Analysis of Long-term Observational Data on Pedestrian Road Crossings at Unmarked Locations

Authors: Anastasis Gerogiannis[1], Nikolai W.F. Bode[1]*

Author affiliations: [1]School of Engineering Mathematics and Technology, Faculty of Engineering, University of Bristol, Bristol, BS8 1TW

Corresponding author: * Nikolai W.F. Bode, nikolai.bode@bristol.ac.uk, School of Engineering Mathematics and Technology, Faculty of Engineering, University of Bristol, Ada Lovelace Building, Tankard's Close, Bristol, BS8 1TW

Declaration of competing interest

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

Funding

This project is funded by the National Institute for Health Research (NIHR) [Public Health Research Rapid Funding Scheme (project reference NIHR131048)]. The views expressed are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care.

Abstract

Crossing roads ranks amongst the most dangerous activities for pedestrians. Roads can be crossed at controlled, signalised locations, where traffic lights or zebra crossings regulate the behaviour of all traffic participants, or at unmarked locations, where pedestrians typically do not have priority. Technological advances mean it is now possible to record observational data on pedestrian road crossing behaviour from public roads almost continuously using commercially available sensors. Here, we report on such a data collection campaign in Bristol, UK. We record the movement paths of traffic participants within the field of view of commercial camera-based sensors at two unmarked crossing locations. Between January and April 2022, we detect over 30,000 pedestrian road crossings across the two locations. We first explore the time series of hourly crossing counts, finding pronounced and regular temporal patterns that differ between locations, and that have not been reported before. We then investigate the relationship of crossing numbers with road traffic characteristics and extraneous factors, such as university term dates, confirming previous findings on traffic volume reducing crossing frequency and the differences between our study sites. Finally, by studying the timing and distance between consecutive crossings we find evidence for social crossing behaviour, such as groups crossing synchronously. We conclude that temporal patterns in road crossing behaviour exist, and that they can differ across locations, which is relevant for research and road safety design. In addition to the specific findings on road crossing behaviour of our study, a key contribution of our work is a case study for how to work with large-volume, low-fidelity observational data on pedestrian behaviour that is becoming increasingly available and has the potential to transform pedestrian road safety research.

Keywords: pedestrian, mid-block, jaywalking, traffic patterns, road safety

1. Introduction

Pedestrian safety on roads remains an important global issue (Theofilatos et al., 2021). For example, in the year 2021 in the United Kingdom alone, 361 pedestrians were killed, 5,032 were reported to be seriously injured, and 11,261 were slightly injured in road accidents (Department for Transport, 2022). Most of these accidents involve cars or other motorised vehicles, meaning that entering or crossing roads rank amongst the most dangerous activities for pedestrians. Consequently, much research has been directed at understanding road crossing behaviour, both at an individual level and including driver behaviour, and at the level of different pedestrian facilities, such as signalised, unsignalized, marked, and unmarked road crossing locations (e.g., reviewed in Amini et al., 2019; Ghomi & Hussein, 2022; Theofilatos et al., 2021).

A particular safety risk arises when pedestrian behaviour deviates from the expectations of drivers (Sheykhfard et al., 2021). Such behaviours can be violations of traffic rules, or simply unexpected actions, such as suddenly running onto an unsignalized marked crossing without signalling an intent to do so. Such pedestrian behaviour can be broadly split into two categories (Ghomi & Hussein, 2022): temporal deviations (e.g., when pedestrians cross the road when traffic signals are on red for them) and spatial deviations (e.g., when pedestrians cross the road at a location where they should not cross the road). Here, we investigate the latter case, spatial deviations, which is often termed jaywalking, even though the appropriateness of this term is debated (e.g., Hough, 2022; Norton, 2007). To avoid using this contentious term, we will refer to road crossings throughout, even though we focus on the specific scenario that has often been described as jaywalking or road crossings at unmarked midblock locations.

To understand and ultimately predict or prevent unsafe situations arising from such road crossing behaviour, studies have focussed on determining the factors responsible for it and several comprehensive reviews of this literature are available (Anik et al., 2021; Ghomi & Hussein, 2022; Theofilatos et al., 2021). An important group of factors relates to the physiological and psychological characteristics of the pedestrians themselves. The speed at which pedestrians walk (or can walk) determines how long it takes them to cross the road. This is used in the design of green-red phases for traffic signals and importantly it determines the minimum time gap between consecutive vehicles on the road that pedestrians require to be able to complete their crossing (Forde & Daniel, 2021; Amini et al., 2019). Gap acceptance theory has been developed to explain which gaps between vehicles pedestrian accept as large enough to cross the road and what aspects influence these crossing decisions (Kadali & Vedagiri, 2013; Theofilatos et al., 2021). Aspects studied related to pedestrian characteristics include communication during/prior to crossing, gaze directions, age, gender, walking in a group with others, social norms, time pressure, mobile phone use, trip purpose, and even socio-economic factors, such as vehicle ownership (Dommes et al., 2012; Amini et al., 2019; Ghomi & Hussein, 2022; Theofilatos et al., 2021 ; Anik et al., 2021). Effects found are diverse and not necessarily consistent across studies (see also below).

The gap sizes between vehicles are determined by the traffic conditions and there appears to be consensus that traffic conditions influence the frequency of road crossings. However, the precise nature of these effects is less clear (Ghomi & Hussein, 2022). For example, average traffic speeds may influence different age groups in different ways (Ghomi & Hussein, 2022), and while traffic volume has been found to reduce the number of road crossings (Wang et al., 2021), other studies highlight the importance of time gaps between vehicles which depends on traffic speed and density, vehicle types (Ghomi & Hussein, 2022), and the noise emitted by vehicles, comparing combustion to electric engines (Soares et al., 2021). Environmental factors, such as weather conditions, have also been considered (Amini et al., 2019; Ghomi & Hussein, 2022; Theofilatos et al., 2021). Models have been developed to predict aspects of the decisions or movement paths of people crossing the road as a function of some of these traffic and pedestrian characteristics to better understand this behaviour and to make it more predictable (Kadali & Vedagiri, 2013; Amini et al., 2019; Anik et al., 2021; Zhu et al., 2021).

Another line of investigation that we also follow in this contribution shifts the focus from individual behaviour to comparing aggregated behaviour across locations. For example, findings suggest that average crossing speeds differ across locations (Govinda et al., 2020), and a comparison of road crossing intentions between two different cities (Dalian, China and Djibouti, Djibouti), suggests that social norms, the perceived ability to judge the situation, and goals can differ across cultures (Aden et al., 2021). Higher average traffic speeds have been suggested to reduce the volume of road crossings (Acharya & Marsani, 2019), and the installation of traffic signals has been suggested to reduce the walking speed and increase the waiting times of pedestrians at crossing locations (Asaithambi et al., 2016). Characteristics of the built environment, including land use (e.g., residential vs commercial), the number of lanes, bus stops, and the presence of traffic islands in the middle of the road have been suggested as being relevant (Amini et al., 2019; Ghomi & Hussein, 2022; Theofilatos et al., 2021). The concept of gap acceptance theory has been further developed into a proactive method for pedestrian safety that computes safety margins based on the difference between gap sizes and pedestrian crossing times to assess the accident risks at different road crossings (Kadali & Vedagiri, 2016). In a different vein, the crash severity at midblock locations has received considerable attention with models being developed to predict the severity of crashes at locations, based on built environment, socio-demographic, and other features (e.g. Pour et al. (2017), and references therein).

