A Hybrid Extreme Value Theory Framework for Adaptive Pedestrian Crash Risk Estimation

Amir Hossein Kalantari[1*[0000-0001-5256-8069]], Amna Chaudhry[1[0009-0003-2426-7025]], Stella Roussou[2[0009-0002-3180-8291]], Apostolos Ziakopoulos [2[0000-0001-6252-6743]] and Amir Pooyan Afghari[1[0000-0003-1158-880X]]

[1] Safety and Security Section, Faculty of Technology, Policy and Management, TU Delft, Delft Netherlands
`*a.h.kalantari@tudelft.nl, A.Chaudhry@tudelft.nl, A.P.Afghari-1@tudelft.nl`

[2] Department of Transportation Planning and Engineering, National Technical University of Athens, Athens, Greece
`s_roussou@mail.ntua.gr, apziak@central.ntua.gr`

Abstract. Proactive road safety approaches have been mostly used for assessing crash risk for motorized vehicles but remain underexplored for pedestrian–vehicle interactions. This study introduces a crash–conflict framework for pedestrian–vehicle interactions that infers risk from behavioral anomalies rather than historical crash data. Classical Extreme Value Theory (EVT) relies on assumptions regarding independence, threshold selection, and tail behavior; when data are noisy, temporally dependent, or sparsely sampled, as in many pedestrian–vehicle interactions, its estimates can be unstable or biased. To address these challenges, a diagnostic, data-driven EVT workflow is employed, integrating principled extreme-event selection with declustering and independence checks, systematic threshold evaluation to reduce subjectivity, and flexible tail modeling for robust and interpretable crash risk estimation. Applied to naturalistic pedestrian–vehicle conflicts at a crosswalk in Athens, the framework produces stable and consistent risk estimates across model formulations. The framework supports scalable, evidence-based pedestrian safety, enabling credible prioritization of high-risk sites, continuous monitoring, and efficient before–after evaluation in complex urban environments.

Keywords: Pedestrian safety; Traffic conflicts; Crash risk estimation, Anomaly detection; GAMLSS

1 Introduction

Although road safety has improved in many areas through advances in vehicle design and infrastructure, pedestrians remain among the most vulnerable road users, accounting for a disproportionate share of traffic fatalities worldwide. Traditional crash-based road safety analysis has long suffered from a fundamental data scarcity problem: serious pedestrian crashes are rare and spatially dispersed, meaning that statistical models built solely on historical crash counts are often underpowered and site-specific. This limitation motivates the use of proactive road safety analysis, which seeks to infer crash risk from traffic conflicts rather than relying solely on historical crash data.

While traffic conflicts are more frequent than crashes, serious conflicts that may be highly correlated with crashes are still rare. A common approach for quantifying these rare, high-risk conflicts is Extreme Value Theory (EVT), which provides a statistical framework for extrapolating beyond observed data to the tails of safety surrogate measures (SSMs) such as time to collision (TTC). Traditionally, studies have employed either the block maxima (BM) or the peaks-over-threshold (POT) approach. These methods have been applied to driver behavior, mixed-traffic conditions at intersections, and pedestrian crossing scenarios [1, 2].

Applying EVT to naturalistic pedestrian conflicts, however, has several practical limitations. A key assumption is that extreme events are independent, whereas in reality, interactions often occur in clusters, for example when pedestrians cross in groups or vehicles arrive in clusters [3]. In addition, the definition of extremes depends critically on threshold selection. In practice, thresholds are often specified using fixed quantiles (e.g., the 90th, 95th, or 99th percentile of the distribution), but such choices are not uniquely determined and can lead to substantially different risk estimates under otherwise identical conditions [4, 5]. This sensitivity is further compounded by the statistical characteristics of pedestrian conflict data, which are often temporally dependent, noisy due to measurement and tracking uncertainty, and limited in effective sample size. As a result, EVT-based estimates may exhibit instability, wide or poorly characterized uncertainty, and sensitivity to modeling choices. Furthermore, previous studies suggest that tail behavior may vary across traffic environments, challenging the robustness of simple, fixed EVT specifications [6].

Recent literature has increasingly acknowledged these challenges. A comparative review of EVT in traffic conflict analysis concluded that modeling choices strongly affect the estimated crash risk [7], and that practical uptake has been limited because uncertainty in extreme-event estimation can remain large for safety decision-making highlighting that uncertainty and model sensitivity remain a barrier to operational use.

