Statistics of Parked Cars for Urban Vehicular Networks

A. B. Reis[†*] and S. Sargento[*]

[†]Dept. of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA 15213-3890, USA
[*]Instituto de Telecomunicações, Universidade de Aveiro, 3810-193 Aveiro, Portugal

***Abstract*—The ability to predict the behavior of cars that are parked in an urban area can be very useful to the development of vehicular networks that leverage these parked cars. In this paper, we analyze the mobility patterns of people living in US cities who use cars as their primary means of transportation. We process and analyze survey data from the metropolitan areas of Atlanta, Chicago, and Knoxville, to extract statistics on the parking behaviors of individual cars.**

We accompany our study with a synopsis of how parked cars are being leveraged in urban vehicular networks, and show possible use cases for fine-grained parking models. We then provide daily and hourly analytical models of parking events, along with useful derivations of key parking statistics.

The data we present in this study conclusively shows that parking events can be classified into two major, clearly-identifiable groups according to the time cars spend parked; that each group is modeled by distinct distributions of probability; that these patterns vary substantially throughout the day and, therefore, are better modeled through time-evolving processes; and finally, that these trends are very similar among different cities.

***Index Terms*—Urban Transportation, Urban Parking, Travel Surveys, Vehicular Ad-Hoc Networks.**

I. Introduction

Regional travel surveys have been commissioned by city transportation committees from as early as 1965 [1]. These surveys consist of randomly-sampled person-to-person interviews that inquire about a person's travels in a specific date, collecting departure and arrival times, locations, and means of transportation, among others. Earlier surveys consisted of in-person interviews – nowadays, Computer-Assisted Telephone Interviews (CATI) are standard.

Studies that make use of these travel surveys are often tailored towards the development of urban parking spaces and parking lots [2], and so data is aggregated in metrics such as *parking spot occupation* throughout the day. While this gives us a broad picture of how parking spaces are used, it does not let us infer the parking behavior of each individual car.

In this study, using data from recent travel surveys, we draw statistics on how single cars park in urban areas. These statistics are of particular use to vehicular urban network research. Recent studies have brought forth the idea of leveraging parked cars as active nodes in an urban network, increasing network connectivity [3], relaying messages across intersections [4], and acting as supporting Road-Side Units [5]. So while the potential for parked cars in these networks seems clear, a number of pitfalls need to be addressed as well – for example, the electronics that enable the network must run limited by the power available from the car's battery, and cars can vacate their spots and become mobile at any time.

With this paper, we provide a probabilistic view of how individual cars park, allowing for informed decisions to be made concerning these cars. Specifically, we analyze how total time parked is distributed, how parking behaviors evolve throughout the day, and how these trends differ from one city to another. We then show how parking time can be accurately modeled through a dual Gamma stochastic process, to characterize distinct short-term and long-term parking behaviors, and provide derivations for estimating parking duration and remaining parking time of individual cars.

The remainder of this paper is organized as follows. First, in Section II, we summarize recent work on leveraging parked cars in urban networks, and show how the data in this study can be used to improve the network. An overview of urban travel surveys is then given in Section III. Section IV draws initial observations on the data; then, an hour-by-hour breakdown of parking trends is given, in Section V. Section VI introduces a probabilistic model of parking behavior, along with derivations of key parking statistics, and details how the survey data was used to fit the model. Finally, concluding remarks are presented in Section VII.

II. Applications in Vehicular Networks

Research has shown that parked cars can be leveraged to improve urban vehicular networks, simply by activating the radios in these otherwise unused entities. For example, in [3], a 3.3x improvement in node density is seen when 10% of all parked cars are activated; [4] suggests using parked cars to act as relays between vehicles whose line of sight is blocked by urban buildings, resulting in nearby emergency messages being delivered up to 17 seconds faster; and [6] shows how parked cars can be used to assist content downloading, by caching content from nearby Road-Side Units (RSUs). Our own work in [5] shows how parked cars can be a powerful and capable substitute to expensive deployments of urban Road-Side Units.

