Joint Estimation of Signal Timings for all Directions of an Intersection Using Low-Frequency GPS Data

Xin QI, Jin ZHOU, Hai JIANG*

Department of Industrial Engineering, Tsinghua University, Beijing 100084, China, haijiang@tsinghua.edu.cn

ABSTRACT

This paper studies the signal timing estimation problem for intersections using sparse GPS data. Judging the signal status by the GPS points after the stop line, we first apply Fast Fourier Transform to obtain the cycle length in the frequency domain. Then, we put all GPS data in one cycle length to deal with the shortage of GPS data. Considering the conflicts of signal phases, we jointly minimize the total error to get the actual signal timing of the intersection. This problem is formulated as a Binary-Mix-Integer-Linear-Program, which can be solved efficiently by the standard branch-and-bound method. Finally, the effectiveness and the robustness of our approach are validated by the numerical experiments.

Keywords: cycle length, signal timing, GPS, BMILP

INTRODUCTION

Accurate signal timing information, that is, the length of the cycle and the start and duration of green times for the different traffic movements through the intersection, can significantly improve the efficiency of the transportation system. Such information can be used to develop speed advisory algorithms that guide drivers or autonomous vehicles through traffic signals without unnecessary braking and acceleration, therefore improve fuel efficiency (Koonce et al. 2008, Osorio and Nanduri 2015). Moreover, this kind of information also serves as crucial input to the calculation of steady-state signal performance measures such as intersection delay and queue lengths (TRB 2010). However, as is pointed out by Hao et al. (2012), it is not only challenging but also costly to collect signal timing information from traffic signal controllers managed by multiple agencies in a wide area such as a region or nationwide.

Recently, researchers have started to tap into GPS data reported by probe vehicles, such as taxicabs or buses, to estimate signal timings at an intersection. When the reporting frequency of the GPS data is high, for example, 1 Hz, the trajectories of probe vehicles can be fully constructed. Cheng et al. (2010) leverage such information to detect *critical points* in vehicle trajectories and apply shockwave analysis to recover signal timings. Kerper et al. (2012) extract the green and red states of the traffic light from vehicle trajectories and develop a nonlinear optimization model involving modulo operations to estimate signal timing parameters. Hao et al. (2012) calculate intersection travel times between upstream and

downstream *Virtual Trip Line* locations using vehicle trajectories. They then apply the Support Vector Machine (SVM) technique to estimate the cycle start and end times and use methods developed in Ban et al. (2009) and Ban, Hao, and Sun (2011) to estimate the effective red (or green) times. When the reporting frequency of GPS data is low, for example, about 30 seconds per record), Fayazi et al. (2015) consider the case the influence of queue is negligible and estimate signal timing parameters by reconstructing vehicle kinematics. Fayazi and Vahidi (2016) extend this method to the case when the influence of queue is explicitly accounted for.

One of the major limitation of existing studies is that they only estimate signal timings for a specific traffic direction of the intersection. As a result, those independently estimated start and duration of green times from different directions may conflict with each other. To address the above challenge, we propose to jointly estimate the signal timings on all directions of an intersection. We develop a binary mixed integer linear programming (BMILP) model that exploits the incompatibility among the different traffic movements at the intersection. This not only ensures the feasibility of the estimated signal timings but also allows us to utilize GPS data from directions with heavy traffic (that is, more GPS data) to improve the estimation accuracy for directions with light traffic (that is, little GPS data). In our approach, we first estimate the instants when the traffic light is green. We then estimate the cycle length using Fast Fourier Transform, which is more straight-forward and computationally efficient than methods proposed in existing literature. We proceed to develop a *folding technique* to collapse events over multiple cycles to a single cycle to virtually increase the number of green instants in a cycle. These green instants serve as input to our BMILP model, which outputs the signal timings for all directions simultaneously. We validate our model using real taxi GPS data from Beijing. We choose three cross-intersections to estimate their cycle lengths and signal timings. All the results of cycle lengths are accurate to seconds, and all the errors of signal phase are above 97%.

The remainder of this paper is organized as follows. Section 2 presents how we figure out the cycle of intersections. We build up model to obtain the duration of green lights and formulate problems as Binary-Mix-Integer-Linear-Programs (BMILP) to figure out signal timing information, respectively. Numerical experiments are conducted in Section 3 to verify the effectiveness of our method. Finally, we make the conclusion and come up with future research directions in Section 4.

MODELING APPROACH

The input to our model is the GPS data reported by probe vehicles, or taxicabs, that originate from all the inbound arms and head toward all the outbound arms in the same period of time. Our main objective is to figure out signal timing information of an intersection, which includes its cycle length, starts and durations of green lights, etc. The data we use is collected from in-car GPS.

Model Input

Each probe vehicle reports at regular intervals, for example, every 30 seconds. Its GPS data includes vehicle ID, the current time, its location in terms of longitude and latitude, travel speed, and direction of travel. Consider the four-way intersection shown in Figure 1, where the signal operates with a fixed signal timing plan. Three trajectories are shown in this Figure. They are labeled as vehicles 1, 2, and 3 respectively. The dots on the trajectories corresponding to the locations where they report their GPS data. Note that these three vehicles may traverse the intersection at different times.

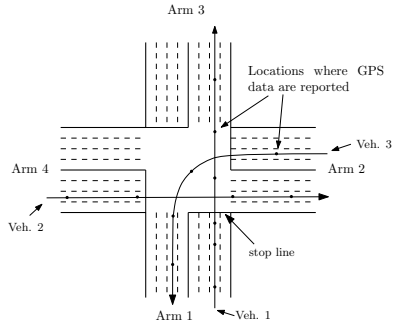


Figure 1: GPS data reported by probe vehicles traversing an intersection.

