Bus Running Time Distributions on a Section Level

Beda Büchel

Francesco Corman

Institute for Transport Planning and Systems

May 2019

STRC | 19th Swiss Transport Research Conference
Monte Verità / Ascona, May 15 – 17, 2019

Institute for Transport Planning and Systems

Bus Running Time Distributions on a Section Level

Beda Büchel
IVT
ETH Zürich
CH-8093 Zürich
phone: +41-44-633 31 93
beda.buechel@ivt.baug.ethz.ch

Francesco Corman
IVT
ETH Zürich
CH-8093 Zürich
phone: +41-44-633 33 50
francesco.corman@ivt.baug.ethz.ch

May 2019

Abstract

Understanding the variability of public transport travel times is essential for various reasons, e.g. for gaining knowledge of the deteriorations and ameliorations in daily traffic, for providing adequate (real time) information to customers and for optimizing transit schedules.
This paper deals with this issue by modeling the day-to-day variability of running times of urban buses on a section level. The investigation bases on planned and actual arisen arrival and departure time data of a selected bus route in Zurich and is conducted by the aid of statistical distributions. In order to find the most appropriate distribution models for running times, we present methodologies to test and choose "good" distributions and fit conventional as well as mixture distributions.
Mixture distributions provide an improved fit to the data in terms of BIC and effect size, but the components need to be justified. This work also introduces an approach to fit meaningful component distributions.

Keywords

public transport, travel time variability, travel time distributions, section running time, mixture distribution

1 Introduction

The travel time of public buses is subject to variability, which arises due to the stochastic nature of various factors influencing bus operations. This variability is perceived inconvenient, as it introduces uncertainty and additional cost to travelers and operators. Therefore, the service quality of public transport depends to a great extent on the travel time variability. Hence it is important to have precise knowledge about the variability of bus travel times.

Statistical distributions can describe the nature and pattern of the travel time variability. However, the modeling of distributions of bus travel time gained of importance only recently, as a big amount of data is nowadays easily available and needs to be interpreted. Statistical distributions are a powerful tool, as they can describe the inherent variability in data with a limited amount of parameters. An appropriate choice of the travel time distribution is an essential input to simulation of transit systems, reliability analysis and prediction of delays.

Various studies put significant effort in fitting travel time distributions (e.g. Mazloumi *et al.* (2010), Xue *et al.* (2011), Kieu *et al.* (2014), or Ma *et al.* (2016)). The findings are strongly influenced by the type and aggregation of the tested data (Büchel and Corman, 2018).

In the manuscript at hand, we focus on section running times (i.e. the time that a bus takes for traveling form a station to the next station). We assess the fitting performance of traditional (unimodal) and mixture distributions to a urban bus line in Zurich. For the fitting of distributions we introduce a training and testing approach, as the "best fitting" distribution should be reproducible. Furthermore, we put special focus on applying mixture distributions, and in future work we try to find meaningful component distributions, so that the final distribution doesn't just show a good fit, but have meaningful components in addition.

2 Literature Review

Studies that investigate the probability distribution of bus running times on a section level are Taylor (1982), Xue *et al.* (2011), Cats *et al.* (2014), Kieu *et al.* (2014), and Ma *et al.* (2016). Generally, it is reported that running time on section levels are highly variable and depend on temporal data aggregation (Büchel and Corman, 2018). Often studies could not identify one distribution that provided the best fit on all considered sections. Normal (Taylor (1982), Xue *et al.* (2011)) or log-normal distributions (Cats *et al.* (2014), Kieu *et al.* (2014) are the most proposed distributions. When buses have similar characteristics as cars (e.g. acceleration, maximal speed) and share the same road space, the statistical modeling of buses and cars can be similar as well. Travel times of cars are often modeled right-tailed by the aid of log-normal distributions. Therefore, it is comprehensible that today, to statistically model running times of buses, mostly log-normal distributions are proposed

Van Lint and Van Zuylen (2005) identified in car traffic four phases (free-flow conditions, congestion onset, congestion, and congestion dissolve) that resulted in distinctively different distribution patterns of travel time variability. Based on this finding, distributions can be assumed to be a sum of component distributions representing different phases of traffic. Various studies have investigated the performance of mixture distributions in fitting travel time distributions compared to traditional models and suggested a superior performance (see e.g. Guo *et al.* (2010), or Susilawati *et al.* (2013)). Based on those findings, Ma *et al.* (2016) suggested the use of mixture distributions for the bus mode as well.

For modeling the distributions, Ma *et al.* (2016) considered Gaussian mixture distributions and set the maximum number of component distributions to be 3. The parameters can be related to free flow, recurrent, and non-recurrent service states. However, it remains unclear in this study how many components were actually used. Also Chen and Sun (2017) used Gaussian mixture distributions and described the same three service states. Their experimental results, however, pointed out that depending on the period 1-2 components (off-peak period) or 3-4 components (peak period) yield the best fit to the data used. Applying mixture distributions posed new questions. The components, which represent service states, should be explainable. Also the amount of needed components should be a priori determinable.

