Traveller recurrence and inter- versus intra-traveller speed variability: Analysis with Bluetooth data

Erik Jenelius
Department of Civil and Architectural Engineering
KTH Royal Institute of Technology
Brinellvägen 23, SE-100 44 Stockholm, Sweden
Email: erik.jenelius@abe.kth.se
ORCiD: 0000-0002-4106-3126

February 2, 2022

Abstract

This paper proposes a linear mixed model of route speed distributions that separates the variability into an inter-traveller component, consistent across days and time intervals for each recurrent traveller, and an intra-traveller component representing uncertainty. The intra-traveller variability corresponds to travel time uncertainty, while the total variability is typically captured by empirical measurements and used in travel time reliability assessments. The intra-traveller and the total variability differ if there are systematic differences in speed between different recurrent travellers. The paper also investigates to what degree vehicles traversing a route during the morning or evening peak over multiple days are recurrent travellers. Using data from Bluetooth and Wifi sensors on 26 routes in Stockholm, Sweden over a three-month period, we find that the traveller recurrence is higher towards the city in the morning peak and out from the city in the afternoon. Model estimation results show that the relative intra-traveller variability is also significantly higher in the commute direction (towards the city in the morning and out from the city in the afternoon) and on routes with high congestion levels. The relations revealed in this paper may be used to estimate the relevant intra-traveller variance based on the total variance and readily available route attributes. Without this correction, the costs associated with travel time variability may be overestimated.

1 Introduction

It is well known that the travel time on a given route is not a constant but typically varies between trips. Some part of the variability is usually predictable by an experienced traveller based on systematic demand variations across hours, weekdays, months and years. Part of the variability, however, arises from factors that are difficult for travellers to foresee, such as demand fluctuations, weather conditions, incidents, etc. This part of the travel time variability gives rise to travel time *uncertainty*. The extent to which the travel time can be predicted in advance is referred to as travel time *reliability*. Low reliability is associated with considerable costs due to late arrivals and the need for safety margins in departure times [1, 2].

It is by now widely accepted that travel time variability, and not only the expected travel time, is associated with economic costs. A large literature has appeared in recent years, covering the theoretical aspects of reliability, traveller behaviour under travel time uncertainty, valuation studies through stated and revealed preference experiments, as well as empirical investigations of travel time variability [1, 2, 3].

A common metric of travel time variability borrowed from general statistics is the standard deviation. However, this metric can be difficult to interpret for travellers and practitioners, and various other metrics have been proposed [3]. The works of [4, 5] derived a foundation for using the standard deviation based on micro-economic scheduling models, in which costs arise from early or late arrivals. Later extensions have found that other metrics such as the travel time variance can be motivated given other assumptions about travellers' scheduling preferences [6].

Many studies have assessed the behaviour of travellers facing travel time uncertainty, in particular their trade-offs between uncertainty, average travel time and travel cost. The value of travel time reliability has been studied in the context of mode choice [7], route choice [8, 9], departure time choice [10], and toll road usage and road pricing [11, 12]. Across studies, the ratio between the value of reliability and the value of travel time has been found to lie in the range between 0.2 and 1.5 with a typical value just under 1 [2].

Economic appraisals of transport policies and infrastructure projects are generally based on travel demand forecasting and traffic assignment models that treat link travel times as deterministic or, interpreted differently, only consider expected travel times [2]. In order to incorporate travel time reliability in appraisals, a number of studies have therefore sought to establish analytical relations between the mean and the standard deviation (or some other metric of variability) of travel time at the link level [13, 14, 15]. Other studies have modelled the correlations between link travel time distributions along a route or in a network [16, 17, 18].

On the empirical side, a long line of research has sought to model and characterise the variability of travel times on different types of roads, links and routes. Many studies have found travel time distributions to be asymmetric with long upper tails [19, 20]. Proposed closed-form distribution functions include log-normal [21], stable [19], gamma [20] and Burr Type XII [22]. Other approaches model the asymmetric distributions as mixtures of multiple underlying distributions representing different traffic states

[23, 24, 25]. A non-parametric approach to estimating route travel time distributions based on floating car data is proposed in [26].

A taxonomy of sources of travel time variability is provided in [27], who separate them into (1) traffic influencing events, including traffic incidents and accidents, road construction work, weather and environmental conditions, (2) traffic demand, including day-to-day fluctuations and special events, and (3) physical road features, including traffic control infrastructure and road capacity. Several authors note that travel time variability can be separated into day-to-day variability, within-day (interval-to-interval) variability, and vehicle-to-vehicle variability [28, 29]. A mixture model of compound Gamma-Gamma distributions to jointly represent day-to-day and vehicle-to-vehicle variability under different traffic states is proposed in [24]. The model is further developed and applied to empirical data in [20].

For some types of trips such as the commute to and from work, travellers tend to use the same route around the same time of day repeatedly over multiple days. An important distinction can be made between the individual traveller's uncertainty across days and the total variability across all trips and travellers. While the individual uncertainty determines the associated reliability cost, empirical travel time measurements typically provide the total travel time variability without distinguishing between recurrent vehicles. The intra-traveller and the total variability differ if there are systematic differences in speed between different recurrent travellers. These differences, referred to here as inter-traveller variability, could be due to heterogeneous preferred driving speeds, or because frequent travellers may be able to strategically plan their departure times to reduce travel time variability. The total travel time variability across trips, travellers and days can thus be conceptually decomposed as

$$\begin{aligned}\text{Total variability} = {} & \text{Inter-traveller variability} \\ & + \text{Intra-traveller variability (uncertainty)}\end{aligned} \tag{1}$$

If the relative magnitude of the inter-traveller variability component is significant, this must be taken into account in economic valuations [2] and modelling [30] of travel time reliability.