This body of research indicates the difficulty of pinpointing the key drivers for pedestrian crossing behaviour. Recognising this problem, authors have started to conduct meta-analyses to combine the insights gained from different studies. Ghomi & Hussein (2022) conduct a meta-analysis to examine the effect of some of the abovementioned factors on the frequency of pedestrian crossings that violate rules across published studies with a focus on historical collision records. This analysis finds agreement across studies on some factors, such as the presence of bus stops increasing road crossings, but is inconclusive on other factors, such as vehicle speeds, age, and gender. A different meta-analysis focusses on factors influencing gap acceptance probabilities in pedestrians, finding that vehicle speed, gap size, and frequency of attempts but not waiting time had significant effects (Theofilatos et al., 2021). The substantial variability across studies found in these meta-analyses and the findings on differences between locations suggests that road crossing behaviour depends on many factors that may not always have been measured in previous work. The main contribution of our work is that we focus on an aspect that has received little attention to date: temporal variation in road crossing behaviour.

Data collection in studies on pedestrian road crossing behaviour has employed observations, such as videographic surveys (Asaithambi et al., 2016, Acharya & Marsani, 2019), surveys (Aden et al., 2021), and controlled experiments, for example using virtual reality (Feldstein & Dyszak, 2020; Soares et al., 2021) or mixed reality (Dommes et al., 2012; Dommes et al., 2014). Studies that consider aggregate behaviour at urban locations typically involve monitoring over limited time periods, such as an hour a day for several days (Asaithambi et al., 2016; Acharya & Marsani, 2019). Consequently, temporal patterns in road crossing numbers are not well understood. For example, the underlying motivation of pedestrians, or other behavioural patterns may change systematically throughout the day, and this may influence the frequency with which pedestrians cross the road at unmarked locations. In one of the first studies considering this issue, it has been argued that not accounting for such effects could mask other relevant factors and a model was proposed to account for temporal variation in crossing behaviour across discrete time periods, without studying these changes explicitly (Zhang & Fricker, 2021). Here, we record data continuously over four months and can thus, to the best of our knowledge, for the first time, explicitly investigate the nature of temporal patterns in pedestrian road crossing behaviour at different locations.

The three main contributions of our work to research on pedestrian road crossings at unmarked locations are as follows. First, a substantial data set on pedestrian road crossing behaviour spanning four months at two different urban locations. Second, a proof of principle for analyses considering the relationship between road crossing numbers and traffic and extraneous variables and considering the spatial and temporal coincidence of crossings. Third, the explicit investigation of temporal patterns in road crossing numbers. In the remainder of this manuscript, we present our data collection and analysis methods, followed by our findings, and we conclude by discussing our contribution in the context of previous work.

2. Methods

2.1. Data collection

Data were collected using two commercial traffic sensors installed at distinct locations in the city of Bristol, UK. The traffic sensors, supplied by the company Vivacity Labs Ltd (URL: https://vivacitylabs.com/, accessed 23rd August 2023), continuously process images from an on-board digital camera and record the time, location, and type of all detected road users. Due to the commercial nature of the sensors, the exact algorithms for image and data processing on sensors are not available. For privacy reasons, sensors do not store image data. Time points are given in milliseconds and road user positions can be tracked at a frequency of up to 10Hz, although this is not necessarily achieved consistently, as explained in the following. A proprietary image processing algorithm is used to detect road users in video frames and to classify the transport mode into a range of categories, including pedestrian, cyclist, car, light goods vehicle, heavy good vehicle, and more. For this study, we only consider three categories: pedestrians, cyclists, and all other motorised vehicles. Incomplete detections or misclassifications can result in road users not being tracked consistently (information on sensor validation is provided in section 2.3). Road user positions are provided as pixel coordinates on video frames and as latitude longitude coordinates. Two types of output from the sensors are available. First counts of road users crossing virtual count lines are provided every 5 minutes. This data is obtained from processing the raw position data and validated by the company (see below). Second, files containing tracklets (trajectory segments) of road users are saved in 5 minute intervals. This data relies on a proprietary tracking algorithm that stitches recorded positions together into tracklets. We use both types of output in this study. We use the positions provided by sensors directly without applying a smoothing to tracklets, as we have no ground truth to determine the adequacy of such approaches.

The two locations selected for this study are at different points along the same road running through central Bristol (B4051). Both locations cover a stretch of road without a signalised road crossing for pedestrians, although these are available nearby. The locations were chosen because they show high levels of pedestrian traffic, and because they differ in the provision of shops, education, and employment providers nearby. Location 1 is close to the campus of the University of Bristol (51.455592861621525°N, 2.6003341981723747°W), and location 2 is immediately adjacent to a major hospital, the Bristol Royal Infirmary (51.45773547023158°N, 2.596965343668497°W). At both locations there are pavements on both sides of the road, and the road has two lanes, one for each direction of traffic. For location 1, the virtual count lines cover both pavements and the width of the road, whereas for location 2, only the pavement and road lane nearer to the sensor are covered by virtual count lines. The lighting conditions at both locations mean that data collection is possible day and night. This data collection was approved by the Faculty Research Ethics Committee in Engineering at the University of Bristol (application ID: 2021-9472-9213)

Figure 1 provides an overview of the data collected and the two locations. The trajectories shown in Figure 1(a,b) demonstrate the level of noise in the accuracy of positional measurements. At location 1 (university periphery), the northern side of the road borders the university campus, whilst the southern side borders a row of shops, including cafes and food outlets. The stretch of road covered by the sensor covers bus stops on both sides of the road. A signalised pedestrian crossing is located nearby, but not in the field of view of the traffic sensor (see Figure 1(c) for details). At location 2 (hospital periphery), the north-western side of the road borders the main hospital building and the south-eastern side of the road borders one café, other shops, and a university medical education and research institute. As for location 1, a signalised pedestrian crossing is located nearby, but not in the field of view of the traffic sensor (see Figure 1(d) for details). For this study, we use data from 1am on the 1st of January 2022 up to 12am on the 30th of April 2022. During this period, no travel restrictions or shop opening restrictions were in place due to the Coronavirus pandemic (Brown & Kirk-Wade, 2021). Teaching at the university was delivered in person, although many students (especially non-UK students) made use of online teaching provision rather than attending classes in person.

