

1 **SAFETY-INTEGRATED CALIBRATION AND VALIDATION OF SIMULATION**
2 **MODELS FOR COMPLEX INNOVATIVE INTERSECTIONS USING BIVARIATE**
3 **EXTREME VALUE THEORY**

4

5

6

7 **Kiran Neupane** (Corresponding Author)

8 Department of Civil and Architectural Engineering and Construction Management

9 University of Wyoming, Laramie, WY 82071

10 kneupane@uwyo.edu

11

12 **Ramesh Dhimal**

13 Department of Civil and Architectural Engineering and Construction Management

14 University of Wyoming, Laramie, WY 82071

15 rdhimal@uwyo.edu

16

17 **Yu (Fred) Song**

18 Department of Civil and Architectural Engineering and Construction Management

19 University of Wyoming, Laramie, WY 82071

20 ysong@uwyo.edu

21

22

23 Word Count: 5157 words + 5 table(s) \times 250 = 6407 words

24

25

26

27

28

29

30 Submission Date: July 30, 2025

1 ABSTRACT

2 This paper presents a safety-integrated calibration and validation framework for microscopic traffic
3 simulation models applied to two unconventional intersection designs: Diverging Diamond Inter-
4 changes (DDI) and Continuous Flow Intersections (CFI). Sites from Utah, United States, were
5 modeled in PTV VISSIM and calibrated against three key Measures of Effectiveness (MOEs):
6 Traffic Volume, Average Speed, and Crash Frequency. Unlike traditional calibration approaches,
7 the proposed framework incorporates surrogate safety analysis directly into the calibration process
8 by embedding bivariate Peak Over Threshold (POT) modeling from Extreme Value Theory (EVT).
9 This method jointly captures conflict severity using Time to Collision (TTC) and Maximum De-
10 celeration (MaxD), allowing crash frequency estimation from simulated conflicts. The calibration
11 results showed strong agreement with field data, achieving GEH values below 5 for all movements
12 and minimal speed errors. Simulation-based crash estimates differed from observed crash frequen-
13 cies by only 0.07 crashes per year for the CFI. Validation using 2019 data further confirmed the
14 predictive capability and generalizability of the models. This study contributes a robust and trans-
15 ferable framework for evaluating both operational efficiency and safety performance in innovative
16 intersection design.

17

18 *Keywords:* calibration, validation, VISSIM, DDI, CFI, surrogate safety, extreme value theory

1 INTRODUCTION

2 Transportation agencies in the United States have been increasingly implementing innovative in-
3 tersection designs that change the traditional traffic flow pattern and significantly reduce conflict
4 points and increase operational efficiency while providing safety to road users (1, 2). . Two of the
5 most promising designs are Diverging Diamond Interchanges (DDIs) and Continuous Flow Inter-
6 sections (CFIs). The DDI allows vehicles to temporarily shift to the opposite side of the road to
7 ease the direct left turns without conflicting with oncoming traffic. This design improves safety
8 and reduces signal phases, which enhances throughput and reduces delay (3–5). On the other
9 hand, the CFI allows left-turn movements to occur upstream of the main intersection at the sig-
10 nalized crossover points. This design runs the left and through movements at the same time at the
11 intersection, reducing the cycle length significantly (6, 7).

12 Although prior studies have established the theoretical and practical benefits of these de-
13 signs, accurate performance evaluation and safety assessment under real-world conditions remain
14 complex. The non-standard traffic movements and confusing lane configurations in DDIs and CFIs
15 necessitate the use of microscopic traffic simulation models, which offer numerous features for the
16 user to replicate real-world scenarios by detailing vehicle interactions, signal phasing logic, and
17 driver behavior (8, 9). However, simulation models are only reliable if the calibration and valida-
18 tion are accurate and have been conducted rigorously using data accumulated in the field to ensure
19 the reliability of the model and to accurately reflect real-world conditions.

20 A major limitation in many existing simulation-based studies is the sole focus on traffic
21 volume calibration, with the calibration and validation of safety metrics often overlooked. Re-
22 searchers in (10, 11) emphasized that strong calibration should encompass multiple metrics such
23 as traffic volume, vehicle mix type, travel time, average speed, and driver behavior distribution to
24 enhance model trustworthiness.

25 Safety evaluation in simulation poses additional challenges due to limited availability of
26 high-quality crash data and the long-term nature of crash occurrences. Using actual crash counts
27 for calibration and validation is often impractical. To bridge this gap, surrogate safety measures
28 have been developed, which identify vehicle conflicts as proxies for actual crashes. One of the most
29 widely adopted tools in this area is the Surrogate Safety Assessment Model (SSAM), developed
30 by (12). SSAM extracts vehicle trajectories from simulation models built on platforms such as
31 PTV VISSIM and identifies conflicts based on metrics such as Time to Collision (TTC), Post-
32 Encroachment Time (PET), and Maximum Deceleration Rate (MaxD).

33 While surrogate conflict counts provide a valuable approximation of safety conditions, their
34 direct correlation to real-world crash frequency remains complex. To improve this, researchers
35 in (13) have employed the Extreme Value Theory (EVT), particularly the Peak Over Threshold
36 (POT) approach, to model the probabilistic behavior of high-severity conflicts. Researchers in (14–
37 16) introduced the use of Bivariate POT EVT models to estimate crash likelihood by capturing the
38 joint extremity of two conflict variables, offering a more sophisticated understanding of crash risk
39 compared to univariate methods.

40 Regarding the use of traffic microsimulation to evaluate innovative intersection design,
41 operations, and safety, several gaps exist:

- 42 • **Limited multi-dimensional calibration:** Many studies calibrate only for traffic volumes
43 without accounting for speed accuracy and safety fidelity.
- 44 • **Underutilization of bivariate EVT:** Although surrogate conflicts are increasingly used,
45 the application of bivariate EVT models to simulate crash frequency remains rare, par-

1 ticularly for unconventional intersections like DDIs and CFIs.

- 2 • **Lack of validation using independent datasets:** Few studies have validated simulation
3 output using temporally distinct field data, raising concerns about overfitting or underfit-
4 ting to calibration conditions.

5 This study addresses these gaps by proposing a detailed calibration and validation frame-
6 work applied to both DDI and CFI, with study sites located in the state of Utah. The framework
7 integrates both traffic operational and safety performance measures. Calibration is conducted us-
8 ing a tri-metric approach, simultaneously aligning simulated traffic volume, average speed, and
9 surrogate crash frequency with field observations to ensure a more holistic representation of real-
10 world conditions. To assess the generalizability and robustness of the calibrated models, validation
11 is performed using independent traffic and crash datasets from a different year (2019) than that
12 used for calibration (2016). The framework further advances surrogate safety modeling by incor-
13 porating a bivariate POT EVT methodology, which jointly analyzes surrogate conflict indicators
14 (TTC and MaxD) to estimate crash frequencies. The calibrated and validated models exhibit strong
15 agreement with observed crash data. Overall, this study represents one of the earliest systematic
16 efforts to concurrently calibrate traffic operations and safety outcomes using surrogate-based crash
17 modeling through bivariate EVT, with independent validation enhancing the generalizability and
18 practical applicability of the developed models.

19 The remainder of this paper is organized as follows. The methodology section outlines
20 the study site selection, data acquisition, microsimulation model setup, and the calibration and
21 validation framework. This is followed by the analysis results and discussion. The paper concludes
22 by summarizing key findings, highlighting their significance, and suggesting directions for future
23 research.