A support network that consists of parked cars will necessarily require some degree of self-organization, as cars park and unpark often and unpredictably. When selecting a parked car for a specific role in the network, with the models presented in this paper one is able to prioritize cars that are statistically more likely to stay parked for longer, so the chosen car can fullfil its role for longer as well.

Another concern with this approach is that power consumption of the radio electronics must be kept in check, in order not to risk draining the car's battery (that is also needed to start the engine). In this context, a system to rotate the supporting roles in the vehicular network among nearby parked cars can be implemented. Using the models in this work, an algorithm can rotate roles among cars based on how long each car is likely to remain parked, and how much of its battery is left.

In general, we believe that most systems that leverage parked cars will benefit by taking into consideration the statistics presented in this work.

III. Data Sources

Performing a metropolitan travel survey demands a substantial amount of work and logistics, along with adequate funding to employ interviewers and collect data on a household-by-household basis. Figure 1 shows the number of surveys performed in the USA, from 1965 to date. The 1990s and 2000s saw a significant growth in the number of surveys performed per year, which has then since declined.

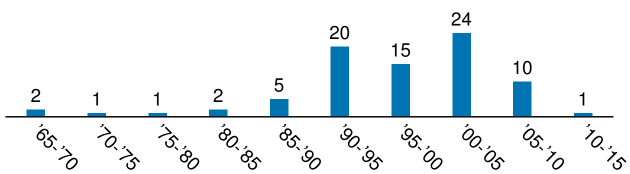

Figure 1. Histogram of travel surveys conducted by metropolitan areas, states and localities of the United States over the past 50 years.

In this work, we analyze data from three specific surveys – they are: Atlanta 2011 [7], Knoxville 2008 [8], and Chicago 2007 [9]. In order to draw statistics that reflect modern parking trends, we specifically select surveys that are no older than 10 years, from medium- and large-sized urban areas. At the time of this writing, these were the surveys we were able to locate that fit the criteria.

As our goal is to analyze the individual parking behavior of vehicles in a city, we process this survey data, and exclude samples in which the mode of transportation does not cause a parking event. Traveling on foot, by bicycle or by collective public transportation (bus, train, subway) are some examples of samples that were excluded. Table I lists the surveys, the number of samples contained in each survey, the percentage of these samples where the person traveled by car, and the population size in each metropolitan area.

IV. Parked Vehicle Statistics

We begin by analyzing the complete 24-hour set of data in each survey, as a whole, and deriving important characteristics in this data. For improved presentation, we apply kernel density estimators to the survey data, using Normal distributions as the kernel function $K(\cdot)$, its bandwidth calculated through the normal distribution approximation [10]. Figure 2 shows how the duration of parking events in a single day is distributed, for each survey's urban area.

Important observations can be drawn from Figure 2, the first one being that the main characteristics of parking duration are remarkably similar among all three cities. A second observation is that a large mass of cars park for 3 hours or less (the first peak from 0 to 180 minutes), which are then followed by a series of smaller overlapping peaks that represent longer-term parking.

Using the 180-minute mark as a classifier, we now analyze the short-term and long-term parking event groups. Figure 3 plots the distribution of the times of day at which vehicles first park, for the two groups, in the city of Chicago. The data shows that short-term and long-term parking events are also very distinct in terms of the time at which the car is parked. Short-term parking (in red) is mostly consistent throughout the daytime, while long-term parking (in blue) peaks substantially between 8 and 9 A.M., with a second smaller peak occurring again around 1 P.M.

The probability distribution of the time cars spend parked, in Figure 2, is useful to vehicular research, and this 24-hour aggregate of events can be fitted to known probability distribution functions. We fitted the survey data in the Chicago dataset, the largest of our three chosen surveys in terms of sample size, to various well-known distributions and applied the Kolmogorov-Smirnov test (which quantifies the distance between two distributions) to judge the best fit.

