Red or Green: Estimating the Patterns of Traffic Signal through Cyclists’ GPS Tracks for Real Time Navigation

Robert Schönauer (Corresponding Author)
mobimera Fairkehrstechnologien
Taborstrasse 11b/19
A-1020 Vienna, Austria
Phone: +43 (0) 650 8415746
schoenauer@mobimera.at

Gerald Richter
Dynamic Transportation Systems
Austrian Institute of Technology
Giefinggasse 2
A-1210 Vienna, Austria
Phone: +43 (0) 50550-6072
gerald.richter@ait.ac.at

Markus Straub
Dynamic Transportation Systems
Austrian Institute of Technology
Giefinggasse 2
A-1210 Vienna, Austria
Phone: +43 (0) 50550-6289
markus.straub@ait.ac.at

Christian Rudloff
Dynamic Transportation Systems
Austrian Institute of Technology
Giefinggasse 2
A-1210 Vienna, Austria
Phone: +43 (0) 50550-6341
christian.rudloff@ait.ac.at

Melitta Dragaschnigg
Dynamic Transportation Systems
Austrian Institute of Technology
Giefinggasse 2
A-1210 Vienna, Austria
Phone: +43 (0) 50550-6255
melitta.dragaschnigg@ait.ac.at

Submitted: August 1, 2013

Word Count: 4836
Table: 1
Figures: 9
Total no. words: 7336
ABSTRACT

Advanced routing and navigation incorporating real time traffic information is state of the art - at least for car drivers. The number of cyclists using digital navigation even to reach familiar destinations is increasing with the availability of waterproof smartphones and smartphone mounts. However, for realistic routing and real-time cycling assistance, profound knowledge about expected waiting times at traffic signals and cycling behavior between the stops is essential. GPS tracks of cyclists can be used to obtain both even when real time information of the traffic control is not provided. This paper presents a first step in this direction by suggesting methods to analyze waiting times at signaled intersections.

GPS tracks have been acquired by cyclists riding specific routes between two locations in the city center of Vienna. Methods to preprocess the GPS tracks are applied. An algorithm to identify signal control cycle lengths for intersections is proposed. Waiting times and speed gradients are analyzed to estimate the offset times and green-light duration.

Cyclists’ profile data are derived from these cleaned tracks and analyzed for significance to represent cycling behavior in free flow traffic between traffic lights, thus allowing for prediction of arrival times at signals which is the larger goal of ongoing research.

Finally, for ten traffic lights along a bidirectional route, results of both estimated signal cycle lengths and signal patterns are shown. Positive verification of the GPS track derived results is achieved by comparison against real signal programs.
INTRODUCTION

Current State Of Bicycle Navigation Tools

Technological evolution within the mobile consumer market lead to a global smartphone penetration rate of around 30% (1). Easy setup and secondary use of smartphones’ location services both offer immense benefits.

Faster hardware and feature-rich applications lead to an increasing use of navigation devices in cars (2) and also show the potential for cyclists: in recent years usage of bicycle navigation devices, especially in the form of smartphone apps, is growing. Touristic usage on mountain bike tours is proposed in (3), apps for urban sightseeing trips (4) and leisure tours (5) already exist.

Challenges

Urban traffic control systems are primarily optimized for motorized traffic. In comparison cyclists typically have to stop more often and wait longer at traffic signals, especially at big intersections.

Calculating pleasurable and time-saving routes for cyclists is difficult without knowledge about traffic signals and user profiles. Both traffic signals patterns (i.e. how long is the green phase for cyclists) and synchronization of the traffic control at consecutive intersections (i.e. at which speeds can the next traffic signal be reached during the green phase) are of relevance. User profiles should contain abstract information about fitness and preferred cycling styles (e.g. acceleration, sporty vs. relaxed, comfortable speed, bicycle type used).

Traffic Signals

In many cities traffic signals are operated using a fixed schedule rather than adapting to traffic situations. While traffic flow optimization systems like SCOOT (6) would allow such adaptation their application is not commonplace. The programming of green waves through a controlled urban network often is based on a central clock and a fixed cycle length.