24 **METHODOLOGY**

25 The methodology adopted for this research involves a systematic approach to calibrate and val-
26 idate microscopic traffic simulation models using PTV VISSIM for two innovative intersection
27 designs: DDI and CFI. Two representative sites (one DDI and one CFI) were modeled and ana-
28 lyzed independently to capture the unique operational and safety characteristics and ensure realistic
29 simulation output. The methodology is structured into four main stages: (1) Pre-modeling Stage,
30 (2) Base Model Stage, (3) Calibration Stage, and (4) Validation Stage. The goal is to develop mod-
31 els that replicate real-world field conditions with good accuracy. The sequence of methodological
32 steps is shown in Figure 1.

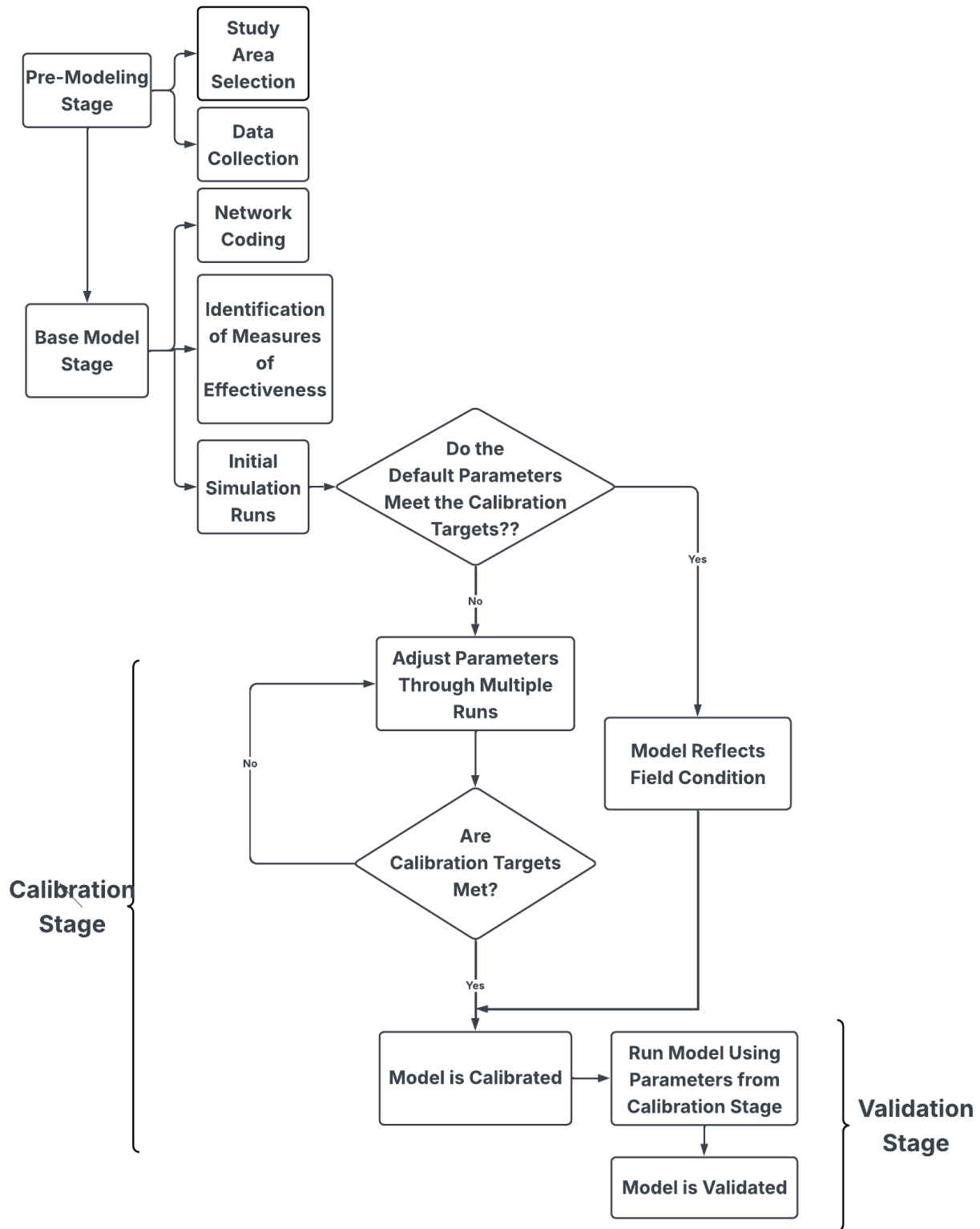


FIGURE 1: Flowchart of methodological framework.

1 Pre-Modeling Stage

2 This stage includes the selection of study sites and data collection for simulation modeling.

3 *Study Area Selection*

4 The study focuses on two unconventional intersection designs located in Salt Lake City, Utah,
5 USA. Each site was selected for its distinctive geometry and operational complexity. The first site
6 is a DDI at the intersection of Bangerter Highway and State Route 201 (SR-201). As shown in
7 Figure 2, the DDI configuration was implemented for the free-flowing movement of left-turning
8 vehicles by briefly switching traffic to the opposite side of the road, reducing conflict points and
9 enhancing signal efficiency. The second site is a CFI at the intersection of Bangerter Highway and
10 3100 South, depicted in Figure 2. The CFI is designed to facilitate continuous left turns by shifting
11 these movements to signalized crossovers upstream of the main intersection. The design improves
12 intersection capacity by allowing simultaneous through and left-turn movements, thus reducing
13 overall delay.

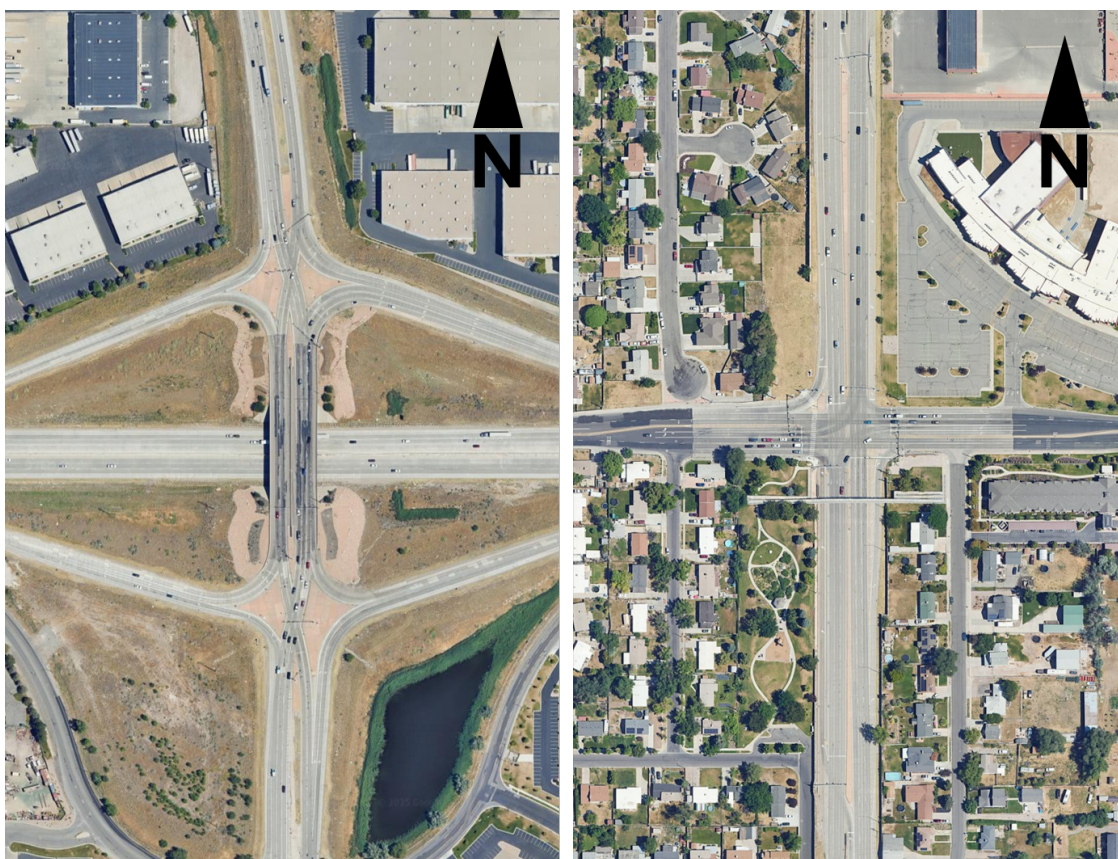


FIGURE 2: Study sites: Left: DDI at Bangerter and SR-201; Right: CFI at Bangerter and 3100 South (17).