The Nakagami probability distribution, a distribution related to the Gamma distribution (that often models waiting times between events), exhibited the best fit to the data shown in Figure 2. Table II shows the Nakagami parameters of shape (m) and spread (Ω) resulting from the fit, and the upper and lower bounds of the Kolmogorov-Smirnov test. The error between the empirical survey data and the fitted probability distribution does not exceed 6%.

Having a single mathematical model capable of describing the time vehicles will spend parked can be very useful – however, we will now show how parking behaviors vary throughout the day, which will prompt for more complex models.

V. Hour-by-Hour Analysis

We know, intuitively, that parking trends vary significantly along the day, matching people's daily routines and habits. We

Survey	Number of samples	Car-only samples	% Car samples	Metro. pop.
Atlanta 2011	119,478	81,863	68%	5.5 m
Knoxville 2008	15,313	11,535	75%	0.9 m
Chicago 2007	159,856	103,964	65%	9.5 m

Table I
Characteristics of travel survey data

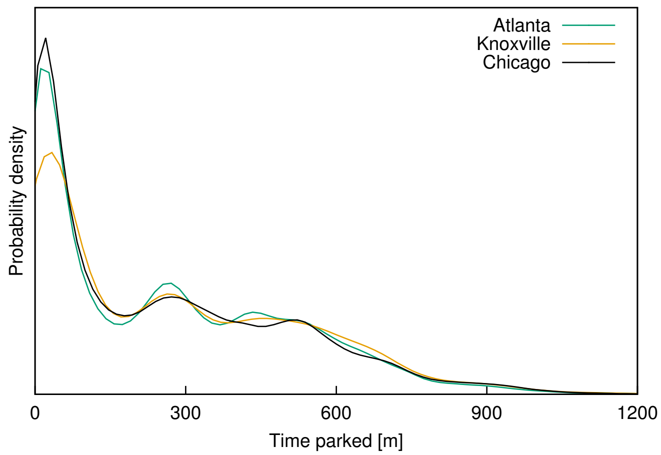

Figure 2. Total time parked, in minutes – probability density functions for the three survey sets.

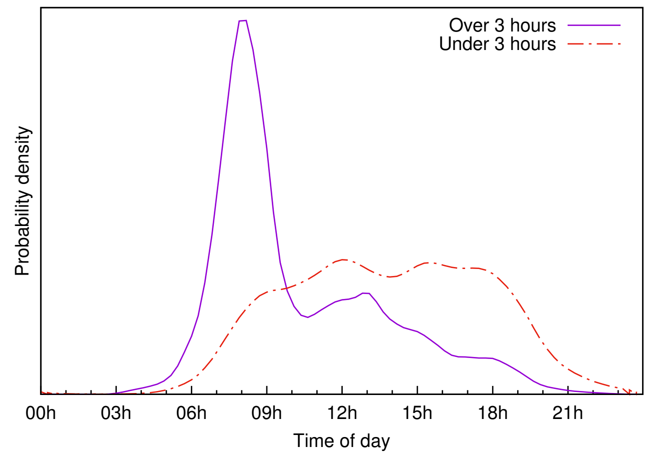

Figure 3. Time of day at which vehicles park, for short-term (under 3 hours) and long-term (over 3 hours) parking.

saw already, in Figure 3, that most long-term parking happens in the early hours of the morning, which coincides with the hours at which most day jobs begin. In this section, we show these trends in finer detail.

Figure 4a shows probability density functions of the total time vehicles stay parked, in an hour-by-hour basis: for example, the top figure pertains to the vehicles that parked between 4:00 and 4:59. Once again, this analysis is repeated for the three chosen surveys. Data begins at 4 A.M., as there were insufficient data points in earlier hours for us to be able to perform meaningful statistical analyses.

These detailed plots let us draw important conclusions. First, we can see that the split between short-term and long-term parking is only meaningful until 10 A.M. Past that hour, long-term parking events almost cease to exist. Second, the peak duration of long-term parking events grows shorter each passing hour: at 4 A.M., long-term parking averages 10 hours of parking time, but four hours later, at 8 A.M., the density peaks at 7.5 hours. The data shows very clearly that the earlier a vehicle is parked, the longer it will stay parked for. And Figure 4a again shows that major parking trends are consistent across different cities, though the Knoxville survey data (the smallest of the three) may be sufficiently distinct from Chicago and Atlanta to warrant further research into smaller-sized urban areas.

