- In 2019, traffic volumes in the Rivolta direction were significantly higher than in 2024, especially during the spring and early summer months. Peak values in 2019 exceed those observed in 2024 by approximately 8,000–10,000 veh/day in some months.
- The Milano direction also shows higher peak values in 2019 compared to 2024, although the difference is less pronounced than in the Rivolta direction.

These differences indicate that traffic demand in 2024 has not fully returned to pre-pandemic levels, particularly for through and regional traffic movements.

COVID-19 Impact on Traffic Patterns

The sharp contrast between 2019 and 2024 traffic volumes can be attributed to the effects of the COVID-19 pandemic, which caused:

- severe mobility restrictions in 2020–2021,
- long-term changes in commuting behavior (e.g., remote work),
- and reduced regional and discretionary travel.

Although 2024 shows a clear recovery compared to pandemic years, traffic volumes remain lower and more evenly distributed across the year. In particular, the extreme spring peaks observed in 2019 are no longer present, suggesting a structural change in travel demand rather than a temporary disruption.

Directional Differences Over Time

- The Rivolta direction experienced the largest absolute reduction in traffic volumes from 2019 to 2024, indicating that this direction was more affected by changes in long-distance and regional travel.
- The Milano direction shows a more stable trend, with smaller differences between pre- and post-pandemic years, likely due to the persistence of essential commuting flows.

Implications for Traffic and Safety Analysis

The comparison between 2019 and 2024 demonstrates that traffic exposure levels have changed substantially over time. Since crash occurrence is strongly related to traffic volume, these variations must be explicitly considered in crash prediction modeling. Using multiple years of traffic data allows the models to capture both pre-pandemic conditions and post-pandemic traffic behavior, improving the robustness of the safety analysis.

4.4.2 Traffic Flow Analysis for Each Segment

Traffic flow analysis was conducted for all sections (Sections 1, 2, and 3) and for both travel directions (Milano → Rivolta and Rivolta → Milano) using the complete dataset available for the period 2018–2024. As no missing values were identified within this time frame, traffic volumes were directly allocated to segments without requiring interpolation or reconstruction procedures.

Traffic Flow Methodology for Each Segment.

Given the availability of complete monthly traffic data for all sections and both travel directions, segment-level traffic assignment was performed through a structured allocation procedure based on observed section-level volumes.

Step 1: Direct Use of Complete Traffic Data (2018–2024)

The full monthly traffic dataset was applied directly to each segment according to its corresponding section and travel direction. No estimation techniques were required, as the dataset contained continuous and complete observations for the entire study period. This approach ensures consistency between recorded section-level flows and segment-level representation.

Step 2: Consistent Distribution Across Segments

The spatial configuration of Sections 1,2,3 and Segments A,B,C,D,E is illustrated in Figures 4-1 and 4-2. Traffic assignment depends on the spatial relationship between segments and sections:

• Segments fully contained within a single section (Segments A, C, and E):
For segments entirely located within one section, traffic flow was directly assigned from that section for both travel directions. No proportional adjustment or redistribution was necessary.

• Segments spanning two adjacent sections (Segments B and D)**:**
For segments located between two sections, traffic volumes were proportionally allocated using data from the adjacent sections. The distribution accounted for traffic transitions between sections, considering entrances and exits to ensure a realistic representation of flow continuity along the corridor.

Step 3: Annual and Monthly Distribution

For all years (2018–2024), the recorded monthly traffic values were directly applied at the segment level. Seasonal patterns, including peak and off-peak periods, were already embedded within the observed dataset; therefore, no additional seasonal adjustment was required.

The resulting allocation framework guarantees coherence between section-level observations and segment-level traffic representation, thereby providing a consistent basis for subsequent capacity estimation, V/C classification, and probabilistic crash prediction modeling.

4.5 Weekday, Weekend, and Flow Profiles Analysis

Traffic demand along the Rivoltana Road varies significantly not only across months and years but also according to the type of day and the temporal distribution of flow within the day. For this reason, the traffic analysis was extended beyond average daily values to explicitly distinguish between weekday traffic, weekend traffic, and daily flow profiles. This approach allows a more realistic representation of traffic exposure and operational conditions, which is particularly relevant for capacity evaluation and crash prediction.

Weekday Traffic Flow Estimation

Weekday traffic represents the dominant component of traffic demand along the Rivoltana Road and is mainly associated with commuting and work-related trips. To estimate representative weekday traffic levels, the available traffic data were analyzed by isolating weekday observations.

For each selected sample, the weekday average flow was calculated as the arithmetic mean of daily traffic volumes observed from Monday to Friday. This approach reduces the influence of day-to-day variability and provides a stable estimate of typical weekday demand. The resulting weekday flow values were found to be consistent across different samples, confirming the reliability of the method.

Weekday traffic is characterized by:

- higher overall daily volumes,
- strong directional imbalance during peak hours,
- and pronounced temporal concentration during commuting periods.

These features make weekday traffic particularly relevant for identifying capacity constraints and congestion-related safety risks.

Weekend Traffic Flow Estimation

Weekend traffic exhibits different characteristics compared to weekdays, reflecting a shift from commuting trips to leisure, recreational, and discretionary travel. Using the same dataset, weekend traffic levels were computed as the average of daily flows observed on Saturday and Sunday for multiple representative samples.

To quantify the relationship between weekend and weekday traffic, a weekend factor was derived by calculating the ratio between weekend and weekday average flows. The analysis consistently showed that weekend traffic volumes are lower than weekday volumes. Across the analyzed samples, the weekend flow was found to be approximately 75% of the weekday flow, leading to the following relationship:

$$K_{\text{weekend}} \approx 0.75$$

This result indicates that, on average, weekend traffic demand along the Rivoltana Road is about one quarter lower than weekday demand.

The weekend factor was then applied to estimate weekend flows from weekday values, ensuring consistency across sections, directions, and years. This approach allows weekday and weekend traffic to be represented separately, even when only aggregate daily values are available.

Daily Flow Profiles

While weekday and weekend averages describe total daily demand, they do not capture how traffic is distributed over the course of the day. For this reason, daily flow profiles were used to represent the temporal evolution of traffic within a typical day.

Flow profiles describe the percentage of daily traffic occurring during each hour and differ between:

- weekdays, which typically show distinct morning and evening peak periods, and
- weekends, which are characterized by flatter profiles with less pronounced peaks.

Based on the available flow profile data, weekday profiles were associated with:

- a sharp morning peak linked to inbound commuting,
- a secondary evening peak linked to return trips,
- and lower flows during mid-day and night-time periods.

Weekend profiles, in contrast, show:

- delayed peak hours,
- more evenly distributed daytime traffic,
- and reduced peak intensities.

These differences highlight the importance of accounting for flow profiles when analyzing operational performance and safety, as traffic concentration during peak hours can significantly influence congestion and crash likelihood.

Figure 4-5 presents the hourly flow profiles for weekday and weekend traffic, expressed as the proportion of daily traffic occurring in each hour. The profiles highlight substantial differences in the temporal distribution of traffic demand between weekdays and weekends.

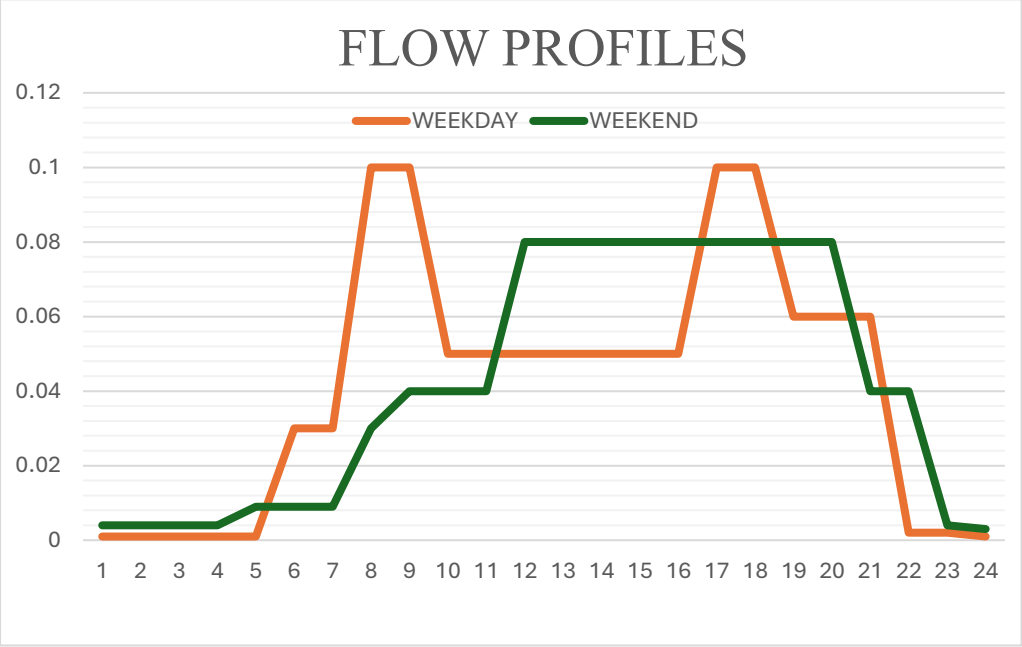


Figure 4-5: flow profiles

Integration of Weekday, Weekend, and Flow Profiles

The combined use of:

- weekday average flows,
- weekend average flows derived using the weekend factor,
- and representative daily flow profiles,

allows a comprehensive characterization of traffic demand along the Rivoltana Road. This integrated approach ensures that both traffic volume and temporal distribution are accurately represented.

Such a detailed traffic characterization is particularly important for:

- identifying peak operating conditions,
- computing volume-to-capacity ratios under realistic scenarios,
- and improving the accuracy of crash prediction models by accounting for variations in traffic exposure over time.

5 Crash Analysis

5.1 Introduction

A rigorous crash analysis is essential to understand the safety performance of a roadway and to support the development of reliable crash prediction models. This chapter focuses on the preparation, verification, and analysis of crash data for the SP14 Rivoltana corridor, with particular emphasis on ensuring spatial and temporal consistency between crash records, road segmentation, and traffic conditions.

Given that the probabilistic modeling framework adopted in this study operates at the segment level, accurate crash–segment association represents a critical prerequisite. Any spatial misalignment or uncertainty in crash location may lead to incorrect attribution of crashes, potentially biasing the estimation of crash likelihood and the identification of high-risk segments. For this reason, the crash dataset is subjected to a systematic verification and filtering process prior to any modeling activity.

The chapter first describes the procedures used to validate crash locations and align them with the segmented road network, including visual inspection and spatial correction where possible. Crashes that cannot be reliably associated with a specific segment due to incomplete or ambiguous location information are identified and excluded to preserve data quality. Subsequently, the filtered crash dataset is analyzed to characterize crash frequency, spatial distribution, and contextual factors relevant to traffic conditions and road geometry.

By establishing a high-quality, spatially consistent crash database, this chapter provides the empirical foundation for the probabilistic crash prediction models presented in the subsequent sections. The analysis ensures that observed crash patterns reflect real roadway conditions and that the resulting model outputs can be interpreted with confidence for safety assessment and decision-making.

5.2 Crash Data Verification and Spatial Filtering

The crash dataset used in this study originates from the official road accident database of Regione Lombardia and was previously available as a public dataset. Although the dataset is no longer publicly accessible, it represents the official source of crash information for the SP14 Rivoltana corridor and was made available for academic research purposes.

As a first step, the dataset was visually inspected and spatially verified to ensure consistency with the adopted road segmentation. To this end, the MATLAB routine plot_segments_withcheck.m was executed, allowing a direct comparison between crash locations and the predefined road segments.

Based on this verification process, the segmentation boundaries were adjusted where necessary to ensure correct alignment with the road geometry and full coverage of the intended study area. After completing these adjustments, the crash–segment association procedure was repeated.

The updated spatial analysis identified a total of 41 crashes located within 21 segments, of which 18 crashes were fully contained within segment boundaries. The relatively limited number of crashes assigned to segments is primarily due to incomplete or imprecise geographic coordinates in several crash records. Consequently, crashes that could not be reliably associated with a specific segment were excluded from the segment-level analysis to avoid introducing spatial uncertainty.

This filtering procedure ensures that only crashes with a clear and reliable spatial correspondence to the road segments are considered in the subsequent analysis. Although this reduces the total number of usable crash records, it enhances the robustness and credibility of the segment-based safety assessment and the crash prediction modeling developed in the following sections.

5.3 Crash Relocation and Data Quality Assessment

Following the initial spatial verification, a detailed review of crash locations was conducted to improve the alignment between crash records and the road segmentation framework. An initial screening revealed that approximately **70 crashes** were not spatially located within the defined road segments. These crashes were identified and marked accordingly (highlighted in red in the figure).

	crash code	COORD_X	COORD_Y
2	155633	9.2845700000	45.4656500000
3	165062	9.4128400000	45.4675900000
4	12803	9.2617900000	45.4713300000
5	44071	9.2616600000	45.4714900000
6	32568	9.2767248407	45.4717535423
7	17743	9.2750207972	45.4717688189
8	133068	9.2720000000	45.4723100000
9	155589	9.2718700000	45.4724500000
10	155629	9.2722200000	45.4726100000
11	129713	9.2750600000	45.4726200000
12	155620	9.2750600000	45.4726200000
13	134442	9.2742000000	45.4726300000
14	164568	9.2732400000	45.4726400000
15	154834	9.2750600000	45.4726500000
16	160948	9.2739500000	45.4727000000
17	167061	9.2739500000	45.4727000000
18	20109	9.2856068646	45.4727747923
19	50766	9.285662114792	95.472760302080
20	18890	9.2801584873	45.4729061069
21	19087	9.2802029528	45.4729210293
22	20112	9.2864097184	45.4730211165
23	138912	9.2863000000	45.4730500000
24	155580	.286605193882475	5.473641242321
25	29742	9.2829356300	45.4733238336
26	18584	9.2872624404	45.4733871115
27	133194	.285462578762135	5.473718153889
28	155588	9.2798600000	45.4735900000
29	155601	9.2798400000	45.4736000000
30	155603	9.2796700000	45.4736200000
31	155577	9.2798700000	45.4736500000
32	172940	9.2867000000	45.4737100000
33	155596	9.2845000000	45.4738500000
34	138440	9.2851000000	45.4738700000
35	179676	9.2863400000	45.4739000000
36	155632	9.2863756000	45.4739164000
37	155606	9.2847800000	45.4739400000
38	182869	9.2853000000	45.4739500000
39	155622	9.2837200000	45.4739600000
40	155576	9.2826800000	45.4742300000
41	45553	9.2890354834	45.4745980284
42	180039	9.2894500000	45.4754200000
43	39922	9.2928040586	45.4755682666
44	175730	9.2901400000	45.4756600000
45	39269	9.2942758663	45.4757469322
46	19573	9.2945578921	45.4757573664
47	39649	9.2975768597	45.4760111971
48	26057	.299236709212285	5.477048326347

67	20191	.304431973377445	45.47750203211116
68	28143	9.3067887232	45.4770777342
69	11437	.301239522065765	45.4772292127588
70	181139	.303640741441675	45.4773768313943
71	34347	.308904455645745	45.4779468602265
72	18143	9.3010474000	45.4772620000
73	155593	9.2991600000	45.4772700000
74	155635	9.3005800000	45.4773000000
75	134629	9.3007600000	45.4773342000
76	155582	9.3009600000	45.4775600000
77	155640	9.3009800000	45.4775700000
78	172893	9.3009500000	45.4775900000
79	20497	9.3042300000	45.4778300000
80	155636	9.303601185493215	45.4777670817485
81	155634	.304278954416725	45.4778456695541
82	174901	9.3086500000	45.4779300000
83	174445	.304078921070025	45.4778242209579
84	134341	9.314957315509575	45.4783563416555
85	10837	.325616150457855	45.4792261586711
86	11967	.325760135137595	45.4792307680888
87	10844	9.325673937391115	45.4792158779859
88	10835	.325901533015295	45.4792408120804
89	41582	9.325815607118705	45.4792357686727
90	30735	.325991999438955	45.4792547734341
91	180179	9.4844485000	45.4786798000
92	20608	9.4820200000	45.4791400000
93	171200	9.3342475000	45.4796986000
94	49973	9.4149183696	45.4801484630
95	138712	9.4630000000	45.4807500000
96	163077	9.4777900000	45.4808400000
97	173519	9.3517948000	45.4810967000
98	165442	9.4638700000	45.4815400000
99	138690	9.4638810000	45.4815900000
100	178611	9.4639971000	45.4816578000
101	138697	9.4663333000	45.4817340000
102	164991	9.4692400000	45.4819000000
103	20509	9.464008099525	55.4816173203257
104	48030	.386701660568325	45.4837448428334
105	180353	9.3508200000	45.4834500000
106	178593	9.3803800000	45.4836700000
107	32236	9.3818860000	45.4838690000
108	164302	9.3635662000	45.4838858000
109	138875	9.3740000000	45.4853000000
110	20570	9.3766100000	45.4854500000
111	20647	8.9350300000	45.5009100000
112	168866	8.8749200000	45.5822600000

Table 5-1 crash data filtering

A systematic relocation procedure was then applied to these crashes in order to maximize the number of usable records while preserving data reliability. As a result of this process:

- **22 crashes** were successfully relocated and reassigned to appropriate segments (marked in green),
- **48 crashes** could not be reliably relocated and were therefore excluded from further analysis (marked in yellow), due to incomplete, ambiguous, or inconsistent location information.
- **41 crashes** were already located (marked in blu)

Crash Relocation Procedure

For each crash considered for relocation, a multi-step verification process was followed. First, the descriptive attributes available in the original crash dataset were examined, with particular attention to the road name and roadway characteristics associated with each crash. These characteristics included the road type (one-way or two-way), road function (mainline or intersection area), and geometric context (rectilinear section, curve, or roundabout).