While some existing studies distinguish between day-to-day and vehicle-to-vehicle variability, no study that we are aware of has considered that some vehicles, i.e., travellers, repeatedly traverse the same route multiple days. A reason for this is presumably that data on travellers' mobility over multiple days have been lacking historically. Analysis of public transport smart card data reveal that travel patterns exhibit strong regularity between days [31], but evidence from private cars is so far limited due to difficulties in data collection. Several studies have used travel diaries to study the regularity of individuals' activity-travel patterns across days and have found the highest degree of repetition for essential activities such as commuting [32]. Further, the level of repetition of daily activity-travel patterns is more correlated with commitments and obligations than with travel mode choice [33]. Meanwhile, studies have shown that familiarity with a route is coupled with higher driving speed [34], which suggests that there is a consistent vehicle-specific component of travel time variability.

Given that there are consistent travel time variations among different travellers, this means that the vehicle-to-vehicle and day-to-day dimensions are intertwined. To correctly capture the part of the total travel time variability that is experienced by an individual traveller a more elaborate model is needed, which decomposes the total variability into day-to-day variability, interval-to-interval variability, traveller-to-traveller variability, and residual (within-day-period-traveller) variability.

The aim of this paper is highlight the distinction between inter versus intra-traveller travel time or speed variability, to assess the prevalence of recurrent travellers on urban motorway and arterial routes, to extract the inter and intra-traveller variability from the total variability and to assess their relative magnitudes. The analysis utilizes disaggregate travel time observations from Bluetooth and Wifi devices installed on multiple routes.

The paper is organized as follows. Section 2 describes the methodology for analysing traveller recurrence and speed variability and describes a case study on multiple arterial and motorway routes in Stockholm, Sweden. Section 3 presents and discusses the results from the case study and Section 4 concludes the paper.

2 Method and Case Study

This section proposes a method for analysing variations in traveller recurrence between routes based on disaggregate travel time observations with associated (possibly rehashed) device IDs from Bluetooth and Wifi data. Further, a linear mixed model of route space-mean speeds is proposed which separates total variance into inter-traveller and intra-traveller variance components.

2.1 Disaggregate travel time data

The methodology is based on route travel time measurements from individual trips. For each measurement, three items of information are assumed to be available: (1) the measured travel time, (2) the date and time of the measurement, and (3) a consistent identifier (ID) for the vehicle. This type of data may be collected through various technologies, e.g., Bluetooth and Wifi sensors or automatic number plate recognition (ANPR) cameras. The data set used in the case study has some limitations with respect to the third item, consistent IDs, and the implications for the analysis are addressed in the following.

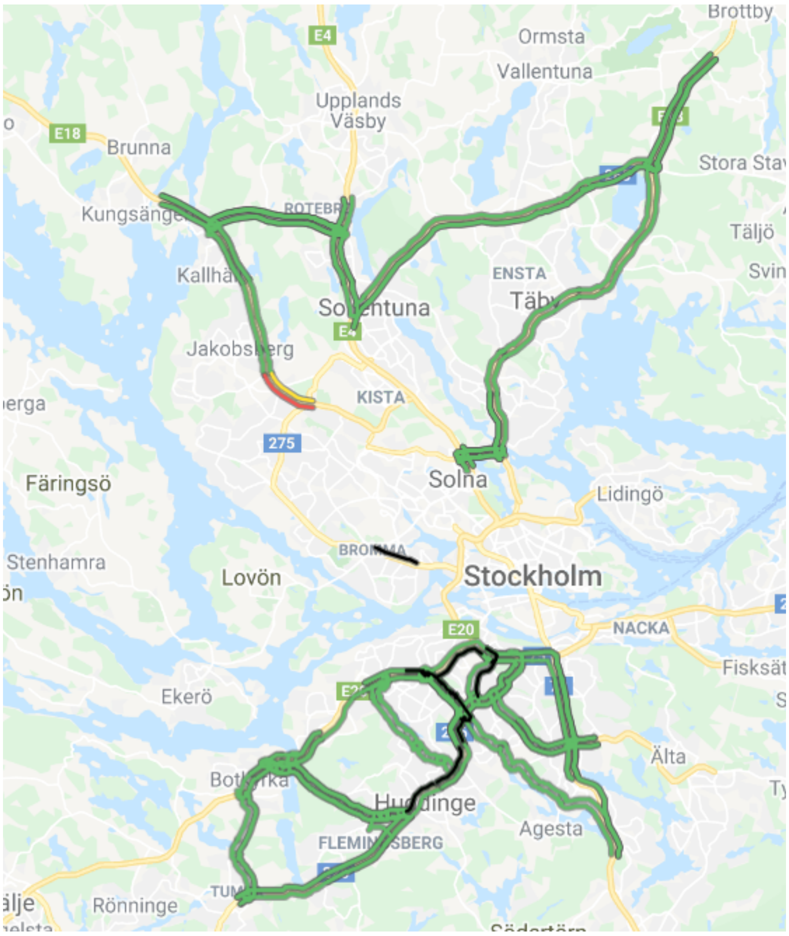

Figure 1: Map of routes equipped with Bluetooth sensors.

We use travel time data from a set of routes in the Stockholm region, shown in Figure 1. Each route is defined as a pair of detectors that capture the MAC addresses of Bluetooth and Wifi devices in vehicles (including mobile phones and other devices inside the vehicles). Travel times are measured by matching the MAC addresses and associated time stamps between the pair of detectors. The data cover the three-month period from 1 January 2019 to 27 March 2019, which includes 61 work days (Monday-Friday).

Each row of data contains information about

- Route id
- Anonymized device ID
- Trip time stamp (the time passing the upstream detector)
- Trip travel time (the time difference between passing the two detectors)

- Level of outlier

The MAC addresses have been hashed into anonymous ID numbers before they are accessible for our analysis. The extent to which the addresses are re-hashed over time is not known to us, which is a disadvantage of the data set. Further, it is possible that mobile devices internally resample their MAC addresses at certain time intervals. These factors imply that the same vehicle may be recorded under multiple IDs over the three-month period, which means that the true number of unique vehicles is lower than the number of IDs. In any case, it can be assumed that this bias is similar across all routes, and we can study variations in traveller recurrence and speed variability components across different route characteristics.