For the purposes of vehicular network research, knowing these distributions and the time at which a vehicle parked will allow one to estimate that vehicle's parking duration, and how likely it is that it will park for a short or a long period. But if one cannot determine when a vehicle first parked, important information can still be obtained.

Figure 4b shows the distribution of time that a parked vehicle has left in that state (until it becomes mobile again), at specific hours of the day. This data is obtained by taking a snapshot of all parked cars at the beginning of the hour, and determining how long each vehicle will remain parked for. This way, knowing only the current time of day, one can scan a car or a group of cars and take an informed estimate of when these cars will be leaving their parking spots. The probability density functions are similar to the ones in Figure 4a, with important differences. The average remaining parked time on each vehicle is higher, which means that, from a random sample of cars that are parked, one is more likely to find cars parking for a longer time. This effect propagates throughout the day: at 1 P.M., while most new cars are parking for an hour or less, the ones already parked will remain in that state for up to 5 hours.

VI. Data Model

We now attempt to obtain a mathematical model for a car's parking behavior, from the data that was just presented. From the hourly data in Figure 4a, and by splitting the data points at the first observable trough in the distribution, we were able to heuristically determine that both short-term and long-term parking can be accurately modeled with a dual Gamma probability distribution. This also tells us that our initial analysis of the whole data set – which, absent of this classification, appeared to be best fit by a Nakagami distribution – can be improved upon.

From Figure 4a we observe that the distribution of short-term parking mimics an exponential distribution (with the distinction that very short parking events (e.g. under 1 minute) are empirically rare), which is itself a special case of the Gamma distribution. Long-term parking events resemble a normally-

	Parameters		K-S Test	
	m	Ω	D^+	D^-
Nakagami Distribution	0.282598	125,292	5.48%	5.97%

