ADVANCES IN TRANSPORTATION STUDIES
An International Journal

Editor in Chief: Alessandro Calvi

Section A & B

Volume XLIV April 2018

Contents

Section A

Section B

ADVANCES IN TRANSPORTATION STUDIES

An International Journal

Section A

Advances in Transportation Studies: an international Journal Section A 44 (2018)

ISBN 978-88-255-1434-6– ISSN 1824-5463 - 18001– DOI 10.4399/97888255143461 – pag. 5-20

Asymmetry of pedestrian influence area in unidirectional and bidirectional crowd flows: empirical evidence and statistical modeling

F. Giannis E.I. Vlahogianni

National Technical University of Athens, 5, Iroon Polytechniou str, 15773, Zografou Campus, Greece, tel. +30-210-7721369, fax +30-210-7721424
email: elenivl@central.ntua.gr

subm. 17th March 2017 *approv. after rev. 11st October 2017*

Abstract

Every pedestrian forms an influence area of variable extent at the emergence of an interaction with another pedestrian. The characteristics of this influence area are usually discussed in the framework of pedestrian simulation models. The present work addresses the statistical modeling of this influence area with the aim to reveal the critical microscopic factors that may affect its type and extent using a multivariate structural equation framework. The analysis focuses on overtaking patterns in unidirectional and bi-directional normal crowd flow conditions based on pedestrians' trajectories derived from video recordings at a highly visited metro station of Athens (Greece). Findings reveal asymmetries between the rear and front part of the influence area. Moreover, both the speed difference between interacting pedestrians and the type of flow (unidirectional or bi-directional) significantly affect the observed asymentries.

Keywords – pedestrian flow, influence area, microscopic traffic analysis, structural equation modeling

1. Introduction

Large transportation infrastructures usually accommodate a significant number of visitors on a daily basis. Thus, they should fulfil specific requirements related to safety and comfort. Safety and comfort require the minimization of interactions between pedestrians. In typical crowd flow conditions, when a pedestrian reaches the influence area of a lead pedestrian - the area around a pedestrian where interacting forces with the rest of pedestrians emerge-, several repulsive forces drive the pedestrian to overtake or decelerate to avoid the lead pedestrian [1].

The literature on the complexity of crowd flows is extensive. Several models have been considered, such as regression models [2], queuing models [3], route choice models [4], discrete choice models [5, 6], heuristics-based modeling [7], game theoretic models [8], macroscopic models [9, 10], and microscopic simulation approaches [11-18]. The most popular models are the social force models [1, 19], and the agent-based models [20-22].

The spatial patterns that emerge during overtaking conditions have been rarely researched and included in the modelling [14], [19], [22, 23]. Pedestrians navigate through traffic with speeds of ranging magnitude and variable headways. Those who choose higher speeds, at some point, will have to overtake those with lower speeds in order to maintain their desired speed or decelerate to avoid collision. An overtaking manoeuvre may occur when a specific pedestrian deviates from his original path, which is determined by the minimum distance between his origin and the desired

destination. In contrast to roadway traffic, where overtaking is determined by the road geometry and the opposing traffic, if any [24], overtaking in crowd flows can emerge in numerous ways depending on the desired destination and the path of "least effort", in time or space [1]. Moreover, the reaction times are much lower than in the case of vehicles.

The pedestrian overtaking behavior is rarely statistically modeled. Thompson et al. [23] analysed the possibility to overtake by defining the direction angles from a lead pedestrian and the route choice based on the minimum deviation from the original, optimal direction. Yuen and Lee [25] proposed a modified social force model to represent the overtaking behavior of pedestrians in unidirectional normal crowd flow conditions; modifications refer to equations that describe the direction and magnitude of overtaking. Ji et al. [26] extended Yuen and Lee [25] work by introducing the pedestrians' anisotropy, a term to describe the effect of the pedestrian behind the subject pedestrian. Hussein and Sayed [22] introduced an agent based model with complex behavioural rules to describe the pedestrian interactions in uni-directional (overtaking) and bi-directional (collision avoidance) flow conditions.

Developing statistical models to explain crowd flow complexity emerging from pedestrians' interactions based on real world collected data can increase the understanding of the behavioral dynamics that may govern crowd flows under different flow conditions, walking environments and transportation facilities. This understanding may be used to improve the parametrization of existing simulation models, assess the level of service of existing transportation facilities and improve their design and safety. To this end, the present work focuses on the statistical modeling of the influence area of a pedestrian in unidirectional and bi-directional normal crowd flow conditions at the emergence of an interaction (overtaking) with an adjacent pedestrian. Structural equation models are developed to reveal the critical factors that may influence the asymmetrical characteristics of a pedestrian's influence area. Data are derived using computer vision and reflect flow conditions near the entrance of a highly visited metro station in the centre of Athens (Greece).