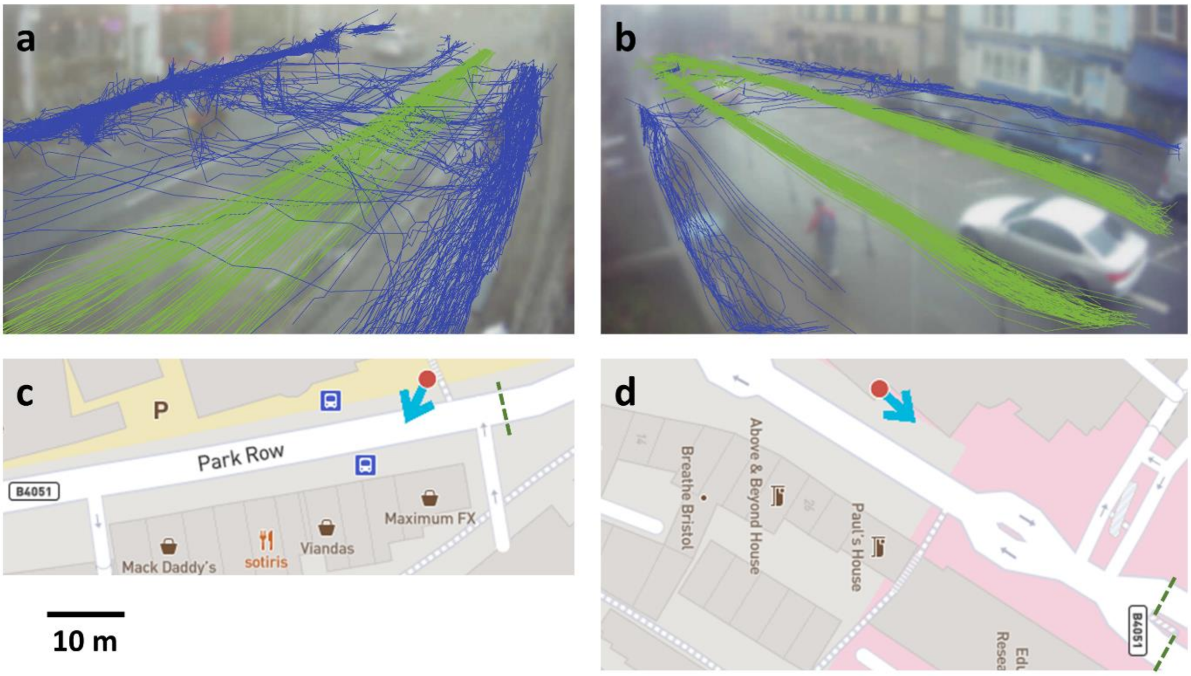

Figure 1: Data collection. (a) pedestrian (blue) and car (green) trajectories captured 8.10-8.20am on the 13th of October 2022, superimposed onto a traffic sensor camera still image (blurred for privacy) at locations 1 (university periphery). (b) the same as (a) but for location 2 (hospital periphery) and 6.10-8.40am on the 13th of October 2022. (c) and (d) show overview maps on the same scale for locations 1 and 2, respectively. The red dot indicates the sensor location and the blue arrow the viewing direction of the sensor camera. Dashed green lines indicate signalised pedestrian crossings. (d) is rotated clockwise by 90 degrees and the writing runs west to east in both (c) and (d). In (d) traffic islands are indicated but these do not substantially change the width of the road contrary to the display on the map. Maps in (c) and (d) are from www.openstreetmap.org (accessed 22nd of August 2023).

2.2. Data preparation

All data preparation and analysis are conducted in the R programming environment, version 4.3.1 (R Core team, 2023). Detecting road crossings by pedestrians first requires projecting latitude longitude coordinates onto a two-dimensional coordinate system. For this, we use the Universal Transverse

Mercator (UTM) projection and UTM zone 30 implemented in the R package "sf", version 0.9-4 (Pebesma, 2018). We then identify the centre line of the road for both locations, project each observed pedestrian position onto this line, and determine the direction along the centre line orthogonal from the pedestrian position to the projected point to determine which side of the road the pedestrian position is on. Road crossings by pedestrians are identified as tracklets which contain positions on both sides of the centre line of the road. We measure speed observations of pedestrians by dividing their displacement between consecutively recorded positions by the time difference between these observations. We find that some speeds exceed a reasonable threshold, even when allowing for measurement errors. Tracklets are defined as speed outliers when one of the observed speeds in the tracklet exceeds 10 m/s. We repeat the entire analysis of this study twice, once without speed outliers, and once for all data. As our findings are broadly comparable, we only report the former analysis (without outliers) in the main text and show results from the latter analysis in the supplementary material. The approximate timing and location of road crossings is identified as the first observation after pedestrians have changed which side of the road centre line they are on. This is only an approximation but determining the exact centre line crossing point and time would require interpolating between observed locations using velocity estimates that are subject to measurement errors. Aggregating approximate locations and timings over many observed road crossings will nevertheless provide meaningful insights.

To study correlations between the occurrence of pedestrian road crossings, other traffic characteristics, and extraneous factors, we compute several summary statistics. All summary statistics are initially obtained for 5 minute intervals and subsequently averaged across hours of the day to obtain hourly data for our entire study period resulting in time series of n=2,879 data points (one hour is lost due to the clocks changing from winter to summer time). The main variable of interest in this study is the time series of the number of pedestrian crossings. We also consider the time series of pedestrian, cyclist, and motorised vehicle counts. For these, we use the data from the virtual count lines, as it has been validated by the sensor provider (see below). For location 2 (hospital periphery), the count lines only cover part of the road which is a limitation of our study. To further characterize traffic, we calculate time series of average speeds and the coefficient of variation of speeds for motorised vehicles and cyclists. According to the speed-density relationship in road traffic, average speeds for motorised vehicles depend on the overall traffic volume (counts). Therefore, we do not consider both counts and average speeds at the same time in our statistical analysis. The coefficient of variation of speeds indicates the variability of speeds relative to the average speed and could be useful for distinguishing stop-and-go traffic from smoothly running traffic. As extraneous factors, we consider weekends and holidays (3rd of January, instead of New Years Day and 15th, and 18th of April 2022 for Good Friday and Easter Monday, respectively), and term time at the university when classes are running (24th of January – 1st of April and 25th of April – 30th of April 2022 for our data collection period).

2.3. Sensor validation

The visible measurement errors in road user positions provided by the sensors highlights the importance of validating sensor accuracy. We do so by confirming that speed measurements taken from the sensors are reasonable, by reporting the road user count accuracy provided by the sensor provider, and by comparing our automated detection of road crossings to manual counts.

The average speeds for pedestrians crossing the road take values 2.48±1.17 m/s (mean ± standard deviation) at location 1 and 1.82±1.05 m/s at location 2 (including outliers average speeds are 4.79±6.22 m/s and 4.40±8.27 m/s, respectively). These speeds are higher than average walking speeds at crossings reported elsewhere to be between 1.1-1.55 m/s (Amini et al., 2019; Forde & Daniel 2021), although not far off the maximum crossing speeds of around 2.4 m/s that have been observed in studies at midblock crossings (Govinda et al., 2020). Average hourly speed and standard deviation is 3.71±1.45 m/s for cylists and 8.46±1.86 for cars at location 1 (university periphery) and 4.34±1.30 m/s for cyclists and 6.84±1.64 for cars at location 2 (hospital periphery). The speed limit at both locations is 20 mph (miles per hour, approximately 8.94 m/s).

The sensor provider conducted a validation of counts reported by sensors by comparing them to manual counts. This validation was completed between 12noon and 5pm on either the 30th or the 31st of August 2022 for 10 minute intervals for each count line. At location 1, sensor accuracies were 98% for vehicles (202 true positives, 5 false negatives), 100% for pedestrians (34 true positives), and 100% for cyclists (14 true positives). At location 2, sensor accuracies were 99% for vehicles (132 true positives, 1 false positive), 94% for pedestrians (73 true positives, 5 false positives), and 100% for cyclists (4 true positives).

On the 14th of June 2022, we conducted two manual counts of pedestrian crossings at location 1 from 12.40pm-12.55pm and from 1.00pm-1.15pm. Across both time periods, a total of 26 pedestrian crossings were counted manually (this includes 11 crossings completed by one of the researchers). Out of these, 24 were correctly detected by the sensor (true positives), 2 were not detected (false negatives), none that did not happen were detected by the sensor (false positives), and a total of 97 pedestrians passed the location without crossing (true negatives). This implies an accuracy of 98%.

2.4. Statistical analysis

Our statistical analysis of the road crossing number time series considers its characteristic temporal patterns and its correlation with traffic characteristics and extraneous factors.

To determine the characteristic temporal patterns, we consider time of day, day of the week, and monthly variation. We compute the averaged pattern at these timescales alongside bootstrapped 95% confidence intervals using the R package "openair", version 2.9-1 (Carslaw & Ropkins, 2012).

We use regression models to investigate correlations of road crossing numbers with traffic characteristics and extraneous factors. As the road crossing time series shows clear and repeating patterns and thus strong autocorrelation, it is not appropriate to apply standard regression models directly to the counts of road crossings. To overcome this problem, we consider the number of crossings divided by the pedestrian count for the corresponding hour in our regression analysis Instances where no pedestrians are present in a given hour are given the value zero. The clear and repeating patterns in the pedestrian count time series render this new time series of the proportion of pedestrian crossings suitable for standard regression models and we verify this as explained below. For simplicity and ease of interpretation, we use Linear models in our analysis, even though there are known problems in applying them to proportion data. Importantly, Linear models assume that residuals (the difference between model fit and observation) follow a Normal distribution with mean zero and constant variance. We assess whether these assumptions are met using residual plots that are discussed in the text and reported in the supplementary information. We use the Akaike Information Criterion to determine if vehicle count or mean vehicle speed is a more suitable predictor, and we use Likelihood ratio tests to determine if summary statistics on bike traffic improve model fit. All models include an intercept, and having assessed that it is not correlated to other predictors, all models also include the vehicle speed coefficient of variation (vehicle speed CV). All models also include binary predictors for weekend and term time, where the estimated coefficient measures the effect of weekends and term time, as opposed to weekdays and time periods without teaching at the university, respectively. We emphasise that all measured effects are only indicative of correlations and not causal relationships.

