Analysis of vehicle-bicycle interactions at unsignalized crossings: A probabilistic approach and application

Ary P. Silvano, Haris N. Koutsopoulos, Xiaoliang Ma

A Department of Transport Science, Division of Transport Planning, Economics and Engineering, KTH Royal Institute of Technology, Teknikringen 72, Stockholm 10044, Sweden
b Department of Civil and Environmental Engineering, Northeastern University, Boston, MA 02115, United States

ABSTRACT

In the last decades, bicycle usage has been increasing in many countries due to the potential environmental and health benefits. Therefore, there is a need to better understand cyclists’ interactions with vehicles, and to build models and tools for evaluating multimodal transportation infrastructure with respect to cycling safety, accessibility, and other planning aspects. This paper presents a modeling framework to describe driver-cyclist interactions when they are approaching a conflicting zone. In particular, the car driver yielding behavior is modeled as a function of a number of explanatory variables. A two-level hierarchical, probabilistic framework (based on discrete choice theory) is proposed to capture the driver’s yielding decision process when interacting with a cyclist. The first level models the probability of the car driver perceiving a situation with a bicycle as a potential conflict whereas the second models the probability of yielding given that a conflict has been perceived by the driver. The framework also incorporates the randomness of the location of the drivers’ decision point. The methodology is applied in a case study using observations at a typical Swedish roundabout. The results show that the conflict probability is affected differently depending on the user (cyclist or driver) who arrives at the interaction zone first. The yielding probability depends on the speed of the vehicle and the proximity of the cyclist.

© 2016 Elsevier Ltd. All rights reserved.

1. Introduction

Public agencies around the world promote bicycle usage because of the potential environmental and health benefits (Edwards and Mason, 2014; Fishman et al., 2012). In Stockholm, for example, the number of cyclists has increased more than 100% in the last 20 years for trips crossing the city cordon (Börjesson and Eliasson, 2012). In order to promote bicycle usage, the Swedish government has established two main goals: (i) make bicycling safer; and (ii) increase the bicycling share as a mode of transport (Trafikverket, 2011). To accomplish these goals the Swedish Transport Administration and cities traffic departments have made considerable efforts to introduce safe bikeways and improve the bicycle network accessibility and associated facilities.

However, bicycling in many cases is at a disadvantage as a viable transportation mode due to lack of network accessibility, impact of weather conditions, safety, and lack of support for multimodal trips that include bicycling as one of the modes. Recently, some of these disadvantages have been addressed with the introduction of bicycle sharing systems (BSS), which allow for spatial and temporal usage and multimodal connectivity providing for short-term one-way transportation. Once properly introduced, BSS can provide support for opportunistic trips complementing public transport networks (Vogel et al., 2011). Such systems have been introduced in more than 400 cities around the world, and many other cities are also planning their introduction (Faghhi-Imani et al., 2014; Frade and Ribeiro, 2014; Jäppinen et al., 2013).

Network accessibility is often a serious impediment that reduces bicycle use. In order to increase bicycle usage a well-connected bicycle network is essential. Common accessibility problems include pavement condition, network discontinuities, bus-stops blocking bicycle lanes or paths, congestion, work zones, and in general interactions with vehicles (Gustafsson and Archer, 2013). Therefore, infrastructure planning, maintenance programs, and policies and regulations are important factors to create a good environment for cycling. For example, in Europe higher costs of car ownership, limited parking spaces, car-free zones, traffic calming measures, and lower speed limits encourage bicycle usage (Pucher et al., 2010).
Cyclists are vulnerable road users, and cycling safety is an essential concern in traffic planning. Among others, the interaction with motorized vehicles has attracted a lot of attention. For instance, one of the most common vehicle-bicycle accidents is with driver turning right and cyclists approaching from behind on the right side of the driver (Räsänen and Summala, 1998). Kim et al. (2007) state that speed is the key factor for serious and fatal outcomes in vehicle-bicycle accidents. Furthermore, children and elderly cyclists are at most risk. The authors find that higher vehicle flows and higher cyclist volumes increase the accident rate. On the other hand, the authors state that the proper design of bicycle facilities (e.g., illumination and wider paths) is an important means to improve cyclist safety. Minikel (2012) points out that higher speeds, higher traffic volumes, and presence of heavy vehicles are detrimental to cyclist safety. Different transport facilities have also been evaluated for cyclist safety. For instance, at roundabouts, vehicle speed and the age of the roundabout are important predictors of bicycle related accidents (Hels and Orozova-Bekvold, 2007).