Determining Instants When The Traffic Light Is In Green

If we consider a series of GPS points reported by the same vehicle one by one, each observation point is associated with a velocity and a set of longitude and latitude. According to this information, we can calculate the distances between locations of the vehicle and the stop-line and further reconstructing the trajectory of a car passing the intersection. Considering the performance of vehicles when approaching the intersection, we define two types. The first kind of cars (denoted as *Type-A*) crosses the stop line without delay or stop cause the signal is green and there are no left vehicles in the queue from the last cycle. It's with high probability that this type of vehicles will keep a relatively stable velocity or even accelerate from a high speed during the time when they pass the intersection. The other type of vehicles (denoted as *Type-B*) are those who delay either due to the red light or remaining queue. Their time-velocity graphs tend to follow a pattern of an acceleration after a deceleration to stop or a small velocity. Within *Type-B* vehicles, there are 3 subtypes we need to consider. Cars who stopped before the stop-line from a relative long distance may have completed acceleration when they are passing the stop-line, so they behave just like those *Type-A* vehilces. The other kind of vehicle is in acceleration state when passing the stop-line and arrives free flow speed before its first GPS is recorded after the stop-line. There is another kind of abnormal cars who is still in very low speed even after passing the stop-line. We abandon the third type with a filter.

Based on the above classification, we detect the time instant for passing the stop-line of each vehicle using the velocity and distance information. For the

reason that a vehicle can only pass the stop-line when the intersection signal state is green, we can get the estimated green time in other words. In Figure 2, we use a time-space diagram to analyze the behavior of one vehicle at the intersection. (t_2, v_2) is the first GPS point recorded after the stop-line. Denote the acceleration rate as a and the distance between the stop-line and the first GPS point after the stop-line as d. If $v_2^2 - 2ad >= 0$, that is the distance a vehicle can go if it starts acceleration from 0 to v_2 with acceleration rate a is larger than d, then we think the car is *Type-B* car. By solving a function of time and velocity we get the time when the car passing the stop-line: $t_{pass} = t_2 - \frac{v_2 - \sqrt{v_2^2 - 2ad}}{a}$. Otherwise, the car is treated as *Type-A* who has achieved acceleration before the stop-line. In this case, the green instant can be calculated from $t_{pass} = t_2 - d/v_2$ as t_A in Figure 2.

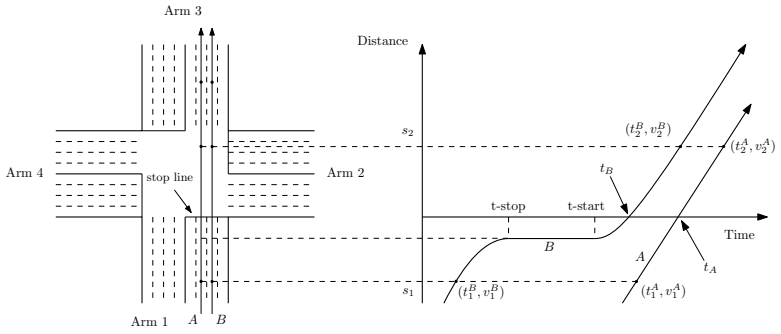


Figure 2: Time-space diagram of two kinds of vehicles at the intersection.

However, the actual data of a single date do not have sufficient vehicles to get so many observations of the green instants, so we need to combine data of parallel days. According to Time-of-Day rule in normal signal timing plan, parallel days have the same signal timing, so we can regard the vehicle data points as in the same day. In our practice, data of about 30 days is enough to achieve the rest of processing. After this processing, we can get a series of t_i, where t_i is an estimated green instant of vehicle i.

An FFT-based Technique To Estimate Cycle Length

In this subsection, we will process the estimated green time instant data. In Figure 3, each bar is the number of vehicles passing the stop line every 5 seconds from 9:00 to 11:00 in the morning. We can see from the graph that the occurrence of green intervals is periodic. Moreover, the data can be transformed into a sequence of uniformly sampled square wave by applying indicator function whose output is 1 if the input count of vehicles is greater than 0. What Fourier transform does is decomposing a function of time (a signal) into frequencies that make the signal up. In this way, the transformation result from our green light instants is supposed to have apparent high amplitudes in frequencies composing the signal. In light of the discreteness of our data as well as the fact that only locations of peak amplitudes need to be known, Discrete Fourier Transform is the justified method. In this paper, we follow the notation in Mitra and Kuo (2006). Suppose the discrete signal is like $x[n]$, $n \in [0, N-1]$, where N is the total number

of discretized seconds in a specific range we care about. $x[t] = 1$ if there is at least one car passing the stop line at time t and $x[t] = 0$ otherwise. DFT can be written as:

$$X[k] = \sum_{n=0}^{N-1} x[n] e^{-j\frac{2\pi kn}{N}}$$

, where $k = 0, 1, ..., N-1$. $X[k]$ is the power or the amplitude of frequency component k. It's worth noting that the index k corresponds to angular frequency $\omega_k = \frac{2\pi k}{N}$ or frequency $f_k = \frac{k}{N}$ and $j = \sqrt{-1}$.

In practice, DFT is calculated by a method called Fast Fourier Transform. FFT is just a set of quicker implementation of DFT computation. It exploits the symmetry in DFT equation, and achieves a computation complexity of $O(NlogN)$ by divide and conquer compared with that of $O(N^2)$ in direct DFT. By applying FFT, the estimated green time instant signal in a finite time period is transformed into a function of frequency with several peaks in certain points, which is shown in Figure 3 and 4. Those peak values, also called sinusoidal components, in Figure 4 come in multiples and are descending in amplitude as frequency increases. This phenomenon is called harmonic and is also a property inherent in our signal. By knowledge from signal processing, the first fundamental frequency, which frequencies of other sinusoidal components are integral multiples of, corresponds to the cycle length of the original signal in time domain. Other peak values with smaller power in Figure 4 are harmonic to the first one.