14 *Data Collection*

15 For the simulation modeling, traffic volumes, signal timing plans, vehicle speeds, and crash data
16 were collected.

17 **Traffic data:** Annual Average Daily Traffic (AADT) data for the years 2016, 2019, and
18 2024 were obtained from the Utah Department of Transportation (UDOT), AADT Map (18) and

1 the Automated Traffic Signal Performance Measures (ATSPM) system (19). Intersection approach
 2 volumes and turning movement counts from 2024 were first extracted and used as a base refer-
 3 ence. The turning movements for 2016 and 2019 were then estimated based on the 2024 turning
 4 movement proportions and the corresponding AADT values, using a scaling equation:

$$5 \quad \text{Movement Volume}_{\text{target year}} = \text{Movement Volume}_{2024} \times \left(\frac{\text{AADT}_{\text{target year}}}{\text{AADT}_{2024}} \right) \quad (1)$$

6 This method enabled a reasonable approximation of traffic volumes and turning movements
 7 for 2016 and 2019, which were then used for the calibration and validation of the VISSIM models.
 8 However, due to the unavailability of complete traffic data, the eastbound and westbound through
 9 movements at the DDI, and the northbound and southbound right-turn movements at the CFI, were
 10 excluded from the analysis.

11 **Signal timing data:** Signal timing plans, phasing sequences, and controller logic for both
 12 intersections were sourced from official UDOT documentation. These plans included informa-
 13 tion on minimum green times, yellow and red clearance intervals, coordination splits, and phase
 14 compatibility.

15 **Speed data:** Vehicle approach speeds were collected from the UDOT ATSPM, which also
 16 provides the corresponding speed detection location information.

17 **Crash data:** Crash data related to the study sites, for the years 2016 and 2019, were
 18 sourced from UDOT's crash records (20). The dataset included a comprehensive range of crash
 19 attributes, including classifications of injury severity and manner of collision.

20 To ensure comparability between calibration and validation datasets, traffic data were ex-
 21 tracted for the same date of January 15 for both years: 2016 for calibration and 2019 for validation.
 22 The selected peak hour of analysis was from 4:30 PM to 5:30 PM, reflecting the period of highest
 23 traffic demand at both locations. This approach ensured consistent seasonal and temporal condi-
 24 tions across both intersections.

25 **Base Model Stage**

26 In this study, the initial simulation models for both study sites were developed using default driving
 27 behavior parameters and field-based geometry to accurately replicate the physical layout and traffic
 28 control elements of the intersections.

29 *Network and Simulation Setup*

30 The traffic networks for both intersections were set up in VISSIM using satellite imagery (21) to
 31 ensure proper alignment. Traffic inputs, turning movements, vehicle speeds, and signal phasing
 32 and timing plans were set up based on UDOT data.

33 Vehicle composition was defined based on observed traffic trends and was consistent across
 34 both DDI and CFI networks. The model included two main vehicle types: 90% passenger cars and
 35 10% heavy goods vehicles (HGVs). Each vehicle type was assigned to the desired speed distri-
 36 bution that matched the expected behavior for different roadway segments. For the DDI network,
 37 the desired speed along Bangerter Highway ranged from 15–55 mph, while SR-201 was modeled
 38 with desired speeds ranging from 35–65 mph. Turning movements, including both left and right
 39 turns, were assigned reduced speed zones ranging from 15–25 mph to account for deceleration

1 and maneuvering behavior. Similarly, in the CFI model, Bangerter Highway had a speed range of
2 15–55 mph, while 3100 South was assigned a narrower range between 30–40 mph. Left and right
3 turns were also modeled with reduced speed areas of 15–25 mph to enhance realism in vehicle
4 trajectories and interactions.

5 For traffic operational data collection, multiple detectors and nodes were integrated into the
6 network at strategic locations. Vehicle detectors were placed at 3 feet and 15 feet upstream of the
7 signal heads at both DDI and CFI intersections to capture high-resolution arrival data. Additionally,
8 data collection points were positioned to measure average vehicle speeds at specified locations:
9 380 feet upstream of the signal head for the DDI southbound approach and 350 feet for the CFI
10 northbound approach. These placements align with typical field sensor configurations for speed
11 monitoring. Furthermore, nodes were added around the intersections to record detailed vehicle
12 performance metrics such as traffic volume, travel time, vehicle count, and queue length for a
13 comprehensive analysis of operational efficiency under varying traffic conditions.

14 For traffic safety data collection, safety-critical interactions were isolated to ensure consis-
15 tency in conflict analysis. A circular buffer zone was defined around the center of each intersection,
16 as shown in Figure 3. For the DDI, a radius of 280 meters was used, while for the CFI, a smaller
17 buffer of 80 meters was applied due to its more compact geometry. These buffer zones were over-
18 laid directly onto the VISSIM network and utilized during post-processing in SSAM. Only vehicle
19 conflicts that occurred within these defined zones were considered in the safety evaluation, ensur-
20 ing that the analysis focused on the core operational area where the majority of merging, diverging,
21 and crossing maneuvers occurred.

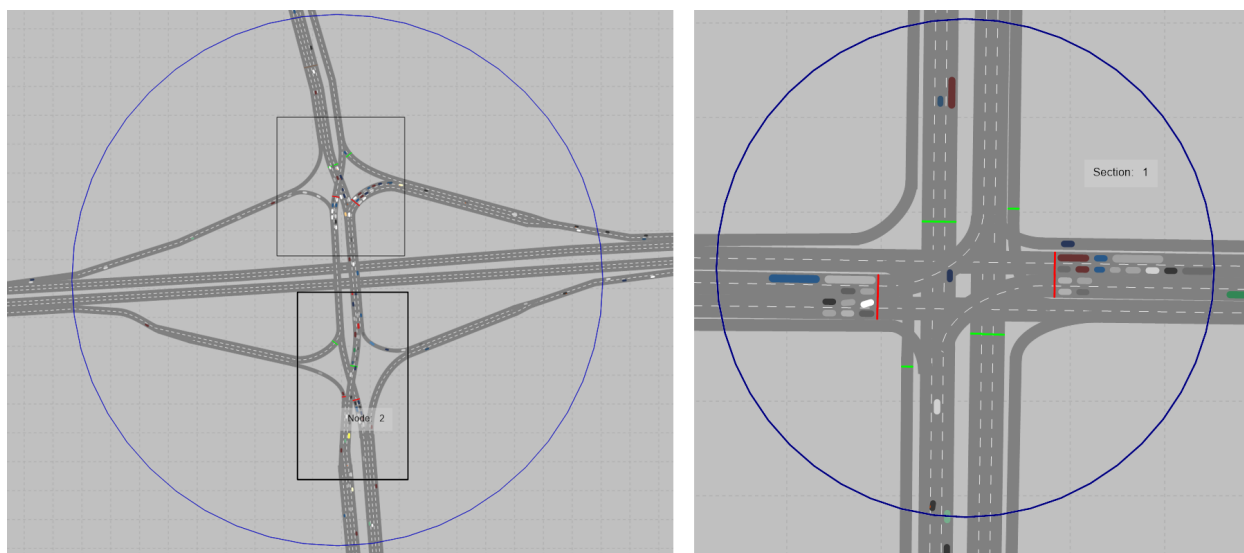


FIGURE 3: Buffer zones for conflict data collection: Left: DDI model; Right: CFI model.

