Modeling crossing behavior and accident risk of pedestrians

George Yannis[1], John Golias[2], Eleonora Papadimitriou[3]

Abstract

This paper presents a methodology for modeling pedestrians crossing behavior along an urban trip, as well as an algorithm for the estimation of accident risk along the trip. For that purpose, existing models are exploited and further developed. In particular, a nested logit model and a linear regression model are merged and adapted to develop a hierarchical crossing behavior model, allowing for the estimation of a distribution of crossing probabilities on an urban road link among junctions and various mid-block locations. The explanatory variables concern a set of directly measurable geometric and traffic characteristics. A second model is then developed for the estimation of the distribution of crossing probabilities along a trip in relation to the distance from the trip

[1] Assistant Professor, Department of Transportation Planning and Engineering, National Technical University of Athens, 5, Iroon Polytechniou str., GR-15773 Athens, Greece, Telephone: +302107721326, Fax: +302107721454, E-mail: geyannis@central.ntua.gr

[2] Professor, Department of Transportation Planning and Engineering, National Technical University of Athens, 5, Iroon Polytechniou str., GR-15773 Athens, Greece, Telephone: +302107721326, Fax: +302107721454, E-mail: igolias@central.ntua.gr

[3] Research Assistant, Department of Transportation Planning and Engineering, National Technical University of Athens, 5, Iroon Polytechniou str., GR-15773 Athens, Greece, Telephone: +302107721326, Fax: +302107721454, E-mail: nopapadi@central.ntua.gr

origin. Both models were sufficiently validated by means of appropriate surveys. On the basis of these models, a complete framework for the assessment of pedestrians crossing behavior in urban areas is developed. Moreover, a set of formulae for the calculation of accident risk along a trip in relation to the estimated crossing behavior of pedestrians is proposed.

Key-words: pedestrian; crossing behavior; crossing probability; hierarchical model; accident risk.

CE database subject headings: pedestrians; accidents; models; probability.

Introduction

In road accident analyses it is usually considered that pedestrians' risk exposure when moving along network segments is equal to zero, which is quite realistic since "hit-along-roadway" accidents are in fact a minor percentage of pedestrians' accidents (Duncan et al. 2002). On the contrary, pedestrians' risk exposure is important when moving across network segments, since there is significant interaction between vehicles' and pedestrians' flows.

Previous research on pedestrians' safety in urban environment is extensive and ranges from accidents analyses and modeling to safety treatments evaluation. However, not many attempts of modeling pedestrians' crossing behavior have been made, probably because this behavior is strongly related to complex human factors. In contrast to vehicles flows, which are distributed along fixed corridors of the road environment and are subject to specific traffic rules, pedestrian flows are characterized by a significant degree of randomness (Mitchell and Smith, 2001), so that one could consider that each individual's trip is unique.

However, a pedestrian's need to balance the possibility of an accident with the cost of waiting to cross the road and within the framework of social acceptance yields a non-trivial decision problem, whose analysis may shed light on how humans value their time and their lives, how they perceive their environment and how they interact with it (Manuszac et al., 2005). Moreover, the difficulty to cross a street reflects an aspect of accessibility, which is equally important to pedestrians as to any other transportation mode (Crider et al. 2001).

Within this context, several studies were devoted to the investigation of the impact of various roadway and traffic control features on the behavior and safety of pedestrians. The implementation of signs prompting motorists to yield for pedestrians, in conjunction with advance stop lines and pedestrian-activated amber flashing lights, brought a positive safety effect at multilane crosswalks (Van Houten, Malenfant, 1992). A positive effect was also associated with the construction of speed humps downstream uncontrolled pedestrian crosswalks (Dixon et al. 1997). The effectiveness of fluorescent strong yellow-green pedestrian warning signs in improving safety at mid-block pedestrian crossing areas was also successfully evaluated (Clark et al. 1996). Nee and Hallenbeck (2003) identified significant changes in motorist and pedestrian movements, which mainly resulted from the construction of refuge islands. Keegan and O'Mahony (2003) report a positive impact of timers (waiting countdown) on the number of pedestrians' illegal crossings at traffic controlled junctions. Hakkert et al. (2002) evaluated the effects of a system for detecting pedestrians near the crosswalk zone and for warning drivers by means of flashing lights embedded in the pavement. A similar system, based on microwave pedestrian detectors mounted on traffic signals, also providing an earlier activation or an extension of the pedestrian stage, were evaluated in different sites and countries (Carsten et al., 1998). These results are very useful from a traffic engineering viewpoint; however, several authors noted that, despite the improvements of the road and traffic features creating a safer environment, the unsafe behavior of pedestrians was less affected (Hakkert et al. 2002, Nee and Hallenbeck, 2003). It is therefore necessary to further analyze and the behavior of pedestrians itself, in order to better integrate it into the traffic features evaluations.

Existing research on pedestrian movement and behavior models allows for a general classification of the proposed approaches into two large groups: microscopic simulation models and statistical and econometric models. Pedestrian movements are mainly simulated in cellular automata, through basic kinematics and traffic rules (Schaefer et al. 1998, Blue and Adler, 2001, Hoogendoorn and Bovy, 2002, Weifeng et al. 2003). Moreover, another promising approach concerns multi-agent systems (Batty and Jiang, 1999, Dijkstra et al. 2001, Bierlaire et al. 2003), which treat pedestrians as autonomous agents with vision, cognition and learning capabilities, allowing for more complex interactions and dynamics to be modeled. Although simulation tools are advantageous as far as flexibility is concerned, they are mainly based on rather simple sets of rules and are usually focused on crowd dynamics or route finding models, hardly taking into the crossing behavior of pedestrians.

On the other hand, several interesting and crossing-behavior-specific approaches have been proposed within the framework of statistical models. In most of the cases crossing behavior is considered to be largely determined by the gap-acceptance theory, according to which each pedestrian has a critical gap to cross the street (Manuszac et al., 2005, Hamed, 2001). Another interesting yet more limited approach for identifying crossing preferences concerns pedestrian level-of-service assessment models (Sarkar, 1995, Baltes and Chu, 2002, Guttenplan et al., 2001). In addition, a promising approach is offered by discrete choice models, which correlate crossing decisions to a utility function (Chu et al. 2003, Evans and Norman, 1998, Hine and Russel, 1993). Most of these researches, although providing useful insight in different aspects of pedestrians' behavior, are limited to a 'local behavior' level, while behavior along an entire trip is seldom explored.

Summarizing, there is a need for an overall approach on modeling pedestrian crossing decisions, investigating both where pedestrians are more likely to cross (i.e. crossing probabilities) and why (i.e. the related determinants), allowing for application along an entire trip and enabling a direct link to accident risk analysis. These issues are examined in the present research and a modeling framework is proposed.

