<http://www.diva-portal.org>

Postprint

This is the accepted version of a paper published in *Transportation Research Part F: Traffic Psychology and Behaviour*. This paper has been peer-reviewed but does not include the final publisher proof-corrections or journal pagination.

Citation for the original published paper (version of record):

Ma, X., Luo, D. (2016)

Modeling cyclist acceleration process for bicycle traffic simulation using naturalistic data.

Transportation Research Part F: Traffic Psychology and Behaviour, 40: 130-144

Access to the published version may require subscription.

N.B. When citing this work, cite the original published paper.

Permanent link to this version:

<http://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-189206>

Modeling cyclist acceleration process for bicycle traffic simulation using naturalistic data

Xiaoliang Ma^{a,*}, Ding Luo^a

^a*System Simulation & Control (S2CLab), Department of Transport Science, KTH Royal Institute of Technology, Teknikringen 10, SE-100 44 Stockholm, Sweden.*

Abstract

Cycling is a healthy and sustainable form of transportation. The recent increase of daily cyclists in Sweden has triggered broad interest in finding how policies and measures may facilitate the planning of bicycle traffic in the urban area. However, in comparison to car traffic, bicycle traffic is still far from well understood. This study is part of the research effort to investigate microscopic cyclist behavior, model bicycle traffic and finally build a simulation tool for applications in transport planning. In particular, the paper focuses on representing bicycle movements when the cyclist doesn't interact with others. The cyclist acceleration behavior is modeled using naturalistic GPS data collected by eleven recruited commuter cyclists. After filtering the large amount of data, cyclist trajectories are obtained and acceleration profiles are abstracted. A mathematical model is proposed based on the dataset, and three model forms are estimated using the maximum likelihood method with Laplace and Normal error terms. While the model with more parameters shows superior performance, the simplified ones are still capable of capturing the trends in the acceleration profiles. On the other hand, the study also introduces social economic characteristics of cyclists to explain the model parameters and they show significant effects. However, the cyclist population being investigated in the study is still limited, and more convincing results can be obtained when the data collection effort is extended to larger population with more variable cyclist characteristics.

Keywords:

Bicycle traffic simulation, cyclist behavior, acceleration model, naturalistic data, GPS.

1. Introduction

Many cities in Europe and US have witnessed the growing cyclist population over the past decades. Although the size of this community is still small in comparison to motorized vehicles, increasing attention has been given to its high level of vulnerability in traffic safety research. In the meanwhile, a tremendous demand for more knowledge of the characteristics specific to this traveler group has also been identified provided the new policy trend for sustainable transport development. As the popularity of cycling increases at a fast pace in most urban areas, both traffic planners and policy makers have also been seeking useful analytical tools which can effectively facilitate addressing bicycle-related planning and operational issues. Since the development of these tools, such as bicycle traffic simulation models, highly depends on sufficient understanding of cyclist behavior, it becomes even more urgent for researchers to initiate related studies.

In comparison to the overwhelming research effort on car traffic, the study on bicycle traffic is still far behind. However, both car and bicycle traffic flows show complex patterns due to their frequent interactions within the traffic system. In particular, the behavior of cyclist and driver plays a central role. Therefore, in order to improve the planning practice of bicycle traffic, it is important to understand how cycle traffic is represented by individuals, how cyclists manipulate their speed levels and interact with each other, and moreover how they interact with cars.

*Corresponding author. Tel: +468 7908426. Email: liang@kth.se.

1.1. Relevant studies

As an effective and efficient tool, traffic microscopic simulation models are extensively used to appraise proposed transport policies and measures before they are implemented. Nevertheless, the accuracy and reliability of these appraisals are greatly dependent on the understanding of the road users' behavior, as well as modeling of their movements and interactions among one another. While the significance of bicycle as a convenient and environmentally friendly non-motorized transport modes has been highlighted by many researchers, there is still a considerable gap in the current literature in comparison to more fruitful results of driver behavior and vehicle traffic flow. [Twaddle et al. \(2014\)](#) suggested several facts making bicycle traffic quite different from vehicle traffic, i.e. the dynamic characteristics of cyclists, including their speed, acceleration and deceleration profiles, and the physical characteristics, including size, flexibility and capability.

Experiences from research on motorized vehicles indicate that the availability of sufficient field data, or naturalistic data is extremely important to the understanding and modeling of cyclist behavior. Given the demand of high quality data, the Instrumented Probe Bicycle (IPB) became an attractive and reliable tool. Although currently there is no standardization of IPB, this technology is developing at a fast pace and many researchers have attempted to carry out bicycle-related studies using data collected by various IPBs. Early on the IPB was quite basic with a few types of equipment mounted and most of those studies merely aimed at investigating the cyclist behavior on a relatively macroscopic level, such as identifying the risk factors existed in the bicycle activities (e.g. [Walker, 2007](#); [Johnson et al., 2010](#)) and obtaining cycling routes and speed distributions (e.g. [Parkin and Rotheram, 2010](#); [Addison and Low, 1998](#)). Most recently, however, more developed IPBs emerged and deeper insights into bicycle mobility have also been presented. For instance, [Joo and Oh \(2013\)](#) employed an IPB equipped with a set of sensors including a GPS receiver, accelerometer, and gyro sensor. According to their report, useful data for identifying longitudinal, lateral, and vertical maneuvers in bicycle movement could be obtained from the IPB. Owing to the abundant data, an intelligent system called Bicycle Monitoring Index (BMI) which can be used to evaluate the safety and mobility of the bicycle environment was developed using the fault tree analysis (FTA) technique. Later, [Dozza and Fernandez \(2014\)](#) applied an advanced IPB with multiple sensors to study bicycle dynamics and cyclist behavior. Longitudinal, lateral, and vertical accelerations were measured and further processed by a specialized software. The main conclusion was that high quality naturalistic data from the IPB laid a solid foundation for further development of bicycle-related applications.