3 Data and Methodology

3.1 Data

For this study we consider an urban bus line in Zurich, Switzerland, belonging to the network of Verkehrsbetriebe Zurich (VBZ). The line 32, which is a trolleybus line that crosses the city, as seen in Fig. 1 is chosen for the analysis. This bus line is round 11 km long and runs from Zurich, Stassenverkehrsamt (STRV) to Zürich, Holzerhurd (HOLZ). It serves Goldbrunnenplatz, where a number of buses depart to the agglomeration of western Zurich. At Kalkbreite / Bhf Wiedikon it connects to the tram lines 2 and 3 and to the train network. It passes Langstrasse, a highly populated nightlife area and connects at Limmatplatz to tram 4, 13 and 17. At Bucheggplatz and at Glaubtenstrasse Süd it connects to bus lines. The line operates from 5 am to 1 am, with a minimal planned headway of 6 min. The service frequency varies from 10 bus/h (peak) to 6 bus/h (evening off-peak). This line is one of the most delayed line in the public transport network of Zurich. VBZ issues all planned and real occurred arrival and departure times of the buses and trams at the stations of the network. This data is publicly available and can be found on the open data portal of the city of Zurich (https://data.stadt-zuerich.ch). The data is captured to the accuracy of seconds. For this study, the data of 2018 is used. The raw data is filtered for regular trips on line 32, meaning special routes to and from depots / garages are neglected. It is important to note that due to this filter the occurrence of extreme values (i.e. high delays) need to be interpreted with care as buses with high delays are often subject to some dispatch measure.

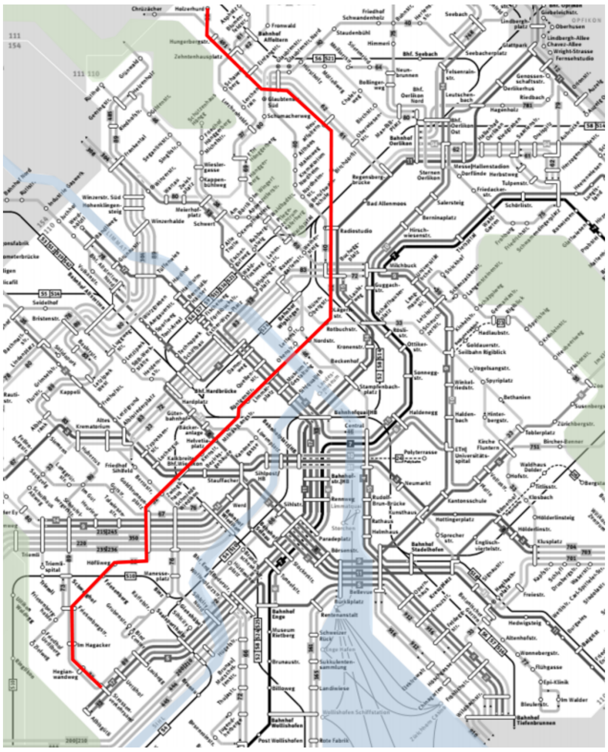

Figure 1: Public transport network of Zurich with highlighted line 32 (red).
Source: https://www.stadt-zuerich.ch/vbz/de/index/fahrplan/liniennetzplaene.html; 27.03.2018, adapted)

In Fig. 2 the characteristics of the sections of the line 32 are presented in the direction of travel Strassenverkehrsamt - Holzerhurd. The table informs about traffic signals (TraSig), pedestrian crossings (Ped) and no right of way conflicts (nRoW) on the sections. At all traffic lights, public transport is treated with priority, however, there are sometimes conflicts between different lines / modes of public transport.

Before analyzing the day-to-day variability of running times on a section level, we first have a preliminary look towards the available data. Therefore, we first analyze the variation of section running times over the course of the year. Fig. 3 shows the 25th, 50th, and 75th percentile

	From	To	TraSig	Ped	nRoW
1	STRV	HEGA	1	3	
2	HEGA	IHAG	1	3	
3	IHAG	FRIE	1	2	
4	FRIE	FRIB	1	2	
5	FRIB	HOEF	1	4	
6	HOEF	GOLP		5	
7	GOLP	ZWIN	1	2	1
8	ZWIN	KALK	2	3	
9	KALK	KERN		2	1
10	KERN	HELV		1	
11	HELV	MILA	3	1	
12	MILA	ROEN	2		
13	ROEN	LIMM		9	1
14	LIMM	NORS	2	1	
15	NORS	ROTB	1	1	
16	ROTB	LAEG	4	3	
17	LAEG	BUCH	1		
18	BUCH	RADI	1		
19	RADI	BIRD	2	3	
20	BIRD	NEUA	2	2	
21	NEUA	GLAU	1		
22	GLAU	EINF	3		
23	EINF	ZEHN	4		
24	ZEHN	HUNG	3		
25	HUNG	HOLZ			1

Figure 2: Characteristics of the analyzed sections of the bus route 32. Amounts of traffic signals (TraSig), Ped (Pedestrian Crossings), no right of ways (nRoW).

of all running times at the section Zehntenhausplatz - Hungerbergstrasse measured in 2018. The time series of the running time data is not a strictly stationary process, meaning that as its unconditional joint probability distribution changes to some extent when shifted in time. However, we can visually assume a wide-sense stationarity for the percentile values with the exception of the summer holiday period (July / August). Meaning that the first moment (i.e. the mean) and autocovariance are constant with respect to time. In the summer holiday period (July / August) are the 50th and 75th percentile values lower than during the rest of the year. Generally, the higher percentile values show more fluctuation.