For fixed timing traffic signals, the cycles are precisely known, i.e. at what time and for how long in a cycle are certain relations of in- and out-lanes of an intersection traversable. The cycle length is defined as the time it takes a traffic signal to complete one full signal program cycle.

Modern traffic management is often coordinated centrally, but lacks publicly accessible interfaces. However, there exist several other possibilities to acquire real time data or stored data of traffic signals. Vision technology is an exemplary approach introduced by smartphone apps like SignalGuru (7) where signal patterns are extracted from videos and GPS-tracks provided by the app’s users. With this approach, however, synchronization is difficult.

Outline Of Our Approach

This paper presents an approach to solely use GPS tracks of cyclists to search for patterns in waiting times and create user profiles. A pattern contains green and red as well as intermediate phases. These patterns can be used to improve routing and to create driving assistance, providing hints on how to cycle efficiently by arriving at traffic signals when they are green. Efforts to acquire tracks from cyclists are described, preprocessing is applied to the obtained data and cyclists’ velocity profiles are calculated. A “signal pattern finder” generates estimates for signal periods and green times. The results are discussed, further research and applications are proposed.
1 SIGNAL ESTIMATION METHODS

2 Estimated Parameters

3 The duration of the green period $T^g$ with $t^o$ as its begin lies between intermediate phases $T^{i1}$ and $T^{i2}$ that contain blinking or yellow signals - the specifics of these phases vary from country to country. $T^c$ is the time span of the whole signal cycle. In most countries the possible cycle lengths range from about 40 to 120 seconds (8).

4 In the presented paper the goal is to find $T^c$, $t^o$ and $T^g$. Those parameters in most cases are sufficient to estimate waiting times at signals for a given time of arrival.

9 Methods

10 In this chapter cyclists are represented as agents with discrete state vectors $S_i(t)$ for agent $i$ at time $t$. Each of these states $S_i(t)$ contains vectors for position $\vec{x}_i = \vec{x}_i(t)$, velocity $\vec{v}_i = \vec{v}_i(t)$ and acceleration $\vec{a}_i = \vec{a}_i(t)$ as functions of time and

$$S_i(t) = S_i(\vec{x}_i, \vec{v}_i, \vec{a}_i).$$

11 To identify the relevant stopping events of the agents two filters are applied to the trajectories.

14 I. Spatial filter

15 The filter is related to position $\vec{c}_i$ of signal $j$. The threshold $d_{th}$ defines a spatial tolerance in considering approaching, decelerating agents.

$$F_1 = \{t | |\vec{x}_i - \vec{c}_i| < d_{th}\}$$

17 II. Velocity filter

18 If speed $|\vec{v}_i|$ drops under the defined threshold $v_{th}$, the related track section is considered to be relevant:

$$F_2 = \{t | |\vec{v}_i| < v_{th}\}$$

20 The resulting state vector set $S^*_{i,j} = \{S_i(t) | t \in F_1 \cap F_2 \}$ contains the track segments relevant for further processing. In FIGURE 1 a sample topology is shown indicating $S^*_{i,j}$ referencing two opposing signals for the horizontal crossings.

23 Orientation of $\vec{v}_i$ is another factor which could be considered to avoid misinterpretation of nearby tracks. Small chosen values in $d_{th}$ obtain the desired driving directions and yield the same result because this avoids consideration of tracks on the lane in the opposing direction.

26 Cycle Length Identification

27 Before estimating the green and red phases of a signal the periodicity of the signal program has to be identified. The assumption for application of the presented method is the utilization of a central clock - thus likely resulting in an identical signal phase being active every day at the same time. Even if there are dynamic offsets or subprograms the central control retains a fixed periodicity.
FIGURE 1 A sample intersection topology, all black and colored dots are positions with speeds under the threshold \( S_i(t) \cap S_i(t) \mid t \in F_s \). In addition, the colored positions represent \( S_{ij}^- \) for magenta and cyan indicate different signals \( j_1 \) and \( j_2 \).

The main idea underlying the identification of the cycle length relies on the fact that cyclists arriving at green light show no waiting time while cyclists arriving at red light will need to wait and hence show more observations in the set \( F_1 \cap F_2 \). Thus ideally when providing a histogram of the waiting periods of these observations we would see green light times as bins with very low number of observations while red light times will feature eventual observations.