Next, the recorded crash location was compared with the nearest road segment in the current segmentation framework. A crash was relocated only if the descriptive attributes of the crash were consistent with the geometric and operational characteristics of the candidate segment. In cases where the road name and roadway characteristics matched the segment context, the crash location was adjusted accordingly and reassigned to the segment.

If inconsistencies were identified—such as a mismatch between the reported road type and the geometric features of the nearby segment—the crash was excluded to avoid introducing spatial or contextual errors into the analysis.

Final Dataset for Segment-Level Analysis

This relocation and filtering process resulted in a refined crash dataset containing only records with a clear and reliable spatial association to the Rivoltana Road segments. Although a portion of the original crash records could not be used, the adopted approach ensures that the final dataset is based on high-quality spatial information, which is essential for meaningful segment-level safety analysis and crash prediction modeling.

The resulting dataset forms the basis for the crash frequency analysis, exposure-based comparisons, and probabilistic crash modeling presented in the subsequent sections.

5.4 Analysis of Crash Frequency per Segment

Figure 5-1 illustrates the distribution of crashes across the road segments of the SP14 Rivoltana corridor after the crash relocation and spatial filtering process. The results show that crash occurrence is not uniformly distributed along the corridor, but instead varies significantly from one segment to another.

The majority of segments experienced a low number of crashes, with most segments recording one or two crashes over the analysis period. This indicates that, for a large portion of the corridor, safety performance is relatively stable and that crash events are infrequent at the segment level. Such behavior is typical of linear road infrastructures where exposure is spread across multiple homogeneous sections.

In contrast, a limited number of segments exhibit noticeably higher crash frequencies. In particular, Segment 19 presents the highest number of crashes, followed by Segments 120, 1, 9, 70, and 104, which also show elevated crash counts compared to the rest of the corridor. These segments can be interpreted as potential safety-critical locations, where local geometric features, traffic demand variations, or operational conditions may increase crash risk.

The presence of a small number of high-crash segments alongside many low-crash segments suggests a heterogeneous safety performance along the Rivoltana Road. This heterogeneity justifies the adoption of a segment-based crash analysis approach, as aggregated corridor-level indicators would mask localized safety issues. Moreover, the observed variability provides a strong motivation for developing crash prediction models that explicitly account for segment-level traffic exposure and roadway characteristics.

Overall, the crash frequency distribution highlights the importance of identifying and analyzing individual segments rather than relying solely on average crash rates, thereby enabling more targeted and effective safety evaluations and interventions.

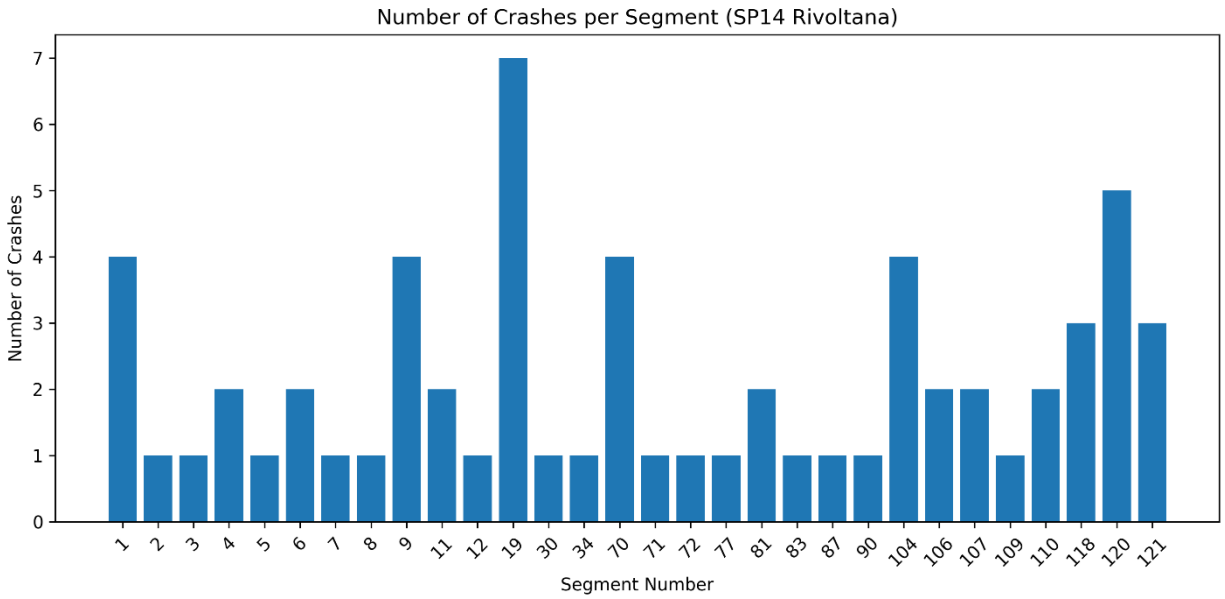


Figure 5-1 : number of crashes per segment sp14

Temporal Analysis of Crashes

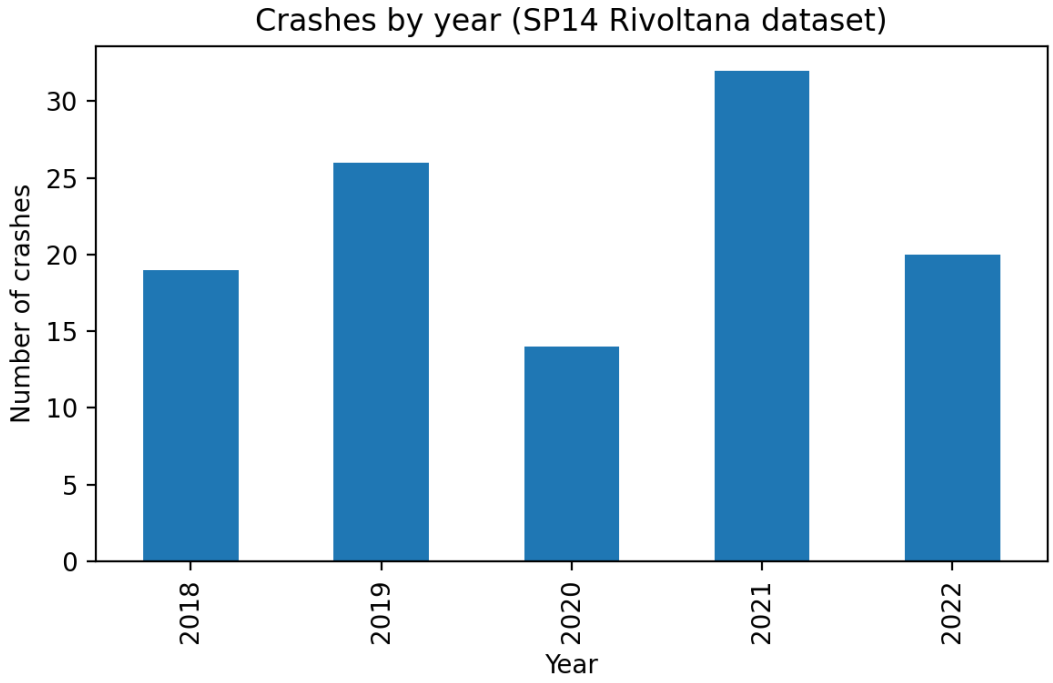
This subsection analyzes the temporal distribution of crashes based on the crashes that were reliably assigned to road segments. The analysis considers yearly and monthly patterns, in order to identify long-term trends and seasonal effects in crash occurrence along the SP14 Rivoltana corridor.

Crash Occurrence by Year

The distribution of crashes by year shows that crash events are spread across the period 2018–2022, with noticeable variability between years. A relatively higher number of crashes is observed in the years 2018 and 2019, followed by a clear reduction in 2020. This decrease is consistent with the reduction in traffic volumes and mobility restrictions associated with the COVID-19 pandemic.

In the subsequent years (2021 and 2022), crash frequency increases again, reflecting the gradual recovery of traffic demand. This trend confirms the strong relationship between traffic exposure and crash occurrence and highlights the importance of accounting for traffic flow variations when interpreting safety performance.

Figure 5-2 Bar chart – Number of crashes per year (segment-assigned crashes)

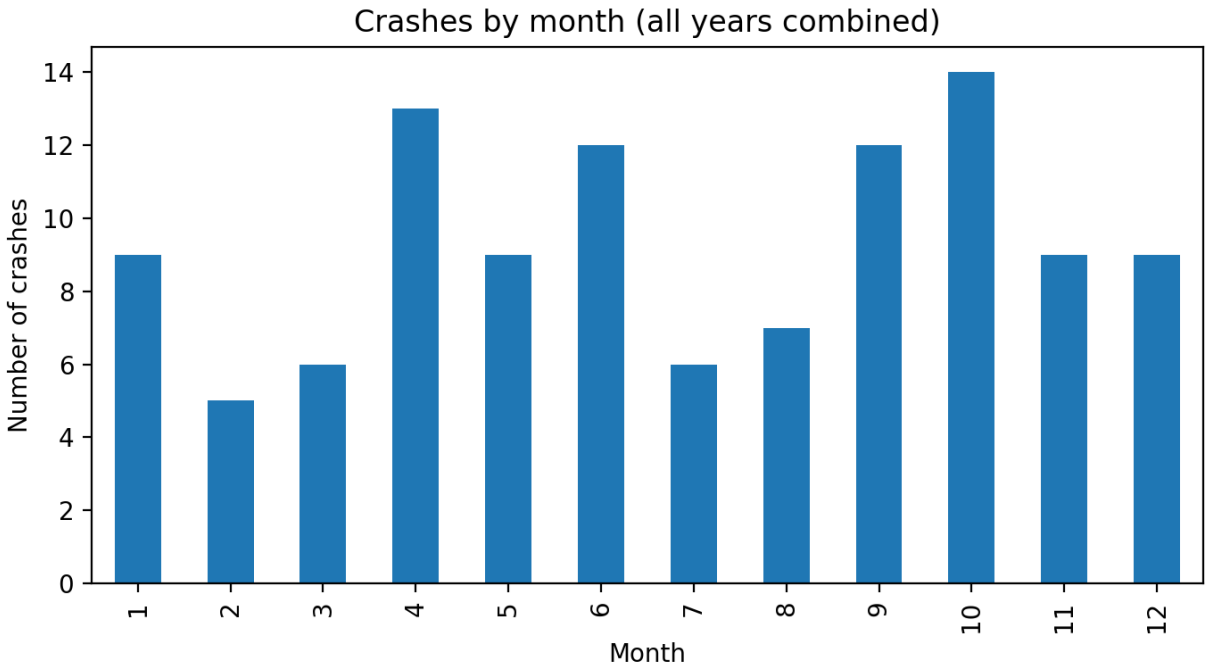


Crash Occurrence by Month (Seasonal Effects)

The monthly distribution of crashes reveals a non-uniform seasonal pattern. Crashes tend to occur more frequently during spring and autumn months, while relatively fewer crashes are observed during winter. This behavior can be explained by a combination of factors, including seasonal variations in traffic demand, daylight conditions, and driving behavior.

Summer months also show a moderate number of crashes, which may be related to increased leisure travel and heterogeneous traffic conditions. The absence of a single dominant month indicates that crash risk is present throughout the year, but with varying intensity depending on seasonal conditions.

Figure 5-3: Bar chart – Number of crashes by month (all segments combined)



Temporal Distribution within the Study Period

The temporal analysis confirms that crashes are not randomly distributed over time but are influenced by broader traffic and mobility trends. The reduction observed in 2020, followed by a recovery in later years, aligns with the traffic flow analysis presented in previous chapters. This consistency strengthens the reliability of the dataset and supports the integration of traffic exposure variables in subsequent crash prediction modeling.

Link with Traffic Analysis

The observed temporal patterns justify the use of multi-year traffic data and hourly flow profiles in the crash modeling framework. Since crash occurrence varies over time in response to changes in traffic demand, capacity conditions, and operational characteristics, the inclusion of temporal variability is essential for accurate estimation of crash risk at the segment level.

Crash Timing versus Peak and Off-Peak Traffic

To further investigate the relationship between traffic conditions and crash occurrence, crashes were analyzed with respect to peak and off-peak traffic periods. Peak periods were defined based on the previously established weekday traffic flow profiles, which show two dominant peaks corresponding to morning commuting hours and late afternoon commuting hours, while the remaining hours of the day were classified as off-peak.

Based on the traffic analysis, peak hours were identified approximately as showing

- Morning peak: 07:00–09:00
- Midday peak : 12:00-02:00
- Afternoon peak: 16:00–18:00

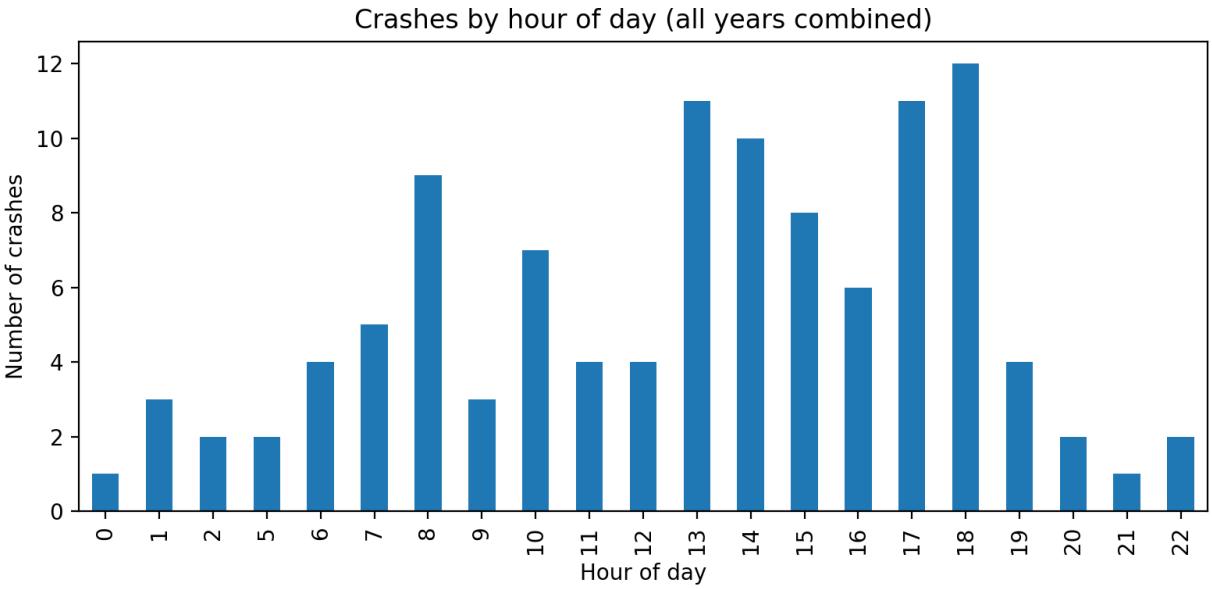


Figure 5-4: crash by hour of day (all years combined)

All other hours were classified as off-peak, including night-time and mid-day periods with lower or more stable traffic demand.