At unsignalized intersections governed by priority rules, the interaction between car drivers and cyclists is often based on expectations. For example, drivers often yield to cyclists. Björklund, 2005 states however, that sometimes the expectations of the road users can be wrong: some fail to look for a specific road user (e.g. failing to give way) and some fail to look in some specific direction. T-junctions are examples where car drivers usually fail to fulfil other road users’ expectations. For instance, drivers, when approaching from the connecting street and turning right at T-junctions, pay more attention to cars coming from the left and less to cyclists and pedestrians on the right (Björklund, 2005).

It has been reported in the literature that the most frequent accident type occurring in interactions between bicycles and vehicles is that of the car turning right and meeting a cyclist approaching from behind on the right side of the driver (Räsänen and Summala, 1998). In a later study, Räsänen and Summala (2000) also report that vehicles travelling at a high speed fail to look for cyclists. According to the literature, high speed vehicles have lower probability to yield to cyclists due to less time to detect and react to the cyclists’ presence (Silvano et al., 2014, 2015). In Sweden, in about 42% of the cases drives do not yield to cyclists (Svensson and Pauna, 2010). Furthermore, the common expectation in vehicle-bicycle interactions at unsignalized intersections, including roundabouts with crosswalks, is that drivers exiting the roundabout yield to cyclists. However, drivers may fail to fulfil such expectation due to lack of attention (e.g. drivers do not realize the presence of the cyclist due to limited vision or lack of attention). Cyclists may also misinterpret the situation (often assuming that drivers are aware of their presence). These cases have been documented in the traffic safety literature as ‘looked-but-failed-to-see-errors’ (Herslund and Jörgensen, 2003). In addition, cyclists overestimate by a factor of 2 the distance at which they would be recognized by drivers (Wood et al., 2009).

Although the interaction between drivers and cyclists is important for traffic safety, the decision process of a driver to yield to a cyclist is not well understood and has received little attention in the literature. The main objective of this paper is to develop a theoretical framework to model the driver-cyclist interaction process and apply it to a specific case where actual data of such interactions were collected.

The rest of the paper is organized as follows: Section 2 presents a theoretical framework using a probabilistic approach to model the driver yielding decision process; Section 3 illustrates the proposed methodology through an application in modeling vehicle yielding probability at a typical roundabout in Stockholm using actual data collected at the facility; Section 4 discusses the corresponding results, and Section 5 concludes the paper.

2. Methodology

2.1. Vehicle-bicycle interaction and driver yielding decision process

Vehicle-bicycle interaction is triggered by the potential collision course on which the vehicle and bicycle would have been involved if current trajectories had been maintained. Normally, at unsignalized intersections and roundabouts in urban environments, the areas where vehicle and bicycle trajectories intersect are the crosswalks. Therefore, the crosswalk is considered as the Conflict Zone (CZ) where the vehicle and bicycle trajectories may intersect with a potential collision. In order to avoid an accident due to potentially intersecting trajectories, a driver’s decision process begins at some distance upstream the CZ with the driver deciding whether the situation presents a potential conflict or not. If drivers perceive the interaction as a potential conflict, they have to further decide whether to yield or not. The decisions are impacted by, among other factors, the vehicle and bicycle speeds, the vehicle and bicycle relative distances to the crossing, etc. Vehicle and bicycle Interaction Zones (IZ) are also defined and their lengths may differ depending on geometric design and other factors. Fig. 1 depicts a vehicle and bicycle interaction as they approach a conflict zone at a typical unsignalized crossing.

For the purposes of this paper, the decision process by the driver to yield to the bicycle is a hierarchical process as follows:

- A conflict decision (C) with potential collision.
- A yielding decision (Y) given that the conflict is perceived by the driver.

The conflict decision (C) is not observed, and is modeled as a latent state. On the other hand, the yielding decision can be observed i.e., a driver stops or adjusts the speed to allow a cyclist to traverse safely the conflict zone. The point where the yielding decision process commences is also not observable and varies from driver to driver and impacted by the actual conditions. It is therefore, modeled as a random variable following a given distribution whose parameters can be estimated from observation of the trajectories and yielding decisions.

The driver-cyclist interactions are similar in some aspects to pedestrian-car interactions. In the pedestrian case, driver behavior has been explained by factors such as vehicle speed, relative position of the pedestrian, pedestrian Platooning and geometric characteristics (Schroeder and Roup hail, 2011). However, the two
problems have a number of differences as well. Pedestrians have different dynamics compared to cyclists (for example, pedestrian cruising speed is on average 1.2 m/s, HCM, 2010). As a result, in the pedestrian case, the stop/yielding event is triggered by the pedestrian being close to the crossing zone, while the driver pays attention to the crossing to detect the pedestrian. In the vehicle-vehicle case though, drivers need to pay attention not only to the crossing but farther into the sidewalk or bicycle path for the possible presence of cyclists in order to be able to come to a full stop or decrease their speed accordingly. Furthermore, general traffic models (vehicle–vehicle interactions), such as gap acceptance models, are not appropriate to apply in the vehicle-bicycle problem because cyclists are vulnerable users (who have priority in many cases) and the involved dynamics are of different scale. According to the traffic rules in Sweden, drivers should yield to cyclists when entering/exiting a roundabout without traffic lights. In other words, cyclists have priority over the traffic stream and drivers are therefore expected to yield or adjust their speed when interacting with cyclists (Svensson and Pauna, 2010).