The data are separated into two time period categories: morning peak (Monday-Friday, 8–9 am) and afternoon peak (Monday-Friday, 4–5 pm). For each route and time period we remove all observations marked as outliers. After this trimming, we discard any routes with less than 100 observed trips or observed trips form less than 15 distinct days in either the morning or the afternoon peak. This produces a final set of 26 routes, listed in Table 1. The table also shows the length of each route and three attributes used in the subsequent analysis: the geographical region (north or south of the city center), road type (motorway or arterial) and direction (towards or away from the city center). The number of observed trips for each route in the morning and afternoon peaks, respectively, is shown in Table 2. The average number of days per route with available data is 57.1 and 58.0 for the morning and afternoon periods, respectively.

2.2 Traveller recurrence

As indicator of the recurrence of travellers on a route and time period we use the mean number of trips per vehicle ID and day. A recurrence value of 1 thus means that the traveller uses the route in the same time period once per day on average. Due to rehashing of vehicle IDs, we expect that the actual recurrence is higher than what is observed from the data. In any case, we are interested in the variation in recurrence across different routes and time periods, in particular the morning and afternoon peaks. To assess the influence of route characteristics and peak period on the recurrence, a linear regression model is estimated,

$$\begin{aligned}\text{Recurrence}_{pr} = \alpha_0 + \alpha_1 \text{Direction}_r + \alpha_2 \text{Region}_r + \alpha_3 \text{Roadtype}_r \\ + \alpha_4 \text{Period}_p + \alpha_5 \text{Direction}_r \cdot \text{Period}_p + \nu_{pr},\end{aligned} \tag{2}$$

where $p \in \{\text{am}, \text{pm}\}$ indicates the time period, $r \in \{1, \ldots, R\}$ indicates the route and $R = 26$ is the number of routes. The error terms ν_{pr} are i.i.d. Normal. The parameters $\alpha_0, \ldots, \alpha_5$ are to be estimated. The model includes the region, road type and direction of the route, the time period, as well as

Table 1: Routes used in the case study.

Route	Length (m)	Region	Type	Dir.
01. E4/E20 S Västberga–Västertorp	1225	south	motorway	out
02. E4/E20 S Lindvreten–Fittja	2178	south	motorway	out
03. E4/E20 S Fittja–Botkyrka	1572	south	motorway	out
04. 226 S Lännavägen–Glämstavägen	2046	south	arterial	out
05. E4/E20 N Västertorp–Västberga	1237	south	motorway	in
06. E4/E20 N Fittja–Lindvreten	2175	south	motorway	in
07. E4/E20 N Botkyrka–Fittja	1562	south	motorway	in
08. 226 N Glömstavägen–Lännavägen	2044	south	arterial	in
09. 226 S Glömstavägen–Hälsovägen	2083	south	arterial	out
10. 226 N Hälsovägen–Glömstavägen	2083	south	arterial	in
11. 73 S Sofielundsplan–Gubbängen	2083	south	arterial	out
12. 73 N Gubbängen–Sofielundsplan	2083	south	arterial	in
13. 73 S Gubbängen–Farsta	2083	south	arterial	out
14. 73 N Farsta–Gubbängen	2083	south	arterial	in
15. 73 N Skogås–Farsta	2083	south	arterial	in
16. E18 O Kalhäll–Jakobsberg	3200	north	motorway	in
17. E18 O Jakobsberg–Barkarby	1100	north	motorway	in
18. E18 O Barkarby–Hjulsta	2500	north	motorway	in
19. E18 V Hjulsta–Barkarby	2500	north	motorway	out
20. E18 V Barkarby–Jakobsberg	1100	north	motorway	out
21. E18 V Jakobsberg–Kalhäll	3200	north	motorway	out
22. E18 O Mörby–Lahäll	2022	north	motorway	out
23. E18 O Lahäll–Viggbyholm	4620	north	motorway	out
24. E18 V Viggbyholm–Lahäll	4629	north	motorway	in
25. E18 V Lahäll–Mörby	2000	north	motorway	in
26. E18 V Mörby–Ålkistan	3000	north	motorway	in

the interaction between direction and time period, as explanatory variables. This allows us to distinguish between the morning and afternoon commutes towards and out from the city center.

2.3 Linear mixed model of route speed

In order to model the speed variability on each route and time period, we divide both morning and afternoon peak periods into time intervals of approximately 15 minutes. Let T_{ijk} denote the travel time on a certain route and time period on day i during time interval j for vehicle k (we omit route and time period indices here for clarity). We model the space-mean speed $v_{ijk} = L/T_{ijk}$, where L is the length of the route. The reason for this choice is two-fold. First, it normalizes the measurements across different routes. Second, the space-mean speed distribution tends to be more symmetric and similar to a normal distribution for the routes considered in the case study (compare Figure 3).

The space-mean speed v_{ijk} is modelled in a linear mixed model (LMM) framework [35] as the sum of a deterministic component μ_{ij} and a zero-mean random term u_{ijk},

$$v_{ijk} = \mu_{ij} + u_{ijk}. \tag{3}$$

The deterministic term μ_{ij} is used to control for parts of the variability that can be predicted based on systematic temporal features. It is modelled as a linear function of a set of predictors (fixed effects),

$$\mu_{ij} = \beta_0 + \beta_1 \text{Interval}_j + \beta_2 \text{Weekday}_i + \beta_3 \text{Month}_i, \tag{4}$$

where Interval_j, Weekday_i and Month_i are categorical variables for the time interval, weekday and month of the observation, respectively, represented as sets of dummy variables with associated parameter vectors β_1, β_2 and β_3.

