

A copula-based spatial justice index for pedestrian-vehicle interaction on school streets

Yuri Calleo

yuri.calleo@unimercatorum.it

Mercatorum University

Federica Mastro

Polytechnic University of Bari

Nadia Giuffrida

Polytechnic University of Bari

Michele Ottomanelli

Polytechnic University of Bari

Article

Keywords: Spatial justice, Road safety, Pedestrian-vehicle interaction, Copula models, Citizen science, Sustainable Development Goal 11.

Posted Date: September 25th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11142595/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

A copula-based spatial justice index for pedestrian-vehicle interaction on school streets

Yuri Calleo^{1*}, Federica Mastro², Nadia Giuffrida²,
Michele Ottomanelli²

^{1*}Department of Economics, Statistics, Business, Universitas
Mercatorum, Rome, Italy.

²Department of Civil, Environmental, Land, Construction and
Chemistry, Polytechnic University of Bari, Bari, Italy.

*Corresponding author(s). E-mail(s): yuri.calleo@unimercatorum.it;

Abstract

Streets around schools are where children meet motor traffic, and making them safe is central to Sustainable Development Goal 11 on safe, inclusive and sustainable cities. Yet communities rarely have tools to measure whether drivers actually slow down when pedestrians are present. We propose a Spatial Justice Index that uses low-cost citizen-science traffic sensors to estimate how the probability of cautious driving changes with the number of pedestrians on a street. The index combines a filter that excludes congested periods, a low percentile of the speed distribution as a measure of driver caution, and a statistical model that links pedestrian counts to speed while accounting for the discrete nature of counts. Applied to two Italian school streets monitored during 2025, the index rose from 0.47 to 0.88 across the observed range of pedestrian activity in Andria, with most of the change occurring between empty and lightly used periods rather than at high crowd densities. In Milan, pedestrian activity was too sparse for any association to be detected. The index offers communities a scalable way to monitor how street space is shared, and it signals explicitly when the available data cannot answer the question.

Keywords: Spatial justice, Road safety, Pedestrian-vehicle interaction, Copula models, Citizen science, Sustainable Development Goal 11.

047 Introduction

048
049 Road traffic crashes involve about 1.19 million people each year, and road traffic injury
050 is the leading cause of death for children and young people aged 5-29 years [1]. The
051 2030 Agenda addresses this burden directly in Sustainable Development Goal (SDG)
052 11: target 11.2 calls for safe and sustainable transport systems that improve road safety
053 “with special attention to the needs of those in vulnerable situations, women, children,
054 persons with disabilities and older persons”, and target 11.7 calls for universal access
055 to safe, inclusive and accessible public spaces, in particular for children [2]. The streets
056 around schools are where these two targets meet. They are public spaces used daily
057 by children on foot, and they are also links in the road network, where the physical
058 space of the city is contested between competing mobility needs [3]. Whether such
059 streets are safe depends, to a large extent, on how drivers behave when pedestrians
060 are present - and that is precisely what local communities and administrations usually
061 cannot measure.

062 In traffic engineering, the presence of pedestrians, cyclists and curbside activity is
063 conventionally treated as “*side friction*”: an impedance that reduces link capacity and
064 vehicle speed, to be minimised in the interest of throughput [4–6]. Urban scholarship
065 on spatial justice proposes a different reading [7, 8]. From this perspective, the same
066 slowdown is evidence of spatial negotiation: by occupying the street edge, vulnerable
067 road users project a “*visual wall*” that induces drivers to moderate their speed, in
068 the same way that street-edge activity and enclosure are known to reduce driving
069 speeds [9, 10]. Under this reading, side friction is not an operational nuisance but a
070 measurable indicator of how fairly street space is shared - and a quantity that SDG-
071 oriented street design seeks to make independent of whether pedestrians are numerous
072 enough to command it.

073 Turning this idea into a community-level measurement faces three methodologi-
074 cal obstacles. (1) The first concerns the scale of observation and the choice of metric.
075 Microscopic studies of pedestrian - vehicle interaction reconstruct individual trajec-
076 tories and crossing decisions with high fidelity [11–13], but require data that are costly
077 and rarely available beyond a few sites. Macroscopic studies use the aggregated counts
078 and speeds that fixed sensors routinely provide [14, 15], but typically summarize speed
079 by its mean or median[16], which smooths out the localized decelerations through
080 which caution is expressed; aggregation makes such approaches comparatively insen-
081 sitive to behavioral perturbations produced by active mobility [17–19]. (2) The second
082 obstacle is confounding by congestion. When pedestrian activity and vehicular density
083 peak together, as they do at school entry and exit times, a speed reduction may reflect
084 interaction between vehicles rather than with pedestrians, and the effect of roadside
085 activity is known to vary with traffic volume [5, 6]. (3) The third obstacle is statistical.
086 Most studies link pedestrian activity to speed through linear regression or symmetric
087 dependence measures [5, 6], which cannot represent a mechanism that might be negli-
088 gible at low pedestrian levels and binding above a threshold - a pattern that field studies
089 of side friction have reported [5]. Copula models, which separate the marginal behav-
090 ior of each variable from their dependence, have proved valuable across transport and
091 safety applications for exactly this reason [19–34], including settings where congestion,
092 road-user mixing and behavioral adaptation overlap [18, 35, 36]. However, transport

data often combine discrete counts with continuous indicators [22, 23, 37], and with discrete margins the copula is not uniquely identified, and rank-based methods are affected by ties [21, 38, 39].

At the same time, the data needed for community-level monitoring are becoming widely available. Low-cost sensors installed by residents in their own windows now record multimodal traffic counts and speed distributions on thousands of streets, as part of citizen-science programs that place the production of mobility evidence in the hands of the communities concerned [40]. These data are aggregated and coarse, but they are continuous in time, spatially widespread and cheap - the properties that SDG monitoring at the neighborhood scale requires.

This paper proposes a probabilistic framework that turns such data into a Spatial Justice Index (SJI). The framework adopts a regime filter that retains intervals with vehicles present but without congestion, so that speed reductions are less likely to reflect interaction between vehicles. The 15th percentile of the interval speed distribution, reconstructed from binned sensor data, serves as a “cautionary speed” that targets the slowest part of the traffic stream, where caution is expected to appear. Finally, a copula model links pedestrian counts, made continuous by jittering [41, 42], to cautionary speed. In this model, a rotated Clayton copula that places dependence in the joint high-pedestrian/low-speed corner - the signature of a threshold-type visual wall - is tested against a symmetric Gaussian benchmark rather than assumed. The SJI is then defined as the conditional probability that cautionary speed falls below the street’s own median, as a function of pedestrian count, together with its gradient. We apply the framework to two Italian school streets monitored by citizen-science sensors during 2025. The results show that the index is informative where pedestrian exposure varies, that the dependence it captures is located at low pedestrian counts rather than in a high-density tail, and that the framework identifies explicitly the streets on which the question cannot be answered with the available data.