In a first step we find the waiting times in a specific stopping area at a specific signal \( j \) corresponding to some assumed cycle length \( T_j^c \):

\[
T_{i,j}^w = t(S_{i,j}^+) \mod T_j^c
\]

Here, \( T_{i,j}^w \) is the set of (discrete) time values containing the waiting periods of all relevant agents within the signal cycle length. For the sake of brevity all further notations are considering only the single signal \( j \) at a specific intersection and we thus write \( T_{i,j}^w = T_{i,j}^w \) and \( T_j^c = T_j^c \) for short.

\( T_{i,j}^w \) is obtained by a modulo division of the times spent waiting \( t(S_{i,j}^+) \) by a chosen \( T_j^c \). Consecutive histogramming is used to structure the calculated waiting times as

\[
h_b = \text{hist}(T_{i,j}^w; T_b, T_j^c) \quad \text{with} \quad b = \{1, \ldots, n_b\}
\]

which depends on \( T_b \) (being the time-extent of a histogram bin) and a cycle length. \( n_b \) is the number of bins in the histogram and evaluates to \( n_B = T_j^c / T_b \). Each of the bins \( h_b \) contains the count of \( T_{i,j}^w \) falling within its range \( \Delta T_b \) which are used for identifying the signal cycle length.

Thereafter we find neighboring bins having values below a threshold \( h^{th} \) which we denote as the set

\[
\{N_b^{th}\} := \#(\text{neighboring bins } h_b \mid h_b \leq h^{th})
\]
Obtaining the maximum value of $N_{b}^{th}$ is indicating the highest likelihood of an uninterrupted green phase. Thus,

$$T^g = \max \{\{N_{b}^{th}\}\} \cdot T_b$$

is the green time duration resulting from the longest group of bins containing no observed waiting agents and therefore represent the green phase in the signal’s program. Setting $T^g$ in relation to $T_c$ expresses a relative green time within the signal period.

$$R^g = \frac{T^g}{T_c}$$

Given a fixed signal program, a balanced green/red relation for the crossings and a large number of observations $R^g$ would yield values close to $\frac{1}{2}$. In order to identify the most likely cycle length $T^{c*}$, while considering that $T^g = T^g(T_b, T_c)$, $R^g$ can be maximized under variation of $T_c$:

$$T^{c*} = \arg\max_{T_c} (R^g(T_c))$$

Having found $T^{c*}$ leads to a signal waiting pattern for further processing, by histogramming as before:

$$h_{b}^{*} = \text{hist}(t(S^{*}) \mod T^{c*})$$

This approach might lead to misclassification, especially at low data density and small real $R^g$. FIGURE 2 shows the typical shape of such a cumulative histogram.

![Histogram](image)

FIGURE 2: Example of a histogram $h_{b}^{*}$ showing the typical saw tooth pattern. The yellow bar indicates $h_{b}^{*}$.

It has to be noted that the presented approach is not only capable to detect signals at intersections with traffic signals but also waiting times at intersections without signal control, where traffic flow characterized by control patterns of neighboring signals propagates to the subsequent intersections.
Estimating Green Phase Starts

A difference quotient of the pattern histogram yields the change of the number of waiting cyclists. For our purposes, we assume its minimum value to define the time bin $h_g^*$ in which most agents end waiting after the signal turns green (marked yellow in FIGURE 2) where index $g$ defines as:

$$g \mid \frac{\Delta h_{i}^*}{T_b} = \min \left( \frac{\Delta h_{i}^*}{T_b} \right)$$

For cyclists behaving corresponding to signal indication this method is deemed reliable. Considering red light runners and queuing, the slope is smoothened, especially at the right-hand side of the histogram.

Estimating Red Phase Starts

Knowing the time-span $T^g*$ of the longest identified green time and a corresponding specific start bin $h_g^*$ we can search for the beginning of the red light in the periodic pattern.

