hEART 2020,
9th Symposium of the European Association for Research in Transportation

ROADWAY TRAVEL TIMES: MAXIMUM LIKELIHOOD ESTIMATION BASED ON FLOATING CAR DATA INTERVALS

Fabien Leurent [a],.Danyang Sun [a*], Xiaoyan Xie [a]

[a] *Ecole des Ponts ParisTech- Laboratoire Ville Mobilité Transport, 6-8 Avenue Blaise Pascal – 77455 Marne-la-Vallée cedex 2, France*
* *Corresponding author. E-mail address: danyang.sun@enpc.fr*

ABSTRACT

Massive Floating Car Data (FCD) datasets have become available for roadway networks, which contain travel time information on short spatial intervals between pairs of successive observations along individual trips. This paper brings about a stochastic model of travel times with a Maximum Likelihood estimation method to exploit FCD material. Probabilistic specifications are put forward for link travel times as Gaussian random variables along with standard error of each estimator. This allows for simple estimation of link attributes based on "Link FCD intervals" and their confidence intervals. An application instance is dealt with for one motorway and one urban avenue in the Grand Paris area with results showing better accuracy than automotive methods based on pointwise average speed.

Keywords: Roadway Travel Time; Gaussian Link Time; Maximum Likelihood Estimation; Floating Car Data Intervals

INTRODUCTION

Roadway networks are purported to be travelled along by different kinds of vehicles. On any usage occurrence, the individual user makes his or her trip along a selected path. The path travel time is a major characteristic as the path costs time to its individual user: it is usually the main basis for path choice and also for departure time choice and travel mode choice (Ortuzar and Willumsen, 2004). Automated personal travel assistants such as Google Maps, Waze and so on provide path advice on the basis of local travel time as the first and foremost criterion. Thus, local travel times determine path choice, hence the formation of trip flows and in turn the local traffic conditions. At the same time, with the great diffusion of GPS technology, massive Floating Car Data datasets have become available for roadway networks. They contain travel time information on short spatial intervals between pairs of points that are successively recorded along individual trips, which enables to recover local traffic characteristics.

Travel time estimation as a key factor in understanding traffic patterns has been a recurrent research issue in the last few decades. Many well-established technologies have been developed based on loop detectors, vehicle diaries and video cameras (Mori et al., 2015). However, those traditional data collection methods, are inherently limited for wide application concerning its spatial-temporal coverage (Sun et al., 2014). Most recently, with the increasing diffusion of GPS technologies, Floating Car Data is emerging as massive available for collecting traces of a wide range of network all day long, which shows a great potential to resolve the data concern in travel time estimation (Jenelius and Koutsopoulos, 2013). Although much more attention has been aroused recently to studying FCD for traffic analysis, the literature for it on travel time estimation is still limited, in particular on the use of low-frequency floating car data, a more practical trend for the data source nowadays (Jenelius and Koutsopoulos, 2013; Mori et al., 2015).

This paper focuses on the local travel time estimation on the link level. In the literature, the current studies can be generally divided into two streams: data-based approaches and model-based approaches. For the prior ones, taking the average/median of all observed points to recover space-mean speed for link travel time estimation is widely adopted in many FCD based studies (Cheng et al., 2015; Ehmke et al., 2012; Fusco et al., 2016; Long Cheu et al., 2002; Ran et al., 2016; Shen and Ban, 2016; Wang et al., 2015). To be simplistic, we define this way as "pointwise average speed". This method has the advantage of being straightforward but is limited to the case when there are sufficient observations over the targeted areas. As for the model-based approaches, a few studies built probabilistic graphical models from observed probe traces to obtain the travel time probability distributions in terms of a series of spatial and temporal traffic variables. A Bayesian Network was proposed by Hunter et al. (Hunter et al., 2009) for structuring the probabilistic model using low-frequency sparse taxi probe data to estimate the historical link travel time distributions. Development of such an approach was conducted by Hofleitner et al. (Hofleitner et al., 2012a) to focus on travel time forecasting, which proposed a dynamic Bayesian network model to model the state transition between neighboring segments. In another study by Hofleitner et al. (Hofleitner et al., 2012b), the authors did a further development by incorporating the traffic physics, the flow theory and state variables considering the number of querying vehicles and turning fractions at intersections. In addition to the Bayesian network, Ramezani and Geroliminis (Ramezani and Geroliminis, 2012) proposed a Markov chains model to estimate the arterial route travel time distribution. Other than probabilistic models, a study by Jenelius and Koutsopoulos (Jenelius and Koutsopoulos, 2013) developed a statistical regression model based on taxi probe vehicle data to estimate the travel time on urban road network as well as analyzing the impacts of the corresponding influencing variables.