The focus of our analysis is on the random component u_{ijk}, which is split into an inter-traveller part ε_k (consistent across days and time intervals), and an intra-traveller part ε'_{ijk}. The inter-traveller part represents systematic differences in travel speed across travellers, and each traveller is associated with a specific realization of the term. The inter-traveller part is assumed not known by the traveller before the trip and reflects the individuals' travel time uncertainty. Speeds are further assumed to vary randomly between days, between time intervals within each day, and even between all trips within each day and time interval. The intra-traveller component is thus decomposed into a day-to-day component ε_i, a within-day interval-to-interval component ε_{ij}, and a residual component ε_{ijk}. The final specification of the (random effects) model is

$$u_{ijk} = \varepsilon_k + \varepsilon_i + \varepsilon_{ij} + \varepsilon_{ijk}. \tag{5}$$

The random terms are assumed to be mutually independent and distributed normal. This assumption implies that the individual variance is equal for every traveller. The inter-traveller component ε_k has variance σ_t^2, the day-to-day component ε_i has variance σ_d^2, the interval-to-interval component ε_{ij} has variance σ_i^2, and the residual component ε_{ijk} has variance σ^2.

The speed model is estimated separately for each time period (morning and afternoon peak) and route combination (p, r), $p \in \{\text{am}, \text{pm}\}$, $r \in \{1, \ldots, R\}$ with the restricted maximum likelihood (REML) method [35]. The estimation is carried out using the `fitlme` routine in Matlab R2018b with the `fminunc` unconstrained optimization route.

2.4 Intra-traveller speed variability

Across all days, periods and travellers, the magnitude of the intra-traveller variance is $\sigma_\text{d}^2+\sigma_\text{i}^2+\sigma^2$ while the total travel time variance is $\sigma_\text{t}^2+\sigma_\text{d}^2+\sigma_\text{i}^2+\sigma^2$.

We define the *relative intra-traveller variance* (RIV) as the ratio between the intra-traveller and total variances,

$$\mathrm{RIV} = \frac{\text{Intra-traveller variance}}{\text{Total variance}} = \frac{\sigma_{\mathrm{d}}^2 + \sigma_{\mathrm{i}}^2 + \sigma^2}{\sigma_{\mathrm{t}}^2 + \sigma_{\mathrm{d}}^2 + \sigma_{\mathrm{i}}^2 + \sigma^2}. \tag{6}$$

This ratio, which lies between 0 and 1, captures the degree to which the total speed variability across all trips represents the travellers' uncertainty.

Rehashing of vehicle IDs implies that some of the intra-traveller variability is incorrectly attributed to inter-traveller variability. This means in turn that the relative intra-traveller variance is underestimated. However, the influence of route and time period characteristics can be assessed qualitatively. For this purpose a linear regression model is estimated,

$$\begin{aligned} \mathrm{RIV}_{pr} &= \gamma_0 + \gamma_1 \mathrm{Direction}_r + \gamma_2 \mathrm{Region}_r + \gamma_3 \mathrm{Roadtype}_r \\ &+ \gamma_4 \mathrm{Period}_p + \gamma_5 \mathrm{Direction}_r \cdot \mathrm{Period}_p + \gamma_6 \mathrm{Congestion}_{pr} + \omega_{pr}, \\ & p \in \{\mathrm{am}, \mathrm{pm}\},\ r \in \{1, \ldots, R\} \end{aligned} \tag{7}$$

The parameters $\gamma_0, \ldots, \gamma_6$ are to be estimated. The error terms ω_{pr} are i.i.d. Normal. In addition to the previously introduced variables, $\mathrm{Congestion}_{pr}$ is defined as the ratio between the route space-mean speed during off-peak hours (measured on Sundays 8–9 am) and during period p, shown in Table 2. The value 1 corresponds to the same speed as during off-peak while higher numbers indicate higher congestion. As can be seen, some routes even display higher speeds during peak hours than off-peak hours, which could be due to more aggressive driving behaviour.

3 Results and Discussion

This section presents results from the analysis of traveller recurrence and the relative intra-traveller speed variance in Stockholm.

3.1 Day-to-day traveller route recurrence

Figure 2 shows the distribution of number of trips per vehicle ID and per route across all 26 routes during the analysis period. Blue and red bars indicate the morning and afternoon peak hours, respectively. Around one third of all observed trips are generated by vehicles whose ID appear only once on the same route during the analysis period. Thus, around two thirds of the trips are generated by vehicles who are observed at least twice on the same route. The average number of trips per ID is 2.22 in the morning peak and 2.05 in the afternoon peak, which corresponds to average recurrence 0.0391 and 0.0355, respectively. The lower recurrence in the afternoon could reflect more varied travel habits compared to the morning.

Table 2: Case study route statistics.

Route	AM			PM		
	Num. trips	Recurrence	Congestion	Num. trips	Recurrence	Congestion
1	46710	0.0366	0.9962	67038	0.0355	1.1893
2	44320	0.0380	0.9999	60825	0.0359	1.2068
3	43348	0.0380	1.0019	57008	0.0362	1.1672
4	10252	0.0359	1.0437	11361	0.0468	0.9373
5	54858	0.0385	1.4135	51720	0.0310	1.0375
6	46556	0.0386	1.0361	51228	0.0322	1.0477
7	40889	0.0382	1.0358	48866	0.0320	1.0570
8	10807	0.0403	1.1305	15198	0.0387	1.0834
9	14750	0.0367	1.0540	17639	0.0392	1.0173
10	11808	0.0351	1.0512	19336	0.0343	1.1041
11	22093	0.0314	1.0052	47861	0.0399	1.3701
12	35537	0.0396	1.3324	23637	0.0309	1.0949
13	19345	0.0334	1.0012	35760	0.0401	1.2457
14	38272	0.0421	1.2249	26694	0.0329	0.8799
15	26328	0.0403	1.0992	23541	0.0337	0.9155
16	25545	0.0378	1.4603	20693	0.0300	0.8820
17	25417	0.0356	1.7851	17153	0.0280	0.8467
18	27080	0.0384	1.7403	10766	0.0245	0.8759
19	15123	0.0317	0.9970	26887	0.0331	1.5122
20	15403	0.0284	1.0069	34023	0.0348	1.1074
21	3503	0.0500	0.9936	27399	0.0352	1.1393
22	25050	0.0401	1.0154	57416	0.0466	1.7802
23	19236	0.0472	1.0204	38370	0.0448	1.1503
24	25332	0.0458	1.8027	19352	0.0353	0.8224
25	53353	0.0533	1.2636	27219	0.0367	0.9556
26	25268	0.0444	1.4114	7814	0.0355	0.8531

Table 2 shows the mean recurrence for each route and the morning and afternoon peaks separately. The recurrence ranges from 0.0245 for route 18 in the afternoon to 0.0472 for route 23 in the morning. The linear regression model in equation (2) is estimated based on the information in Tables 1 and 2. Estimation results are shown in Table 3.