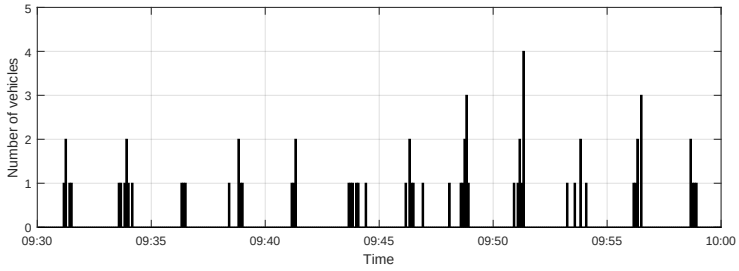


Figure 3: Sampled green time instants between 9:00 and 11:00.

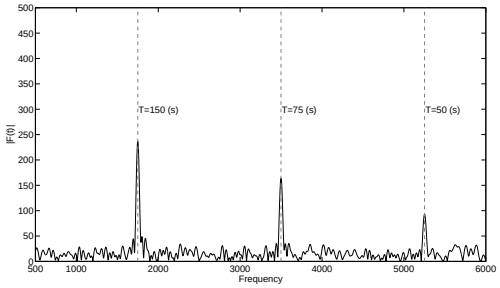


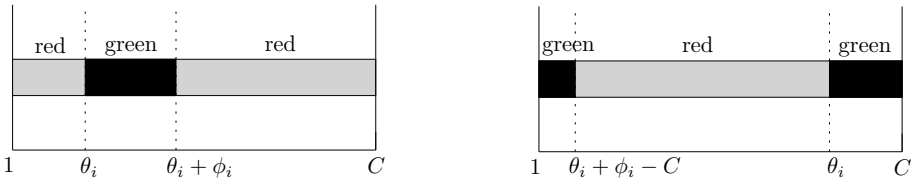
Figure 4: FFT result of sampled green light time between 9:00 and 11:00.

In reality, the signal timing scheme is not always the same among different period, because it needs to be adjusted according to traffic demand which varies through time. Using FFT method introduced earlier, the issue of detecting cycle-length-change can be solved. This is further discussed in Case Studies.

The BMILP Model

We now estimate the duration of the green lights based on the estimated cycle length in this subsection. Unlike traditional method (like Fayazi et al. (2015)) which estimate the signal timing plan of single direction, we formulate a Binary-Mix-Integer-Linear-Program to characterize the duration of the green lights for all directions in the intersection at the same time considering the possible conflicts among all directions.

Let I be the set of all possible directions in the intersection, C be the cycle length measured in seconds. We discrete the entire cycle into C intervals with a duration of 1 second. For a given direction i, let $\theta_i \in \{1, 2, \cdots, C\}$ be the start of green and $\phi_i \in \{1, 2, \cdots, C\}$ be the duration of green. Since the signal operates cyclically, depending on whether $\theta_i + \phi_i$ exceeds C, the green time for direction i can take two possible types as are illustrated in Figure 5. Figure 5(a) shows the first type, where $\theta_i + \phi_i \leq C$. This means that the signal is green in interval $k \in [\theta_i, \theta_i + \phi_i)$, or $k \in [\theta_i, \theta_i + \phi_i - 1]$. Figure 5(b) shows the second type, where $\theta_i + \phi_i > C$. This means that the signal is green in interval $k \in [1, \theta_i + \phi_i - C) \cup [\theta_i, C]$, or $[1, \theta_i + \phi_i - C - 1] \cup [\theta_i, C]$.



(a) Type 1: $\theta_i + \phi_i \leq C$, which means that the signal is green in interval $[\theta_i, \theta_i + \phi_i)$.

(b) Type 2: $\theta_i + \phi_i > C$, which means that the signal is green in interval $[0, \theta_i + \phi_i - C) \cup [\theta_i, C)$.

Figure 5: Two possible types of signal timing for direction i. The entire cycle is descritized into C intervals and the values on the $x-$axis indicate the start of the corresponding intervals.

For a given direction $i \in I$, let N_i be the number of green instants observed for direction i and N_{ik} be the number of green instants observed during time interval k. That is, $\sum_{k=1}^{C} N_{ik} = N_i$. Let binary variable s_{ik} be 1 if the signal for direction i is green at instant k; and 0, otherwise.

Our objective is to maximize the fraction of vehicles that traverse the intersection when the traffic light is green:

$$\max \quad \sum_{i \in I} \sum_{k=1}^{C} \frac{N_{ik}}{N_i} \cdot s_{ik} \tag{1}$$