22 Simulation parameters were configured to cover a total period of 4,201 seconds, comprising
23 a 600-second warm-up period, a 3,600-second peak hour simulation period, and a 1-second cool-
24 down phase. The warm-up period allowed the system to reach equilibrium before data collection
25 began. The simulation resolution was set to 10 steps per simulation second to ensure fine-grained
26 tracking of vehicle movements. A random seed value of 42 was used to ensure repeatability across
27 10 simulation runs, with a random speed increment of 1 for variability assessment. The simulation

1 was executed at maximum speed using all available processor cores to optimize computation time
2 without compromising result accuracy.

3 *Identification of Measures of Effectiveness (MOEs)*

4 Calibration and validation of simulation models must be conducted using a set of well-defined
5 MOEs. In this study, traffic volume, average approach speed, and crash frequency were selected
6 as the MOEs. Traffic volume was used to assess the model's ability to reproduce observed flow
7 patterns and demand across the network. Average speed provided insights into prevailing operat-
8 ing conditions, reflecting driver behavior and congestion levels. Crash frequency was estimated
9 using conflict data extracted from SSAM, offering a data-driven approach to evaluate the model's
10 safety performance. These three indicators were chosen for their ability to holistically capture both
11 operational and safety characteristics of the intersections. Importantly, the same MOEs were ap-
12 plied during validation to ensure consistent predictive accuracy when models were tested against
13 independent data from a different year.

14 *Initial Simulation Runs*

15 Initial simulation runs were conducted to ensure that the network was functioning correctly and
16 that the base model was appropriately configured before beginning the calibration and validation
17 process. Careful analysis of the geometric layout, signal timing implementation, vehicle routing,
18 and traffic flow behavior was conducted under typical flow conditions. A key aspect of this phase
19 was the verification of lane assignments, detector placement, and traffic control logic.

20 Several operational issues were identified and addressed. To evaluate speed consistency
21 and detect outliers in vehicle performance, a random S-curve speed distribution analysis was per-
22 formed. Unusual congestion patterns were also observed at specific locations, prompting refine-
23 ment of traffic inputs and signal coordination. Signal-related problems, such as phase mismatches
24 or overlaps, were examined and corrected to align with the provided signal plans. All driver be-
25 havior parameters, including car-following and lane-changing models, were initially kept at their
26 default values as defined by VISSIM, which provided a baseline representation of driver tenden-
27 cies.

28 **Calibration**

29 Calibration is a critical step to fine-tune the microsimulation model inputs so that the simulated
30 outputs align with the real-world conditions observed during the simulation period. This calibra-
31 tion effort focused on three selected MOEs: traffic volumes, average approach speed, and crash
32 frequency. To account for simulation randomness, each scenario was executed 10 times using
33 different random seed values, and average results were used for evaluation.

34 *Traffic Volume Calibration*

35 To assess the accuracy of the simulated traffic volumes, the following metrics were used:

36 **Geoffrey E. Havers (GEH) statistic:** The GEH statistic is widely used in traffic simulation
37 to assess the accuracy of individual movement volumes.

$$38 \quad \text{GEH}_i = \sqrt{\frac{2(S_i - O_i)^2}{S_i + O_i}} \quad (2)$$

39 where O_i = observed traffic volume for turning movement i , and S_i = simulated traffic volume

1 for movement i . A GEH value below 5 indicates a good match; between 5 and 10 indicates an
2 acceptable match, and above 10 suggests a poor match.

3 **Root Mean Square Error (RMSE):** RMSE is used to quantify the difference between
4 observed and simulated traffic volumes by measuring the square root of the average squared devi-
5 ations. It gives more weight to large errors, making it sensitive to outliers.

$$6 \quad \text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (O_i - S_i)^2} \quad (3)$$

7 **Mean Absolute Percentage Error (MAPE):** MAPE expresses the average magnitude of
8 the error between observed and simulated traffic volumes. It provides a normalized measure of
9 prediction accuracy that is easy to interpret. A value below 10% is considered excellent for traffic
10 simulation studies.

$$11 \quad \text{MAPE} = \frac{1}{n} \sum_{i=1}^n \left| \frac{O_i - S_i}{O_i} \right| \times 100 \quad (4)$$

12 A MAPE below 10% indicates a very good match; between 10% and 20% indicates an acceptable
13 match, between 20% and 30% suggests the need for calibration tuning, and above 30% indicates
14 that the model accuracy needs improvement.

15 *Traffic Speed Calibration*

16 Speed accuracy was evaluated using average speeds collected from field detectors and their corre-
17 sponding simulated values from VISSIM's data collection points. The metrics used were absolute
18 error and percent error:

$$19 \quad \text{Absolute Error} = |S_i - O_i| \quad (5)$$

$$20 \quad \text{Percent Error} = \left| \frac{S_i - O_i}{O_i} \right| \times 100 \quad (6)$$

21 An error below 10% was considered acceptable for model calibration. To assess the accuracy of
22 simulated average speeds, absolute and percent errors were selected over RMSE or MAPE due to
23 their straightforward interpretability and consistency with the volume calibration metrics used in
24 the study. These error types provide direct, easily understandable comparisons for practical eval-
25 uation of traffic model performance. While RMSE and MAPE are also valid statistical measures,
26 the chosen metrics align better with the reporting format used throughout the calibration process.
27 Future work may explore the use of RMSE/MAPE for additional validation layers if needed.

28 For adjustments, rather than modifying car-following or lane-change aggressiveness pa-
29 rameters without the support of field-observed data, the calibration process retained the default
30 VISSIM driver behavior settings. Instead, since reliable speed data were available, the desired
31 speed distribution curves were adjusted to better replicate observed speed conditions and flow be-
32 havior in the field.

33 *Crash Frequency Calibration*

34 A key component of the model calibration is ensuring that the simulated crash frequency aligns
35 with field-observed crash data. This was achieved through a structured surrogate safety analysis.

36 **Conflict detection using SSAM:** Vehicle trajectories from VISSIM were exported and an-

1 analyzed in SSAM to detect conflicts. To convert the hourly number of conflicts from the simulation
 2 into an estimate of the annual number of conflicts, a conversion equation was developed as follows:
 3

$$4 \quad \text{Conflicts}_{\text{year}} = \left(\frac{C_{\text{hour}}}{P_{\text{wd}}} \times D_{\text{wd}} \right) + \left(\frac{C_{\text{hour}}}{P_{\text{nw}}} \times D_{\text{nw}} \right) \quad (7)$$

5 where C_{hour} = total conflicts per peak hour; P_{wd} = percentage of daily traffic occurring during the
 6 peak hour on workdays; P_{nw} = percentage of daily traffic occurring during the peak hour on non-
 7 workdays; D_{wd} = number of workdays in a year; and D_{nw} = number of non-workdays in a year.
 8 This study assumes a typical year has 250 workdays and 115 non-workdays.

9 **Crash estimation using EVT:** Crash estimation in this study was performed using EVT,
 10 a statistical approach that focuses on modeling the behavior of the extreme tail of a distribution.
 11 One widely used method under EVT is the POT approach. POT models the tail of a distribution by
 12 focusing only on values that exceed a selected threshold. For conflict data, this method helps isolate
 13 the most severe interactions, which are most likely to lead to crashes. Let Z be a variable, such
 14 as inverse TTC, and u be a threshold. The exceedances over this threshold follow a Generalized
 15 Pareto Distribution (GPD), which is defined as:

$$16 \quad P(Z > z | Z > u) = 1 - \left(1 + \frac{\xi(z-u)}{\sigma} \right)^{-1/\xi} \quad (8)$$

17 where ξ = shape parameter; σ = scale parameter; u = threshold; and z = exceedance value over u .