Crash Occurrence during Peak Periods

A significant share of crashes occurred during peak traffic periods. This concentration is expected, as peak hours correspond to the highest traffic volumes, increased congestion, and greater interaction between vehicles. During these periods, traffic conditions are often characterized by:

- reduced headways,
- frequent speed changes,
- stop-and-go conditions,
- and higher driver workload.

These factors increase the likelihood of rear-end collisions and lateral conflicts, which are among the most frequent crash types observed along the corridor. The coincidence of high crash occurrence with peak periods highlights the strong role of traffic exposure and operational stress in influencing crash risk.

Crash Occurrence during Off-Peak Periods

Crashes occurring during off-peak periods are fewer in number but remain relevant from a safety perspective. Off-peak crashes are more dispersed throughout the day and are often associated with:

- higher operating speeds,
- reduced congestion,
- and lower traffic density.

Under such conditions, crashes may be less frequent but potentially more severe, particularly in cases involving loss of control, run-off-road events, or front-lateral collisions. Night-time off-peak crashes, although limited in number, may also be influenced by reduced visibility and driver fatigue.

Comparison between Peak and Off-Peak Conditions

The comparison between peak and off-peak crash occurrence indicates that:

- Peak periods account for a disproportionate share of crashes, primarily due to increased exposure and vehicle interactions.
- Off-peak periods show lower crash frequencies, but crashes in these periods may be associated with higher speeds and different crash mechanisms.

This distinction confirms that crash risk is not solely a function of traffic volume, but also depends on the traffic state (congested versus uncongested) and driver behavior under different operating conditions.

Implications for Traffic and Safety Modeling

The observed relationship between crash timing and peak/off-peak traffic conditions supports the use of hourly traffic volumes and V/C ratios in crash prediction modeling. By disaggregating traffic demand by time of day, it becomes possible to:

- identify critical operating periods,
- evaluate crash risk under different congestion levels,
- and improve the explanatory power of segment-level safety models.

In particular, peak-period conditions emerge as a key contributor to crash frequency, while off-peak conditions require attention due to their association with higher speeds and potentially higher crash severity.

Interpretation of Crash Type Distribution

As illustrated In Figure 5-5, we report the distribution by type of crashes along the SP14 Rivoltana corridor. The most frequent crash type is rear-end collisions (Tamponamento), which account for a substantial share of the total crashes. This result is consistent with traffic conditions characterized by variable speeds, frequent braking, and congestion, particularly during peak hours. The second most common category is front-lateral collisions (Scontro frontale-laterale), typically associated with conflicts at intersections, access points, and roundabouts. Other crash types, such as run-off-road events (Fuoriuscita) and impacts with obstacles (Urto con ostacolo accidentale), occur less frequently but remain relevant from a safety perspective, as they are often associated with loss of control or higher speeds. Overall, the crash type distribution suggests that both congestion-related mechanisms and geometric conflict points play an important role in crash occurrence along the corridor.

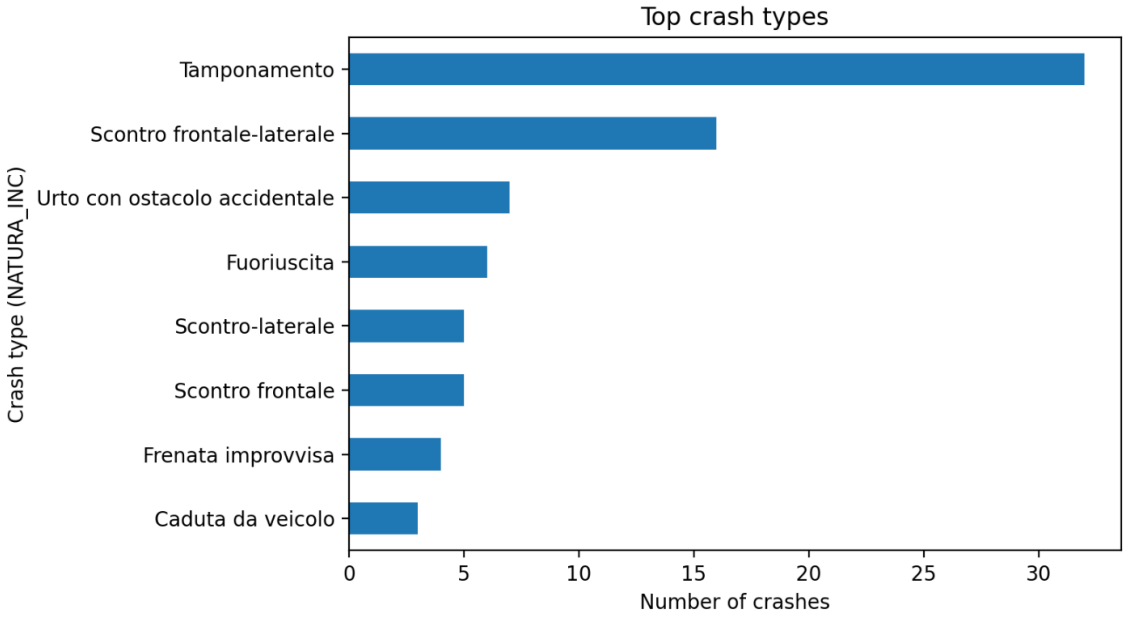


Figure 5-5 : numbers of crashes per each crash type

Interpretation of Crash Location Context (Intersection Type)

The following figure 5-6 presents the distribution of crashes by road element and intersection context. The majority of crashes occurred on rectilinear segments (Rettilineo), which is expected given that these segments represent the largest share of the corridor length and traffic exposure. Roundabouts (Rotatoria) constitute the second most frequent crash location, highlighting their role as concentrated conflict areas involving merging and yielding maneuvers.

Crashes on curved segments (Curva) and signalized intersections (Intersezione segnalata) are less frequent but indicate localized safety concerns related to reduced visibility, speed adaptation, and complex vehicle interactions. The limited number of crashes occurring at narrow sections or traffic-calming features suggests either lower spatial prevalence of these elements along the corridor or reduced exposure levels.

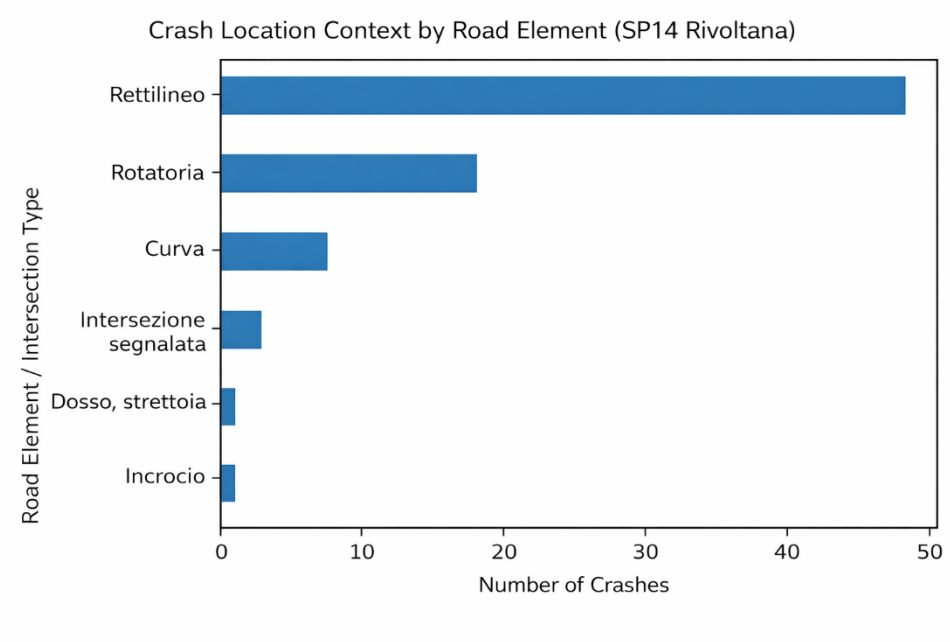


Figure 5-6 : distribution of crashes by road element and intersection context

Interpretation of Road Surface Conditions

The following Figure 5-7 shows crash occurrence by pavement surface condition. Most crashes occurred on dry pavement (Asciutto), followed by a smaller share on wet pavement (Bagnato). This result reflects exposure rather than risk, as dry conditions represent the majority of driving time. Nevertheless, the presence of crashes under wet, icy (Ghiacciato), and slippery (Sdrucciolevole) conditions indicates that reduced friction and vehicle stability can contribute to crash occurrence when adverse surface conditions are present. Although the absolute number of such crashes is low, their potential severity warrants attention, particularly during winter and rainy periods.

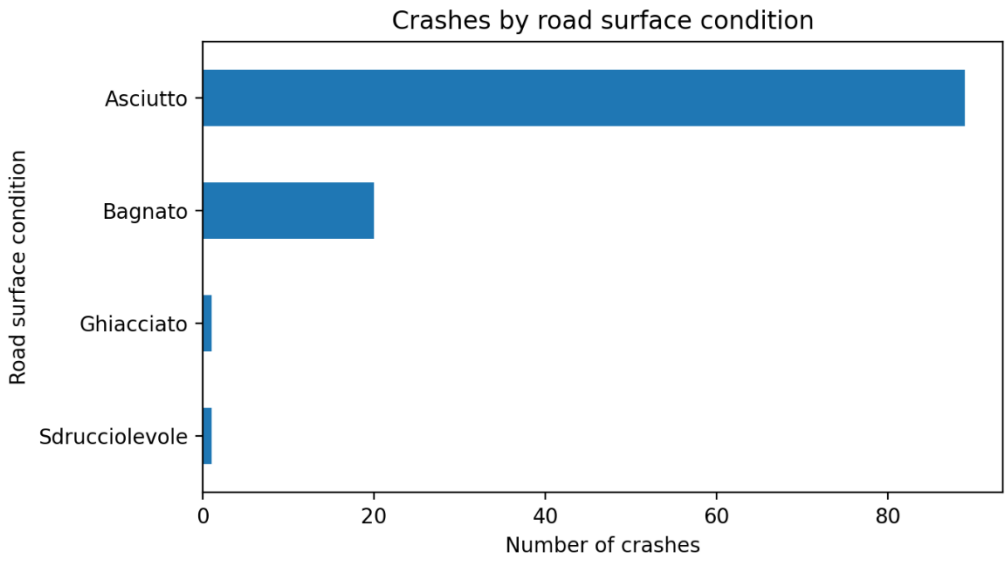


Figure 5-7: crash occurrence by pavement surface condition

Interpretation of Weather Conditions

The following Figure 5-8 presents the distribution of crashes by weather condition. The vast majority of crashes occurred under clear weather (Sereno) conditions, which again reflects higher exposure during normal driving conditions rather than safer operation. Crashes during rain (Pioggia), fog (Nebbia), and snow (Neve) are less frequent but remain significant because adverse weather can reduce visibility and road friction, increasing crash risk. The presence of crashes under fog and snow, although limited in number, highlights the importance of considering environmental factors in safety evaluations, particularly for high-speed or rural segments.

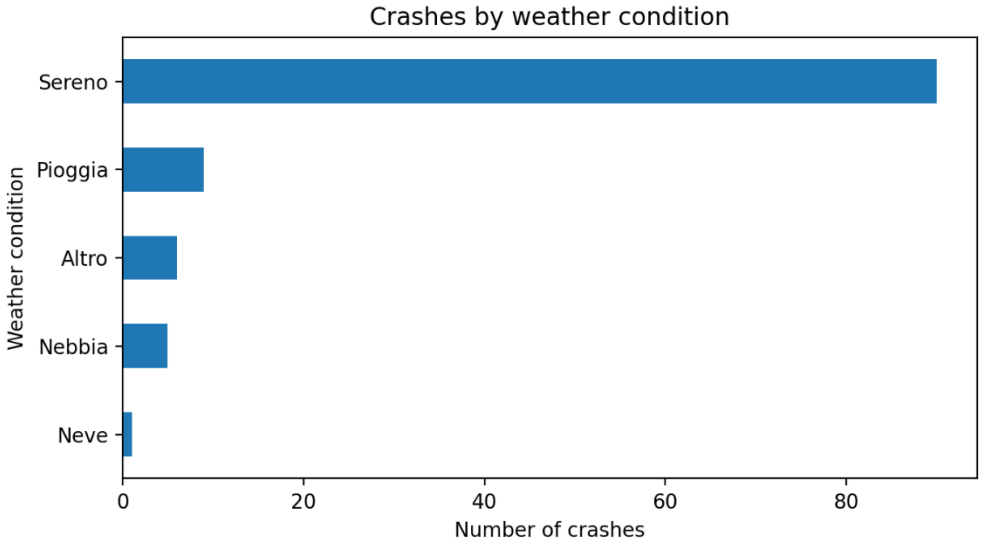


Figure 5-8: distribution of crashes by weather condition

Overall Interpretation and Link to Modeling

Taken together, these figures indicate that crash occurrence along the SP14 Rivoltana corridor is influenced by a combination of traffic conditions, road geometry, and environmental factors. Rear-end crashes on straight segments under clear and dry conditions dominate the dataset, suggesting that congestion, speed variability, and driver behavior are key contributors to crash risk. At the same time, the occurrence of crashes at roundabouts, curves, and under adverse conditions supports the inclusion of geometric and environmental variables in the subsequent crash prediction models

5.5 Conclusion

The crash analysis conducted along the SP14 Rivoltana corridor highlights clear and consistent patterns in crash occurrence that are strongly influenced by traffic demand, roadway characteristics, and temporal conditions. The results show that crash frequency is not uniformly distributed across the corridor, but instead concentrated in specific segments, particularly those characterized by higher traffic volumes, geometric complexity, or proximity to intersections and roundabouts.

Temporal analysis indicates that a substantial proportion of crashes occurs during peak traffic hours, when congestion, speed variability, and frequent vehicle interactions increase the likelihood of conflicts, especially rear-end and lateral collisions. In contrast, crashes during off-peak periods are less frequent but may be associated with higher operating speeds and different crash mechanisms. Seasonal and annual trends further demonstrate the close relationship between traffic exposure and crash occurrence, with a noticeable reduction in crashes during 2020, corresponding to the mobility restrictions imposed during the COVID-19 pandemic, followed by an increase in subsequent years as traffic volumes recovered.

Environmental conditions also play an important role in crash occurrence. Although most crashes occur under dry pavement and clear weather conditions—reflecting normal exposure levels—the presence of crashes during wet pavement, rain, fog, and icy conditions highlights the increased vulnerability of the roadway under reduced friction and visibility. Additionally, the predominance of rear-end crashes on rectilinear segments suggests that traffic flow instability and driver behavior are key contributors to crash risk along the corridor.

Overall, this analysis confirms that crash occurrence on the SP14 Rivoltana is governed by a combination of traffic exposure, operational conditions, roadway geometry, and external factors. These findings provide a robust foundation for the development of segment-level crash prediction models and support the use of detailed traffic flow and capacity indicators, such as hourly volumes and V/C ratios, to better understand and manage safety performance along the corridor.

6 Methodology for Crash Prediction on the SP14 Rivoltana Road

6.1 Introduction

The objective of this methodological framework is to identify road segments along the SP14 Rivoltana Road that exhibit an elevated probability of crash occurrence. To achieve this goal, crash prediction models are developed and applied at the segment level, integrating historical crash data, roadway geometric characteristics, and traffic flow information.

The approach aims to support the identification of critical segments by analyzing both the magnitude and the stability of predicted crash counts. In particular, segments that consistently present higher predicted crash values with low variability across multiple simulations are considered statistically significant and indicative of increased safety risk. This methodology provides a quantitative basis for comparing safety performance across segments and for prioritizing targeted safety interventions.

6.2 Data Sources

Three primary datasets were used to develop the crash prediction models for the SP14 Rivoltana Road:

Crash Data (Rivoltana40.mat)

This dataset contains detailed information on individual crash events occurring along the study corridor. The data were collected by ISTAT and provided by Regione Lombardia, and include key attributes such as crash identification code, geographic coordinates, date and time of occurrence, crash type, and severity indicators. These data form the temporal and spatial foundation of the crash analysis and were previously filtered and validated to ensure consistency with the adopted segmentation framework.

Segment and Traffic Data (segments_all.mat)

This dataset describes the geometric and operational characteristics of the road segments defined in the segmentation procedure presented in Chapter 3. Each segment is represented by a polygonal shape defined through coordinate points extracted during the segmentation process. The dataset includes geometric attributes such as segment length, alignment type, and segment category.