The intercept represents the recurrence in the baseline case: an arterial route in the north region aligned towards the city center during the morning peak. There is no statistically significant difference between the north and south regions, nor between arterials and motorways. However, there is a significantly lower recurrence in the afternoon than in the morning towards the city (p-value 7.9e-5). Further, an F-test shows that the recurrence is significantly higher out from the city than towards the city in the afternoon (F-statistic 11.5, p-value 0.0015). Finally, the recurrence out from the city in the afternoon peak is not significantly different from towards the city in the morning peak (F-statistic 0.600, p-value 0.443). The R^2 of the model is 0.33, which indicates that a large portion of the variation in recurrence is

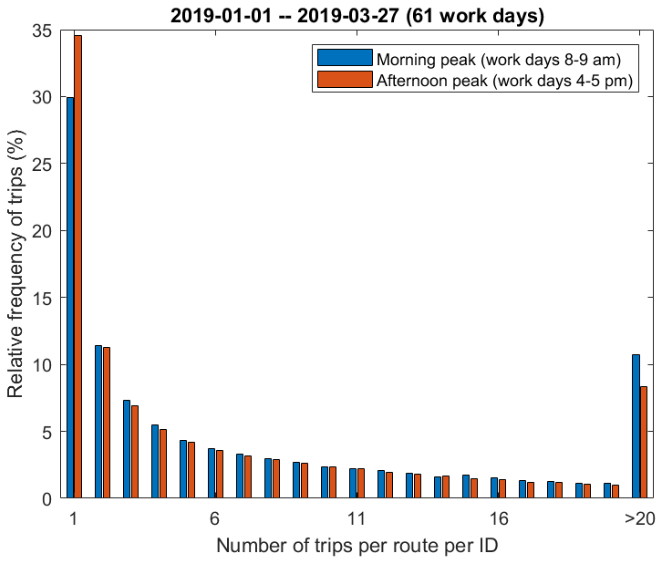

Figure 2: Distributions of number of trips per route per vehicle ID during the morning and afternoon peak hours.

Table 3: Estimation results for traveller recurrence model.

Parameter	Est.	SE	p-value
Intercept	0.0428	0.00236	1.48e-22
Direction out	-0.00323	0.00193	0.102
Region south	-0.00223	0.00176	0.211
Roadtype motorway	-0.00149	0.00183	0.419
PM peak	-0.00803	0.00185	7.92e-05
Direction out & PM peak	0.00976	0.00273	0.000830
Number of observations	52		
R^2	0.328		
F-statistic vs. constant model	4.5		0.00202

not explained by the included factors. The rehashing of device IDs could be a partial explanation.

All in all, the results show that a relatively small part of the traffic flow on urban routes consists of recurring travellers, although the rehashing of vehicle IDs biases the estimated number of trips per traveller downwards. The recurrence is higher towards the city in the morning peak and out from the city in the afternoon peak compared to other situations. This is in line with expectations that commute trips tend to be regular and follow these directions.

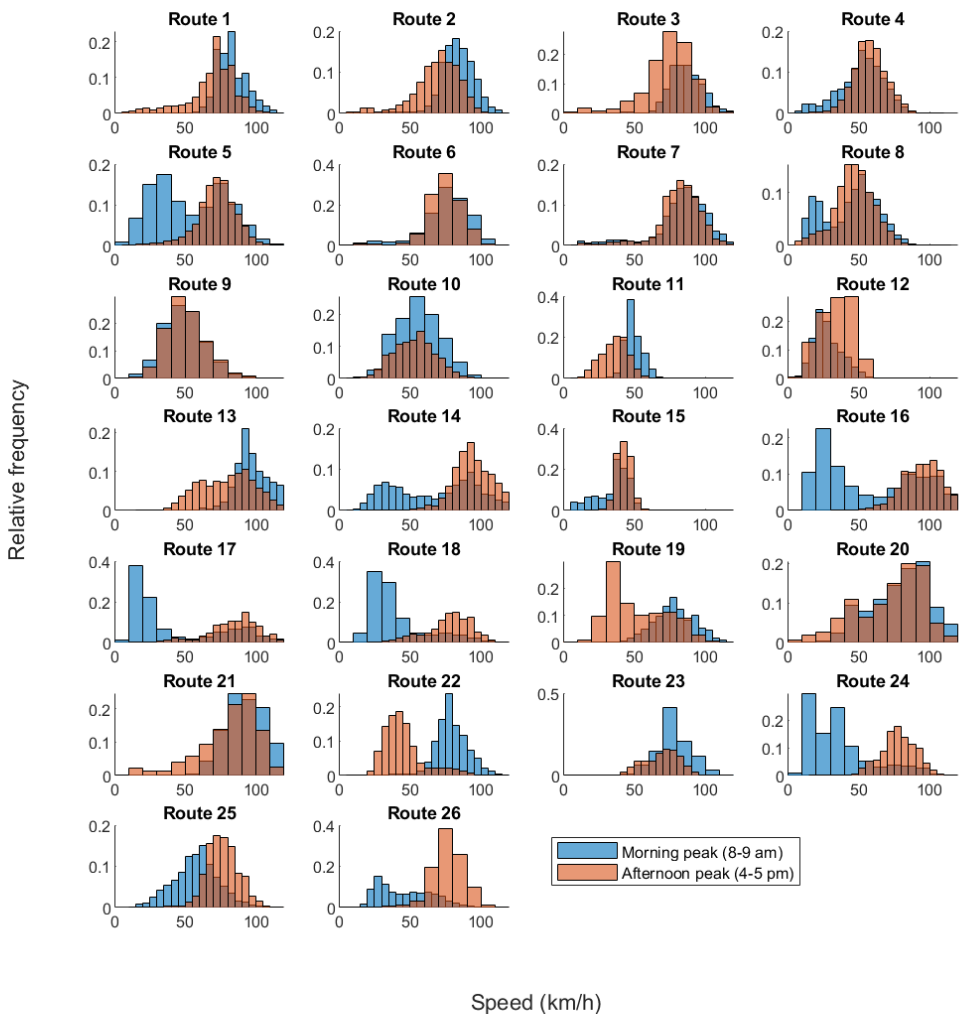

Figure 3: Space-mean speed distributions on case study routes.