Status of the signal at given time instant

We now develop constraints that ensure $s_{ik} = 1$ if the signal for direction i is green during interval k; and $s_{ik} = 0$, otherwise. This is achieved by Constraints (2) through (11) as follows, where x_i, y_{ik}, z_{ik} are binary variables, and M is a sufficiently large positive integer:

$$\theta_i + \phi_i \leq C + Mx_i, \qquad \forall i \in I \quad (2)$$

$$k \geq -x_i M + \theta_i - (1 - s_{ik})M, \qquad \forall i \in I, k \in \{1, 2, \cdots, C\} \quad (3)$$

$$k \leq \theta_i + \phi_i - 1 + (1 - s_{ik})M + x_i M, \qquad \forall i \in I, k \in \{1, 2, \cdots, C\} \quad (4)$$

$$k \leq \theta_i - 1 + s_{ik}M + y_{ik}M + x_i M, \qquad \forall i \in I, k \in \{1, 2, \cdots, C\} \quad (5)$$

$$k \geq \theta_i + \phi_i - s_{ik}M - (1 - y_{ik})M - x_i M, \qquad \forall i \in I, k \in \{1, 2, \cdots, C\} \quad (6)$$

$$\theta_i + \phi_i \geq C - (1 - x_i)M, \qquad \forall i \in I \quad (7)$$

$$k \geq \theta_i + \phi_i - C - (1 - x_i)M - s_{ik}M, \qquad \forall i \in I, k \in \{1, 2, \cdots, C\} \quad (8)$$

$$k \leq \theta_i - 1 + s_{ik}M + (1 - x_i)M, \qquad \forall i \in I, k \in \{1, 2, \cdots, C\} \quad (9)$$

$$k \leq \theta_i + \phi_i - C - 1 + (1 - s_{ik})M + z_{ik}M + (1 - x_i)M, \quad \forall i \in I, k \in \{1, 2, \cdots, C\} \quad (10)$$

$$k \geq \theta_i - (1 - s_{ik})M - (1 - z_{ik})M - (1 - x_i)M, \qquad \forall i \in I, k \in \{1, 2, \cdots, C\} \quad (11)$$

Order of signal displays for incompatible movements

When two movements are incompatible of each other, that is, if the vehicle trajectories of these two movements intersect at the intersection, we need to ensure that their green times do not overlap for safety purposes. Let Ψ be the set of incompatible movements, that is, $\Psi = \{(i, j) | \text{movements } i \text{ and } j \text{ are incompatible}; i < j; i, j \in I\}$. We introduce binary variable Ω_{ij} to indicate the order of signal displays for movement $(i, j) \in \Psi$. $\Omega_{ij} = 0$ if the start of green for movement i precedes that of movement j, that is, $\theta_i \leq \theta_j$; and $\Omega_{ij} = 1$ if the opposite is true. The order of signal displays for incompatible movements is enforced by the following constraints:

$$\theta_i + \phi_i \leq \theta_j + \Omega_{ij}M, \qquad \forall (i, j) \in \Psi \quad (12)$$

$$\theta_j + \phi_j - C \leq \theta_i + \Omega_{ij}M, \qquad \forall (i, j) \in \Psi \quad (13)$$

$$\theta_j + \phi_j \leq \theta_i + (1 - \Omega_{ij})M, \qquad \forall (i, j) \in \Psi \quad (14)$$

$$\theta_i + \phi_i - C \leq \theta_j + (1 - \Omega_{ij})M, \qquad \forall (i, j) \in \Psi \quad (15)$$

Note that when $\Omega_{ij} = 0$, Constraints (14) and (15) are redundant; and when $\Omega_{ij} = 1$, Constraints (12) and (13) are redundant.

Other constraints on the variables

The following constraints are for the decision variables:

$$\theta_i, \phi_i \in \{1, 2, \cdots, C\}, \qquad \forall i \in I \tag{16}$$

$$s_{ik} \in \{0, 1\}, \qquad \forall i \in I, k \in \{1, 2, \cdots, C\} \tag{17}$$

$$\Omega_{ij} \in \{0, 1\}, \qquad \forall (i, j) \in \Psi \tag{18}$$

$$x_i, y_i, z_i \in \{0, 1\}, \qquad \forall i \in I \tag{19}$$

CASE STUDIES

In this section, we conduct a series of numerical experiments on three real intersections based on GPS data collected in September 2012, Beijing, China. The numerical experiments can be divided into three parts. First, we give our data source a brief description in Section , and then green intervals of each direction is estimated using FFT in Section . Lastly, we solve the corresponding BMILP model based on obtained cycle lengths, and report final results.

Input

GPS data we use in this paper is provided by AutoNavi and is gathered from float vehicles mounted with GPS devices. Covered area and time are restricted within Zhongguancun Area from in the morning. Three four-arm intersections we choose are crossings between Zhichun Road and Zhongguancun Street, Chengfu Road and Zhongguancun East Road, as well as Zhongguancun South Road and Zhongguancun Street. Later in the passage, we refer them as Intersection 1, 2 and 3 in sequence. We compare our experimental results with true values, which are gathered through field investigation in those specific locations.

We aggregate the input in any intersection monthly and divide traffic flows into eight movements. We only use data in weekdays in case that signal schedule is different from that of weekends. In grasping an understanding of traffic flow in those three intersections, we select a typical non-peak hour, 10:00 to 11:00, and count the number of unique vehicle ID in separate movements during this time in Table 1. We can tell from the percentage of cars in each direction that there is variance in both traffic demand and flow capacity in directions. Take Intersection 1 as an example, the number of cars moving from North to East is significantly less than that of cars moving from South to North. This imbalance among movements is a characteristic of our data feed, and it's still true even put into most other intersections within the city. It has to be stated that applying methods brought up by published papers can be especially difficult in those directions with sparse flow, and this is exactly where direction-joint comes into play.