Starting at the bin $h_g^*$ we cumulatively sum up the bin entries and define a sensitivity factor $f_r$ as a certain fraction of the maximum cumulative waiting time, which is used as a threshold value. This yields the bin $h_r^*$ which identifies the beginning of the red light phase, where index $r$ is

$$r \mid \left( \sum_{i=g}^{r} h_i^* \right) \geq f_r \times \max (h_i^*)$$

With the identification of the histogram indices $g, r$ and their respective times for a determined cycle length $T^{c*}$, a signal pattern is defined under the assumptions stated before.
DATA ACQUISITION, PREPROCESSING AND CYCLISTS’ PROFILES

Data Acquisition

GPS Track Collection

20 cyclists, recruited with help of a bicycle advocacy group, were paid to cycle equally often along four predefined routes (two courses, both direction) in Vienna (Austria) for five hours using different bicycles. FIGURE 3 shows the routes leading from Museumsquartier to the main building of the University of Vienna either via Ringstraße or Museumstraße and vice versa. Data acquisition took place in the first half of May 2013 during warm but windy weather. Teltonika GH3000 GPS trackers set to a sampling frequency of 1 Hz were used.

Several constraints on cycling were defined. To ensure that identical signal programs are active and traffic jams are avoided, times were restricted to work days from 10am to 3pm and from 7pm to 10pm. Traffic regulations were to be followed at all times with a special emphasis on not ignoring red lights, since red light running individuals can distort the data and derived results. Extreme weather situations like strong winds or heavy rain were to be avoided. Cycling for more than one hour at once was forbidden to avoid exhaustion.

Different bicycles (the cyclists’ own bicycles, bicycles of the Viennese bike-sharing program and e-bikes) were used for two reasons: Firstly to get more diverse arrival times at traffic signals for each person and secondly to analyze the effect of different bicycles in user profiles.
Ground Truth for Traffic Signals

For all traffic signals along the four test routes we requested the official signal plans from the municipal administration of Vienna (Magistratsabteilung 33). We received a pdf file for each traffic signal and manually extracted offset, green- and red-light durations for cyclists, and a Boolean flag if the green-light parameters change due to public transport priority. The cycle length for all traffic lights in the main arterial roads is known to be set to 100 seconds.

Preprocessing

Preprocessing of the acquired GPS data consisted of the following steps:

1. Cleaning (remove "tumbleweed" and smooth GPS inaccuracies)
2. Splitting (split tracks at the defined trip start and end points)
3. Joining split tracks with logbook entries.

The first step in our algorithm is data cleaning. First we delete all points that are clearly outside the study area by defining a rectangle around it. Second, single points that can only be reached with an unrealistic speed (50 m/s) are deleted. When the trajectory drifts away from the original path, it often ends up jumping back to the right position once the reception improves. We delete the point before the unrealistic jump and recalculate the speeds between the new pair of points. This is repeated until the remaining speeds are all below the threshold. To remove some of the random errors from the GPS trajectories, the trajectories are smoothed using cubic spline methods.

The second preprocessing step is to cut trajectories into pieces between longer stops and breaks. First, tracks were cut at the two corners of the route. For that the cyclists were asked to take small detours away from the planned route into a designated area (circles in FIGURE 2). Whenever the GPS track enters the zone, the tracks are cut. The next trajectory starts when the track leaves the zone again. Furthermore, stops were detected following an algorithm of (9). The algorithm checks at a point \( pt1 \) if the GPS trajectory does not leave a circle with radius 30 meters for at least 5 minutes. If this is the case, \( pt1 \) is the last point of the current trip. The next point \( pt2 \) outside the circle is the start of the new trajectory.

Finally unlikely points are deleted and replaced. In this preprocessing step loops in the GPS tracks are replaced. Loops and unlikely parts often happen at stops due to the normally distributed error in the GPS data. This leads to accumulations of GPS points ("tumbleweed"). FIGURE 4 shows an example of the cleaning and "tumbleweed" removal.

FIGURE 4 Trajectories after cleaning and splitting process.
To find these, it is checked if the direct distance between two points divided by the time it takes to get from the first to the second point along the trajectory is less than 0.5 m/s. If pairs of points are found that are in this category all the points in between are deleted and replaced by the midpoint between the first and last point of the loop. This is done starting at the pair with the lowest fraction of distance and time.

