Modeling and Characterization of Vehicular Density at Scale

Gautam S. Thakur
CISE, University of Florida,
Gainesville, Florida.
Email: gsthakur@cise.ufl.edu

Pan Hui
Telekom Innovation Laboratories,
Berlin, Germany.
Email: pan.hui@telekom.de

Ahmed Helmy
CISE, University of Florida,
Gainesville, Florida.
Email: helmy@cise.ufl.edu

***Abstract*—Future vehicular networks shall enable new classes of services and applications for car-to-car and car-to-roadside communication. The underlying vehicular mobility patterns significantly impact the operation and effectiveness of these services, and hence it is essential to model and characterize such patterns. In this paper, we examine the mobility of vehicles as a function of traffic density of more than 800 locations from six major metropolitan regions around the world. The traffic densities are generated from more than 25 million images and processed using background subtraction algorithm. The resulting vehicular density time series and distributions are then analyzed. It is found using the goodness-of-fit test that the vehicular density distribution follows heavy-tail distributions such as Log-gamma, Log-logistic, and Weibull in over 90% of these locations. Moreover, a heavy-tail gives rise to long-range dependence and self-similarity, which we studied by estimating the Hurst exponent (H). Our analysis based on seven different Hurst estimators signifies that the traffic patterns are stochastically self-similar ($0.5 \leq H \leq 1.0$). We believe this is an important finding, which will influence the design and deployment of the next generation vehicular network and also aid in the development of opportunistic communication services and applications for the vehicles. In addition, it shall provide a much needed input for the development of smart cities.**

I. Introduction

Research in the area of vehicular networks has increased dramatically in recent years. With the proliferation of mobile networking technologies and their integration with the automobile industry, various forms of vehicular networks are being realized. These networks include vehicle-to-vehicle, vehicle-to-roadside, and vehicle-to-roadside-to-vehicle architectures. Realistic modeling, simulation and informed design of such networks face several challenges, mainly due to the lack of large-scale community-wide libraries of vehicular data measurement, and representative models of vehicular mobility.

Earlier studies in this area have clearly established a direct link between vehicular macro-mobility based on density distribution and the performance of vehicular network primitives and mechanisms [1], [2], including broadcast and geocast protocols [3]. Although good initial efforts have been exerted to capture realistic vehicular density distributions, such efforts were limited by availability of sensed vehicular data. Hence, there is a real need to conduct vehicular modeling and characterization using larger scale and more comprehensive data sets. Furthermore, commonly used assumptions, such as exponential distribution [4] have been used to derive many theories and conduct several analyses, the validity of which bears further investigation.

In this study, we systematically examine the modeling and characterization of vehicular mobility using a family of heavy-tail and memoryless theoretical distributions. To avoid the limitations of sensed vehicular data, we instead utilize the existing global infrastructure of tens of thousands of video cameras providing a continuous stream of street images from half a dozen regions around the world [5], [6]. We processed millions of images, captured from publicly available traffic web cameras, using a novel density estimation algorithm to help investigate and understand the traffic patterns of cities and major highways. Our algorithm employs simple, scalable, and effective background subtraction techniques to process the images and build an extensive library of spatio-temporal vehicular density data [7]. The resultant dataset of 25 million records used, has traffic density time series from 819 locations belonging to six major metropolitan regions around the world.

As the first step towards realistic vehicular network modeling, we aim to provide a comprehensive view of the fundamental statistical characteristics of the vehicular traffic density exhibited by the data. We conducted two main sets of statistical analyses: the first includes an investigation of the best-fit distribution and goodness-of-fit test using a family of heavy-tail and memoryless models, while the second is a study of the long range dependence (LRD) and self-similarity observed in that data. Our analysis shows two main results: i) the empirical data of vehicular densities in most of the locations follow heavy-tail distributions such as 'Log-gamma', 'Log-logistic', and 'Weibull'. ii) the data consistently showed a high degree of self-similarity. This may suggest a long-range-dependent process governing the vehicular arrival process in many realistic scenarios. Such result is in sharp contrast to the assumptions of memoryless processes commonly used for modeling the vehicular mobility.

The rest of the paper is organized as follows: In section II, we detail our vehicular dataset, its pre-processing method to extract traffic density, and density validation using ground truths. Statistical analysis of measurements and modeling is illustrated in section III. Vehicular time series following self-similar processes is discussed in section IV. Finally, we conclude our paper in section V.