3.2 Inter and intra-traveller speed variability

Figure 3 shows the observed space-mean speed distributions for all 26 routes and the morning and afternoon peaks separately, encapsulating both the inter-traveller and the intra-traveller variability. For each histogram the number of bins is selected based on Sturges' rule. Many routes show distinctly different speed distributions in the morning and the afternoon peaks, which indicates a clear difference in directionality of the traffic between the two periods. With a few exceptions such as routes 13 and 14, most such routes are on motorways. Some routes display bimodal distributions which suggests that traffic conditions vary between congested and uncongested.

The speed model in Section 2.3 is estimated separately for each route and time period combination. Table 4 shows the R^2 coefficient, the speed

Table 4: Speed variance components (km^2/h^2) and relative intra-traveller variance.

Route	AM						PM					
	R^2	σ^2_t	σ^2_d	σ^2_i	σ^2	RIV	R^2	σ^2_t	σ^2_d	σ^2_i	σ^2	RIV
1	0.52	62.11	2.99	1.02	61.06	0.51	0.80	32.54	138.99	63.94	71.67	0.87
2	0.70	59.52	30.60	10.51	50.62	0.61	0.80	29.90	102.59	61.07	59.75	0.84
3	0.67	111.84	4.78	1.54	64.38	0.39	0.77	51.62	122.87	52.72	83.95	0.80
4	0.75	41.36	82.10	27.10	73.19	0.82	0.38	22.72	22.75	10.21	83.70	0.82
5	0.87	30.88	378.89	91.86	91.47	0.95	0.66	52.78	36.46	30.87	74.42	0.68
6	0.83	51.50	211.54	35.48	63.20	0.86	0.75	38.82	71.13	8.13	47.28	0.75
7	0.78	77.86	165.87	65.28	106.61	0.81	0.76	71.36	104.14	30.81	78.55	0.72
8	0.78	30.79	119.75	70.44	84.50	0.90	0.59	24.33	66.16	25.87	82.34	0.86
9	0.59	93.90	13.03	5.16	90.48	0.54	0.38	80.59	6.71	4.19	123.92	0.62
10	0.52	103.99	5.55	3.54	114.93	0.54	0.56	88.97	16.99	10.76	97.97	0.56
11	0.63	20.40	1.92	0.71	16.53	0.48	0.79	10.61	32.94	13.19	20.70	0.83
12	0.82	19.08	43.82	19.13	21.38	0.82	0.76	19.53	20.98	33.41	34.79	0.74
13	0.49	96.25	1.98	1.24	99.06	0.52	0.74	41.70	109.51	60.16	104.66	0.84
14	0.83	32.13	365.55	166.89	163.43	0.96	0.54	71.65	25.28	23.77	103.95	0.64
15	0.88	9.13	57.03	21.78	17.88	0.91	0.64	11.83	2.67	0.20	10.76	0.53
16	0.93	35.05	393.68	257.58	97.50	0.96	0.61	99.37	17.66	1.97	87.42	0.51
17	0.90	12.95	355.47	239.64	115.67	0.98	0.49	177.75	9.89	2.07	180.50	0.52
18	0.88	13.87	127.74	110.70	48.74	0.95	0.66	135.31	6.60	1.72	98.42	0.44
19	0.62	108.56	2.86	2.55	87.01	0.46	0.85	40.85	170.29	66.85	76.16	0.86
20	0.49	205.75	2.39	1.70	229.69	0.53	0.61	148.29	109.10	51.88	229.65	0.70
21	0.76	143.27	0.49	1.67	75.52	0.35	0.82	62.64	232.32	44.39	95.69	0.84
22	0.54	66.34	5.18	1.11	63.65	0.51	0.78	6.95	160.91	18.50	48.29	0.97
23	0.62	80.36	5.33	0.95	57.02	0.44	0.85	80.67	27.20	4.01	30.95	0.42
24	0.86	45.90	270.87	95.95	77.34	0.91	0.65	68.76	5.57	7.31	60.05	0.49
25	0.76	18.81	130.15	46.08	60.93	0.93	0.55	59.07	6.80	2.70	62.08	0.54
26	0.89	12.73	185.34	52.29	44.36	0.96	0.69	62.81	5.13	1.04	47.28	0.45

variance components and the relative intra-traveller variance (RIV) according to (6). Some patterns can be observed from the results. For example, routes 14–18 have lower inter-traveller variance σ^2_t than intra-traveller variance components (σ^2_d, σ^2_d and σ^2) in the morning period but higher inter-traveller variance in the afternoon period. For routes 19–23 the opposite pattern holds. Table 1 shows that the former group are aligned towards the city center while the latter go out from the city center. Thus, there appears to be systematic variations in the nature of speed variability depending on the direction of the route and time period.

The RIV values range from 0.35 for route 21 in the morning to 0.98 for route 17 in the morning, with average values 0.71 in the morning peak and 0.69 in the afternoon peak. The differences in variance decomposition among the routes observed above is also manifested in the RIV values, which are higher when the intra-traveller variance components are larger and vice versa.

Table 5: Estimation results for relative intra-traveller variance model.