18 In the bivariate case, two conflict severity variables are modeled together, such as TTC
 19 and relative speed, or TTC and MaxD. This joint modeling approach accounts for the combined
 20 risk of both variables becoming extreme, which is often more predictive of actual crashes than
 21 univariate analysis. The bivariate distribution of threshold exceedances is expressed through a
 22 copula function or a joint GPD. For threshold exceedances ($X > u_1, Y > u_2$), the joint survivor
 23 function is:

$$24 \quad P(X > x, Y > y | X > u_1, Y > u_2) = \bar{G}(x, y) \quad (9)$$

25 This function is typically estimated using parametric or semi-parametric models, which
 26 allow us to derive the annual probability of crashes based on the joint severity of conflicts. Figure 4
 27 and Figure 5 show the DDI and CFI conflicts per hour plotted in the bivariate space formed by TTC
 28 and MaxD.

$$29 \quad \text{Crashes}_{\text{year}} = \text{Conflicts}_{\text{year}} \times P_{\text{crash}|\text{conflict}} \quad (10)$$

30 where $P_{\text{crash}|\text{conflict}}$ = the probability of a conflict leading to a crash, as estimated from the bivariate
 31 EVT model.

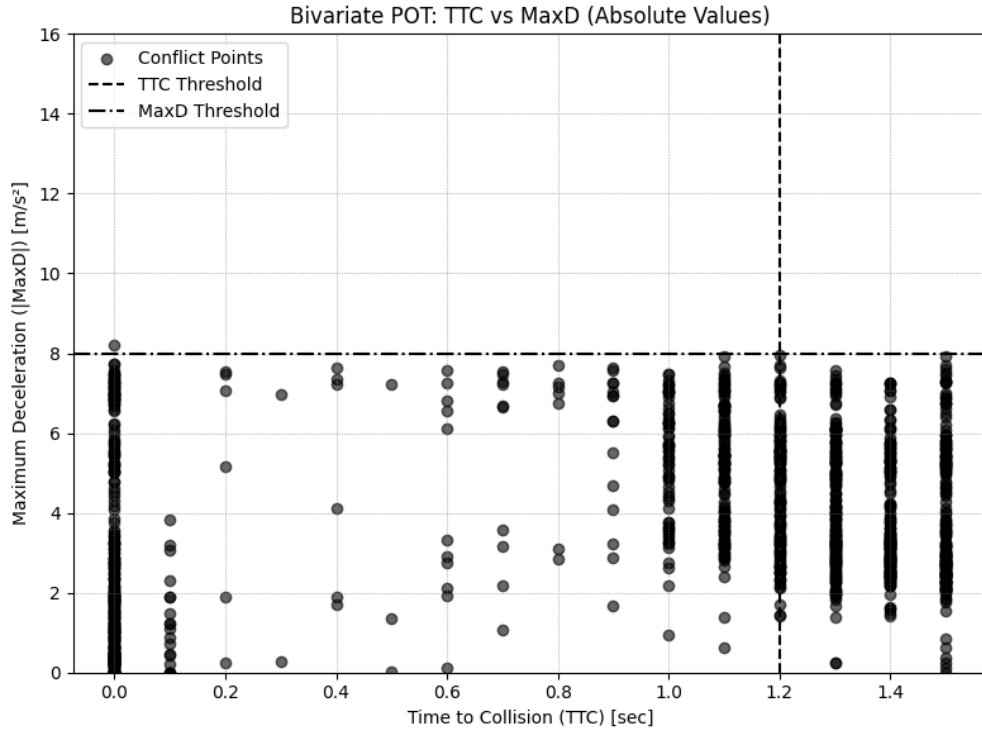


FIGURE 4: Bivariate POT: DDI conflicts plotted using TTC and MaxD with thresholds at $TTC \leq 1.2$ s and $MaxD \geq 8$ m/s^2 . The upper-left quadrant indicates high-risk conflicts exceeding both thresholds.

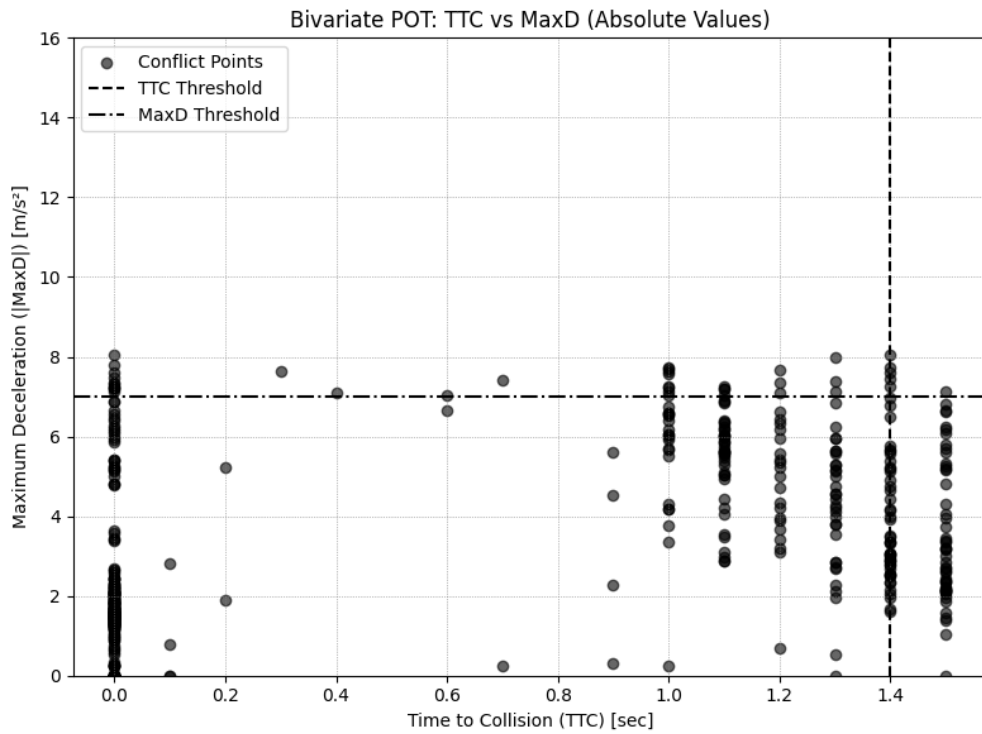


FIGURE 5: Bivariate POT: CFI conflicts plotted using TTC and MaxD with thresholds at $TTC \leq 1.4$ s and $MaxD \geq 7$ m/s^2 . The upper-left quadrant indicates high-risk conflicts exceeding both thresholds.

1 The selection of TTC and MaxD thresholds used in the bivariate EVT model was guided
 2 by prior studies and empirical analyses of conflict severity within the surrogate safety literature.
 3 Specifically, threshold values of $TTC \leq 1.2\text{--}1.5$ seconds and $MaxD \geq 6\text{--}8$ m/s² have been widely
 4 adopted to identify critical vehicle interactions that are most likely to lead to crashes, particularly
 5 in signalized intersections and complex roadway geometries. For this study, a sensitivity analysis
 6 was first conducted to examine the influence of different threshold combinations on conflict filter-
 7 ing and crash estimation outcomes. The final thresholds, $TTC \leq 1.2$ s and $MaxD \geq 8$ m/s² for the
 8 DDI and $TTC \leq 1.4$ s and $MaxD \geq 7$ m/s² for the CFI, were chosen to balance model calibration
 9 accuracy, empirical consistency, and the physical interpretability of conflicts. These values pro-
 10 duced distributions that focused on high-risk interactions without overrepresenting benign events.
 11 The selected thresholds are consistent with definitions and practices used in previous studies such
 12 as those by (22) and (23), reinforcing their relevance for use in bivariate crash prediction models.