2. Pedestrian influence area

An important feature of crowd flow modeling is the influence area formed around a pedestrian. In literature, this area is described as being square [26], [27], cyclic or elliptical [1], [28]. The last is the prevailing form and is used in the popular social force model. The influence area dictates the manner a pedestrian maneuvers through crowd flows and is the reason why the overtaking maneuver may occur. When a pedestrian reaches the influence area of a lead pedestrian, the interaction between them - usually described as the repulsive force - will force the pedestrian to overtake. The overtaking maneuver will end when the pedestrian departs from the influence area of the lead pedestrian.

Most existing research papers treat the influence area as being symmetrical. Although this approach may reduce the complexity of the modeling, it provides an unrealistic behavior of pedestrians' interactions. Recent evidence has demonstrated the importance of both the lateral and longitudinal spatial distance between interacting pedestrians that are correlated to walking speeds [29]. Campanella et al. [30] suggest that the influence area in front of a pedestrian is different from the one behind him; empirical findings of the specific study reveal that the pedestrians react stronger to events in front of them, as well as those that occur closer to their axis of movement. Daamen et al. [31] used empirical data from different flow configurations, population characteristics and traffic conditions from various experiments to calibrate pedestrian simulation models.

2.1. Modeling approach

Every pedestrian produces an influence area of variable extent around him. This influence area impacts the surrounding pedestrians and affects their way of moving, for example the way an upstream pedestrian will overtake the subject pedestrian to maintain his desired speed. Let this influence area be asymmetrical. Moreover, let the "true" spatial extent of this influence area be unknown or difficult to measure (latent), yet, identified by the major semi-axes of two ellipses - rear and front - formed by the overtaking trajectories of other pedestrians. Figure 1 graphically depicts the above conditions; the dark arrow represents the direction of movement for the subject pedestrian and the dashed line represents the trajectories of the pedestrian who overtakes the subject pedestrian. This asymmetrical influence area is influenced by various traffic and geometry related characteristics.

In the present paper, the above-described problem is modeled using a Multiple Indicators-Multiple Causes (MIMIC) latent variable model a special class of Structural Equation Models (SEM). SEM has been previously applied to transit system quality of service analysis [32, 33], secondary incident traffic influence assessment [34], travel behavior modeling [35], driver's behavior modeling [36] and public acceptability analysis of new technologies for traffic management [37]. SEM models can be considered as a generalized case of multivariate classical statistical models and suffer from similar constraints as classical statistical models, but outperform similar techniques due to their ability to treat auto-correlated errors, non-normal data and latent variables [38].

SEMs consist of two components: a measurement model, which defines the relations between a latent variable and its indicators and a structural model, which specifies the casual relationships among latent variables and explains the casual effects [38]. The MIMIC model considers the latent variable η to be scalar and relates the vector of indicators y and the observed exogenous variables x that cause η by the following system of equations:

$$\begin{aligned} \eta &= \Gamma x + \varepsilon \\ y &= \Lambda \eta + \zeta \end{aligned}, \tag{1}$$

where Γ and Λ are matrices of unknown parameters to be estimated and ε and ζ are the error terms.

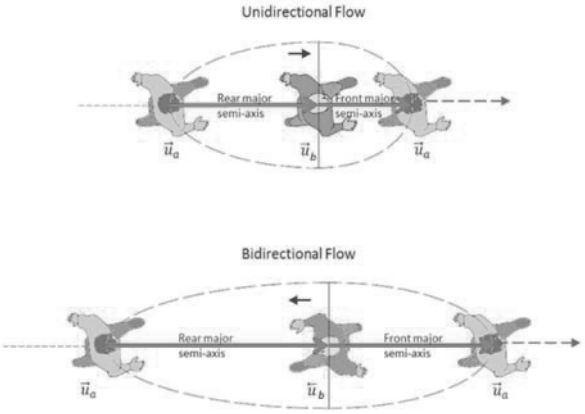

Fig. 1 – Visualization of principle of combined SPF, where intersections are considered only in terms of their frequency (1); segment-only SPF (2a); and intersection SPF (2b)

Various approaches have been used to estimate SEM models; the most popular is the Maximum Likelihood method. Golob [35] provides a thorough literature review on the estimation methods with emphasis on their relation to the sample size.