Although such models have the advantage to take many comprehensive factors into account, they also require more external data in terms of the physical and spatial parameters to express the functions more precisely, which limits the large-scale applicability. In the meantime, the complexity of model structuring also restricts the transition to other cities for the variance of network structure and huge computation workload. Another persistent issue in most of the existing studies is the lack of a measure of reliability in the travel time estimation (Mori et al., 2015). Confidence intervals rather than just a unique value of average time would be especially helpful to provide a more complete information to the road users.

Acknowledging the needs to address the above-mentioned problems, this paper aims to build a stochastic model of local travel times together with a Maximum Likelihood estimation method to exploit FCD material. Probabilistic specifications are put forward for link travel times as Gaussian random variables. This allows for simple estimations of link attributes based on "Link FCD intervals". Analytical properties are to be obtained specifically at sub-links along with variance models dealt with postulates. An application study is dealt with for a major motorway segment as well as an urban link for comparison in the Grand Paris area.

METHODOLOGY

In the stochastic model, the travel time is analyzed as a random variable that adds up local random variables that involve local characteristics: we introduce a set of assumptions and derive some theoretical properties, including a Probability Density Function (PDF) for the travel time. At the link level, the local characteristics include the mean and standard deviation of local speed. This is for homogenous sections excluding link endpoints.

The estimation method takes the travel time PDF as a likelihood function for field observations of individual travel times. The network framework enables us to gather large samples of individual trips and extract the associated information by using an ad-hoc method of Maximum Likelihood Estimation. As for application instance, we have availed ourselves of an FCD dataset provided by the Coyote firm: car trajectories are monitored with one geolocation time stamp per half minute. Every pair of two successive individual timestamps contains information on the network conditions in-between. Our method to exploit such information is complementary to the link time estimation methods based on instant speeds monitored at timestamp points (Cheng et al., 2015; Long Cheu et al., 2002).

Let us consider travel times $\Delta h \equiv h' - h$ between point pairs $(\mathrm{M}, \mathrm{M}')$ along link a, separated by spatial length $\Delta s \equiv s' - s$. We model any Δh as a random variable, with stochastic characteristics that depend on the link conditions and the associated parameters. Our modelling assumptions are:

- (L1) that the average time $\mathrm{E}[\Delta h]$ is proportional to the spatial length Δs, with factor coefficient τ_a:

$$\mathrm{E}[\Delta h] = \tau_a \Delta s$$

- (L2) That the variations of the travel time come from a stochastic process along space with autocorrelation function $\chi_a(s, s')$: then,

$$\mathrm{V}[\Delta h] = \chi_a(s, s')$$

Special instances will be considered to make the model simpler. Our basic specification is that local variations are mutually independent and identically distributed per unit of distance: then, denoting by σ_a the standard deviation of local variations (per length unit), it holds that

$$\mathrm{V}[\Delta h] = \sigma_a^2 \, \Delta s.$$

The reason is that the variance of the sum of independent local variables is the sum of their respective variances, therefore leading to linear dependence according to length Δs under the assumption of homogenous distribution. The product form relies upon the hypothesis that successive intervals are statistically independent. Under the Gaussian assumption, the likelihood function of a link interval is simply:

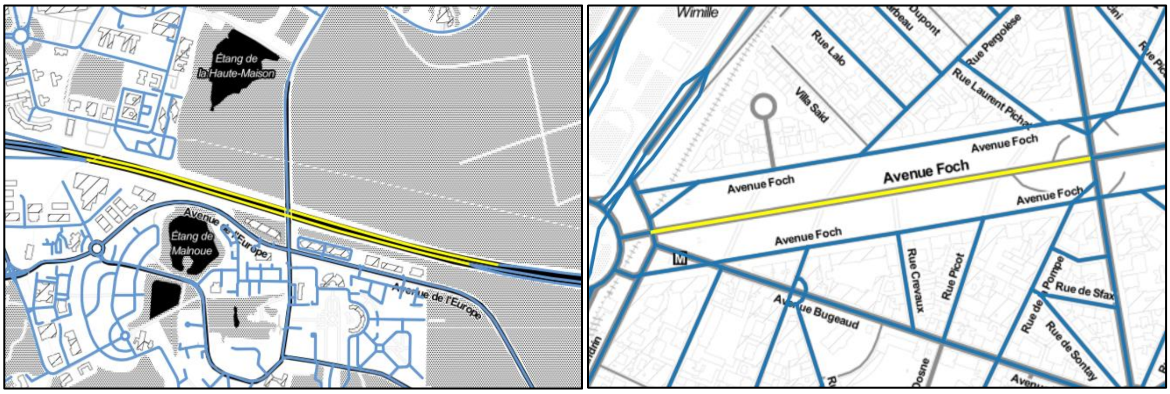
$$L_{ui}(\Theta_a) = \mathrm{f}(h_i^+, h_i^-, s_i^+, s_i^-, \Theta_a) = \frac{\exp\left(-\frac{1}{2}\frac{(\Delta h_i - \tau_a \Delta s_i)^2}{\sigma_a^2 \Delta s_i}\right)}{\sigma_a \sqrt{\Delta s_i}\sqrt{2\pi}}$$

We obtained analytical formulas joint the Maximum Likelihood estimation of the average time and variance parameters, as well as the standard error of estimation associated to each estimator. All formulas are easy to calculate so that the estimation method is straightforwardly applicable, and its accuracy can be controlled.

APPLICATION AND RESULTS

Study Location

The proposed method was applied on two different roadway segments, with the aim to compare the experimental results between highway setting and urban setting. The highway segment was selected from a major link along the motorway A4 in Great Paris region, which performs as a main arterial serving the traffic between the center and eastern sub-regions. The urban segment is chosen from on the Avenue Foch, which a major avenue in Paris. Geographical layouts are shown as in Figure 1. Travel time of the two-directional movement was studied separately. No ramp access or intersection was included in this application as the model focuses on the link travel time. The segment length is 1445m and 1635m for the eastbound and the westbound direction respectively on the A4 motorway segment, and 607m for both directions on the urban segment.



(a) Segment along A4 motorway (b) Segment along Avenue Foch (urban setting)

Figure 1 Studied roadway segments

Data Set

The FCD was obtained by onboard GPS devices from Coyote, which is a major roadway navigation service provider in France. The raw FCD is organized by a sequence of vehicle logs, each of which represents an instantaneous trace of a vehicle, containing its location, time stamp, speed and travelling direction etc. The sampling frequency is around 30s subject to the variation of signal transmission condition. Data over two normal weekdays (February 05 and February 06, 2019) on the selected segments were analyzed. The network roadway data were extracted from OpenStreetMap. Due to imperfect recording of GPS coordinates, the deviation between FCD points and road network is quite common. Numerous effective map-matching algorithms were developed by previous studies (Liu et al., 2017; Newson and Krumm, 2009). In this study, the FCD was map-matched to the nearest roadways according to the travelling directions and re-projected the locations to the nearest foot-points on the segment.

Link Interval Extraction

Link intervals along the segments were extracted for the two directions in two days respectively. Each interval consists of a pair of two successive FCD timestamps. Distance travelled along the road to the starting node was also calculated based on geo-coded coordinates using geo-packages in Python. Anonymized vehicle ID was used to track different vehicles. Invalid pairs were excluded if the trajectory time span was abnormal, setting the rule as less than 300s considering consecutive sampling frequency is around 30s. Tolerance was made for in-stable signal condition. A descriptive summary of extracted intervals and all the point-wise observations is given in Table 1.