Estimation Results Of Cycle Lengths

We set the acceleration rate of a car accelerating from the stop as a stable value 1.44 m/s, which is widely applied in existing literature like (Long 2000). After calculating the time that each vehicle passing the stopp-line, given the first GPS point reported after the stop line, we get a bunch of those passing times which

Mvt.	Intersection 1		Intersection 2		Intersection 3	
	No. of Car	Pct.(%)	No. of Car	Pct.(%)	No. of Car	Pct.(%)
EW	247	17.66	123	12.68	16	1.84
ES	88	6.29	126	12.99	93	10.69
NS	340	24.30	119	12.27	408	46.90
NE	148	10.58	51	5.26	50	5.75
SN	189	13.51	157	16.19	229	26.32
SW	161	11.51	197	20.31	56	6.44
WE	183	13.08	150	15.46	10	1.15
WN	43	3.07	47	4.85	8	0.92
Total	1399	100.00	970	100.00	870	100.00

Table 1: Counts of Unique ID in 3 intersections of all directions in 10:00-11:00.

are discrete samples of green intervals. Here for example, the length of sampled time series during 9:00 and 11:00 is 7200s and our signal is a discrete function of time. According to basics of FFT and the necessary accuracy of our problem, we divide our sample length that is 1 second into 2^N (here we use $N = 18$) samples of equal interval, and then conduct FFT. The transformation results of sampled green interval of Intersection 2 during 9:00 and 11:00 are shown in Figure 4 in the previous section. Here, in this case, the frequency with the most power is 1750Hz and the corresponding cycle length of this signal is 150s, by formula $T = 2^N/f$.

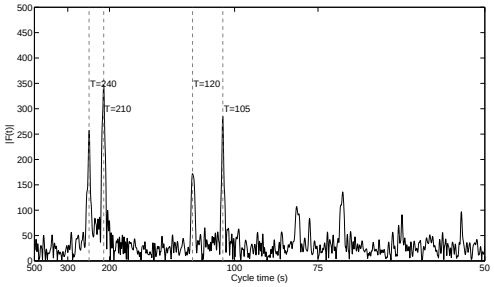
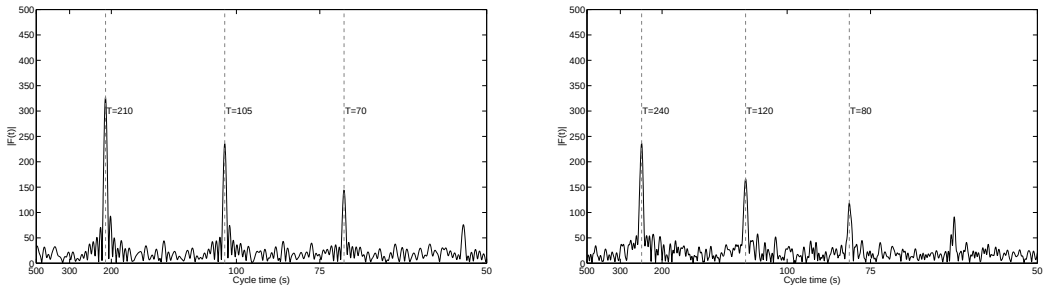


Figure 6: Results of cycle length estimation in Intersection 1, results from FFT during 7:00-11:00.



(a) Results from FFT during 9:00-11:00.

(b) Results from FFT during 7:00-9:00.

Figure 7: Results of cycle length estimation in Intersection 1.

In Figure 7, we reported three pictures as a result of transformed sampled passing times from time domain to frequency domain by FFT in Intersection 1.

No.	Period	Observed (s)	Raw Output		Rounded Output	
			FFT (s)	Error (%)	Rounded (s)	Error (%)
Intersection 1	7:00- 9:00	240	240.28	0.12	240	0.00
	9:00-11:00	210	209.88	0.06	210	0.00
Intersection 2	7:00- 9:00	180	179.80	0.11	180	0.00
	9:00-11:00	150	149.71	0.19	150	0.00
Intersection 3	7:00- 9:00	180	179.67	0.18	180	0.00
	9:00-11:00	150	150.05	0.04	150	0.00
Average				0.12		0.00

Table 2: Estimated and observed cycle lengths in three intersections during 7:00-9:00 and 9:00-11:00.

To better relate the result of FFT with the estimation of cycle length, we have transformed the horizontal axis to the time corresponds to the frequency. Figure 6 shows two cycle lengths $T_1 = 210s, T_2 = 240s$ at the same time cause it's generated from data in whole morning. The cycle length estimation method proposed can easily detect the schedule change of cycle length from peak hour to non-peak hour by dichotomy. And we find the change of cycle length happens around 9 a.m. Figure 7(a) and 7(b) separately shows the power of a specific cycle length in those two periods, normally referred as peak hour and non-peak hour. Choosing first the cycle length with the most power can give us the cycle length during this time period. Complete results of cycle length estimation are gathered in Table 2. Comparing the raw output with observed truth, the error is lower than 0.2%. While we rounded the raw output into integers, the final estimation towards cycle lengths matches exactly to the observed cycle lengths.

Estimation Results Of Signal Timing

In this subsection, we will demonstrate our results of signal timing estimation based on the estimated cycle lengths of Intersection 1, 2 and 3. Because of the change of the signal schedule, we divide the data associated with each intersection into two parts, that is, data integrated during 7:00-9:00 and 9:00-11:00. By mod all those passing times in the set by the cycle length (T) we fold all reappearing cycles during this period in a single cycle. Then we get eight histograms, as shown in Figure 8, whose vertical axis telling the number of vehicles who passing the stop line in each second in a cycle. It shows our estimation of green intervals in eight movements of Intersection 1 during 9:00 and 11:00 by integrating data during a month. In a cycle, the signal state at a time slot when there are more vehicles passing the stop line is more likely to to be green. Looking at those histograms, one can not only get the schedule of traffic signals transformation, as shown up to down, but also movements in the same phase. Histograms of two directions in the same row, like (NS, SN) or (NE, SW), are green periods belong to the same phase and are discharged together. That's why their graphs look in the same pattern, have similar peak position and span of bars. It's easily noticed that directions in the same phase may have different height of bars, dramatically shown in Figures 8(e) and 8(f). This is caused by variation and imbalance in traffic demand from directions, and it further supports our direction-joint mixed-integer model by compensating the absence of vehicle flows in some movements. There are also some outliers whose heights are one or two and lies out the main green period.