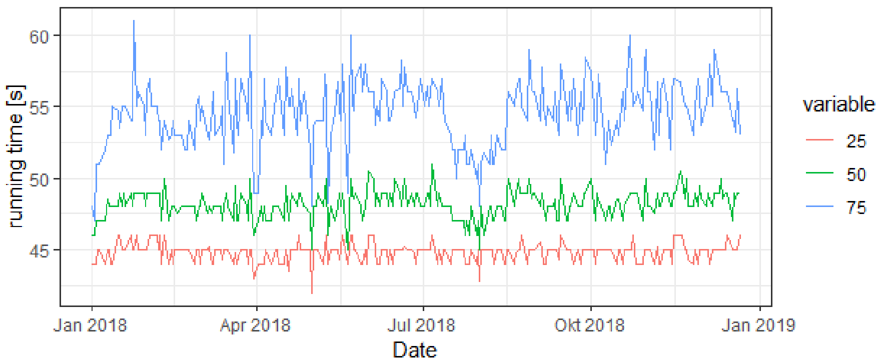

Figure 3: Percentile values of the running time over the course of the year 2018 at the section 24 (Zehntenhausplatz-Hungerbergstrasse)

In Fig. 4, the 25th, 50th, and 75th percentile of the running times at the section Friesenbergstrasse - Friesenberg is shown. The picture is quite different from the previous one. There are some times periods where the percentile values only fluctuate a little, whereas some other times there is quite a lot of fluctuation. Furthermore, the percentile values seem similar for periods of some weeks / months and there are distinct changes between such periods. The fluctuations are due to construction going on over the course of the year 2018. The construction lead to different restrictions to the buses. We can not assume the curves to be stationary.

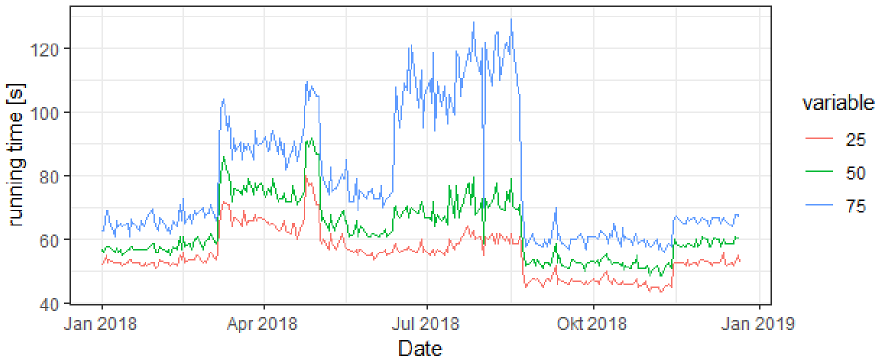

Figure 4: Percentile values of the running time over the course of the year 2018 at the sectionn 4 (Friesenbergstrasse-Friesenberg)

Next, we investigate the variation of running times over the course of a day. In order to do this, we look at running time data of sections whose yearly pattern can be assumed to be stationary. Fig. 5 shows the variation of the running times over the course of the day at the section Lägernstrasse - Bucheggplatz. The figure indicates, that the time of the day has almost no influence to the percentile values of the running times in this section.

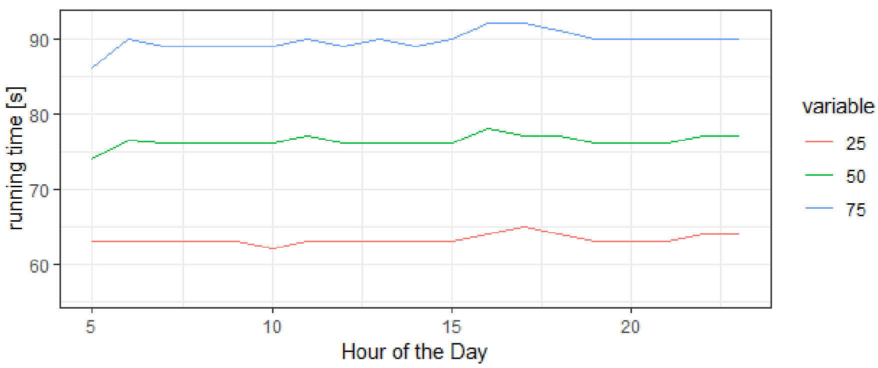

Figure 5: Course of the day section 17 (Lägernstrasse - Bucheggplatz)

In Fig. 6, on the other side, we see a very distinct daily pattern. In the morning and in the afternoon, the percentile values are remarkably higher than during the rest of the day. Especially for the higher percentiles (50th and 75th) the effect is clear.

The reasons for this two preliminary investigations is straight forward. We don't want to aggregate obviously different processes together. We want to find distributions that characterize more or less stationary events. As in many of the 25 sections of the investigated bus line there was construction going on, we focus in the following on the first 9 weeks of 2018, as in this

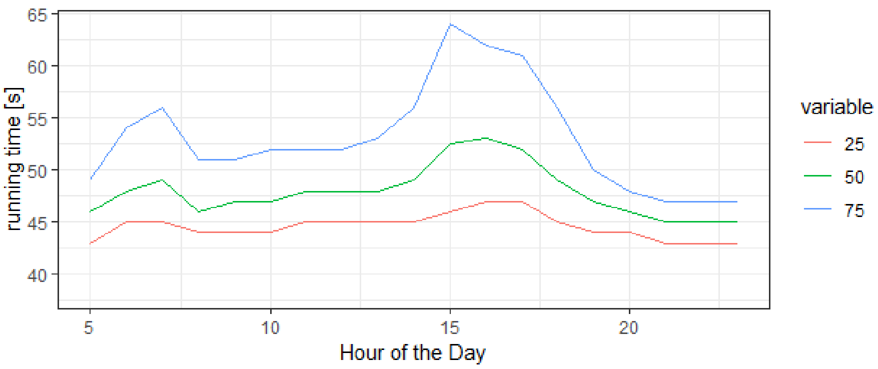

Figure 6: Course of the day section 24 (Zehntenhausplatz-Hungerbergstrasse)

weeks the variations seem stationary. We study the inter-peak (10:00-14:00) and the peak (16:00-18:00) running times.