A critical step in SEM modeling is the assessment of the model's fit. Due to the inherent complexity of such models, a tedious evaluation should be followed in order to assure the model's fit. This includes measures to select the best modeling structure, such as the Akaike's Information Criterion (AIC) or Bayesian information criterion (BIC), as well as likelihood ratio tests for comparing the proposed model to the saturated one (the model that fits the covariances perfectly) and baseline models (model that includes the means and variances of all observed variables plus the covariances of all observed exogenous variables). Other means for model evaluation include the Root Mean Squared Error of Approximation (RMSEA) along with the probability of RMSEA being below 0.05, the Standardized Root Mean Squared Residual (SRMR) and the Coefficient of Determination (CD) of the various models. A model fits the real data adequately, if the lower bound of the 90% CI of RMSEA is below 0.05 and is poor if the upper bound is above 0.10 [39]. A good fit provides SRMR less than 0.08 [40]. CD is similar to R^2 for the entire model.

Finally, a supplementary issue in statistical modeling – especially when complex modeling structures are involved - is overfitting. Overfitting occurs when models have proportionally larger number of parameters in relation to the sample size. The outcome is having models that may replicate in detail the characteristics of the data they are asked to model, but cannot efficiently generalize on out-of-sample data of the same phenomenon [41]. Various rule-of-thumb approaches exist in literature as remedies against developing erroneous models [42]. The simplest of all are: i. to randomly split the sample and use the largest part for model building and evaluation and the other for testing (holdout method), and ii. to estimate n times the model leaving each time a data point for validation (leave-one-out cross validation approach). These approaches are very popular in machine learning, but rarely used in statistical modeling [41].

2.2. Video-based pedestrian flow experiment

Video-based data collection approaches are very popular in pedestrian traffic flow analysis and modeling [43-45]. This is mainly because video recordings – regardless of some major limitations that may be alleviated through proper camera settings and video post processing - are a safe and non-disruptive approach to study pedestrians' behavior and crowd flow dynamics [45]. The study area is approximately 150 square meters and is located outside a highly visited Metro station (Figure 2). In the study area, high pedestrian flows - both unidirectional and bi-directional – are reported and a limited number of stationary obstacles exist.

Fig. 2 – Study area. In light grey the area where pedestrian trajectories are recorded

For the experiment, video recording took place at a pedestrian bridge adjacent to the study area. The camera was placed approximately 8 meters higher than the study area. The collected data include video streams from the morning peak period (8:00 a.m. – 12:00 a.m.) for one week. From the video streams, pedestrian trajectories were extracted by a semi-automated process using a trajectory extraction system [46]. Specifically, each pedestrian's location on screen was be converted into real-world coordinates by using the two equations that follow:

$$x_{real} = \frac{a_1 x_{screen} + a_2 y_{screen} + a_3}{a_4 x_{screen} + a_5 y_{screen} + 1}$$
$$y_{real} = \frac{a_6 x_{screen} + a_7 y_{screen} + a_8}{a_4 x_{screen} + a_5 y_{screen} + 1} \quad (2)$$

where (x_{real},y_{real}) is the real-world coordinate, (x_{screen},y_{screen}) is the video image coordinate and α_1 to α_8 are coefficients that were estimated based on the coordinates of four reference points in the real world. In the video detection process, the frame rate was set to 25 frames per second, which is equal to 0.04 seconds time interval. Since the video frame rate were predefined, the real position of the pedestrian per frame is known, which allows the researcher to extract all necessary kinematic characteristics. For example, speed is calculated as the derivative of the position with respect to time, while speed difference is obtained by comparing speed changes per time intervals. The extracted trajectories depicted 5 origin/destination points that - in turn - form 6 routes of pedestrian flow (either unidirectional or bidirectional). Some examples of extracted trajectories of the subject pedestrian are seen in Figure 3.

Overall 804 cases of pedestrians' interactions that included an overtaking are considered in the further statistical analysis. Table 1 summarizes the variables that are considered in modeling. The variables *spatial_rear* and *spatial_front* describe the asymmetric influence area around a pedestrian who is been overtaken. The first is calculated at the beginning of the passing of the subject pedestrian (pedestrian b), while the second is calculated at the end of the overtaking attempt (Figure 1). The variable *s_a-s_b* denotes the speed difference between two interacting pedestrians and takes on positive values. This difference is calculated at the beginning of an overtaking attempt.