13 *Thresholds*

14 The calibration process was guided by specific threshold targets and evaluation metrics suggested
 15 by FHWA Traffic Analysis Volume III, as shown in Table 1.

TABLE 1: Calibration parameters and goals.

Calibration Parameters	Calibration Target/Goal
Traffic Volume	Simulated and measured link volume for more than 85% of the links to be: 1) Within 100 vph for volumes less than 700 vph 2) Within 15% for volumes between 700 and 2700 vph 3) Within 400 vph for volumes greater than 2700 vph Simulated and measured link volumes for more than 85% of links to have a GEH value ≤ 5 .
Speed	Modeled average link speeds to be within ± 10 mph of field speeds on at least 85% of links.
Visualization Check	Check field consistency: queuing, weaving, signal patterns, lane usage, bottlenecks, etc.
Crashes/Safety	Simulated conflicts converted to crash frequencies should be within $\pm 25\%$ of observed field crash frequency.

16 *Calibration Framework and Parameter Tuning*

17 The calibration process in this study followed a sequential approach rather than a multi-objective
 18 optimization framework. Specifically, the model was first calibrated for traffic volumes using the
 19 GEH statistic, RMSE, and MAPE to ensure alignment with observed turning movement counts.
 20 Once volume targets were met, the average approach speeds were calibrated by adjusting the de-
 21 sired speed distribution curves in VISSIM. The adjustments involved tuning the mean and stan-
 22 dard deviation of the assumed Gaussian distribution for desired speeds across various roadway
 23 segments. No other driver behavior parameters (e.g., car-following or lane-change aggressiveness)
 24 were modified. This limitation was intentional to avoid overfitting and to ensure generalizability,
 25 particularly due to the absence of detailed field data for those parameters. Once the traffic oper-
 26 ational aspects were calibrated, we moved on to calibrate the surrogate crash frequency. Conflict

1 data from SSAM were transformed into annual crash estimates using a bivariate POT EVT ap-
2 proach. The bivariate EVT model jointly analyzed TTC and MaxD to estimate the likelihood of
3 high-severity crashes. Each calibration step was repeated iteratively, using average outputs from
4 ten simulation runs, until the model achieved close agreement with field observations across all
5 three metrics.

6 All calibration performance metrics were computed based on aggregated values over the
7 entire 3,600-second simulation period, rather than at individual time steps. This strategy ensured
8 consistency in model tuning and supported the application of field-representative summary statis-
9 tics.

10 *Validation*

11 The validation process assessed both the predictive capability and generalizability of the calibrated
12 DDI and CFI microsimulation models. Independent datasets from 2019 were used to determine
13 whether the models, calibrated using 2016 data, could accurately reproduce real-world traffic con-
14 ditions and safety outcomes without further model adjustment. For validation, only the input ap-
15 proach volumes were updated to reflect the 2019 observed volumes, specifically for the 4:30–5:30
16 PM peak hour on January 15, 2019. All other model elements, including driver behavior paramet-
17 ers, traffic composition, and speed distributions, were held constant from the calibration phase to
18 ensure an unbiased assessment of model predictive performance.

19 We compared simulated and observed turning movement volumes, approach speeds, and
20 estimated crash frequencies, using the same set of metrics applied during the calibration process
21 to maintain consistency in performance evaluation. Furthermore, the conflict-to-crash conversion
22 functions developed during calibration were applied without modification. This approach ensured
23 that the validation rigorously evaluated the models' ability to generalize to new traffic conditions
24 based solely on changes in traffic demand, without relying on further tuning or recalibration.

25 While the current study validated the crash prediction model using a single year of addi-
26 tional data, we recognize that assessing the predictive accuracy of surrogate safety-based models
27 over time is crucial for broader applicability. The proposed modeling framework, particularly the
28 use of bivariate EVT, is designed to capture the statistical behavior of extreme conflict events,
29 which are generally stable unless significant changes occur in traffic operations or roadway geom-
30 etry. Although only two years of crash data were available for comparison, the model demonstrated
31 strong consistency in predicting crash frequencies. In future research, we plan to expand the val-
32 idation effort by incorporating multi-year crash datasets to systematically evaluate the temporal
33 robustness of the model and enhance its applicability for long-term safety forecasting.

34 *Impact of Safety Metrics in Calibration*

35 To evaluate the added value of incorporating safety metrics in the calibration process, a qualitative
36 ablation analysis was conducted. We examined how the model would perform if calibrated solely
37 on traffic volume and speed, excluding crash frequency as a target metric. In such a scenario, the
38 desired speed distributions would still be adjusted to align simulated traffic flow and speeds with
39 field observations; however, no effort would be made to match simulated surrogate conflict severity
40 with observed crash data.

41 Based on the final calibration results, the full tri-metric approach, including crash fre-
42 quency, yielded crash estimates that closely matched field conditions, with only a 2.69 crashes/year
43 difference for the DDI and a 0.07 crashes/year difference for the CFI. Preliminary tests during

1 model development (prior to crash tuning) showed that omitting crash frequency led to consis-
2 tent overestimation of crash risk, likely due to unrefined vehicle speed behavior in conflict-prone
3 zones. Although these early results were not retained for formal comparison, the trend was consis-
4 tent across multiple iterations.

5 RESULTS

6 Calibration Results

7 Traffic Volume

8 For the DDI model, a high degree of correlation was achieved between observed and simulated
9 volumes, with an R^2 of 0.9986. All movement volumes had GEH values below 5, indicating
10 a strong match. The RMSE was 39.26 and the MAPE was 16.05%. Similarly, the CFI model
11 demonstrated an excellent fit with an R^2 of 0.9976. All movements also had GEH values below 5,
12 with an RMSE of 24.41 and a MAPE of 8.13%. Both models met the calibration targets, suggesting
13 that the simulated traffic volumes closely replicate observed field conditions. The regression results
14 are shown in Figure 6 and Figure 7, and the GEH results are shown in Table 2 and Table 3 for the
15 DDI and CFI, respectively.



FIGURE 6: Regression plot for DDI.

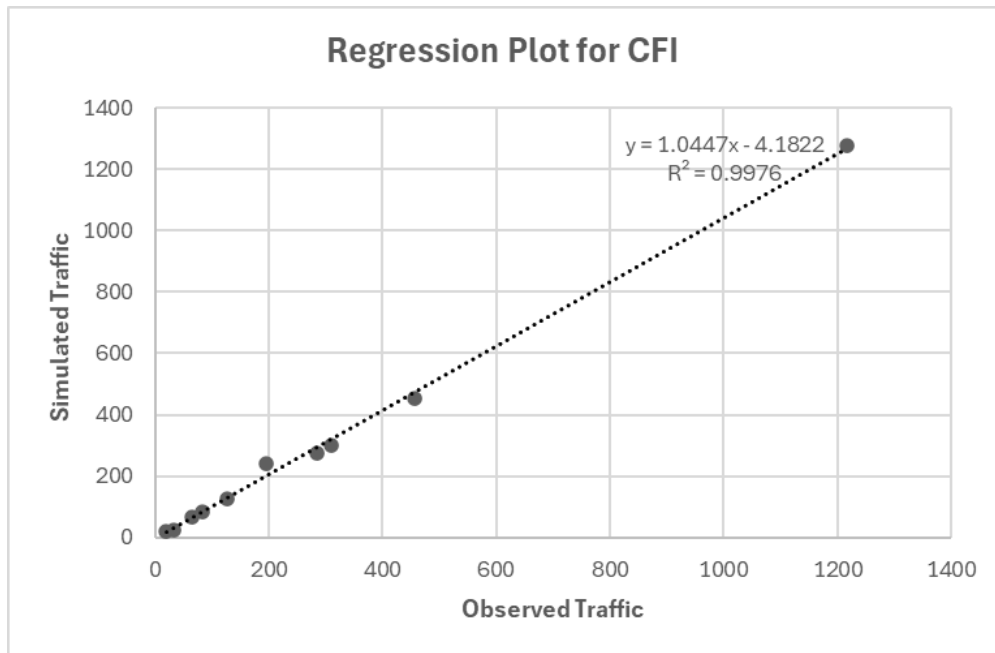


FIGURE 7: Regression plot for CFI.