In comparison to the overwhelming research effort on driver behavior and vehicle traffic simulation, the modeling of cyclist behavior and bicycle traffic is still a relatively undeveloped area. Early attempts were made (e.g. [Ferrara and Lam, 1979](#)) to study cyclists' crossing behavior at intersections using computer simulation. In Europe where cycling has always been more prevailing than in the U.S, research on bicycle traffic and operation turns out to be more thriving. For example, researchers in the Netherlands developed a simulation tool to examine the level of service of separate bicycle facilities with mopeds included ([Botma and Papendrecht, 1991, 1993](#); [Botma, 1995](#)). In Finland, [Kosonen \(1999\)](#) developed a simulation program mainly for vehicle traffic but bicycle mode was included based on the same car-following logic. [Taylor and Davis \(1999\)](#) reviewed many studies on bicycle transport and pointed out that limited computer simulation program was available by that time. [Faghri and Egyházióvá \(1999\)](#) then identified the requirement of mixed simulation models involving motor vehicles and bicycle traffic. They thereof proposed a microscopic simulation model BICSIM. A simple biking following logic based on the total safety distance model was developed and implemented in the simulation tool.

Most recently, several microscopic traffic simulation tools, widely used in road traffic modeling and analysis, have been extended with the bicycle mode to meet the requirements for scientific and commercial applications. VISSIM, AIMSUN and SUMO are typical examples of such tools. [Twaddle et al. \(2014\)](#) did a comprehensive review of the latest work on bicycle and mixed traffic modeling. In particular, they summarized the modeling approaches for bicycle movement into four categories: longitudinally continuous models, cellular automata models, social force models and logic models. The majority of current microscopic traffic simulation tools model the longitudinal and lateral motion of road users separately (e.g. car-following and lane changing models are often applied for vehicle traffic). Therefore, the same approach can be naturally extended for modeling cyclist behavior. Moreover, [Oketch \(2000\)](#) proposed an approach for modeling mixed-traffic streams which allows the inclusion of different kinds of transportation modes, such as motorcycles, bicycles, three-wheeled vehicles and so on.

Cellular automata (CA) are discrete models not only on time but also on space ([Nagel and Schreckenberg, 1992](#)). The approach has been extensively applied to represent vehicle traffic from microscopic to more aggregate models

(Gould and Karner, 2009). In order to describe bicycle movement, cyclists can also be assumed to move across the discrete cells of bicycle path based on the rules established. The approach provides an effective way to investigate the interactions between road users. For instance, Zhao et al. (2013) attempted to model passing events in mixed bicycle traffic with CA.

The social force model has been originally used in pedestrian dynamics. The concept has been further extended for modeling bicycle's move in reaction to a number of attractive and repulsive forces. Movement within social force models is not bounded to the longitudinal and lateral directions. So the approach is usually computationally more expensive. Meanwhile, attempt has been made to represent the behavior of and interaction among cars, pedestrians and bicycles by a social force model (Schönauer et al., 2012). Logic models tend to include the detailed tactical behavior of cyclists. Comparatively, the models often have more complex structure to represent advanced strategies of human behavior such as cyclist path planning. However, the computational requirements mitigate the applicability of the approach in modeling large-scale traffic network.

1.2. Research objectives

As the overview of the literature shows, few studies tried to build high fidelity models for bicycle movement which could promote microscopic simulators on the basis of the understanding of cyclist behavior. While some bicycle models have been implemented in existing traffic simulation, few of them were carefully calibrated using real field measurement due to the lack of high quality cycling data. Given the increasing requirement to develop bicycle traffic models and availability of naturalistic data, the goal of this paper is therefore twofold. First, the study looks for analytical approaches to understand cyclist behavior under common circumstance based on high frequency GPS data, especially for the acceleration (and deceleration) behavior. Many detailed cycling characteristics are expected from data analysis and modeling work. Second, the paper targets at developing bicycle movement model which can be implemented in a dedicated in-lab bicycle traffic simulator. Moreover, the acceleration models are expected to be calibrated and validated using the processed naturalistic datasets.

This paper is therefore structured in the following manner. In Section 2, the general modeling framework of the simulator is introduced; Section 3 presents a data-driven approach for modeling cycling behavior; Section 4 offers a summary of the entire work.

2. Simulation modeling framework

In traffic microscopic simulation, agent-based approach provides a convenient framework to represent detailed human behavior. In order to build an agent-based bicycle traffic simulator, the modeling structure has to be planned in advance. The simulator adopts discrete time mechanism in which time step is used to advance simulation time. While the paper focuses on the development of bicycle movement model using naturalistic data collected by GPS devices, this section overviews the general modeling framework and essential components.

