European Electric Vehicle Congress
Brussels, Belgium, 3rd – 5th December 2014

On the distribution of individual daily vehicle driving distances

Patrick Plötz[1], Niklas Jakobsson[2], Frances Sprei[2], Sten Karlsson[2]
[1]*Fraunhofer Institute for Systems and Innovation Research ISI, Breslauer Str. 48, 76139 Karlsruhe, Germany, patrick.ploetz@isi.fraunhofer.de*
[2]*Chalmers University of Technology, Energy and Environment, 412 96 Göteborg, Sweden*

Abstract

The vehicle kilometres travelled (VKT) by individual passenger cars vary strongly between days. This is important for electric vehicles since trips larger than the electric range reduce their utility. Here we analyse different distribution functions for the variation in daily VKT with three sets of travel data. In contrast to the literature, no analysed distribution stands out best. We apply our findings for the distribution functions to estimate the number of days per year with driving distance larger than 100 km and find that the distributions differ in their predictions of the number of such days.

Keywords: daily driving, limited range, battery electric vehicle, driving data

1 Introduction

The limited electric driving range of battery electric vehicles (EVs) is a major hurdle for many consumers and the electric range of plug-in hybrid EVs strongly impacts the utility of EVs [1]. Taking into account the multitude of vehicle usage scenarios, a major question is: What range is required for EVs? Accordingly, the variation in distances travelled by one individual on different days of the year is important for the utility of EVs. The distribution of daily vehicle kilometres travelled (VKT) has thus found new attention by the market introduction of EVs and GPS based measurements [2–6]. Ideally the distribution of VKT should be based on empirical data from long observation times, however this may be cumbersome and expensive to realise. Shorter measurement periods and travel diaries can be helpful if distributions are fit to them. However, several distribution functions have been discussed without conclusive evidence. There seems to be agreement that both individual-longitudinal and cross-sectional VKT distributions are peaked and right skewed. Greene (1985) and Lin et al. (2012) analyse the Weibull, log-normal and Gamma distribution and argue that the Gamma distribution is most suitable [2, 3]. On the other hand, Plötz (2012) identifies the log-normal as a good approximation to the cross-sectional distribution of daily VKT and Blum (2013) finds the log-normal distribution to be the best fit for the majority of analysed driving profiles. Each of these studies only looks at one specific data set and thus it may be hard to generalize the conclusions; e.g., the observation time might influence the choice of best distribution.

Here, we analyse daily VKT distributions and their impact on the likelihood of rare long-distance trips. The present work differs from previous studies in several aspects. First, we compare different data sets from different countries and with different observation times. To these data sets, three different distributions are fitted as well as extrapolations for the longer data sets We are thus able to draw a more comprehensive picture of longitudinal daily VKT. Second, we analyse several goodness-of-fit (GoF) measures. Third, we combine the abstract analysis of the best-fitting distribution functions with consequences for the utility of electric vehicles and their limited range.

2 Data and Methods

2.1 Driving data

We use three data sets to analyse the goodness of fit (GoF) of different distributions. The data sets comprise vehicle motion from Germany [9], Sweden [10] and Canada [11] and the average observation periods range from 7 to more than 200 days. Each data set contains all daily VKT of each individual user where the number of users in the data sets ranges from 75 for the Canadian data set to 6,339 for the German data. The different data sets are summarised in Table 1. We thus compare daily driving data from three different countries. The average daily and annual VKT are not the same in these countries, yet the variation and fluctuation of daily VKT should follow similar patterns in different countries since they are caused by living and working conditions as well general vehicle usage patterns. Since all three countries are developed western nations with high vehicle ownership ratios, we expect these countries to be comparable with respect to the distribution of daily VKT.

Table 1: Summary of data sets.

Name of data set	Mobility Panel	SCMD	Winnipeg data
Location	Germany	Sweden	Canada
Method	Questionnaire	GPS	GPS
Sample Size	6339	429	75
Avg. obser-vation period	7 days	58 days	216 days

The variability of driving behaviour between different individuals is an important factor in the different data sets. Table 2 contains summary statistics of the driving behaviour on the individual user level. The large sample from Germany contains a very broad range of driving with annual VKT ranging from 15 to more than 200,000 km per year. Please note that all German driving profiles contain exactly seven days of observation by design (mobility questionnaires were used). Furthermore, share of driving days is typically higher than in the Canadian data set. The smaller Winnipeg data set also contains a noteworthy range of vehicle usages. However, the mean and median VKT for the Canadian data set is rather below the north-American average, possibly due to the geographically isolated position of the City of Winnipeg.