Objectives and methodology

Background and basic principles

The objective of this research is the detailed investigation of pedestrians' crossing behavior in urban areas. In particular, this research aims to explore pedestrians' choice of a crossing location from the various options along an urban road link and to identify the main related determinants. Furthermore, it is attempted to extend the analysis and investigate pedestrians crossing behavior along an entire trip. For that purpose, a classical statistical modeling approach is adopted, in order to achieve a sufficient level of detail. However, it is attempted to obtain modeling results that would be exploitable in a pedestrians' simulation framework i.e. that would enable the derivation of appropriate rules for the crossing behavior of pedestrians both at road link and at trip level.

A road link is defined, as in the usual traffic engineering sense, as a road segment between two junctions. Junctions may be traffic light controlled or not, whereas other pedestrian facilities (e.g. marked crosswalks) may be present at junctions or at mid-block. Moreover, a pedestrian trip is defined as a movement along a section including more than one road link, in which a crossing opportunity would be sought.

Concerning the possible crossing options available to a pedestrian moving along a road link, two options {*J1, J2*} are considered for junctions (one for each junction) and three options {*a, b, c*} for the mid-block section. The mid-block crossing options, in particular, are chosen to reflect the variation of traffic conditions within the mid-block section perceived by a pedestrian. Consequently, the section downstream the first junction and the section upstream the second junction are considered to be two options (*a* and *c*), as they correspond to vehicles accelerating from the first junction or decelerating towards the second junction respectively, and these speed variations are perceived by pedestrians. Accordingly, a third mid-block option (*b*) concerns the remaining intermediate section, which corresponds to the average cruise speed of all vehicles moving along the road link.

According to the above, the decision making process as regards road crossing is considered to follow a two-level structure, as shown in Figure 1:

Level 1: Junction or mid-block?

Level 2.1: If junction, which one?

Level 2.2: If mid-block, which section?

Figure 1 to be inserted here

In order to translate this hierarchical structure into a modeling framework, existing models are considered, exploited and further developed.

Merging a nested logit model and a linear regression model

A Nested logit model

With reference to the first level of pedestrian crossing decisions (junction or mid-block), a nested logit model proposed by Chu et al. (2002) is exploited. In this study, a crossing scenario was presented to survey participants for soliciting their stated crossing preferences. The options available included two options for crossing at a junction and up to four options for crossing at a mid-block location. Thus, the nested logit model developed has a two-level structure. The top level has two branches: junctions (J) and mid-block locations (*M*). The bottom level has two options in the junction branch (*A, E*) and up to four options (*B*: cross first - walk later, *C*: jaywalk, *D*: walk first - cross later, *F*: use mid-block crosswalk) in the mid-block branch. Most explanatory variables concern the road street environment, so they can be directly measured for model applications.

The calculation of probabilities also follows a two-level process. The utility U_O for option *O* (*O = A, E; B, C, D, F; I, M*) is defined as the sum of the products of all variables values with the corresponding coefficients. The probability of a crossing option being chosen is the product of its marginal and conditional choice probabilities. The marginal probability represents the probability of choosing junctions or mid-block options. The conditional probability represents the probability of choosing a particular crossing option once the choice has been made between junctions or mid-block options.

The model is well behaved and consistent with utility maximization. All parameter estimates are statistically significant and intuitive, and a satisfactory overall fit (adjusted ρ^2= 0.452) is obtained (Chu et al 2002). Although the model follows the basic structure of the present analysis, an important limitation rises from the fact that the nested level for mid-block includes options describing "how to cross", rather than "where to cross". On the other hand, the top level and the nested level for junctions are in full accordance with the proposed structure. Therefore, the methodology concerning the top level of the model can be applied in the present research for the "junction or mid-block" pedestrian decision. Additionally, the junction options of the bottom level of the nested logit model can be exploited in the framework of the second level of the present research in the "which junction" investigation.

A linear regression model

With reference to the second level "which section at mid-block" investigation, an existing linear regression model will be exploited. In particular, Baltes and Chu (2002) proposed a level of service estimation for mid-block crossings based on pedestrians' perceived difficulty to cross (NCTR, 2001). Crossing difficulty was explained to participants as the risk of being hit by a vehicle. Participants then rated the difficulty to cross at several mid-block locations in a scale from 1 to 6 without actually crossing, after a three minutes observation of traffic conditions.

The ordinary least squares (OLS) statistical method was used to develop a crossing difficulty model in relation to personal, roadway, crosswalks and traffic control characteristics of each mid-block location. All parameter estimates were found to be significant, and a reasonable overall fit (adjusted R^2=0.34) was obtained (Baltes, Chu, 2002). The OLS method was used under the assumption that the ranking is made on the entire (continuous) scale from 1 to 6. Accordingly, by applying the model, one obtains continuous values of crossing difficulty. Modeling results can then be transformed to express a pedestrian level-of-service designation. Each variable-coefficient product may be interpreted as the contribution of each variable to the overall level of difficulty at a particular spot.

However, in order to determine crossing behavior, one needs a mechanism to convert this predicted level of difficulty into a particular crossing probability. Consequently, a method was developed within the present research, in order to transform these modeling results to crossing probabilities at various mid-block locations. If (*n*) crossing locations along a link are considered, the ratio of the perceived difficulty to cross (D_i) at location (*i*) to the perceived difficulty to cross (D_j) at location (*j*) reflects a relative crossing difficulty (RD_{ij}). By expressing all crossing difficulty values in relation to a crossing difficulty of reference (D_r), we obtain relative difficulty ratios among the (*n*) locations. The crossing difficulty of reference should be equal to the crossing difficulty of one of the (*n*) locations.

$$RD_{ir} = \frac{D_i}{D_r} \qquad (2.1)$$

Moreover, one can accept that the ratio of the probability (P_i) to cross at location (*i*) to the probability (P_j) to cross at location (*j*) is equal to the relative crossing difficulty between locations (*i*) and (*j*). It is therefore possible to express all crossing probabilities in relation to a crossing probability of reference (P_r), which should be the probability that corresponds to the location adopted as difficulty of reference, and so:

$$P_i = RD_{ir} * P_r$$

Given that:

$\sum P_i = 1$

$\sum RD_{ir} \times P_r = 1$

Therefore: $P_r \sum RD_{ir} = 1$

$$P_r = \frac{1}{\sum RD_{ir}} \qquad (2.2)$$

According to the above, the value of the crossing probability of reference (P_r) can be calculated when all relative crossing difficulty ratios (RD_{ir}) are known. All crossing probabilities at mid-block can then be calculated on the basis of the crossing probability of reference. It should be noted that these probabilities correspond in fact to conditional probabilities of the mid-block options. In order to obtain the final crossing probabilities along the road link, for each crossing option the product (marginal probability)*(conditional probability) needs to be calculated.

Development of a generalized crossing probability distribution along a trip

In order to model pedestrians' behavior along a trip including more than one road links, it would be necessary to account for the different weights that different links may have in the crossing preferences of a pedestrian. This parameter is less related to the traffic and geometry parameters of the road, making a section more likely to be chosen than another, and more related to the natural tendency of an individual to adopt a particular behavior in relation to the phase of a trip in which he or she is more likely to cross.