Similar to other transport simulation, bicycle network can be represented by nodes and links. Link represents bicycle path on which cyclist runs, and it has properties like direction, width and road grade. Node connects links, and cyclists at different links (and directions) may meet each other on nodes. Traffic on links can be represented either aggregately or in details. In our case, individual cyclist and their decision and movement are modeled at the microscopical level. Similar to driver behavior in car traffic, cyclist behavior can also be described at different levels. At the strategic level, cyclist, when generated, has to choose route according to distance, his habit and other accessible information. In the simulator, route is selected from the two shortest paths before departure. The cyclist dynamics is mainly affected by the decision on movement and speed at the tactical level. Depending on road and cyclist physical conditions, desired speed will be assigned to cyclists for different road links. Change of the desired speed under various conditions will lead to acceleration or deceleration process. This crucial behavior is indeed what we want to model for free-flow bicycle movement in this paper.

On the other hand, when bicycles meet each other at links, they need to either carry out overtaking or control speed and maintain a safe distance from the leading one. In car traffic, overtaking is a maneuver that requires modeling of detailed lane changing behavior. With the consideration on computational efficiency during simulation, simple model is preferred in our development. The detailed lane changing maneuver is therefore neglected in the current simulator implementation. On the other hand, cyclist-following model, similar to driver-following model in car traffic, is crucial

Figure 1: An illustration of the general modeling framework for cyclist behavior (upper) and visualization of simulated bicycles in the modeled network (lower).

for representing bicycle speed and movement when they interact with each other on road links. It contributes for modeling congestion and capacity at the link level. Nevertheless, modeling of cyclist-following behavior is beyond the scope of this paper.

Figure 1 summarizes our approach on modeling cyclist movement for the bicycle traffic simulator as well as the visualization of simulated bicycles on the Stockholm bicycle network. When a cyclist is generated, he will be assigned with a random number for representing his agility. A person with higher agility will chose higher desired speed at different links, depending on the free-flow speed distribution at each link. When cyclist with higher speed approaches a peer with lower speed at the same link, he can probabilistically choose to keep his speed and overtake or simply decide to follow the leader. In the latter case, his dynamics will be based on the prediction of the cyclist-following model.

In general, as described, the simulator still requires further extension with other aspects of modeling and calibration such as cyclist-following behavior. A prototype software has been developed using Java and it can read network topology from the Open Street map as well as the shape file of the Stockholm bicycle digital network. The model developed in this paper has been implemented in the simulator. The later part of this paper focuses on development of bicycle acceleration models.

3. Development of acceleration models

In this section, a data-driven approach for modeling cyclist behavior is presented. As described, straightforward bicycling regimes are proposed, followed by the presentation of an acceleration model capable of predicting both acceleration and deceleration processes.

3.1. Data Collection and Processing

Current study has been greatly benefited from a large amount of naturalistic data collected in the urban area of Stockholm (see Figure 2). The data collection was organized twice, respectively in the autumn 2013 and spring 2014. Eleven commuter cyclists, three female and eight male, were recruited for the data collection. All participants were provided with handlebar mountable devices including the Garmin Oregon and Garmin Edge 500 GPS devices. All participants were also required to record their normal cycling trips as many as possible using the provided equipments. Devices were regularly retrieved to the lab in order to upload the data on the server. By the end of the data collection, 126 available bicycle trips, including 141, 142 observations, had been acquired and the raw data was then stored and managed in a PostgreSQL database. All adopted devices were able to provide GPS and altitude measurements with a time interval of one second owing to the high-sensitivity integrated GPS receivers and the internal barometric altimeters. Some representative bicycling trajectories are also visualized in Figure 2.

The altitude measurements were smoothed using locally weighted regression (Cleveland and Devlin, 1988) and this processing resulted in useful smoothed gradient profiles (Luo and Ma, 2014). As to the data of GPS coordinates, not only were the original measurements of latitude and longitude with eight digits in the fractional part available to the users, but also the derived information including distance and speed were presented. However, further data analysis revealed a main issue of the raw GPS data: occasional missing observations to a relatively large extent undermined the consistency of measurements which turned out to be critical to the current case given our objective of understanding and modeling cyclist behavior using naturalistic data. Therefore, after trying some estimation approaches from the simple autoregressive-moving-average (ARMA) filters to more advanced filtering techniques, the Kalman filter (KF) based approaches brought the most consistent estimation results.

The KF filter is an advanced method which is able to give optimal estimates of data modeled by the state space relation. The approach was first suggested by Kalman (1960) and has been widely used in signal processing and control systems over the past few decades. In this study, using the method allowed us to take advantages of high-frequency GPS measurements. As a result, not only was the problem of missing observations addressed by the “Predict” procedure of the KF filter, but also the measurement noise was largely removed from the filtering procedure.

The specific implementation of the KF filter referred to previous studies on driver behavior using an instrumented vehicle (Ma and Andréasson, 2005, 2007) in which the extended Kalman smoothing algorithm was applied to estimate the collected car-following data. Similarly, the bicycle tracking problem for one dimension (the direction of the road)

Figure 2: The digitalized bicycle road network in Stockholm(upper) and example trips for data collection (lower).

can be represented by a state space model (it is a linear system in this case) as follows:

$$\begin{aligned}\mathbf{x}(t+1) &= \mathbf{A} \cdot \mathbf{x}(t) + \mathbf{w}(t) \\ \mathbf{y}(t) &= \mathbf{H} \cdot \mathbf{x}(t) + \mathbf{v}(t)\end{aligned}\tag{1}$$

where $\mathbf{x}(t) = [s_n(t) \ v_n(t) \ a_n(t)]^T$ denotes the state vector of the cyclist n at time t including position, speed and acceleration respectively; $\mathbf{y}(t) = \bar{s}_n(t)$ denotes the position measurement (i.e. distance from the starting point) at time t ; $\mathbf{w}(t) = [0 \ 0 \ \theta(t)]^T$ denotes the process noise; $\mathbf{v}(t) = \varepsilon_s(t)$ denotes the measurement noise; $\mathbf{H}$ denotes the relation matrix between the measurement and state vector and indeed is simply a row vector $[1 \ 0 \ 0]$. $\mathbf{A}$ denotes the state