To assess the temporal coincidence of road crossings, we consider the times at which pedestrian road crossings were observed. If road crossings are clustered temporally, this indicates that pedestrians cross the road together, either because they are walking together in a social group, or because traffic conditions are favourable for crossing the road, for example. We construct a regular binary time series of time bins and determine for each bin whether at least one crossing occurs within the time bin (time series value 1) or not (time series value 0). We then use runs tests, as implemented in the R package “DescTools” (Signorell et al., 2021), to assess if the distribution of runs of consecutive zeros or ones is equal to the number expected under a random distribution of the same number of zeros and ones as in the time series considered. We consider time bins of 5 s, 10 s, and 50 s in this analysis.

3. Results

3.1. Temporal characteristics of traffic and road crossing numbers

Over the entire study period, we count 160,654 cyclists, 692,597 pedestrians, and 1,712,188 vehicles at location 1 (university periphery). The counts for location 2 (hospital periphery) are lower with 100,859 cyclists, 559,585 pedestrians, and 1,426,130 vehicles but this is likely due to the count lines not covering the entire width of the road at this location. The difference between locations is reversed for the number of observed pedestrian road crossings. At location 1, we count 9,921 crossings, and at location 2, we count 28,472. For both locations, there are a considerable number of speed outliers, and we consequently only use data from 5,521 and 20,990 crossings for locations 1 and 2, respectively (we repeat our analysis for all data). These numbers suggest that road crossings occur frequently at both locations, even though signalised road crossings are available nearby.

Figure 2 shows that traffic at location 1 shows clear temporal patterns. All transport modes show the expected diurnal pattern with low traffic volumes at night. On weekdays but not on weekends, cycle traffic shows distinct morning and afternoon peaks indicative of commuter traffic. On most weekdays but not on weekends and possibly Wednesdays, pedestrian traffic shows three peaks. The peaks in the morning and afternoon are likely to be linked to working or studying patterns, and the midday peak may be due to people getting lunch. Vehicle traffic also shows one early and one late peak, although these are not as clearly delineated as for cycle traffic, and vehicle traffic remains high throughout the day. On weekends, traffic generally increases later and is somewhat higher at night which may be due to people going out (bars and nightclubs are close to location 1). January has lower traffic volumes than the other months, although this could be largely due to the holidays around Christmas and New Year, and the fact that university teaching only started on the 24th of January 2022. The temporal traffic patterns at location 2 are comparable to those at location 1, except for location 1 having a stronger weekday midday peak in pedestrian traffic (supplementary figure S1).

The temporal patterns in road crossings differ between locations 1 and 2 (Figures 3 and 4). At location 1, the peak in crossings occurs on or just after midday on weekdays, whereas at location 2, crossing number is the highest in the morning around 7am or 8am. Location 1 shows increased levels of crossings during the weekend between 2am and 4am. This does not occur at location 2, where the weekday morning peak persists on weekends and there is an additional evening peak at around 7pm. While we can only speculate, it is reasonable to suggest that these differences could be due to the proximity of different service providers to the two locations. While the university is closed over the weekend, the hospital is open, and there are more bars and nightclubs near to location 1 that location 2. The temporal patterns in road crossings at both locations are the same when outliers are not removed (supplementary figure S2 and S3).

We observe that even though the shape of the temporal pattern in pedestrian counts is similar for both locations, except for a more pronounced weekday midday peak at location 1, the temporal pattern in road crossing numbers is not. This suggests that there are location-specific behavioural drivers for crossing the road and these could be linked to the desire for reaching/leaving workplaces or food outlets quickly, as indicated above. Removing speed outliers does not affect the shape of temporal patterns. We also investigate if speed outliers occur at specific places along the road centre line. For example, tracklets further away from sensors may involve higher measurements errors. However, we do not find consistent patterns along those lines. For location 1, speed outliers occur more frequently at intermediate distances from the sensor, whereas at location 2 they occur more frequently closer to the sensor (Supplementary figures S4 and S5).

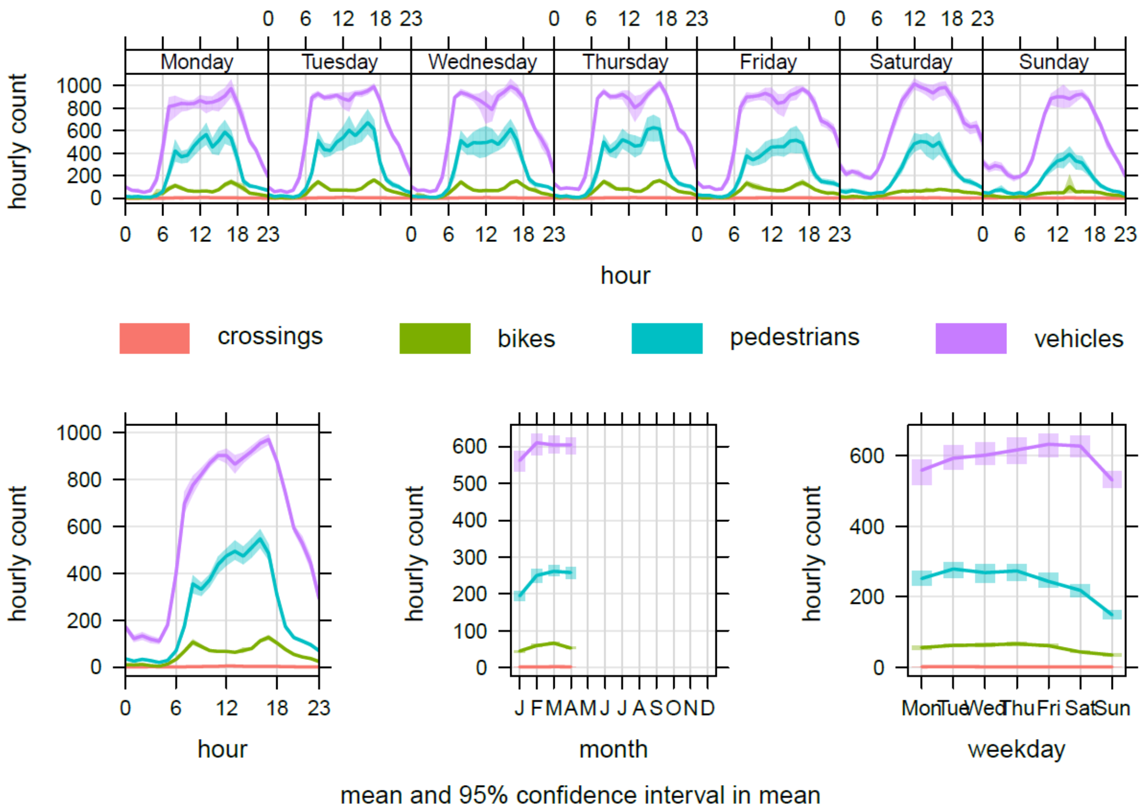

Figure 2: Characteristic traffic patterns from 1st of January to 30th of April 2022 at location 1 (university periphery). Average counts for road crossings, bikes, pedestrians, and motorised vehicles are shown with bootstrapped 95% confidence intervals (n=100 bootstrap samples). Counts of pedestrians crossing the road are substantially lower that the other counts. Starting at the top, and moving in an anti-clockwise direction, the average hourly counts for the different days of the week, the average hourly count pattern, the average monthly counts, and the average daily counts for the different days of the week are displayed.

