Spatial Point Process Modeling of Vehicles in Large and Small Cities

Qimei Cui, Ning Wang
School of Information and Communication Engineering
Beijng University of Posts and Telecommunications
Beijing, 100876, China
Email: {cuiqimei, wangning}@bupt.edu.cn

Martin Haenggi
Department of Electrical Engineering
University of Notre Dame
Notre Dame, IN 46556, USA
Email: mhaenggi@nd.edu

***Abstract*—The uncertainty in the locations of vehicles on streets induced by vehicles passing and queueing make the spatial modeling of vehicles a difficult task. To analyze the performance of vehicle-to-vehicle (V2V) communication for vehicular ad hoc networks (VANETs), accurate spatial modeling is of great importance. In this paper, we concentrate on spatial point process modeling for random vehicle locations in large and small cities, performing empirical experiments with real location data of mobile taxi trajectories recorded by the global positioning system (GPS) in Beijing city of China and Porto city of Portugal. We find that the empirical probability mass functions (PMFs) of the number of taxis in test sets in different regions of Beijing or in Porto all follow a negative binomial (NB) distribution. Based on the above, we show that the Log Gaussian Cox Process (LGCP) model, whose empirical PMF nicely fits the NB distribution, accurately characterizes diverse spatial point patterns of random vehicle location in both large and small cities. This is verified by the minimum contrast method. The LGCP model can be applied to analyze performance metrics (i.e., connectivity, coverage, and capacity) and optimize the practical deployments of VANETs.**

I. Introduction

A. Motivation

Spatial stochastic models have become increasingly popular in telecommunication networks in recent years. By taking the geometric structure of network architectures into consideration, stochastic geometry models offer a more relevant view to location-dependent network characteristics than conventional hexagon network models. Similarly, the dynamics of the vehicles in vehicular ad hoc networks (VANETs) create rapid changes of network topology, leading to inevitable uncertainty in the spatial pattern of the vehicle density on a road. To describe the vehicular patterns and characterize the performance of vehicle-to-vehicle (V2V) communication, spatial stochastic models are crucial.

B. Related Work

A road system is first added to wireless communication systems in [1] for strategic planning and economic analysis, and the Poisson line process is proposed to model the road position, and the Poisson point process (PPP) is used to model traffic on the road. Reference [2] models the telecommunication networks including the road system as random geometric graphs, on the edges of which the locations of the network nodes are modeled as linear Poisson processes. Reference [3] observes that the vehicles that enter the highway through one of the traffic entry points form a Poisson process and studies the impact of vehicle mobility on the connectivity of VANETs. Connectivity is also analyzed in [4] and [5], where the same model for vehicles entering the highway is used and, in addition, it is assumed that the speed of the vehicles is constant or normally distributed. In [6] and [7], the vehicle locations are assumed to form a Poisson point process in one dimension, and the authors perform an analytical evaluation of broadcast protocols. In [8], the random locations of the vehicles in a street are modeled as a one-dimensional stationary Cox process with Fox's H-distributed random intensity. In [19], graph-based optimal revenue packet scheduling in Vehicle-to-Infrastructure communication is discussed.

Besides the above theoretical studies, empirical research based on real data recently has focused on spatial point process modeling and analysis of real road systems or random vehicle location or traffic. For example, [9] fits Poisson line tessellations, Poisson-Voronoi tessellations, and Poisson-Delaunay tessellations to real data of the road system in Paris for the purpose of cost analysis and strategic planning of telecommunication networks. The authors in [10] further utilize the random tessellations to model the road systems in fixed-access networks, and random points are added to the tessellation edges to represent network components. Reference [11] uses a nonhomogeneous Poisson process model for vehicles at traffic signals, and shows that the introduction of traffic signals does not affect the Poisson property of the stochastic model when the vehicles are with deterministic velocity profile, which is validated against empirical data in London city.

In [12], the authors develop different spatial point process models for base station in cellular networks based on the empirical data. In [13], by analyzing real data of traffic counts in highway from the Bureau of Transport Statistics, New South Wales, the vehicles are modeled as a PPP, and numerical evaluation also indicates the proposed model is more accurate than existing one-dimensional models under sparse vehicle traffic densities. It is verified that the two-dimensional PPP assumption is only accurate for low vehicle density in [14]. In [15], experimental results for single-vehicle data of the Dutch

978-1-5090-5019-2/17/$31.00 ©2017 IEEE

freeway A9 and the German freeway A5 display that when the distances between cars are correlated due to traffic congestion, the vehicles do not follow a Poisson distribution.