In addition, the file contains traffic-related variables for multiple years, including estimated Average Daily Traffic (ADT) values and corresponding capacity estimates. Traffic data are provided separately for weekday and weekend conditions, allowing the computation of volume-to-capacity (V/C) ratios under different operating scenarios. These variables are fundamental inputs for the crash prediction models, as they represent traffic exposure and congestion levels at the segment scale.

Crash–Segment Mapping (crash.mat)

This dataset was generated through a dedicated preprocessing procedure designed to associate crashes with road segments while accounting for spatial uncertainty in crash location data. A buffer region was created around each segment polygon to capture crashes whose recorded coordinates may not lie exactly on the segment centerline due to GPS inaccuracies.

Crashes falling within a segment buffer were assigned to that segment, provided that no overlap with other segment buffers occurred. This process ensured a unique and consistent crash–segment association. The resulting dataset includes the following output variables:

- **`crash`**: number of crashes assigned to each segment;
- **`crashcode`**: mapping between individual crash identification codes and segment indices.

These variables serve as the dependent inputs for the crash prediction models developed in the subsequent sections.

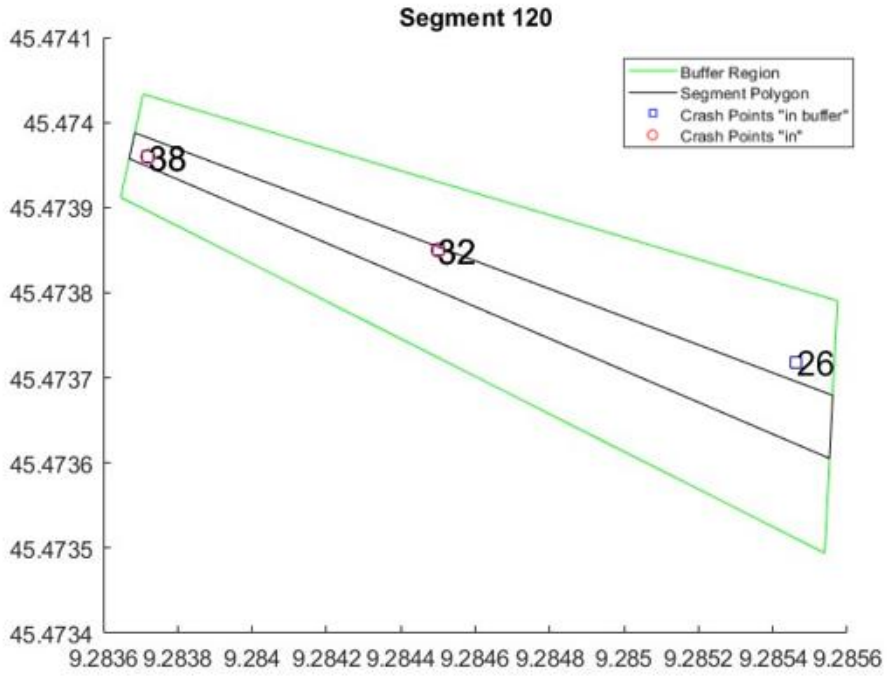


Figure 6-1: crash–segment mapping for Segment 120, showing the segment polygon, buffer region, and associated crash points.

6.3 Volume-to-Capacity Classification

For each crash recorded in the Rivoltana crash dataset, the first step of the analysis consists of identifying the road segment on which the crash occurred. This information is obtained from the crash–segment mapping procedure described in Section 4.2, which assigns each crash to a unique segment through the `crashcode` variable. Once the segment identifier is retrieved, the temporal attributes of the crash—including date, year, time of occurrence, and day type (weekday or weekend)—are extracted from the crash dataset.

Based on this temporal information, the corresponding traffic demand and capacity values are selected for the identified segment. Average Daily Traffic (ADT) values and segment capacities are chosen according to the year of the crash and the type of day, ensuring that the estimated traffic conditions are representative of those prevailing at the time of the event. This step is particularly important given the observed temporal variability in traffic volumes across years and between weekday and weekend conditions along the SP14 Rivoltana corridor.

To further refine the estimation of traffic conditions at the exact time of each crash, the selected ADT value is disaggregated into hourly traffic flow using the standardized weekday and weekend flow profiles developed in Chapter 3. These profiles define the proportion of daily traffic occurring in each hour of the day and allow the estimation of the hourly flow corresponding to the crash time.

Once the hourly traffic flow and the segment capacity are available, the volume-to-capacity (V/C) ratio is computed as the ratio between hourly flow and capacity. The V/C ratio provides a normalized indicator of the operational state of the segment at the time of the crash, allowing traffic conditions to be compared consistently across different segments and time periods.

Based on the calculated V/C ratio, each crash is classified into one of four traffic condition classes:

- VC Class 1: uncongested conditions ($V/C < 0.25$);
- VC Class 2: low to moderate traffic conditions ($0.25 \leq V/C < 0.50$);
- VC Class 3: moderate to high traffic conditions ($0.50 \leq V/C < 0.75$);
- VC Class 4: congested conditions ($V/C \geq 0.75$).

This classification provides a structured representation of the traffic context in which each crash occurred and forms a key explanatory input for the crash prediction models developed in the following sections. By linking crash occurrence to specific congestion levels, the methodology enables a more nuanced understanding of segment-level safety performance under varying traffic conditions.

6.4 Crash Matrix Construction

To support probabilistic crash modeling along the SP14 Rivoltana Road, a three-dimensional crash matrix, denoted as:

$$\text{mat\_inc(type,segment,VC)}$$

was constructed. Each element of this matrix represents the number of crashes observed for a given segment type, specific road segment, and Volume-to-Capacity (V/C) class.

The matrix was populated by iterating over all crashes retained after the relocation and filtering procedures described in Chapter 3. For each crash:

- The associated segment ID was retrieved from the crash–segment mapping dataset.
- The segment type and geometric attributes were identified from the segmentation framework.
- The crash year and day type (weekday/weekend) were extracted.
- The corresponding ADT and segment capacity were obtained using the traffic assignment procedure described in Section 4.3.
- ADT was disaggregated into hourly flow using standardized daily profiles.
- The V/C ratio was computed and classified into one of four predefined congestion classes.

Each crash contributed one unit to the corresponding matrix element:

$$\text{mat\_inc(segment\_type,crashSeg,VC\_class)} = \text{mat\_inc(segment\_type,crashSeg,VC\_class)} + 1$$

The resulting matrix captures the empirical crash distribution as a function of geometry, segment identity, and traffic congestion. It constitutes the empirical foundation for the probabilistic models described below.

6.5 Method 1: Length- and Volume-to-Capacity-Based Weighted Model

6.5.1 Model Description

Method 1 estimates crash likelihood by combining geometric exposure and congestion-related effects. The method relies directly on the previously constructed crash matrix (mat_inc), from which two probability components are derived:

- a length-based component, and
- a V/C-based congestion component.

Their relative contribution is controlled by the weighting parameter α.

6.5.2 Crash Likelihood Estimation

- **Length-Based Probability (matpb)**

For each segment, total crashes (summed across all V/C classes) are normalized by segment length to obtain crash density. This density is compared with the average density of the corresponding segment type. If the segment-level value is lower than the type-level average, it is adjusted upward to mitigate data sparsity effects.

- **V/C-Based Probability (matpl)**

For each segment, crashes are weighted according to their V/C class using predefined multipliers:

$$w = [0.25 - 0.50 - 0.75 - 1.00]$$

These weights assign progressively greater importance to crashes occurring under higher congestion levels. The weighted crash counts are summed and normalized to obtain the V/C-based probability measure.

6.5.3 Combined Likelihood Estimation

The final crash likelihood for each segment is obtained through a convex weighted combination of the two probability components:

$$matF = \alpha \cdot matpl + (1 - \alpha) \cdot matpb$$

where:

- $matpl$ is the V/C-based probability vector,
- $matpb$ is the length-based probability vector,
- $\alpha \in [0,1]$controls the relative influence of congestion and geometric exposure.

The resulting vector is normalized to obtain a probability distribution over all segments:

$$pdf_1 = \frac{matF}{\sum matF}$$

The vector pdf_1 represents the probability of crash occurrence for each segment and serves as the input distribution for the subsequent Monte Carlo simulation.

6.5.4 Crash Simulation via Monte Carlo

To simulate the stochastic nature of crash occurrence and evaluate the stability of the predicted crash distribution, a Monte Carlo simulation is performed using the probability distribution pdf_1. The simulation is executed for 20,000 iterations, following these steps:

1. In each iteration, a set of random numbers uniformly distributed between 0 and 1 is generated;
2. Each random number is mapped to a road segment using a cumulative probability distribution constructed from pdf_1;
3. A counter is incremented to record the number of simulated crashes assigned to each segment during that iteration;
4. After completing all iterations, two statistical indicators are computed for each segment:

- the mean predicted number of crashes, representing the expected crash frequency;
- the standard deviation of the predicted crash counts across iterations, representing the variability of the estimate.

This simulation framework allows the assessment of both expected crash frequency and the uncertainty associated with the prediction for each segment.

6.5.5 Identification of Significant Segments

To identify road segments with stable and reliable crash predictions, a coefficient of variation (CV) is computed for each segment as the ratio between the standard deviation and the mean of the predicted crash counts:

$$\mathrm{CV} = \frac{\sigma}{\mu}$$

where $\mu$is the mean predicted crash count and $\sigma$is the corresponding standard deviation.

A segment is classified as significant if its standard deviation is less than 70% of its predicted mean, indicating relatively low variability and higher statistical confidence in the predicted crash frequency. Segments satisfying this criterion are considered robust candidates for safety assessment and prioritization.

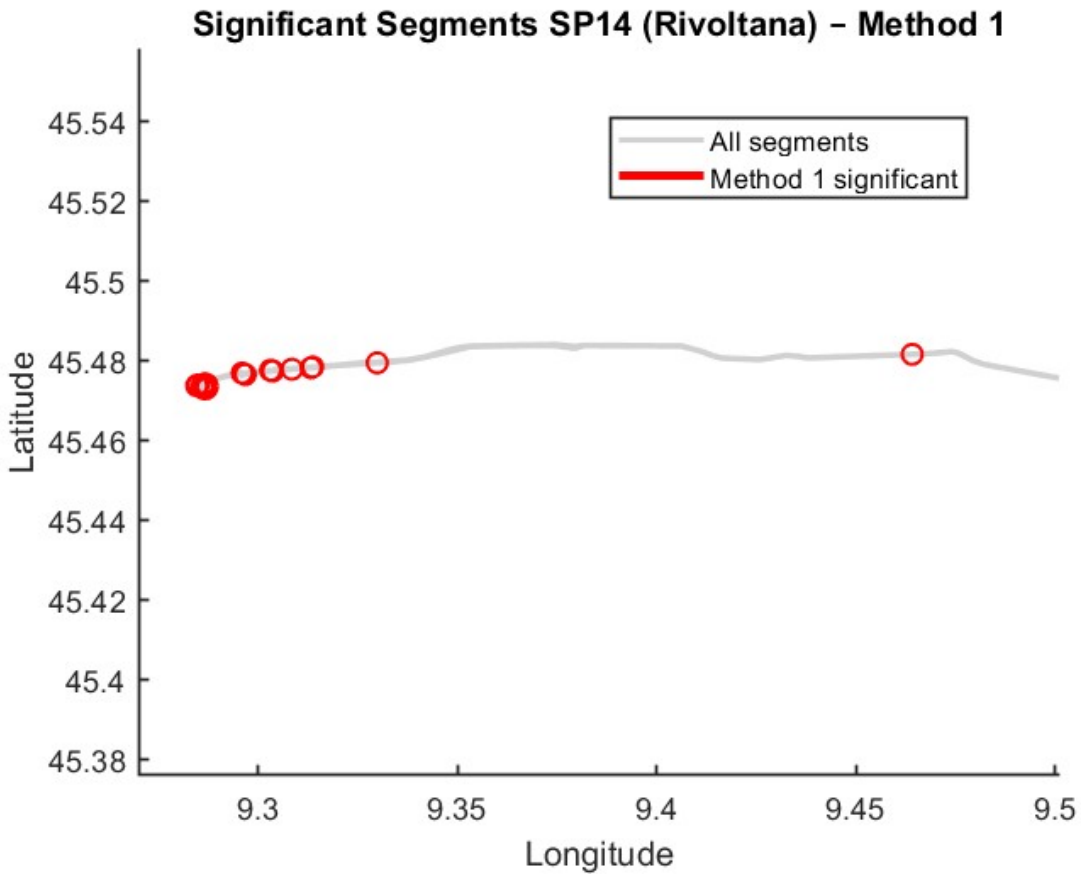
The identified significant segments are stored in the output matrix `outc`, which includes:

- the segment index,
- the segment type,
- the mean predicted crash count,
- the observed number of crashes.
-

The resulting output, summarized in Table 6.1, provides a concise representation of segments exhibiting elevated and stable crash risk according to Method 1.

Segment ID	type	Predicted Crashes	Actual Crashes
1	9	4.412	4
6	9	2.8229	2
9	9	5.9857	4
11	9	3.2338	2
12	9	2.041	1
19	9	9.1805	7
70	5	2.8171	4
103	9	4.3738	4
117	9	2.915	3
119	9	3.4599	5
120	9	2.7238	3

Table 6-1: significant segment output 'outc' (Method 1).



6.6 Segment-Level Interpretation of Method 1 Results

Table 6.1 reports the predicted and observed crash counts for the segments identified as statistically significant by Method 1. Overall, the results show a good agreement between predicted and actual crashes, confirming the ability of the Length–VC-based model to capture segment-level crash exposure along the Rivoltana Road.

Most significant segments belong to Type 9, which corresponds to acceleration/deceleration lanes and similar transition elements. These segments consistently exhibit elevated crash risk due to frequent speed changes, merging and diverging maneuvers, and increased driver workload. For example, Segment 1 and Segment 103 show predicted crash values (4.41 and 4.37, respectively) that closely match the observed counts (4 crashes each), indicating high model accuracy for these locations.

Segment 19 stands out as the most critical location, with the highest predicted crash count (9.18) and a similarly high observed value (7 crashes). This confirms that the model correctly identifies this segment as a dominant hot spot, likely due to a combination of geometric exposure and traffic saturation effects.

Several segments show slight overestimation, such as Segments 9, 11, and 12, where predicted crashes exceed observed values by 1–2 crashes. This behavior is consistent with the conservative nature of Method 1, which prioritizes geometric exposure and tends to assign higher risk to longer or operationally complex segments, even when crash data are limited.

In contrast, Segment 70 (Type 5) represents a different geometric configuration, likely associated with a crossroads or intersection area. Here, the model slightly underestimates crash occurrence (2.82 predicted versus 4 observed), suggesting that intersection-related conflict dynamics may not be fully captured by length and VC effects alone.

Finally, segments such as 117, 119, and 120 show close correspondence between predicted and actual crashes, reinforcing the robustness of Method 1 in identifying stable crash-prone locations when sufficient exposure and traffic variability are present.

Overall, the segment-level analysis demonstrates that Method 1 provides reliable and interpretable predictions, particularly for acceleration/deceleration segments, while minor discrepancies highlight the influence of localized intersection dynamics and limited crash samples .

6.7 Model Accuracy Assessment

The accuracy of the proposed crash prediction model is evaluated by comparing the predicted crash counts obtained from the Monte Carlo simulation with the observed crash counts derived from the original dataset. Model performance is quantified using the Root Mean Square Error (RMSE), which provides a measure of the average deviation between predicted and observed values.

Two RMSE indicators are computed:

- RMSE_all: the RMSE calculated across all road segments, providing an overall assessment of the model's predictive capability along the entire SP14 Rivoltana corridor;
- RMSE_seg: the RMSE calculated considering only the significant segments, identified based on low variability in predicted crash counts across Monte Carlo iterations. These segments are stored in the output matrix `outc`, and their RMSE is computed by comparing the predicted mean crash counts with the corresponding observed values.

This dual evaluation approach allows the assessment of both the global performance of the model and its reliability in high-confidence segments, where predictions are more stable and statistically robust.

6.7.1 Parameter Tuning and Sensitivity Analysis

The proposed model includes parameters that influence the final crash probability distribution used in the Monte Carlo simulation. In particular, the weighting parameter αplays a central role in balancing the contribution of traffic-related effects and geometric exposure.

Weighting Parameter α

The parameter αcontrols the relative importance of the two probability components:

- the V/C-based probability (matpl), which captures the influence of traffic congestion and operating conditions;
- the length-based probability (matpb), which represents geometric exposure and segment size.