In order to incorporate this parameter in the model, some features of the nested logit model are used. In particular, the mid-block options *B, C, D* and *F* of the nested logit model refer to a "how to cross" behavior and were not considered so far. The options *B* and *D* are considered to express the tendency of an individual to cross earlier or later along a link. Two utility functions correspond to these options. When isolating these two functions from the rest of the model, one can notice the following: when the walking distance is different for the two options, different utilities are obtained. However, when the walking distance is the same for the two options, the same utilities are obtained, regardless of the value of the walking distance.

It can therefore be considered that these "cross earlier or later" options of the nested logit model, although obtained through observations on a single road link, may also refer to a pedestrian trip along several urban road links. Therefore, considering a trip of a given length (walking distance) , along a number of links, on which a pedestrian would have to cross, on the basis of the respective utility functions, one can calculate the following crossing probabilities:

B: cross first and walk later P= 0.618

D: walk first and cross later P= 0.382

Moreover, it can be considered that the probability *B* refers to the tendency of crossing at the very beginning of the trip and the probability *D* refers to the tendency of crossing towards the very end of the trip. By attributing these probability values to the starting point (0% of the total length) and the ending point (100% of thetotal length) of the trip respectively, one can obtain a generalized probability distribution for choosing a crossing location along the trip in relation to the total trip distance.

Another consideration may result by including the option *C*: jaywalking in the model, as an expression of the general tendency to randomly cross at any location along the trip and could be attributed to correspond to the middle of the trip (50% of the total distance). In this case, a more complex distribution of crossing probabilities would be obtained.

On the basis of the above, a framework for the assessment of pedestrians crossing behavior along a trip is proposed. Demonstrations of the methodology developed above, as well as results from models validation, are presented in the following sections.

Models demonstration and validation

Modeling pedestrian behavior along a single road link

It was shown that, merging the pedestrians crossing choice nested logit model and the pedestrians level-of-service linear regression model, all the levels of the proposed hierarchy of crossing decisions can be considered, and the related crossing probabilities can be estimated for a road link. Table 1 shows the proposed hierarchical crossing behavior model, with the respective variables and values. More specifically, the first columns shows which model is used (nested logit or OLS), the second column includes the respective explanatory variables, the third column describes the variables definition (i.e. measurement unit and scale) and the fourth column includes the parameter estimate for each variable (as obtained from the nested logit and OLS models). The following columns correspond to the different crossing options of the different levels of the model, and a bullet is used to indicate which variables of which model are used for each crossing option of each level.

In the proposed model, the initial decision (Level 1) is made between junction (*J*) and mid-block (*MB*). A second (Level 2.1) decision is made between the two junctions {*J1, J2*}, whereas three options {*a, b, c*} are considered for mid-block, each one corresponding to the respective sections presented above i.e. (a) is the section downstream junction 1, (b) is the intermediate section and (c) is the section upstream junction 2).

It should be noted that a number of variables' sensitivity tests were carried out for the initial model, resulting in a more flexible final model, which includes fewer variables.

Table 1 to be inserted here

The methodology presented above can be applied to any urban road link, defined among two junctions. By entering the appropriate values in the nested logit model, the respective utilities are obtained and the marginal probabilities for crossing at junction or at mid-block are estimated. Conditional probabilities are then estimated for the junction options. It is noted that utilities for the mid-block options of the initial nested logit model (*B, C, D, F*) are included in the calculations only in order to obtain the related marginal probabilities; however, as explained in previous section (A nested logit model), their probabilities are not meaningful in this research (and so the respective columns in Table 1 are shaded).

For each one of the mid-block options, crossing difficulty is estimated on the basis of the level of service model. By entering appropriate values for each variable at each location in the combined crossing difficulty model, the respective crossing difficulty is obtained. Relative crossing difficulties in relation to crossing difficulty of mid-block option (*a*) as difficulty of reference are then calculated. On the basis of the formulae presented above, the probability to cross at section (*a*) and then the probabilities at all other sections can be estimated. These probabilities correspond to the conditional probabilities of the mid-block options.

An example of the use of the hierarchical crossing behavior model, including calculations of marginal, conditional and final crossing probabilities, using indicative traffic and roadway variables' values, is presented in Table 2. The nested logit and OLS models are used to calculate the various marginal and conditional probabilities, providing the final set of probabilities for all crossing options. It is shown that crossing probabilities are higher at junctions compared to mid-block locations, and even higher at locations with pedestrian facilities (crosswalk markings, pedestrian signal etc.). Moreover, mid-block locations close to the junction areas correspond to higher crossing probabilities, while the pedestrian facilities of junctions appear to also affect the nearby mid-block locations (e.g. increased probability for option *c*).

Such results are demonstrated in Figure 2, where indicative crossing probabilities are estimated for each one of three consecutive road links, assuming that a pedestrian would have to pick a crossing option along these road links.

Table 2 to be inserted here

Figure 2 to be inserted here

Modeling pedestrian behavior along an entire trip

From the previous example it was shown that the application of the hierarchical crossing behavior model on each one of three consecutive road links results to a uniform probability distribution along the trip (e.g. as in Figure 1). According to the demonstration presented so far, all links along a pedestrian trip were considered to be equivalent, in the sense that no preference for a particular link was taken into account. For that purpose, parameters of the nested logit model (options *B* and *D*) were exploited to develop a generalized probability distribution along a pedestrian trip in relation to trip length. The general (linear) formula of this distribution is the following:

$$P(x) = -0.00236x + 0.618 \qquad (3.1)$$

A graphical representation is also given in Figure 3. In the above formula, the value x corresponds to the percentage of the total length of the trip.

Figure 3 to be inserted here

As mentioned in the previous section, a second, more complex consideration can be obtained by including options *B, C* and *D* of the nested logit model, resulting in a 'cross earlier - jaywalk - cross later' distribution. In this case, the general formula of the distribution would be non-linear; more specifically, a parabolic curve would be considered. Within this research, the first, simpler consideration of the distribution is adopted and validation efforts are focused on this linear formula.

The above proposed equation is generic and is considered to apply to any urban pedestrian trip. Considering then that each trip includes several links, the crossing probabilities of the various locations within each link can be weighted in relation to the location of this location within the route.

It should be noted that, whilst the hierarchical crossing behavior model yields dependent (conditional) probabilities summing up to one for each road link, the final crossing probabilities, obtained as the product of crossing probabilities from the hierarchical model and the generalized distribution, produces independent (unconditional) probabilities. It will be shown in the following section that such a consideration is often realistic, especially in relation to accident risk. However, it is a matter of the needs and assumptions of each particular analysis to accept this; otherwise, these final probabilities can be rescaled.