2.2. Modeling framework

Based on the discussion in the previous section the driver decision process during the vehicle-bicycle interaction is represented by decision tree in Fig. 2.

In addition to the conflict decision being latent, as mentioned in the previous section, the location \( l_n \) where the yielding decision is made is also unobserved and varies across drivers. In general, it is a function of various factors, such as driver characteristics, intersection design, visibility, weather conditions, relative speeds and locations of the entities involved, etc. Therefore, the driver’s yielding decision process (at unsignalized intersections) can be generally modeled by the stochastic nature of the yielding decision and randomness of the location where the decision process starts.

Given the yielding decision point located at distance \( l_n \) from the conflicting zone, the probability of observing yielding decision \( Y_n \) is given by:

\[
P(Y_n = 1|l_n) = P(Y_n = 1|C_n = 1; l_n) \cdot P(C_n = 1|l_n)
\]

(1)

And

\[
P(Y_n = 0|l_n) = P(Y_n = 0|C_n = 1; l_n) \cdot P(C_n = 1|l_n) + P(C_n = 0|l_n)
\]

(2)

The probabilities at any level of the decision tree in Fig. 2 can be modeled using the discrete choice framework (Ben-Akiva and Lerman, 1985). This framework has been applied in the literature to a wide range of decision problems, including drivers’ decisions in other settings e.g., decisions to overtake (Farah and Toledno, 2010); decisions to accelerate/do-nothing/decelerate (Koutsopoulos and Farah, 2012); and lane changing decisions (Karandeep and Baibling, 2012). In the proposed modeling framework, each alternative has a utility associated with it and assuming that the error term is Gumbel distributed the corresponding probabilities are given by a binary logit model.

At the first level of the decision tree (Fig. 2) the probability of the driver deciding that a conflict exists is given by a binary logit model.

\[
(C_n = 1|l_n) = \frac{e^{V_{1n}l_n}}{e^{V_{1n}l_n} + e^{V_{0n}l_n}}
\]

(3)

where,

\( V_{1n} \): systematic utility of decision i (i = 1 conflict) by driver n;

a function of explanatory variables: \( V_{1n} = \alpha \cdot X_{1n} \); \( \alpha \) is a vector of parameters to be estimated and \( X_{1n} \) a vector of explanatory variables related to decision i by driver n.

One of the decisions \( (C_n = 0) \) is used as the basis and its utility is set to 0. Hence,

\[
P(Y_n = 1|l_n) = \frac{e^{V_{1n}l_n}}{1 + e^{V_{1n}l_n}} = \frac{e^{\alpha X_{1n}}}{1 + e^{\alpha X_{1n}}}
\]

(4)

Similarly, at the second level of the decision hierarchy the probability of yielding \( (Y_n = 1) \) is given by

\[
P(Y_n = 1|C_n = 1; l_n) = \frac{e^\beta Z_{in}}{1 + e^\beta Z_{in}}
\]

(5)

where,

\( \beta \) : Vector of parameters to be estimated

\( Z_{n} \); Vector of explanatory variables related to the yielding decision i (i = 1 or 0) by driver n.

Observed data that can be used for the estimation of such a model typically include vehicle and bicycle speeds and trajectories, and the observed decisions (i.e. decision to yield, \( y_n \)). Demographic characteristics of the driver and cyclist are not easily available.

Most explanatory variables \( X_{0n} \) and \( X_{1n} \) depend on the distance of the decision point \( l_n \) from the conflicting zone. To capture the stochasticity of the point where the decision process starts \( l_n \) is modeled as a random variable across the population of drivers. For example, \( l_n \) may follow a normal distribution with the mean influenced by driver characteristics (e.g. age, gender, trip purpose), intersection design (e.g. number of lanes, lane width), vehicle speed, etc. If related data is available (e.g. from facilities with diverse geometric characteristics), the impact of different factors influencing the location of the decision point can also be estimated along with the other model parameters. In general, \( l_n \) may be bounded between \( l_{min} \) and \( l_{max} \) values due to physical limitations, visibility distance, etc. Hence, the distance decision \( l_n \) to \( f(l_n) \) the conflict zone across the population of drivers has a probability density function

\[
f(l_n) = \begin{cases} f(\cdot) & l_{min} < l_n < l_{max} \\ 0 \text{otherwise} & \text{otherwise} \end{cases}
\]

(6)

In general, as mentioned above, the mean of the distribution is defined as a function of explanatory variables and parameters to be estimated. In the absence of such data the distribution can be characterized by its standard parameters with fixed values (to be estimated).