Based on the above, the previous research work has not produced a widely accepted spatial model that well characterizes the random vehicle locations, although the two-dimensional model is closer to reality. In urban VANET scenarios, it is of great importance to account for random vehicle locations in V2V communication. The dynamics of vehicle passing and queueing as well as high mobility create inevitable uncertainty in the spatial pattern of the vehicles on the road. Therefore, spatial point process modeling of random vehicle locations based on real data analysis and mining is still a crucial problem to be addressed.

C. Contribution and Paper Organization

In this paper, we concentrate on spatial point process modeling and analysis of random vehicle locations in large and small cities—Beijing city of China with a large-scale and regular road system, and Porto city of Portugal with a small and irregular road system. We perform empirical experiments with the global positioning system (GPS) traces of taxis in Beijing and Porto. The data of Beijing records realtime GPS location information for 12509 taxis over a month, and the data of Porto consists of realtime GPS location information for 442 taxis over a year. Our goal is to identify a planar point process to accurately model random vehicle locations using tools from stochastic geometry.

We firstly analyze the empirical probability mass functions (PMFs) of point counts of the point pattern generated by the vehicles. We find that the empirical PMFs of the number of taxis in test sets in different regions of Beijing or in Porto all follow a negative binomial (NB) distribution. On the basis of the above, the conclusion is obtained that the LGCP model whose empirical PMF nicely fits the NB distribution can characterize diverse spatial point pattern of random vehicle location very well, which is verified by the minimum contrast method and Monte Carlo test. The general LGCP model can be applied to analyze performance metrics (i.e., connectivity, coverage and capacity) and optimize the practical deployment for VANET.

The rest of the paper is organized as follows. Section II presents spatial point models and characteristics. Section III is devoted to the analysis of the empirical PMFs of the counting measure of the vehicle point process. We fit several spatial point process models to real vehicle location data by the minimum contrast method in Section IV. Finally, concluding remarks are given in Section V.

II. Spatial Point Process Models and Characteristics

A. Spatial Point Process Models

There are several kinds of spatial point processes to describe a collection of points in two dimensions, such as PPPs, cluster processes, hard-core processes, Cox processes, and Gibbs processes [16, Ch. 3]. Each of them have different characteristics—PPPs exhibit complete spatial randomness due to their independence property; Cox processes and cluster processes are overdispersed relative to PPPs, i.e., they are more irregular; hard-core processes have a minimum distance between points and thus are more regular than the PPPs; Gibbs processes may be overdispersed or underdispersed. Vehicles are often clustered due to traffic congestion and intersections, so Thomas cluster processes, Matérn cluster processes, and LGCPs are promising candidates modeling the vehicle pattern.

B. Spatial Point Process Characteristics

There are five classical statistics called G, F, J, K and L functions to describe the inter-point "dependence" and "clustering" [16, Ch. 3]. The G function is the nearest-neighbor distance distribution. The F function, also called the empty space function, is the cumulative distribution function of the distance from a fixed location to the nearest point of the spatial point process. The J function is defined as $J(r) \triangleq \frac{1-G(r)}{1-F(r)}$, which is a measure of how close a process is to a PPP. Ripley's K function is defined that $\lambda K(r)$ is the expected number of additional random points within a distance r of the typical point of the point process, where λ is the intensity of the process. The L-function is a transformation of Ripley's K function, defined as $L(r) \triangleq \sqrt{\frac{K(r)}{\pi}}$.

III. PMF Analysis of Real Location Data

A. Spatial Point Specifications

A single realization of a spatial point process is called a deterministic point pattern [16]. We treat the taxi locations as a realization of a point process and aim at finding the point process that has the highest likelihood of reproducing a realization containing the taxi locations of Beijing and Porto.

1) Beijing: The data set of Beijing comprises GPS information of 12509 taxis for a month (from 2012/11/01 to 2012/11/27). It contains 785.4 million entries, each one comprised of the taxi location and metadata information associated with each taxi, as shown in Table I. The frequency of recording position information varies from one to six times per minute.