Results

Two school streets with contrasting speed regimes

The two monitoring sites are Via XX Settembre in Andria, a medium-sized city in Apulia, and Via Gentilino in Milan; both are streets adjacent to schools (see Methods). The raw data describe two structurally different streets (Table 1). Via XX Settembre operates as an urban traffic corridor, with an average of 73.8 vehicles and 10.3 pedestrians per 15-minute interval and an average 85th percentile operating speed (V_{85}) of 43.6 km/h. Via Gentilino behaves as a strongly traffic-calmed street, with 13.2 vehicles and 0.7 pedestrians per interval on average and an average V_{85} of 14.4 km/h (Supplementary Fig. S1). At Via Gentilino, 45% of raw intervals record no vehicles and most record no pedestrians, whereas Via XX Settembre has a substantially longer right tail of pedestrian activity (Supplementary Fig. S2). The Andria data also contain a small number of implausible V_{85} values above 100 km/h, which we attribute to detection errors in low-volume intervals.

The regime filter affects the two streets in opposite directions. At Via Gentilino, the free-flow mean volume (21.4) exceeds the raw mean (13.2) because the filter removes

Table 1 Descriptive statistics of the two study streets, before (raw) and after (free-flow) regime filtering. Values are means per 15-minute interval unless stated otherwise; n , number of intervals; SD, standard deviation; V_{85} , 85th percentile operating speed; V_{15} , cautionary speed (15th percentile of the interval speed distribution). The free-flow regime retains intervals with $0 < Q_t < 50$ vehicles.

	Andria (Via XX Settembre)	Milan (Via Gentilino)
<i>Raw data</i>		
Intervals, n	4,845	27,964
Monitoring days	53	297
Vehicles	73.8	13.2
Pedestrians	10.3	0.7
V_{85} (km/h)	43.6	14.4
<i>Free-flow regime</i>		
Intervals, n (% of raw)	1,756 (36%)	14,199 (51%)
Vehicles	10.9	21.4
Pedestrians (maximum)	2.05 (46)	1.23 (73)
Intervals with no pedestrians (%)	52.7	57.7
V_{15} (km/h), mean (SD)	26.94 (16.71)	2.62 (2.25)

the many zero-volume intervals; at Via XX Settembre, the free-flow mean (10.9) is far below the raw mean (73.8) because the filter removes the many high-volume intervals that characterise the corridor. The Andria free-flow sample is therefore drawn from the low-traffic periods of the day and week, a feature that matters for interpretation (Discussion).

Within the free-flow regime, cautionary speed differs sharply between the streets (Fig. 1). At Andria, V_{15} has a mean of 26.94 km/h and is widely dispersed. At Milan, 97% of free-flow intervals place V_{15} in the lowest (0-10 km/h) speed bin, with a strong concentration near 1.5 km/h, the value returned by the reconstruction when all detected cars fall in that bin (Methods). The low Milan values therefore reflect a structurally slow street, measured on a scale whose effective resolution is very coarse.

Cautionary speed declines with the first pedestrians at Andria

At Andria, cautionary speed decreases as pedestrian activity increases (Fig. 2a). The locally weighted regression (LOESS) [43] curve falls from about 30 km/h in intervals without pedestrians to a minimum of about 12 km/h at 15-20 pedestrians; the apparent rebound beyond about 25 pedestrians rests on very few intervals and lies within a wide confidence band. The tie-corrected Kendall's rank correlation between pedestrian count and cautionary speed is $\tau_b = -0.203$ ($n = 1,756$; two-sided test, $P < 10^{-29}$). At Milan, cautionary speed is compressed near the bottom of its range at all pedestrian levels (Fig. 2b), and the association is not distinguishable from zero ($\tau_b = -0.003$, $n = 14,199$, $P = 0.63$). These P values assume independent intervals and are therefore optimistic (Methods); at Andria, however, the association is strong enough that serial dependence is unlikely to account for it.

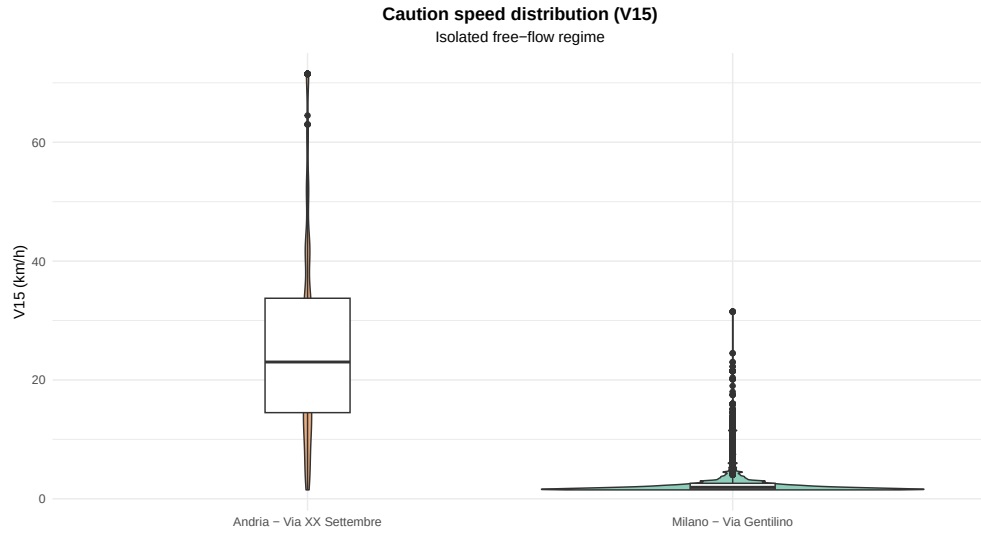


Fig. 1 Cautionary speed under free-flow conditions. Distribution of the cautionary speed V_{15} (15th percentile of the car speed distribution in each 15-minute interval) at Andria ($n = 1,756$ intervals) and Milan ($n = 14,199$ intervals). Violins show kernel density estimates; internal boxplots show the median and the interquartile range.