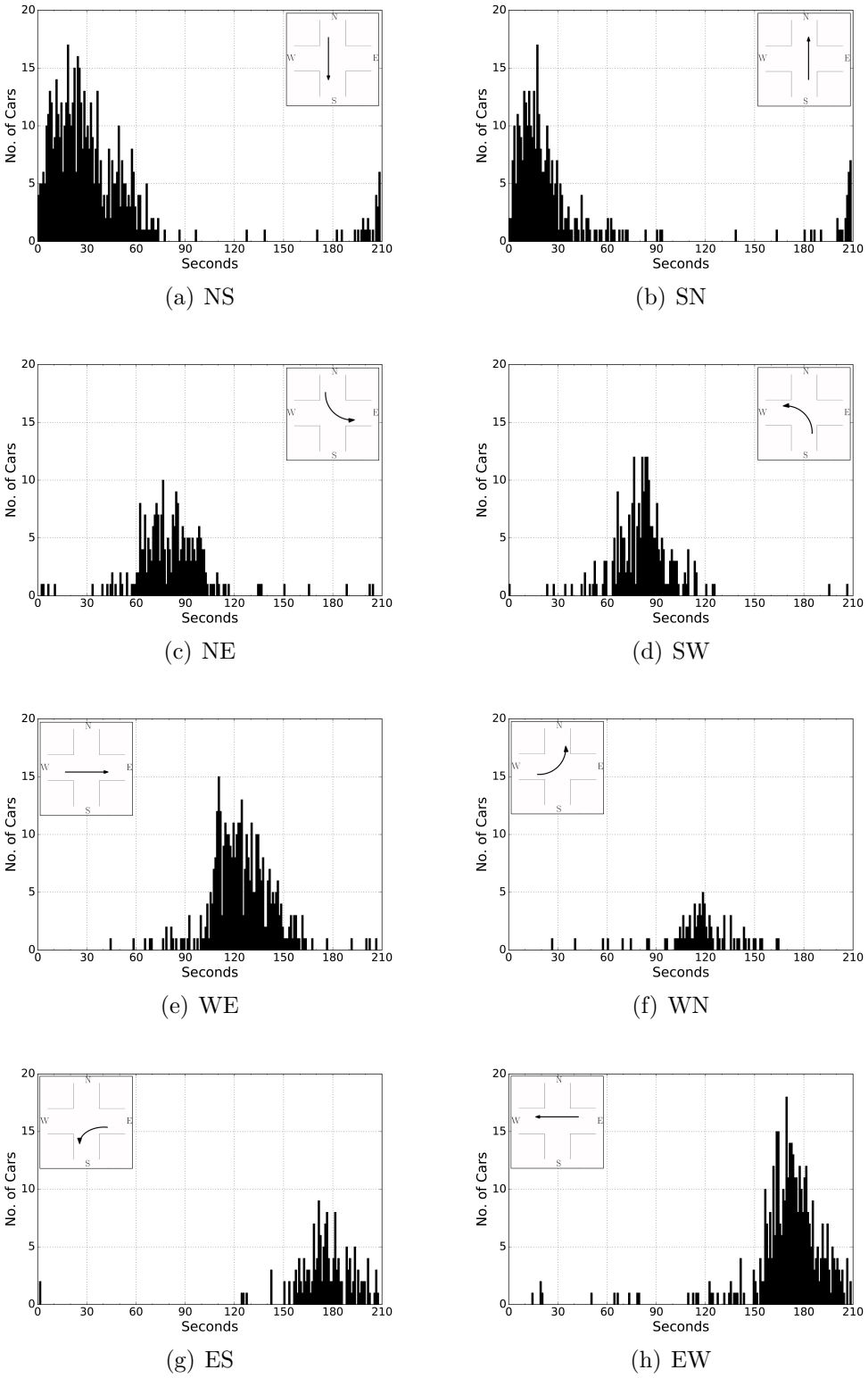
This results from irrational driving behaviors like violation of traffic regulation or error stem from the time-velocity model.

In order to compare the estimation results to the observed truth, some evaluate indices have been introduced. We adopt a criteria called *the Percentage of Correct State* (PSC) from Kerper et al. (2012). By signal state, it means if we think the signal of each second in a cycle as a state, which is either green or red, then there are T signal states in a cycle, where T is cycle length in seconds. Hence, the percentage of correct states is defined as the percentage of the number of correctly estimated states in a cycle. Besides, we also calculated the absolute error between true values and estimated green light start time (θ_e), green time duration (ϕ_e) and the time green light ends (η_e). The estimation of those parameters in eight directions can recover the whole signal timing plan in this direction.

As shown in Figure 1, a typical crossroad intersection has 4 arms and 8 phases and some phases may conflict with each other. We use the case of 4 arms here, and other sorts of pre-timed signal timing cases are certainly suitable. The result can be easily ported to other kinds of intersections. For notational brevity, we use N, S, E, W to represent Arm north, south, east, and west respectively. Then the direction characterized by two arms can be denoted as an acronym of two characters. For example, NE means the direction from north to east. Let I be the set of all possible directions in the typical 4-arm intersection, that is, I={NE, NS, EW, ES, SN, SW, WE, WN}. So for this case, we have S_c={[SW, NS], [SW, WN], [SW, EW], [SW, ES], [SW, WE], [NS, WN], [NS,EW], [NS, ES], [NS, WE], [WN, EW], [WN, NE], [WN, SN], [EW, NE], [EW, SN], [NE, SN], [NE, ES], [NE, WE], [SN, ES], [SN, WE], [ES, WE]}. Note that each direction has exactly 5 conflicting directions. Each straight line direction does not conflict with the left turn direction of the same upstream and the opposite straight line direction.