The probability of a driver n making a yielding decision \( y_n \) is obtained by combining Eqs. (1) and (6):

\[
P(Y_n = 1) = \int_{l_{min}}^{l_{max}} P(Y_n = 1|l_n) \cdot f(l_n) \, dl_n
\]

(7)

The likelihood function of obtaining the observed decisions is given by the joint probability of observing the \( N \) independent vehicle-bicycle events.

\[
P(y_1, \ldots, y_N) = \prod_{n=1}^{N} \left( \int_{l_{min}}^{l_{max}} (P(Y_n = 1|l_n))^{y_n} (P(Y_n = 0|l_n))^{1-y_n} \cdot f(l_n) \, dl_n \right)
\]

(8)

where, \( N \) is the sample size of observed vehicle-bicycle interactions, and \( y_n \), as noted before assumes a value 1 if a yielding decision is observed (and 0 otherwise).

The log-likelihood function can be expressed as:

\[
\mathcal{L} = \sum_{n=1}^{N} \log \left( \int_{l_{min}}^{l_{max}} (P(Y_n = 1|l_n))^{y_n} (P(Y_n = 0|l_n))^{1-y_n} \cdot f(l_n) \, dl_n \right)
\]

(9)
The model parameters are estimated by maximizing (9).

3. Case study

For model estimation purposes, vehicle and bicycle trajectory data, subjects’ speeds, yielding decisions, and intersection design characteristics are required. Data from different unsignalized intersections could provide variability in geometric characteristics to estimate their impact on behavior. However, the collection of such data over a number of unsignalized intersections is costly and time consuming task. Consequently, data from a single intersection, a one-lane roundabout in Stockholm, where vehicle-bicycle interactions are frequently observed during the rush hour, were used.

The data were extracted from video recordings. Due to technical constraints detailed vehicle and bicycle trajectories were not extracted. Only data about the presence of vehicles and bicycles at certain discrete locations in the intersection were recorded. Due to these data limitations, the application of the model developed in Section 2 was simplified by assuming that the yielding decision was made when the vehicle is at a given fixed point \( l \). Under this assumption the likelihood function becomes:

\[
L = \sum_{n=1}^{N} \log \left( P \left( Y_n = 1 | l_n = l \right) Y_n \right) \left( P \left( Y_n = 0 | l_n = l \right) \right)^{1-Y_n}
\]  

(10)

3.1. Data description

The data used in this case study was collected during the autumn 2013 by means of video cameras at an unsignalized roundabout (Fig. 3). In the last decades, roundabouts have become a popular traffic facility in Sweden and are commonly used to implement speed reduction measures. According to the Swedish Statistik Central Byrå (2016), the number of roundabouts in the country was doubled over a period of 5 years, from 1360 roundabouts in 2005–2010 to a number of 2800 roundabouts in 2010. In suburban areas, roundabouts are the place where vehicle-bicycle interactions take place routinely.

As Fig. 4(a) illustrates, drivers exiting the roundabout primarily consider the incoming traffic in the entry lane, though they have priority over that traffic stream. Afterwards, drivers have a clear sight into the sidewalk to detect the presence of cyclists upstream on the sidewalk or bicycle path. The driver’s sight line is crucial for detecting the presence of cyclists as well as the angle of the bicycle path with the conflict zone. Their visibility distance depends on several factors such as driver characteristics (e.g., experience, age), vehicle speed, facility conditions (e.g., illumination), and intersection design characteristics. The decision to yield takes place a few meters away from the CZ (Fig. 4b). Therefore, based on the intersection design of the roundabout, the decision point was set to 10 m away from the CZ. This distance allows drivers a clear sight upstream the sidewalk. It should be pointed out that the sidewalk is a shared space. The inner half is dedicated to pedestrians and the outer half (next to the road) is used by cyclists. The intersection design encourages cyclists to ride in a parallel trajectory with the approaching vehicles and then make a sharp left turn to cross the crosswalk. This situation makes cyclists, in many cases, ride on the sidewalk halfway the crosswalk which results in the bicycle IZ starting earlier relative to the vehicle IZ (which starts at the beginning of the crosswalk). Additionally, the sidewalk follows a curve at 90° when reaching the minor street as shown in Fig. 4(a) and (b). Drivers can detect and track a cyclist from the minor street all the way to the crosswalk. Therefore, the bicycle IZ begins at the middle of the crosswalk and extends through the sidewalk parallel to the road and finally turns into the minor street. The bicycle IZ is 30 m long comprising 3 segments of 10 m each.