Table 2: Summary statistics of driving behaviour in the data sets

	Min	0.25	Median	Mean	0.75	Max
SCMD data (N = 429)						
Observation period [days]	30	51	59	587	64	147
Share of driving days	0.21	0.67	0.83	0.8	0.96	1
Average daily VKT [km]	6.9	38.36	51.9	57.1	72.3	172.0
Average annual VKT [km]	1,715	9,570	14,933	17,154	21,903	71,347
Winnipeg data (N = 75)						
Observation period [days]	3	108	238	216	325	448
Share of driving days	0.13	0,.58	0.67	0.67	0.81	1
Average daily VKT [km]	13.6	22.5	28.9	33.2	39.3	91.8
Average annual VKT [km]	1,640	5,300	7,330	9,280	11,260	30,400
Mobility panel data (N = 6339)						
Observation period [days]	Seven for all drivers by design					
Share of driving days	1/7	6/7	7	0.92	7	7
average daily VKT [km]	0.29	22	28.3	50.6	65	469
Annual VKT [km]	15	8,000	12,000	13,830	17,000	260,000

The Swedish Car Movement Data (SCMD) consists of GPS measurements of 429 privately driven cars in western Sweden. Measurements were evenly distributed over the years 2010-2012. The cars were randomly sampled from the Swedish vehicle registry with an age restriction on the car of maximum 8 years. Western Sweden is representative for Sweden in general in terms of urban and rural areas, city sizes and population density. The sample is representative in terms of car size and car fuel type. In relation to the household of the cars there is a slight overrepresentation of cars being a first car in a household compared to the national average due to the age inclusion criteria in the sampling. Similarly the cars in the data have a higher average annual VKT of 17,154 km compared to about 13,000 km for the national average also due to the younger age of the cars compared to the national average. With regards to the age of the drivers, there is a slight over-representation of senior citizens. A full description of the data including pre-processing is available in [10].

2.2 Distribution functions and Goodness-of-fit (GoF) measures

The daily VKT r_l are assumed to be independent and identically distributed (iid) random variables. Similar to Lin et al. (2012), we analyse three right-skewed two-parameter distributions. These distributions are the log-normal $f(r) = \exp[-(\ln r-\mu)^2/(2\sigma^2)]/(r\sqrt{2\pi}\sigma)$, Gamma $f(r) = r^{k-1}\exp[-r/\theta]/(\Gamma(k)\theta^k)$, and Weibull $f(r) = (k/\lambda)(r/\lambda)^{k-1}\exp[-(r/\lambda)^k]$ distribution. All three distribution functions assign zero probability to daily VKT of zero km length and fall off slower than a normal distribution for long distances.

For each individual vehicle, the parameters for scale and variation of the longitudinal daily VKT are obtained by maximum likelihood estimates for all three distribution functions.

Several GoF measures are applied to the distribution of daily VKT:

(1) the Akaike information criterion (AIC) is a penalized log-likelihood $\mathrm{AIC} = -2\,\mathrm{LL} + 2\,(p+1)$, where p is the number of the model parameters and LL the log-likelihood;
(2) the root mean squared error $\mathrm{RMSE} = \sum_i (y_i - f_i)^2/n$,
(3) the mean average percentage error $\mathrm{MAPE} = \sum_i |(y_i - f_i)/f_i|/n$, and
(4) the χ2 statistic $\chi^2 = \sum_i (y_i - f_i)^2/f_i$ where n is the number of driving days, y_i the observed and f_i the expected value at r_i.

These measures are used to analyse the difference between the estimated and empirical distribution functions.

2.3 Number of days per year with more than L km

An understanding of the distribution of daily VKT allows us to estimate the probability of rare long-distance travel [7]. Here and in the following, we only consider daily VKT instead of the length of individual trips.