Table II
Fitting parameters and Kolmogorov-Smirnov test results

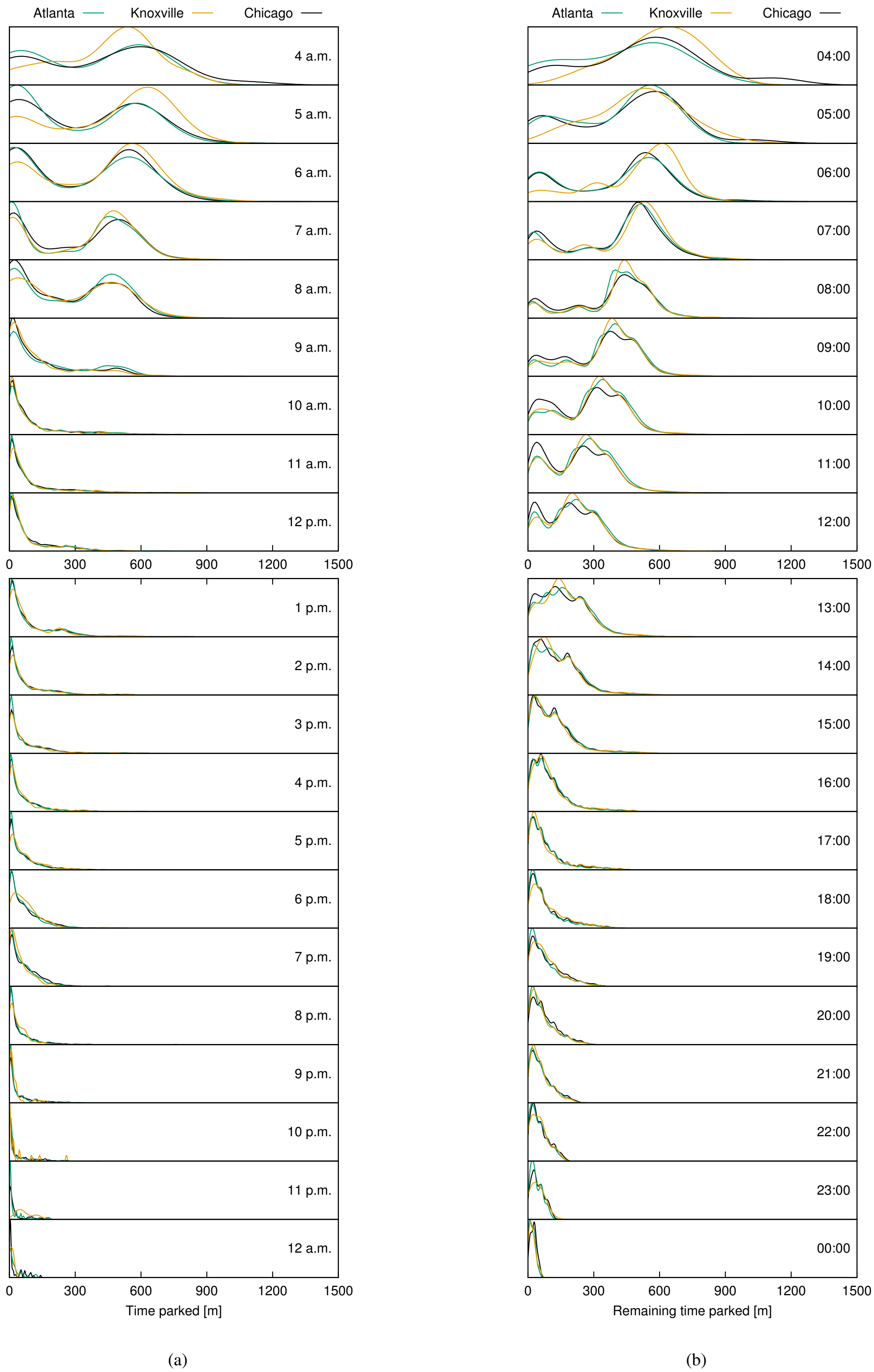

Figure 4. (a) Probability distribution of total time parked, grouped by the time of day at which parking occurred. (b) Distribution of remaining time until parking ends, from global snapshots at the start of each hour.

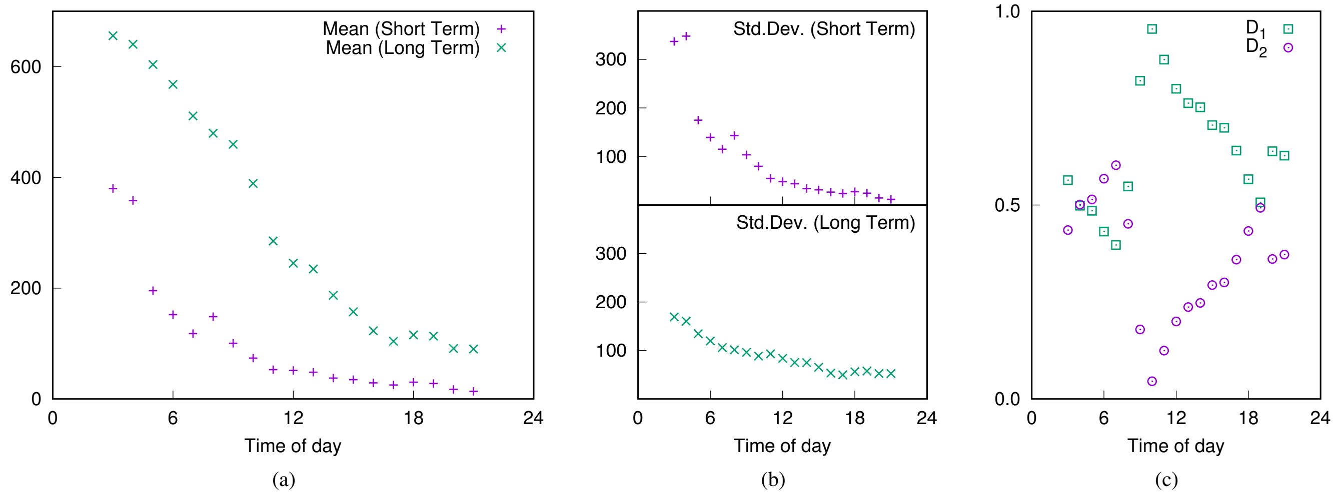

Figure 5. (a) Expected value of each distribution. (b) Standard deviation of each distribution. (c) Weight coefficients.