Cyclists’ Profiles

It is an intention of the presented research to provide information about the waiting at the intersections while also determining important factors regarding the cycling behavior of individuals under varying conditions. Velocity profiles are a central factor in the prediction of probable arrival times at signal controlled intersections along the remaining route. This section provides some details on influencing parameters regarding those velocity profiles.

Since constant time interval logging of GPS trackers results in a large amount of positions recorded at low velocities due to slow acceleration, essential information regarding a cyclist’s velocity range and behavior is contained in the velocity per distance traveled. A useful variable for our purposes is obtained by resampling the uniformly time-distributed measurements to measurements uniformly distributed along the traveled path. To extract distributions of desired traveling segments with speeds below 2 m/s are excluded. This parameterization is more symmetric around a mean with reasonable variance. Figure 5 shows exemplary distributions of this variable.

![FIGURE 5 Distributions of measured resampled data and respective densities.](image)

In the case of sampled cycling data the three most influential discrimination parameters for profiling a cyclist were identified by visual analysis of distributions:

- extreme weather conditions (strong winds, precipitation) : yes/no
- the bicycle type being used : 7 values
- the cyclist's personal driving characteristics : 20 cyclists.

Because consecutive speed measurements along a route are correlated, the estimation of the subsample variances is realized using block bootstrapping. After picking a random track for a fixed set of parameters, a block of speed values is chosen from this track and its mean value is determined. A variance is calculated from the means successively.

Although the distributions and thus the cycling profiles cannot be cleanly distinguished in the stronger sense of a statistical test on their dissimilarity, a t-test is applied to yield a quantitative description. Figure 6 shows the results of the discrimination analysis. Although the same weather allows most combinations, its inter-quartile distance is comparable to that of the same cyclist.
FIGURE 6 Dissimilarity measure under different parameter restrictions. The left shows a box plot where only one parameter is fixed, while the others are t-tested in their combinations. In the right plot all parameters but one are fixed, again comparing combinations.

The strongest expression of profile dissimilarities results from the variation of the bicycle type used for traveling which seems to dominantly influence the speed distributions. A simple 2 value characterization of the weather shows the least discriminatory power. This result indicates the necessity for precisely time aware routing applications to especially consider the bicycle type besides the individual cyclist’s behavior. Since the cyclists that took part in the data acquisition are frequent riders the effects on individual cycling profiles for data comprising more diverse fitness conditions have to be analyzed in more detail.

APPLICATION OF THE SIGNAL PATTERN FINDER

In the following section, the introduced algorithms are applied to the cleaned GPS tracks of the bidirectional course. The cycling course directions are named path 1 and path 2 respectively.

Results On Cycle Length Identification

Previously described methods to obtain $T^c_*$ are applied to all signals at the intersections along the paths. After filtering about 100 relevant tracks on each path remain.

Threshold parameters are chosen such that:

\[ v^{th} = 1.25 \frac{m}{s} \]

equals slow walking speed, which turned out to yield consistent results.

\[ d^{th} = 5 - 25 m \]

decreases with the cumulated number of seconds spent waiting $\Sigma h$. It is chosen to result in a total $\Sigma h \geq 1500$.

\[ h^{th} = \max(h) / 8 \]

For the chosen locations the bin width $T_b$ was defined as 5 seconds. Cycle lengths in road signals usually are multiples of 5 seconds although GPS trackers can provide higher sampling rates. For our purposes the stop positions at the signals were manually located on the map.
Path 1 runs from south to north and passes signals V06006, V07027 ... V08015, V01055 (see FIGURE 2, left blue course). Path 2 runs from north to south and passes signals V01055, V08015 ... V07027, V06006 (see FIGURE 2, left blue course). The analyses to identify the cycle length

FIGURE 7 shows the combined $R_g \times \sum h$ where $R_g = T_g / T_c$ for all the relevant crossings on this route for cycle lengths from 60 to 120 seconds in $T^b$ steps. The green squares show results at the correct cycle length. At the signals V07027, V07024 and V08004 a lack of data shows $R_g > 0$ for false $T_c$. The cumulative number of seconds spent waiting ranges from 50 to 5000. Correct cycle lengths are found at $T^c = 100$ seconds except for intersections V01055 ($T^c = 75$seconds) and V6006, where no pattern could be identified. Intersection V06006 lies very close to the start of the test course, therefore it includes GPS noise from departure or arrival movement. Intersection V07027 has a fixed signal program - yet red is only triggered by pedestrian demand (push button).