Another challenge is the potential presence of bounded tails in pedestrian surrogate safety measures. Empirical studies have occasionally reported negative GP shape parameters, implying a finite upper bound, which is physically plausible given the geometric and kinematic constraints of vehicle–pedestrian interactions. In practice, however, tail behavior is often uncertain, and assumptions regarding boundedness can influence the stability and interpretation of extrapolated crash risk. This highlights the need for modeling approaches that remain robust under different tail behaviors rather than relying on a single distributional assumption.

Overall, these streams of literature point to a clear research gap. While detailed observations of pedestrian–vehicle interactions are increasingly available, the challenge lies in applying EVT in a way that is robust to the statistical characteristics of such data, including temporal dependence, bounded supports, heterogeneous tails, and limited effective sample sizes. Although extensions such as non-stationary and multivariate EVT have been proposed, their practical application remains challenging and often difficult to interpret for decision-making in road safety. Therefore, there is a need for a unified, data-driven framework that systematically evaluates EVT assumptions, reduces subjectivity in key steps such as threshold selection, enforces independence and tail consistency, quantifies uncertainty, and adapts modeling choices to the statistical properties of the data. This need is addressed through the introduction of a hybrid crash–conflict framework, where hybrid refers to the integration of anomaly-based extreme-event identification, multi-criteria threshold diagnostics, and alternative tail modeling strategies within a single EVT workflow. Modeling choices are adapted to the data rather than relying on a fixed EVT specification, and the framework is demonstrated on naturalistic pedestrian–vehicle data from Athens.

2 Methodology

2.1. Naturalistic data collection and metric extraction

Pedestrian–vehicle data for this study were collected from Vasilissis Sofias–Panepistimiou, a central corridor in Athens with high pedestrian activity and traffic volumes. Video recordings were obtained during peak and off-peak periods, totaling four hours at a signalized intersection, to capture pedestrian–vehicle interactions. The footage was processed with a computer-vision pipeline comprising YOLOv8 for detection of pedestrians and vehicles, ResNet-50 for feature embedding [8], a homography transform to map image coordinates to the ground plane [9], the Hungarian algorithm for data association, and a Kalman filter for trajectory smoothing [10]. The pipeline also incorporated traffic-light color detection to classify crossings by pedestrians and vehicles as legal or illegal [10]. In addition, it extracted trajectories, speeds, and SSMs such as TTC. Figure 1 shows the crosswalk region of interest (ROI) and the gating scheme.

2.2. EVT Framework

The proposed EVT framework consists of five main stages. In the first stage, relevant safety indicators are selected to capture critical interactions. Candidate extreme events are then identified in the second stage using pre-filtering and anomaly detection techniques, followed by tail-direction restriction and transformation into an EVT-compatible space. In the third stage, the identified events are declustered to enforce approximate statistical independence. In the fourth stage, EVT fit diagnostics and threshold selection procedures are performed to assess model validity. Finally, crash risk and frequency are estimated in the fifth stage based on the fitted distributions. The details of the proposed five-stage framework are outlined below.

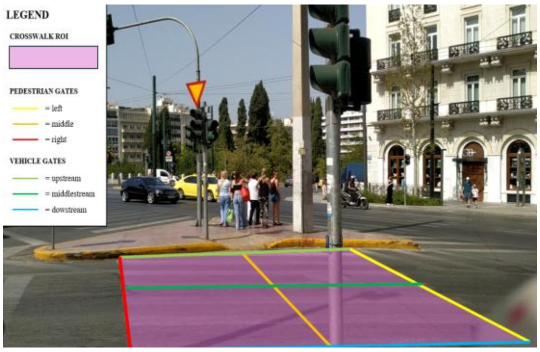

Fig 1. Crosswalk ROI on Vasilissis Amalias Avenue.

Stage 1: Safety indicator

Minimum TTC (TTC_{min}) is chosen as the primary safety indicator for vehicle-pedestrian interactions, as it is well-suited for crosswalks where short horizons drive evasive action or impact likelihood, and enables comparability across heterogeneous agents and speeds. TTC_{min} is computed by projecting detections onto the ground plane via homography and deriving inter-agent distances and relative velocities from Kalman-filtered trajectories, frame by frame at 29.74 frames per second for each pedestrian–vehicle pair.