3. Empirical microscopic features of pedestrian interactions

The study of the microscopic features of pedestrian flows is important for the development and calibration of efficient car following models [28].

Fig. 3 – Examples of pedestrian trajectories that are recorded

Tab. 1 – Description of variables

Variable	Description
Continuous	
spatial_rear	The major semi-axis of the elliptical influence area at the back (unidirectional flow) or the front (bi-directional flow) of pedestrian been overtaken
spatial_front	The major semi-axis of the elliptical influence area at the front (unidirectional flow) or the back (bi-directional flow) of pedestrian been overtaken
speed_a	Speed of subject pedestrian
speed_b	Speed of pedestrian been overtaken
(s_a)-(s_b)	Speed difference between pedestrian a and pedestrian b at the beginning of an overtaking attempt
Categorical	
comp	Subject pedestrian grouping (0 for single pedestrian, 1: otherwise)
dir	The direction of travel of the interacting pedestrian (0: bi-directional flow, 1: otherwise)
stationary	Pedestrian been overtaken (1: the pedestrian b has zero speed, 0: otherwise)
obs_right	The existence of obstacle at the right of the subject pedestrian ((1: obstacle in less than 6 m from the subject pedestrian, 0: otherwise)
obs_left	The existence of obstacle at the left of the subject pedestrian (1: obstacle in less than 6 m from the subject pedestrian, 0: otherwise)
gender	The gender (1: male, 0: female)

This section focuses on speed and spacing at the emergence of an interaction between two pedestrians (unidirectional flow) or an opposing pedestrian (bi-directional flow). The spacing is the calculated rear major semi-axis for pedestrian *a* and front major semi-axis for pedestrian *b*.

Figure 4 graphically depicts the continuous variables through their quartiles (1st quartile, median and 3rd quartile). In x- axis the variables; names are depicted, whereas in y-axis is the value of each variable (in meters for spatial_front and spatial_rear variables and m/sec for speeds). The single points in speed_a variable represent outliers falling outside the range of ±1.5IQR, where IQR is the interquartile range. As can be observed, on average, *spatial_rear* takes on higher values compared to those of *spatial_front*. This means that pedestrians seem to change their course earlier when deciding to overtake, in comparison to when they return to their original course at the end of the overtaking maneuver. This is a first indication of the asymmetry in the influence area of a pedestrian at the emergence of an overtaking maneuver.

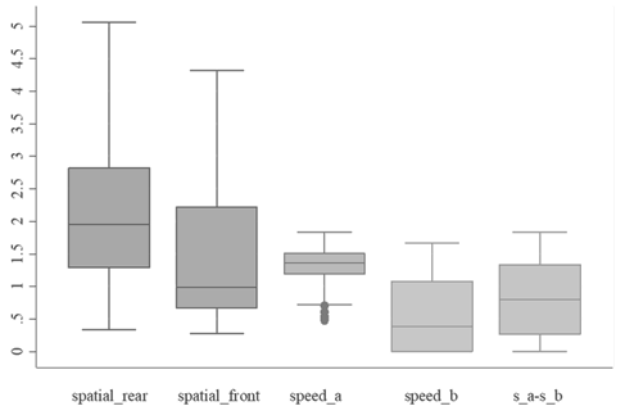

Fig. 4 – Box plot graph for the continuous variables

Moreover, the longitudinal spacing for walking at the beginning of an overtake is smaller than the typical longitudinal spacing for walking indicated in Fruin [47]. This may be the reflection of cultural differences. Regarding speeds, the average speed of pedestrian a is expected to be much higher than the one of pedestrian b, as the experiment depicts interactions during overtaking of the pedestrian b by the pedestrian a.

The first question, which emerges, is whether speed and spacing differ with respect to the type of flow (unidirectional or bi-directional). Figure 5 and Figure 6 demonstrate the variability in the distributional characteristics of the rear and major semi-axis with respect to the type of flow and whether pedestrian b is stationary or not.