II. Measurements and Vehicular Density Estimation

Table I details the six regions used in this study and the extent of the data and time span of the sample. The traffic snapshots in form of images taken at few seconds apart

978-1-4673-1017-8/13/$31.00 ©2013 IEEE

TABLE I. GLOBAL WEBCAM DATASET

Region	# of Cameras	Duration	Interval	Records	Database Size
Connecticut	120	21/Nov/10- 20/Jan/11	20 sec.	7.2 million	435 GB
London	182	11/Oct/10 - 22/Nov/10	60 sec.	1 million	201 GB
Seattle	121	30/Nov/10 - 01/Mar/11	60 sec.	8.2 million	600 GB
Sydney	67	11/Oct/10 - 05/Dec/10	30 sec.	2.0 million	350 GB
Toronto	89	21/Nov/10 - 20/Jan/11	30 sec.	1.8 million	325 GB
Washington	240	30/Nov/10 - 01/Mar/11	60 sec.	5 million	400 GB
Total	**819**	**-**	**-**	**25.2 million**	**2311 GB**

from every camera (at intervals ranging from 20-60 seconds), first pass a background estimation and subtraction phase. These are then used to estimate the *traffic density* arriving per unit time as opposed to a car count. While a car count might seem preferable to a traffic density measure, there are several practical challenges. A car count requires a far greater computational cost due to the effort required to isolate each object. Traffic congestion further complicates matters when cars occlude each other, making it difficult to segregate cars based on edge structures. In addition, vehicles at the far end of the road are small in the image and cannot be detected by these algorithms.[1] Since these cameras do not have night vision, we limit our study to 7am-6pm. On average, we download 15 gigabytes of imagery data per day from over 2,700 traffic cameras, with an overall dataset of 7.5 terabytes containing around 125 million images. In this paper, for a fair comparison, we have selected only six regions with nearly similar time granularity of traffic snap shot, as shown in Table I.

A. Background Subtraction

Background subtraction is a standard method for object localization in image sequences with fixed cameras, where the frame rate is lower than the velocity of the objects to be tracked (i.e. cars move out of the scene typically at a rate exceeding 1 minute). The basis for models of background are based on the observation that *background* does not change significantly (in comparison to foreground/objects) across time. Any part of an image that does fit with that model is deemed as *foreground/object*. These foreground regions are then further processed for the detection of desired objects.

The background model used here assumes that the distribution of background pixel values may be modeled as a weighted sum of Gaussian distributions. Our approach follows closely to those proposed by [8], [9], [10] because of their reliability and robustness to sensitive changes in the lighting conditions. In our approach, the observed pixel value is modeled by a weighted sum of Gaussian kernels. Let x_t represent a pixel value in the t^{th} frame, then the probability of observing this value is assumed to be:

$$p(x_t) = \sum_{i=1}^{K} w_i^t * \mathcal{N}(\mu_{i,t}, \Sigma_{i,t}) \qquad (1)$$

where $\mathcal{N}(\mu_{i,t}, \Sigma_{i,t})$ is the i^{th} kernel with mean $\mu_{i,t}$ and covariance matrix $\Sigma_{i,t}$, and w_i^t is the weight applied to that

[1]Another solution could be to only count cars that are close to the camera; while this is definitely an option for video data, for snapshot data it would result in those distant cars having left the scene before the next snapshot; the net effect being that the maximum observed car count at a junction is truncated causing problems in the multivariate analysis later on.

TABLE II. SUMMARY OF REGRESSION ANALYSIS

Camera	df	$\beta_0(\alpha = 0.95)$	$\beta_1(\alpha = 0.95)$	R^2	p	ρ
1	100	-1.19±0.046	0.03±0.003	0.7922	0	0.91
2	100	-3.25±0.130	0.09±0.007	0.8579	0	0.92
3	100	8.16±0.045	0.10±0.005	0.9308	0	1.00
4	100	8.16±0.045	0.10±0.005	0.9308	0	1.00
5	100	8.16±0.045	0.10±0.005	0.9308	0	1.00
6	100	-2.13±0.112	0.07±0.008	0.7499	0	0.88

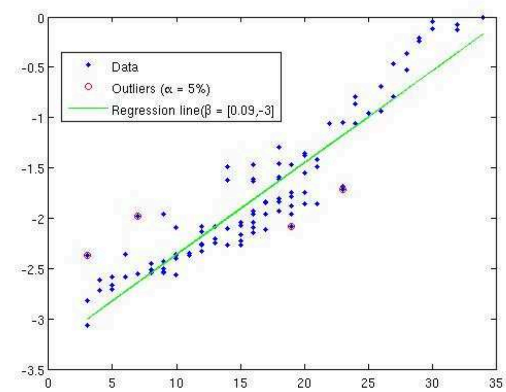


Fig. 1. A comparison of empirical traffic densities with number of cars.