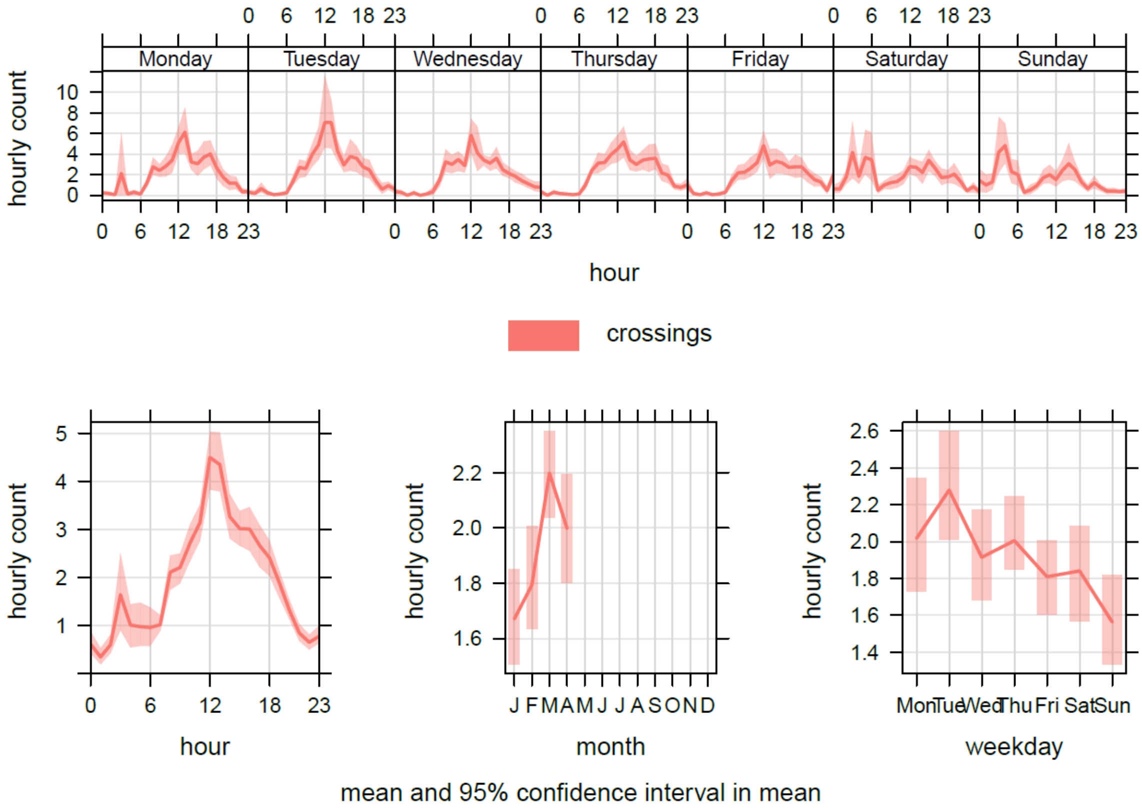

Figure 3: Characteristic patterns in road crossing counts from 1st of January to 30th of April 2022 at location 1 (university periphery). Average counts for road crossings with bootstrapped 95% confidence intervals are shown (n=100 bootstrap samples). The same summary plots as in figure 2 are shown.

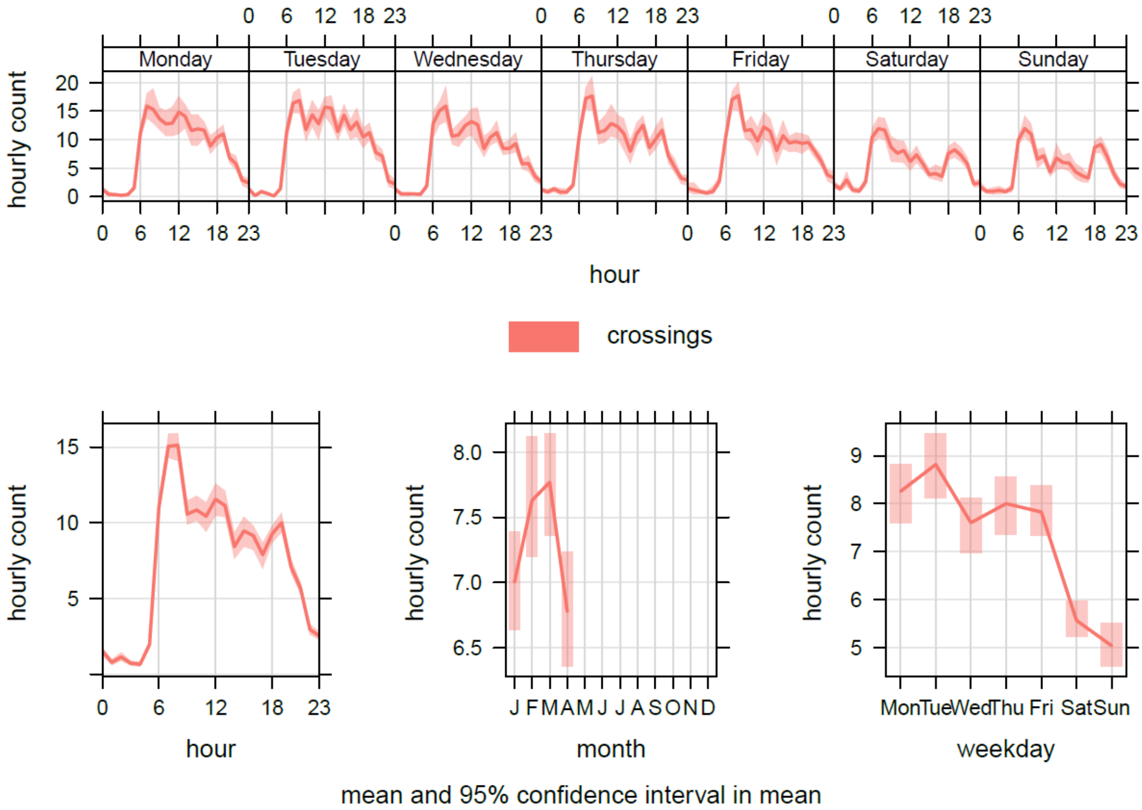

Figure 4: Characteristic patterns in road crossing counts from 1st of January to 30th of April 2022 at location 2 (hospital periphery). Average counts for road crossings with bootstrapped 95% confidence intervals are shown (n=100 bootstrap samples). The same summary plots as in figure 2 are shown.

3.2. Relationships between road crossings, traffic, and extraneous variables

To study the correlation between the proportion of road crossings of the total pedestrian count and traffic variables, we first fit two models to our data, one including vehicle speed CV and vehicle count as predictors, the other with vehicle speed CV and average vehicle speed as predictors. The resulting differences in AIC values between models, ΔAIC, indicate that vehicle count is a more suitable predictor (ΔAIC=109 and ΔAIC=216 for locations 1 and 2, respectively). Considering cycle traffic, we only consider the number of cyclists present, as other cycle traffic characteristics leads to many missing values due to the many hours when no cyclists pass our study locations, and find that it does not substantially improve model fit (Likelihood-ratio test, location 1: $X_1^2 = 2.1416$, $p = 0.1434$; location 2: $X_1^2 = 0.3771$, $p = 0.5392$). We then add the binary predictors for weekend and term time to our model and report the results of our model fit in table 1 for location 1 and table 2 for location 2. We confirm that the predictors included in our models are not correlated, as required for the type of regression model we use here.