Parameter	Model 1			Model 2		
	Est.	SE	p-value	Est.	SE	p-value
Intercept	0.837	0.0602	4.42e-18	0.284	0.125	0.0275
Direction out	-0.376	0.0491	9.62e-10	-0.245	0.0486	7.90e-06
Region south	0.0680	0.0448	0.136	0.108	0.0377	0.00635
Roadtype motorway	0.0190	0.0466	0.685	0.0141	0.0382	0.713
PM peak	-0.285	0.0472	2.54e-7	-0.134	0.0498	0.0101
Direction out & PM peak	0.556	0.0695	2.93e-10	0.316	0.0756	0.000135
Congestion	—	—	—	0.397	0.0822	1.613e-05
Number of observations	52			52		
R^2	0.621			0.750		
F-stat. vs. constant model	15.1		9.47e-9	22.5		4.65e-12

To assess the influence of route characteristics and time periods, the linear regression model in equation (7) is estimated using the independent variables in Tables 1 and 2. Two model versions are estimated, without (Model 1) and with (Model 2) the congestion variable. Estimation results are shown in Table 5. In Model 1, there is no statistically significant difference between arterials and motorways. The relative intra-traveller variance in the morning peak is significantly higher towards the city center than out from it (p-value 9.6e-10). Further, the RIV is significantly higher in the morning peak than in the afternoon peak towards the city (p-value 2.5e-7), but significantly higher in the afternoon than in the morning out from the city (F-statistic 28.26, p-value 3.03e-6).

In Model 2, the relations found for Model 1 above are also present. Further, the level of congestion on the route has a significant positive impact on the relative intra-traveller travel time variance (p-value 1.6e-5). This implies that each driver has less influence on the chosen speed in congested traffic conditions. An F-test for the combined effect of route direction and peak period reveals that the relative intra-traveller variance out from the city in the afternoon peak and towards the city in the morning peak are not significantly different (F-statistic 2.34, p-value 0.133). Unlike in Model 1 the RIV is significantly higher in the south region than the north, but this result may not be robust.

An alternative model formulation using the traveller recurrence as explanatory variable finds no significant effect of this variable. This indicates that there is no clear link between the average familiarity of the travellers with the variance composition of the route, at least at the aggregate level.

The analysis reveals that the relative intra-traveller variance varies systematically with route and time period characteristics. Specifically, it tends to be higher in circumstances associated with heavy commuting traffic and congestion. This likely reflects that each driver has less influence on the chosen speed in such traffic conditions. Meanwhile, the type of road (motorway or arterial) has no impact, which suggests that the traffic characteristics of the routes are more important for the speed variance composition than the infrastructural characteristics.

4 Conclusions

This paper has investigated to what extent the vehicles traversing a route are recurring travellers depending on attributes such as road type, direction relative to the city center and time of day. Using data from Bluetooth and Wifi sensors over a three-month period, we have found that the average number of trips per vehicle ID is higher towards the city in the morning peak and out from the city in the afternoon, which is consistent with the knowledge that commute trips tend to have the highest regularity across days.

Motivated by the finding that a non-vanishing share of trips are made by recurrent travellers, the paper has proposed a model of route speed distributions that separates the variability into an inter-traveller component, consistent across days and time intervals, and an intra-traveller component. The intra-traveller component is further split into day-to-day, interval-to-interval and residual variability. Model estimation results show that the relative intra-traveller variance is significantly higher in the commute direction (towards the city in the morning and out from the city in the afternoon) and on routes with high congestion levels. This is consistent with the intuition that more congestion leads to lower flexibility in the speed choice.

Due to some rehashing of vehicle IDs in the case study data, the precise magnitudes of inter-traveller and intra-traveller variances are difficult to estimate. However, the results indicate that a distinction must be made between the intra-traveller variance, which corresponds to travel time uncertainty, and the total variance which is typically used in travel time reliability assessments. The relative magnitudes of the two terms vary systematically with route characteristics (direction and congestion) and time periods. The relations revealed in this paper may be used to estimate the relevant intra-traveller variance based on the total variance and readily available route attributes. Without this correction, the costs associated with travel time variability may be overestimated.

Further research is needed to assess the generality of the findings in varying settings. The robustness of the results should also be verified by applying the analysis to data that do not suffer from limitations of rehashed vehicle IDs. Other topics for future work include exploring the speed variance model structure and potentially extending the linear mixed model formulation proposed here, and extending the time frame of the analysis to incorporate seasonal variations. Finally, an interesting research direction is to investigate the causes for the inter-traveller speed variability and the relation between frequency of recurrence and speed.

Data Availability

The Bluetooth/Wifi travel time data used to support the findings of this study are proprietary of Trafik Stockholm and so cannot be made freely available. Requests for access to these data should be made to the author and are subject to approval by Trafik Stockholm.

Conflicts of Interest

The author declares no conflict of interest.

Funding Statement

This research was funded by the Swedish National Transport Administration under grant no. TRV 2018/16380.

Acknowledgments

The authors would like to thank Trafik Stockholm for kindly providing the data used in the study, and Maria Börjesson, Leonid Engelson and Karin Brundell-Freij for valuable comments on the work. The authors also wants to thank two anonymous reviewers for their helpful feedback on an earlier version of the paper.

References

[1] C. Carrion and D. Levinson, "Value of travel time reliability: A review of current evidence," *Transportation Research Part A*, vol. 46, pp. 720–741, 2012.

[2] G. C. de Jong and M. C. J. Bliemer, "On including travel time reliability of road traffic in appraisal," *Transportation Research Part A*, vol. 73, pp. 80–95, 2012.

[3] M. A. P. Taylor, "Travel through time: the story of research on travel time reliability," *Transportmetrica B: Transport Dynamics*, vol. 1, no. 3, pp. 174–194, 2013.

[4] R. B. Noland and K. A. Small, "Travel-time uncertainty, departure time choice, and the cost of morning commutes," *Transportation Research Record*, vol. 1493, pp. 150–158, 1995.

[5] M. Fosgerau and A. Karlström, "The value of reliability," *Transportation Research Part B*, vol. 44, no. 1, pp. 38–49, 2010.

[6] M. Fosgerau and L. Engelson, "The value of travel time variance," *Transportation Research Part B*, vol. 45, no. 1, pp. 1–8, 2011.