Beijing city is the capital of China with a population of 21 million on an area of 16,000 km^2; its road system is large, and relatively regular. The intensity of the taxis in Beijing is not constant and varies with the location due to the impact of the road systems. Fig. 1 gives a realization of the point process comprised of 6927 vehicles in Beijing at 08:30 am on Nov. 2, 2012. From Fig. 1, a stationary point process appears unsuitable to accurately model the entire data set. Hence we partition the city into 9 regions in which the point pattern is relatively homogeneous. By doing so, we can fit a stationary point process model to each region. The points in Region 2 and Region 5 are densely distributed, the points in Region 9 exhibits strong clustering, and the patterns in the other regions are relatively sparse.

TABLE I
FORMAT FOR BEIJING DATA SET

taxi ID	status	time	longitude	latitude	velocity
470341	0	20121102084536	116.5713	39.8063	39.8

TABLE II
FORMAT FOR PORTO DATA SET

trip ID	taxi ID	timestamp	polyline
T16	20000440	1408037740	[-8.618,41.136],[-8.618,41.135]$\cdots$

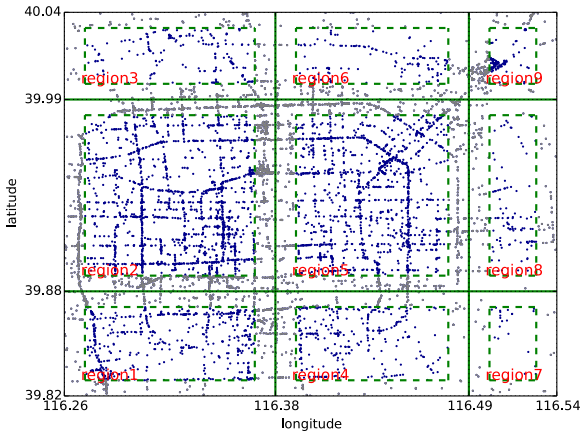

Fig. 1. The taxi distribution at 8:30 am on 2012/11/02 in Beijing (116.26E-116.54E, 39.82N-40.04N). The green dashed line are sampling window in which test set centers are constrained.

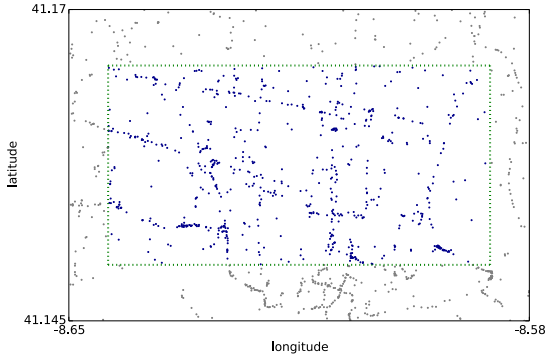

Fig. 2. The taxi distribution at 9:00 am (2013/07/01-2013/07/31) in Porto (8.65W-8.58W, 41.145N-41.17N). The green dashed line are sampling window in which test set centers are constrained.

2) Porto: The data set of Porto includes GPS information of 442 taxis during a complete year (from 2013/07/01 to 2014/06/30). There are 1.7 million entries, of which each one represents a complete taxi trajectory, as shown in Table II. The complete taxi trajectory is a sequence of GPS positions measured every 15 seconds.

Porto is a small port city in Portugal with a population of 0.26 million on an area of 41 km^2; its road system is small and irregular. The spatial point pattern generated by taxi data of Porto is more homogeneous. A single realization is depicted in Fig. 2, which consists of 1245 taxi locations at 9:00 am over a month (2013/07/01 to 2013/07/31). Different from Beijing, we choose the union of 31 daily snapshots taken at 9:00 am because the number of points at 9:00 am on a single day is too small (less than 50) to be statistically significant.

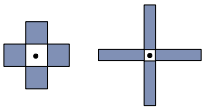

Fig. 3. The cross rectangle in the Palm sampling method. The ratio of arm length to arm width in the left subfigure is 1, and that in the right subfigure is 4. Note that the center part is excluded.

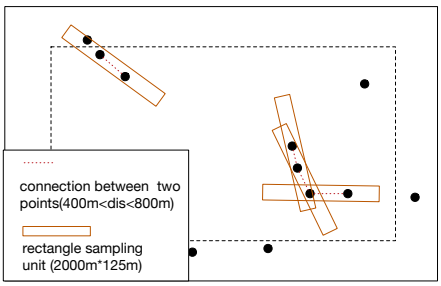

Fig. 4. The 2-Palm sampling methods. 'dis' in the legend refers to the distance between two points.