distributed random variable, and in fact a Gaussian model will reasonably fit the data – however, such a distribution is defined in $\mathbb{R}$, which does not apply here as parking events cannot have a negative duration. A Gamma distribution, defined in $\mathbb{R}^+$, models long-term behavior equally well, and the resulting dual model is substantially more tractable as we no longer need to be concerned with negative tails in the normal distribution.

From our data, a car's parking behavior can then be described by a stochastic process X, indexed by discrete time t, where each X_t is a continuous random variable representing the time the vehicle will spend parked, when it parks at hour t. The individual random variables follow a mixture distribution of the aforementioned Gamma models. The first-order density of X is given by

$$\begin{aligned} f(\mathbf{x},\mathbf{t}) =& D_{1,t} \times \frac{1}{\Gamma(\kappa_{s,t})\theta_{s,t}{}^{\kappa_{s,t}}} x^{\kappa_{s,t}-1} e^{-\frac{x}{\theta_{s,t}}} \\ &+ D_{2,t} \times \frac{1}{\Gamma(\kappa_{l,t})\theta_{l,t}{}^{\kappa_{l,t}}} x^{\kappa_{l,t}-1} e^{-\frac{x}{\theta_{l,t}}} \\ &\mathbf{x} > 0 \quad , \quad \mathbf{t} = \{0, 1, 2, \ldots, 23\} \quad , \end{aligned} \tag{1}$$

where κ_s and θ_s are the shape and scale parameters of the Gamma distribution that models short-term parking, and κ_l and θ_l are their equivalent, but for the distribution that models long-term events. Coefficients D_1 and D_2 weight each distribution, as a valid density function must always integrate to one (and therefore, $D_1 + D_2 = 1$). The t subscript indicates that the variable is specific to time t.

The first-order distribution $F_X(x)$ of the stochastic process X_t is then given by

$$\begin{aligned} F(\mathbf{x},\mathbf{t}) =& D_{1,t} \frac{\gamma\left(\kappa_{s,t}, \frac{x}{\theta_{s,t}}\right)}{\Gamma(\kappa_{s,t})} + D_{2,t} \frac{\gamma\left(\kappa_{l,t}, \frac{x}{\theta_{l,t}}\right)}{\Gamma(\kappa_{l,t})} \\ &\mathbf{x} > 0 \quad , \quad \mathbf{t} = \{0, 1, 2, \ldots, 23\} \quad , \end{aligned} \tag{2}$$

where $\Gamma(\cdot)$ and $\gamma(\cdot)$ are the upper and lower incomplete Gamma functions, respectively.

Knowing the hour t at which a car parked, the expected time that the car will be parked for can be shown to be:

$$\begin{aligned} E[X_t, \mathbf{t} = t] &= \int_0^\infty x f(x, \mathbf{t} = t) \mathrm{d}x \\ &= D_{1,t}\kappa_{s,t}\theta_{s,t} + D_{2,t}\kappa_{l,t}\theta_{l,t} \quad , \end{aligned} \tag{3}$$

where $\kappa\theta$ is, by definition, the expected value of the Gamma distribution.