3.2 Distribution Fitting

For each section, individual distributions and mixture distributions are fitted at peak and inter-peak conditions. For individual distributions, normal, log-normal, Cauchy, Weibull and logis distributions are chosen, as literature suggest that they perform the best under specific testing conditions. The parameters are estimated using the parametric maximum likelihood method. For mixture distributions, we fit models with normal and log-normal component distributions. Normal component distributions are chosen, as they are used in Ma *et al.* (2016) and Chen and Sun (2017), as well as in the vast majority of car traffic travel time studies. Log-normal component distributions are chosen, as log-normal distributions provide in general an improved fit compared with normal distributions and as they are right tailed. The parameters of the mixture distributions are estimated for 1, 2, and 3 component distributions.

In many studies, hypothesis tests are performed. We do not do this here due to a philosophical reason: We are not looking for a "true" distribution, but for a good and practical approximation to the data. With hypothesis tests, if is usually tested if the difference between the data and the candidate distribution might be random. If a lot of data is available, differences between data and candidate distributions are almost always significant. If hypothesis tests are carried out, they should consider that the theoretical distribution may not be fitted from the empirical data. Approaches to overcome this issues are parametric bootstrap approach (see e.g. Stute *et al.* (1993)), or a training-testing approach.

3.3 Assessing Fit

We assess the quality of the fit and decide on the best fitting distributions by two means: by the Bayesian information criterion (BIC) and by the effect size.

In order choose the best performing distribution out of a set of candidate distributions, it is often made use of the information creation technique. The Bayesian information criterion (BIC) measures the relative quality of a statistical model by trading off the complexity (by considering the number of parameters) and goodness-of-fit of the fitted distribution (by considering the maximized value of the log-Likelihood). The criteria is given by:

$$BIC = \ln(n)k - 2\ln(\hat{L}) \tag{1}$$

where k the number of parameters to be estimated, $\hat{L}$ the maximized value of the likelihood function of the estimated distribution, and n is the number of observations. The model with the the lowest criterion value is defined as the best model.

We calculate the BIC statistic in the following way: out of our dataset we choose $N = 500$ observation at random, fit the candidate distribution to it, and calculate the BIC statistic. We repeat this process $M = 100$ times and calculate the average BIC number.

Furthermore, we calculate the effect size. We measure the effect size by the Quantile Absolute Deviation (QAD). This measure aims to compare quantiles of the two distribution based on the the entire range of probabilities $[0, 1]$. The quantile deviation of two populations is the average absolute distance between the quantiles of two populations. Suppose F^{-1} and G^{-1} are quantile functions for the two statistical populations corresponding to the cumulative distribution functions, F and G respectively, then QAD is given by:

$$QAD = \int_0^1 \left|F^{-1}(p) - G^{-1}(p)\right| dp \tag{2}$$

The effect size is evaluated in the following: out of our dataset we choose $N_1 = 500$ and $N_2 = 500$ mutual exclusive observations at random. One can think of N_1 as training data and N_2 as validation data. The candidate distribution is fitted to the N_1 observation. We repeat this process $M = 100$ times and calculate the effect size based on fitted distribution and the data N_2. An example of the calculation is given in Fig. 7. The red curve is the quantile function of the empirical data (N_2), the green curve the is the fitted distribution based on data N_1, and finally the blue curve is the difference. The integral of the blue curve is the QAD.

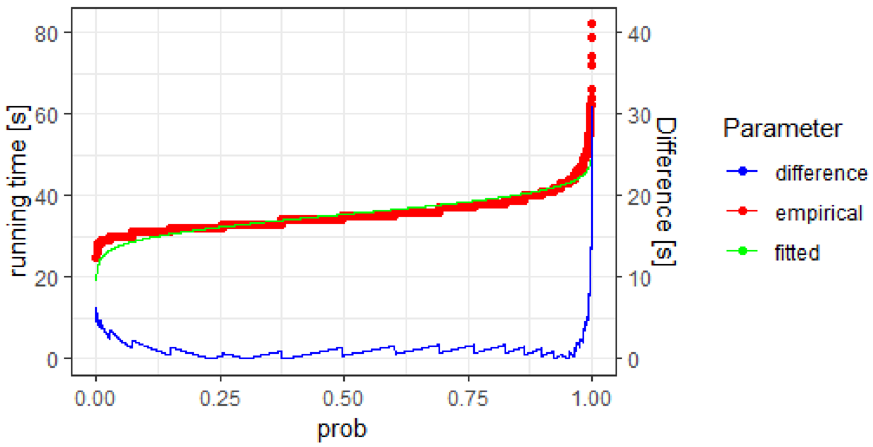

Figure 7: Example of determination of QAD

4 Results

4.1 Individual Distributions

In a first step, we fit individual (=non mixture) distributions to the section running times of bus line 32. We look at the inter peak (10:00 - 14:00) running times of all sections belonging to this route. The resulting BIC-values and QAD-values are shown in Fig. 8. The BIC value is a criterion for model detection, which is based on the likelihood function. The model with the lowest BIC is preferred. QAD, on the other side, is a measure of effect size. The smallest effect size is in general preferred.