B. Empirical PMFs with Different Sampling Methods

In a spatial model, the countable random collection of the vehicles on the Euclidean space $\mathbb{R}^2$ is regarded as a point process $\Phi = \{x_1, x_2, \cdots\}$, where x_i denotes a vehicle location in the point process. The counting measure $\Phi(B) \triangleq \#\{\Phi \cap B\}$ is a random variable that denotes the number of vehicles in a Borel set $B \subset \mathbb{R}^2$ [16]. Similarly, we use the deterministic counterpart $\varphi(B) \triangleq \#\{\varphi \cap B\}$ to denote the counting measure of a point pattern φ.

To calculate the empirical PMFs of the point counts $\varphi(B)$, the first step is to sample and count the number of points generated by the taxi data. The basic principle of sampling and counting is to choose many test sets B_i so that they cover an entire region in a uniform manner, and count the number of points falling in B_i.

We consider four shapes—circle, square, rectangle, and cross rectangle—for sampling. The cross rectangle is the union of a horizontally and a vertically oriented rectangle where the center square is excluded, as illustrated in Fig. 3. As shown in Table III, the area of the test sets is fixed to 0.25 km^2 to maintain consistency. To eliminate boundary effects, we use a sampling window that is smaller than the region. The gap between the sampling window boundary and the region boundary is set as 1000 m for Beijing or 500 m for Porto, so that the test sets do not exceed the region. The four sampling methods are:

1) Lattice Sampling: A sampling method in which the center of B follows a square-lattice distribution.

2) Uniform Random Sampling: A sampling method where the center of B follows a uniform PPP.

3) Reduced Palm Sampling: A sampling method in which B is centered at the location of the taxis.

4) Reduced 2-Palm Sampling: A sampling method where the center of B is at the midpoint of two taxi locations. The distance between any two taxis vary from 400 m to 800 m. The sampling method is shown in Fig. 4.

Fig. 5 shows the empirical PMFs and the corresponding cumulative distribution functions (CDFs) of the point counts with lattice sampling and uniform-random sampling methods in Region 1. Fig. 6 shows the results of reduced Palm sampling methods based on four shapes and reduced 2-Palm sampling method in Region 1. The Palm sampling and 2-Palm sampling are used to observe the impact of the street. Note that 'Palm-Xrectangle1' refers to Palm-sampling method for the cross-rectangle case in which the ratio of arm length to arm width is 1, while the ratio is 4 in 'Palm-Xrectangle4' as shown in Fig. 3. For Palm sampling, the cross rectangular shape of test set is to covers the street that the taxi lie on, given that most of the streets in Beijing are north-south or east-west. For 2-Palm sampling, the rectangular shape is the optimal choice because this way the test set covers the street that the two taxis lie on.

The empirical PMF curves in Fig. 5 are monotonically decreasing, while the empirical PMF curves in Fig. 6 first increase then decrease with increasing point number. This indicates that lattice sampling and uniform random sampling are definitely different from Palm and 2-Palm sampling and that different shapes have little impact on the empirical PMF calculation. Observing both results, we find that the empirical PMF of the point number less than 3 in Palm and 2-Palm sampling is smaller than that in lattice sampling and uniform random sampling because the former is constrained to sample in the vicinity of taxis. The same conclusion can be drawn for the other 8 regions. Moreover, random sampling and uniform random sampling are similar, and the shapes of the test sets barely affect the PMFs. Hence the following experiments only adopt the lattice sampling method.

TABLE III
SAMPLING METHODS

Sampling method	Shape of test set	Size of test set	Area of test set (km^2)
Lattice sampling	circle	radius=281m	0.25
	square	500m×500m	0.25
Uniform random sampling	circle	radius=281m	0.25
	square	500m×500m	0.25
Palm sampling	circle	radius=281m	0.25
	square	500m×500m	0.25
	cross ectangle	250m×250m×4	0.25
		500m×125m×4	0.25
2-Palm sampling	rectangle	2000m×125m	0.25

C. Statistics on the Basis of the PMFs

On the basis of the empirical PMFs, we evaluate three statistics—empirical mean, variance, and intensity. We use φ to denote the point pattern generated by taxi data. The empirical mean in Region i is given by

$$M = \sum_{k=1}^{K} \frac{\varphi_i(B_k)}{K}, \tag{1}$$

where φ_i is the data set of Region i, K is the number of test sets $B \subset \mathbb{R}^2$, and B_k denotes the kth test set.