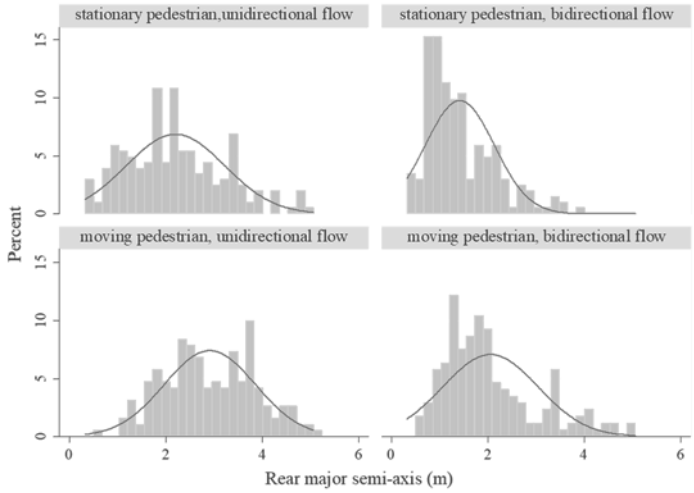

Fig. 5 – Distribution of rear major semi-axis for the type of flow and the subject pedestrian motion (pedestrian b) (the dark line depicts the normal distribution)

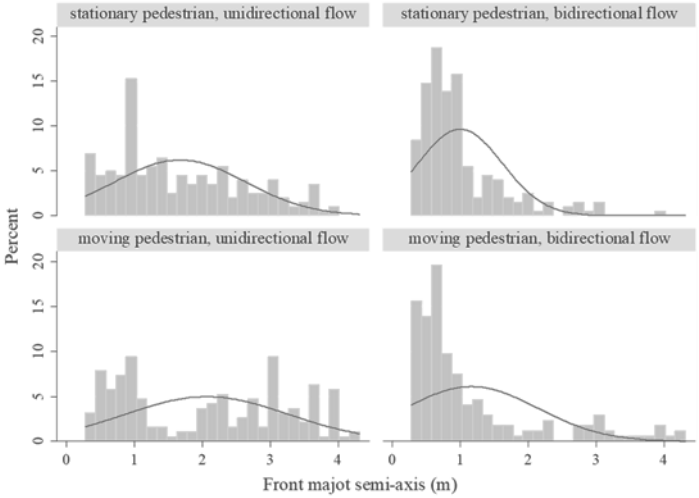

Fig. 6 – Distribution of front major semi-axis for the type of flow and the subject pedestrian motion (pedestrian b) (the dark line depicts the normal distribution)

A non-parametric two sample Mann–Whitney U test is conducted to test the hypothesis that speed and spacing come from the same distribution for unidirectional and bi-directional crowd flows. Findings show that the above hypothesis can be rejected for the cases of rear and front major semi-axes (test values z = 10.795 (p>|z| = 0.000) and z = 11.425 (p>|z| = 0.000) respectively). For speeds, the hypothesis can be rejected for the pedestrian a (z = 3.940 (p>|z| = 0.000)), whereas the hypothesis of similar distributional characteristics for unidirectional and bi-directional flow can be accepted for the speeds of pedestrian b (z = 1.145 (p > |z| = 0.252)). This implies that the subject pedestrian been overtaken (pedestrian b) has similar behavior in terms of speed in both flows. This may relate to a tactical behavior during overtaking conditions between an active and passive partner; the active pedestrian will choose a speed at the emergence of an overtake which may vary, whereas variations of the average speed of the passive partner (the pedestrian b) will probably be less critical. Figure 7 and Figure 8 depict the fundamental microscopic diagram for pedestrian a and pedestrian b.