The optimization model we developed in Section 2.3 is used to minimize the overlapped green periods of all conflict phase pair. In other words, it gives a signal plan of eight directions which maximize the number of cars passing the stop line during the green period of this plan. We have reported our final results in six tables from Table 3 to Table 8. For example, Columns 2 to 4 in Table 3 presented the observed parameters including θ_o, ϕ_o, η_o in eight moving directions. Columns 5 to 7 give our estimation results regards these three parameters, while Columns 8 to 10 is the absolute error between observed truth and our estimation. The last Column is the index the Percentage of Correct State that we have introduced. In the last row of each table, we calculated the averaged absolute start time error, duration error, end time error, and percentage of correct state of eight directions.

In all our estimation results, even the intersection with the worst percentage of correct states is 97.33%. It has improved a lot compared to an average of 90% generated by simulation data in Kerper et al. (2012). In average, the difference between our estimation of the start of green and the true green start time is smaller than 2.5s. It's the same for the difference between the estimated and observed time when green ends. In summary, our model has a fairly high accuracy for all the testing intersections in both peak hours and non-peak hours and the obtained signal timing has no conflicts.



(a) NS (b) SN

(c) NE (d) SW

(e) WE (f) WN

(g) ES (h) EW

Figure 8: Estimated green intervals of eight directions in Intersection 1.

CONCLUSIONS

In this paper, we develop a model that estimates signal cycle length and jointly estimate phase timing using GPS data with low transmission frequency. The input is GPS data which has low penetration rate and low transmission frequency. The output is the entire signal timing plan of one intersection, including cycle length, start time and duration of each signal phase for every period of a day. Since the data of a single date is not enough, we implement a data fusion step for parallel days as data preprocessing.

In the estimation of cycle length, we use Fourier Transform to transform the "signal" in our data (observed GPS points) from time domain to frequency domain. Due to the periodicity of signal timing, we can get "peaks" in the frequency domain. These peaks correspond to the periods of our data, and we can find cycle length in these peaks. When there is more than one peak that could be the signal cycle, we decompose the time domain to analyze changes of these peaks. In this way, we can detect the change in cycle length between different periods, such as peak and non-peak hours.

In the estimation of signal timing, we use an optimization model to jointly estimate signal timing for all the phases. We put the conflicts of phases into consideration, and add these constraints in our optimization model. Therefore, the obtained signal timing plan is always feasible. We model this problem as a Binary-Mix-Integer-Linear-Programming and solve it by branch-and-bound. Results show that the degree of accuracy is fairly high. PCS for all the test intersections is above 97%.

References

Ban X, Herring R, Hao P, Bayen A, 2009 *Delay pattern estimation for signalized intersections using sampled travel times. Transportation Research Record: Journal of the Transportation Research Board* (2130):109–119.

Ban XJ, Hao P, Sun Z, 2011 *Real time queue length estimation for signalized intersections using travel times from mobile sensors. Transportation Research Part C: Emerging Technologies* 19(6):1133–1156.

Cheng Y, Qin X, Jin J, Ran B, 2010 *An exploratory shockwave approach for signalized intersection performance measurements using probe trajectories. Transportation Research Board 89th Annual Meeting*, number 10-1617.

Fayazi SA, Vahidi A, 2016 *Crowdsourcing phase and timing of pre-timed traffic signals in the presence of queues: algorithms and back-end system architecture. IEEE Transactions on Intelligent Transportation Systems* 17(3):870–881.

Fayazi SA, Vahidi A, Mahler G, Winckler A, 2015 *Traffic signal phase and timing estimation from low-frequency transit bus data. Intelligent Transportation Systems, IEEE Transactions on* 16(1):19–28.

Hao P, Ban X, Bennett KP, Ji Q, Sun Z, 2012 *Signal timing estimation using sample intersection travel times. Intelligent Transportation Systems, IEEE Transactions on* 13(2):792–804.

Kerper M, Wewetzer C, Sasse A, Mauve M, 2012 *Learning traffic light phase sched-*

ules from velocity profiles in the cloud. 2012 5th International Conference on New Technologies, Mobility and Security (NTMS), 1–5 (IEEE).

Koonce P, Rodegerdts L, Lee K, Quayle S, Beaird S, Braud C, Bonneson J, Tarnoff P, Urbanik T, 2008 *Traffic signal timing manual*. Technical report.

Long G, 2000 *Acceleration characteristics of starting vehicles. Transportation Research Record: Journal of the Transportation Research Board* (1737):58–70.

Mitra SK, Kuo Y, 2006 *Digital signal processing: a computer-based approach*, volume 2 (McGraw-Hill Higher Education).

Osorio C, Nanduri K, 2015 *Energy-efficient urban traffic management: a microscopic simulation-based approach. Transportation Science* 49(3):637–651.

TRB, 2010 *Highway Capacity Manual* (Transportation Research Board), fifth edition edition.

Mvt.	Observed			Estimated			Error Analysis			PCS
	θ_o	ϕ_o	η_o	θ_e	ϕ_e	η_e	$\|\theta_o - \theta_e\|$	$\|\phi_o - \phi_e\|$	$\|\eta_o - \eta_e\|$	
EW	230	50	280	226	51	277	4	1	3	97.08%
ES	230	50	280	230	47	277	0	3	3	98.75%
NS	40	75	115	41	80	121	1	5	6	97.08%
NE	115	55	170	114	56	170	1	1	0	99.58%
SN	40	75	115	37	77	114	3	2	1	98.33%
SW	115	55	170	121	49	170	6	6	0	97.50%
WE	170	60	230	170	60	230	0	0	0	100.00%
WN	170	60	230	170	56	226	0	4	4	98.33%
Average							1.875	2.75	2.125	98.33%

Table 3: Observed signal timing, estimated signal timing and the error analysis for data during 7:00-9:00 in Intersection 1 ($a_{acc} = 1.44m/s$), T = 240s.