An equally important derivation is the probability that a car still has n more hours left parked – useful if, e.g., one knows that a car has been parked for t hours, and wishes to know the probability that it will stay parked for at least n more hours. Let t_a be the time the car has been parked for, and $t_p = t_a + n$ the parked time we wish to know the probability of. This conditional probability will be given by

$$P[X_t > t_p | X_t > t_a] = \frac{1 - F_X(t_p)}{1 - F_X(t_a)} \quad , \quad t_a < t_p$$

$$= \frac{D_1\gamma\left(\kappa_s, \frac{t_p}{\theta_s}\right)\Gamma(\kappa_l) + D_2\gamma\left(\kappa_l, \frac{t_p}{\theta_l}\right)\Gamma(\kappa_s) - \Gamma(\kappa_l)\Gamma(\kappa_s)}{D_1\gamma\left(\kappa_s, \frac{t_a}{\theta_s}\right)\Gamma(\kappa_l) + D_2\gamma\left(\kappa_l, \frac{t_a}{\theta_l}\right)\Gamma(\kappa_s) - \Gamma(\kappa_l)\Gamma(\kappa_s)}, \tag{4}$$

where $\{\kappa_s, \theta_s, \kappa_l, \theta_l, D_1, D_2\}$ are specific to the time t when the vehicle in question parked.

Data Fitting

We fit the hourly survey data points shown in Figure 4a to the probabilistic model of Equation (1), through an Expectation-Maximization (EM) routine. Due to the considerable number of parameters $(\kappa_s, \theta_s, \kappa_l, \theta_l, D_1, D_2)$, we opted to augment the EM process with an iterative fitting routine, drawing from concepts of evolutionary computation. As all three surveys show similar parking trends, we fitted only the data in the Chicago 2007 set, the largest of the three.

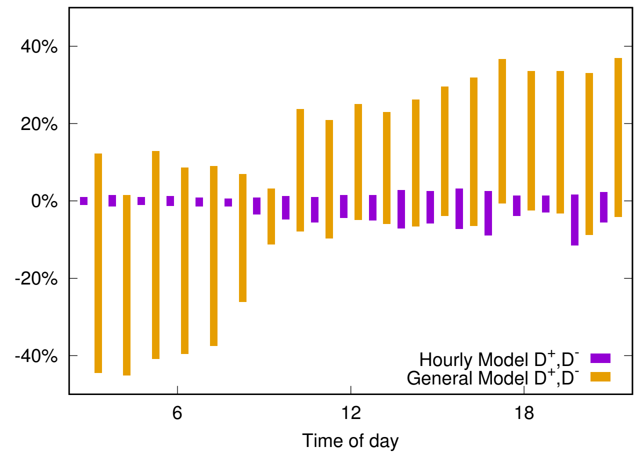

Figure 6. Kolmogorov-Smirnov test results for the data against the general and hourly models.

The fitting routine works by mutating one variable on each iteration, and then determining the log-likelihood of the resulting analytical model, to evaluate its *goodness of fit* to the survey data. If a particular mutation improves the goodness of fit, it is kept – if not, it is discarded. As the routine iterates, it converges towards a set of variables that is a better fit to the data. We applied this process to every hourly subset of data, and drew the values of all variables once fitness stabilized. A table with the raw fitting results can be found in the Appendix.

From the results of the fitting process, Figure 5 shows the main characteristics of the short-term and long-term Gamma-distributed parking behaviors, and their evolution over time. Figure 5a plots the mean parking time which, as could be observed earlier in Figure 4a, grows shorter throughout the day. The variance of parking time also grows tighter throughout the day, with brief increases at 9 A.M. and around 6 P.M.

Validation

To validate the hourly model that resulted from the fitting process, we performed the Kolmogorov-Smirnov test on each hourly snapshot of data. Figure 6 shows the $\{D^-, D^+\}$ range of the K-S test for both the general (daily) and the hourly models. We can see that the hourly model fits the data with errors not exceeding 10%, and stays under 5% error in many of the hourly slices. In comparison, the general (daily) model will undershoot and overshoot on most of the partitioned data.

VII. Conclusion

In this paper, we studied vehicular parking trends in medium-sized urban areas. By analysing recent travel surveys from the cities of Atlanta, Chicago, and Knoxville, we provide a probabilistic view of how individual cars park. We first showed how parking events naturally separate into two major groups of short-term and long-term parking; then, we conducted an hour-by-hour analysis of the data, revealing how trends evolve throughout the day. Modeling each event group separately, we determined that short-term parking resembles an exponential distribution, while long-term parking is predominantly gaussian in its nature.