kernel such that $\sum_i w_i^t = 1$. We assume that RGB channels are uncorrelated thus the covariance matrix for each kernel is diagonal.[2] When a new frame arrives, the pixel values are compared to the kernels to determine if it is likely that this value was drawn from a distribution with $\mathcal{N}(\mu_{i,t}, \Sigma_{i,t})$ (using for example a 95% confidence interval). If so, $\mu_{i,t}$, $\Sigma_{i,t}$ and w_i are updated using exponential filters; if not, a new kernel is created and the existing kernel with the lowest w_i is eliminated (see [8] for specifics). Short lived kernels and their associated pixels are deemed to be possibly foreground producing a binary map. Morphological operations are then applied to this map to remove noise and any blobs with area smaller than a certain threshold.

The view of most cameras used in this study is along the direction of the road and this perspective skews the size of objects on an image [11]. To counter this effect, we weigh each foreground pixel with the exponent of it's distance from the bottom of the image. Thus a pixel in the bottom of the image will be weighted less (object appear larger at the bottom than on the top) than a pixel at the top. While this weighting is not exact and does produce some warping as we shall see in the next section, it is not excessive but is simple and does not require manually tuning each camera.

[2]Thus reducing the number of unknown parameters.

B. Ground Truth for Validation

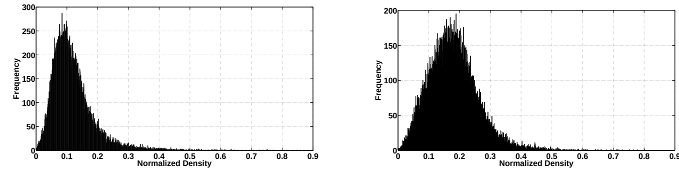
To test the performance of the car density capture, six cameras were selected at random and 102 images from each were examined by hand to produce a *ground truth* count for the number of cars. This ground truth was then regressed against the measured car density to check that the relationship is linear. The regression from one of the cameras is shown in Figure 1 and has a reasonable fit. There are some outliers, especially at low levels of traffic and there also appears to be a slight non-linear relationship between the ground truth and measured car density due to the warping effect of perspective (discussed above). Table II shows the summary statistics for the regression analysis including Spearman's correlation coefficient, ρ, which seems to imply that there is a perfect non-linear correlation for camera's 3 to 5.[3] Overall, the analysis shows that while there are some errors, the relationship between the actual and measured number of cars is sufficiently clear to allow analysis at a network level. For more information, we suggest interested readers to refer to our technical report [7].

III. TRAFFIC MODELING

In this section, we perform modeling and characterization of vehicular traffic densities. We show that memoryless models such as exponential distribution do not capture the traffic trends, instead heavy-tail distributions such as Weibull are better at estimating the parameters of empirical traffic data. We use goodness-of-fit test to support our analysis of using such distributions for traffic modeling purposes.

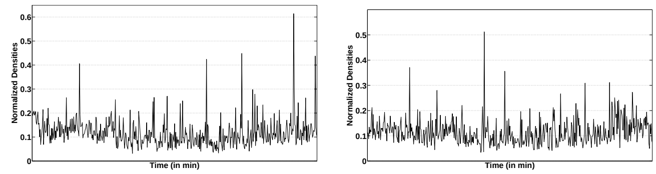
Evaluation Approach: Earlier, we have shown that traffic at each location is linearly correlated to number of vehicles at that location. The next step is to study the underlying statistical patterns through a sequence of observations. We achieve this by modeling the empirical vehicle traffic densities using a family of heavy-tail and memoryless distributions. A heavy-tail distribution such as Pareto, is characterized by a density function that converges less rapidly than an exponential function. For a random variable exhibiting heavy-tail waiting time, the larger its already passed waiting time value, the lower it's likelihood of future arrival in given time interval. In case of memoryless processes such as exponential, models' subsequent events are completely independent from the previous events. The distribution models we consider are Exponential, Log-gamma, Log-logistic, Normal, Poisson, and Weibull distribution. For the analysis purposes, we follow the approach suggested by Clauset et. al. to ensure that the parameters of the theoretical models are not estimated from the observed data [12]. First, we measure the distribution parameters by using the Maximum Likelihood Estimation (MLE). Second, we validate the significance level of estimated model parameters using the graphical properties and goodness-of-fit measures based on statistical theory. We use Kolmogorov-Smirnov test (KS-test) and evaluate its D-statistics (estimate maximum absolute difference between the empirical and theoretical distribution) on the CDFs of estimated parameters and of empirical vehicle densities. We rank models based on their significance and accuracy in modeling empirical data. We also report models at 3% and 5% of conservative deviation in order to show the efficacy of more than one distribution in modeling the empirical data. Finally, we compute an aggregate statistic for all 819 locations that shows widespread applicability of heavy-tail distributions in modeling empirical data.