As mentioned above, the individual daily VKT r_l are assumed to be iid random variables. Let $f(r)$ denote the user-specific distribution of daily VKT. The probability of driving more than L km on a driving day is then given by $\int_L^\infty f(r)dr = 1 - F(L)$ where $F(r)$ is the cumulative distribution function of $f(r)$. Let n denote the number of driving days out of N days of observation such that $\alpha = \mathrm{n/N}$ is the share of driving days. Thus, $D(L) = 365(n/N)[1-F(r)]$ is the number of days per year with more than L km of daily VKT. Accordingly, $D(L)$ is the number of days requiring adaptation for a potential BEV user. For each vehicle the share of driving days is estimated as $\mathrm{n/N}$ and the vehicle-specific log-normal, Weibull and Gamma parameters are estimated from likelihood maximisation. Using the cumulative distribution functions, e.g. for the log-normal distribution $F(x) = \frac{1}{2}\left[1 + \mathrm{erf}\left(\frac{\ln x - \mu}{\sqrt{2}\sigma}\right)\right]$, the user-specific number of days requiring adaptation $D_i(L)$ is calculated. This procedure is repeated for each vehicle in the data base. The different CDFs for each user can be compared to the empirical cumulative distribution function (ECDF) $\hat{F}(r) = \frac{1}{n}\sum_{l=1}^{n} 1(r - r_l)$ where $1(\mathrm{x})$ is the indicator function. The latter is related to the survivor function $S(r) = 1 - F(r)$ for which 95% confidence intervals $S \pm 1.96\sigma$ can be obtained.

For the German driving data, there are very few cases (37 out of 6,339) with no variation in daily driving between the days reported, i.e. $\sigma_i = 0$. We set σ_i equal to the sample mean in these cases. However, the results reported below are robust against the exclusion of the vehicles.

The number of days with more than L km of daily VKT is a useful quantity to systematically compare different users and data set. However, as a single aggregated quantity it cannot account for other potentially important details. For example, the daily VKT alone does not distinguish between different destinations or return to home at night, including no information about the availability of charging infrastructure to recharge the vehicle for the following day. Furthermore, holiday travel might require a long ranged vehicle for more than the main travel days alone. The long-distance travel might occur only on the first and last day of the holiday – and these would be counted as days with more than 100 km – yet the owners do not return home in between intermediate trips or days and would require e.g. a rental car for the days in between as well. Furthermore, the potential electric driving of an EV user can be strongly impacted by several elements, such as aggressiveness in driving or seasonal factors requiring intense use of auxiliaries. For the focus of the present work, the threshold L for long-distance trips is kept fixed and seasonal factors are not analysed.

3 Results

3.1 Goodness-of-fit statistics

For each individual vehicle of the different data sets each of the distribution functions has been fitted for the daily VKT using maximum likelihood estimates. Table 2 summarises the GoF results by indicating which share of individual daily VKT were best according to which GoF measure. The GoF statistics are applied to each individual vehicle. Reported in the table is the share of vehicles for which the stated distribution had the best fit. The best distribution for most users is in bold face.

For the Mobility Panel data [9] with only seven days of observation, the log-normal distribution fits most daily VKT best according to three out of four GoF measures. The picture is less clear for the SCMD and Winnipeg data: Each distribution is best for most of the driving profiles in at least one measure and data set. For the SCMD data the Weibull distribution performs relatively well for all measures (none below 30%), while both the log-normal and the gamma distributions have measures that are the best fit for only 20% of the observed vehicles. Additionally, the SCMD data with more than 56 days of observation have been analysed separately (not shown). In that case, the Weibull distribution is best for most daily VKT according to three out four GoF measures. Also for the Winnipeg data the Weibull distribution shows slightly better performance. Both the log-normal and Weibull distribution perform best for two of the measures. Overall, no clear pattern is discernible for a GoF measure to prefer any of the distributions.

Our result contrast previous research that finds the gamma distribution as the most suitable [2, 3]. Our results are more comprehensive by analysing different data sets from different regions with different observations periods. We find no evidence for the Gamma distribution to be the best function for the random variation of daily VKT. For the data analysed here, the Weibull and log-normal distributions perform better than the Gamma distribution and the data sets with longer observation periods yield better GoF statistics for the Weibull distribution.

A more detailed analysis of the GoF statistics within a data set confirms the difficulties in deciding between the best distribution function. Figure 1 shows the performance of the different distribution functions for the four GoF measures for the Winnipeg data set.

Figure 1 shows that the actual values of the four GoF measures achieved by the different distributions are similar for the distributions explaining why no distribution dominates the overall GoF statistics for the Winnipeg data as stated in Table 3. Furthermore, the MAPE seems to be flawed by outliers in case of the log-normal distribution. Furthermore, Figure 1 indicates that the mean GoF statistics over the individual vehicles (the mean value in each panel of Figure 1) is lower for Weibull and Gamma than for the log-normal distribution.