Figure 3: The trajectory estimated from GPS coordinate measurement using KF(upper) and the measured and estimated speed profiles (lower).

transition matrix and is specified as follow:

$$\mathbf{A} = \begin{pmatrix} 1 & \Delta t & \Delta t^2/2 \\ 0 & 1 & \Delta t \\ 0 & 0 & \phi \end{pmatrix}.$$

In (Ma and Andréasson, 2007), a first-order autoregressive (AR) model described by the following equation was used to represent the acceleration time series

$$a_n(t+1) = \phi a_n(t) + \theta(t) \quad (2)$$

where $\theta(t)$ denotes a white noise following the Gaussian distribution $N(0, \sigma^2)$; ϕ denotes a parameter which determines the AR process. When ϕ is set to 1, the model represents a random walk process in which the acceleration at the current point of time depends on both the point of time and a random noise term $\theta(t)$.

The cyclist GPS device brings for each trip a sequences of (x, y) coordinate measurements . This can also be easily fit for the presented KF estimation scheme. However, the state vector includes states from two different dimensions i.e. $\mathbf{x}(t) = [s_{n,x}(t) \ s_{n,y}(t) \ v_{n,x}(t) \ v_{n,y}(t) \ a_{n,x}(t) \ a_{n,y}(t)]^T$, which includes position, speed and acceleration for the axis of x and y . So $\mathbf{A}$ and $\mathbf{H}$ have to be adapted to the case of the 2D coordinate measurement (Luo and Ma, 2014). Figure 3 shows a bicycle trajectory estimated by the 2D KF filtering scheme using GPS measurement and a comparison of estimated and measured bicycle speed profiles.

Figure 4: An illustration of the identification of cycling regimes from the filed data. (a) Examples of identified acceleration, deceleration and cruising processes; (b) Detailed speed and acceleration profiles extracted from (a); (c) Detailed speed and deceleration profiles extracted from (a).

3.2. Cycling regimes

While looking through the speed and acceleration profiles, it was found that, when significant increases in speed were carried out by the cyclists, the patterns of the change in acceleration rates during the same period showed great similarity regardless of other factors. Specifically, the U-shape acceleration-time profiles were frequently observed together with the S-shape speed-time profiles. Likewise, the same patterns could be observed in the deceleration cases. Figure 4a illustrates a speed profile of 400 seconds, together with its corresponding acceleration profile. Besides the normal curves, a significant acceleration process and a deceleration process are also marked by red and green dash

lines, respectively. Detailed profiles for both acceleration and deceleration cases are further demonstrated in Figure 4b and Figure 4c, respectively.

Apart from those two types of processes which could almost be certainly identified, some processes during which no striking change in speed were also observed with their corresponding acceleration profiles merely exhibiting insignificant fluctuations around zero. In the current study, such processes are labeled as cruising processes. Figure 4a shows some examples for each kind of process. As a summary, three bicycling regimes including acceleration, deceleration and cruising were proposed. Given the regimes, the cycling behavior can be roughly described by a way that a cyclist always endeavors to achieve and maintain a desired speed which varies in real time depending on multiple factors, such as the weather, road grade and road width (external factors) as well as the age and gender of the individual (internal factors). The processes during which the cyclist maintains her present desired speed corresponds to the cruising behavior, whereas the processes during which she tries to achieve desired speed refers to either acceleration or deceleration behavior. The approach can be analytically represented by

$$a(t) = \begin{cases} a_{acc}(t) & \text{if } v(t) < v_{des} \\ a_{dec}(t) & \text{if } v(t) > v_{des} \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

where $a(t)$ denotes the acceleration rate at time t ; $a^{acc}(t)$ and $a^{dec}(t)$, respectively, denote the acceleration rate and deceleration rate at time t ; v_n^{des} denotes the desired speed of the cyclist n at time t which depends on a series of unknown factors.

An empirical framework was applied to distinguish significant acceleration or deceleration profiles from the original dataset. While consecutive speed-up and slowing-down clusters were first identified, more detailed selection was carried out to find potential acceleration processes. Based on the overall observation, some empirical rules are applied to facilitate the search:

- The acceleration time of a process t_a is in the range of 5 seconds and 16 seconds;
- The total distance during the acceleration process d_a should be longer than 5 meters;
- The road gradient over the entire acceleration process g_a should not be greater than 10%;
- The variance in speed of an acceleration process is significant in a way that the value of a newly defined index τ should not be lower than 0.5:

$$\tau = \frac{\Delta V}{\max(v_f, v_i)} \quad (4)$$

$$\Delta V = |v_f - v_i| \quad (5)$$

where v_i and v_f , respectively, denote the initial speed and final speed of a process.

The entire procedure was implemented by integrating MATLAB scripts and SQL queries, mainly invoking combined SQL queries through MATLAB. As a result, 831 acceleration profiles (8,005 observations) and 704 deceleration profiles (5,559 observations) respectively stood out. Analysis on these profiles then revealed multiple characteristics of acceleration and deceleration processes which are in part illustrated in Figure 5. Due to the resemblance between the acceleration case and deceleration case, only plots for the former one are presented in the article. It can be seen that in most cases, the variance in speed during the acceleration process lies in a range between 2 m/s and 6 m/s, while the maximum acceleration rate mainly varies between 0.5 m/s² and 1 m/s².