For location 1, we find that all predictors, except for term time, have low p-values, suggesting they are correlated with the number of road crossings divided by the total pedestrian count (table 1). In contrast, at location 2, only the vehicle count has a very low p-value (table 2). Vehicle speed CV and term time also have p-values lower than 0.1 but given the number of data points, a clearer indication of a statistically significant correlation would be expected. The signs of model parameter estimates are the same for both locations suggesting that the relationships between variables is similar for both locations. Vehicle count has the strongest effect and coincides with a lower value of the dependent variable. Vehicle speed CV and weekends coincide with higher values of the dependent variable. The parameter associated with term time is negative, this p-values for this variable suggest we cannot reject the null hypothesis that it is not correlated with the dependent variable at either location. While the proportion of road crossings of the total pedestrian count cannot be compared directly to the temporal patterns discussed in section 3.1, our findings support our earlier observation that weekends lead to changes in crossing patterns at location 1, but not at location 2. In line with previous work on road crossings at uncontrolled locations, we find that vehicle traffic volume is relevant for crossing behaviour, even at the aggregated level that we consider here. The fact that cycle traffic does not play the same role is perhaps not surprising, given the difference in vehicle and cycle traffic at out study locations (see Figure 2 and Supplementary figure 1).

Assessing whether the assumptions of linear regression hold for our data using residual plots reveals that dividing road crossing numbers by the total pedestrian count appears to successfully mitigate autocorrelation issues, but the distributional assumptions are not met, as residual distributions are more skewed that would be expected under a Normal distribution (Supplementary figures S6 and S7). As such, the outcomes of our hypothesis tests should be treated with caution. As we only use regression models for an exploratory analysis and we can only study correlations with this analysis, we refrain from implementing an alternative approach, but we discuss this further in section 4.

Repeating our regression analysis for all data, including speed outliers, yields qualitatively similar results to the analysis reported here (see Supplementary tables S1 and S2). The only qualitative difference is that for location 2, the p-value associated with term time is lower than 0.05 (see Supplementary table S2). Thus, there could be an indication that during term time the number of crossings per observed pedestrian is somewhat lower at location 2.

Table 1

Multiple linear regression model fit to road crossing data for location 1 (university periphery). Data without speed outliers is used (n=2,877 due to missing values). The dependent variable is the hourly number of road crossings, divided by the pedestrian number in the corresponding hour. The model includes an intercept, two continuous independent variables (hourly vehicle count and hourly average of vehicle speed coefficient of variation, CV), and two binary independent variables (weekend and term time; the measured effect is for weekend days and national holidays, and for days during term-time when teaching is taking place at the university, respectively). The table shows parameter estimates, estimated standard errors (SE), the test statistic for the parameter specific test (T), and the corresponding p-value (P; null hypothesis, H_0*: parameter = zero*). P-values $<2\times10^{-16}$ are smaller than the numerical precision of the statistical software. P-values lower than 0.05 are shown in bold. Effect ranges for vehicle counts and coefficient of variation in vehicle speeds are [-0.0430,0] and [0.00318,0.0455], respectively (rounded to 3 significant figures).

Coefficient	Estimate	SE	T	P
Intercept	1.565×10^{-2}	2.351×10^{-3}	6.655	**3.38×10^{-11}**
Vehicle count	-3.002×10^{-5}	2.706×10^{-6}	-11.093	**$< 2\times10^{-16}$**
Vehicle speed CV	1.603×10^{-2}	2.248×10^{-3}	7.128	**1.28×10^{-12}**
Weekend	7.799×10^{-3}	1.780×10^{-3}	4.381	**1.22×10^{-5}**
Term time	-3.085×10^{-4}	1.711×10^{-3}	-0.180	0.857

Table 2

As table 1 but for location 2 (hospital periphery). Data without speed outliers is used (n=2,867 due to missing values). Effect ranges for vehicle counts and coefficient of variation in vehicle speeds are [-0.0918,0] and [0.00583,0.0327], respectively (rounded to 3 significant figures).

Coefficient	Estimate	SE	T	P
Intercept	9.612×10^{-2}	8.542×10^{-3}	11.253	**< 2×10^{-16}**
Vehicle count	-9.952×10^{-5}	6.675×10^{-6}	-14.909	**< 2×10^{-16}**
Vehicle speed CV	1.215×10^{-2}	6.795×10^{-3}	1.788	0.0739
Weekend	4.512×10^{-3}	3.889×10^{-3}	1.160	0.2461
Term time	-6.561×10^{-3}	3.562×10^{-3}	-1.842	0.0656

3.3. Temporal and spatial coincidence of consecutive crossings

Our runs tests on binned time series for pedestrian road crossing occurrence show that crossings are more clustered in time than would be expected if time bins with crossings occurred randomly and independently from each other. Considering time bins of 5 s, at location 1 we observe r=9,992 runs (segments of time series without change in value) for a time series with k=1,030,455 bins without crossings and m=5,203 bins with crossings (some bins include more than one crossing), which results in a test statistic Z=-53.847 and a p-value lower than 2.16×10^{-16}, below the numerical precision of the statistical software (for location 2, we find r=37,486, k=2,052,685, m=19,769, Z=-61.689, $p<2.16 \times 10^{-16}$). Changing the bin size of the time series to 10 s or 50 s yields the same results, qualitatively (details not reported here).

To investigate the coincidence of observed road crossings further, we consider the bivariate distribution of the time difference and distance between consecutive crossings (Figures 5 and 6). We first consider the marginal distributions. At both locations the frequency of distances between consecutive crossings decreases monotonically as distances increase. Such a distribution is consistent with the situation when crossings occur approximately uniformly randomly along the stretch of road. For comparison, consider a scenario with only two crossing sites. This would result in peaks in the distribution around zero and around the distance between the crossing sites. This suggests that overall pedestrians cross the road along the entire stretch covered by the sensors (see also Supplementary figures S4 and S5 which provide more information on where pedestrians cross). The frequency of time differences shows a clear mode at around 2 minutes for both locations. However, the distribution also has a pronounced tail towards zero and the distribution for location 1 has a second peak at around 2 seconds (Figure 5). These skewed or even multimodal distributions indicate temporal clustering of crossings, as also suggested by our runs test analysis. The bivariate distributions for both locations show consecutive road crossings that are close in time and space. This likely suggests that pedestrians either walk together in social groups or groups created externally (e.g., by individuals disembarking from a bus and wanting to cross the road), or that vehicle traffic conditions lead to situations particularly suited for crossing the road. Consecutive road crossings that are close in time but distant in space are not observed, even though this could be a plausible scenario if traffic is stopped by a nearby traffic light, creating opportunities for crossing along the entire stretch of road.