[7] J. Bates, J. Polak, P. Jones, and A. Cook, "The valuation of reliability for personal travel," *Transportation Research Part E*, vol. 37, no. 2–3, pp. 191–229, 2001.

[8] A. de Palma and N. Picard, "Route choice decision under travel time uncertainty," *Transportation Research Part A*, vol. 39, no. 4, pp. 295–324, 2005.

[9] N. Y. Tilahun and D. M. Levinson, "A moment of time: Reliability in route choice using stated preference," *Journal of Intelligent Transportation Systems*, vol. 14, no. 3, pp. 179–197, 2010.

[10] T. A. Li, Z. and D. A. Hensher, "Embedding risk attitudes in a scheduling model: Application to the study of commuting departure time," *Transportation Science*, vol. 46, no. 2, pp. 170–188, 2012.

[11] C. R. Bhat and R. Sardesi, "The impact of stop-making and travel time reliability on commute mode choice," *Transportation Research Part B*, vol. 40, no. 9, pp. 709–730, 2006.

[12] P. C. Devarasetty, M. Burris, and W. Douglass-Shaw, "The value of travel time and reliability- evidence from a stated preference survey and actual usage," *Transportation Research A*, vol. 46, no. 8, pp. 1227–1240, 2012.

[13] H. Tu, *Monitoring Travel Time Reliability on Freeways*. PhD thesis, Delft University of Technology, 2009.

[14] Y. C. Stogios, A. Brijmohan, H. Mahmassani, J. Kim, Y. Chen, and P. Vovsha, eds., *Incorporating Reliability Performance Measures into Operations and Planning Modeling Tools*. Washington, DC: The National Academies Press, 2014.

[15] S. Peer, C. Koopmans, and E. Verhoef, "Prediction of travel time variability for cost-benefit analysis," *Transportation Research Part A*, vol. 46, pp. 79–90, 2012.

[16] M. Ramezani and N. Geroliminis, "On the estimation of arterial route travel time distribution with markov chains," *Transportation Research Part B*, vol. 46, pp. 1576–1590, 2012.

[17] E. Jenelius and H. N. Koutsopoulos, "Travel time estimation for urban road networks using low frequency probe vehicle data," *Transportation Research Part B*, vol. 53, pp. 64–81, 2013.

[18] Z. Ma, H. N. Koutsopoulos, L. Ferreira, and M. Mesbah, "Estimation of trip travel time distribution using a generalized Markov chain approach," *Transportation Research Part C*, vol. 74, pp. 1–21, 2017.

[19] M. Fosgerau and D. Fukuda, "Valuing travel time variability: characteristics of the travel time distribution on an urban road," *Transportation Research Part C*, vol. 24, pp. 83–101, 2012.

[20] J. Kim and H. S. Mahmassani, "Compound Gamma representation for modeling travel time variability in a traffic network," *Transportation Research Part B*, vol. 80, pp. 40–63, 2015.

[21] A. J. Richardson and M. A. P. Taylor, "Travel time variability on commuter journeys," *High-Speed Ground Transportation*, vol. 6, pp. 77–99, 1978.

[22] Susilawati, M. A. P. Taylor, and S. V. C. Sekhar, "Distributions of travel time variability on urban roads," *Journal of Advanced Transportation*, vol. 47, no. 8, pp. 720–736, 2013.

[23] F. Guo, H. Rakha, and S. Park, "Multistate model for travel time reliability," *Transportation Research Record*, vol. 2188, no. 1, pp. 46–54, 2010.

[24] J. Kim and H. S. Mahmassani, "A finite mixture model of vehicle-to-vehicle and day-to-day variability of traffic network travel times," *Transportation Research Part C*, vol. 46, pp. 83–97, 2014.

[25] S. Yang and Y.-J. Wu, "Mixture models for fitting freeway travel time distributions and measuring travel time reliability," *Transportation Research Record*, vol. 2594, no. 1, pp. 95–106, 2016.

[26] M. Rahmani, E. Jenelius, and H. N. Koutsopoulos, "Non-parametric estimation of route travel time distributions from low-frequency floating car data," *Transportation Research Part C*, vol. 58B, pp. 343–362, 2015.

[27] J. Kwon, R. Hranac, K. Petty, and N. Compin, "Decomposition of travel time reliability into various sources," *Transportation Research Record*, vol. 2229, pp. 28–33, 2011.

[28] R. B. Noland and J. W. Polak, "Travel time variability: A review of theoretical and empirical issues," *Transport Reviews*, vol. 22, no. 1, pp. 39–54, 2002.

[29] M. Yildirimoglu, Y. Limniati, and N. Geroliminis, “Investigating empirical implications of hysteresis in day-to-day travel time variability,” *Transportation Research Part C*, vol. 55, pp. 340 – 350, 2015.

[30] Y. Hollander and R. Liu, “Estimation of the distribution of travel times by repeated simulation,” *Transportation Research Part C*, vol. 16, pp. 212–231, 2008.

[31] G. Goulet-Langlois, H. N. Koutsopoulos, Z. Zhao, and J. Zhao, “Measuring regularity of individual travel patterns,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 5, pp. 1583–1592, 2018.

[32] C. R. Bhat and R. Misra, “Discretionary activity time allocation of individuals between in-home and out-of-home and between weekdays and weekends,” *Transportation*, vol. 26, pp. 193–209, 1999.

[33] Y. O. Susilo and K. W. Axhausen, “Repetitions in individual daily activity-travel-location patterns: a study using the Herfindahl-Hirschman index,” *Transportation*, vol. 41, pp. 995–1011, 2014.

[34] P. Colonna, P. Intini, N. Berloco, and V. Ranieri, “The influence of memory on driving behavior: How route familiarity isrelated to speed choice. An on-road study,” *Safety Science*, vol. 82, pp. 456–468, 2016.

[35] C. E. McCulloch, S. R. Searle, and J. M. Neuhaus, *Generalized, Linear, and Mixed Models*. John Wiley & Sons, Inc., Hoboken, New Yersey, 2nd ed., 2008.