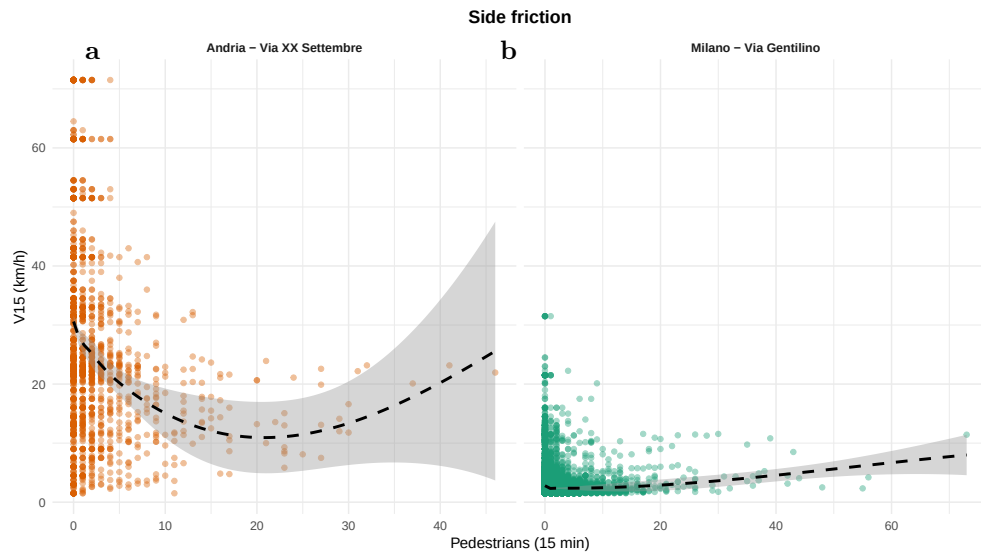


Fig. 2 Cautionary speed against pedestrian count under free-flow conditions. **a**, Andria ($n = 1,756$ intervals). **b**, Milan ($n = 14,199$ intervals). Each point is a 15-minute interval. Dashed lines show LOESS fits and grey bands their 95% confidence intervals. Horizontal bands of points (for example at 41.5, 51.5 and 61.5 km/h) correspond to intervals in which all detected cars fall in a single 10 km/h speed bin.

The dependence is not concentrated in the high-density tail

The copula model was fitted separately at each street (Table 2). Jittering the pedestrian counts behaved as expected. At Andria, the jittered Kendall's τ averaged -0.168 (SD 0.006 over $m = 50$ replications), smaller in absolute value than the raw τ_b by a factor of 0.83 . This attenuation is predicted by the tie structure of the counts: with 52.7% of intervals recording no pedestrians, the factor cannot exceed $\sqrt{1 - 0.527^2} \approx 0.85$ (equation (4) in Methods). The small standard deviation shows that the result does not depend on a particular jitter draw.

Table 2 Model selection and validation metrics. Kendall's τ is reported for the raw (tie-corrected τ_b) and jittered data (mean $\pm$ SD over $m = 50$ replications). The dependence parameter $\hat{\theta}$ of the rotated Clayton copula is reported with a 95% percentile bootstrap interval ($B = 150$ resamples). The implied Kendall's tau is $-\hat{\theta}/(\hat{\theta} + 2)$ and the tail dependence coefficient is $2^{-1/\hat{\theta}}$. The Vuong statistic Z compares the rotated Clayton with the Gaussian copula fitted to the same data; negative values favour the Gaussian. P values are two-sided.

	Milan	Andria
τ_b , raw (P)	-0.003 (0.63)	-0.203 ($< 10^{-29}$)
τ , jittered (mean $\pm$ SD)	-0.002 ± 0.003	-0.168 ± 0.006
Margin for V_{15} (AIC)	Lognormal ($41,676.8$)	Weibull ($14,499.6$)
Margin for jittered pedestrian count	Gamma	Gamma
Rotated Clayton $\hat{\theta}$ (95% interval)	0.0001 (boundary)	0.137 (0.104 - 0.191)
Implied Kendall's tau	≈ 0.000	-0.064
Tail dependence $\hat{\lambda}$	≈ 0	0.006
AIC, rotated Clayton / Gaussian	$2.15 / 1.94$	$-35.74 / -101.34$
Cramér-von Mises distance S_n	64.9	1.9
Vuong Z (P)	-0.40 (0.69)	-6.40 (1.6×10^{-10})

At Via Gentilino, the estimated dependence collapsed to independence: the jittered τ is indistinguishable from zero, the bootstrap distribution of $\hat{\theta}$ sits at the lower boundary of the parameter space, and the comparison with the Gaussian copula is uninformative (Vuong $Z = -0.40$, $P = 0.69$). The large Cramér-von Mises distance at this site does not indicate a missed dependence. It most plausibly reflects the discreteness of the Milan speed data: with V_{15} concentrated at a few values, no continuous parametric margin can reproduce the point masses, and the empirical copula departs from any continuous model.

At Via XX Settembre, the rotated Clayton parameter is estimated away from independence ($\hat{\theta} = 0.137$, bootstrap interval 0.104 - 0.191), and its density is highest in the intended high-pedestrian/low-speed corner (Fig. 3b). However, three diagnostics indicate that the dependence is not located where the threshold hypothesis places it. First, the symmetric Gaussian copula fits markedly better: its AIC is lower by 65.6 units, and the Vuong test favours it ($Z = -6.40$, $P = 1.6 \times 10^{-10}$). Second, the Kendall's tau implied by the fitted rotated Clayton (-0.064) accounts for less than 40% of the observed jittered concordance (-0.168). Third, the implied tail dependence ($\hat{\lambda} = 0.006$) is negligible. A one-parameter family that can only place dependence

in the high-pedestrian tail therefore cannot reproduce the dependence in these data, which is spread over the body of the distribution or located at low pedestrian counts. This agrees with the raw data, in which most of the decline in cautionary speed occurs between zero and a few pedestrians (Fig. 2a).

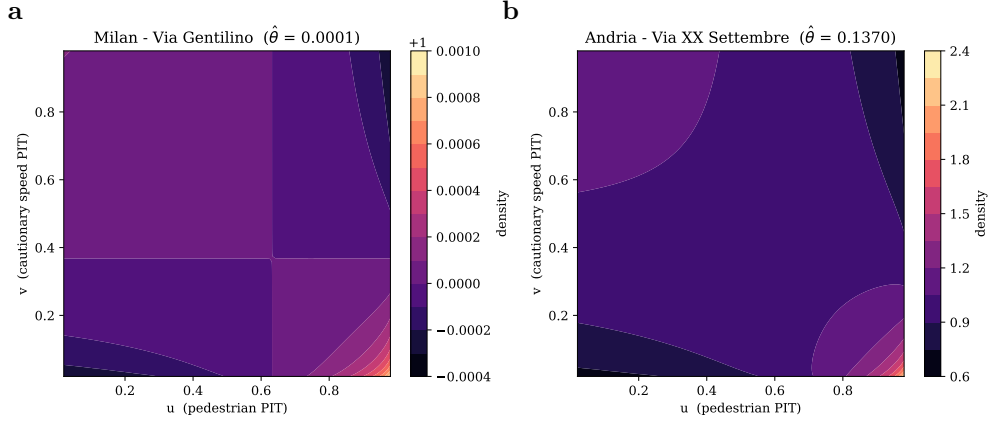


Fig. 3 Fitted rotated Clayton copula densities. **a**, Milan ($\hat{\theta} \approx 0$). **b**, Andria ($\hat{\theta} = 0.137$). The horizontal axis (u) is the probability integral transform of the jittered pedestrian count and the vertical axis (v) that of the cautionary speed. In **b** the density is highest in the lower-right corner (many pedestrians, low speed), as intended by the construction; in **a** it is flat to within 10^{-3} (note the colour scale, centred on 1).