TABLE 1 Descriptive summary of extracted intervals

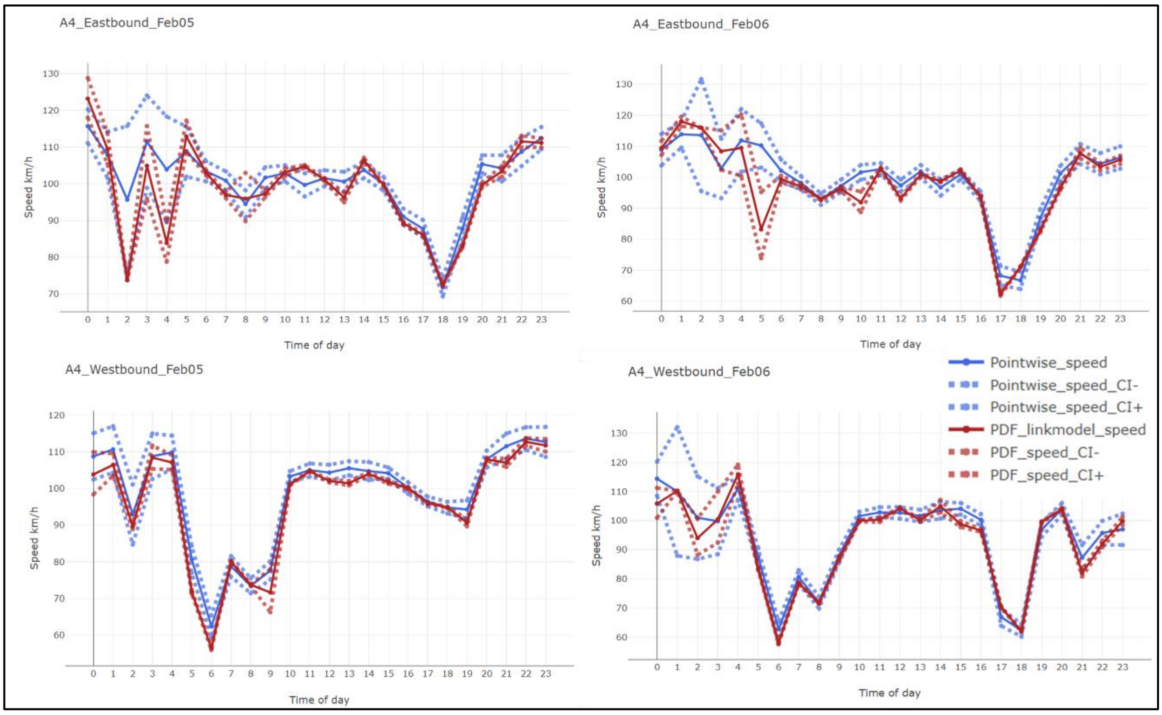
Setting		**Eastbound 05**	**Eastbound 06**	**Westbound 05**	**Westbound 06**
A4	**Link intervals** (count)	1248	1327	2014	2073
	Pointwise observations	3472	3605	4401	4416
	Average speed overall	96.7 km/h	93.0 km/h	91.5 km/h	86.5 km/h
AF*	**Link intervals** (count)	787	983	452	464
	Pointwise observations	1364	1692	1019	1117
	Average speed overall	23.7 km/h	24.1 km/h	36.7 km/h	34.5 km/h