[3] The other notation in Table II is standard regression notation: df denotes the degrees of freedom. α and β are the regression coefficients as $y = \alpha x + \beta$, R^2 is the % of variance explained, see Equation eqn:r2, p is the p-value.



(a) (b)

Fig. 2. Histogram of empirical traffic densities of a location.



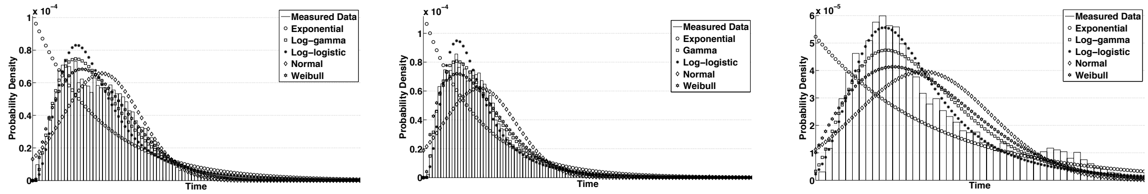
(a) Sydney (Nov' 10) (b) Sydney (Dec' 10)

Fig. 3. Traffic densities during two different time periods.

Data Preparation: We prepare traffic density data of each location as an individual time series. As shown in Table-I, we model 819 time series from six regions for several months. Specifically, we define $y_i(t)$ as the time series of vehicular traffic densities associated with i^{th} location at time t. Note that $y_i(t)$ is a time series of traffic density at each traffic location and is linearly related to the number of vehicles at that traffic location as described in the previous section. For each $y_i(t)$, we systematically calculate distribution parameters by using maximum likelihood estimation and estimate goodness-of-fit in the empirical data of each location against all the theoretical models using the Kolmogorov-Smirnov (KS) test [13].

Skewed Distribution: We start our observation by looking at the histogram of traffic densities of two different location as shown in Figure 2. A fairly smooth histogram is skewed-right with a possibility of large frequency of traffic occurring in the first half (average density value of 0.45) of the density distribution. For both the locations, the frequency mean is centered around 30 and the median bin is at 0.50. The spread of the empirical values shows that a wide range of traffic is present at these locations. We have recorded similar observations for other traffic locations as well. Previous studies have shown that skewed-right data are better modeled using Weibull-like distribution that have two parameters (shape and scale), unlike the exponential distribution, which has only one parameter (rate) [14]. Next, we use the methods proposed in [15] to check the stationarity, where the mean, variance, and autocorrelation of density distribution are all constant over time. In Figure 3, we have shown the distribution of traffic for two months for a location that despite few spikes looks stationary. This is an important step, as examination of fractal behavior in traffic requires stationarity criterion to be fulfilled.

Curve Fitting: Next, we consider the univariate distribution fitting using our theoretical models to the empirical traffic densities. In Figure 4, we show the empirical probability density function (PDF) plot for the fitted distribution of three different locations' traffic densities together with five other theoretical models. In Figure 4(a), Weibull distribution has estimated the parameters for empirical data quite well and the fitting agrees with the empirical PDF. In Figure 4(b), Log-



(a) Weibull (b) Log-gamma (c) Log-logistic

Fig. 4. Curve fitting for three different location time-series. In (a) Weibull, (b) Log-gamma, and (c) Log-logistic better models the empirical data.

gamma distribution has the least deviation and is able to model the empirical data quite well, the fitting shows that Log-gamma distribution is able to agree with the empirical PDF of traffic densities for that location. The last Figure 4(c) shows that the fitted log-logistic distribution very well models the empirical densities of the final location. Although, one may reason that the three samples corroborates efficacy of the heavy-tail, the analysis is trying to focus that memoryless distribution such as exponential deviate largely in accurately modeling empirical data. It can very well be said that model parameters of exponential distribution have underestimated the empirical data in all three cases, while the normal distribution has overestimated the skewed-right section of the empirical data. Thus, regression analysis indicates that heavy-tailed models such as Log-gamma, Log-logistic, and Weibull are better in estimating the parameters of empirical data.