In this study, the parameter is set to:

$$\alpha = 0.65$$

This choice assigns 65% weight to the V/C-based probability and 35% weight to the length-based probability. The selected value reflects the assumption that, along the SP14 Rivoltana

Road, geometric characteristics and exposure play a dominant role in crash occurrence, while traffic congestion still contributes meaningfully to crash risk.

The adopted weighting represents a compromise between purely geometry-driven and purely traffic-driven models and ensures that both dimensions are incorporated into the crash prediction framework. The sensitivity of the results to the choice of $\alpha$is discussed in the following sections, where the stability of predicted crash distributions is evaluated.

alpha values	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5	0.55	0.6	0.65	0.7	0.75	0.8	0.85	0.9	0.95	1
rsme all	1.0143	0.9639	0.9147	0.8675	0.8192	0.777	0.7344	0.6954	0.66	0.6267	0.6	0.5757	0.5595	0.5471	0.5387	0.5426	0.5504	0.5653	0.5818
rsme sig	2.5437	2.4475	2.3426	2.1124	1.9786	1.7595	1.6226	1.5277	1.4447	1.3707	1.3309	1.2121	1.2812	1.3338	1.3922	1.4801	1.5691	1.6677	1.758

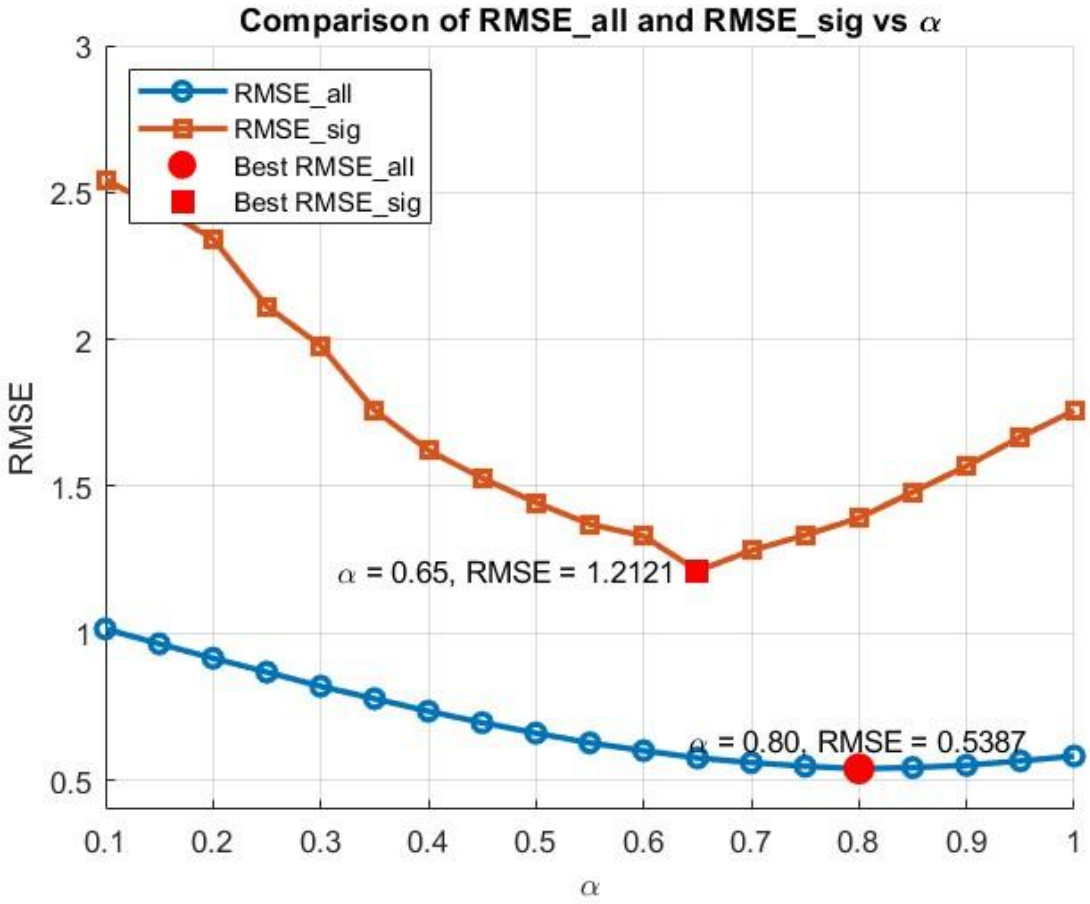


Figure 6-2: Sensitivity analysis of Alpha

6.7.2 Results and Discussion – Method 1

The performance of Method 1 (Length and VC-Based Weighted Model) was evaluated using the optimal parameter value $\alpha = 0.65$, which assigns greater importance to traffic congestion effects (VC-based probability) while still accounting for geometric exposure. This weighting reflects the strong influence of traffic operating conditions on crash occurrence along the SP14 Rivoltana corridor.

Overall Model Performance

The overall Root Mean Square Error, calculated across all segments, is:

$$\mathrm{RMSE}_{\mathrm{all}} = 0.576$$

This relatively low value indicates a good agreement between predicted and observed crash counts at the network level. The model is therefore effective in capturing the general spatial distribution of crashes along the road and provides a reliable representation of average crash frequency across segments.

Performance on Significant Segments

For the subset of significant segments, identified based on low variability in Monte Carlo predictions, the computed RMSE is:

$$\mathrm{RMSE}_{\mathrm{sig}} = 1.211$$

The higher RMSE observed for significant segments is expected and can be explained by the fact that these segments typically exhibit higher crash frequencies, making absolute prediction errors more pronounced. Nevertheless, the model successfully identifies segments with elevated risk, and the predicted values remain within a reasonable range of the observed crash counts.

Importantly, the higher RMSE for significant segments does not indicate poor model performance; rather, it reflects the increased crash intensity and variability associated with these locations. These segments are precisely the ones of greatest interest for safety assessment and intervention.

Impact of the Weighting Parameter (α = 0.65)

The optimal value α = 0.65 suggests that traffic conditions play a dominant role in explaining crash occurrence on the SP14 Rivoltana Road. This outcome highlights the relevance of congestion-related factors such as high volume-to-capacity ratios and peak-hour conditions in shaping crash risk. At the same time, the inclusion of the length-based component ensures that geometric exposure is not overlooked.

The balance achieved by this weighting confirms that crash risk on the Rivoltana is driven by an interaction between traffic demand and road geometry, rather than by a single dominant factor.

6.8 Method 2: Type–VC Matrix-Based Crash Prediction Model

Description

Method 2 adopts a categorical and hierarchical probabilistic approach to model crash occurrence along the SP14 Rivoltana Road. Unlike Method 1, which combines continuous predictors such as segment length and volume-to-capacity (VC) ratio through weighted averaging, this method relies exclusively on observed historical crash distributions across discrete categories.

Crash prediction is carried out through a sequence of probabilistic selections that mirror the empirical structure of the crash data:

1. Segment type selection: a road segment type is selected based on its historical share of crashes.
2. VC class selection: a traffic saturation class (VC Class 1 to 4) is selected, either globally or conditioned on the chosen segment type.
3. Segment selection: a specific road segment is selected within the chosen type and VC class, weighted by its observed crash frequency.

By following this hierarchical structure, the model distributes simulated crashes in a way that preserves the relative importance of segment type and traffic conditions, directly reflecting the observed categorical patterns in the Rivoltana crash dataset.

6.8.1 Methodological Steps

1. Crash Matrix Definition and Normalization

The same three-dimensional crash incidence matrix `mat_inc`, previously constructed for Method 1, is used as the foundational input for Method 2. This matrix stores the number of observed crashes indexed by:

- Segment type,
- VC class (four categories from uncongested to congested),
- Segment index.

Prior to simulation, the matrix is normalized along different dimensions to derive probability distributions required for the hierarchical selection process.

2. Probability Distribution Functions

The simulation framework is based on three nested probability distribution functions (PDFs), applied sequentially for each simulated crash event.

Type Selection PDF
A probability vector is constructed for segment type selection by dividing the total number of crashes associated with each segment type by the total number of crashes observed along the SP14 Rivoltana corridor. This PDF governs the first selection step and reflects the relative crash proneness of different road typologies.

VC Class Selection PDF
For the selected segment type, a four-element probability vector is derived representing the likelihood of each VC class. These probabilities are computed by normalizing the crash counts across VC classes within the selected type. This step captures how traffic congestion levels influence crash occurrence for each road typology.

Segment Selection PDF
Given a selected segment type and VC class, a final probability vector is constructed over all segments belonging to that category. Each segment's probability is proportional to its observed crash frequency within the corresponding type–VC group. This ensures that segments with historically higher crash occurrence have a higher chance of being selected.

These probability distributions are recomputed dynamically during the simulation process to maintain consistency with category-conditioned crash patterns.

3. Monte Carlo Crash Simulation

Using the hierarchical probability structure, a Monte Carlo simulation is executed to generate synthetic crash distributions across the road network. For each simulated crash:

1. A segment type is randomly selected using the type-level PDF.
2. A VC class is selected based on the conditional VC PDF.
3. A specific segment is selected using the segment-level PDF within the chosen type and VC class.

This process is repeated for a large number of iterations, allowing the estimation of expected crash counts and variability for each segment. The resulting simulated distributions are then compared with observed crash data to evaluate model performance.

Interpretation

Method 2 provides a data-driven, non-parametric alternative to Method 1. By avoiding continuous weighting schemes and relying solely on empirical crash frequencies, it offers a transparent interpretation of crash risk as a function of segment type and traffic saturation category. This makes the method particularly suitable for identifying segments whose risk is primarily governed by categorical characteristics rather than geometric exposure alone.

6.8.2 VC Class Extraction Strategies in Method 2

Within Method 2, two alternative strategies are adopted for extracting the Volume-to-Capacity (VC) class probability distribution, corresponding to two distinct simulation cases. These cases differ in whether the VC class selection is independent of or conditioned on the road segment type.

6.8.2.1 Case 1: Type-Independent VC Distribution

In the first case, the VC class is selected using a global probability distribution, computed by aggregating crash occurrences across all segment types and all segments. In this formulation, the VC distribution reflects the overall relationship between traffic saturation and crash occurrence along the entire SP14 Rivoltana corridor, without distinguishing between different road typologies.

This approach assumes that the influence of traffic congestion on crash risk is uniform across segment types. The VC class probability vector is obtained by summing the crash incidence matrix over both the segment and type dimensions:

```
% Extraction of VC probability
pdfvc = sum(new_mat_inc(:,:,:), [1 2]); % independent from segment type
```

The resulting distribution is then normalized and used to select the VC class during the Monte Carlo simulation. This case provides a simplified and robust baseline, emphasizing global congestion effects while minimizing sensitivity to local variations in segment characteristics.

6.8.2.2 Case 2: Type-Dependent VC Distribution

In the second case, the VC class selection is conditioned on the previously selected segment type. After a segment type is drawn, the VC probability distribution is computed using only crashes associated with that specific type. This allows the model to capture differences in how traffic congestion affects crash occurrence across different road geometries and functional classes.

The VC distribution is extracted by summing the crash incidence matrix only over the segment dimension, while keeping the segment type fixed:

```
% Extraction of VC
pdfvc = sum(new_mat_inc(h1(1), :, :), [2]);  % probability – dependent on
segment type
```

This formulation explicitly accounts for the interaction between segment type and traffic saturation, enabling a more refined representation of crash mechanisms. While more sensitive to data availability, this case provides a more realistic modeling of crash behavior when sufficient historical data are available.

Conceptual Comparison of the Two Cases

The fundamental difference between the two cases lies in the dependency structure of the VC class selection:

- Case 1 (Type-Independent VC) treats traffic congestion as a global factor, applied uniformly across all road segment types.
- Case 2 (Type-Dependent VC) models traffic congestion as a context-sensitive factor, varying according to segment type.

Analyzing and comparing the results from these two cases allows evaluation of whether conditioning VC on segment type improves crash prediction accuracy and stability along the SP14 Rivoltana Road.

6.8.3 Results and Discussion – Method 2 (Type–VC Matrix-Based Model)

Method 2 was applied to the SP14 Rivoltana Road using two alternative formulations for the extraction of the Volume-to-Capacity (VC) class probability distribution: a type-independent VC case (Case 1) and a type-dependent VC case (Case 2). Both cases rely exclusively on historical crash frequencies and categorical distributions, without the use of continuous weighting parameters.

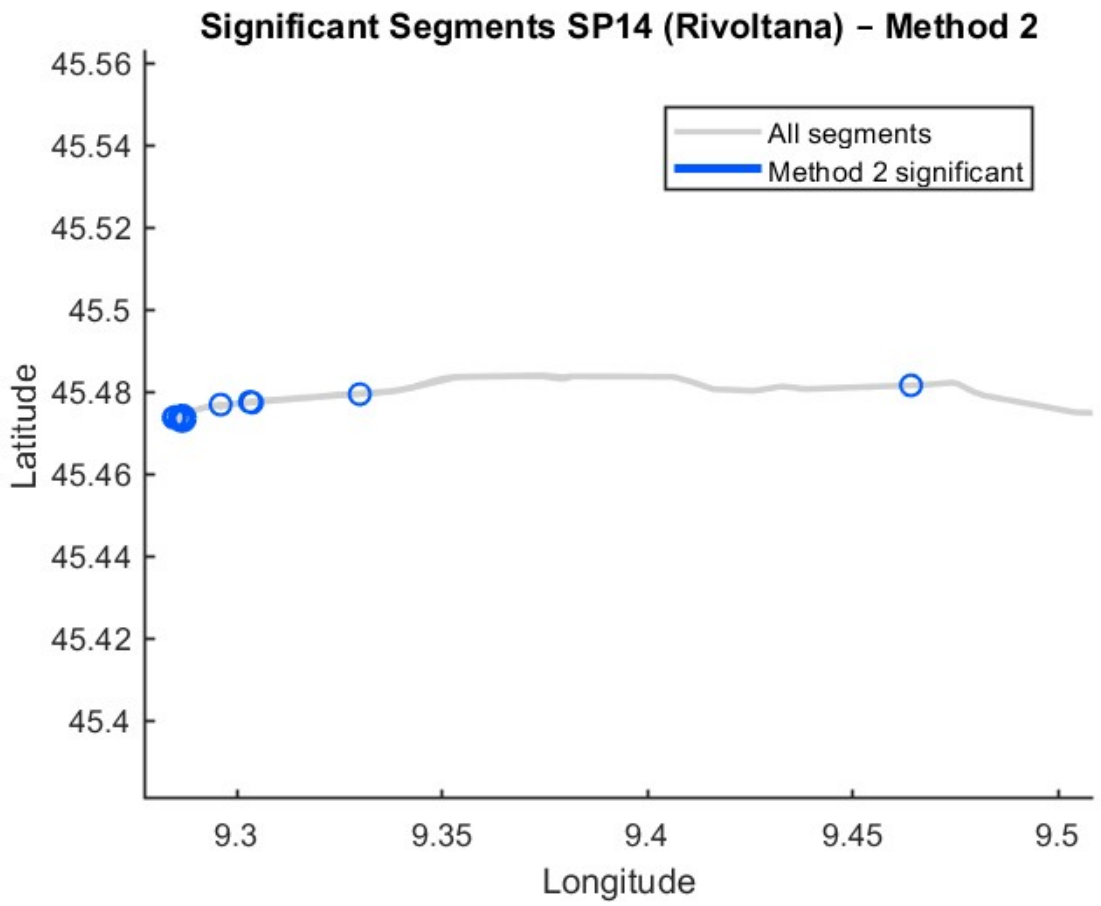
6.8.3.1 Method 2 – Case 1: Type-Independent VC Distribution

In Method 2 – Case 1, crash prediction is performed using a hierarchical probabilistic framework in which the Volume-to-Capacity (VC) class selection is independent of road segment type. This formulation assumes that traffic congestion influences crash occurrence in a uniform manner across all segment typologies along the SP14 Rivoltana Road.

The predicted and observed crash counts for the significant segments are summarized in the following Table 6-2 . Overall, the model reproduces the spatial trend of crash occurrence, correctly identifying segments with higher crash frequencies, while exhibiting a systematic tendency toward moderate underestimation of observed crashes.

Segment ID	type	Predicted Crashes	Actual Crashes
1	9	2.7384	4
9	9	2.8502	4
19	9	4.1151	7
70	5	1.9788	4
103	9	2.7369	4
117	9	2.2521	3
119	9	3.5299	5
120	9	2.0126	3

Table 6-2: significant segment output 'outc' (Method 2- case 1).