Stage 2: Exceedance sampling

Anomaly detection is used to recast exceedance selection as an unsupervised outlier detection problem. Three complementary approaches are considered: Isolation Forest (IForest), which isolates rare behaviors through recursive partitioning; the Minimum Covariance Determinant (MCD), which detects multivariate outliers via robust covariance estimation; and a hybrid strategy combining both. Prior to detection, optional conflict prefiltering and one-sided EVT support restriction are applied to exclude non-critical anomalies. For EVT fitting, the method with the highest post-declustering effective sample size is selected, reflecting the number of usable, approximately independent tail events.

Stage 3: Threshold Selection and EVT Diagnostics

After declustering, candidate exceedances are evaluated through a diagnostic threshold sweep to identify tail regions suitable for POT (GP) modeling. Candidate thresholds are assessed using a hybrid scoring framework that combines the Kolmogorov–Smirnov statistic, the Akaike Information Criterion (AIC), the Bayesian Information Criterion (BIC), and cross-validated predictive diagnostics, together with penalties for insufficient exceedances, unstable local tail behavior, and inappropriate tail mass. To improve robustness, a stability-based threshold selection criterion is applied, favoring regions where goodness-of-fit remains consistent across neighboring thresholds. Mean Residual Life (MRL) behavior and parameter stability are used as supporting diagnostics.

Stage 4: Model fitting and fallback strategy

Extreme-value fitting in naturalistic traffic data can become unstable due to limited exceedances, bounded or irregular tails, and violations of model support conditions. To address this, the framework adopts a mode-aware fitting strategy that retains baseline models while augmenting them with auxiliary tail models for comparison and fallback. For the POT branch, exceedances are constructed in EVT space as

$$Y = u - X \, (\textit{left tail}), \tag{1}$$

ensuring $Y \geq 0$ and consistency with one-sided tail modeling. The GP distribution is then fitted to these exceedances,

$$f(y \mid \xi, \sigma) = \frac{1}{\sigma}\left(1+\frac{\xi y}{\sigma}\right)^{-(1+1/\xi)}, y \geq 0, \tag{2}$$

where ξ is the shape parameter and $\sigma > 0$ is the scale parameter, with explicit support checks applied when $\xi < 0$ to enforce a finite upper endpoint.

In parallel, the GEV distribution is fitted in a mode-aware block-maxima branch using likelihood-based estimation. In practice, this fit is obtained through standard maximum-likelihood estimation and, when necessary, refined via bounded-shape likelihood optimization to improve numerical stability.

When baseline GEV/GP fitting is weak or when broader comparison is required, auxiliary tail models are introduced. A Bayesian GP is estimated using Markov chain Monte Carlo (MCMC) inference to obtain posterior distributions of the tail parameters. In addition, flexible parametric tail models are fitted through a Generalized Additive Models for Location, Scale and Shape (GAMLSS) framework applied directly in exceedance space; in the full implementation this is handled via penalized-likelihood estimation, with a likelihood-based surrogate fallback when the full GAMLSS specification is not available.

Stage 5: Crash risk estimation

Crash risk is defined as the probability that the safety indicator (TTC_{min}) reaches a crash-relevant severity level relative to a selected EVT threshold. For a threshold u and crash cutoff x_c, a margin δ ($\delta = u - x_c$) quantifies the position of the crash condition within the tail. For POT-based models, risk is computed in exceedance space as

$$P(X \leq x_c) = \pi_{\text{tail}} \cdot P(Y \geq \delta), \tag{3}$$

where π_{tail} is the empirical tail probability and $P(Y \geq \delta)$ is the model-based survival beyond δ. For GEV models, risk is obtained from the fitted distribution evaluated at x_c. Bayesian GP extends the same formulation by propagating posterior uncertainty to derive credible intervals, while GAMLSS models follow an analogous exceedance-based approach.

In addition to probability estimates, anchor-calibrated crash frequency (ACCF) is derived by translating EVT-based crash risk into an occurrence rate over the observed period through anchor calibration, linking tail probabilities to expected crash counts under the observed interaction intensity. This provides a direct link between extreme-value modeling and safety-relevant metrics suitable for practical interpretation.

3. Results

The framework was applied to 2,844 interaction episodes under a univariate, left-tail configuration using episode-level TTC_{min}. The hybrid anomaly detection method was selected based on effective sample size (ESS ≈ 2,149) and internal consistency. After declustering and preprocessing, 2,348 observations were retained and transformed into EVT space for tail modeling.