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure7.png}
\caption{FIGURE 7 Hilton plot showing the combined goodness of fit of identified cycle length on path 1 and path 2. The green squares indicate the "real" cycle length - except for one signal all cycle lengths could be correctly retrieved.}
\end{figure}

Considering the path follows the primary direction of an arterial road the maximum $R_g$ was found to be around 0.5. Comparing the results of paths 1 and 2 gives a second verification: We obtain a perfect match regarding identified $T^c$. The ground truth data described in the chapter *data acquisition* assumingly validates the estimates of the cycle lengths on the course along both paths. The confidence of $T^c$ is strictly determined by the number of cyclists who have stopped at a specific signal. Having asserted valid cycle length estimates allows extracting further information.
Signal Green Time Estimation Results

Further signal pattern analysis is done using the histograms of cumulative waiting periods. All parameters were estimated using the data set of path 1 and path 2.

Path 1

Using a $T_{c^*} = 100$ (75) seconds and setting $f_r = 1\%$ for path 1 yields the histograms $h_{b^*}$ shown in FIGURE 8. The distinctive bulks of waiting time slots within the signal cycle can be seen at most waiting spots.

Assuming correct estimation of the signal cycle parameters the shape of $h_{b^*}$ does conform to a theoretically expectable single sawtooth shape for $[t^0 + T^g, t^0]$. Conformity increases with sample size, while for small samples the stochastic noise rather distorts the sawtooth shape.

Path 2
The cumulative waiting periods for path 2 are shown in FIGURE 9. Identical $T^{c*}$ and similar deviations as for path 1 are visible.

FIGURE 9 Results signals 1 to 10 direction "North - South ". The solid vertical lines represent the models result, the dashed are the "known" times. $t^{o*}$ in GREEN is indicating the start of a green phase. The RED $T^{b*} + t^{o*}$ indicate the begin of RED (YELLOW) phases.
Discussion

To summarize the findings, Table 1 provides the errors of the estimation and an attempt at interpreting them. The green phase's start can be estimated more precisely than that of the red phase. The results show multiple aspects which might influence estimation quality:

- Number of stopped and waiting cyclists
- Distance to neighboring intersections
- Distance to potential trip destination/origin
- Duration $T^g_*$ of signals at neighboring intersections.

The number of GPS tracks is a critical factor to yield exact parameters for a pattern. Tests carried out while using only half of the available data set resulted in doubled average errors on the estimated quantities. Simulation of 100 bicycle tracks on path 1 led to more accurate results with average errors below 2 seconds for each time parameter. Under non-perfect conditions, more data is required to achieve this accuracy.

<table>
<thead>
<tr>
<th>Junction Nr.</th>
<th>$T^g_*$ count</th>
<th>error $T^g_*$</th>
<th>error $T^g_*$</th>
<th>$T_c^*$ count</th>
<th>error $T_c^*$</th>
<th>error $T_c^*$</th>
</tr>
</thead>
<tbody>
<tr>
<td>V01055</td>
<td>75</td>
<td>986</td>
<td>0</td>
<td>999</td>
<td>3</td>
<td>-3</td>
</tr>
<tr>
<td>V06006</td>
<td>-</td>
<td>4285</td>
<td>2</td>
<td>1662</td>
<td>1</td>
<td>22</td>
</tr>
<tr>
<td>V07023</td>
<td>100</td>
<td>1550</td>
<td>1</td>
<td>1532</td>
<td>1</td>
<td>-6</td>
</tr>
<tr>
<td>V07024</td>
<td>75</td>
<td>104</td>
<td>1</td>
<td>90</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>V07027</td>
<td>100</td>
<td>24</td>
<td>-21</td>
<td>607</td>
<td>1</td>
<td>-47</td>
</tr>
<tr>
<td>V07032</td>
<td>100</td>
<td>1669</td>
<td>0</td>
<td>317</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>V08003</td>
<td>100</td>
<td>1606</td>
<td>0</td>
<td>866</td>
<td>1</td>
<td>38</td>
</tr>
<tr>
<td>V08004</td>
<td>100</td>
<td>218</td>
<td>2</td>
<td>79</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>V08005</td>
<td>100</td>
<td>529</td>
<td>2</td>
<td>811</td>
<td>0</td>
<td>-1</td>
</tr>
<tr>
<td>V08015</td>
<td>100</td>
<td>88</td>
<td>3</td>
<td>1562</td>
<td>1</td>
<td>-5</td>
</tr>
<tr>
<td>Average (err.&lt;8)</td>
<td>597</td>
<td>4.3</td>
<td>4.9</td>
<td>714</td>
<td>1.0</td>
<td>10.1</td>
</tr>
</tbody>
</table>