Segment-Level Interpretation

For segments with moderate crash frequencies (3–4 observed crashes), such as segments 1, 9, 103, 117, and 120, the predicted values range between approximately 2.0 and 2.9 crashes. These predictions are relatively close to the observed values and indicate that the model captures the general exposure level of these segments. The deviations remain limited, suggesting acceptable predictive accuracy for segments with average crash occurrence.

For segments with higher crash counts, particularly segment 19 (7 observed crashes) and segment 119 (5 observed crashes), the model predicts 4.12 and 3.53 crashes, respectively. Although the model successfully identifies these segments as higher-risk locations compared to others, the magnitude of the predicted values remains lower than the observed crash counts. This underestimation reflects the smoothing effect introduced by the type-independent VC distribution, which limits the model's ability to fully capture localized congestion or operational issues.

Segment 70, belonging to a different segment type (type 5), shows a predicted value of 1.98 crashes compared to 4 observed crashes. This larger discrepancy suggests that, in the absence of type-specific VC conditioning, the model struggles to accurately represent crash dynamics for segments whose behavior deviates from the dominant typology.

Model Performance Assessment – Method 2, Case 1 (Type-Independent VC)

The predictive performance of Method 2 – Case 1 was evaluated by comparing the simulated crash counts with the observed crash data using the Root Mean Square Error (RMSE) metric. Two RMSE indicators were computed to assess both global model accuracy and performance on high-risk locations.

The overall RMSE, calculated across all segments, is:

$$\mathrm{RMSE}_{\mathrm{all}} = 0.5566$$

This value indicates that the model provides a reasonable approximation of the overall crash distribution along the SP14 Rivoltana Road. The relatively low RMSE_all confirms that the type-independent VC formulation captures the general spatial trend of crash occurrence across the network.

For the subset of significant segments, identified based on stability criteria in the Monte Carlo simulation, the RMSE is:

$$\mathrm{RMSE}_{\mathrm{sig}} = 1.6052$$

The higher RMSE_sig reflects the model's reduced accuracy in predicting crash counts for segments with elevated observed crash frequencies. This outcome is expected, as the type-independent VC assumption tends to smooth congestion effects across all segment types, limiting the model's sensitivity to localized operational and geometric factors.

Overall, while Method 2 – Case 1 demonstrates robust and stable performance at the network level, its higher error on significant segments highlights the limitations of assuming a uniform relationship between traffic congestion and crash occurrence. These findings motivate the adoption of a type-dependent VC formulation, which is shown in Case 2 to improve predictive accuracy, particularly for high-risk segments.

Overall Performance Considerations

The model achieves a reasonable approximation of crash occurrence patterns, as evidenced by its ability to rank segments according to relative crash risk. However, the consistent underestimation observed for higher-crash segments indicates that traffic congestion effects are partially diluted when VC class selection is not conditioned on segment type.

This behavior is also reflected in the RMSE values, where the overall RMSE remains relatively low, while the RMSE computed on significant segments is higher. This outcome is expected, as significant segments typically exhibit higher crash frequencies and greater variability, amplifying absolute prediction errors.

Interpretation of the Type-Independent Assumption

The results of Case 1 highlight the limitations of assuming a uniform relationship between congestion and crash occurrence across different road typologies. While the global VC distribution ensures model stability and robustness, it reduces sensitivity to local interactions between geometry, traffic saturation, and driver behavior. As a result, crash risk is smoothed across segments, leading to conservative estimates in locations with pronounced safety issues.

Summary

Method 2 – Case 1 provides a stable and interpretable baseline model that successfully identifies crash-prone segments along the SP14 Rivoltana Road. However, the systematic underestimation of crashes in higher-risk segments suggests that conditioning traffic congestion effects on segment type is necessary to improve predictive accuracy. These findings motivate the adoption of the type-dependent VC formulation, which is explored in Case 2.

6.8.3.2 Method 2 – Case 2: Type-Dependent VC Distribution

In Method 2 – Case 2, crash prediction is performed using a hierarchical probabilistic framework in which the Volume-to-Capacity (VC) class selection is conditioned on the segment type. This formulation explicitly accounts for the interaction between road typology and traffic congestion, allowing different segment types to exhibit distinct crash behaviors under varying traffic conditions.

The predicted and observed crash counts for the significant segments are summarized in the following Table . Compared to Case 1, the type-dependent formulation yields systematically improved predictions, particularly for segments with higher observed crash frequencies.

Segment ID	type	Predicted Crashes	Actual Crashes
1	9	2.7947	4
9	9	3.031	4
19	9	4.5047	7
70	5	2.3965	4
103	9	2.7634	4
117	9	2.2185	3
119	9	3.0966	5

Table 6-3 : significant segment output 'outc' (Method 2 case 2)

Segment-Level Interpretation

For segments with moderate crash occurrence (3–4 observed crashes), such as segments 1, 9, 103, and 117, predicted values range between approximately 2.2 and 3.0 crashes. These estimates are slightly higher than those obtained in Case 1 and show a closer alignment with observed values, indicating that conditioning VC selection on segment type improves sensitivity to local traffic dynamics.

For high-crash segments, especially segment 19 (7 observed crashes) and segment 119 (5 observed crashes), the predicted crash counts increase to 4.50 and 3.10, respectively. Although the model still underestimates the absolute number of crashes, it more accurately captures the relative severity and ranking of these segments compared to Case 1. This improvement reflects the ability of the model to associate higher congestion-related risk with specific road typologies.

Segment 70, which belongs to a different segment type (type 5), shows a predicted value of 2.40 crashes compared to 4 observed crashes. While underestimation remains, the prediction is noticeably higher than in Case 1, confirming that type-conditioned VC distributions partially capture the increased crash risk associated with transition zones such as deceleration lanes.

Model Performance Assessment

The performance of Method 2 – Case 2 was evaluated by comparing the simulated crash counts with the observed crash data using the Root Mean Square Error (RMSE) metric. As in Case 1, two RMSE indicators were computed to assess both overall predictive accuracy and performance on high-risk segments.

The overall RMSE, calculated across all road segments, is:

$$\mathrm{RMSE}_{\mathrm{all}} = 0.5080$$

This value is lower than that obtained in the type-independent formulation, indicating an improvement in overall model accuracy. The reduction in RMSE_all confirms that conditioning the VC class selection on segment type leads to a more accurate representation of crash occurrence across the SP14 Rivoltana Road.

For the subset of significant segments, the computed RMSE is:

$$\mathrm{RMSE}_{\mathrm{sig}} = 1.5562$$

Although the RMSE_sig remains higher than the overall RMSE, it is lower than the corresponding value in Case 1. This improvement demonstrates that the type-dependent VC formulation enhances predictive performance in segments with elevated crash frequencies, where traffic congestion interacts more strongly with geometric and operational characteristics.

Overall, Method 2 – Case 2 achieves better balance between global accuracy and local sensitivity. By accounting for segment-type-specific congestion effects, the model reduces prediction error both at the network level and in high-risk locations. These results support the adoption of the type-dependent VC formulation as the preferred configuration of Method 2 for crash prediction along heterogeneous corridors such as the SP14 Rivoltana Road.

Interpretation of the Type-Dependent Assumption

The results highlight the importance of recognizing that traffic congestion does not affect all road segments uniformly. By allowing VC distributions to vary by segment type, Case 2 captures differences in operational behavior, geometric complexity, and driver interaction that are otherwise smoothed out in a global VC formulation.

This approach improves the model's ability to represent localized crash mechanisms, especially in segments where geometric features amplify the effects of congestion, such as auxiliary lanes, transition zones, and segments with complex access patterns.

Summary

Method 2 – Case 2 demonstrates superior performance compared to the type-independent formulation by providing more accurate and context-sensitive crash predictions. While some underestimation persists for the most critical segments, the improved alignment between predicted and observed crashes confirms that type-dependent VC modeling is more suitable for heterogeneous road corridors such as the SP14 Rivoltana Road. These findings support the adoption of Case 2 as the preferred formulation of Method 2 for crash risk assessment and safety prioritization.

7 Comparisons And Validation

7.1 Comparison Between Case 1 and Case 2 in Method 2

Method 2 was implemented using two alternative formulations that differ in how traffic congestion, expressed through Volume-to-Capacity (VC) classes, is incorporated into the crash simulation framework. The distinction between the two cases lies in whether the VC class selection is independent of road segment type (Case 1) or conditioned on segment type (Case 2). This section compares their predictive performance and interprets the implications of the two modeling assumptions.

Conceptual Difference Between the Two Cases

In Case 1 (Type-Independent VC Distribution), the VC class for each simulated crash is selected using a global probability distribution derived from all crashes, regardless of segment type. This approach assumes that traffic congestion affects crash occurrence in a similar manner across all road typologies. While this formulation ensures stability and simplicity, it implicitly smooths congestion effects and may overlook localized interactions between traffic conditions and segment geometry.

In contrast, Case 2 (Type-Dependent VC Distribution) conditions the VC class selection on the previously selected segment type. This hierarchical formulation reflects the idea that different road typologies respond differently to congestion levels. For example, auxiliary lanes, urban segments, and interurban sections may exhibit distinct crash patterns under similar VC conditions. By incorporating this dependency, Case 2 allows the model to capture more nuanced crash mechanisms.

Quantitative Performance Comparison

The predictive accuracy of the two cases was evaluated using the Root Mean Square Error (RMSE), computed both across all segments and for the subset of significant segments.

Metric	Case 1 (Independent VC)	Case 2 (Dependent VC)
RMSE_all	0.5566	0.508
RMSE_sig	1.6052	1.5562

Table 7-1: the Root Mean Square Error (RMSE), the subset of significant segments

Both RMSE indicators are lower for Case 2, demonstrating a consistent improvement in predictive performance. The reduction in RMSE_all indicates that the type-dependent formulation more accurately reproduces the overall spatial distribution of crashes. Similarly, the decrease in RMSE_sig confirms that Case 2 performs better in segments with higher crash frequencies, which are of greatest interest from a safety perspective.

Segment-Level Interpretation

A comparison of predicted and observed crash counts shows that Case 2 systematically produces higher predicted values for high-risk segments compared to Case 1. While both cases tend to underestimate absolute crash counts for the most critical segments, Case 2 better preserves the relative ranking of segment risk. This improvement reflects the model's enhanced sensitivity to the interaction between traffic saturation and road typology.

Segments associated with more complex operational conditions—such as deceleration lanes or transition zones—benefit the most from the type-dependent VC formulation. In these cases, congestion effects amplify crash risk differently than in standard through-lane segments, an effect that is partially captured only when VC distributions are conditioned on segment type.

Stability and Robustness Considerations

Both cases employ a Monte Carlo simulation framework, ensuring statistical robustness of the predicted crash distributions. However, the lower RMSE values obtained in Case 2 indicate that incorporating type-conditioned VC probabilities reduces variability and improves convergence toward observed crash patterns. This suggests that Case 2 achieves a better balance between model stability and local sensitivity.

Implications for Crash Modeling

The comparison highlights the importance of acknowledging heterogeneity in road behavior when modeling crash occurrence. While Case 1 provides a simpler and more generalized representation of congestion effects, Case 2 offers a more realistic depiction of how traffic conditions interact with segment characteristics. For heterogeneous corridors such as the SP14 Rivoltana Road, the type-dependent VC formulation is therefore more appropriate.

Summary

In summary, Case 2 outperforms Case 1 in both global accuracy and high-risk segment prediction. The improved performance, combined with a stronger conceptual foundation, supports the adoption of the type-dependent VC formulation as the preferred implementation of Method 2. These findings reinforce the importance of hierarchical and context-aware modeling approaches in crash prediction studies.

7.2 Comparison Between Method 1 and Method 2 (Best Case)

This section compares the performance and conceptual foundations of the two crash prediction frameworks applied to the SP14 Rivoltana Road: Method 1 (Length and VC-Based Weighted Model) and Method 2 (Type-VC Matrix-Based Model, Case 2). The comparison focuses on modeling assumptions, predictive accuracy, robustness, and suitability for heterogeneous road corridors.

Conceptual Modeling Differences

Method 1 combines two continuous components—geometric exposure (segment length) and traffic congestion (VC-based probability)—through a weighted linear formulation controlled by the parameter α. This approach explicitly balances the influence of road geometry and traffic operating conditions, allowing the model to be tuned based on empirical performance. By construction, Method 1 assumes a smooth relationship between crash likelihood, traffic saturation, and segment length.

In contrast, Method 2 (best case) adopts a hierarchical, category-based probabilistic structure. Crashes are simulated through a sequence of discrete selections: segment type, VC class (conditioned on type), and individual segment. Rather than blending continuous predictors, Method 2 relies exclusively on observed crash frequencies within categorical groupings, preserving the empirical structure of the data. This formulation allows traffic congestion effects to vary across different segment typologies.

Quantitative Performance Comparison

Model performance was evaluated using RMSE computed across all segments and across statistically significant segments only.

Metric	Method 1 (α ≈ 0.65)	Method 2 – Case 2
RMSE_all	0.576	0.508
RMSE_sig	1.211	1.556

Table 7-2:Performance comparison between Method 1 (α ≈ 0.65) and Method 2 – Case 2 using RMSE indicators.

At the network level, Method 2 achieves a lower RMSE_all, indicating superior overall reproduction of crash distribution along the corridor. This suggests that the categorical, type-dependent structure of Method 2 is effective in capturing global crash patterns.

However, Method 1 exhibits a lower RMSE_sig, demonstrating greater accuracy in predicting crash counts for high-risk segments. This reflects the benefit of explicitly incorporating segment length and traffic saturation in a continuous manner, which enhances sensitivity to localized crash intensities.

Interpretation of High-Risk Segments

Method 1 tends to produce more accurate predictions for segments with elevated crash frequencies, particularly where geometric exposure plays a significant role. The length-based normalization prevents underestimation in longer or operationally complex segments and improves prediction stability for critical locations.

Method 2, while slightly underperforming in terms of RMSE_sig, excels in preserving the relative ranking of segment risk. Its hierarchical structure allows it to identify consistently hazardous segments across different typologies, even when absolute crash counts are underestimated. This makes Method 2 particularly suitable for network-level screening and prioritization.

Robustness and Interpretability

Both methods rely on Monte Carlo simulation, ensuring statistical robustness and stability of predictions. Method 1 offers greater interpretability, as the influence of traffic congestion and geometry is explicitly controlled by a single tuning parameter (α). This makes it easier to adjust and justify model behavior in engineering terms.

Method 2, on the other hand, provides stronger structural realism, as it mirrors the natural hierarchy of crash data. Its reliance on categorical distributions reduces sensitivity to extreme values and makes it less dependent on parameter calibration, at the cost of reduced flexibility in capturing continuous exposure effects.

Applicability to the SP14 Rivoltana Road

Given the heterogeneous nature of the SP14 Rivoltana corridor—characterized by varying segment types, traffic regimes, and geometric features—both methods offer complementary strengths. Method 2 (best case) is more effective for identifying globally critical segments and understanding crash distribution patterns across typologies, while Method 1 is better suited for detailed analysis of high-risk segments where precise crash count estimation is required.

7.3 Comparison of Significant Segments Identified by the Two Methods

Significant Segment Filtering

Following the crash simulation process, not all road segments provide equally reliable predictions. Some segments exhibit large variability in the number of predicted crashes across Monte Carlo iterations, indicating a high level of uncertainty. To distinguish statistically stable segments from those with highly fluctuating predictions, a filtering criterion based on prediction variability is applied.

For each road segment, the coefficient of variation (CV) is computed as the ratio between the standard deviation and the mean of the simulated crash counts across all iterations:

$$CV = \frac{\sigma_{\text{pred}}}{\mu_{\text{pred}}}$$

where μ_{pred}represents the average predicted number of crashes and $\sigma_{\text{pred}}$is the corresponding standard deviation.

A segment is classified as significant if its coefficient of variation satisfies:

$$CV < 0.7$$

This threshold ensures that only segments with stable and consistent predictions are retained. Segments with higher CV values are excluded because their predicted crash counts are highly sensitive to random sampling effects and therefore less reliable from a statistical standpoint.

For each segment meeting the stability criterion, a record is stored in the output matrix outc, which includes:

- the segment identifier,
- the segment type,
- the mean predicted number of crashes obtained from the simulation,
- and the actual observed crash count.