Goodness-of-Fit Test: We extend our study of modeling to all the locations and perform a goodness-of-fit test as explained previously. The result of goodness-fit-test ranks various distributions based on the deviation of curves found and the values of estimators. We have observed that the traffic at individual locations can vary a lot, but in general Log-gamma, Log-logistic, and Weibull distribution can capture the key trends. In Table III, we have ranked the top three distributions that very well estimate the parameters of empirical data. Log-logistic distribution yields a much better fit for four out of six regions, while most of the traffic locations in London and Seattle are best described using Log-gamma and Weibull distribution respectively. The results of goodness-of-fit test also show that 91% of Connecticut, 70% of Sydney, and 80% of Washington D.C locations' empirical traffic data are better modeled using Log-logistic distribution. In Table III, we show dominant distributions at 3% and 5% deviation from the empirical data distribution, with most of the cases showing heavy-tail models are better suited for characterizing the empirical data. For example, at 3% deviation, 50% of locations' traffic can be modeled using Log-logistic distribution while at 5% deviation 93% locations' traffic can be modeled. These results strongly indicate that traffic at several locations in those six regions lasts for a long time that leads to congestion-like situation.

In Figure 5, we show the aggregate results of goodness-of-fit criterion for all 819 locations. Our results show that the distributions with heavy-tail properties are prominent in modeling the empirical data. In general, our results show that memoryless distributions such as exponential are insufficient to explain empirical distribution of traffic densities. We find that empirical values are better modeled using distribution with heavy-tails such as Log-gamma, Log-logistic, and Weibull. The Log-logistic distribution, particularly yields better fit for 57% of all aggregated locations in comparison to Log-gamma and

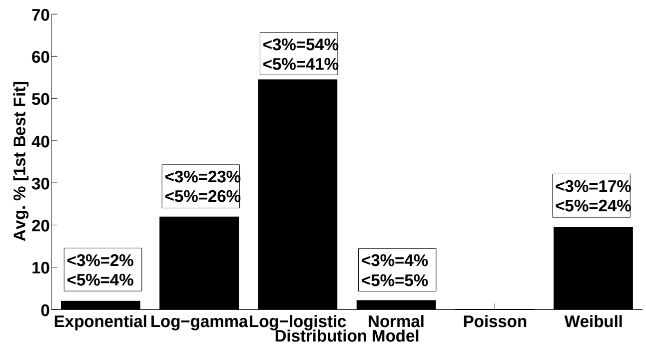


Fig. 5. The percentage for the distributions that cover all 819 locations from six regions. The values in the box show average percentage estimates deviation from empirical data.

Weibull, which are both around 20% mark. The Log-logistic also has less deviation at both 3% and 5% error level compared to all other models. These results imply that most of the urban traffic patterns are bursty in nature and traditional memoryless distributions are inadequate in realistically capturing the traffic patterns. Previous studies have shown the presence of heavy-tail indicate a self-similar behavior with noticeable bursts at a wide range of time scale [16], [17]. Next, we examine this conjecture by estimating the value of Hurst exponent.

IV. Analysis of Self-Similarity

In the previous section, we saw that theoretical distributions with heavy-tail properties are better at modeling the empirical traffic densities. Such distributions exhibit extreme variability in sampling. For example, sampling such distributions result in large quantities of very small values and few samples with extremely large values. Such type of behavior is known to cause long-range dependence and self-similarity of the network traffic [16], [17]. Here, we examine this self-similar nature in a vehicular setting and estimate the Hurst exponent. In [18], authors have studied dynamic behavior of a single vehicle moving through a sequence of traffic lights on a single-lane highway that has demonstrated the self-similar behavior. In [19], Meng et. al. examined the quantitative characteristics of the self similar vehicle arrival pattern on highways and headway distribution in traffic data provided by the Texas Department of Transportation. Using cellular automata model, Campari et. al. showed that highway traffic exhibits self-similarity in both car density and flow. They also concluded the fractal dimension increases from free flow to congested flow [20]. Although these studies strongly indicate that modeling arrival patterns is a challenging task, they have several short comings. First the dataset used were limited to few regions, second they were sampled from same type of traffic conditions (either highways or urban streets), and third they have been focusing on the arrival patterns of individual vehicles without

TABLE III. DOMINANT DISTRIBUTION AS BEST FITS[RANKED AND % DEVIATION]

Region	1^{st} Best Fit	2^{nd} Best Fit	3^{rd} Best Fit	$\leqslant$3%	$\leqslant$5%
Connecticut	L[91%]	G[5%]	W[4%]	L[50%], W[2%], G[1%]	L[93%], W[13%], G[10%], E[5%]
London	G[38%]	L[29%]	W[26%]	G[20%], L[15%], W[10%], N[8%]	G[55%], L[51%], W[44%], N[23%]
Sydney	L [70%]	G[17%]	W[14%]	L[65%], G[22%], W[8%],	G[49%], W[37%], N[6%]
Toronto	L[40%]	G[27%]	W[26%]	G[18%], W[17%], L[9%], E[3%]	W[72%], L[69%], G[63%], E[24%], N[1%]
Washington D.C.	L[80%]	W[11%]	G[7%]	L[60%], W[8%], G[6.54%], E[4%]	L[91%], W[35%], G[30%], E[14%]
Seattle	W[36%]	L[34%]	G[29%]	W[16%], G[14%], L[4%]	G[55%], W[47%], L[35%]
E = Exponential, G = Log-gamma, L = Log-logistic, N = Normal, W = Weibull					