Considering BIC values, the lowest big values for each section is represented in the darkest green. According to Kass and Raftery (1995), a difference in BIC values between 0 and 2 is considered as *"not worth more than a bare mention"*, and a difference in BIC values between 2 and 6 is considered as *"postive"*, and a difference in BIC values between 6 and 10 is considered as *strong"*. As the calculated BIC value is a mean value, it can be expected to grow linearly with more sample data. Therefore, the exact value as well as differences depend on the sample size. Hence, we don't look at the absolute difference in BIC but at relative differences. Candidate distributions with a difference in BIC values to the lowest one of less than 0.1% are shown in a lighter green, and such with differences in BIC values to the lowest one between 0.1% and 1% shown in the lightest green. We see that the log-normal distribution is in most of the times the best ranked distribution (in 19 out of 25 cases). If it is not the best ranked distribution, it has in four out of six cases a difference in BIC of less than 1% to the best performing distribution. Often, when the log-normal distribution is not the top ranked distribution, the logis distribution is the best performing distribution. Additionally, the logis distribution is often the second best ranked distribution. Rarely the Chauchy is best ranked, but mostly the difference in BIC value between the Cauchy candidate distribution and the best ranked distribution is more than 1%. It seems not adequate to suggest its use for all sections. On the right size of the Fig. 8 we see the QAD value (=effect size) of the unimodal candidate distributions at inter-peak conditions. The image is similar - log-normal seems to perform best in average, followed by the logis distribution. Cauchy performs quite bad in terms of the measure of effect size, this as it provides a very bad fit for extreme (very high and very low) values. The color-code of the effect size is a continuous scale from 0s (dark green) to 4s (white). The effect size is remarkably small in many cases.

	BIC					QAD				
	norm	lnorm	cauchy	weibull	logis	norm	lnorm	cauchy	weibull	logis
1	2933	2906	3028	3086	2908	0.70	0.57	5.80	1.72	0.62
2	2796	2767	2851	2954	2764	0.74	0.62	4.56	1.63	0.65
3	3101	3055	3184	3268	3067	0.92	0.68	6.75	2.14	0.80
4	3849	3769	3943	3938	3832	2.66	1.84	14.99	3.76	2.53
5	3435	3260	3137	3707	3208	2.88	2.02	5.42	5.83	1.76
6	2994	2937	3026	3214	2927	0.94	0.72	5.67	2.51	0.73
7	3897	3825	4004	3988	3874	2.21	1.32	15.84	3.59	1.99
8	4503	4469	4687	4551	4505	2.42	1.80	33.88	3.78	2.57
9	3613	3625	3748	3646	3610	0.88	1.04	11.64	1.42	0.83
10	2622	2583	2712	2778	2593	0.61	0.49	4.23	1.32	0.56
11	4307	4206	4296	4418	4249	4.74	3.21	19.39	7.20	3.91
12	3701	3601	3709	3814	3645	2.48	1.63	11.03	3.78	2.01
13	4116	4041	4199	4213	4087	3.10	2.03	18.97	4.82	2.85
14	3737	3668	3814	3858	3714	2.34	1.76	12.94	3.82	2.23
15	3204	3168	3379	3294	3210	0.91	0.65	9.00	1.58	0.93
16	3642	3609	3803	3748	3644	1.37	1.01	13.49	2.67	1.36
17	4243	4215	4468	4281	4259	2.02	2.02	28.59	2.48	2.52
18	3666	3572	3639	3842	3580	2.37	1.68	9.91	4.65	1.77
19	3657	3620	3772	3779	3642	1.61	1.18	12.41	3.11	1.49
20	3225	3167	3220	3482	3147	1.31	1.02	6.58	3.61	0.99
21	3265	3180	3203	3453	3185	1.87	1.46	6.29	3.37	1.52
22	3273	3243	3429	3409	3265	0.84	0.64	9.35	2.04	0.84
23	3820	3759	3894	3944	3793	2.47	1.83	13.94	4.12	2.26
24	3320	3210	3264	3529	3215	1.79	1.23	7.04	3.70	1.34
25	4376	4093	4049	4457	4150	7.86	4.61	13.70	10.45	4.97

Figure 8: BIC- and QAD value of individual candidate distributions at inter-peak conditions

We perform same procedure for the peak period (16:00-18:00) running times of all sections. Still log-normal candidate distributions seem to perform in general the best. It is top ranked in 18 out of 25 cases. The sections, in which the log-normal distribution is not the best performing distribution at peak period are similar to the sections in which the log-normal distribution is not the best performing distribution at inter-peak period. Also Cauchy distribution is sometimes the top-ranked distribution in terms of BIC. However, in general it lays underneath the other candidate distributions. Logis is again often the second best distribution after log-normal, however the differences in BIC compared with the best performing distribution are bigger than during inter-peak conditions. In terms of effect size, log-normal performs best followed by logis. Cauchy, again, doesn't perform good, due to big differences at very high and low percentile values. However, it is possible that with an other measure of effect size, cauchy could perform better. Compared to the inter-peak conditions, the effect size is for sections 11-14 and 21-25 much higher, whereas for the other sections the difference is not that high. This is not astonishing, as in the preliminary investigates we have seen that these sections have a higher percentile running times at peak conditions.