The Spatial Justice Index

Figure 4 shows the SJI and its gradient at both streets. The SJI is the probability that cautionary speed falls below the street’s own median, so it equals 0.5 when pedestrian presence is uninformative. At Via Gentilino, where no dependence is estimated, the curve is flat at 0.5 by construction. This is the signature of an inconclusive case study, not of a street where safety is guaranteed regardless of pedestrian presence. That stronger claim may well be true of Via Gentilino, given its generally low speeds, but it cannot be tested there, because pedestrian activity hardly varies.

At Via XX Settembre, the SJI rises from about 0.47 with no pedestrians to about 0.88 at the largest observed count (46 pedestrians per 15 minutes): the probability of cautious driving increases materially with pedestrian activity. The curve is concave, and its gradient is largest near zero and declines thereafter (Fig. 4b). This shape must be read with care. On the pedestrian-count scale, the gradient is the product of a dependence factor and the density of pedestrian counts, which is strongly right-skewed and decreasing (equation (9) in Methods). The concavity is therefore partly inherited from the distribution of counts and does not, on its own, distinguish between mechanisms. The direct evidence on mechanism comes from the copula diagnostics and the raw data, which together indicate that driver caution at Andria is associated mainly with the difference between empty and non-empty intervals, with little evidence of an additional effect that switches on above a high pedestrian threshold. The upper

end of the curve rests on few intervals and is determined largely by the fitted models, so it should not be read as a precisely estimated asymptote.

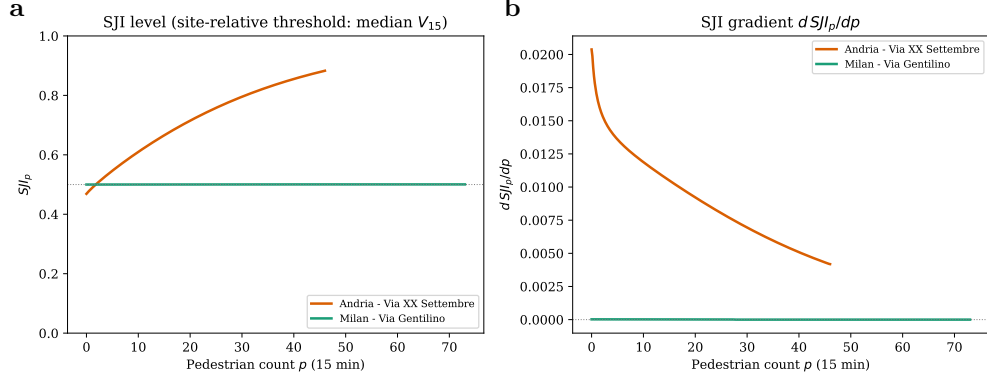


Fig. 4 Spatial Justice Index. **a**, Probability that the cautionary speed falls below the street’s median value, as a function of the number of pedestrians per 15-minute interval (SJI), based on the rotated Clayton copula. **b**, Gradient of the SJI with respect to pedestrian count. Orange, Andria; green, Milan. The Milan curve is flat at 0.5 because no dependence is estimated at that street. The shape of the Andria gradient near zero reflects the distribution of pedestrian counts (Methods).

Finally, low cautionary speeds occur both in intervals whose operating speed V_{85} exceeds 30 km/h and in those below it (Supplementary Fig. S3). The SJI and V_{85} therefore describe different aspects of street performance.

Discussion

The two school streets studied here tell different stories, and the contrast between them is informative for both the SDG agenda and the design of the index. At Via XX Settembre, drivers are markedly more likely to drive cautiously when pedestrians are present, and most of this change occurs between empty and lightly used intervals. This finding carries a specific implication for SDG targets 11.2 and 11.7. Where speed moderation depends on pedestrian presence, safe conditions are least available precisely when pedestrians are few - for example, when a single child walks to school early or leaves late. A street of this kind offers protection that is contingent on the protected party being present to command it, and it is the street’s baseline speed regime, not the crowd, that governs risk during the low-exposure periods that make up most of the day. The evidence does not support the stronger, threshold-type form of the visual-wall hypothesis, in which caution switches on only when pedestrians are numerous: the copula that encodes that mechanism is outperformed by a symmetric benchmark and reproduces less than half of the observed concordance. If there is a visual wall at Andria, it appears to be raised by the first pedestrians rather than by a crowd.

An alternative explanation must be weighed. Because the regime filter retains only intervals with fewer than 50 vehicles, the Andria free-flow sample is drawn from low-traffic periods, when the absence of pedestrians and higher speeds may share common causes such as time of day, lighting and the school timetable. Filtering on vehicle

volume does not remove this form of confounding, and part of the association between empty intervals and faster cautionary speeds may reflect it. The associations reported here are conditional associations within a controlled volume regime, not causal effects. Separating a pedestrian-presence effect from its temporal correlates requires analyses stratified by hour and day type, which the framework accommodates directly and which are the most important extension of this work.

At Via Gentilino, the framework returns no identifiable dependence, and we have distinguished this throughout from a finding of structural safety. A flat index is what one would observe both on a street whose design enforces low speeds regardless of who is present and on a street where pedestrian exposure is too sparse for an effect to be estimated. Via Gentilino is plausibly the former, but with 57.7% of free-flow intervals recording no pedestrians and a monitoring period that includes school holidays, the data cannot tell the two apart. This has a methodological corollary that we regard as a contribution in its own right: indices of this family are identifiable only where the conditioning variable varies. Defining the threshold relative to each street’s own median makes this failure visible. Under a common absolute threshold, the Milan index would have been close to one everywhere and would have invited exactly the reassuring interpretation that the data do not support. Communities and administrations applying the SJI should therefore report pedestrian-exposure diagnostics alongside the index and treat a flat curve on a low-exposure street as a null result.

From the perspective of sustainable communities, the framework’s main value lies in what it asks of the data and of the people who produce them. It requires only binned speeds and modal counts from sensors that residents install in their own windows [40], so the evidence on whether a school street is shared fairly can be generated, inspected and contested by the community that uses it. This supports the kind of community-level measurement and citizen participation that local SDG monitoring needs, and it complements conventional indicators. The operating speed V_{85} describes how fast traffic moves; the SJI describes whether that speed responds to the people who share the street. A street may perform acceptably on V_{85} while showing a steep SJI gradient, which indicates that its safety depends on pedestrian presence rather than on design. This reading is consistent with Vision Zero and traffic-calming principles, which hold that safety should be built into street geometry rather than left to negotiation between the parties at risk [9, 10].