We assume that a vehicle–bicycle interaction exists whenever a driver and a cyclist are in their interaction zones (mainly because of the limitations of the available data). For instance, if a vehicle arrives at the car IZ and no cyclist occupies the bike IZ, no interaction

---

**Figure 2.** Vehicle overall yielding decision process.

**Figure 3.** Study area.
event occurs. On the other hand, if a vehicle and bicycle are present in the corresponding interaction zones simultaneously, an interaction event takes place (and a conflict situation may be perceived by the driver).

The data was collected over a period of 2 h in the afternoon peak. Semi-automated video analysis software, SAVA, was used to extract section speeds (Archer, 2003). After distance calibration, virtual lines are drawn at desired sections and speeds calculated from the distance between the virtual lines drawn and the times the vehicle crossed the corresponding lines. Fig. 4(a) illustrates the locations where vehicle and bicycle speeds were measured respectively. In particular, vehicle speeds were measured at the decision point (i.e., at 10 m before the CZ); whereas, bicycle speeds were measured at 10, 20 and 30 m away from the CZ. Based on the above assumptions vehicle–bicycle interactions were extracted. During the two hours of data collection, 187 interaction events were identified out of which 37 yielding decisions were observed.

Fig. 5 illustrates vehicle and bicycle speeds based on observed yielding decisions. For example, Fig. 5(a) shows the cumulative distribution of speeds for drivers who yielded to cyclists compared to drivers who did not. The group of drivers who did not yield comprises drivers who did not encounter any approaching cyclist thus travelling undisturbed through the CZ and drivers who encountered a cyclist but did not yield. Fig. 5(b) depicts the cumulative distribution of the bicycle speeds. Two distributions are shown, one for the speed of cyclists when drivers yielded to them and one for the speeds of cyclists when drivers did not yield. In general, Fig. 5(a) shows, on average, a speed difference of 9 km/h between drivers who yielded to an interacting cyclist and drivers who did not. On the other hand, Fig. 5(b) points out that only a small difference of 1 km/h between cyclists involved in yielding decisions and those who were not was observed. The small difference can be an indication of risky behavior from cyclists who expect that vehicles will yield in their presence and hence do little to adjust their speeds.

3.2. Model specification

3.2.1. Conflict decision

A number of explanatory variables can be used to explain the probability of a conflict. The arrival-time at the boundary of each interaction zone is a key factor. Moreover, the probability of the driver perceiving a conflict is impacted by whether the car or the bicycle arrives first at the interaction zones. It is expected that the arrival time difference (ATD) between the car and the bicycle impacts the probability of conflict. Smaller differences in arrival times imply stronger perception of conflict. However, the model specification is quite general and can capture non-symmetric effects, depending on whether the vehicle or the bicycle arrives first. In order to capture these effects, two variables are defined as follows:

\[ x_{ATD}^{1} = \max(0, T_{car} - T_{bike}) \]  
\[ x_{ATD}^{2} = \min(0, T_{car} - T_{bike}) \]

Where \( T_{car} \) is the arrival time of the vehicle at the car IZ and \( T_{bike} \) is the arrival time of the bicycle at the bicycle IZ. \( x_{ATD}^{1} \) represents the case when the bicycle arrives first with only positive values by definition; and \( x_{ATD}^{2} \) denotes the situation where the vehicle arrives first at the boundary of the interaction zone.

The probability that a driver perceives a potential conflict, assumes maximum value when the car and bicycle arrive at their interaction zones at the same time. However, it can also be argued that in the cases where cars and bicyclists arrive at the interaction zones at approximately the same time, drivers may not be able to perceive the arrival time exactly. Therefore, and also because of the risk involved, the decision of a conflict may be impacted equally for a range of arrival time differences. This general situation is captured by the following variables:

\[ x_{ATD}^{1} = \max(t_{1}, T_{car} - T_{bike}) \]  
\[ x_{ATD}^{2} = \min(t_{2}, T_{car} - T_{bike}) \]

Where, \( T_{car}, T_{bike}, x_{ATD}^{1} \), and \( x_{ATD}^{2} \) are defined as previously. \( t_{1} \) and \( t_{2} \) represent thresholds of fixed values.

Therefore, the impact of the ATD on the decision that a conflict may exist, is constant for a range of ATD values near zero defined by thresholds \( t_{1} \) and \( t_{2} \), as shown in Fig. 6.

Table 1 summarizes a number of explanatory variables that were tested during the model specification phase. The variables capture different aspects of the interaction between cars and bicycles.