One important finding that we figured out from data analysis is that the maximum acceleration a_{max} of an acceleration process could be, to a large degree, explained by the difference ΔV between the initial and final speed of the process. Figure 5c shows a strong linear correlation between the logarithm transform of the two variables. In addition, the distribution of acceleration time shown in Figure 5d indicates that cyclists normally require around 10 seconds to complete an acceleration process. In order to model the cyclist acceleration process, it is important to first predict the duration of an acceleration (or deceleration) process. This study shows that the duration can be estimated by the

Figure 5: Some characteristics of cyclists' acceleration behavior. (a) Distribution of the variance in speed (ΔV); (b) Distribution of the maximum acceleration a_{max} ; (c) The linear relationship between transformed ΔV and a_{max} ; (d) Distribution of the acceleration time t_a .

initial speed and final desired speed (also the difference between them). Using the collected data, the acceleration and deceleration durations are estimated respectively and lead to the following two empirical models:

$$t_a^{acc} = \frac{\Delta V}{0.1334 \cdot \Delta V^{0.8069} + 0.003 \cdot v_i} \quad (6)$$

$$t_a^{dec} = \frac{\Delta V}{0.1284 \cdot \Delta V^{0.7773} + 0.0113 \cdot v_f}. \quad (7)$$

3.3. Acceleration model

Given the aforementioned cycling regimes, one of the objectives of the current study becomes to develop models which are capable of depicting the particular acceleration and deceleration profiles. Besides the intuitive feature of U-shape curve, another one from a mathematical perspective should also be taken into consideration that zero acceleration, $a = 0$, and zero jerk, $da/dt = 0$, at the start and end of both acceleration and deceleration cases (time 0 and time t) must be satisfied. As a result, a polynomial model introduced by Akçelik and Biggs (1987) in an attempt to model driver's acceleration profiles was adopted in our study. The form of the polynomial model is shown as follow:

$$y = x^n(1 - x^m)^2 \quad (8)$$

where parameter n and m are to be estimated and they together determine the shape of the curve. One of the merits of using this polynomial model is that it meets two requirements on acceleration and speed profiles: zero acceleration at the beginning and end of an acceleration process; zero jerk ($dy/dx = 0$) at the beginning and end of an acceleration

process. Based on the model form and intuition on the relation between maximum acceleration and speed difference during an acceleration process, a new acceleration model is proposed as follows:

$$a_g(t) = \lambda_g \cdot \Delta V^{\beta_g} \cdot \theta(t)^{p_g} (1 - \theta(t)^{q_g})^2 + \varepsilon_g(t) \quad (9)$$

$$\theta(t) = \frac{t}{t_a^g} \quad (10)$$

where $a_g(t)$ denotes the acceleration rate at time t and $g \in \{acc, dec\}$; λ_g , β_g , p_g and q_g denote parameters to be determined. ΔV denotes the discrepancy between the cyclist's current speed and her final speed which refers to the desired speed in the simulation for the ongoing process. $\theta(t)$ denotes the ratio of the current acceleration time against the total duration time predicted by Equation (6) and (7). $\varepsilon_g(t)$ denotes the random term associated with the acceleration rate at time t . It captures the effect of omitted variables and is assumed to be independently and identically distributed. Given the proposed model form labeled as Model 3, another two simplified model forms (Model 1 and Model 2) with each omitting certain parameters were also suggested with the aim of model comparison:

Model 1

In the first model, only indispensable parameters including λ_g and q_g were reserved.

$$a_g(t) = \lambda_g \cdot \Delta V \cdot \theta(t)(1 - \theta(t)^{q_g})^2 + \varepsilon_g(t) \quad (11)$$

Model 2

The second model lets parameter β_g enter the model so that ΔV 's effect on the final acceleration can be reflected through β_g .

$$a_g(t) = \lambda_g \cdot \Delta V^{\beta_g} \cdot \theta(t)(1 - \theta(t)^{q_g})^2 + \varepsilon_g(t) \quad (12)$$

Model 3

The third model specification retains all the parameters that show up in the general form. The purpose of this specification is to examine parameter p_g 's influence on the shape of the model curve.

$$a_g(t) = \lambda_g \cdot \Delta V^{\beta_g} \cdot \theta(t)^{p_g} (1 - \theta(t)^{q_g})^2 + \varepsilon_g(t) \quad (13)$$

3.4. Model estimation using Maximum Likelihood method

The Maximum Likelihood (ML) method is known as a general strategy for obtaining asymptotically optimal estimators and has been a popular model identification approach in transport and operation research. By assuming that the random term in Equation (9) follows different distributions, different probability density functions of the variable $a_g(t)$ can be obtained, thus leading to different formulation of likelihood functions and estimation results. Specifically, normal distribution and Laplace distribution were taken into account for the random term in the current study based on preliminary numerical experiments.