Table 3: Summary of goodness of fit (GoF) statistics

	Mobility Panel			*SCMD*			*Winnipeg*		
GoF	$\ln N$	Weib.	Γ	$\ln N$	Weib.	Γ	$\ln N$	Weib.	Γ
AIC	32%	**60%**	8%	**44%**	37%	19%	**40%**	35%	25%
RMSD	**74%**	12%	14%	20%	**45%**	36%	**36%**	35%	29%
$\chi 2$	**88%**	9%	3%	30%	34%	**37%**	17%	**41%**	**41%**
MAPE	**75%**	22%	3%	35%	**39%**	26%	9%	**51%**	40%

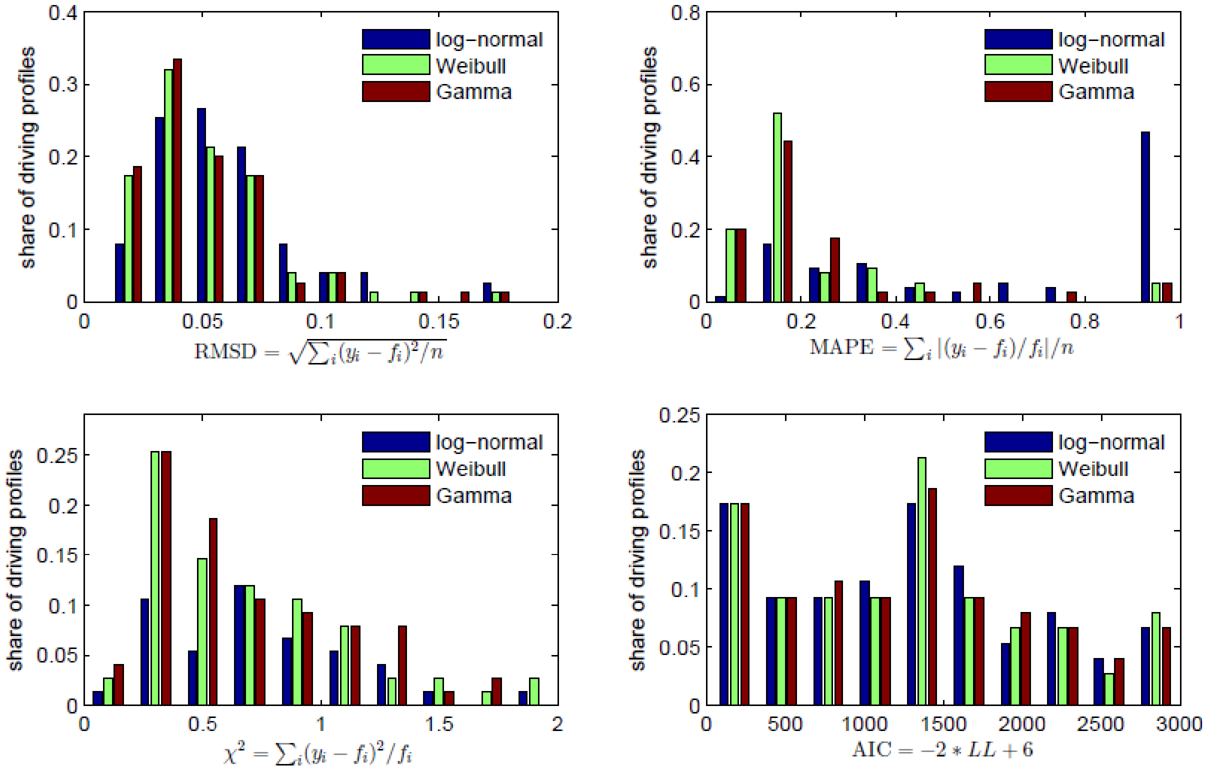


Figure 1: Distribution of GoF measures for the Winnipeg data.