*AF stands for Avenue Foch

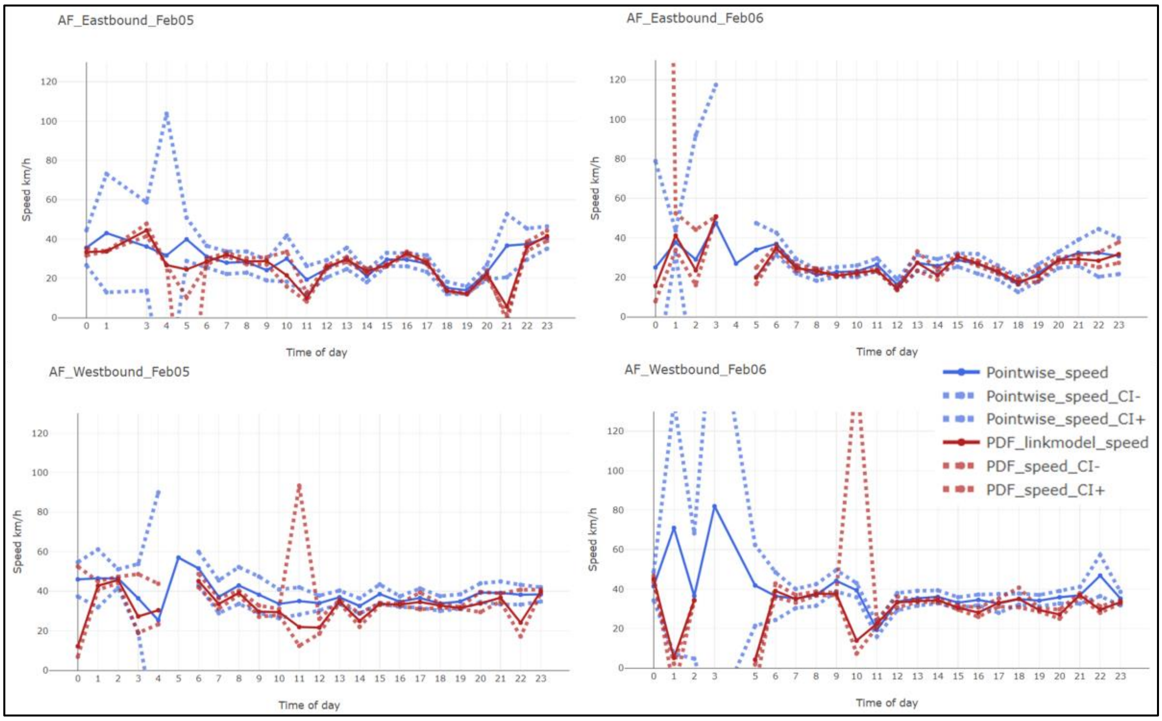
Link Analysis Resutls

The stochastic parameters of the link model were estimated based on the extracted data by different time of the day. Besides, the corresponding pointwise average speed was also computed. To measure the reliability of the estimation, confidence intervals were computed stemming from those estimated parameters. As a result, line-charts were plotted to shows a detailed comparison between the interval estimation of PDF link model and the point-wise estimation for both the two segments, shown in Figure 2. Space mean speed was used for the comparison in the plot, as for a given length, modeling space mean speed is essentially equivalent as modeling the link travel time (Hall, 1996; Mori et al., 2015).

As can be seen from the two plots, the space mean speed estimated by the interval estimation is generally consistent with pointwise average speed with a similar fluctuating trend. Significant speed reductions were observed on the motorway segment during peak hours along the tidy movement to and from the city center. The urban segment was observed with less fluctuation but overall with relatively low speed. However, the interval estimation was found more likely to estimate a lower speed than pointwise average especially on the urban segment which involves more congested scenarios with higher variation in vehicular motion. Moreover, the confidence intervals were significantly narrower than those of pointwise average speed for most of the situations, which indicates that the interval estimation could provide a more reliable estimation of the travel time. It was also found that the more data available, the more precise results on the estimation. Nevertheless, the interval estimation would require less data to reach a higher precision level.



(a) A4 motorway segment



(b) Avenue Foch urban segment

Figure 2 Result comparision between the interval estimation and the pintwise average estimation

CONCLUSION

This paper puts forward a stochastic model of local travel time estimation along roadway links on the network. Basic modeling assumptions were postulated to model link travel times as random variables. Building upon the stochastic model, we have devised a Maximum Likelihood estimation method that can be applied to FCD trajectories along the network. Intervals in time and space between two successive timestamps monitored along the trajectories constitute the basic data. The practicality of the estimation was demonstrated in a case study. Estimations were computed and compared between our stochastic model and straightforward conventional pointwise average method along with confidence intervals. Results indicate that the stochastic model is able to deliver a more reliable estimation and require fewer observations to reach a higher precision. Moreover, the pointwise average estimation was found tending to provide a higher speed than the stochastic model with less certainty. This may lead to an underestimation of the travel time, implicating the limitations of the current applications adopting such a straightforward estimation.