Table 3 shows a complete example of modeling pedestrians' crossing behavior for a trip along three consecutive road links. The first two columns concern the results of the calculation of crossing probabilities for each crossing option of each road link, as obtained by using the hierarchical crossing behavior model and the third column includes the rescaled figures, summing up to one for the sequence of the three road links. Moreover, the next two columns include the results of the generalized linear probability distribution, obtained in relation to the distance from the trip origin for each crossing option. Finally, the last two columns concern the final crossing probabilities (in independent and rescaled form) of the various options of the trip. The results are also graphically presented in Figure 4.

In this example, traffic lights and pedestrian signals were considered in Junction 2 of the first road link (also affecting junction 1 of the second road link) and in Junction 2 of the third link, and consequently the probabilities of these options being chosen are higher, as shown in the single links probability distribution of Figure 4. However, when applying the generalized linear probability distribution along the trip to weight the single road links, an additional aspect of crossing behavior is highlighted. More specifically, the options at and around junction 2 of the first link present a clearly increased final probability of being chosen, as they not only correspond to better pedestrian facilities but they are also located relatively close to the trip origin, and are therefore favored by the general tendency to cross earlier (than later) along the trip. On the contrary, the final crossing probabilities at and around junction 2 of the third road link are slightly reduced, as they correspond to the end of the trip.

Table 3 to be inserted here

Figure 4 to be inserted here

Summary and validation

This research provides a framework for modeling pedestrians' behavior along a trip in urban areas. The proposed approach is hierarchical and includes the following steps:

- Estimation of initial probabilities for each crossing option of each road link of the trip, through the hierarchical crossing behavior model (junction or mid-block, which junction, which section at mid-block), in relation to a number of traffic and road geometry characteristics.
- Estimation of the crossing probabilities for each crossing option of the trip, through a generalized probability distribution, in relation to the distance from trip origin.
- Calculation of the final crossing probabilities, by weighting the initial crossing probabilities with the generalized distribution.

In the framework of possible model validation, three options were exploited. The first two concern surveys in Athens, Greece and in Florence, Italy, in order to validate the hierarchical crossing behavior model. The third one concerns a survey in Lille, France, in order to validate the generalized crossing probability distribution. The results are promising, especially as far as the hierarchical model is concerned. However, it is indicated that further improvement and calibration of the models could be possible.

In particular, the Athens survey was carried out on a road link between traffic light controlled junctions on Solonos street by means of ad-hoc video recordings. The

Florence survey was carried out on a road link between non traffic light controlled (but crosswalk-marked) junctions on Via Galliano by means of video recordings from a Noise Monitoring Station. In each case, the crossing behavior of about 1,800 pedestrians was recorded. In both cases, the estimated crossing probabilities for each option were obtained from the hierarchical model by introducing average traffic values, and were compared to the actual (observed) crossing distributions, as collected from the related sites. Results are presented in Table 4.

Table 4 to be inserted here

The Florence results show that 59.5% of pedestrians actually crossed at junction and 40.4% actually crossed at mid-block sections, while the models average estimated probabilities are 68.1% and 31.9% respectively. At Level 1 (junction or mid-block) and Level 2.1 (which junction) the results are satisfactory, taking into account the related standard deviations. At Level 2.2 (which section at mid-block), though, the estimated probability for option (*b*) is somewhat different from the observed one. According to the Athens results, 87.3% of pedestrians actually crossed at junction and 12.6% actually crossed at mid-block, while the model average estimated probabilities are 91.5% and 8.5% respectively. Taking into account the related standard deviations, the differences between the actual crossings and the estimated distribution are not significant too. Again, the estimated probability is significantly different from the observed one only for mid-block option (*b*). As all other characteristics are the same for all mid-block options, it appears that the measurement of speed variation may be the main factor that limits the model's performance at mid-block option (*b*).

In general, the probabilities to cross at junction are higher on Solonos street than on Via Galliano, due to the presence of traffic lights and pedestrian signals at the junctions. Moreover, on Via Galliano the predicted probabilities for the two junctions are almost equal due to the highly symmetrical link characteristics (no traffic lights at both junctions, marked crosswalks at both junctions).

In the framework of the Lille survey, for the validation of the generalized crossing probability distribution, 79 pedestrians' trips were monitored as far as crossing behavior along the trip is concerned. The survey was carried out in the Villeneuve d' Ascq suburban area of Lille, inside and around the Lille-1 University campus. In particular, 79 pedestrians exiting from the University, the nearby metro station and the City Hall, were followed for five minutes, and the number and location of their road crossings were recorded in relation to the distance from trip origin. From this data, a distribution of the percentage of crossings by the percentage of trip length was calculated.

The results, presented in Figure 5, confirm the general form of the estimated linear distribution and the basic assumption of a tendency to cross earlier (than later) along a trip. Although the linear form of the proposed distribution can be further discussed, the validation results (under the specific conditions) are promising. Moreover, the estimated probability (utility) of crossing at the very beginning of the trip (i.e. trip origin) is fully validated by the observations. However, the estimated probability of crossing at the very end of the trip (i.e. trip destination) is not fully validated, as a difference of around 16% amongst the two distributions is obtained. Although this difference is not dramatic, a potential for further improvement of the distribution could be considered.

Figure 5 to be inserted here

It should be noted that the Lille survey area is not considered to be typically urban, as it includes a part of the university campus with low traffic and few traffic lights and pedestrian facilities, and this may be one reason for the less satisfactory validation of the estimated linear distribution in the second half of the trip length. A more extensive survey, including more observations in a larger and more urban area would certainly allow for more insight in this issue.

Further validation should also focus on the final (weighted) crossings distribution. The Lille data were not suitable for that purpose, as they included a limited number of observations, on trips of different length, number of road links, origin / destination etc. A sufficient number of observations for a number of specific trips would be required, in order to calculate the detailed crossings distribution (i.e. probabilities for each crossing option of the trips). However, the above validation results provide sufficient evidence of satisfactory performance of the model.

Estimation of accident risk of pedestrians along a trip

The proposed model can be applied for the assessment of accident risk along a pedestrian trip, through weighting of the individual accident risks of each crossing option along the trip by its respective probability of being chosen. However, a number of issues need to be addressed for the calculation of the total accident risk.

As mentioned previously, the crossing probabilities for a single road link are dependent probabilities ($\Sigma P_i=1$), in the sense that opting for a particular crossing location presupposes rejection of the other crossing locations on the road link. Assuming that the accident risk R_o of crossing at each location (o) along an urban road link (i) is known, and by calculating the respective crossing probabilities P_o through the hierarchical crossing behavior model, the total accident risk for road link (i) is equal to:

$$R_i = P_{j1}R_{j1} + P_aR_a + P_bR_b + P_cR_c + P_{j2}R_{j2} = \sum P_oR_o \qquad (4.1)$$

However, when examining a sequence of two or more road links, the above consideration would not be realistic. It was mentioned previously that the final crossing probabilities along a trip may or not be considered to be independent, according to the context of the analysis. However, in most cases these probabilities will have to be considered as independent, and be further transformed. This can be explained as follows: A pedestrian moving along a sequence of (n) road links, is seeking an appropriate crossing option in order to move 'across' this trip direction. However, while moving along the (n) links, he or she will probably have to cross roads 'along' the trip direction; these roads are arms of the junctions that separate the (n) road links i.e. the perpendicular road links of the examined three road links (see Figure 6). Therefore, a pedestrian would have to carry out up to (k)=(n)-1 crossings 'along' the trip direction, regardless of the number and location of crossings 'across' the trip direction. Consequently, the crossing probabilities of the $(i)^{th}$ road link (i=1, …, n) are affected by the accident risks of the previously made crossings 'along' the trip direction.