3.2.2. Yielding decision

Contrary to the conflict decision, yielding events can be observed, i.e., a vehicle stops or adjusts its speed to let a cyclist traverse safely the conflict zone. Therefore, provided that a conflict is perceived, the driver needs to make a decision with two possible mutually exclusive outcomes: the driver yields to the conflicting cyclist or not (see Fig. 2).

Factors that potentially influence yielding decisions include the speed of the subject vehicle, the speed of the bicycle, the relative positions of the vehicle and bicycle with respect to the conflict zone, and the relative speed and position of the bicycle with respect to the vehicle at the decision point. Table 2 summarizes various variables and their definitions.
Fig. 5. Vehicle and bicycle cumulative speed distributions.

Table 1
Potential explanatory variables for modeling the decision that a conflict is present.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Parameter</th>
<th>Unit</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x_1^{TD}$</td>
<td>$\alpha_1$</td>
<td>sec</td>
<td>Arrival time difference when bicycle arrives first</td>
</tr>
<tr>
<td>$x_2^{TD}$</td>
<td>$\alpha_2$</td>
<td>sec</td>
<td>Arrival time difference when car arrives first</td>
</tr>
<tr>
<td>$T_{Tcar}$</td>
<td>$\alpha_3$</td>
<td>sec</td>
<td>Vehicle travel time within the interaction zone</td>
</tr>
<tr>
<td>$T_{Tbike}$</td>
<td>$\alpha_4$</td>
<td>sec</td>
<td>Bicycle travel time within the interaction zone</td>
</tr>
<tr>
<td>$T_{TCar}$</td>
<td>$\alpha_5$</td>
<td>sec</td>
<td>Vehicle time to reach the conflict zone if speed is maintained from the boundary of the car IZ</td>
</tr>
<tr>
<td>$T_{Tbike}$</td>
<td>$\alpha_6$</td>
<td>sec</td>
<td>Bicycle time to reach the conflict zone if speed is maintained from the boundary of the bike IZ</td>
</tr>
</tbody>
</table>

Other variables were extracted from the video recordings and tested as well, such as expected arrival time at the CZ zone, number of cyclists in each yielding event, etc. Categorical variables were also considered, e.g. vehicle speed lower than selected thresholds, vehicles travelling in platoon. Site geometric characteristics play an important role in the yielding process (e.g., 1-lane, 2-lane...
facilities; raised or non-raised crosswalks; painted, non-painted bicycle lanes; angle of intersecting approaches; road and sidewalk width and gradient, etc.). However, data from different facilities are needed to account for such variability, which were not available for this study.

Maximum likelihood, implemented in MATLAB®, was used to estimate the simplified formulation of the problem given by Eq. (10). The value of parameters $\tau_1$ and $\tau_2$ were estimated through a grid search. For each combination of values ($\tau_1$, $\tau_2$), the remaining model parameters were estimated using maximum likelihood (with $\tau_1$ and $\tau_2$ fixed). The combination ($\tau_1$, $\tau_2$) that led to the highest likelihood value was selected.

4. Estimation results and discussion

Table 3 summarizes the estimation results. In model I the thresholds $\tau_1$ and $\tau_2$ are set equal to zero whereas in model II the optimal thresholds are estimated with the corresponding specifications. Different vehicle and bicycle threshold combinations were investigated ranging from 0 up to 6 s in increments of 0.5 s. The maximum log-likelihood was achieved for $\tau_1 = 4$ and $\tau_2 = 0$ s.

The results suggest that Model II is better based on the various metrics reported in the table e.g., the likelihood ratio index (corrected$^2$) and the Akaike Information Criterion (AIC).

4.1. Conflict probability

The estimation results are consistent with a-priori expectations. The sign of the parameters are correct, indicating that the probability of perceiving a conflict increases as the arrival time difference between the car and the bike decreases. Fig. 7 shows the probability of perceiving a conflict estimated by Models I and II as a function of the arrival time difference. If a vehicle arrives at the interaction zone first (negative values), the probability that a conflict is perceived is negligible for differences larger than 2 s. However, if a cyclist arrives at the interaction zone less than 2 s after the driver arrives there, the probability of perceiving a conflict increases very rapidly. For instance, a vehicle travelling at 20 km/h in the roundabout can completely traverse the car IZ within 2 s with no cyclist in sight. The probability of perceiving a conflict is around 0.50 if a cyclist arrives at the bike IZ after 1.0 s.

According to the results from Model I, when a bicycle arrives first at the interaction zone, the probability that a conflict takes place is non-zero even if the arrival time difference is up to 16 s. A slow cyclist travelling at the speed of 10 km/h takes about 12 s to reach the conflict zone. Thus, a driver arriving later can still catch up with the cyclist in a collision course. The threshold-based model (Model II) estimates lower probabilities of conflict decisions for ATD values greater than the threshold of 4 s. Model II estimates that a driver perceives a non-zero conflict probability for differences in arrival times up to 12–14 s. At a difference of 6.0 s, the conflict probability is about 0.50. The probability of conflict is very high, almost 1, if the vehicle arrives within 4 s after the bicycle.