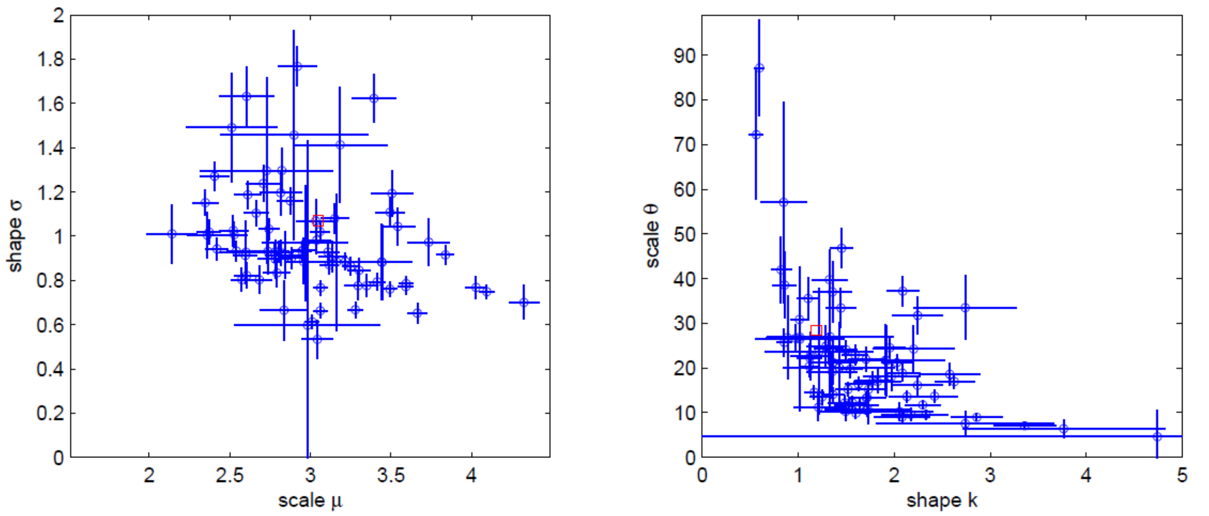


Figure 2: Distribution of scale and form parameters for the Winnipeg data set.

We study the results of the individual fits more closely in Figure 2. Figure 2 shows the maximum-likelihood estimates of the scale and shape parameters of the log-normal and Weibull distributions including 68% confidence intervals for the Winnipeg data set. The mean scale and shape parameters are marked by red squares in Figure 2. The parameters for the Gamma distribution have been omitted since they are similar to the Weibull parameters. We observe different widths of the individual confidence bands ranging from different observation periods, i.e. longer observation corresponds to more driving days and better estimates. Future research could use weighted averages of GoF measures explicitly taking into account the higher accuracy of estimates based on more data.

In summary, the log-normal distribution seems to best for the German data set. The evidence is less conclusive for the other two data sets but the Weibull seems to perform slightly better. Again, this is in clear contrast to the works of Greene (1985) and Lin et al. (2012).

3.2 Number of days requiring adaption

The best-fitting distribution function does not necessarily provide the best estimate for the probability of long-distance trips. Accordingly, we analyse the three distributions according to their ability to correctly estimate the number of days per year with daily VKT exceeding a threshold L.

Exemplary empirical distribution and estimates of the number of days requiring adaptation, $D(L)$, for different electric ranges (L) are shown in Figure 3 for real users from the Winnipeg data set. Shown are non-parametric empirical estimates (solid blue lines) with 95% confidence bands (dashed blue lines) for four individual users. Also shown are the individual log-normal (red) and Gamma (green) and Weibull (magenta) estimates (dashed lines). User data characterised by number of driving days n and number of observation days N: $n = 24; N = 186$ (top left), $n = 320; N = 340$ (top right), $n = 211; N = 324$ (lower left), $n = 156$; $N = 331$ (lower right).

We observe that all three theoretical estimates fall within the 95% confidence bands but tend to deviate from the empirical data for large driving distances L. The log-normal assumptions seems to overestimate the number of days requiring adaptation $D(L)$ for $L > 150$ km and the Gamma and Weibull estimates seem to underestimate $D(L)$. Thus, Figure 3 demonstrates that the individual distribution functions differ in the estimated probability of long-distance trips. Accordingly, the identification of best distribution function describing daily VKT is also of practical relevance for the potential utility of EVs since they estimate different probabilities for the days requiring adaptation.

We calculated the number of days requiring adaption for the Swedish and German data sets. The Swedish data set has an observation period long enough to extrapolate the actual number of days with more than 100 km daily driving distance from the observed driving behaviour to one year. The resulting number of days per year requiring adaptation is shown as a function of annual VKT for the Swedish data in Figure 4. We find that approximately 25% of the vehicles exceed 100 km of daily VKT on less than one day per month or less than twelve days per year. However, the fraction of users requiring adaptation on more than two days per month grows with increasing annual VKT.