Figure 2 presents the hybrid threshold diagnostics, combining the MRL curve with confidence bands and the corresponding number of exceedances $\mathrm{n_{exc}}$, together with complementary metrics including the KS statistic, GOF p-value (model adequacy), a relative tail-risk proxy, and information criteria (AIC and BIC). The lower panels show the evolution of the GP shape $\xi(\mathrm{u})$ and scale $\sigma(\mathrm{u})$ parameters across candidate thresholds, with the selected threshold indicated by a vertical dashed line. A total of 12 candidate thresholds were evaluated. The selected threshold, $\mathrm{u_z} = 0.971$, corresponds to a physical cutoff of $\mathrm{u_x} = 1.013$ s. At this point, diagnostics remain acceptable (KS = 0.084, AIC = 141.91, BIC = 145.89), with 54 exceedances ($\pi \approx 0.023$). As shown in Figure 2, the selected threshold lies within a transitional tail region. The MRL curve begins to stabilize beyond the initial non-linear segment, while higher thresholds lead to a rapid reduction in exceedances and increasingly restrictive π_{tail} constraints. The hybrid diagnostics reflect this trade-off: information criteria improve slightly at higher thresholds, but at the cost of reduced sample support. In this region, $\sigma(\mathrm{u})$ decreases smoothly and $\xi(\mathrm{u})$ increases gradually without abrupt changes, indicating entry into a consistent tail regime, albeit with limited data

Table 1 summarizes the left-tail fits at the selected threshold (u = 1.013 s). The whole-distribution GEV provides a stable benchmark, fitted on block maxima from the full sample, with moderate tail agreement and a bounded risk estimate (0.069; 95% CI: 0.045–0.100), corresponding to ACCF of 1.903 crashes/year (95% CI: 0.519–4.873). For reference, the observed pedestrian–vehicle crash frequency at the site was 0.571 crashes/year (95% CI: 0.156–1.463). In this setting, the block-maxima formulation appears to overestimate risk relative to exceedance-based models, likely because it is less targeted to the lower tail of interest and more sensitive to limited effective sample size. Conditioning on exceedances via the GP substantially improves tail fit, yielding a lower risk (0.021) and ACCF of 0.572 crashes/year (95% CI: 0.156–1.466). Its main limitation, however, lies not in the central estimate itself, but in the lack of reliable asymptotic uncertainty quantification under sparse exceedances. The Bayesian GP preserves these central estimates almost exactly (risk ≈ 0.021; ACCF ≈ 0.571, 95% CI: 0.156–1.463), while providing coherent posterior-based uncertainty quantification, making it the most reliable formulation in the present analysis. The GAMLSS fallback remains numerically well-defined but shows weaker tail agreement (KS ≈ 0.145) and a more conservative estimate (risk ≈ 0.006; ACCF ≈ 0.169, 95% CI: 0.046–0.433), suggesting that the selected family was less compatible with the observed tail structure than the GP-based models in this application. This should not be interpreted as a limitation of GAMLSS itself, since a different family choice could plausibly yield stronger tail representation. Overall, the results indicate that moving from global to threshold-based EVT improves tail fidelity,

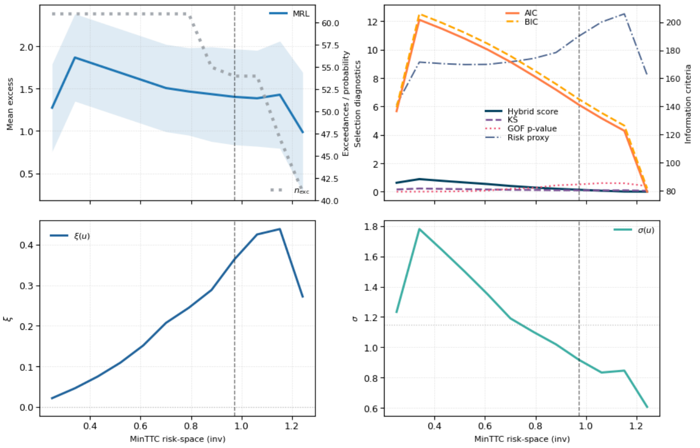

Fig. 2. Threshold diagnostics in TTC_{min} risk space (inverse scale).

while Bayesian regularization strengthens inferential reliability, and that robust pedestrian safety assessment requires balancing tail sensitivity, statistical stability, and model parsimony.