For example, as shown in Figure 6, the crossing probabilities corresponding to road link 2 are affected by the accident risk R_{K1} of the crossing K_1. In particular, crossing at some location on road link 2 excludes crossing on road link 1 but also requires that no accident has occurred on K_1. Accordingly, the crossing probabilities corresponding to road link 3 are affected by the accident risk of the crossings K_1 and K_2.

Figure 6 to be inserted here

It can be said that, while the crossing decisions 'across' the trip direction are probabilistic, in the sense that the various locations have different probabilities of being chosen, the crossing decisions 'along' the trip direction are deterministic, in the sense that a pedestrian is obliged to carry out each one of these crossings in order to follow his route (the only decision concerns on which side of the trip direction they will be made i.e. above or below the trip direction).

Therefore, crossing probabilities for the entire sequence of link sections are independent. In particular, crossing probabilities are considered to be conditional in relation to the accident risk R_{ki} of the previous 'intermediate' crossings (*k*). Therefore, the crossing probabilities corresponding to each option (*o*) of a road link (*i*) of a particular sequence of (*n*) road links are calculated as follows:

$$P_o' = [\prod_{i=1}^{n-1} (1-R_{ki})]\, P_o \qquad (4.2)$$

The accident risk R_i for a link (*i*) is then estimated according to the formula presented above in relation to the accident risks R_o of the crossing options P_o' that take into account the accident risk of the previous 'intermediate' crossings.

$$R_i = \sum P_o' R_o \qquad (4.3)$$

One can therefore consider (2n-1) sequential independent accident risks R_i, corresponding to the (2*n*-1) sequential situations encountered, one for each link (*i*) and one for each 'intermediate' crossing (*k*), as shown in Figure 6. It should be noted that, for practical reasons, R_{Ki} is considered not to vary on either side of the road. Accident risk along the trip is then calculated as follows:

$$R_{1\rightarrow K1} = R_1 + (1-R_1)\, R_{K1}$$

$$R_{1\rightarrow 2} = R_1 + (1-R_1)\, R_{K1} + (1-R_1)(1-R_{K1})\, R_2$$

$R_{1\rightarrow K2}$ = R_1+(1-R_1) R_{K1}+(1-R_1)(1-R_{K1}) R_2+(1-R_1)(1-R_{K1})(1-R_2) R_{K2}

And so on, therefore for the entire trip:

$$R_{1\rightarrow n} = \sum_{i=1}^{n}\left[\prod_{i=1}^{n} R_i(1-R_{i-1})\prod_{i=1}^{n-1}(1-R_{ki})\right] \qquad (4.4)$$

On the basis of this equation, it is possible to estimate the total accident risk on any location along a pedestrian trip. In particular, once the individual accident risks of each location along a trip are known, the calculation of the total accident risk at some location along the trip involves the following steps:

- Estimation of the crossing probabilities of each location along the trip from the hierarchical model of Table 1 and the generalized distribution of Figure 3, as described previously.
- Estimation of the weighted accident risk along the trip from the origin up to the selected location from equation 4.4

The application of this methodology for different planning scenarios would allow for a comparative assessment of planning alternatives in terms of pedestrians' safety.

Discussion

In this research, a hierarchical modeling framework for pedestrians' crossing behavior is developed. The first hierarchical process considered concerns the calculation of crossing probabilities on an urban road link, between junctions and various mid-block locations. A second hierarchical process refers to the adaptation of the model from link-level to trip-level. The complete modeling process allows for a detailed estimation of crossing probabilities for the different options available along a pedestrian trip. Some basic formulae for the calculation of accident risk (and its evolution) along a trip in accordance with the assumptions of the analysis are also provided.

It is demonstrated that accident risk at a particular crossing location also depends on the previous crossing decisions, which make certain crossings (and risks) 'inevitable' (deterministic crossing behavior) while other crossings (and risks) remain subject to behavioral parameters (probabilistic crossing behavior). It was further shown that the first type of behavior concerns crossings 'along' the trip direction, whereas the second type concerns crossings 'across' the trip direction. An algorithm for the aggregation of risk exposure along a pedestrian trip is developed, taking into account the sequence of crossings and related risks within the trip, according to the parameters presented above.

As regards the crossing behavior modeling, two existing models were exploited. Merging, adapting and further developing these models is not always straightforward, and validation is necessary to ensure consistency and validity of the final outcome in relation to the initial assumptions. Given that the validation results of the proposed models are promising and in full accordance with the basic assumptions adopted, it can be said that the exploitation of existing models allowed overcoming a number of limitations in modeling pedestrians' behavior, namely the difficulty in collecting and coding adequate and detailed data on pedestrians' behavior, especially for entire trips.

The proposed methodology, can be used for the calculation of the total accident risk of pedestrian trips within the assessment of different urban and transportation planning schemes. For instance, this methodology could be applied to assess the benefit in terms of pedestrians' safety from the implementation of traffic control at uncontrolled junctions, from the development of pedestrian zones, from the construction of medians etc. Applying the model for "before" and "after" scenarios, one could assess the changes in pedestrians' behavior in terms of crossing decisions and subsequently calculate the related changes in accident risk.

Nevertheless, the proposed methodology could be further improved. In particular, further research should focus on the consideration of more complex trips; the present examples' results refer to the simple case of pedestrian trips on sequential road links along a single direction, on which a crossing opportunity is sought. However, more complex trips including several changes of direction are often carried out in practice. It is explored whether the proposed methodology could also be applied on such trips.

The implementation of the proposed methodology for modeling pedestrians' crossing behavior within a micro-simulation framework would be an interesting field for further research, and would also allow for a more sophisticated consideration of aggregate accident risk. However, further development of the approach (as discussed above), as well as full validation of the complete modeling framework would be required.

Acknowledgements

This research was supported by the European Commission within the "HEARTS - Health effects and risks of transport systems" project of the 5th Framework Programme. The authors would also like to thank Mr. Sylvain Lassarre, Research Director in the French National Institute of Transport and Safety Research (INRETS) for his useful comments and suggestions throughout this research.

References

Baltes M., Chu X. (2002). "Pedestrian level of service for mid-block street crossings". *Proc: TRB 81st Annual Meeting*, *Transportation Research Record No 1818*, Transportation Research Board, Washington, 2002.