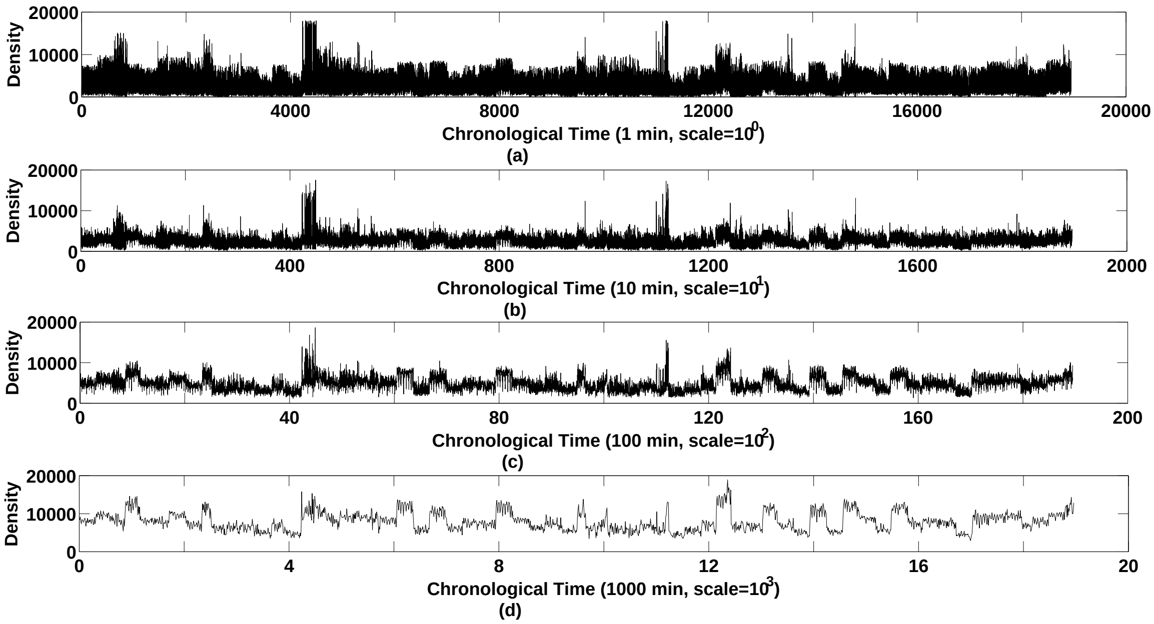


Fig. 6. Scaling of stochastic self similar vehicular traffic.

considering lane capacities and traffic densities that are important to study the vehicular congestion and future aspects of vehicular networking. In this paper, we specifically focus on examining the distribution of traffic densities on a planet scale level that help study road capacity and investigate the causes of widespread congestion in major metropolitan areas.

We benchmark the estimation of Hurst exponent using seven different estimators to study whether the observed traffic density time series of all 819 locations is self-similar in nature. The estimators we have used are: (i) Absolute Value Method, (ii) Aggregate Variance Method, (iii) Variance of Residuals Method, (iv) R/S Method, (v) Periodogram, (vi) Whittle Method, and (vii) Abry-Veitch Method. In general, the Hurst exponent is asymptotically estimated, which is prone to statistical uncertainty and errors. By applying many estimators we take a comprehensive approach in analyzing the self-similar behavior. A detailed information about these estimators is given in [16], [17], [21].

The first indication that traffic time series have a stochastic self-similar process is visually depicted in a series of plots in Figure 6. In Figure 6 (a) the traffic trace is plotted against a time granularity of 1 minute. Figure 6 (b) is the same traffic time series aggregated by a factor of 10 (i.e. the time scale is compressed at 10 minutes). Subsequent plots of Figure 6 (c) and (d) are aggregated by increasing the granularity by two more orders. These plots look very similar to long range dependence and are invariant to the chosen time granularity.