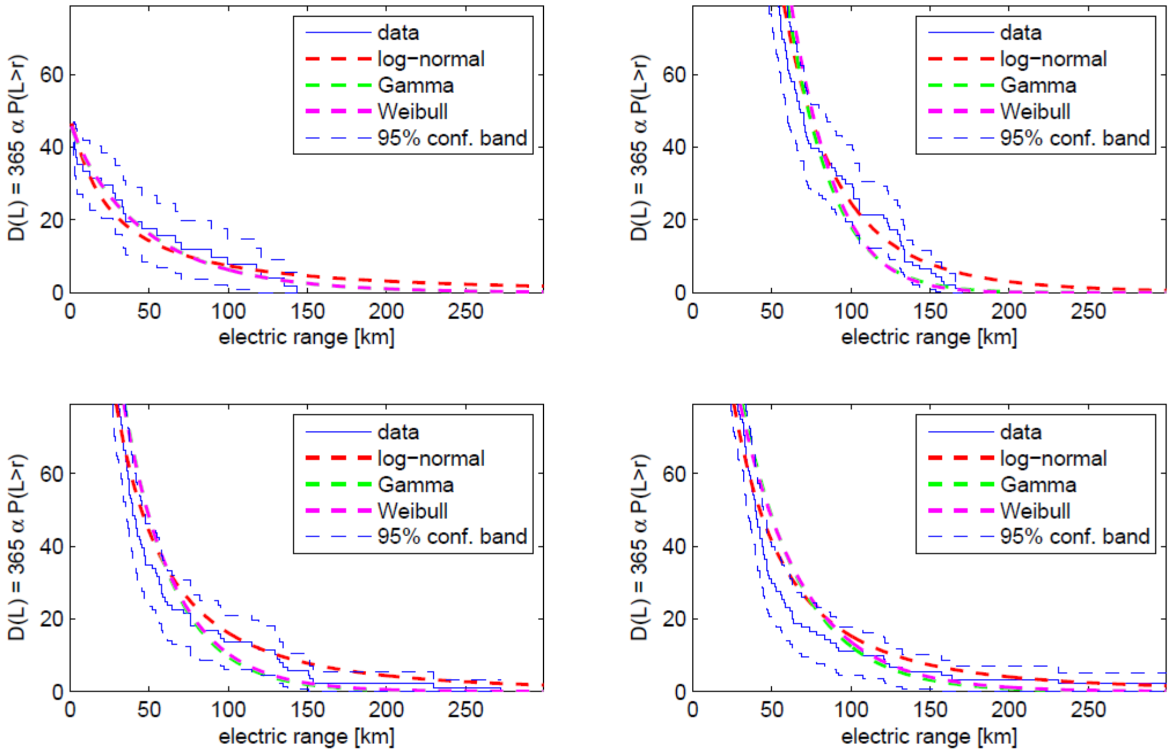


Figure 3: Comparison of empirical distribution with 95 % confidence interval and estimates of number of days requiring adaptation, *D(L)*, for different electric ranges (*L*)

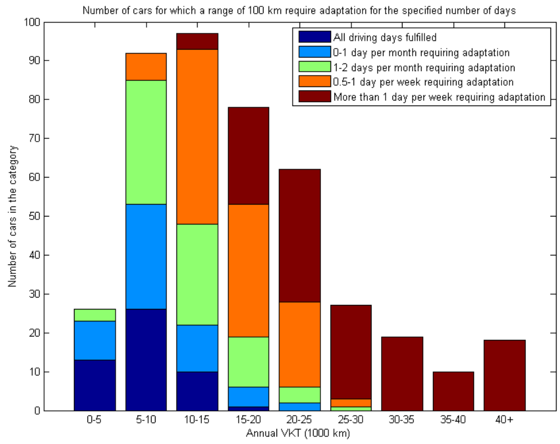


Figure 4: Share of users with days per year requiring adaptation for the Swedish data set.

Additionally, we estimated the number of days per year requiring adaptation from the large sample short observation period Mobility panel data from Germany. For this, a log-normal distribution has been fitted by maximum likelihood estimation for each individual user and the number of days requiring adaptation $D(L)$ has been calculated. Figure 5 shows the share of users with days per year requiring adaptation as a function of annual VKT for the German data set. The overall shares and pattern of the days requiring adaption is similar among the German and Swedish data set, yet a large number of days with more than 100 km of daily VKT seems to occur more frequently in the Swedish data, this is most likely an effect of that the Swedish data contain younger cars then the German data.