Batty, M., Jiang, B.(1999). "Multi-Agent Simulation: New Approaches to Exploring Space-Time Dynamics Within GIS". Working Paper 04/99. Centre for Advanced Spatial Analysis, University College London.

Bierlaire, M., Antonini, G., Weber, M., (2003). *Behavioral Dynamics for pedestrians*. In K. Axhausen (ed): Moving through nets: the physical and social dimensions of travel, Elsevier, 2003.

Blue, V. J., Adler, J.L., (2001). "Cellular Automata microsimulation for modeling bi-directional pedestrian walkways". *Transportation Research Part B* 35.

Carsten O.M.J., Sherborne D.J., Rothengatter J.A. (1998). "Intelligent traffic signals for pedestrians: evaluation of trials in three countries", *Transportation Research Part C* 6 pp. 213-229.

Chu, X., Guttenplan, M., Baltes, M., (2003). "Why people cross where they do - The role of street environment". *Proc: TRB 82nd Annual Meeting*, Transportation Research Board, Washington, 2003.

Clark, L.K., Hummer, J.E., Dutt, N. (1996), "Field evaluation of fluorescent strong yellow-green pedestrian warning signs", *Transportation Research Record* No 1538.

Crider, L., Burden, J., Han, F., (2001). *Multi-modal level of service - 'Point' level of service project - Final Report*. FDOT - Florida Department of Transportation.

Dijkstra, J., Jessurun, A.J., Timmermans, H.J.P., (2001). *A multi-agent cellular automata model of pedestrian movement*. In M. Schreckenberg and S.D. Sharma (ed.): Pedestrian and evacuation dynamics, Springer - Verlag, Berlin, 2001.

Dixon M.A, Alvarez J.A., Rodriguez J., Jacko J.A. (1997). "The effect of speed reducing peripherals on motorists' behavior at pedestrian crossings", *Computers & Industrial Engineering* 33 (1-2) pp. 205-208.

Duncan, C., Khattak, A., Hughes, R., (2002). "Effectiveness of pedestrian safety treatments for hit-along-roadway crashes". *Proc: TRB 81st Annual Meeting*, Transportation Research Board, Washington, 2002.

Evans D., Norman P. (1998). "Understanding pedestrians road crossing decisions: an application of the theory of planned behavior". *Health Education Research* 13 (4).

Guttenplan, M., Landis, B., Crider, L., McLeod, D. (2003), "Multi-modal level of service - Analysis at a planning level", *Proc: TRB 82nd Annual Meeting*, Transportation Research Board, Washington, 2003.

Hakkert, S., Gitelman, V., Ben-Shabat, E. (2002), "An evaluation of crosswalk warning systems: effects on pedestrian and vehicle behavior", *Transportation Research Part F* 5.

Hamed M. (2001). "Analysis of pedestrians behavior at pedestrian crossings". *Safety Science* 38.

Hine J., Russel J. (1993). "Traffic barrier and pedestrian crossing behavior". *Journal of Transport Geography* 1 (4).

Hoogendoorn, S.P., Bovy, P.H.L., (2002). "Normative pedestrian behavior theory and modeling". *Proc: of the 15th International Symposium on Transportation and Traffic Theory in the 21st Century*, Australia, 2002.

Keegan O., O'Mahony M. (2003), "Modifying pedestrian behavior", *Transportation Research Part A* 37*.*

Manuszac M, Manski C., Das S., (2005). "Walk or wait? An empirical analysis of street crossing decisions". *Journal of Applied Econometrics* 20 (4).

Mitchell, D.H., Smith, J.M.G., (2001). "Topological network design of pedestrian networks". *Transportation Research Part B* 35.

NCTR - National Center for Transit research, University of South Florida (2001). "Pedestrian mid-block crossing difficulty", Final Report, Florida Department of Transportation.

Nee, J., Hallenberg, M. (2003), "A motorist and pedestrian behavioral analysis relating to pedestrian safety improvements", Final Report, Research Project T1803, Task 16, Washington State Transportation Commission, USDOT- Federal Highway Administration, 2003.

Sarkar, S., (1995). "Evaluation of safety for pedestrians at macro- and microlevels in urban areas". *Transportation Research Record* No 1502.

Schaefer, L.A., Mackulak, G.T., Cochran, J.K., Cherilla, J.L., (1998). "Application of a general particle system model to movement of pedestrians and vehicles". *Proc: the 1998 Winter Simulation Conference*.

Van Houten R., Malenfant L. (1992), "The influence of signs prompting motorists to yield before marked crosswalks on motor vehicle-pedestrian conflicts at crosswalks with flashing amber", *Accident Analysis & Prevention* 24 (3) pp. 217-225.

Weifeng, F., Lizhong, Y., Weicheng, F., (2003). "Simulation of bi-direction pedestrian movement using a cellular automata model". *Physica A* 321.

Table 1. Crossing options, variables and values of the hierarchical crossing behavior model for a single road link

Model	Variable	Definition	Coefficient	B	C	D	F	J	MB	J1	a	b	c	J2
			Level					LEVEL 1		LEVEL 2				
			Crossing option	B	C	D	F	J	MB	J1	a	b	c	J2
LOGIT	Constant	1	1.000	2.208		1.727	1.388	2.233						
	Walking distance	m	-0.011	•	•	•	•			•				•
	Crossing distance	m	-0.009	•	•	•	•			•				•
	O/D* both at mid-block	1, 0	1.572						•					
	O at midblock-D at junction	1, 0	0.842						•					
	Traffic volume	veh/h	0.000	•	•	•	•							
	Crosswalk marking	1, 0	1.000	•	•	•	•			•				•
	Traffic signal	1, 0	0.750							•				•
	Pedestrian signal	1, 0	1.235							•				•
	Vi		0.759					•						
	Vm		0.834						•					
OLS	Constant	1	-2.478								•	•	•	
	Nears Traffic Vol	1000 veh/h	-0.116								•	•	•	
	Fars Traffic Vol	1000 veh/h	0.267								•	•	•	
	Speed	km/h	0.007								•	•	•	
	Nears Cros Width	m	-0.280								•	•	•	
	Fars Cros width	m	0.407								•	•	•	
	Crosswalk	1, 0	-0.276								•	•	•	
	Pedestrian signal	1, 0	-0.493								•	•	•	
	Signal spacing	m	0.002								•	•	•	