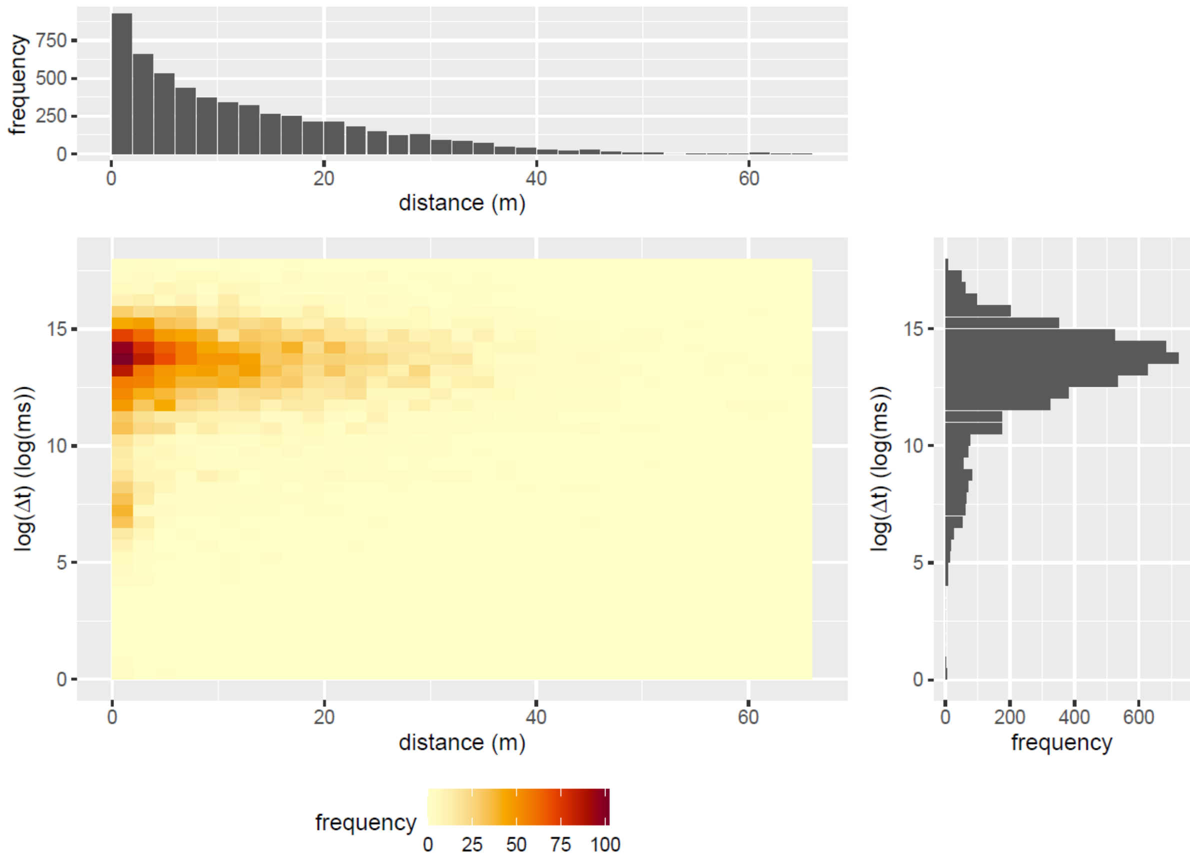

Figure 5: Bivariate distribution of the time difference, Δt, and distance between consecutive road crossings and corresponding marginal distributions for location 1 (university periphery). Note the log scale for time differences (values of 7.5 and 11.7 on this axis correspond to 2 seconds and 2 minutes, respectively).

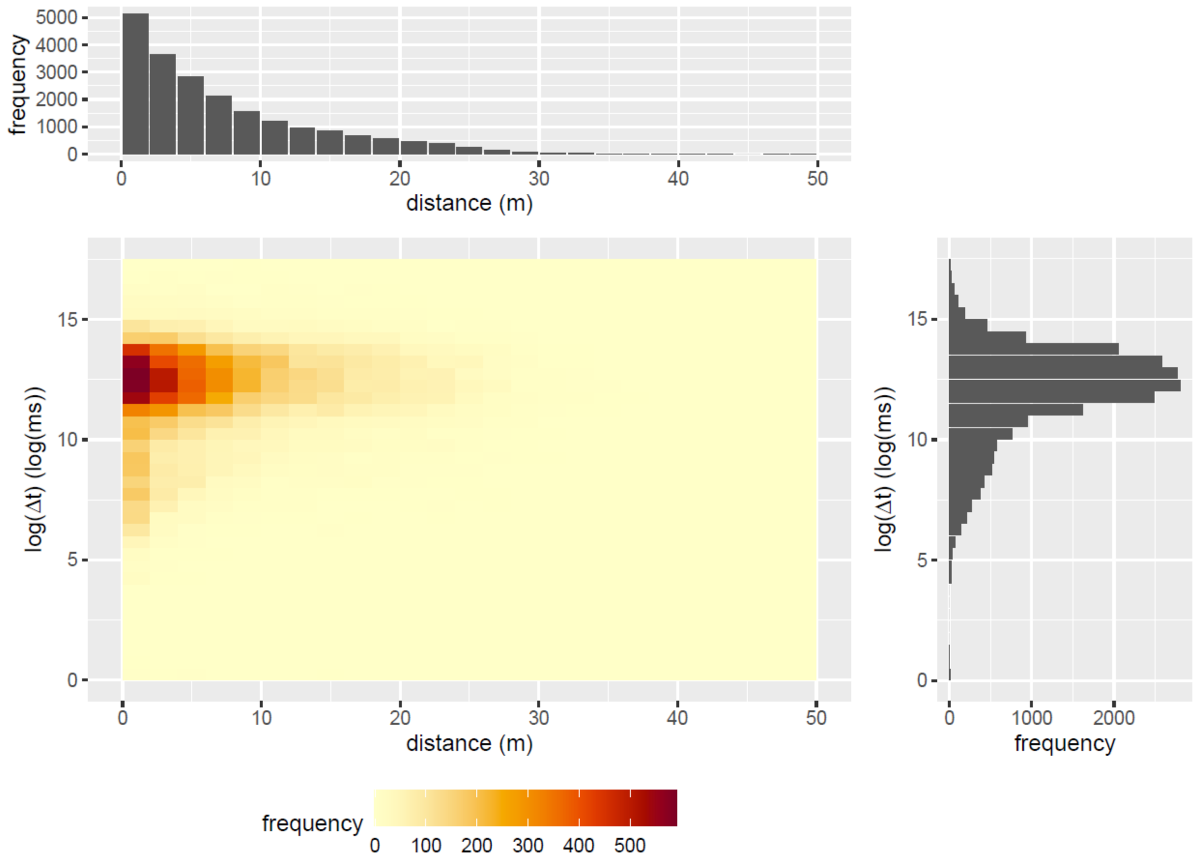

Figure 6: Bivariate distribution of the time difference, Δt, and distance between consecutive road crossings and corresponding marginal distributions for location 2 (hospital periphery). Note the log scale for time differences.

4. Discussion

We find clear evidence for temporal patterns in the number of pedestrians crossing the road at unmarked locations. These patterns differ across the two locations studied, even though overall temporal traffic patterns including pedestrian traffic are broadly similar for both locations. To the best of our knowledge, this has not been reported before, and we suggest it is indicative for changes in pedestrian road crossing behaviour throughout the day. Based on our study sites, we hypothesise that a larger proportion of pedestrians wanting to reach and leave work or food outlets could be motivating factors for changes in aggregated observed behaviour. Further work is needed though to ascertain this. Nevertheless, the difference in temporal patterns across locations has implications for research and road safety design. Considering research, our findings support Zhang & Fricker (2021), who suggest that temporal patterns could mask other factors. For example, suppose a peak in road crossings coinciding with rush hour for traffic at one location, but not at another location, similar to what we observe here. If this temporal pattern was not accounted for, different relationships between road crossing numbers and traffic characteristics could be found for the two locations and may even average out, if data from both locations was combined. Considering road safety design, peaks in crossing behaviour could be de-risked by enforcing time-limited speed restrictions on roads, as is commonly done to reduce noise or air pollution.

We also investigate the relationship of road crossings as a fraction of the total pedestrian numbers with traffic characteristics and other extraneous variables. We find differences between locations that are in line with the differences in observed temporal patterns, namely that weekends have no effect at location 2 near to the hospital. In line with previous work (Wang et al., 2021), traffic volume appears to be relevant consistently with higher traffic volumes coinciding with a lower proportion of pedestrians crossing. The effect of other variables, such as term time or vehicle speed CV is less clear, especially when also considering the analysis that includes speed outliers. This echoes aspects of the findings of meta-analyses on factors important in determining pedestrian crossing behaviour where the effect of several factor was inconclusive even when considering the evidence from many studies (Ghomi & Hussein, 2022; Theofilatos et al., 2021). We suggest that our findings in this regard should be treated with caution for two main reasons. First, studies on observational data always risk missing important factors, the inclusion of which in a statistical analysis could change the observed patterns substantially. Second, as discussed above, our choice of regression model was driven by interpretability and simplicity. Alternative regression models, such as the ones suggested by Zhang & Fricker (2021) to account for temporal variability, or autoregressive models fitted to the crossing count time series, could yield a statistically more robust model fit and inference. Another limitation of our data that warrants caution is the fact that the count lines at location 2 do not cover the entire width of the road. As such, the estimated coefficients should not be compared quantitatively across locations. The quantitative findings should also be considered in the context of the time period when data was recorded. Whilst no travel restrictions due to the Coronavirus pandemic were in place in the UK (Brown & Kirk-Wade, 2021), they were in force elsewhere, impacting the ability of a proportion of the large population of international students in Bristol to attend classes in person. An offering of online teaching may have also been taken up by other students, possibly reducing the overall movement of students to and from the university. Whilst it is important to consider them, these aspects do not invalidate our findings or methodology.