Laplace distribution

Assuming that the random term follows a zero-mean Laplace distribution shown as follow:

$$\varepsilon_g(t) \sim \text{Lalace}(0, \sigma_g) \quad (14)$$

where σ_g denotes the standard deviation. Then, the probability density function of the acceleration $a_g(t)$ is given by

$$f(a_g(t)) = \frac{1}{\sqrt{2}\sigma} \exp\left(-\frac{\sqrt{2}|a_g(t) - \lambda_g \cdot \Delta V^{\beta_g} \cdot \theta(t)^{p_g} (1 - \theta(t)^{q_g})^2|}{\sigma}\right) \quad (15)$$

The likelihood function is given by

$$\mathcal{L} = \prod_{t=1}^N \frac{1}{\sqrt{2}\sigma} \exp\left(-\frac{\sqrt{2}|a_g(t) - \lambda_g \cdot \Delta V^{\beta_g} \cdot \theta(t)^{p_g} (1 - \theta(t)^{q_g})^2|}{\sigma}\right) \quad (16)$$

The log likelihood function is given by

$$\mathcal{LL} = -N \ln(\sqrt{2}\sigma) - \frac{\sqrt{2}}{\sigma} \sum_{t=1}^N |a_g(t) - \lambda_g \cdot \Delta V^{\beta_g} \cdot \theta(t)^{p_g} (1 - \theta(t)^{q_g})^2| \quad (17)$$

Table 1: Estimation results of all acceleration models with the random term assumed to follow Laplace distribution.

Parameter	Acceleration			Deceleration		
	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
λ	0.5381	0.7430	0.8247	-0.3615	-0.5664	-0.7435
q	2.2773	2.2373	2.1023	3.7801	3.7868	3.0999
β	–	0.7877	0.8274	–	0.6942	0.6967
p	–	–	1.0585	–	–	1.2058
σ (m/s ²)	0.2543	0.2520	0.2520	0.2353	0.2302	0.2294
Log Likelihood	146.8	206.6	207.2	555.9	675.5	695.3
Observations	6503	6503	6503	5532	5532	5532

Note: all parameters are statistically significant

Table 2: Estimation results of all acceleration models with the random term assumed to follow normal distribution.

Parameter	Acceleration			Deceleration		
	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
λ	0.5053	0.6679	0.4482	-0.3735	-0.5505	-0.4638
q	2.4126	2.4029	3.2184	3.5442	3.5790	4.0977
β	–	0.8247	0.8274	–	0.7474	0.7495
p	–	–	0.7466	–	–	0.8649
σ (m/s ²)	0.2512	0.2497	0.2488	0.2273	0.2241	0.2238
Log Likelihood	-244.4	-204.2	-181.9	345.0	424.9	433.1
Observations	6503	6503	6503	5532	5532	5532

Note: all parameters are statistically significant

Normal distribution

When the random term $\varepsilon_g(t)$ is assumed to follow a zero-mean Normal distribution as follow:

$$f(a_g(t)) = \frac{1}{\sigma_g} \phi \left(\frac{a_g(t) - \lambda_g \cdot \Delta V^{\beta_g} \cdot \theta(t)^{p_g} (1 - \theta(t)^{q_g})^2}{\sigma_g} \right) \quad (18)$$

the log likelihood function for all observations is then given by

$$\begin{aligned} \mathcal{LL}(\lambda_g, \beta_g, p_g, q_g, \sigma_g | a_g(t)) &= \sum_{n=1}^{N_g} \ln [f(a_g(t))] \\ &= -\frac{N_g}{2} \ln(2\pi) - \frac{N_g}{2} \ln \sigma_g^2 - \frac{1}{2\sigma_g^2} \sum_{n=1}^{N_g} [a_g(t) - \lambda_g \cdot \Delta V^{\beta_g} \cdot \theta(t)^{p_g} (1 - \theta(t)^{q_g})^2]^2 \end{aligned} \quad (19)$$

where N_g denotes the number of all observations in each case.

3.5. Estimation and validation results

Maximizing the likelihood functions formulated above would provide the ML estimates of the model parameters. The current study applies 80% of the randomly selected acceleration, or deceleration, profiles of the total cyclist population whereas the rest 20% profiles are used for subsequent model validation. The ML estimation was implemented using [MATLAB \(2013\)](#). The estimation results for both acceleration and deceleration cases are summarized in Table

Figure 6: Examples of the residual distributions with normal and Laplace distribution fits for model 3 of the acceleration case. (a) The residual results given that the random term is assumed to follow normal distribution; (b) The residual results given that the random term is assumed to follow Laplace distribution

Figure 7: An illustration of the model validation with the random term assumed to follow normal distribution. (a) Comparison of acceleration and speed profiles for different models for the acceleration case; (b) Comparison of acceleration and speed profiles for different models for the deceleration case.

1 and 2. Figure 6 presents two examples of the residual distributions and the fitted curve for Normal and Laplace distribution. All variables turned out to be statistically significant at the 95% confidence level. A comparison between Model 1 and Model 2 in both acceleration and deceleration cases indicates that the entry of parameter β_g mainly had an impact on λ_g while the magnitude of q_g did not vary much. This is easy to explain because the parameter β_g which is smaller than 1 diminished the effect of ΔV on the final value of acceleration so that the parameter λ_g had to accordingly increase to compensate for the shortage in the estimate. When the complete form of the model with parameter p_g included were estimated, the value of β_g did not change significantly while λ_g and q_g became smaller and larger, respectively. This, however, could be explained by the feature of the original polynomial model.