4.2. Yielding probability

The estimation results of the yielding probability (Table 3) indicate that the speed of the car and the location of the bicycle are the most important explanatory variables. The signs of the coefficients are in line with expectations and previous literature.

As expected, the speed of the vehicle impacts the yielding probability negatively. Higher vehicle speeds result in lower yielding probabilities. At higher speeds drivers have less time to detect cyclists and react accordingly and they also feel that they will be able to clear the intersection before the arrival of the bicycle. This is consistent with findings from other studies reported in the literature (Rasanen and Summala, 2000; Silvano et al., 2014, 2015).

Fig. 8 shows the probability estimates of Models I and II as a function of the vehicle speed with cyclist located beyond $R_2$. The results suggest that there is a probability that a yielding event takes place if the speed of the vehicle is under 20 km/h.

The proximity of the cyclist, captured by the variables $R_1$ and $R_2$, increases the probability of yielding (positive sign) as expected. Furthermore, the coefficient for variable $R_1$ is higher than the coefficient of $R_2$, indicating the higher urgency to yield when the cyclist is closer (also indicating a nonlinearity in the impact of the location of the bicycle on the yielding probability). The presence of a cyclist in Region 1 ($R_1 = 1$) has a strong influence on drivers increasing the yielding probability compared to the situation where there is no interacting cyclist there. Region 2 ($R_2$) shows an increase in the yielding probability as well. However, the impact is lower compared to the previous region since the cyclist is farther away (11–20 m from the CZ). The proximity of the cyclist to the CZ impacts the drivers’ yielding probability as shown by the above results. It
is important to include Region 2 in the final model specification as it shows how the yielding probability decreases if a cyclist is detected further away from the CZ. It is also important as it shows that the model captures drivers’ response variations as cyclists are located further away from the CZ (with impact diminishing with distance). Region 3 ($R_3$) was not significant in any model and did not improve the model performance overall when included in the specification. Therefore, Region 3 was not included in the final model specification.

Fig. 9 shows the yielding probability, according to Model II, as a function of vehicle speed for different locations of the bicycle, namely, $R_1$ and $R_2$. The black line is the case where the bicycle is beyond those regions ($R_1 = R_2 = 0$). According to the results, a driver with speed of 20 km/h has a 60% probability to yield to a cyclist in Region 1 (red line). However, if the bicycle is located in Region 2, the corresponding yielding probability (for the same vehicle speed) is only 25% (blue line) and less than 5% when located further upstream (black line).

From a behavioral point of view, the results, in general, agree with a priori expectations such as the impact of the vehicle speed and the relative position of the bicycle. The presence of a cyclist close to the CZ increases the yielding probability. Vehicle speed also influences heavily yielding decisions. Bicycle speed ($V_{bike}$), was not significant although the a-priori expectation was that it would impact the yielding probability. The data description presented in Section 3.1 showed that bicycle speeds did not have much variability. Therefore, the bicycle speed did not explain drivers’ yielding decisions in this application. The impact of the bicycle speed needs further investigation with data from sites where its variability is greater. It is expected that depending on the cyclist proximity, bicycle speed plays a role e.g. a slow cyclist meters away from the CZ may encourage drivers not to yield. The results may also indicate that drivers have difficulty differentiating between relative small bicycle speed differences.

4.3. Validation and prediction

Due to the limited sample size it is difficult to validate the results by using part of the observations for calibration and the remaining for validation. In such cases, the leave-one-out cross-validation method (Arlot and Celisse, 2010) is often used to evaluate model performance. For a sample of size n, according to the method, the

\begin{table}[h]
\centering
\begin{tabular}{|l|c|c|c|}
\hline
Variables & Parameters & Model I $t_1 = t_2 = 0$ & Model II $t_1 = 4; t_2 = 0$ \\
\hline
Presence of Conflict Decision model & & & \\
Constant & $\alpha_0$ & 3.758 & (2.99)$^b$ & 3.077 & (2.91)$^b$ \\
$Y_{car}^{(2)}$ & $\alpha_1$ & -0.484 & (-1.39) & -0.526 & (-2.21)$^b$ \\
$Y_{car}^{(3)}$ & $\alpha_2$ & 3.677 & (2.11)$^b$ & 3.064 & (2.12)$^b$ \\
\hline
Yielding Decision model & & & \\
Constant & $\beta_0$ & 19.733 & (1.77)$^a$ & 18.505 & (1.61)$^a$ \\
$V_{car}$ & $\beta_1$ & -1.196 & (-1.89)$^a$ & -1.113 & (-1.68)$^a$ \\
$R_1$ & $\beta_2$ & 5.146 & (1.85)$^a$ & 4.326 & (1.38) \\
$R_2$ & $\beta_3$ & 2.829 & (1.46) & 2.426 & (1.25) \\
$LL(b)$ & & -22.445 & & -20.812 & \\
corrected$R^2$ & & 0.762 & & 0.780 & \\
AIC & & 58.890 & & 55.624 & \\
LL(0) & & -94.445 & & \\
No. of observations & 187 & & & & \\
Yielding events & & & & & 37 \\
\hline
\end{tabular}
\caption{Summary of model estimation results ($t$-statistics in parenthesis).}
\end{table}