Several limitations qualify these findings. Cautionary speed is reconstructed from 10 km/h bins of car speeds rather than observed for individual vehicles, so its resolution is coarse where speeds are concentrated in a single bin, as at Via Gentilino, and in intervals with few cars it is determined by one or two detections; sensitivity analyses excluding intervals below a minimum number of cars, and validation against vehicle-level speed data, would strengthen the construct. Camera-based detection is sensitive to lighting, occlusion and calibration, as the implausible operating speeds at Andria illustrate; restricting the analysis to daylight, high-uptime intervals is a natural robustness check, as is documenting the stability of the results over alternative values of the free-flow threshold $Q_{\max}$. The Andria data cover 53 days in two windows (February-March and November-December 2025), so seasonal variation is unobserved; conversely, restricting the Milan data to school days would sharpen the comparison

between the streets. Consecutive 15-minute intervals are serially dependent, so the reported P values and the interval-level bootstrap understate uncertainty; resampling whole days would give more reliable intervals [44]. The Gamma margin for pedestrian counts is applied to a variable with more than half of its mass at zero, and it governs the upper end of the index; rank-based estimation [45] or the exact margin implied by the jittered empirical counts would avoid this assumption. Finally, the family comparison was restricted to the hypothesis-driven rotated Clayton and a Gaussian benchmark. A complete comparison should include copulas that place dependence in the low-pedestrian/high-speed corner and should use calibrated goodness-of-fit tests [46], and model-based curves should be compared with nonparametric estimates of the SJI within classes of pedestrian count and reported also on the probability scale, where their shape depends only on the copula. Two school streets in two Italian cities cannot support general claims about street typologies. The framework, however, is portable to any street monitored by comparable sensors.

In summary, the Spatial Justice Index turns citizen-collected traffic data into a transparent measure of whether drivers yield space to pedestrians. At the street where pedestrian exposure varies, it reveals a clear association between pedestrian presence and cautious driving, concentrated at low pedestrian counts; at the street where exposure is sparse, it declines to give an answer rather than giving a reassuring one. Both outcomes are useful for communities working towards safe and inclusive streets under SDG 11.

Methods

Study sites and data

The first site is Via Gentilino in Milan, at the entrance of the Istituto Comprensivo “P. Thouar e L. Gonzaga”. The street lies in a dense urban area with a compact street network, where local traffic interacts with public transport, pedestrian flows and metropolitan mobility. The school entrance opens directly onto the street, making it a sensitive node where pedestrian flows, cycling and local vehicular circulation converge at school arrival and departure times. Nearby parks and green areas may further encourage active mobility, and the area has been involved in initiatives promoting safer school streets and traffic calming. The second site is Via XX Settembre in Andria (Apulia), near the Primo Circolo Didattico “Guglielmo Oberdan”. Andria has a compact urban core surrounded by a wider peri-urban and rural municipal territory. Via XX Settembre is an urban street in a central and well-connected part of the local network, close to more heavily trafficked roads and main circulation axes, which may contribute to higher vehicular volumes; pedestrian flows and local traffic converge there during school hours.

At both sites, data were collected by Telraam sensors, low-cost devices installed inside a window with an unobstructed view of the street [40]. The sensor detects road users in its field of view, tracks them across frames, classifies them by mode and converts detections into anonymous counts, aggregated at 15-minute resolution for pedestrians, cyclists, cars and heavy vehicles. Pedestrian counts refer to all pedestrians detected in the field of view, on either side of the street. The sensor does not report

individual speeds. For each interval, it reports the share of detected cars in eight 10 km/h speed bins (0-10, 10-20, ..., 70+ km/h) and the 85th percentile operating speed V_{85} . The Milan dataset contains 27,964 records collected between January and 12 December 2025 on 297 distinct days (about 6,991 hours). The Andria dataset contains 4,845 records collected between 19 February and 22 December 2025 on 53 days (about 1,211 hours), concentrated in two windows (19 February-19 March and 28 November-22 December), with no observations from April to October. The study used only aggregated, anonymous counts; no personal data were collected.

Cautionary speed

Let $\mathcal{T} = \{1, \dots, N\}$ index the 15-minute intervals at a site, let $f_{t,k}$ be the share of cars in speed bin $k = 1, \dots, 8$ during interval t , with bin boundaries $b_0 = 0 < b_1 = 10 < \dots < b_8 = 80$ km/h (b_8 is a nominal upper bound for the open 70+ bin), and let $F_{t,k} = \sum_{j \leq k} f_{t,j}$ with $F_{t,0} = 0$. Interpolating linearly within bins gives a piecewise-linear approximation $\hat{F}_{V_t}$ to the distribution function of car speeds in interval t . The cautionary speed is its 15th percentile,

$$V_{15,t} = \inf\{v \geq 0 : \hat{F}_{V_t}(v) \geq 0.15\} = b_{k^*-1} + \frac{0.15 - F_{t,k^*-1}}{f_{t,k^*}} (b_{k^*} - b_{k^*-1}), \quad (1)$$

where k^* is the first bin with $F_{t,k^*} \geq 0.15$. We use the 15th percentile rather than the mean because it targets the slowest part of the traffic stream, where cautionary slowdowns are expected, and it is insensitive to speeding vehicles. Equation (1) implies that $V_{15,t}$ carries a discretization error bounded by the bin width, and that $V_{15,t} = b_{k-1} + 1.5$ km/h whenever all detected cars fall in the same bin k , which is frequent when an interval contains few cars. The cautionary speed therefore has point masses at a small set of values.

The pedestrian variable $P_t \in \{0, 1, 2, \dots\}$ is the number of pedestrians detected during interval t .

Regime filtering

To reduce confounding by interaction between vehicles, we retained only intervals in a free-flow regime. With Q_t the total motorized vehicle count in interval t and $Q_{\max} = 50$ vehicles per 15 minutes,

$$R_t = \begin{cases} 1, & 0 < Q_t < Q_{\max}, \\ 0, & \text{otherwise,} \end{cases} \quad \mathcal{T}^* = \{t \in \mathcal{T} : R_t = 1\}, \quad n = |\mathcal{T}^*|. \quad (2)$$

The upper bound (200 vehicles per hour) is a conservative value for a single urban street, below which interactions between successive vehicles are expected to be infrequent. The lower bound excludes intervals with no recorded vehicles, which comprise both genuinely empty periods and periods of reduced detection, and for which the speed distribution is undefined. Filtering on a volume band reduces, but does not eliminate, confounding: it does not control for time of day or the school timetable, and it

507 is a form of selection on an endogenous variable. All analyses below use the free-flow
 508 subset $\mathcal{T}^*$.

509

510 Continuization of counts by jittering

511

512 By Sklar’s theorem [47, 48], any bivariate distribution H with margins F_P and F_V
 513 can be written as $H(p, v) = C\{F_P(p), F_V(v)\}$ for a copula C , which is unique only
 514 when both margins are continuous. Because P_t is discrete, its copula with $V_{15,t}$ is not
 515 identified [38]. We therefore used the continuous extension (jittering) [41, 42]

516

$$517 \quad P_t^* = P_t + U_t, \quad U_t \sim \mathcal{U}(0, 1), \quad (3)$$

518

519 with U_t independent of $(P_t, V_{15,t})$ and across intervals. The event $\{P_t = p\}$ coincides
 520 with $\{P_t^* \in [p, p + 1)\}$, and (P^*, V_{15}) has a unique copula.