As mentioned, the cross validation technique was applied for the proposed models using 20% of total profiles

Figure 8: Comparison of acceleration and speed profiles for model 3 with different assumptions of the random term. (a) An example of the acceleration case; (b) An example of the deceleration case.

randomly selected from the datasets. This is to evaluate how estimated models generalize to independent dataset. The validation was performed based on the derived speed profiles instead of the estimated acceleration profiles. Examples of simulated profiles can be seen in Figure 7. A number of performance indexes (PIs) were further computed to assess different models, including the root mean square error (RMSE), Theil's inequality coefficient (U), the mean error (ME) and mean absolute percentage error (MAPE). Theil's inequality coefficient quantifies the overall error of the validation. MAPE reflects the existence of systematic under- or over-prediction by the developed models. It is also worth mentioning that a simple remedy was made in this process in order to eliminate observations concerning speeds close to zero when the models were calibrated (a threshold of 0.5 m/s in terms of speed was adopted here). The estimation of PIs are summarized as follows:

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{n=1}^N (Y_n^p - Y_n^o)^2} \quad (20)$$

$$\text{MAPE} = \frac{1}{N} \sum_{n=1}^N \left| \frac{Y_n^p - Y_n^o}{Y_n^o} \right| \quad (21)$$

$$\text{ME} = \frac{1}{N} \sum_{n=1}^N (Y_n^p - Y_n^o) \quad (22)$$

$$U = \frac{\sqrt{\frac{1}{N} \sum_{n=1}^N (Y_n^p - Y_n^o)^2}}{\sqrt{\frac{1}{N} \sum_{n=1}^N (Y_n^p)^2} + \sqrt{\frac{1}{N} \sum_{n=1}^N (Y_n^o)^2}} \quad (23)$$

Table 3 summarizes a comparison of validation results on cyclist speed using various PIs for both acceleration deceleration processes. Although the models with more parameters seem to give slightly better prediction performance, the differences among them are quite small. This implies that Model 2 with nonlinear parameter on the desired speed difference is good enough to be used in the simulator.

Table 3: Validation statistics for speed comparison for both acceleration and deceleration cases.

Statistics	Acceleration			Deceleration		
	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
RMSE (m/s)	0.88	0.83	0.83	0.93	0.87	0.88
MAPE (%)	15.4	15.3	15.3	25.9	23.9	23.9
ME (m/s)	0.15	0.12	0.11	-0.06	-0.02	-0.01
U (fraction)	0.0029	0.0027	0.0027	0.0031	0.0029	0.0029

3.6. Effects of social-economic characteristics

Since the social-economic characteristics of the cyclists such as gender and age are also known for the cyclist population, the proposed model is extended to analyze the effects of those factors. In our further analysis, we mainly consider two factors: gender and agility. Agility is not directly discriminable. But in our population there are cyclists who regularly ride racing bike and are more used to bicycling than the rest, and meanwhile age is correlated to the desired speed level. Therefore, the effects of the factors are expected in the model analysis.

In light of the concerns above, Model 2 is thus extended to incorporate both factors:

$$\begin{aligned}
 a_g(t) &= \lambda_g \cdot \Delta V^{\beta_{1,g}X_1 + \beta_{2,g}X_2} \cdot \theta(t)(1 - \theta(t)^{q_s})^2 + \varepsilon_g(t) \\
 X_1 &= \begin{cases} 1 & \text{male} \\ 0 & \text{female} \end{cases} \\
 X_2 &= \begin{cases} 1 & \text{high agility using racing bike} \\ 0 & \text{others} \end{cases}
 \end{aligned} \tag{24}$$

where X_1 and X_2 denote the explanatory dummy variables of cyclist's gender and agility respectively. $\beta_{1,g}$ and $\beta_{2,g}$ are corresponding model parameters. Indeed, the results from model estimation shows that both gender and agility have positive impacts, and they are statistically significant. However, the Log likelihood suffers a decrease, indicating that the model with extended explanatory variables doesn't provide as desirable results as the simple form with a single parameter. This can be attributed to the limitation of the data diversity given the small cyclist population size. Meanwhile, other factors such as age and physical condition may also play critical roles on cyclist acceleration performance. In the future work, this issue can be examined again when more cyclist behavior datasets are collected.

4. Conclusions

This paper presents a microsimulation-oriented modeling framework for bicycle movement and interaction on links. In particular, the development of the modeling framework is based on in-depth understanding of cyclist behavior from a large amount of naturalistic data. The study bridges the gap by exploring and modeling cyclist behavior with real naturalistic data instead of only experimental data. The naturalistic data comes from two sources: a high-sensitivity integrated GPS receiver and internal barometric altimeter. Collected raw data are processed by applying the Kalman smoothing algorithm to more reliable GPS data, and locally weighted regression to more consistent altitude information, respectively. Information in cycling speed and acceleration rate is in the meanwhile estimated, and gradient profiles are derived based on both altitude and distance data.

In the data analysis, bicycle trajectories are empirically classified into three different regimes: cruising, acceleration and deceleration. The unique U-shape curves for both acceleration and deceleration processes are observed and further used to promote the formulation of dedicated acceleration models for cyclists. As a result, a main contribution of the paper is the acceleration model proposed for representing the longitudinal movement of cyclist. Maximum likelihood estimation approach is applied to identify the model parameters while the residual is assumed to follow different distributions. In model identification, various forms of the proposed model are estimated and compared with their likelihood values. Model validation shows that the proposed model and its variation are all capable to capture the observed acceleration and deceleration processes, although more complex model seems giving slightly better performance. In addition, the social economic factors was analyzed by adding them as explanatory dummy variables in

the model. The estimation shows that both gender and agility of the cyclists demonstrate significant effects on the acceleration decision.