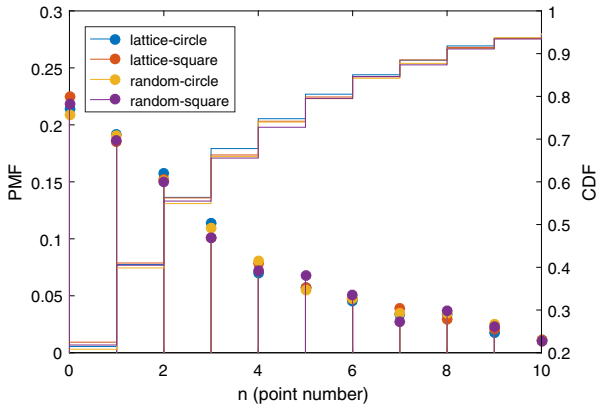

Fig. 5. The empirical PMFs and corresponding CDFs of point number in test set in Region 1 of Beijing with lattice and uniform random sampling methods.

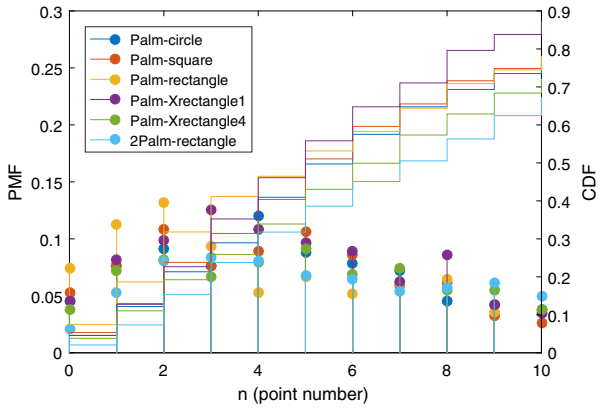

Fig. 6. The empirical PMFs and corresponding CDFs of point number in test set in Region 1 of Beijing with reduced Palm and 2-Palm sampling methods.

The variance is derived using the empirical mean as

$$\sigma^2 = \sum_{k=1}^{K} \frac{[\varphi_i(B_k)]^2}{K} - M^2. \tag{2}$$

The intensity is the expected number of points in a test set B, and is given by

$$\lambda = \frac{M}{|B|} = 4M/\text{km}^2. \tag{3}$$

The results shown in Table IV verify that the taxi points do not follow a Poisson distribution whose mean and variance are the same. NB models are suitable for discrete data like count

TABLE IV
MEAN, VARIANCE AND INTENSITY IN REGION 1 OF BEIJING

Sampling method	Shape of test set	Mean	Variance	Intensity (/km^2)
Lattice sampling	circle	3.22	13.48	12.89
	square	3.24	13.55	12.97
Uniform random sampling	circle	3.31	13.72	13.24
	square	3.32	13.99	13.26
Palm sampling	circle	6.93	28.3	27.73
	square	6.74	27.82	26.96
	cross rectangle	5.81	18.76	23.25
		7.46	26.77	29.85
2-Palm sampling	rectangle	9.21	52.25	36.83

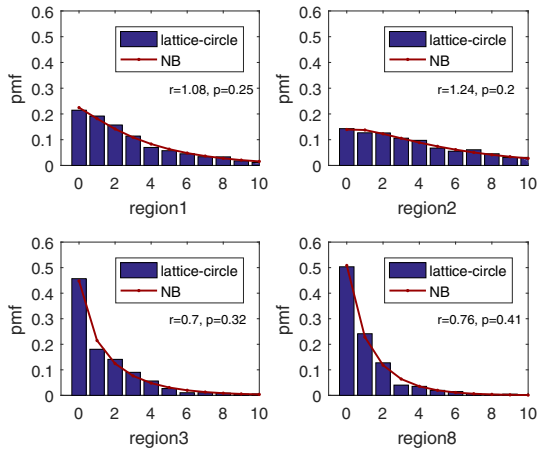

Fig. 7. Empirical PMFs of point number in test set in Regions 1, 2, 3, 8 of Beijing with lattice-circle sampling, and fitted PMFs of NB variables.