We use the Selfis tool [21] to investigate the value of Hurst exponent using seven different estimators at 95% confidence

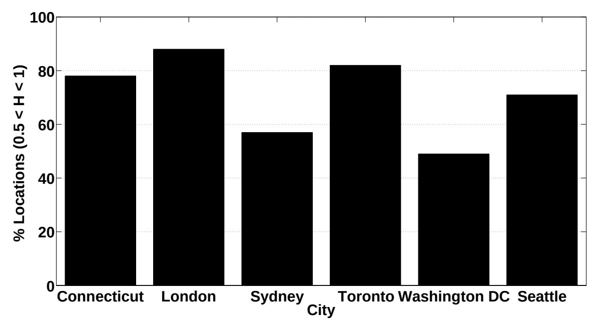
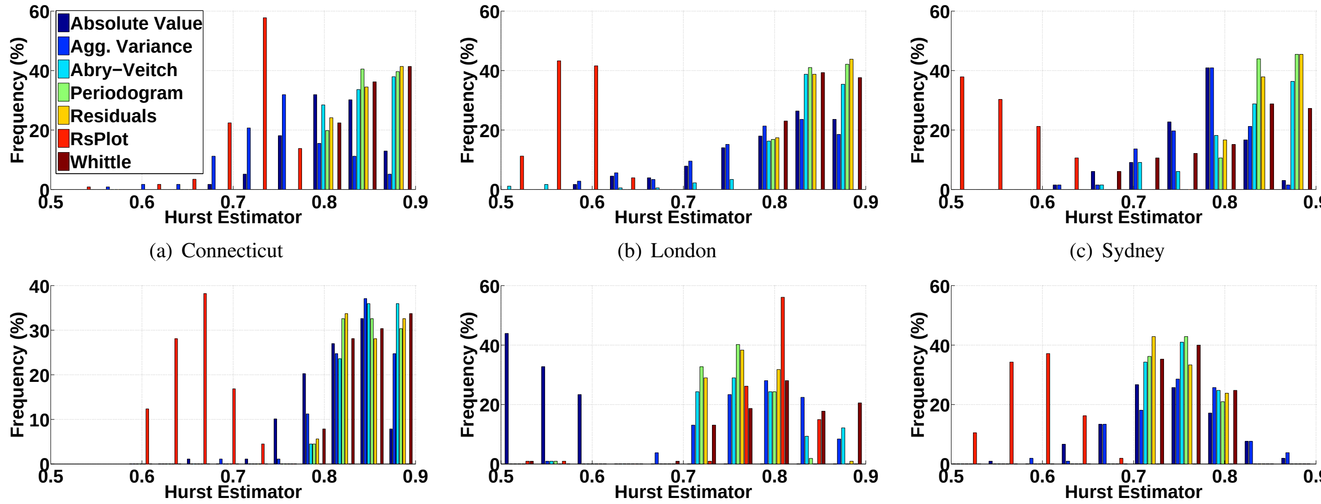


Fig. 8. The percentage of locations from every region that have self-similarity in their traffic patterns.

interval for all 819 time series. A Hurst exponent at (or very close to) 0.5 indicates lack (or weakness of) the long range dependence and suggests a short range memory process, while an exponent higher than 0.5 and closer to 1 indicates a strong long-range dependence and suggests a self-similar process.

In Figure 7, we show the bar-plots of the distributions of estimated Hurst exponent by all the seven estimators for all six regions. As evident from these plots, the Hurst estimators consistently produce results well-exceeding 0.5. These trends support the adequacy of self-similar processes in modeling the vehicular traffic time series over various time scales. This result is also in-line with our previous findings that Log-gamma, Log-logistic, and Weibull (considered in the family of heavy-tail distributions) provide a better distribution fit for observed traffic densities, and also supports the failure of memoryless distributions. As evident, average Hurst exponent is greater than 0.5 with low deviation at 95% confidence interval. Finally,



(a) Connecticut (b) London (c) Sydney

(d) Toronto (e) Washington DC (f) Seattle

Fig. 7. Histograms showing the distributions of Hurst exponent, which is estimated from seven different estimators for the locations of six regions.

we show the aggregate statistics for all self-similar time series in Figure 8. This result indicates that the traffic on more than 70% locations of London, Connecticut and Toronto is self-similar, while less than 50% of Sydney and Washington D. C. traffic is self-similar.

Overall, our analysis from section III and IV shows that the current traffic trends are better modeled using heavy-tail distributions. In general Log-gamma, Log-logistic and Weibull distribution are better in modeling empirical data of traffic densities, as showed by the goodness-of-fit test. Also, the traffic time series exhibit self-similar process.