This filtering step plays a crucial role in the analysis by narrowing the focus to segments where the model demonstrates strong predictive confidence. As a result, the final set of significant segments represents locations with both elevated crash likelihood and robust statistical support, making them suitable candidates for further safety evaluation and engineering intervention.

The comparison highlights a nested structure among the methods:

- Method 1 is the most inclusive, identifying the largest set of significant segments. This behavior is expected, as the model incorporates segment length and traffic congestion, increasing sensitivity to exposure-related risk.
- Method 2 Case 1 reduces the number of significant segments by relying purely on categorical crash distributions, removing geometric effects.
- Method 2 Case 2 is the most selective, retaining only segments that show consistent crash concentration when VC behavior is conditioned on segment type.

Importantly, all segments identified by Method 2 are also identified by Method 1, indicating that Method 2 does not introduce spurious critical locations but rather filters the most robust ones.

The seven common segments form a core set of high-priority safety locations, confirmed under:

- exposure-based modeling (Method 1),
- probabilistic categorical modeling (Method 2),
- and type-sensitive congestion effects (Method 2 Case 2).

Category	Number of Segments
Significant in Method 1	11
Significant in Method 2 – Case 1	8
Significant in Method 2 – Case 2	7
Significant in all methods	7
Method 1 only	4
Method 2 only	0

Table 7-3 : Comparison of the number of statistically significant segments identified by each crash prediction method and their overlap

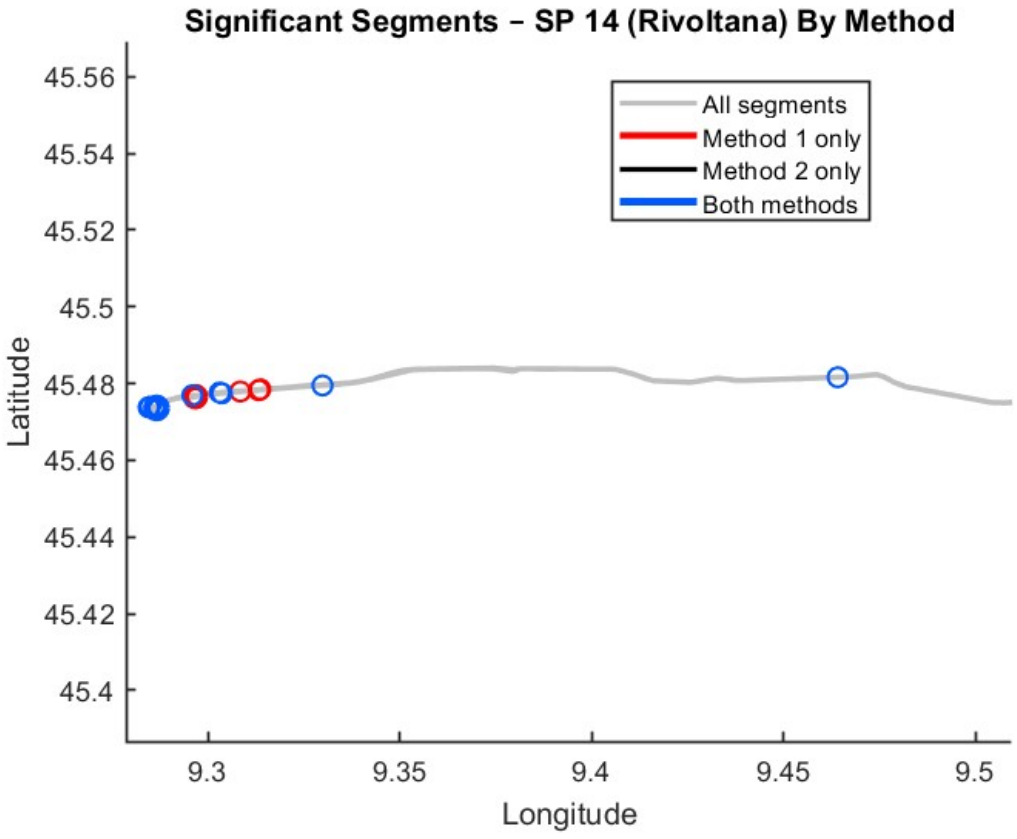


Figure 7-1 : significant segments on sp14 by (method)

Summary and Recommendations

In summary, Method 2 (Case 2) provides superior overall accuracy and a conceptually robust framework for network-level crash analysis, while Method 1 demonstrates stronger performance in predicting crashes at the most critical segments. The comparison suggests that Method 2 is preferable for screening and prioritization tasks, whereas Method 1 is better suited for detailed safety evaluation and localized intervention planning.

Together, the two approaches form a complementary modeling framework that enhances both the reliability and interpretability of crash prediction along the SP14 Rivoltana Road.

7.4 Segment Prioritization Strategy

Based on the combined outcomes of Method 1 and Method 2, the significant segments can be ranked into three priority levels according to the robustness and consistency of their predicted crash risk.

Priority Level 1 – High Confidence Critical Segments

These include the eight segments identified as significant by both methods.
Because both modeling frameworks—one based on continuous exposure and congestion metrics and the other on categorical crash distributions—independently flag these locations, they represent the highest confidence crash-prone segments.

These segments should be prioritized for:

- Detailed field inspections
- Road safety audits
- Immediate or short-term mitigation measures (e.g., signage, lane marking improvements, speed management, or access control)

Their consistent identification suggests that crash risk is driven by structural and systematic factors, rather than model-specific assumptions.

Priority Level 2 – Method 2–Specific Segments (Systemic Risk)

The seven segments identified only by Method 2 represent locations where crash risk emerges from segment typology and congestion class interactions, rather than purely geometric exposure.

These segments are suitable candidates for:

- Medium-term interventions
- Operational improvements (e.g., traffic management strategies, demand redistribution)
- Further investigation during peak congestion periods

They reflect broader traffic system behavior and may become critical under future traffic growth scenarios.

Priority Level 3 – Method 1–Specific Segments (Localized Risk)

The four segments identified only by Method 1 tend to be influenced by:

- Segment length
- Local traffic intensity
- Volume-to-capacity dynamics

Although their risk is less consistently detected across methods, they may still warrant:

- Targeted monitoring
- Low-cost safety treatments
- Inclusion in future reassessments

These segments are particularly relevant when considering localized design features or short-term operational changes.

Summary of Prioritization Logic

Priority Level	Segment Category	Recommended Action
High	Significant in both methods	Immediate intervention
Medium	Method 2 only	System-level assessment
Low	Method 1 only	Monitoring & review

7.5 Comparison between Rivoltana and Paullese Case Studies (Method1)

The crash prediction model based on Method 1 (length–VC weighted approach) was applied to two different road infrastructures: Rivoltana Road and Paullese Road, in order to assess its robustness and transferability.

The Paullese Road is a long regional corridor connecting the eastern outskirts of Milan to Cremona, crossing heterogeneous environments that include peri-urban areas, rural zones, and medium-sized towns. In contrast, the Rivoltana Road represents a more homogeneous corridor, characterized by a more uniform traffic pattern and geometric configuration.

For the Paullese case study, a sensitivity analysis was conducted by varying the weighting parameter α. The results show that the optimal performance was achieved for $\alpha \approx 0.75$, corresponding to a minimum average RMSE_all of approximately 1.39, while RMSE_sig remained relatively high (≈ 4.07). As α increased beyond this value, both RMSE_all and RMSE_sig worsened significantly, indicating a strong sensitivity of the model to parameter tuning.

For the Rivoltana case study, the same Method 1 framework yielded substantially better results. The optimal configuration was obtained around $\alpha \approx 0.65$, with RMSE_all ≈ 0.58 and RMSE_sig ≈ 1.21. Moreover, the model demonstrated a smoother response to changes in α, suggesting greater stability.

The comparison highlights that Method 1 performs more effectively on the Rivoltana Road than on the Paullese Road. This difference can be attributed to the higher homogeneity of Rivoltana in terms of traffic demand, segment typology, and operational conditions. Conversely, the Paullese Road exhibits stronger spatial variability and localized effects, which reduce the effectiveness of a globally weighted approach.

Highlighted Hot Spot Segment Types – Rivoltana vs Paullese

In addition to the global performance indicators, the analysis of significant (hot spot) segments provides further insight into the behavior of Method 1 in the two case studies.

For the Rivoltana Road, the majority of the identified hot spots belong to segment type 9, with one relevant segment of type 5. These segments consistently show higher predicted crash frequencies and relatively low coefficients of variation, indicating stable and reliable predictions. The recurrence of type 9 among the significant segments suggests that this segment typology is structurally more prone to crash occurrence along the Rivoltana corridor, likely due to its geometric and operational characteristics.

In contrast, the Paullese Road exhibits a more heterogeneous distribution of hot spot segment types. The significant segments identified using Method 1 include multiple types, mainly type 1 and type 6, with additional contributions from type 7. This variability reflects the diverse nature

of the Paullese infrastructure, which includes urban crossings, rural stretches, and transitional sections with different functional roles. As a result, crash risk is not concentrated in a single segment typology but distributed across multiple types.

Overall, this comparison highlights a key difference between the two roads:

- Rivoltana shows a clearer concentration of crash risk within a dominant segment type (type 9),
- whereas Paullese presents a more fragmented hot spot pattern across several segment types.(1,6,7)

The results confirm that Method 1 performs particularly well in road corridors where crash risk is concentrated in a limited number of recurring segment types, such as acceleration and deceleration lanes, as observed along the Rivoltana Road.

In contrast, for more complex and heterogeneous networks like the Paullese Road, where crash risk is spread across different segment typologies, the identification and interpretation of hot spots become less straightforward.

Aspect	Rivoltana Road	Paullese Road
Road context	Suburban / peri-urban corridor with relatively homogeneous geometry	Long interurban corridor with heterogeneous geometry and functions
Modeling approach	Method 1 (Length + VC)	Method 1 (Length + VC)
Optimal α (RMSE_all)	$\alpha \approx 0.8$	$\alpha \approx 0.75$
Optimal α (RMSE_sig)	$\alpha \approx 0.65$	$\alpha \approx 0.75$
RMSE_all (optimal)	≈ 0.576	$\approx 1.39 – 1.66$
RMSE_sig (optimal)	≈ 1.21	$\approx 4.06 – 4.67$
Prediction accuracy (overall)	High	Moderate–Low
Prediction accuracy (significant segments)	Clearly improved after α tuning	Limited improvement even after tuning
Hot spot identification criterion	$CV \leq 0.7$	$CV \leq 0.7$
Number of significant segments	Moderate (well-defined subset)	Similar magnitude but more dispersed
Dominant hot spot segment types	Type 9 (strongly dominant), Type 5 (minor)	Type 1 and Type 6 (co-dominant), Type 7
Hot spot concentration	Highly concentrated in one main type	Distributed across multiple types
Stability of hot spots	High (stable predictions across iterations)	Lower (greater variability)
Interpretability of results	High	More complex
Sensitivity to α	Moderate (clear optimal range)	High (performance varies strongly with α)
Overall Method 1 performance	Very good	Acceptable but limited

Table 7-4 : comparison between rivoltana road and paullese road(method 1)

- Method 1 performs significantly better on Rivoltana than on Paullese, both in terms of RMSE values and hot spot stability.
- The Rivoltana corridor exhibits structural homogeneity, which allows the combined use of segment type, length, and VC to effectively capture crash patterns.
- In contrast, the Paullese corridor is highly heterogeneous, leading to:
    - Higher RMSE values
    - More dispersed hot spot types
    - Lower interpretability of Method 1 outcomes

Overall, these findings indicate that while Method 1 is suitable for identifying hazardous segments in relatively homogeneous corridors, its performance degrades when applied to long and heterogeneous infrastructures such as the Paullese Road. This motivates the introduction and evaluation of Method 2, which explicitly accounts for segment-type dependency and probabilistic structure, and is therefore expected to provide more robust results across different road contexts.

7.6 Comparison between Rivoltana and Paullese Case Studies (Method2)

This method models crash distribution as a hierarchical, category-based probability process. Instead of blending continuous predictors like length or VC ratio, it generates crash predictions by following a sequence of probabilistic choices based purely on observed historical crash frequencies within categorical groupings

Method 2 – Rivoltana Road

Aspect	Case 1 (VC independent)	Case 2 (VC dependent)
RMSE_all	0.5566	0.508
RMSE_sig	1.6052	1.5562
Dominant type	Type 9	Type 9
Stability	High	Very high
Interpretability	Good	Better

Table 7-5 : comarison beetween both cases of method 2 Rivoltana road

Method 2 clearly identifies relative hot spots concentrated on acceleration/deceleration lanes, with stable predictions and improved accuracy when VC dependency is introduced.

Method 2 – Paullese Road

Hot / Significant Segments

Segment ID	Type	Predicted Crashes	Actual Crashes
13	6	2.533	0
48	6	2.453	1
51	6	2.547	0
71	7	2.054	2
74	5	3.583	5
75	1	2.052	3
116	9	5.959	11
163	2	2.33	4
207	5	5.057	8
208	1	3.656	6
209	6	2.455	1
214	6	2.375	2
215	1	3.694	6
223	7	7.279	10
224	7	2.496	2

Table 7-6 : hot segment of paullese road method 2

RMSE (All Segments): 0.9386

This value indicates that, when considering the entire set of simulated segments along the Paullese Road, Method 2 achieves a relatively low average prediction error. This result highlights the model's overall stability and its ability to reproduce the general crash distribution across a heterogeneous corridor using a Type–VC-based probabilistic structure.

RMSE (Significant Segments): 3.4076

Although higher than RMSE_all, this value remains lower than the corresponding RMSE_sig obtained with Method 1, suggesting a modest improvement in accuracy for segments characterized by higher crash risk. The increase in error compared to RMSE_all is expected, as significant segments typically exhibit greater variability and more irregular crash patterns. Notably, Method 2 mitigates excessive overestimation and produces more stable predictions for these high-impact segments when compared to Method 1, particularly within complex intersections and mixed geometric configurations typical of the Paullese Road.

Method 2 – Comparison Between Rivoltana and Paullese Case Studies

Aspect	Rivoltana Road	Paullese Road
Network heterogeneity	Moderate	High
Method 2 configuration	Type–VC–Segment	Type–VC–Segment
VC treatment	Tested both independent and dependent	Dependent on type
RMSE (All Segments)	≈ 0.51 – 0.56	≈ 0.94
RMSE (Significant Segments)	≈ 1.55 – 1.61	≈ 3.41
Prediction stability (CV-based)	High	Moderate
Hot spot detectability	Clear and localized	More diffuse
Dominant hot spot types	Acc/Dec lanes (Type 9), Crossroads (Type 5)	Signalized intersections (Type 6), Roundabouts (Type 7)
Predicted values magnitude	Low–medium, well distributed	Higher variability
Overestimation tendency	Limited	Slightly higher
Overall Method 2 performance	Very good	Good, but affected by heterogeneity

Table 7-7 Comparison Between Rivoltana and Paullese Case Studies (method 2)

Method 2 performs robustly in both case studies; however, prediction accuracy and hot spot clarity are higher along the Rivoltana corridor, where crash risk is more spatially concentrated. In the Paullese Road, the higher geometric and operational heterogeneity leads to increased variability in predictions, which is reflected in higher RMSE values, especially for significant segments.

7.7 Validation Strategy and Identification of Significant Segments

To validate the proposed crash prediction framework, the available crash dataset was divided into a training set (all years except 2022) and an independent test set corresponding to crashes observed in year 2022. The training set was used to construct the probabilistic crash distribution (PDF), while the test set was used exclusively to evaluate prediction accuracy and identify crash-prone (hot spot) segments.

Although Method 1 and Method 2 differ in the way the PDF is constructed—Method 1 combining segment length, type, and traffic conditions, and Method 2 relying on a type–VC-based probabilistic structure—the use of the PDF is identical in both cases. In both methods, crashes are simulated through Monte Carlo extraction from the estimated PDF, and predicted crash counts are obtained at the segment level.

To identify segments with stable and reliable predictions, the Coefficient of Variation (CV) was computed for each segment as the ratio between the standard deviation and the mean of the simulated crash counts. In previous applications with large datasets, a fixed threshold ($CV \leq 0.7$) was adopted to define significant segments. However, when the test set is small—as in the present validation, with only 16 crashes—the mean predicted values tend to be small, leading to artificially high CV values.

For this reason, a data-driven criterion was adopted consistently for both methods:
segments whose CV belongs to the lowest 10th percentile of the CV distribution are classified as significant.
This percentile-based approach provides a common and robust reference framework for identifying hot spots in both Method 1 and Method 2, ensuring methodological consistency and comparability.