1 *Average Speed*

2 Average speed calibration was conducted by comparing field-measured speeds with simulated out-
 3 puts for the major approaches at each intersection. For the DDI, the observed average speed was
 4 28.00 mph on the northbound approach, while the simulated speed was 28.29 mph. The absolute
 5 error was 0.29 mph, resulting in a percentage error of 1.04%. For the CFI, the field-measured speed
 6 was 32.31 mph on the southbound approach, and the simulated speed was 32.60 mph. The absolute
 7 error was also 0.29 mph, with a percentage error of 0.90%. In both cases, the percentage error was
 8 well below the 10% threshold, indicating a high degree of accuracy in replicating real-world traffic
 9 speeds.

TABLE 2: GEH values for DDI: Observed vs. Simulated Traffic Volumes

Movement	Simulated Volume	Actual Volume	GEH
EB Left	152	166	1.11
EB Right	111	94	1.67
WB Left	1106	1064	1.27
WB Right	113	95	1.76
NB Left	373	339	1.80
NB Right	790	727	2.28
NB Through	538	496	1.84
SB Left	557	544	0.55
SB Right	7	4	1.27
SB Through	1235	1159	2.19

TABLE 3: GEH values for CFI: Observed vs. Simulated Traffic Volumes

Movement	Simulated Volume	Actual Volume	GEH
EB Left	127	128	0.08
EB Right	68	65	0.36
EB Through	301	309	0.45
WB Left	243	196	3.17
WB Right	23	32	1.71
WB Through	276	286	0.59
NB Left	85	83	0.21
NB Through	454	456	0.09
SB Left	22	20	0.43
SB Through	1275	1216	1.67

1 *Crash Frequency*

2 Crash frequency calibration was conducted using conflict data extracted from SSAM and analyzed
3 via the Bivariate Peak Over Threshold (POT) method from Extreme Value Theory. This approach
4 modeled the joint severity of traffic conflicts using Time to Collision (TTC) and Maximum Decel-
5 eration (MaxD) thresholds, converting hourly conflict counts into annual crash estimates.

6 For the DDI model, using a threshold of $TTC \leq 1.2$ s and $MaxD \geq 8$ m/s², the model esti-
7 mated an annual crash frequency of 22.31 crashes/year, compared to the observed 25 crashes/year
8 in the field, as shown in Table 4. The difference of 2.69 crashes/year indicates strong alignment
9 between the simulated and real-world safety outcomes. For the CFI model, with thresholds of
10 $TTC \leq 1.4$ s and $MaxD \geq 7$ m/s², the estimated crash frequency was 17.93 crashes/year, closely
11 matching the observed value of 18 crashes/year, as shown in Table 5. The difference was only
12 0.07 crashes/year. These results confirm the effectiveness of the surrogate safety-based calibration
13 method, demonstrating that the VISSIM model can realistically simulate crash-prone interactions
14 at both intersection types.

TABLE 4: Crash Frequency Calibration for DDI by Conflict Type

Conflict Type	Conflicts per Hour	Conflicts per Year	Crashes per Year (Simulated)	Crashes per Year (Field)
Crossing	110	604,348.4	2.13	5
Rear-End	1052	5,014,117	17.65	19
Lane Change	131	719,724	2.53	1
Total	1293	7,103,840	22.31	25

TABLE 5: Crash Frequency Calibration for CFI by Conflict Type

Conflict Type	Conflicts per Hour	Conflicts per Year	Crashes per Year (Simulated)	Crashes per Year (Field)
Crossing	9	48,904.6	0.31	3
Rear-End	450	2,445,229.6	15.68	14
Lane Change	56	304,295.2	1.95	1
Total	515	2,798,429.4	17.93	18

1 Validation Results

2 Validation of the calibrated models was conducted using 2019 field data. Results for both intersec-
3 tions show strong agreement between the simulation outputs and observed conditions. Overall, the
4 validation confirms the reliability of the calibrated models in reproducing real-world operational
5 and safety performance for both intersection types.

6 *DDI Model*

7 **Traffic Volume:** The model achieved an R^2 value of 0.9876, with RMSE = 45.73, MAPE = 9.17%,
8 and 9 out of 10 movements having GEH < 5. **Speed:** The simulated average speed of 28.55 mph
9 closely matched the observed speed of 28.00 mph, with a percentage error of 1.96%. **Crash**
10 **Frequency:** Using the bivariate POT method ($TTC \leq 1.2$ s and $MaxD \geq 8$ m/s²), the model
11 predicted 21.9 crashes/year, compared to 8 crashes/year observed. Although the predicted crash
12 frequency notably exceeds the observed value, this overestimation reflects the conservative nature
13 of the bivariate surrogate safety modeling approach. Since the model prioritizes high-severity
14 conflict events, such predictions can be sensitive to the chosen thresholds and the stochastic nature
15 of microsimulation outputs. Importantly, while the absolute values differ, the model preserved the
16 relative patterns of conflict types, which is essential for comparative safety analysis. This result
17 underscores the need for further validation using multi-year crash records or extended empirical
18 datasets to refine model calibration and improve long-term predictive reliability.

19 *CFI Model*

20 **Traffic Volume:** An R^2 of 0.9983 was observed, with RMSE = 36.33, MAPE = 11.86%, and 9 out
21 of 10 movements within the GEH < 5 threshold. **Speed:** The model produced an average speed
22 of 32.44 mph, closely replicating the field value of 32.31 mph, resulting in a percentage error of
23 0.40%. **Crash Frequency:** The estimates were 19.64 crashes/year ($TTC \leq 1.4$ s and $MaxD \geq$
24 7 m/s²), consistent with the 20 crashes/year observed, validating the model's predictive capability
25 for safety performance.

26 CONCLUSION

27 This study developed and validated a comprehensive calibration framework for VISSIM microsim-
28 ulation models applied to two innovative intersection types: a DDI and a CFI. The framework
29 incorporated traffic volumes, average speeds, and surrogate safety metrics into the calibration pro-
30 cess and applied bivariate EVT to estimate crash frequencies.

31 Unlike conventional threshold-based surrogate safety methods, the use of bivariate EVT
32 enabled joint modeling of conflict severity indicators, TTC and MaxD, resulting in more realistic
33 crash predictions, especially in low-crash settings. The calibrated models demonstrated strong
34 alignment with field data, with all volume and speed metrics meeting FHWA calibration targets.
35 Additionally, simulated crash frequencies closely matched observed crash data, highlighting the
36 importance of including safety metrics during model calibration.

37 Validation using independent 2019 field data confirmed the robustness of the models. While
38 the DDI model slightly overestimated crash frequency, the relative trends in conflict severity were
39 preserved, supporting the validity of the surrogate safety-based approach. The CFI model exhibited
40 near-perfect alignment in both operational and safety outputs.

41 This study is limited to two intersections in Salt Lake City, Utah, and therefore the results
42 may not generalize to other locations with different traffic dynamics, geometries, or driver behav-

1 iors. Future work should explore applying this framework to other intersection types and regions,
2 incorporate real-time traffic and conflict data for dynamic model updating, and utilize machine
3 learning to enhance predictive accuracy in crash modeling.