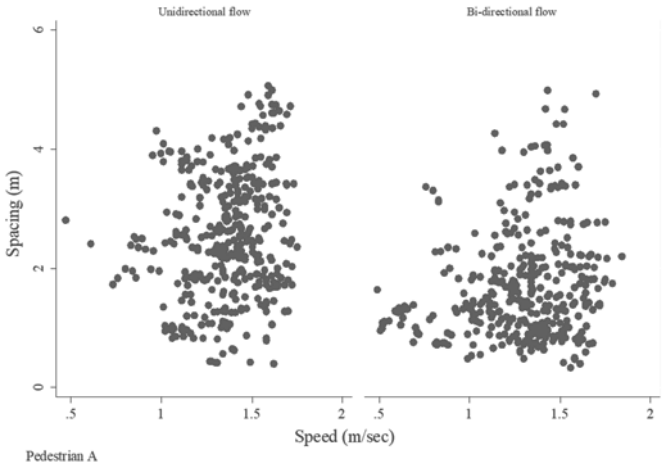

Fig. 7 – Speed-spacing relationship for pedestrian a

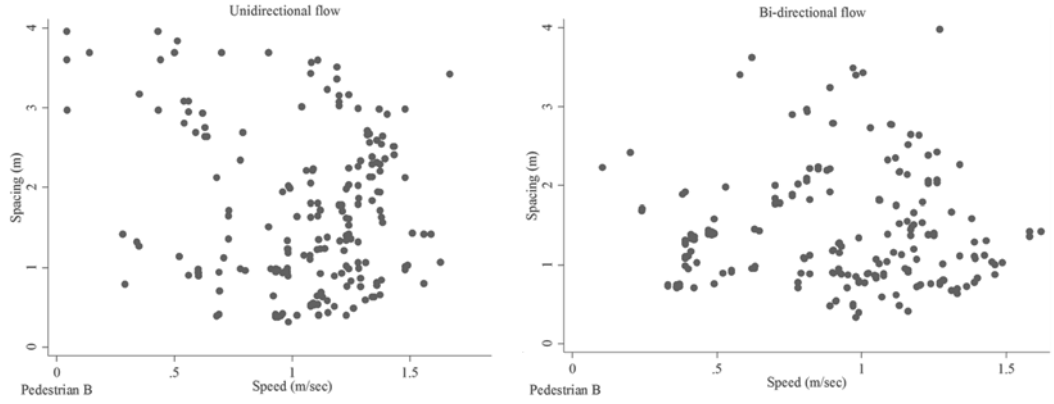

Fig. 8 – Speed-spacing relationship for the subject pedestrian (pedestrian b)

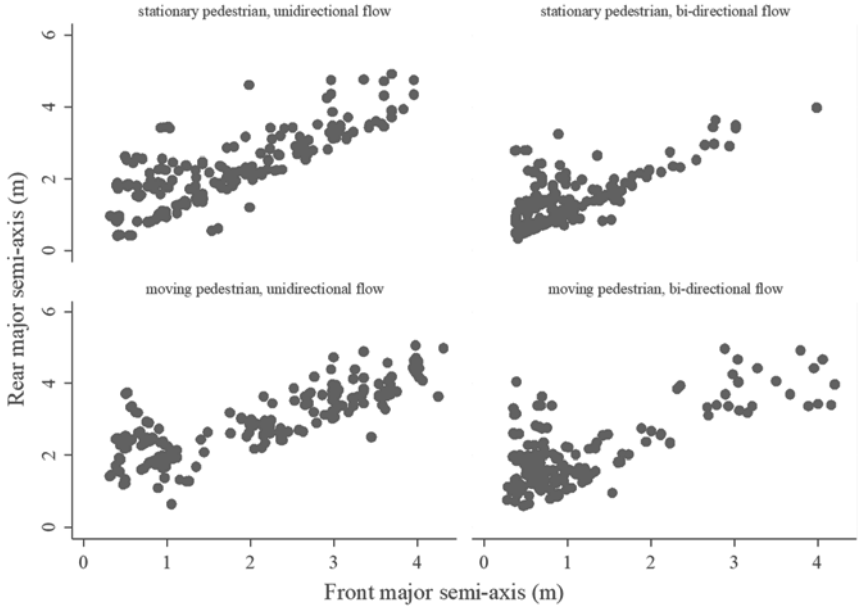

Fig. 9 – Scatter plot of rear and front major semi-axis with respect to the type of flow and the subject pedestrian motion (pedestrian b)

Although a linear relationship between spacing and pedestrian speed has been previously reported [48], in the specific paper, for both types of flow, no clear relationship can be observed. A supplementary pairwise correlation analysis also shows a low linear correlation between speed and spacing (correlation coefficient below 0.4).

Figure 9 shows the relationship between rear and front major semi-axis with respect to the type of flow and whether the subject pedestrian (b) is stationary or not. Although the correlation analysis shows that there exists strong linear relationship between the magnitude of rear and front major semi-axes (correlation coefficient above 0.6), this linearity is not clear, especially in low values of front major semi-axis. For these values, an increased scattering in the relevant rear major semi-axis values is observed. The observed stochasticity in low values of both variables supports the hypothesis of the asymmetrical influence area around a pedestrian when interacting with others in overtaking conditions. Further research is needed to mathematically describe the observed relationship.

4. Multivariate modeling of the asymmetrical influence area

A MIMIC model with a latent variable representing the pedestrian's influence area is constructed. It is assumed that the influence area is not symmetric and unobservable (latent), yet, described based on two indicators; the *spatial_rear* and the *spatial_front* namely the major semi-axis of two elliptical influence areas defined by the spacing of two pedestrians at the emergence and after the completion of the overtaking respectively (Figure 1). The rest of the variables depicted in Table 1 are considered as predictors of the asymmetrical influence area (latent variable).