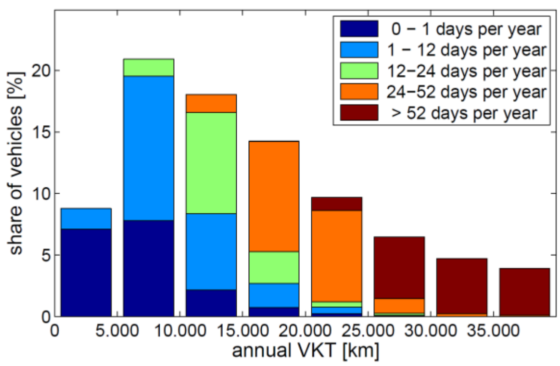


Figure 5: Share of users with days per year requiring adaptation for the German data set.

We observe a clear correlation between the days requiring adaptation and the annual VKT for the Swedish and German data. Thus, Figure 4 and 5 demonstrate that the annual VKT is a major explanatory factor for the number of days requiring adaptation. The higher share of users with a high number of days per year exceeding 100 km of daily VKT in the Swedish data set are accordingly very likely to originate from the higher annual VKTs present in Swedish driving.

The long observation periods in the Swedish data set allow for a direct comparison between the estimated and extrapolated number of days per year requiring adaption. We calculated the CDF($D(L)$) for the Swedish data in two ways. First, an individual log-normal distribution has been obtained from maximum-likelihood estimates for each vehicle. The corresponding number $D(L)$ has been estimated as in the German case by using the log-normal distribution (see Methods section). The CDF($D(L)$) of these log-normal based estimates is shown as blue ine in Figure 6. Second, we linearly extrapolated the number of days requiring adaptation up to one year for each driver (through the factor $365/N$ where N is the measurement period) and calculated the CDF($D(L)$) of these estimates too. The result is shown as the red line in Figure 6. Comparing the log-normal estimate with the linear extrapolation shows that the log-normal approach over-estimates the CDF of the number of days requiring adaptation $D(L)$ for $D < 80$ and agrees with the direct extrapolation for $D > 80$. This is consistent with the tendency of the log-normal distribution to over-count the number of days with a long driving distance over a large part of the DRA parameter space. This result is also consistent with the German case where the log-normal distribution gives a conservative estimate of the share of cars that require few DRAs compared to the Weibull and Gamma distributions.

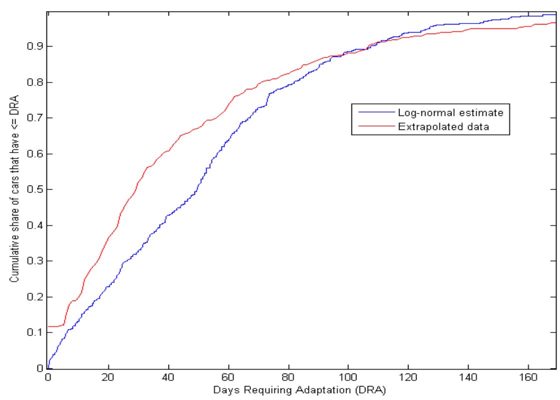


Figure 6: CDF of days with more than 100 km daily VKT for the Swedish data. Log-normal estimate and linearly extrapolated data compared.

In summary, our results indicate that the distribution functions differ in their ability to predict the number of days per year with daily VKT above a threshold. The Weibull and Gamma distribution tend to under-estimate these days potentially requiring adaptation whereas the log-normal distribution is more conservative and shows a tendency to underestimate this quantity.

4 Discussion and Summary

We analysed goodness-of-fit statistics and the ability to predict the days per year with long-distance driving for three distributions to describe daily driving distances of individual car users with driving data from different countries. In contrast to Lin et al. (2012) no single distribution clearly outperforms the others. However, the choice of the best matching distribution function has clear implications for the utility of limited range vehicles such as battery EVs. The log-normal distribution falls off much slower than the Weibull and Gamma distribution for large distances indicating that long-distance trips are much more likely. Thus, the decision between different distributions of daily VKT has direct consequences for calculations of the utility of EVs.

Only three two-parameter distributions have been analysed and many other distributions are possible. However, the set of possible distributions should be limited by general arguments. For example, only distributions that assign zero probability to daily VKT of length zero should be used since a day without driving is not a driving day. Future research should analyse more data and different GoF measures. It is also possible that no single distribution function is best for all drivers and all purposes. For example, to decide about the frequency of long-distance driving only, one could analyse only the tail of the individual daily VKT distribution or take different distribution functions for different users. Furthermore, the distribution function uncertainty can be taken into account by using model uncertainty methods such as Bayesian model averaging [12].