This research is restricted to the link level. However, it could be saved as a modular section. Further research may be invested to build the probabilistic model for node or intersections between different links so as to develop more reliable estimation of path travel time.

REFERENCES

Cheng, X., Zhao, X., Xu, Z., Zhou, J., Yang, N., 2015. Prediction of the shortest travel time based on intersection delay, in: 2015 IEEE First International Smart Cities Conference (ISC2). IEEE, pp. 1–7.

Ehmke, J.F., Meisel, S., Mattfeld, D.C., 2012. Floating car based travel times for city logistics. Transp. Res. Part C Emerg. Technol. 21, 338–352.

Fusco, G., Colombaroni, C., Isaenko, N., 2016. Short-term speed predictions exploiting big data on large urban road networks. Transp. Res. Part C Emerg. Technol. 73, 183–201.
Hall, F.L., 1996. Traffic stream characteristics. Traffic Flow Theory US Fed. Highw. Adm. 36.
Hofleitner, A., Herring, R., Abbeel, P., Bayen, A., 2012a. Learning the dynamics of arterial traffic from probe data using a dynamic Bayesian network. IEEE Trans. Intell. Transp. Syst. 13, 1679–1693.
Hofleitner, A., Herring, R., Bayen, A., 2012b. Arterial travel time forecast with streaming data: A hybrid approach of flow modeling and machine learning. Transp. Res. Part B Methodol. 46, 1097–1122.
Hunter, T., Herring, R., Abbeel, P., Bayen, A., 2009. Path and travel time inference from GPS probe vehicle data. NIPS Anal. Netw. Learn. Graphs 12, 1–8.
Jenelius, E., Koutsopoulos, H.N., 2013. Travel time estimation for urban road networks using low frequency probe vehicle data. Transp. Res. Part B Methodol. 53, 64–81.
Liu, X., Liu, K., Li, M., Lu, F., 2017. A ST-CRF Map-Matching Method for Low-Frequency Floating Car Data. IEEE Trans. Intell. Transp. Syst. 18, 1241–1254. https://doi.org/10.1109/TITS.2016.2604484
Long Cheu, R., Xie, C., Lee, D.-H., 2002. Probe vehicle population and sample size for arterial speed estimation. Comput. Civ. Infrastruct. Eng. 17, 53–60.
Mori, U., Mendiburu, A., Álvarez, M., Lozano, J.A., 2015. A review of travel time estimation and forecasting for Advanced Traveller Information Systems. Transp. Transp. Sci. 11, 119–157.
Newson, P., Krumm, J., 2009. Hidden Markov map matching through noise and sparseness, in: Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems. ACM, pp. 336–343.
Ortuzar, J. de D., Willumsen, L.G., 2004. Introduction in modeling transport. West Sussex, England: John Wiley & Sons, LTD.
Ramezani, M., Geroliminis, N., 2012. On the estimation of arterial route travel time distribution with Markov chains. Transp. Res. Part B Methodol. 46, 1576–1590. https://doi.org/10.1016/j.trb.2012.08.004
Ran, B., Song, L., Zhang, J., Cheng, Y., Tan, H., 2016. Using tensor completion method to achieving better coverage of traffic state estimation from sparse floating car data. PloS One 11, e0157420.
Shen, J., Ban, Y., 2016. Route Choice of the Shortest Travel Time Based on Floating Car Data. J. Sens. 2016.
Sun, D., Zhang, C., Zhang, L., Chen, F., Peng, Z.-R., 2014. Urban travel behavior analyses and route prediction based on floating car data. Transp. Lett. 6, 118–125.
Wang, X., Peng, L., Chi, T., Li, M., Yao, X., Shao, J., 2015. A hidden Markov model for urban-scale traffic estimation using floating car data. PloS One 10, e0145348.