One limitation in the study is that the altitude data, though carefully processed, doesn't show significant impact on the speed profiles. This is mainly because the current device doesn't give us altitude measurement with good quality, and other road databases with relevant information should be applied in the future to enhance the model. The next step for completing the bicycle traffic simulation model includes the development of cyclist-following model using similar bicycle trajectory data. Moreover, the monitoring data on roads should also be collected for better understanding of bicycle traffic patterns on the network.

Acknowledgment

The authors would like to acknowledge the funding support by the Länsförsäkringar Alliance's Research Foundation (P13-12) and J. Gust. Richert Research Foundation (2013). The authors are grateful to Prof. Em. Lars-Göran Mattsson and other members in the project reference group for their advice and interest.

References

- Addison, P., Low, D., December 1998. A novel nonlinear car-following model. *Chaos* 8 (4), 791–799.
- Akçelik, R., Biggs, D., 1987. Acceleration profile models for vehicles in road traffic. *Transportation Science* 21 (1), 36–54.
- Botma, H., 1995. Method to determine level of service for bicycle paths and pedestrian-bicycle paths. *Transportation Research Record* (1502), 38–44.
- Botma, H., Papendrecht, H., 1991. Traffic operation of bicycle traffic. *Transportation Research Record* (1320).
- Botma, H., Papendrecht, H., 1993. Operational quality of traffic on a bicycle path. In: *Compendium of Technical Papers, ITE, 63rd Annual Meeting*.
- Cleveland, W. S., Devlin, S. J., 1988. Locally weighted regression: an approach to regression analysis by local fitting. *Journal of the American Statistical Association* 83 (403), 596–610.
- Dozza, M., Fernandez, A., 2014. Understanding bicycle dynamics and cyclist behavior from naturalistic field data (november 2012). *Intelligent Transportation Systems, IEEE Transactions on* 15 (1), 376–384.
- Faghri, A., Egyházióvá, E., 1999. Development of a computer simulation model of mixed motor vehicle and bicycle traffic on an urban road network. *Transportation Research Record: Journal of the Transportation Research Board* 1674 (1), 86–93.
- Ferrara, T. C., Lam, T. N., 1979. Analysis of bicycle delays at intersections and crossings by computer simulation. *Transportation Research Record: Journal of the Transportation Research Board* 706, 36–44.
- Gould, G., Karner, A., 2009. Modeling bicycle facility operation. *Transportation Research Record: Journal of the Transportation Research Board* 2140 (1), 157–164.
- Johnson, M., Charlton, J., Oxley, J., Newstead, S., 2010. Naturalistic cycling study: identifying risk factors for on-road commuter cyclists. In: *Annals of Advances in Automotive Medicine/Annual Scientific Conference*. Vol. 54. Association for the Advancement of Automotive Medicine, p. 275.
- Joo, S., Oh, C., 2013. A novel method to monitor bicycling environments. *Transportation research part A: policy and practice* 54, 1–13.
- Kalman, R., 1960. A new approach to linear filtering and prediction problems. *ASME Transaction: Journal of Basic Engineering* **D**, 35–45.
- Kosonen, I., 1999. Hutsim - urban traffic simulation and control model: Principles and applications. Ph.D. thesis, Helsinki University of Technology.
- Luo, D., Ma, X., 2014. Analysis of cyclist behavior using naturalistic data: data processing for model development. In: *Proceedings of the 3rd International Cycling Safety Conference (ICSC2014)*. Gothenburg, Sweden.
- Ma, X., Andréasson, I., 2005. Dynamic car following data collection and noise cancellation based on the Kalman smoothing. In: *Proceedings of IEEE International Conference on Vehicular Electronics and Safety (ICVES'05)*.
- Ma, X., Andréasson, I., 2007. Behavior measurement, analysis and regime classification in car-following. *IEEE Transactions On Intelligent Transportation Systems* 8 (1), 144–156.
- MATLAB, 2013. version 8.1.0.604 (R2013a). The MathWorks Inc., Natick, Massachusetts.
- Nagel, K., Schreckenberg, M., 1992. A cellular automaton model for freeway traffic. *Journal de physique I* 2 (12), 2221–2229.
- Oketch, T. G., 2000. New modeling approach for mixed-traffic streams with nonmotorized vehicles. *Transportation Research Record: Journal of the Transportation Research Board* 1705 (1), 61–69.
- Parkin, J., Rotheram, J., 2010. Design speeds and acceleration characteristics of bicycle traffic for use in planning, design and appraisal. *Transport Policy* 17 (5), 335–341.
- Schönauer, R., Stubenschrott, M., Huang, W., Rudloff, C., Fellendorf, M., 2012. Modeling concepts for mixed traffic. *Transportation Research Record: Journal of the Transportation Research Board* 2316 (1), 114–121.
- Taylor, D., Davis, W. J., 1999. Review of basic research in bicycle traffic science, traffic operations, and facility design. *Transportation Research Record: Journal of the Transportation Research Board* 1674 (1), 102–110.
- Twaddle, H., Schendzielorz, T., Fakler, O., 2014. Bicycles in urban areas. *Transportation Research Record: Journal of the Transportation Research Board* 2434 (1), 140–146.
- Walker, I., 2007. Drivers overtaking bicyclists: Objective data on the effects of riding position, helmet use, vehicle type and apparent gender. *Accident Analysis & Prevention* 39 (2), 417–425.

Zhao, D., Wang, W., Li, C., Li, Z., Fu, P., Hu, X., 2013. Modeling of passing events in mixed bicycle traffic with cellular automata. *Transportation Research Record: Journal of the Transportation Research Board* 2387 (1), 26–34.