Table 1: Estimations and respective errors with a short discussion. $count$ represents the cumulative number of seconds spent waiting at the specified signal.
Results indicate that adaptive traffic control which does not alter the $T^c_{gy}$ of an intersection barely distorts the estimations. The error at signals with adaptive control is not significantly larger. It has to be remarked though, that the number of signals in this study is quite small. Another specific strength of the applied methods is the fact that the estimation of the red phase's start is not influenced by red light runners.

Low GPS noise (wide roads) and a high sample rate is a precondition for suitable results. In this study the quality of the GPS signals was no issue. The cyclists had been carefully instructed where to place the receivers and the devices used are more accurate than standard smartphones. Cutting out samples with a positional dilution of precision $pdop > 4 \text{ m}$ triples the error. Every track is needed to improve the estimation - even those with low positional accuracy.

The impact of the single 75 seconds $T^c_e$ between the majority of 100 seconds cycle lengths might be disturbing a green wave but nevertheless has a positive effect on the gathered data. Just from the point of view of pattern reconstruction there is a benefit of both longer waiting times and a heterogeneously approaching of cyclists during the whole cycle length. Different $T^c_e$ or other disturbances of green waves result in a more uniform distribution of arrival times at following signals.

**CONCLUSIONS**

Methods to derive road signal patterns from analysis of cycling GPS tracks were introduced. Filtering based on location and speed was applied to identify the periods when cyclists have to wait at red signals. A signal control paradigm based on fixed time frames was assumed – i.e. at similar times of day the same program phase is active. Based on this assumption the signal patterns of road signals were obtained. The methods were applied to the GPS tracks of 20 cyclists in Vienna (~2x10^6 seconds logged cycling). The actual control program was digitized and used to compare the results. Additionally, cycling behavior and its dependence on several influencing parameters was analyzed to obtain an overview of required information for behavioral profiling in further applications.

This paper demonstrates the feasibility and potential of the introduced algorithms even for limited GPS track sample sizes. The switch from red to green phase can be estimated very accurately; estimation of the end of the green phase is less reliable. Given a sufficient number of GPS tracks a estimation precision below 5 seconds is achieveable. Possible applications comprise routing, navigation or network analyses in transportation planning.

Future research should focus on the topological variety of intersections and their consequences as well as the impact of dynamic traffic control. For the presented work all stopping positions have been manually extracted using GIS tools. Fully automated systems would require some kind of clustering strategy in order to identify stopping positions at intersections. Application of the presented approach to real crowd data with cycling behavior characterization is an ongoing effort to verify the impact of red runners, varying GPS noise and diverse weather conditions.

The presented work demonstrates the methods’ application to cycling data. Because urban bicycle traffic flow is not very sensitive to demand factors the implementation at hand comes across favorable preconditions. In principle, other modes of transport with similar tracking information could be used for further applications.
ACKNOWLEDGEMENTS

The authors thankfully acknowledge Dietmar Bauer for his valuable contribution and support. The thanks are extended to Seraphim Alvanides and Godwin Yeboah from the Faculty of Engineering and Environment at the Northumbria University for organizing the pre-AGILE Conference Workshop Understanding Urban Cycling: A Data Challenge (CDC2013). It provided the first forum for discussing the application of the algorithms introduced in this paper. The work for this paper is part of the project BikeWave, supported by the Austrian Ministry for traffic, innovation and technology (BMVIT) within the national funding framework IV2Splus under contract 835761.

REFERENCES