Our results have consequences for the design of EVs: Based on the share of car users a vehicle manufacturer wishes to address when deciding about the range of an EV, the methods and results described here help to decide on required ranges. Our findings can be applied to individual user groups such as early adopters of EVs or mid-size car users to design a vehicle that fits their range requirements. In this context, the different distributions clearly differ in the estimated share of users that could cover their driving needs on all but a few days per year. We also showed that the total annual vehicle kilometres travelled are a major explanatory factor for the inter-user variation in days with large daily VKT.

Acknowledgments

PP would like to acknowledge funding by the project Get eReady as part of the show case regions of the Federal Government of Germany. The Area of Advance of Transport, Chalmers, is gratefully acknowledged for the funding of NJ, FS and SK.

References

[1] Chan, C., 2007. “The state of the art of electric, hybrid, and fuel cell vehicles.” Proceedings of the IEEE 95, 704–718.

[2] Greene, D.L., 1985. “Estimating daily vehicle usage distributions and the implications for limited-range vehicles.” Transportation Research Part B: Methodological 19, 347–358.

[3] Pearre, N.S., Kempton, W., Guensler, R.L., Elango, V.V., 2011. “Electric vehicles: How much range is required for a day’s driving?” Transportation Research Part C: Emerging Technologies 19, 1171–1184.

[4] Smith, R., Shahidinejad, S., Blair, D., Bibeau, E., 2011. “Characterization of urban commuter driving profiles to optimize battery size in light-duty plug-in electric vehicles.” Transportation Research Part D: Transport and Environment 16, 218–224.

[5] Lin, Z., Dong, J., Liu, C., Greene, D., 2012. “PHEV energy use estimation: Validating the gamma distribution for representing the random daily driving distance”, in: Transportation Research Board 2012 Annual Meeting.

[6] Tamor, M.A., Gearhart, C., Soto, C., 2013. “A statistical approach to estimating acceptance of electric vehicles and electrification of personal transportation.” Transportation Research Part C: Emerging Emerging Technologies 26, 125–134.

[7] Plötz, 2014, “How to estimate the probability of rare long-distance trips.” Working paper on Sustainability and Innovation No. S1/2014. Fraunhofer ISI, Karlsruhe.

[8] Blum, A., 2014. Electro-mobility: statistical analysis of human mobility patterns. Master’s thesis. infernum. Wuppertal.

[9] MOP, 2010. Mobility Panel Germany 1994-2010. Institute for Transportation Reseach,

Karlsruhe Institute of Technology. http://daten.clearingstelle-verkehr.de/192/ .

[11] Auto21 (2011): Monitoring driving behaviour for Plug-in Hybrid Electric Vehicle (PHEV) research http://auto21.uwinnipeg.ca/data.html

[10] Karlsson, S. (2013) The Swedish car movement data project Final report. Göteborg, Chalmers University of Technology.

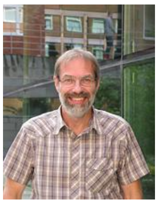
[12] Hoeting, J. A., Madigan, D., Raftery, A. E., & Volinsky, C. T. (1999). Bayesian model averaging: a tutorial. Statistical science, 382-401.

Authors

Patrick Plötz received a PhD in Theoretical Physics. He is a senior scientist in the Competence Center Energy Technology and Energy Systems at the Fraunhofer Institute for Systems and Innovation Research ISI. His current research focuses on energy efficiency and plug-in electric vehicles.

Niklas Jakobsson received his M.Sc. in Industrial Ecology at Chalmers University of Technology and is currently a Ph.D. Student at the division of Physical Resource Theory at the same university. His work is focused on private driving patterns as well as the effects of subsidies on EV sales.

Frances Sprei is an Assistant Professor in Sustainable Mobility at the Department of Energy and Environment, Chalmers University of Technology, Sweden. Her research assesses different innovative personal mobility choices. She received her PhD in 2010 and has been a visiting scholar/post-doc at Stanford University.



Sten Karlsson received a PhD in 1990 and is senior lecturer at the Department of Energy and Environment, Chalmers University of Technology, Sweden. His current research is focusing on energy efficiency and technology assessment, especially concerning private cars and the electrification of vehicles.