4 Overall, the proposed framework advances microsimulation calibration practice by aligning
5 model outputs with both operational and safety performance metrics. It provides a scalable, data-
6 driven tool to inform the planning, evaluation, and design of innovative intersection treatments
7 aimed at improving traffic efficiency and roadway safety.

8 **ACKNOWLEDGMENT**

9 The author used the large language model (LLM), GPT-4o, to conduct grammatical and spelling
10 checks and corrections throughout the paper. The model was not used for any citation of related
11 works, data processing, data analysis, or graphing. The authors understand that LLMs have limi-
12 tations such as potential for bias, errors, and gaps in knowledge, and only used the model to polish
13 the writing. The authors take full responsibility for the content in this paper.

14 **AUTHOR CONTRIBUTION**

15 The authors confirm contribution to the paper as follows: study conception and design: KN, YS;
16 data collection: KN, RD, YS; analysis and interpretation of results: KN, YS; draft manuscript
17 preparation: KN, RD. All authors reviewed the results and approved the final version of the
18 manuscript.

19 **DECLARATION OF CONFLICTING INTERESTS**

20 The authors declared no potential conflicts of interest with respect to the research, authorship,
21 and/or publication of this article.

22 **FUNDING**

23 This work was supported in part by the Center for Transformative Infrastructure Preservation and
24 Sustainability (CTIPS), funded by a grant from U.S. Department of Transportation's University
25 Transportation Centers Program under Grant 69A3552348308 (IIJA/BIL). The work presented in
26 this paper remains the responsibility of the authors.

27 **REFERENCES**

- 28 1. Federal Highway Administration (FHWA), *Innovative Intersection Design*. U.S. Depart-
29 ment of Transportation, Washington, D.C., 2021.
- 30 2. Federal Highway Administration (FHWA), *Alternative Intersections/Interchange: Infor-*
31 *mational Report (AIIR)*. U.S. Department of Transportation, Washington, D.C., 2010.
- 32 3. Hunter, M., A. Guin, J. Anderson, and S. J. Park, Operating Performance of Diverging
33 Diamond Interchanges. *Transportation Research Record: Journal of the Transportation*
34 *Research Board*, Vol. 2673, No. 11, 2019, pp. 801–812.
- 35 4. Edara, P. K., J. G. Bared, and R. Jagannathan, Diverging diamond interchange and double
36 crossover intersection–vehicle and pedestrian performance. In *3rd International Sympo-*
37 *sium on Highway Geometric Design, Chicago, IL*, 2005.
- 38 5. Hummer, J. E., C. M. Cunningham, R. Srinivasan, S. Warchol, B. Claros, P. Edara, and
39 C. Sun, Safety evaluation of seven of the earliest diverging diamond interchanges installed
40 in the United States. *Transportation research record*, Vol. 2583, No. 1, 2016, pp. 25–33.

- 1 6. Federal Highway Administration (FHWA), *Signalized Intersection: Informational Guide*.
2 Report FHWA-HRT-04-091, U.S. Department of Transportation, Washington, D.C., 2004.
- 3 7. Goldblatt, R., F. Mier, and J. Friedman, Continuous flow intersections. *ITE journal*,
4 Vol. 64, 1994, pp. 35–35.
- 5 8. Federal Highway Administration (FHWA), *Traffic analysis toolbox volume III: Guidelines*
6 *for applying traffic microsimulation modeling software*. U.S. Department of Transporta-
7 tion, Washington, D.C., 2004.
- 8 9. Liu, H., H. Deng, J. Li, S. Yang, K. Dong, and Y. Zhao, Calibration method for microscopic
9 traffic simulation considering lane difference. *Simulation*, 2024.
- 10 10. Park, B. B. and J. D. Schneeberger, Microscopic Simulation Model Calibration and Vali-
11 dation: Case Study of VISSIM Simulation Model for a Coordinated Actuated Signal Sys-
12 tem. *Transportation Research Record: Journal of the Transportation Research Board*, Vol.
13 1856, No. 1, 2003, pp. 185–192.
- 14 11. Chu, L., H. Liu, J.-S. Oh, and W. Recker, A calibration procedure for microscopic traf-
15 fic simulation. In *Proceedings of the 2003 IEEE International Conference on Intelligent*
16 *Transportation Systems*, IEEE, Shanghai, China, 2003, Vol. 2, pp. 1574–1579.
- 17 12. Gettman, D. and L. Head, Surrogate Safety Measures from Traffic Simulation Models.
18 *Transportation Research Record: Journal of the Transportation Research Board*, Vol.
19 1840, No. 1, 2003, pp. 104–115.
- 20 13. Songchitruksa, P. and A. P. Tarko, The extreme value theory approach to safety estimation.
21 *Accident Analysis & Prevention*, Vol. 38, No. 4, 2006, pp. 811–822.
- 22 14. Zheng, L., K. Ismail, T. Sayed, and T. Fatema, Bivariate extreme value modeling for road
23 safety estimation. *Accident Analysis & Prevention*, Vol. 120, 2018, pp. 83–91.
- 24 15. Gastaldi, M., F. Orsini, G. Gecchele, and R. Rossi, Safety analysis of unsignalized inter-
25 sections: a bivariate extreme value approach. *Transportation Letters*, Vol. 13, No. 3, 2021,
26 pp. 209–218.
- 27 16. Borsos, A., Application of Bivariate Extreme Value models to describe the joint behavior
28 of temporal and speed related surrogate measures of safety. *Accident Analysis & Preven-*
29 *tion*, Vol. 159, 2021, p. 106274.
- 30 17. Google, *Google Earth*. Available online at [https://earth.google.com/web/@40.](https://earth.google.com/web/@40.72848761,-111.96391478,1291.9119413a,5760.27796359d,35y,0.01755056h,0t,0r/data=CgRCAggB0gMKATBCAggBSg0I_____ARAA)
31 [72848761,-111.96391478,1291.9119413a,5760.27796359d,35y,0.01755056h,](https://earth.google.com/web/@40.72848761,-111.96391478,1291.9119413a,5760.27796359d,35y,0.01755056h,0t,0r/data=CgRCAggB0gMKATBCAggBSg0I_____ARAA)
32 [0t,0r/data=CgRCAggB0gMKATBCAggBSg0I_____ARAA](https://earth.google.com/web/@40.72848761,-111.96391478,1291.9119413a,5760.27796359d,35y,0.01755056h,0t,0r/data=CgRCAggB0gMKATBCAggBSg0I_____ARAA), 2025, accessed April
33 29, 2025.
- 34 18. UDOT, *Utah Department of Transportation AADT Google Map*. Available online at
35 <https://www.udot.utah.gov/connect/docs/aadt-google-map/>, 2025.
- 36 19. UDOT, *Utah Department of Transportation Automated Traffic Signal Performance Mea-*
37 *sures*. Available online at <https://udottraffic.utah.gov/atspm/>, 2025.
- 38 20. Utah Department of Transportation, *Crash data for study sites, 2016 and 2019*, 2016/2019,
39 unpublished raw data, obtained via public records request.
- 40 21. PTV Group, *PTV VISSIM*. PTV Group, Karlsruhe, Germany, 2023.
- 41 22. Gettman, D. and L. Head, Surrogate Safety Measures from Traffic Simulation Models.
42 *Transportation Research Record*, Vol. 2083, No. 1, 2008, pp. 104–115.
- 43 23. Lareshyn, A., Å. Svensson, and C. Hydén, Evaluation of traffic safety, based on micro-
44 level behavioural data: Theoretical framework and first implementation. *Accident Analysis*
45 *& Prevention*, Vol. 42, No. 6, 2010, pp. 1637–1646.