• values to be entered
* O/D: origin / destination

Model	Variable	Definition	Coefficient	B	C	D	F	J	MB	J1	a	b	c	J2
			Level					**LEVEL 1**		**LEVEL 2**				
			Crossing option	B	C	D	F	J	MB	J1	a	b	c	J2
Model	**Variable**	**Definition**	**Coefficient**											
LOGIT	Constant	1	1.000	2.208		1.727	1.388	2.233						
	Walking distance	m	-0.011	0	0	0	0			0				100
	Crossing distance	m	-0.009	12	12	12	12			12				12
	O/D both at mid-block	1, 0	1.572						0					
	O at midblock-D at junction	1, 0	0.842						1					
	Traffic volume	veh/h	0.000	1500	1500	1500	1500							
	Crosswalk marking	1, 0	1.000	0	0	0	0			1				1
	Traffic signal	1, 0	0.750							0				1
	Pedestrian signal	1, 0	1.235							0				1
	Vi		0.759					2.1136						
	Vm		0.834						2.4255					
	Utility Function	∑ (Var * Coef)		**1.652**	**-0.556**	**1.170**	**0.831**	**3.836**	**2.865**	**0.894**				**1.764**
	Marginal probabilities							**0.725**	**0.275**					
	Conditional probabilities									**0.295**				**0.705**
OLS	Constant	1	-2.478								1	1	1	
	Nears Traffic Vol	1000 veh/h	-0.116								1.5	1.5	1.5	
	Fars Traffic Vol	1000 veh/h	0.267								1.5	1.5	1.5	
	Speed	km/h	0.007								10	18	12	
	Nears Cros Width	m	-0.280								6	6	6	
	Fars Cros width	m	0.407								6	6	6	
	Crosswalk	1, 0	-0.276								0	0	1	
	Pedestrian signal	1, 0	-0.493								0	0	1	
	Signal spacing	m	0.002								100	100	100	
	Crossing Difficulty	∑ (Var * Coef)									**-1.189**	**-1.135**	**-1.945**	
	Relative crossing difficulty (a)										1.000	0.955	1.636	
	Crossing probability (a)										0.278			
	Conditional Probabilities										**0.278**	**0.266**	**0.456**	
	Final Probabilities (marginal*conditional)									**0.214**	**0.076**	**0.073**	**0.125**	**0.511**

Table 2. Example of calculation of crossing probabilities for a single road link

Table 3. Example of calculation of crossing probabilities distribution along a trip

	Single links			Entire trip			
Road link	Option	Probabilities A	Scaled probabilities Asc=A/totalA	% trip distance B	Generalized probability distribution C=-0.00236B+0.618	Final probabilities D=A*C	Scaled probabilities Ds=D/totalD
1	J11	0.301	0.100	0	0.618	0.186	0.123
	a1	0.007	0.002	3	0.611	0.004	0.003
	b1	0.015	0.005	11	0.592	0.009	0.006
	c1	0.082	0.027	19	0.572	0.047	0.031
	J21	0.595	0.198	22	0.566	0.337	0.222
2	J12	0.632	0.211	25	0.559	0.353	0.233
	a2	0.082	0.027	28	0.552	0.045	0.030
	b2	0.064	0.021	47	0.507	0.032	0.021
	c2	0.091	0.030	69	0.454	0.041	0.027
	J22	0.131	0.044	72	0.448	0.059	0.039
3	J13	0.266	0.089	75	0.441	0.117	0.078
	a3	0.026	0.009	78	0.434	0.011	0.007
	b3	0.015	0.005	86	0.415	0.006	0.004
	c3	0.085	0.028	97	0.389	0.033	0.022
	J23	0.608	0.203	100	0.382	0.232	0.153
	Total	3.000	1.000		Total	1.514	1.000

Table 4. Validation of the hierarchical crossing behavior model for single links

(Via Galiano - Florence, Solonos str. - Athens)

	LEVEL 1		LEVEL 2					
Via Galliano (Florence)	J	MB	J1	a	b	c	J2	**TOTAL**
Observed Crossings Distribution								
Number of observations	1,114	756	668	226	327	203	446	1,870
%	59.6	40.4	35.7	12.1	17.5	10.9	23.9	100.0
Estimated Crossing Probabilities								
Average (%)	68.1	31.9	36.2	10.8	10.5	10.6	32.0	100.0
Std.Deviation (%)	15.3	15.3	16.2	5.2	5.0	5.1	17.2	
Solonos str. (Athens)								
Observed Crossings Distribution								
Number of observations	1,566	227	1,276	69	125	33	290	1,793
%	87.3	12.7	71.2	3.9	7.0	1.8	16.2	100.0
Estimated Crossing Probabilities								
Average (%)	91.5	8.5	74.4	2.9	2.8	2.8	17.1	100.0
Std Deviation	2.0	2.0	20.5	0.7	0.7	0.7	19.5	

List of Figures:

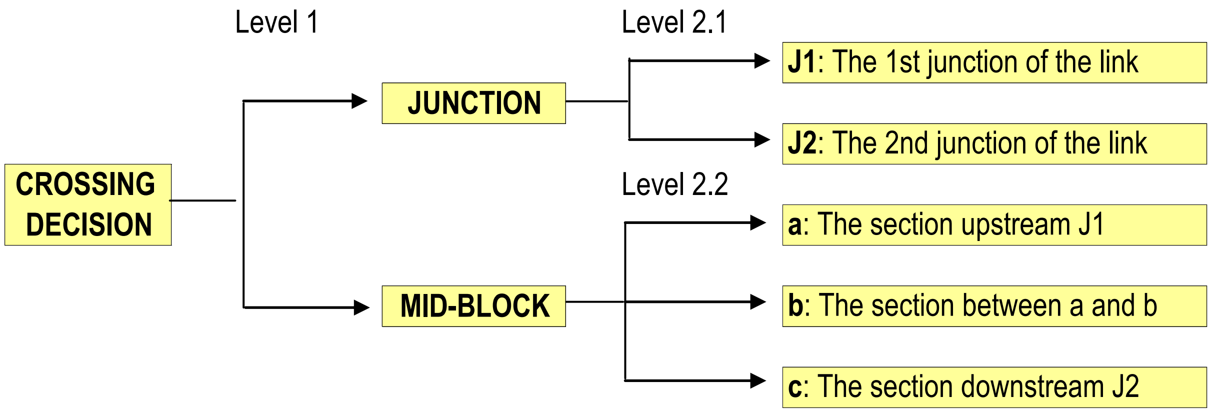

Figure 1. Overview of the hierarchical crossing behaviour model

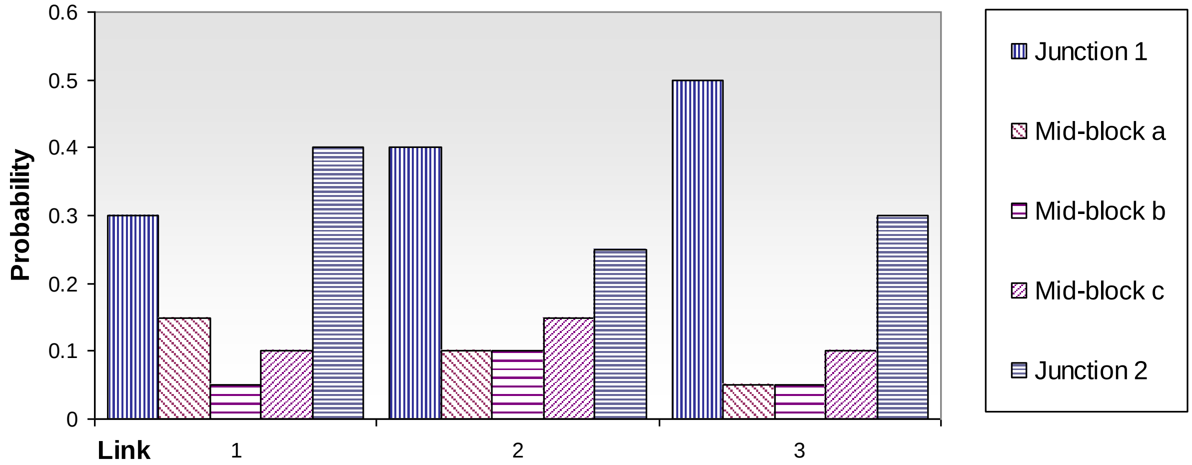

Figure 2. Crossing probabilities (junctions and mid-block) for three consecutive road links

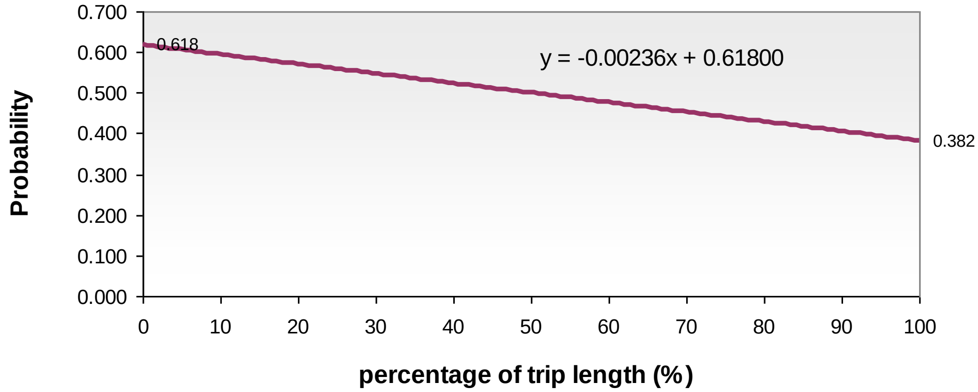

Figure 3. Crossing probability distribution along a trip

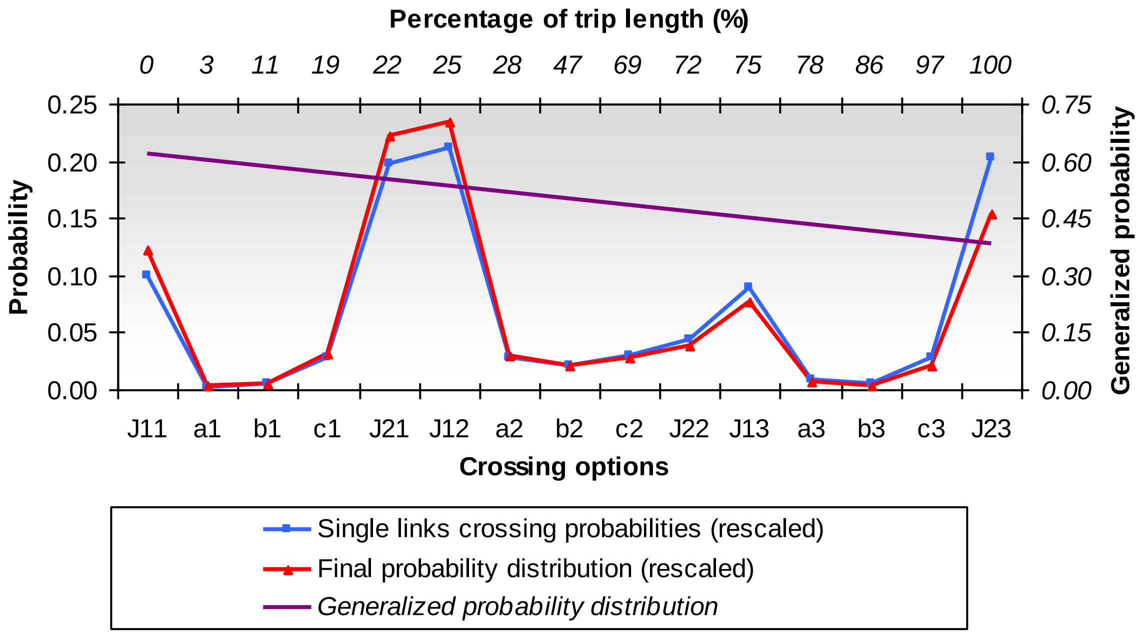

Figure 4. Final crossing probabilities (junctions and mid-block) along a trip

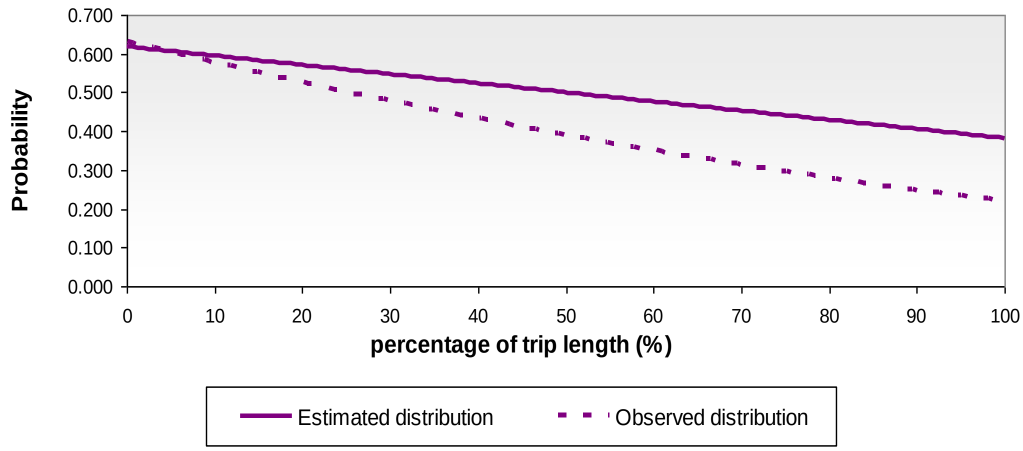

Figure 5. Validation of the crossing probability distribution along a trip (Lille, France)

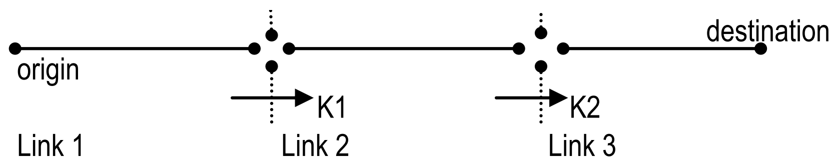

Figure 6. Crossings along a trip direction