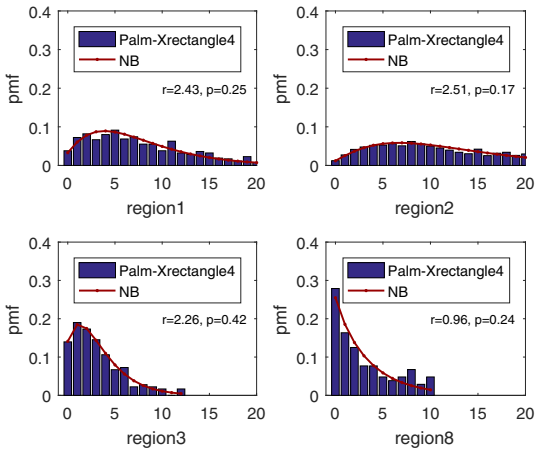

Fig. 8. Empirical PMFs of point number in test set in Regions 1, 2, 3, 8 of Beijing with Palm-Xrectangle4 sampling, and fitted PMFs of NB variables.

data when the variance is larger than the mean. The PMF of a NB random variable N is

$$\mathbb{P}(N=k)=\binom{k+r+1}{k}p^k(1-p)^r, \tag{4}$$

where $r>0$ and $0<p<1$ are parameters of the distribution. Here, we attempt to fit the discrete empirical PMF data to the PMF of the NB distribution.

We use the maximum likelihood method to estimate the parameters r and p of the NB distribution. The experimental results show that the empirical PMFs in Regions 1 and 4 are very close, coinciding with that in Regions 2 and 5 and that in Regions 3, 6 and 7. Thus, only the empirical PMF data in Regions 1, 2, 3 and 8 with three different methods are provided in Fig. 7, Fig. 8, and Fig. 9 as typical representatives. Judging from the results, it is observed that the empirical PMFs in all the regions can fit NB distributions with specific r and p parameters well. The fitted NB distributions in Region 2 and Region 1 that belong to dense regions are both close under different sampling methods. Similarly, the fitted NB distributions in Region 3 and Region 8 that belong to sparse regions are approximate under lattice sampling method, but there exists significant discrepancy in the parameters for Palm and 2-Palm sampling methods due to the distinctive sparsity in regions. Fig. 10 shows the empirical PMF data in Porto, which can also fit NB distributions very well.

Remark 1: The empirical PMFs of the number of points in test set in different regions of Beijing and in Porto follow a NB distribution, reflecting self-similarity in dense regions or in sparse regions and in large-scale and regular road systems or in small and irregular road systems.

IV. Fitting Method and Pattern Modeling

In this section, we develop a point process model Φ, whose statistics match those of the taxi data.

First, the three candidate point process models—the Thomas cluster Process, the Matérn cluster process, and the LGCP—are described in this section. A general cluster Poisson process

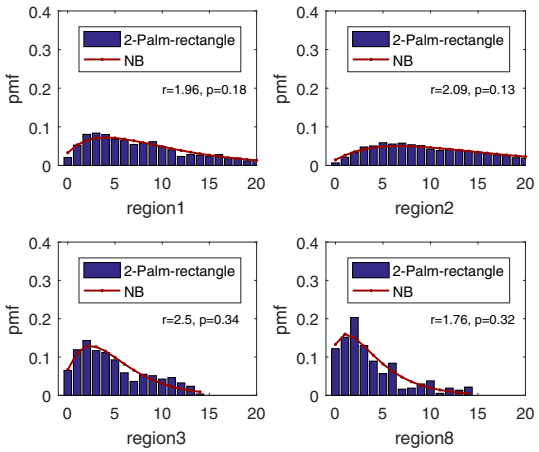

Fig. 9. Empirical PMFs of point number in test set in Region 1, 2, 3, 8 of Beijing with 2-Palm-rectangle sampling, and fitted PMFs of NB variables.