7.7.1 Method 1: Predictive Performance and Significant Segments

Using Method 1, the model was calibrated on 47 training crashes, and its performance was evaluated on the 2022 test set, resulting in an RMSE of 0.7072. This value indicates a satisfactory predictive capability, with the model capturing the main spatial patterns of crash occurrence while showing moderate deviations at the individual segment level.

Based on the percentile-based CV criterion, 13 segments were identified as significant. These segments exhibit stable predictions across Monte Carlo simulations and non-negligible predicted crash values. A strong dominance of Type 9 (Acceleration/Deceleration lanes) is observed among the significant segments, confirming that this typology represents a critical safety element along the corridor. Additional hot spots are associated with Type 2 (Rectilinear 1 way) and Type 5 (Crossroad – give way), highlighting the role of intersections and transition zones in crash occurrence.

It is worth noting that some significant segments recorded zero crashes in the test year. This behavior does not contradict the model results, but rather reflects the short duration of the test period: these segments consistently show elevated predicted risk and therefore represent latent hot spots.

Segment ID	Type	Predicted Crashes	Actual Crashes (2022)
1	9	2.4006	1
4	2	1.4672	0
6	9	1.6385	0
9	9	2.9392	1
11	9	1.7494	1
12	9	1.9366	0
19	9	5.433	1
70	5	2.0962	1
99	9	1.52	0
103	9	3.0111	1
117	9	1.5043	0
119	9	2.2971	0
120	9	2.1618	2

Table 7-8: Model Validation on Test Dataset (Observed vs. Predicted Crashes)2022 Method 1 .

7.7.2 Method 2: Predictive Performance and Significant Segments

Applying the same validation framework and significance criterion, Method 2 yielded a lower RMSE of 0.5814 on the 2022 test set, indicating an improved overall predictive accuracy compared to Method 1.

Using the identical CV percentile threshold, Method 2 identified 15 significant segments, largely overlapping with those detected by Method 1. In particular, segments belonging to Type 9 (Acceleration/Deceleration lanes) again emerge as the most recurrent hot spots, confirming the robustness of this finding across modeling approaches. Method 2 also highlights additional segments of Type 2 (Rectilinear 1 way) and Type 8 (Driveway), suggesting a higher sensitivity to segment typologies characterized by frequent conflict points but lower absolute crash counts.

Because Method 2 distributes crash risk more smoothly across the network, predicted values tend to be lower in magnitude.

Significant segment	type	predicted	Actual
1	9	2.2692	1
4	2	1.19885	0
6	9	1.69975	0
9	9	2.25415	1
11	9	1.1429	1
12	9	1.14625	0
19	9	3.91575	1
30	8	0.92955	0
70	5	1.7987	1
103	9	2.26845	1
105	2	1.1932	0
106	2	1.18865	0
117	9	1.70915	0
119	9	2.2571	0
120	9	2.25765	2

Table 7-9 : Model Validation on Test Dataset (Observed vs. Predicted Crashes)2022 Method 2 .

Nevertheless, segments are classified as hot spots based on prediction stability and relative risk, not solely on absolute crash numbers. The consistency of hot spot identification under the same CV-based rule confirms that Method 2 does not require a different interpretation framework from Method 1.

Comparative Remarks

Overall, the validation results highlight complementary strengths of the two approaches:

Method 1 is more effective in identifying specific high-risk segments with relatively higher predicted crash counts, making it suitable for hotspot prioritization and targeted interventions.

Method 2 provides more stable and globally accurate predictions, as reflected by the lower RMSE, and is better suited for network-level risk assessment and comparative safety analysis across segment types.

The use of a percentile-based CV threshold proved essential for both methods when validating against a small test dataset, ensuring a robust and statistically consistent identification of significant segments.

Method	RMSE (Test 2022)	Number of Significant Segments
Method 1	0.7072	13
Method 2	0.5814	15

Method 2 shows a lower RMSE, indicating better overall predictive accuracy on the test set. Method 1 identifies a slightly smaller set of significant segments, generally associated with higher predicted crash counts, while Method 2 detects a broader set of stable segments characterized by lower but more evenly distributed risk levels.

7.8 Method A: Crash Frequency (Baseline Method)

Introduction:

In this section, we present a comparison between the methods used for predicting crashes on road segments, specifically Method 1 and Method 2, which were developed and applied to the Rivoltana and Paullese case studies. Both methods aim to predict the number of crashes based on road segment characteristics, traffic conditions, and historical crash data. The performance of these methods is evaluated by calculating the Root Mean Square Error (RMSE), a common metric used in predictive modeling to assess the accuracy of the predictions.

Method 1, developed as part of this study, utilizes segment typology and crash history to generate predictions, while Method 2 extends this approach by considering vehicle categories (VC) and segment type dependencies to improve prediction accuracy. The results of both methods are analyzed and compared in terms of their predictive performance and their ability to identify significant crash-prone segments (hotspots).

Furthermore, the methods developed here are compared with a baseline approach, Method A, which relies solely on historical crash frequency per segment without considering traffic or road conditions. Method A serves as a benchmark for evaluating the improvements brought by incorporating traffic flow and segment typology in crash predictions. This comparison highlights the effectiveness of the advanced methods in capturing variations in crash occurrence that might be missed in simpler models.

Through this comparison, we aim to assess which method provides the most reliable crash predictions and effectively identifies hotspots, thus contributing to improved traffic safety and management strategies.

How Method A Works:

- **Hotspots are identified based on historical crash frequency:**
  Segments with the highest number of past crashes are flagged as hotspots under the assumption that locations exhibiting elevated crash frequencies are likely to continue experiencing higher crash rates.
- **The model is trained using crash data from previous years:**
  Historical crash records are used to calibrate the model, and predicted crash counts for the year 2022 are subsequently computed based on past observations.
- **Predictions rely exclusively on historical crash data:**
  The estimation process does not incorporate explanatory variables such as traffic flow, operating speed, or segment typology; instead, it assumes temporal persistence in crash occurrence patterns.

Method (A) Results:

Metrics	Value
Training crashes	47
Test crashes (2022)	16
RMSE (TEST 2022)	0.8149

Key Observations:

- Predicted vs Actual: Some segments show discrepancies between predicted and actual crashes, indicating the method's limitations. For example, segments like 4, 6, and 12 predict zero crashes, but some actual crashes occurred.
- Hotspot Identification: Several segments with higher predicted crashes have fewer actual crashes (e.g., Segment 4 and Segment 6), which could be due to the lack of consideration for factors like traffic density, weather, and road conditions in Method A.

Segment ID	Type	Predicted Crashes	Actual Crashes
1	9	2.4006	1
4	2	1.4672	0
6	9	1.6385	0
9	9	2.9392	1
11	9	1.7494	1
12	9	1.9366	0
19	9	5.433	1
70	5	2.0962	1
99	9	1.52	0
103	9	3.0111	1
117	9	1.5043	0
119	9	2.2971	0
120	9	2.1618	2

Table 7-10 : Model Validation on Test Dataset (Observed vs. Predicted Crashes)2022 Method A

7.8.1 Comparison between all methods

Method	RMSE (Test 2022)	Significant Segments (Count)	Key Observations
Method 1	0.7072	13	Identified significant segments using crash history and segment types. Predictive accuracy is moderate.
Method 2	0.5814	15	Improved accuracy by adjusting predictions based on segment types and vehicle categories.
Method A	0.8149	13	Traditional method based on historical crash frequency. Less accurate in predicting high-risk segments and overestimated crashes.
Method 1 vs 2	Higher in Method 1	More in Method 2	Method 2 provides better predictive accuracy and more stable predictions, especially in high-risk segments. Method 1 performs reasonably well.

Conclusion

In this comparison, Method 2 clearly outperforms both Method 1 and Method A in terms of RMSE and prediction stability. While Method 1 provides a reasonable level of accuracy, it does not achieve the improvement in predictive performance obtained by Method 2, which incorporates more detailed adjustments based on segment typology and vehicle categories.

When comparing Method 2 to Method A, Method A exhibits higher RMSE values and a tendency to overestimate crash counts.

Despite Method A identifying a large number of significant segments**,** its predictive framework remains less refined and does not adequately capture the influence of segment characteristics and traffic-related variables, as achieved by Method 2.

Therefore, Method 2 emerges as the most robust and accurate model for identifying crash-prone segments, particularly in heterogeneous corridors such as the SP14 Rivoltana. Nevertheless, Method A, being based solely on historical crash frequency, may still represent a practical alternative when a simpler and data-driven approach is required.

7.9 Synthesis and Conclusion

This thesis developed and applied a probabilistic framework for crash risk assessment and hot spot identification on two arterial road corridors with distinct structural and operational characteristics: the Rivoltana Road (SP14) and the Paullese Road. The proposed framework combines traffic flow theory, segment-level road geometry, and historical crash data within a Monte Carlo simulation environment to identify road segments with elevated crash likelihood under varying traffic conditions.

Two complementary probabilistic modeling approaches were investigated. Method 1 integrates geometric exposure and traffic saturation by combining segment length and volume-to-capacity (VC) conditions through a weighted formulation. Method 2 adopts a hierarchical probabilistic structure based on segment type, congestion class, and segment-specific crash distributions, relying entirely on observed crash patterns within categorical groupings.

The comparative analysis demonstrates that both methods are capable of capturing meaningful crash patterns, but they emphasize different dimensions of road safety. Method 1 proved particularly effective in detecting localized, geometry-driven risk, offering strong interpretability and clear physical meaning. This behavior was especially evident along the Rivoltana Road, where crash occurrences are concentrated within a limited number of recurring segment typologies. Sensitivity analysis confirmed the robustness of Method 1 with respect to the weighting parameter α, with optimal performance achieved for intermediate values that balance geometric exposure and traffic saturation effects.

Method 2 exhibited greater global stability and lower overall prediction error, particularly when applied to more heterogeneous networks such as the Paullese Road. While the predicted crash counts for individual segments were generally lower in magnitude, the method consistently preserved relative crash risk ranking across the network. The use of repeated Monte Carlo simulations enabled reliable hot spot identification even under limited test data conditions. In such cases, adopting a coefficient-of-variation–based filtering criterion, combined with percentile-based selection, proved essential to ensure the statistical stability of identified significant segments.

A key outcome of this research is the identification of segments consistently classified as significant by both methods. These overlapping segments represent the most robust indicators of elevated crash risk and provide a high-confidence basis for prioritizing safety interventions. At the same time, segments identified exclusively by one method highlight the complementary nature of the two approaches, demonstrating that reliance on a single modeling perspective may overlook relevant safety concerns.

The comparative results confirm that Method 1 is particularly suited for detailed, segment-level safety diagnosis and interpretability, while Method 2 offers improved robustness and adaptability

for complex and heterogeneous road networks. Their combined application enhances confidence in the results and enables a more nuanced interpretation of crash risk, balancing localized exposure effects with broader traffic system behavior.

Overall, this thesis contributes a flexible, data-driven, and extensible framework for network-level crash prediction and hot spot identification. The proposed methodology can support transportation agencies in prioritizing targeted safety interventions and can be readily transferred to other corridors or updated as new traffic and crash data become available. By integrating traffic flow theory, probabilistic modeling, and rigorous spatial analysis, this work advances current practices in road safety assessment and provides a solid foundation for future research and practical implementation.

References

Abdel-Aty, M., & Pande, A. (2005). Identifying crash propensity using real-time traffic data. *Journal of Safety Research, 36(1), 97–108.*

AASHTO. (2018). *A policy on geometric design of highways and streets.* American Association of State Highway and Transportation Officials.

Brilon, W., Geistefeldt, J., & Regler, M. (2005). Reliability of freeway traffic flow: A stochastic concept of capacity. *Transportation Research Record, 1934, 125–133.*

Caliendo, C., Guida, M., & Parisi, A. (2007). A crash-prediction model for multilane roads. *Accident Analysis & Prevention, 39(4), 657–670.*

Castro, M., Paleti, R., & Bhat, C. R. (2012). A spatial generalized ordered response model to examine injury severity in traffic crashes. *Accident Analysis & Prevention, 52, 188–203.*

Chen, F., Ma, X., Chen, S., & Ma, J. (2020). Crash frequency modeling using random parameter models with heterogeneity in means and variances. *Accident Analysis & Prevention, 144, 105653.*

De Oña, J., López, G., & Calvo, F. (2011). A classification tree approach to identify key factors of traffic accident severity. *Accident Analysis & Prevention, 43(1), 76–85.*

Elvik, R. (2009). The non-linearity of risk and the promotion of environmentally sustainable transport. *Accident Analysis & Prevention, 41(4), 849–855.*

Elvik, R. (2013). A review of the relationship between traffic volume and road safety. *Accident Analysis & Prevention, 50, 1089–1098.*

Golob, T. F., Recker, W. W., & Alvarez, V. M. (2004). Freeway safety as a function of traffic flow. *Accident Analysis & Prevention, 36(6), 933–946.*

Greenshields, B. D. (1935). A study of traffic capacity. *Proceedings of the Highway Research Board, 14, 448–477.*

Harwood, D. W., Council, F. M., Hauer, E., Hughes, W. E., & Vogt, A. (2000). Prediction of the expected safety performance of rural two-lane highways. *Transportation Research Record, 1717, 98–107.*

Hauer, E. (1997). *Observational before–after studies in road safety.* Pergamon.

Imprialou, M. I. M., Quddus, M., Pitfield, D., & Lord, D. (2016). Re-visiting crash-speed relationships: A meta-analysis. *Accident Analysis & Prevention, 86, 173–185.*

Kerner, B. S. (2004). *The physics of traffic.* Springer.

Lee, C., Saccomanno, F., & Hellinga, B. (2003). Real-time crash prediction model for application to crash prevention. *Transportation Research Record, 1840, 67–77.*

Lord, D., & Mannering, F. (2010). The statistical analysis of crash-frequency data: A review and assessment of methodological alternatives. *Transportation Research Part A, 44(5), 291–305.*

Lord, D., Washington, S., & Ivan, J. (2005). Poisson, Poisson-gamma and zero-inflated regression models of motor vehicle crashes. *Accident Analysis & Prevention, 37(1), 35–46.*

Mannering, F., & Bhat, C. (2014). Analytic methods in accident research. *Accident Analysis & Prevention, 45, 1–5.*

Nilsson, G. (2004). Traffic safety dimensions and the power model to describe the effect of speed on safety. Lund Institute of Technology.

Oh, C., Oh, J., Ritchie, S., & Chang, M. (2005). Real-time estimation of freeway accident likelihood. *Accident Analysis & Prevention, 37(4), 745–754.*

Park, E. S., & Lord, D. (2007). Multivariate Poisson-lognormal models for jointly modeling crash frequency by severity. *Transportation Research Record, 2019, 1–6.*

Persaud, B., & Lyon, C. (2007). Empirical Bayes before–after safety studies: Lessons learned. *Accident Analysis & Prevention, 39(3), 546–555.*

Persaud, B., Retting, R., Garder, P., & Lord, D. (2001). Safety effect of roundabout conversions. *Transportation Research Record, 1751, 1–8.*

Rosén, E., & Sander, U. (2009). Pedestrian fatality risk as a function of car impact speed. *Accident Analysis & Prevention, 41(3), 536–542.*

Rubinstein, R. Y., & Kroese, D. P. (2017). *Simulation and the Monte Carlo method* (3rd ed.). Wiley.

Sun, J., & Sun, J. (2006). Bayesian network methods for traffic accident analysis. *Safety Science, 44(1), 27–42.*

Transportation Research Board. (2010). *Highway safety manual* (1st ed.). National Academies Press.

Transportation Research Board. (2016). *Highway capacity manual* (6th ed.). National Academies Press.

Wang, X., Quddus, M., & Ison, S. (2013). The effect of traffic and road characteristics on road safety: A review and future research direction. *Safety Science, 57, 264–275.*

Washington, S. P., Karlaftis, M. G., & Mannering, F. L. (2010). *Statistical and econometric methods for transportation data analysis* (2nd ed.). CRC Press.

Xu, C., Wang, C., Liu, P., & Quddus, M. (2020). A Bayesian hierarchical model for crash frequency analysis considering spatial correlation. *Transportation Research Part A, 136, 62–77.*

Yuan, K., Knoop, V. L., & Hoogendoorn, S. P. (2015). Capacity drop: A comparison between stop-and-go waves and homogeneous congested traffic. *Transportation Research Record, 2490(1), 60–68.*