521

522 Jittering fixes the estimand rather than removing the effect of ties. Pairs tied in
 523 P are broken at random by independent noise and contribute no net concordance, so
 524 the population Kendall’s tau of the jittered pair, defined as the difference between the
 525 probabilities of concordance and discordance, satisfies

$$526 \quad \tau(P^*, V_{15}) = \tau_b(P, V_{15}) \sqrt{\left(1 - \sum_k \pi_k^2\right) \left(1 - \sum_l \omega_l^2\right)}, \quad (4)$$

527

528 where τ_b is the tie-corrected coefficient [49], $\pi_k = \Pr(P = k)$, and ω_l are the prob-
 529 abilities of the atoms of V_{15} (a derivation is given in Supplementary Note 1). If the
 530 tie-corrected version is also used for the jittered pair, the second factor cancels. The
 531 jittered tau is therefore attenuated relative to τ_b by a factor that depends only on the
 532 ties. Because a single jitter draw is random, we repeated the procedure over $m = 50$
 533 independent replications and report the mean and standard deviation of the resulting
 534 Kendall’s tau.

535

536 Marginal distributions and copula model

537

538 For V_{15} we compared Weibull and Lognormal distributions and retained the one with
 539 the lower Akaike information criterion ($\text{AIC} = 2k - 2 \ln L$, with k parameters and
 540 maximized likelihood L) separately at each site. For P^* we used a Gamma distribu-
 541 tion. Estimation followed the two-stage inference-functions-for-margins approach [50]:
 542 marginal parameters were estimated by maximum likelihood, and the copula param-
 543 eter by maximum likelihood on the probability integral transforms $u_t = F_{P^*}(P_t^*)$ and
 544 $v_t = F_V(V_{15,t})$.

545

546 The side-friction hypothesis implies that high pedestrian counts ($u \rightarrow 1$) coincide
 547 with low cautionary speeds ($v \rightarrow 0$). A threshold-type mechanism, in which pedestrian
 548 presence becomes binding only above a critical level, would concentrate dependence
 549 in that corner of the unit square. To represent it, we used a rotated Clayton copula
 550 obtained by fitting the Clayton copula $C_{\text{Cl}}(u', v; \theta) = (u'^{-\theta} + v^{-\theta} - 1)^{-1/\theta}$, $\theta > 0$, to
 551 the reflected pair $(u', v) = (1 - u, v)$:

551

$$552 \quad C(u, v; \theta) = v - C_{\text{Cl}}(1 - u, v; \theta). \quad (5)$$

This copula describes negative dependence, with Kendall's tau $-\theta/(\theta + 2)$ and tail dependence coefficient $\lambda = 2^{-1/\theta}$ in the $(u \rightarrow 1, v \rightarrow 0)$ corner; it reduces to independence as $\theta \rightarrow 0$. Because the same rotation label (for example 90° or 270°) may denote different corners in different treatments and software implementations [50, 51], we defined the construction explicitly and verified numerically that the fitted density is highest in the intended corner (Fig. 3b). The conditional distribution function of V given $U = u$ is

$$h(v | u; \theta) = \frac{\partial C(u, v; \theta)}{\partial u} = (1 - u)^{-\theta-1} [(1 - u)^{-\theta} + v^{-\theta} - 1]^{-1-1/\theta}. \quad (6)$$

As a symmetric benchmark we fitted the Gaussian copula $C_\rho(u, v) = \Phi_2\{\Phi^{-1}(u), \Phi^{-1}(v); \rho\}$ to the same data, where Φ_2 is the bivariate standard normal distribution function with correlation ρ ; it implies no tail dependence.

Spatial Justice Index

Because the two streets differ greatly in their speed regimes, a common absolute threshold would lie outside the support of one street's cautionary speeds and make the index uninformative. We therefore defined the threshold relative to each street, as $v_{\text{crit}} = F_V^{-1}(q)$ with $q = 0.5$ (the street's fitted median cautionary speed). The SJI is the conditional probability

$$SJI_p = \Pr(V_{15} \leq v_{\text{crit}} | P^* = p) = h\{q | F_{P^*}(p); \hat{\theta}\}. \quad (7)$$

Under the rotated Clayton copula, $SJI_p \rightarrow q$ for all p as $\theta \rightarrow 0$, and SJI_p increases monotonically from $q^{1+\theta}$ at $u = 0$ towards 1 as $u \rightarrow 1$ for any $\theta > 0$. The upper end of the curve is therefore determined jointly by the copula family and by the upper tail of F_{P^*} . The curves in Fig. 4 evaluate equation (7) at $P^* = p$. The conditional probability of the discrete event $\{P = p\}$ is the average of h over the corresponding interval of u ,

$$\Pr(V_{15} \leq v_{\text{crit}} | P = p) = \frac{C\{F_{P^*}(p+1), q\} - C\{F_{P^*}(p), q\}}{F_{P^*}(p+1) - F_{P^*}(p)}, \quad (8)$$

which differs most from equation (7) at $p = 0$. Under jittering, the distribution function of P^* is also available exactly from the empirical count distribution, $F_{P^*}(x) = \Pr(P \leq \lfloor x \rfloor - 1) + (x - \lfloor x \rfloor) \Pr(P = \lfloor x \rfloor)$, which offers a model-free alternative to the Gamma margin. We complement the level with its gradient,

$$\frac{dSJI_p}{dp} = f_{P^*}(p) \left. \frac{\partial h(q | u; \hat{\theta})}{\partial u} \right|_{u=F_{P^*}(p)}, \quad (9)$$

which measures the local sensitivity of cautious driving to an additional pedestrian and is comparable across streets with different speed regimes. Equation (9) shows that the gradient on the count scale is the product of a dependence factor and the marginal density f_{P^*} . For a right-skewed count distribution, whose density decreases in p , a

declining gradient can therefore arise even when the dependence factor is constant or increasing.