Drawing from these observations, we developed an analytical model that accurately reflects vehicles' parking behavior in an urban area, along with important derivations that are of particular relevance to vehicular network research that intends to leverage parked cars. This is the topic of our future work as well: to use these models to improve algorithms that enable the use of parked cars as urban Road-Side Units.

Acknowledgment

This work was partially funded by the CMU-Portugal Project S2MovingCity: Serving and Sensing a Moving City (CMUP-ERI/TIC/0010/2014), by FCT/MEC through national funds, and co-funded by FEDER – PT2020 partnership agreement under the project UID/EEA/50008/2013.

Appendix

Hour	D_1	D_2	κ_s	θ_s	κ_l	θ_l
3	0.5642	0.4358	1.272	298.7	15.00	43.73
4	0.4984	0.5016	1.059	338.4	15.95	40.14
5	0.4854	0.5146	1.252	156.1	20.09	30.06
6	0.4317	0.5683	1.195	127.5	22.50	25.24
7	0.3972	0.6028	1.057	111.6	23.27	21.96
8	0.5482	0.4518	1.079	137.8	22.36	21.46
9	0.8206	0.1794	0.9479	106.0	22.90	20.08
10	0.9543	0.0457	0.8640	85.58	19.39	20.07
11	0.8755	0.1245	0.9315	56.75	9.388	30.38
12	0.8000	0.2000	1.153	44.69	8.545	28.69
13	0.7631	0.2369	1.224	39.52	9.682	24.23
14	0.7523	0.2477	1.233	30.55	6.246	29.97
15	0.7061	0.2939	1.252	27.84	5.831	27.02
16	0.6995	0.3005	1.224	23.82	5.335	23.06
17	0.6407	0.3593	1.129	22.34	4.384	23.73
18	0.5669	0.4331	1.229	24.59	4.244	27.27
19	0.5071	0.4929	1.309	21.23	3.849	29.47
20	0.6390	0.3611	1.451	11.93	3.031	30.10
21	0.6277	0.3723	1.454	9.482	2.966	30.41

References

[1] C.L. Purvis, "San Francisco Bay Area 1990 Regional Travel Characteristics," Dec. 1994. Available: http://ntl.bts.gov/DOCS/SF.html
[2] C. Morency and M. Trepanier, "Characterizing Parking Spaces Using Travel Survey Data," CIRRELT, Tech. Rep. 2008-15, May 2008.
[3] Nianbo Liu *et al.*, "PVA in VANETs: Stopped cars are not silent," 2011 Proceedings IEEE INFOCOM, pp. 431-435, Apr. 2011.
[4] D. Eckhoff, C. Sommer, R. German, and F. Dressler, "Cooperative Awareness at Low Vehicle Densities: How Parked Cars Can Help See through Buildings," 2011 IEEE Global Telecommunications Conference (GLOBECOM 2011), pp. 1-6, Dec. 2011.
[5] A.B. Reis and S. Sargento, "Leveraging Parked Cars as Urban Self-Organizing Road-Side Units," 2015 IEEE 82nd Vehicular Technology Conference (VTC2015-Fall), Sep. 2015.
[6] F. Malandrino *et al.*, "Content downloading in vehicular networks: Bringing parked cars into the picture," 2012 IEEE 23rd International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC), pp. 1534-1539, Sep. 2012.
[7] Atlanta Regional Commission, "ARC Regional Travel Survey," Nov. 2011.
[8] Knoxville Regional Transportation Planning Organization, "2008 East Tennessee Household Travel Survey," Oct. 2008.
[9] Chicago Metropolitan Agency for Planning, "Chicago Regional Household Travel Inventory," 2007.
[10] B.W. Silverman, "Density Estimation for Statistics and Data Analysis," New York: Chapman and Hall, 1986.