Considering the temporal and spatial coincidence of crossings, we find evidence for pedestrians crossing together with others. Given that many pedestrians walk in groups with others, this is not unexpected. Alternative drivers for this behaviour, such as pedestrians crossing at the same time to catch an arriving bus, are also likely to be relevant. Whilst our work and previous work suggest traffic conditions impact crossing decisions (Ghomi & Hussein, 2022; Theofilatos et al., 2021), we do not find evidence suggesting that crossing decisions occur concurrently along an entire stretch of road, as could be expected under suitable traffic conditions, such as stopped traffic.

5. Conclusions

In summary, we suggest that temporal variation in pedestrian behaviour at unmarked road crossings exists, can vary substantially across locations, and should thus be considered in research and road safety design. Given the variation we observe over a small spatial scale (two locations in the same city on the same road), we propose that substantially more work is needed to be able to predict this temporal variation, even if detailed traffic data is available. Until this is the case, it could be beneficial to monitor locations where interventions are planned continuously for several months, to avoid missing opportunities for finding the most effective road safety solutions. Our work serves as a proof of principle, demonstrating that despite measurement errors, it is possible to detect temporal patterns, investigate driving factors for crossing numbers, and even query behaviours impacting the temporal and spatial coincidence of crossings using commercially available sensors at busy urban locations. If data from such sensors is consistently shared with researchers, it holds the potential to transform our understanding of pedestrian road safety.

References

Acharya, S., & Marsani, A. (2019). Modelling the Relationship between Pedestrian Illegal Mid-Block Crossings with Traffic and Geometric Parameters. International Journal of Advanced Engineering and Management, 4(1).

Aden, W. A., Zhao, S., Subhan, F., Zhou, H., & Ullah, I. (2020). A Comparative Study on Pedestrians' Intention to Violate Traffic Rules: The Case of China and Djibouti. Journal of Road and Traffic Engineering, 66(3), 1-9.

Amini, E.R., Katrakazas, C., & Antoniou, C. (2019). Negotiation and decision-making for a pedestrian roadway crossing: A literature review. Sustainability, 11(23), 6713.

Anik, M. A. H., Hossain, M., & Habib, M. A. (2021). Investigation of pedestrian jaywalking behaviour at mid-block locations using artificial neural networks. Safety science, 144, 105448.

Asaithambi, G., Kuttan, M. O., & Chandra, S. (2016). Pedestrian road crossing behavior under mixed traffic conditions: a comparative study of an intersection before and after implementing control measures. Transportation in developing economies, 2(2), 14.

Brown, J. & Kirk-Wade, E. (2021). Coronavirus: S history of 'Lockdown laws' in England. House of Commons Library, Number 9068.

Carslaw, D.C. & Ropkins, K. (2012). “openair — An R package for air quality data analysis.” Environmental Modelling & Software, 27–28(0), 52–61.

Department for Transport (2022). Reported Road Casualties Great Britain: Pedestrian Factsheet 2021. https://www.gov.uk/government/statistics/reported-road-casualties-great-britain-pedestrian-factsheet-2021 (accessed 31/08/2022).

Dommes, A., Cavallo, V., Vienne, F., & Aillerie, I. (2012). Age-related differences in street-crossing safety before and after training of older pedestrians. Accident Analysis & Prevention, 44(1), 42-47.

Dommes, A., Cavallo, V., Dubuisson, J. B., Tournier, I., & Vienne, F. (2014). Crossing a two-way street: comparison of young and old pedestrians. Journal of safety research, 50, 27-34.

Feldstein, I. T., & Dyszak, G. N. (2020). Road crossing decisions in real and virtual environments: A comparative study on simulator validity. Accident Analysis & Prevention, 137, 105356.

Forde, A., & Daniel, J. (2021). Pedestrian walking speed at un-signalized midblock crosswalk and its impact on urban street segment performance. Journal of traffic and transportation engineering (English edition), 8(1), 57-69.

Ghomi, H., & Hussein, M. (2022). An integrated text mining, literature review, and meta-analysis approach to investigate pedestrian violation behaviours. Accident Analysis & Prevention, 173, 106712.

Govinda, L., Abhigna, D., Nair, P. M., & Shankar, K. R. (2020). Comparative study of pedestrian crossing behaviour at uncontrolled intersection and midblock locations. Transportation Research Procedia, 48, 698-706.

Hough, W. P. (2022). Street Rivalry Reignited? Repealing the Jaywalking Paradigm. UMKC Law Review, 91, 455.

Kadali, B. R., & Vedagiri, P. (2013). Modelling pedestrian road crossing behaviour under mixed traffic condition. European transport, 55(3), 1-17.

Kadali, B. R., & Vedagiri, P. (2016). Proactive pedestrian safety evaluation at unprotected mid-block crosswalk locations under mixed traffic conditions. Safety science, 89, 94-105.

Norton, P. D. (2007). Street rivals: Jaywalking and the invention of the motor age street. Technology and culture, 48(2), 331-359.

Pebesma, E., (2018). Simple Features for R: Standardized Support for Spatial Vector Data. The R Journal 10 (1), 439-446, https://doi.org/10.32614/RJ-2018-009

Pour, A. T., Moridpour, S., Tay, R., & Rajabifard, A. (2017). Modelling pedestrian crash severity at mid-blocks. Transportmetrica A: Transport Science, 13(3), 273-297.

R Core Team (2023). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. URL: https://www.R-project.org/ (accessed 23/08/23).

Sheykhfard, A., Haghighi, F., Papadimitriou, E., & Van Gelder, P. (2021). Analysis of the occurrence and severity of vehicle-pedestrian conflicts in marked and unmarked crosswalks through naturalistic driving study. Transportation research part F: traffic psychology and behaviour, 76, 178-192.

Signorell, A. et al. (2023). DescTools: Tools for descriptive statistics. R package version 0.99.48.

Soares, F., Silva, E., Pereira, F., Silva, C., Sousa, E., & Freitas, E. (2021). To cross or not to cross: Impact of visual and auditory cues on pedestrians' crossing decision-making. Transportation research part F: traffic psychology and behaviour, 82, 202-220.

Theofilatos, A., Ziakopoulos, A., Oviedo-Trespalacios, O., & Timmis, A. (2021). To cross or not to cross? Review and meta-analysis of pedestrian gap acceptance decisions at midblock street crossings. Journal of Transport & Health, 22, 101108.

Wang, Y., Shen, B., Wu, H., Wang, C., Su, Q., & Chen, W. (2021). Modeling illegal pedestrian crossing behaviors at unmarked mid-block roadway based on extended decision field theory. Physica A: Statistical Mechanics and its Applications, 562, 125327.

Zhang, Y., & Fricker, J. D. (2021). Investigating temporal variations in pedestrian crossing behavior at semi-controlled crosswalks: A Bayesian multilevel modeling approach. Transportation research part F: traffic psychology and behaviour, 76, 92-108.

Zhu, H., Almukdad, A., Iryo-Asano, M., Alhajyaseen, W. K., Nakamura, H., & Zhang, X. (2021). A novel agent-based framework for evaluating pedestrian safety at unsignalized mid-block crosswalks. Accident Analysis & Prevention, 159, 106288.