Validation and model comparison

The validation had four components. (i) We computed the Cramér-von Mises distance between the fitted copula $C_{\hat{\theta}}$ and the empirical copula C_n ,

$$S_n = \sum_{i=1}^n \{C_n(u_i, v_i) - C_{\hat{\theta}}(u_i, v_i)\}^2, \quad C_n(u, v) = \frac{1}{n} \sum_{j=1}^n \mathbf{1}\{u_j \leq u, v_j \leq v\}, \quad (10)$$

as a descriptive measure[46]; a calibrated test would require a parametric bootstrap. (ii) We obtained a 95% percentile interval for $\hat{\theta}$ from a nonparametric bootstrap with $B = 150$ resamples of intervals, re-executing the full pipeline (jittering, marginal fitting, probability integral transform, copula estimation) on each resample. (iii) We compared the Kendall's tau implied by the fitted copula with the empirical jittered tau; for a correctly specified one-parameter family both estimate the same quantity, so a large discrepancy indicates that the family cannot reproduce the observed dependence. (iv) We compared the rotated Clayton and Gaussian copulas by AIC and by Vuong's test for non-nested models[52],

$$Z = \frac{\sqrt{n} \bar{\ell}}{s_{\ell}}, \quad \ell_i = \ln c_{\hat{\theta}}(u_i, v_i) - \ln c_{\hat{\rho}}(u_i, v_i), \quad (11)$$

where $c_{\hat{\theta}}$ and $c_{\hat{\rho}}$ are the fitted copula densities and $\bar{\ell}$ and s_{ℓ} are the mean and standard deviation of ℓ_i . Z is asymptotically standard normal when the two models are equally close to the true distribution; positive values favour the rotated Clayton, and no degrees-of-freedom correction is needed because both models have one parameter. The rotated Clayton was retained as the hypothesis-driven specification, and diagnostics (iii) and (iv) were used to judge whether the data support the threshold mechanism it encodes, not to select the family.

Statistical reporting

Associations between pedestrian count and cautionary speed were tested with Kendall's τ_b (two-sided test of $\tau_b = 0$ based on its asymptotic normal distribution), and the copula families were compared with the two-sided Vuong test, at a significance level of 0.05. All P values assume independent 15-minute intervals. Because consecutive intervals are serially dependent, these P values and the interval-level bootstrap intervals should be regarded as lower bounds on the true uncertainty. The sample sizes are $n = 1,756$ (Andria) and $n = 14,199$ (Milan) free-flow intervals. Descriptive statistics are reported as means with standard deviations unless stated otherwise. Analyses were performed in R.

Use of large language models	645
Generative AI was used to review the internal consistency of the manuscript and to edit the language of the manuscript. The authors reviewed and edited all content and take full responsibility for it.	646 647 648 649
Data availability. Telraam sensors collected the data that support the findings of this study and are available from the corresponding author upon reasonable request.	650 651 652
Author contributions. Y.C.: Conceptualization, Methodology, Software, Validation, Formal analysis, Data Curation, Writing - Original Draft, Writing - Review & Editing, Visualization. N.G.: Conceptualization, Methodology, Resources, Data Curation, Writing - Review & Editing. F.M.: Data Curation, Writing - Review & Editing. M.O.: Methodology, Supervision.	653 654 655 656 657 658
Funding. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.	659 660
Competing interests. The authors declare no competing interests.	661 662
Supplementary information. Supplementary Note 1, Supplementary Table S1 and Supplementary Figs. S1-S3 are provided in a separate file.	663 664 665 666
References	667
[1] World Health Organization. <i>Global Status Report on Road Safety 2023</i> (World Health Organization, Geneva, 2023).	668 669 670
[2] United Nations General Assembly. Transforming our world: the 2030 Agenda for Sustainable Development. Resolution A/RES/70/1 (2015).	671 672 673
[3] Allam, Z., Nieuwenhuijsen, M., Chabaud, D. & Moreno, C. The 15-minute city offers a new framework for sustainability, liveability, and health. <i>The Lancet Planetary Health</i> 6 , e181–e183 (2022).	674 675 676 677
[4] Transportation Research Board. <i>Highway Capacity Manual</i> (National Research Council, Washington, DC, 2000).	678 679 680
[5] Gulivindala, P. & Mehar, A. Analysis of side friction on urban arterials. <i>Transport and Telecommunication</i> 19 , 21–30 (2018).	681 682 683
[6] Al-yasee, A. H. H., Leong, L. V. & Al-Jameel, H. A. Comprehensive assessment of side friction impacts on urban traffic flow: A case study of Hilla City, Iraq. <i>Open Engineering</i> 14 (2024).	684 685 686 687
[7] Madanipour, A., Shucksmith, M. & Brooks, E. The concept of spatial justice and the European Union's territorial cohesion. <i>European Planning Studies</i> 30 , 807–824 (2022).	688 689 690

- 691 [8] Choi, J., Kim, S., Min, D., Lee, D. & Kim, S. Human-centered designs, char-
692 acteristics of urban streets, and pedestrian perceptions. *Journal of Advanced*
693 *Transportation* **50**, 120–137 (2016).
694
- 695 [9] Dumbaugh, E. Safe streets, livable streets. *Journal of the American Planning*
696 *Association* **71**, 283–300 (2005).
697
- 698 [10] Ewing, R. & Dumbaugh, E. The built environment and traffic safety: A review
699 of empirical evidence. *Journal of Planning Literature* **23**, 347–367 (2009).
700
- 701 [11] Luu, D. T., Baby, T., Lee, J., Daimon, T. & Lee, S. C. Pedestrian crossing strategy
702 while interacting with multiple autonomous vehicles in local street intersections:
703 Effects of road infrastructure. *Transportation Research Part F: Traffic Psychology*
704 *and Behaviour* **118** (2026).
705
- 706 [12] Lanzaro, G., Sayed, T. & Fu, C. A comparison of pedestrian behavior in inter-
707 actions with autonomous and human-driven vehicles: An extreme value theory
708 approach. *Transportation Research Part F: Traffic Psychology and Behaviour* **99**,
709 1–18 (2023).
710
- 711 [13] Feng, Z. *et al.* Impact of risk perception and trust in autonomous vehicles on
712 pedestrian crossing decision: Navigating the social-technological intersection with
713 the ICLV model. *Transport Policy* **152**, 71–86 (2024).
714
- 715 [14] Geroliminis, N. & Sun, J. Properties of a well-defined macroscopic fundamental
716 diagram for urban traffic. *Transportation Research Part B: Methodological* **45**,
717 605–617 (2011).
718
- 719 [15] Treiber, M. & Kesting, A. *Traffic Flow Dynamics: Data, Models and Simulation*
720 (Springer, Berlin, Heidelberg, 2013).
721
- 722 [16] Alomari, A. H., Khedaywi, T. S., Marian, A. R. O. & Jadah, A. A. Traffic speed
723 prediction techniques in urban environments. *Heliyon* **8** (2022).
724
- 725 [17] Sun, S., Zhang, C. & Yu, G. A bayesian network approach to traffic flow fore-
726 casting. *IEEE Transactions on Intelligent Transportation Systems* **7**, 124–132
727 (2006).
728
- 729 [18] Kim, J. & Wang, G. Diagnosis and prediction of traffic congestion on urban
730 road networks using Bayesian networks. *Transportation Research Record* **2595**,
731 108–118 (2016).
732
- 733 [19] Cervellera, C., Macciò, D. & Rebora, F. Copula-based scenario generation for
734 urban traffic models. *Expert Systems with Applications* **210**, 118389 (2022).
735
- 736 [20] Größer, J. & Okhrin, O. Copulae: An overview and recent developments. *Wiley*
Interdisciplinary Reviews: Computational Statistics **14** (2022).