4. Conclusions

The proposed framework delivers a diagnostic-rich and adaptive approach for modeling extreme pedestrian–vehicle conflicts from naturalistic trajectory data. Its key contribution is a transparent, data-driven EVT workflow combining independence checks, declustering, and a diagnostic threshold sweep that reduces reliance on subjective threshold selection. By systematically evaluating admissible thresholds and alternative tail formulations, the framework improves the stability and reproducibility of extreme-value inference under sparse and noisy observations, enabling consistent interpretation of how modeling choices influence risk estimates. The resulting risk and frequency estimates are suitable for operational use, supporting prioritization of crosswalks, targeted interventions (e.g., signal timing, speed management, refuge design), and staged implementation under uncertainty. Because diagnostic checks and modeling choices are explicit, before–after changes are more likely to reflect interventions rather than artefacts, and estimates can be updated as new data become available. These risk estimates can further be translated into expected crash numbers by incorporating exposure information and aligning with observed crash records, enabling outcome-based targets, performance tracking, and proportionate cost–benefit appraisal.

The study has limitations. Evidence is drawn from a single site with a modest number of independent clusters and tail exceedances, so external validity may be limited. Estimates remain sensitive to the quality of trajectory extraction and TTC measurement, to assumptions of local stationarity and independence. The analysis is univariate (TTC_{min}) and site-covariate effects are not modeled; non-stationary or multivariate EVT could refine inference in future multi-site applications.

Table 1. EVT models comparison

	GEV	GP	Bayesian GP	GAMLSS
Threshold (s)	N/A	1.013	1.013	1.013
Shape	0.350	0.332	0.417	2.001
Scale	0.756	0.947	0.964	2.330
Location	1.094	N/A	N/A	N/A
KS Statistic	0.179	0.084	N/A	0.145
AIC*	80.104	141.908	N/A	142.003
Estimated risk	0.069	0.021	0.021	0.006
Risk Lower bound	0.045	N/A	0.020	N/A
Risk higher bound	0.100	N/A	0.021	N/A
ACCF	1.903	0.572	0.571	0.169

* GEV AIC is computed on the full sample, whereas tail-model AICs use exceedances only; they are not directly comparable across likelihood spaces.

References

1. Ankunda, A., Ali, Y., Mohanty, M.: Pedestrian crash risk analysis using extreme value models: New insights and evidence. Accident Analysis & Prevention. 203, 107633 (2024).
2. Arun, A., Haque, M.M., Washington, S., Sayed, T., Mannering, F.: A systematic review of traffic conflict-based safety measures with a focus on application context. Analytic methods in accident research. 32, 100185 (2021).
3. Niu, D., Sayed, T.: Short-term conflict-based crash risk forecasting: A Bayesian conditional peak-over-threshold approach. Analytic Methods in Accident Research. 46, (2025).
4. Yue, Q., Guo, Y., Sayed, T., Zheng, L., Lyu, H., Liu, P.: Bayesian hierarchical non-stationary hybrid modeling for threshold estimation in peak over threshold approach. arXiv preprint arXiv:2503.14839. (2025).
5. Murphy, C., Tawn, J.A., Varty, Z.: Automated threshold selection and associated inference uncertainty for univariate extremes. Technometrics. 67, 215–224 (2025).
6. Tahir, H.B., Haque, M.M.: A non-stationary bivariate extreme value model to estimate real-time pedestrian crash risk by severity at signalized intersections using artificial intelligence-based video analytics. Analytic Methods in Accident Research. 43, 100339 (2024).
7. Ali, Y., Haque, M.M., Mannering, F.: Assessing traffic conflict/crash relationships with extreme value theory: Recent developments and future directions for connected and autonomous vehicle and highway safety research. Analytic methods in accident research. 39, 100276 (2023).
8. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 770–778 (2016).
9. Hartley, R., Zisserman, A.: Multiple view geometry in computer vision. Cambridge university press (2003).
10. Ventura, R., Roussou, S., Ziakopoulos, A., Barabino, B., Yannis, G.: Using computer vision and street-level videos for pedestrian-vehicle tracking and behaviour analysis. Transportation Research Interdisciplinary Perspectives. 30, 101366 (2025).