\(^a\) Statistically significant at the 5% level.  
\(^b\) Statistically significant at the 10% level.
Fig. 8. Probability as a function of vehicle speed.

Fig. 9. Impact of bicycle proximity on the yielding probability.

Table 4
Parameter results from Leave-one-out cross-validation.

<table>
<thead>
<tr>
<th>Conflict Probability</th>
<th>Yielding Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>α₀</td>
<td>β₀</td>
</tr>
<tr>
<td>α₂</td>
<td>β₁</td>
</tr>
<tr>
<td>α₃</td>
<td>β₂</td>
</tr>
<tr>
<td>β₃</td>
<td>β₃</td>
</tr>
</tbody>
</table>

Model is estimated n-1 times, each time with one of the observations removed from the sample. The estimated model is then used to predict the outcome of the observation that was left out. The approach, when repeated for all observations, provides a good estimate of the performance of the model.

The results of the application of the leave-one-out method are summarized in Table 4, for Model II. The table shows the minimum, maximum and range for each parameter across all 186 leave-one-out estimations, as a measure of model stability. The constant of the yielding probability (β₀) has the widest range, followed by the coefficient for R₁ (Region 1). The ranges in the parameters of the conflict...
decision probability are smaller compared to the parameters in the yielding probability with the largest range for the constant estimate ($\alpha_0$). In general, the results indicate that the model is stable and robust.

Table 5 shows the results of the correctly predicted versus observed events. The model predicts 31 out of 37 observed yielding events correctly and 135 out of 150 non-yielding events, with an overall accuracy of 89%.

Figs. 10 and 11 illustrate the predictive ability of the model, based on the leave-one-out results, using the probability of predicting the actual yielding decisions as a measure of performance. Fig. 10 shows the distribution of the probability of the actual yielding decisions. The x-axis represents the yielding probability, as estimated by the model, while the y-axis is the frequency of observations that were predicted according to the corresponding probability, in percentages (with the actual number on top of each bar). For instance, in 25 cases (out of the 37 actual yielding events) the probability of predicting the actual decision is greater than or equal to 0.90, and in 6 other cases, the model estimates the probability of yielding as 0.70. On the other hand, the model estimates low yielding probabilities for 6 other actual yielding events. In total, the model assigns a probability of at least 0.70 in 31 out of 37 yielding decisions. This illustrates a robust performance.

Fig. 11 illustrates similar results of the distribution of the probability of non-yielding for 150 observations with non-yielding outcome. The results indicate robust performance in this case as well, as the model predicts non-yielding decisions with probability above 0.90 for 127 out of the 150 non-yielding cases. In summary, the model is robust and in most cases the predicted probability of the actual choice is very high (even with the results based on the leave-one-out method).

5. Conclusion

The paper proposes a methodology for modeling the driver-cyclist interactions at unsignalized intersections capturing the driver's yielding decision process. The paper introduces a modeling framework with two levels (conflict decision and yielding deci-
The proposed modeling framework has a number of limitations (many due to the available data). It does not consider the cyclists’ decision process in response to the presence of vehicles (which would have been difficult to estimate given the low variability in observed bicycle speeds). An integrated framework capturing both, driver and cyclist decisions is an interesting direction for further research. The model considers vehicle-bicycle interactions in one traffic direction. In practice, cyclists have interactions in both vehicle directions e.g., vehicles exiting and entering the roundabout (or intersection). Further research can also address some of the limitations of the available data. For example, complete trajectory data over a range of facilities can provide a better understanding of the vehicle-bicycle interactions e.g., the impact of intersection design. Another limitation that may be addressed in future work is the assumption of fixed interaction zones and the location where the decision process starts.

Acknowledgements

The authors would like to thank the financial support by the Foundation of the Swedish National Insurance Co. (Länsförsäkring AB). The data collection was mainly funded by J. Gust. Richter Stiftelse (managed by Sweco AB).

References