	BIC					QAD				
	norm	lnorm	cauchy	weibull	logis	norm	lnorm	cauchy	weibull	logis
1	3073	3029	3162	3242	3038	0.87	0.66	6.72	2.09	0.77
2	2775	2742	2869	2934	2749	0.69	0.56	4.95	1.53	0.63
3	3223	3145	3208	3412	3148	1.52	1.14	6.50	3.03	1.18
4	3790	3724	3920	3875	3784	2.19	1.52	15.21	3.13	2.16
5	3469	3302	3191	3721	3257	2.94	2.08	5.68	5.70	1.78
6	2942	2913	3065	3069	2933	0.75	0.60	6.10	1.54	0.74
7	4004	3959	4159	4070	4000	1.82	1.03	18.99	3.00	1.73
8	4519	4494	4702	4559	4526	2.37	1.72	33.26	3.48	2.55
9	3590	3611	3733	3614	3592	0.84	1.12	11.45	1.18	0.79
10	2678	2639	2796	2822	2659	0.61	0.49	4.82	1.28	0.58
11	4973	4803	4917	5005	4911	10.55	6.36	36.34	11.95	9.35
12	5584	5399	5746	5468	5582	15.63	9.96	102.3	9.77	16.36
13	5429	5355	5595	5420	5418	7.91	5.25	84.53	7.39	7.92
14	4200	4076	4189	4306	4132	4.51	2.99	18.10	6.44	3.68
15	3215	3167	3350	3319	3211	1.22	0.90	8.51	1.96	1.14
16	3659	3619	3794	3788	3648	1.55	1.15	13.09	3.14	1.49
17	4236	4196	4458	4277	4254	2.46	1.93	28.59	2.99	2.80
18	4056	3933	3991	4167	3981	4.18	2.88	13.92	6.00	3.38
19	3674	3646	3831	3785	3672	1.26	0.90	13.79	2.74	1.23
20	3517	3441	3487	3722	3437	2.20	1.72	8.53	4.34	1.72
21	4493	4205	4113	4511	4321	9.59	6.11	14.78	10.55	7.27
22	4329	3780	3599	4497	3719	9.80	3.91	8.73	15.84	2.91
23	5240	4800	4868	5192	4906	19.89	8.50	36.51	21.54	9.26
24	4366	3970	3953	4408	4031	8.47	4.04	13.12	10.80	4.18
25	4596	4197	4102	4644	4253	11.08	5.78	13.74	14.03	5.85

Figure 9: BIC- and QAD value of individual candidate distributions at peak conditions

4.2 Mixture Distributions

Then we fit mixture distribution for normal and log-normal distributions. Normal is chosen, as it is often used in literature (see Ma *et al.* (2016),Chen and Sun (2017)). Log-normal is chosen, as log-normal distributions perform in general the best out of the unimodal distributions (see last section). The following figures show the BIC and QAD values for the candidate distributions at all sections, for peak and inter-peak conditions.

The results for inter-peak conditions are shown in Fig. 10. Generally, it can be said that log-normal clearly outperform normal mixture distributions with the same amount of parameters, which we have already seen in the last section for one-component distributions. This holds true for peak and inter-peak conditions.

	BIC						QAD					
	n(1)	ln(1)	n(2)	ln(2)	n(3)	ln(3)	n(1)	ln(1)	n(2)	ln(2)	n(3)	ln(3)
1	2931	2904	2896	2895	2915	2914	0.70	0.57	0.43	0.43	0.43	0.43
2	2794	2764	2727	2735	2744	2744	0.74	0.62	0.39	0.39	0.39	0.39
3	3098	3053	3036	3030	3045	3043	0.92	0.68	0.48	0.46	0.46	0.45
4	3848	3769	3720	3704	3723	3717	2.66	1.84	0.97	0.91	0.92	0.89
5	3437	3261	3073	3046	3045	3039	2.88	2.02	0.93	0.79	0.68	0.65
6	2995	2936	2901	2894	2908	2906	0.94	0.72	0.49	0.46	0.45	0.45
7	3896	3823	3826	3813	3827	3821	2.21	1.32	1.04	0.91	0.92	0.89
8	4503	4470	4496	4478	4488	4484	2.42	1.80	2.01	1.76	1.59	1.60
9	3613	3624	3619	3611	3615	3612	0.88	1.04	0.75	0.70	0.68	0.65
10	2623	2584	2574	2570	2584	2583	0.61	0.49	0.37	0.36	0.35	0.35
11	4308	4207	4154	4147	4165	4159	4.74	3.21	1.36	1.24	1.30	1.25
12	3702	3603	3563	3551	3569	3563	2.48	1.6	0.84	0.74	0.77	0.75
13	4117	4043	4035	4023	4036	4032	3.10	2.03	1.26	1.10	1.07	1.03
14	3736	3667	3604	3591	3605	3602	2.34	1.76	0.88	0.77	0.73	0.72
15	3203	3168	3163	3155	3170	3168	0.91	0.65	0.50	0.47	0.47	0.46
16	3641	3608	3604	3597	3612	3610	1.37	1.01	0.74	0.69	0.69	0.68
17	4244	4216	4229	4205	4195	4190	2.02	2.02	1.91	1.56	1.23	1.25
18	3667	3573	3510	3497	3510	3509	2.37	1.68	0.81	0.70	0.69	0.68
19	3659	3622	3612	3608	3624	3623	1.61	1.18	0.73	0.70	0.68	0.68
20	3221	3164	3104	3098	3110	3110	1.31	1.02	0.60	0.56	0.55	0.56
21	3264	3179	3066	3055	3062	3060	1.87	1.46	0.60	0.54	0.52	0.52
22	3273	3242	3250	3242	3251	3250	0.84	0.64	0.60	0.55	0.53	0.52
23	3822	3761	3720	3711	3722	3721	2.47	1.83	0.93	0.83	0.80	0.79
24	3320	3209	3140	3118	3125	3123	1.79	1.23	0.80	0.69	0.59	0.57
25	4379	4094	3976	3928	3926	3915	7.86	4.61	2.64	1.98	1.66	1.53