Mvt.	Observed			Estimated			Error Analysis			PCS
	θ_o	ϕ_o	η_o	θ_e	ϕ_e	η_e	$\|\theta_o - \theta_e\|$	$\|\phi_o - \phi_e\|$	$\|\eta_o - \eta_e\|$	
EW	160	50	210	161	49	210	1	1	0	99.52%
ES	160	50	210	163	47	210	3	3	0	98.57%
NS	0	70	70	0	69	69	0	1	1	99.52%
NE	70	40	110	65	46	111	5	6	1	97.14%
SN	0	70	70	0	65	165	0	5	5	97.62%
SW	70	40	110	69	42	111	1	2	1	99.05%
WE	110	50	160	111	52	163	1	2	3	98.10%
WN	110	50	160	111	50	161	1	0	1	99.57%
Average							1.5	2.5	1.5	98.57%

Table 4: Observed signal timing, estimated signal timing and the error analysis for data during 9:00-11:00 in Intersection 1 ($a_{acc} = 1.44m/s$), T = 210s.

Mvt.	Observed			Estimated			Error Analysis			PCS
	θ_o	ϕ_o	η_o	θ_e	ϕ_e	η_e	$\|\theta_o - \theta_e\|$	$\|\phi_o - \phi_e\|$	$\|\eta_o - \eta_e\|$	
EW	5	50	55	6	49	55	1	1	0	99.44%
ES	55	40	95	57	42	99	2	2	4	96.67%
NS	95	50	145	99	47	146	4	3	1	97.22%
NE	145	40	185	149	37	186	4	3	1	97.22%
SN	95	50	145	100	49	149	5	1	4	95.00%
SW	145	40	185	146	40	186	1	0	1	98.89%
WE	5	50	55	6	49	55	1	1	0	99.44%
WN	55	40	95	55	42	97	0	2	2	98.89%
Average							2.25	1.625	1.625	97.85%

Table 5: Observed signal timing, estimated signal timing and the error analysis for data during 7:00-9:00 in Intersection 2 ($a_{acc} = 1.44m/s$), T = 180s.

Mvt.	Observed			Estimated			Error Analysis			PCS
	θ_o	ϕ_o	η_o	θ_e	ϕ_e	η_e	$\|\theta_o - \theta_e\|$	$\|\phi_o - \phi_e\|$	$\|\eta_o - \eta_e\|$	
EW	15	40	55	18	37	55	3	3	0	98.00%
ES	55	40	95	54	37	91	1	4	3	96.67%
NS	95	35	130	92	40	132	3	5	2	96.67%
NE	130	35	165	130	37	167	0	2	2	98.67%
SN	95	35	130	91	39	130	4	4	0	97.33%
SW	130	35	165	132	36	168	2	1	3	96.67%
WE	15	40	55	18	36	54	3	4	1	97.33%
WN	55	40	95	55	36	91	0	4	4	97.33%
Average							2	3.25	2	97.33%

Table 6: Observed signal timing, estimated signal timing and the error analysis for data during 9:00-11:00 in Intersection 2 ($a_{acc} = 1.44m/s$), T = 150s.

Mvt.	Observed			Estimated			Error Analysis			PCS
	θ_o	ϕ_o	η_o	θ_e	ϕ_e	η_e	$\|\theta_o - \theta_e\|$	$\|\phi_o - \phi_e\|$	$\|\eta_o - \eta_e\|$	
EW	150	25	175	143	31	174	7	6	1	95.56%
ES	150	25	175	151	25	174	1	0	1	98.89%
NS	180	80	255	176	80	256	1	0	1	98.89%
NE	80	50	130	78	50	128	2	0	2	97.78%
SN	180	80	225	176	80	256	1	0	1	98.89%
SW	80	50	130	83	41	124	3	9	6	95.00%
WE	130	20	150	129	22	151	1	2	1	98.89%
WN	130	20	150	128	15	143	1	5	7	95.00%
Average							2.25	2.75	2.5	97.36%

Table 7: Observed signal timing, estimated signal timing and the error analysis for data during 7:00-9:00 in Intersection 3 ($a_{acc} = 1.44m/s$), T = 180s.

Mvt.	Observed			Estimated			Error Analysis			PCS
	θ_o	ϕ_o	η_o	θ_e	ϕ_e	η_e	$\|\theta_o - \theta_e\|$	$\|\phi_o - \phi_e\|$	$\|\eta_o - \eta_e\|$	
EW	105	20	125	105	15	120	0	5	5	96.67%
ES	105	20	125	105	25	130	0	5	5	96.67%
NS	145	80	225	146	79	225	1	1	0	99.33%
NE	75	30	105	75	30	105	0	0	0	100.00%
SN	145	80	225	145	80	225	0	0	0	100.00%
SW	75	30	105	75	30	105	0	0	0	100.00%
WE	125	20	145	130	15	145	5	5	0	96.67%
WN	125	20	145	130	15	135	5	5	10	90.00%
Average							1.375	2.625	2.5	97.42%

Table 8: Observed signal timing, estimated signal timing and the error analysis for data during 9:00-11:00 in Intersection 3 ($a_{acc} = 1.44m/s$), T = 150s.