V. Conclusion and Future Work

In this paper, we have shown that the observed values of vehicular traffic densities are better modeled using heavy-tail distributions. Since a heavy-tail distribution also indicates a self-similar process, an investigation into that direction showed self-similarity in the time series of traffic densities. In all, we examined 819 locations' vehicular traffic density data, containing more than 25 million records. Our first analysis on modeling and characterization indicates that heavy-tail distributions such as Log-gamma, Log-logistic, and Weibull better model these observed traffic densities. In the second analysis, we found that time series of traffic densities are self-similar, as estimated by seven different estimators for the Hurst exponent ($0.5 \leq H \leq 1.0$). These results suggest that the traditional notion of using memoryless models for traffic modeling purposes should be revisited. We believe that our study will provide new insight into the development of future vehicular networks and infrastructure design.

References

[1] L. Briesemeister, L. Schafers, and G. Hommel, "Disseminating messages among highly mobile hosts based on inter-vehicle communication," in *Intelligent Vehicles Symposium,*, 2000.

[2] J. Singh, N. Bambos, B. Srinivasan, and D. Clawin, "Wireless lan performance under varied stress conditions in vehicular traffic scenarios," in *Vehicular Technology Conference*, 2002.

[3] F. Bai and B. Krishnamachari, "Spatio-temporal variations of vehicle traffic in vanets: facts and implications," in *ACM VANET*, 2009.

[4] N. Wisitpongphan, F. Bai, P. Mudalige, V. Sadekar, and O. Tonguz, "Routing in sparse vehicular ad hoc wireless networks," *Selected Areas in Communications, IEEE Journal on*, vol. 25, no. 8, oct. 2007.

[5] G. S. Thakur, P. Hui, and A. Helmy, "Modeling and characterization of urban vehicular mobility using web cameras," *IEEE Infocom workshop on Network Science for Communication Networks (NetSciCom)*, 2012.

[6] G. S. Thakur, P. Hui, H. Ketabdar, and A. Helmy, "Spatial and temporal analysis of planet scale vehicular imagery data," in *Data Mining Workshops (ICDMW), IEEE International Conference on*, dec. 2011.

[7] G. S. Thakur, M. Ali, P. Hui, and A. Helmy, "Comparing background subtraction algorithms and method of car counting," *ArXiv*, 2012.

[8] C. Stauffer and W. Grimson, "Adaptive background mixture models for real-time tracking," in *CVPR IEEE*, vol. 2, 1999.

[9] Y. Benezeth, P. Jodoin, B. Emile, H. Laurent, and C. Rosenberger, "Review and evaluation of commonly-implemented background subtraction algorithms," in *ICPR*, 2008.

[10] Y. Sheikh and M. Shah, "Bayesian modeling of dynamic scenes for object detection," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 2005.

[11] D. A. Forsyth and J. Ponce, *Computer Vision: A Modern Approach*. Prentice Hall Professional Technical Reference, 2002.

[12] A. Clauset, C. R. Shalizi, and M. E. J. Newman, "Power-law distributions in empirical data," *SIAM Review*, vol. 51, pp. 661–703, 2009.

[13] J. Massey, Frank J., "The kolmogorov-smirnov test for goodness of fit," *Journal of the American Statistical Association*, vol. 46, 1951.

[14] R. Groeneveld, "Skewness for the weibull family," *Statistica Neerlandica*, vol. 40, no. 3, pp. 135–140, 1986.

[15] R. Clegg, "A practical guide to measuring the hurst parameter," in *Proceedings of 21st UK Performance Engineering Workshop, School of Computing Science, Technical Repo*, N. Thomas. N. Thomas, 2005.

[16] R. J. Adler, R. E. Feldman, and M. S. Taqqu, Eds., *A practical guide to heavy tails: statistical techniques and applications*. Cambridge, MA, USA: Birkhauser Boston Inc., 1998.

[17] K. Park and W. Willinger, *Self-Similar Network Traffic and Performance Evaluation*, 1st ed. New York, NY, USA: John Wiley Sons., 2000.

[18] T. Nagatani, "Self-similar behavior of a single vehicle through periodic traffic lights," *Physica A: Statistical Mechanics and its Applications*, vol. 347, pp. 673 – 682, 2005.

[19] Q. Meng and H. L. Khoo, "Self-similar characteristics of vehicle arrival pattern on highways," *Journal of Transportation Engineering*, 2009.

[20] E. Campari and G. Levi, "Self-similarity in highway traffic," *The European Physical Journal B - Condensed Matter and Complex Systems*, vol. 25, pp. 245–251, 2002, 10.1140/epjb/e20020028.

[21] T. Karagiannis, M. Faloutsos, and M. Molle, "A user-friendly self-similarity analysis tool," *SIGCOMM CCR*, 2003.