Figure 10: BIC- and QAD value of mixture candidate distributions at inter-peak conditions

For inter-peak as well as for peak conditions, log-normal mixture distributions with two components seem to be often a good fit in terms of BIC. However, sometimes also a 3-component model has the lowest BIC-value. In summary, it can be stated that at peak conditions more components seem adequate to model the data.

Considering QAD values, using more components reduces the effect size, which could have been expected. During inter-peak conditions using 1 or 2 components gives already a quite low QAD. During peak conditions, on the other side, sections 21-25 need more components in order to have a low QAD value.

	BIC						QAD					
	n(1)	ln(1)	n(2)	ln(2)	n(3)	ln(3)	n(1)	ln(1)	n(2)	ln(2)	n(3)	ln(3)
1	3073	3029	3025	3018	3035	3030	0.87	0.66	0.49	0.45	0.46	0.46
2	2775	2742	2727	2724	2742	2742	0.69	0.56	0.38	0.37	0.38	0.38
3	3223	3145	3083	3075	3089	3087	1.52	1.14	0.55	0.52	0.53	0.51
4	3790	3724	3693	3678	3694	3688	2.19	1.52	0.92	0.83	0.86	0.82
5	3469	3302	3121	3100	3088	3083	2.94	2.08	0.94	0.84	0.68	0.67
6	2942	2913	2899	2897	2914	2913	0.75	0.60	0.40	0.40	0.39	0.40
7	4004	3959	3975	3966	3985	3982	1.82	1.03	1.12	0.99	1.04	1.01
8	4519	4494	4516	4503	4518	4516	2.37	1.72	1.96	1.67	1.65	1.64
9	3590	3611	3604	3584	3591	3590	0.84	1.12	0.82	0.71	0.68	0.66
10	2678	2639	2635	2629	2641	2639	0.61	0.49	0.39	0.37	0.34	0.34
11	4973	4803	4731	4702	4721	4714	10.55	6.36	3.24	2.50	2.47	2.42
12	5584	5399	5314	5257	5301	5265	15.63	10.0	6.50	4.71	5.19	4.77
13	5429	5355	5397	5365	5370	5359	7.91	5.25	6.36	5.10	4.33	4.06
14	4200	4076	4027	4005	4011	4002	4.51	2.99	1.52	1.26	1.28	1.19
15	3215	3167	3137	3129	3132	3132	1.22	0.90	0.55	0.55	0.48	0.48
16	3659	3619	3616	3609	3619	3617	1.55	1.15	0.77	0.73	0.67	0.66
17	4236	4196	4196	4187	4199	4195	2.46	1.93	1.41	1.46	1.29	1.27
18	4056	3933	3848	3837	3856	3850	4.18	2.88	1.05	0.91	0.97	0.92
19	3674	3646	3649	3645	3660	3659	1.26	0.90	0.78	0.73	0.74	0.73
20	3517	3441	3359	3349	3361	3356	2.20	1.72	0.74	0.68	0.68	0.66
21	4493	4205	3985	3924	3932	3916	9.59	6.11	2.72	1.99	1.77	1.56
22	4329	3780	3549	3566	3507	3492	9.80	3.91	2.40	1.85	1.37	1.23
23	5240	4800	4712	4650	4681	4640	19.89	8.50	5.82	4.83	5.21	4.30
24	4366	3970	3871	3806	3797	3784	8.47	4.04	2.79	2.11	1.70	1.68
25	4596	4197	4045	3992	4015	3996	11.08	5.78	3.22	2.31	2.36	2.03

Figure 11: BIC- and QAD value of mixture candidate distributions at peak conditions

5 Discussion & Further Work

The empirical data does not follow any of the candidate distributions. We can not expect any of the candidate distribution to be the *"true"* distribution. A lot of parameters have an influence on the running time distribution, without (exact) information of those parameters the *"true"* distribution cannot be identified. However, there are candidate distributions which provide a good approximation to the empirical distribution. Considering conventional (unimodal) distributions, log-normal distributions perform well, both in terms of BIC and effect size. By using multiple log-normal component distributions the fit can be improved significantly, which holds specially true for peak running times.