- [21] Bhat, C. R. & Eluru, N. A copula-based approach to accommodate residential self-selection effects in travel behavior modeling. *Transportation Research Part B: Methodological* **43**, 749–765 (2009). 737
738
739
740
- [22] Sun, B., Ermagun, A. & Dan, B. Built environmental impacts on commuting mode choice and distance: Evidence from Shanghai. *Transportation Research Part D: Transport and Environment* **52**, 441–453 (2017). 741
742
743
744
- [23] Ermagun, A. & Samimi, A. Mode choice and travel distance joint models in school trips. *Transportation* **45**, 1755–1781 (2018). 745
746
747
- [24] Yu, Y., Chen, M., Qi, H. & Wang, D. Copula-based travel time distribution estimation considering channelization section spillover. *IEEE Access* **8**, 32850–32861 (2020). 748
749
750
- [25] Zhao, T., Nie, Y., Wu, X. & Zhang, Y. Empirical analysis of the dependence structure in traffic data using copula function. *Proc. 2014 IEEE Int. Conf. Service Operations and Logistics, and Informatics (SOLI)* 38–42 (2014). 751
752
753
754
- [26] Fang, Z. *et al.* Multivariate analysis of traffic flow using copula-based model at an isolated road intersection. *Physica A: Statistical Mechanics and Its Applications* **599**, 127431 (2022). 755
756
757
758
- [27] Zhou, Y., Cheng, Q., Zhang, C., Luo, M. & Liu, Z. Stochastic fundamental diagram modeling using asymmetric vine and nested archimedean copulas. *Transportation Research Part B: Methodological* **203** (2026). 759
760
761
762
- [28] Yasmin, S., Momtaz, S. U., Nashad, T. & Eluru, N. A multivariate copula-based macro-level crash count model. *Transportation Research Record: Journal of the Transportation Research Board* **2672**, 64–75 (2018). 763
764
765
766
- [29] Ahmad, N., Gayah, V. V. & Donnell, E. T. Copula-based bivariate count data regression models for simultaneous estimation of crash counts based on severity and number of vehicles. *Accident Analysis and Prevention* **181** (2023). 767
768
769
770
- [30] Chen, T., Wong, Y. D., Shi, X. & Yang, Y. A data-driven feature learning approach based on Copula-Bayesian Network and its application in comparative investigation on risky lane-changing and car-following maneuvers. *Accident Analysis and Prevention* **154** (2021). 771
772
773
774
- [31] Ghomi, H. & Hussein, M. Moving Vision Zero programs forward: What pedestrian-focused countermeasure combinations work best and where? A dynamic copula-based time-series approach. *Accident Analysis & Prevention* **192**, 107229 (2023). 775
776
777
778
779
- [32] Gu, R. & Sze, N. N. A vine copula-based analysis of spatial dependence of traffic conflict risk at highway ramp areas. *Accident Analysis and Prevention* **225** (2026). 780
781
782

- [33] Hu, Y., Li, Y., Yuan, C. & Huang, H. Modeling conflict risk with real-time traffic data for road safety assessment: A copula-based joint approach. *Transportation Safety and Environment* **4** (2022).
- [34] Nashad, T., Yasmin, S., Eluru, N., Lee, J. & Abdel-Aty, M. A. Joint modeling of pedestrian and bicycle crashes: Copula-based approach. *Transportation Research Record: Journal of the Transportation Research Board* **2601**, 119–127 (2016).
- [35] Cheng, W., Gill, G. S., Vo, T., Zhou, J. & Sakrani, T. Use of bivariate Dirichlet process mixture spatial model to estimate active transportation-related crash counts. *Transportation Research Record: Journal of the Transportation Research Board* **2672**, 105–115 (2018).
- [36] Wali, B., Santi, P. & Ratti, C. A joint demand modeling framework for ride-sourcing and dynamic ridesharing services: A geo-additive Markov random field based heterogeneous copula framework. *Transportation* **50**, 1809–1845 (2023).
- [37] Bilal, M., Son, S. & Jang, K. Traveler’s interactive decision-making behavior between itinerary and mode choice using copula-based discrete-count joint modeling. *Transportation* **52**, 1249–1266 (2025).
- [38] Genest, C. & Nešlehová, J. A primer on copulas for count data. *ASTIN Bulletin* **37**, 475–515 (2007).
- [39] Zilko, A. A. & Kurowicka, D. Copula in a multivariate mixed discrete–continuous model. *Computational Statistics & Data Analysis* **103**, 28–55 (2016).
- [40] Ažman Momirski, L. & Berčič, T. Southern inner ring road in Ljubljana: 2021 data set from traffic sensors installed as part of the citizen science project WeCount. *Data in Brief* **41**, 107878 (2022).
- [41] Denuit, M. & Lambert, P. Constraints on concordance measures in bivariate discrete data. *Journal of Multivariate Analysis* **93**, 40–57 (2005).
- [42] Machado, J. A. F. & Santos Silva, J. M. C. Quantiles for counts. *Journal of the American Statistical Association* **100**, 1226–1237 (2005).
- [43] Cleveland, W. S. Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association* **74**, 829–836 (1979).
- [44] Künsch, H. R. The jackknife and the bootstrap for general stationary observations. *The Annals of Statistics* **17**, 1217–1241 (1989).
- [45] Genest, C., Ghoudi, K. & Rivest, L.-P. A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika* **82**, 543–552 (1995).

[46] Genest, C., Rémillard, B. & Beaudoin, D. Goodness-of-fit tests for copulas: A review and a power study. <i>Insurance: Mathematics and Economics</i> 44 , 199–213 (2009).	829 830 831 832
[47] Sklar, A. Fonctions de répartition à n dimensions et leurs marges. <i>Publications de l'Institut de Statistique de l'Université de Paris</i> 8 , 229–231 (1959).	833 834 835
[48] Nelsen, R. B. <i>An Introduction to Copulas</i> 2nd edn (Springer, New York, 2006).	836 837
[49] Kendall, M. G. The treatment of ties in ranking problems. <i>Biometrika</i> 33 , 239–251 (1945).	838 839
[50] Joe, H. <i>Multivariate Models and Dependence Concepts</i> (Chapman & Hall, London, 1997).	840 841 842
[51] Brechmann, E. C. & Schepsmeier, U. Modeling dependence with c- and d-vine copulas: The R package CDVine. <i>Journal of Statistical Software</i> 52 , 1–27 (2013).	843 844 845
[52] Vuong, Q. H. Likelihood ratio tests for model selection and non-nested hypotheses. <i>Econometrica</i> 57 , 307–333 (1989).	846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryInformation.pdf](#)