The effect size is an adequate tool to understand the quality of fit of candidate distributions. In comparison to hypothesis tests, it does not depend on the amount of data used. Furthermore, if QAD is used, the effect size has some immediate real life meaning, as it can be interpreted as the average deviation between the candidate distribution and the data. However, it is important to note that a major part of the QAD value usually comes from the tails of the distribution (i.e. low and high running time values). This being said, it remains an open question whats an appropriate effect size is. This depends on the study and is to some extent subjective. In this study we considered QAD values smaller than 4s as good (see color code in Fig. 8 - Fig. 11).

The model selection based on the BIC value has the problem of over fitting. The more data is used to estimate the distribution, the higher is the amount of components of the model with the lowest BIC. A better methodology to choose the best model could be to compare the likelihood of the fitted distribution with additional data.

In this study we use a bootstrap approach in order to fit distributions. In the era of big data, it makes sense do so, as the main requirement to a distribution is its reproducibility. Especially when it comes to fitting mixture distributions to data, given a different subset of sampled data, the parameters of distributions vary remarkably. Therefore, further work should address the development of a methodology that fits robust distributions, which are reproducible with additional data.

We show, that log-normal mixture distribution with one or two are a good assumption for off-peak conditions. For off-peak conditions, more components improve the fitting significantly, considering BIC and effect size. By comparing the required modes with the characteristics of the sections (see Fig.2), no clear conclusion can be drawn at the moment. This, however could help understand the modes. Furthermore, the components of the distribution should be meaningful. They should not only be introduced, but be comprehensible and add value in understanding the bus running process.

The data is currently only available for the bus lines of Zurich. The bus system in Zurich has, compared with other cites, a relatively high punctuality and the running times don't vary vastly between peak and inter-peak conditions. Having data from less punctual bus systems would improve the understanding.

In a next step two working packages are addressed: First, we aim to identify characteristic distributions for "normal" and "delayed" section running times, so that we can assign meaning to the components. In a second step, we aim to understand the running time correlation from section to section, putting special focus on partial correlations between the component distributions.

6 Conclusion

We have shown that considering individual (unimodal) distributions, the log-normal distribution provides the best fit to section running time data in general. However, the best fitting distribution depends on the section and the time of the day, which aligns with previous findings. Especially during peak period, introducing additional components, resulting in mixture distributions, provides an improvement to the fit of the distributions. The result indicated, that mixture models with two component distributions describe the inter-peak running times well. For peak running times, additional components improve the fit for some sections. The assessment of the fit was done by BIC, as well as by QAD, which measures the effect size.

In a next step the work should focus on finding meaningful component distributions. It should be investigated if characteristic distributions on a section level can be found. Furthermore it should be investigated, how section running times are correlated with respect to the component distributions.

7 References

Büchel, B. and F. Corman (2018) Modelling Probability Distributions of Public Transport Travel Time Components, paper presented at the *Swiss Transport Research Conference*, 29.

Cats, O., F. M. Rufi and H. N. Koutsopoulos (2014) Optimizing the number and location of time point stops, *Public Transport*, **6** (3) 215–235, October 2014, ISSN 1866-749X, 1613-7159.

Chen, S. and D. Sun (2017) A multistate-based travel time schedule model for fixed transit route, *Transportation Letters*, 1–10, January 2017, ISSN 1942-7867, 1942-7875.

Guo, F., H. Rakha and S. Park (2010) Multistate Model for Travel Time Reliability, *Transportation Research Record: Journal of the Transportation Research Board*, **2188** (1) 46–54, January 2010, ISSN 0361-1981, 2169-4052.

Kass, R. E. and A. E. Raftery (1995) Bayes Factors, *Journal of the American Statistical Association*, **90** (430) 773–795, June 1995, ISSN 0162-1459, 1537-274X.

Kieu, L. M., A. Bhaskar and E. Chung (2014) Establishing definitions and modeling public transport travel time variability.

Ma, Z., L. Ferreira, M. Mesbah and S. Zhu (2016) Modeling distributions of travel time variability for bus operations: Bus Travel Time Distribution Model, *Journal of Advanced Transportation*, **50** (1) 6–24, January 2016, ISSN 01976729.

Mazloumi, E., G. Currie and G. Rose (2010) Using GPS data to gain insight into public transport travel time variability, *Journal of Transportation Engineering*, **136** (7) 623–631.

Stute, W., W. G. Manteiga and M. P. Quindimil (1993) Bootstrap based goodness-of-fit-tests, *Metrika*, **40** (1) 243–256, December 1993, ISSN 0026-1335, 1435-926X.

Susilawati, S., M. A. P. Taylor and S. V. C. Somenahalli (2013) Distributions of travel time variability on urban roads: TRAVEL TIME VARIABILITY ON URBAN ROADS, *Journal of Advanced Transportation*, **47** (8) 720–736, December 2013, ISSN 01976729.

Taylor, M. A. P. (1982) Travel Time Variability - The Case of Two Public Modes, *Transportation Science*, November 1982.

Van Lint, J. and H. Van Zuylen (2005) Monitoring and predicting freeway travel time reliability: Using width and skew of day-to-day travel time distribution, *Transportation Research Record: Journal of the Transportation Research Board*, (1917) 54–62.

Xue, Y., J. Jin, J. Lai, B. Ran and D. Yang (2011) Empirical characteristics of transit travel time distribution for commuting routes, *Transportation Research Board 90th Annual Meeting*.