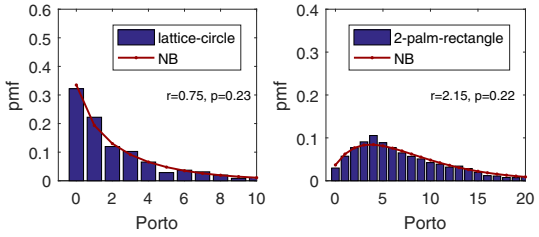

Fig. 10. Empirical PMFs of point number in test set of Porto with lattice-circle and 2-Palm-rectangle sampling, and fitted PMFs of NB variables.

is generated by taking a parent point process and daughter point processes for each parent, and translating the daughter processes to the position of their parent. The cluster process is then the union of all the daughter points. The Thomas cluster Process and the Matérn cluster process are defined as [16, Ch. 3]

1) Thomas Process: A doubly Poisson cluster process, where the intensity function of a cluster is given by

$$\lambda_0(x)=\frac{\overline{c}}{2\pi\delta^2}\exp(-\frac{\|x\|^2}{2\delta^2}), \tag{5}$$

i.e., the daughter points are normally scattered with variance δ^2 around each parent point, and the mean number of daughter point is $\bar{c}$.

2) Matérn Cluster Process: A doubly Poisson cluster process, where the intensity of the cluster can be expressed as

$$\lambda_0(x) = \frac{\bar{c}}{\pi r^2}\mathbf{1}_{b(o,r)}(x), \tag{6}$$

where $\mathbf{1}(.)$ is the indicator function. i.e., the daughter points are uniformly scattered on the ball of radius r centered at each parent point, and the mean number of daughter points is $\bar{c}$.

A general Cox process is a doubly stochastic Poisson process where the intensity measure itself is random. The intensity measure is a realization of a non-negative locally finite random measure [16, Ch. 3]. The LGCP is a specific type of Cox process.

3) Log Gaussian Cox Process: A Cox process where the logarithm of the intensity function is a Gaussian process [18]. Specifically, the random intensity function of a LGCP is given as $\Lambda(s) = \exp\{Y(s)\}$ where $Y = \{Y(s) : s \in \mathbb{R}^2\}$ is a real-valued Gaussian process. $Y(s) \sim \mathcal{N}(\mu, C)$ means the random function Y is distributed as a Gaussian process with mean function μ and covariance function C. The exponential covariance can be parametrized in the form

$$C(r) = \beta \exp\left(-\frac{r}{\alpha}\right). \tag{7}$$

Based on the results we found in Section III-C, we fit the Thomas model, the Matérn cluster model, and the LGCP model to the given point set, using the minimum contrast method [17]. It fits the model by matching the data's summary statistic to its theoretical value. The K function describes the correlation between points which makes it a suitable summary statistic for fitting.

For the ease of notation, the parameters of any point process model are collectively referred to as θ. For example, in the case of LGCP, θ refers to (α, μ, β). We determine the values of the parameters of a point process which give the closest match between the theoretical expected value of the summary statistic $K_\theta(r)$ and the observed value of the summary statistic evaluated from the data, denoted by $\hat{K}(r)$. The best match is determined by minimizing the discrepancy D between two functions, which is defined as

$$D(\theta) = \int_a^b \left|\hat{K}(r)^q - K_\theta(r)^q\right|^p dr, \tag{8}$$

where $0 \leq a < b$, and $p, q > 0$ are exponents. We employ a generic fitting algorithm for the method of minimum contrast, since the theoretical K function of the three models can be computed exactly from the model parameters.

Fig. 11 and Fig. 12 compare the K function of the observed point patterns in Region 3 of Beijing, and Porto, to the K function of the fitted models, and depict the gap between the K function of the observed point pattern and the fitted models. The gap between the $K(r)$ curve of the observed point pattern and the curve of the LGCP model with the NB distributed

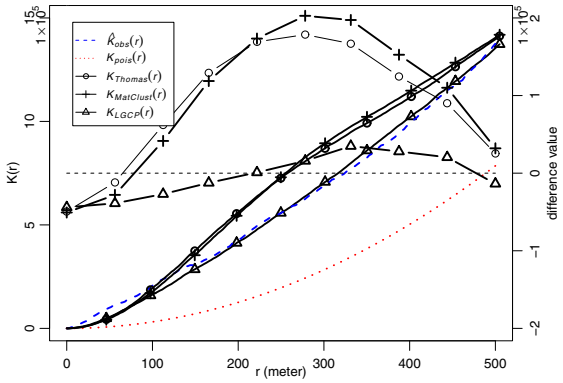

Fig. 11. Left axis:K function of the observed point pattern in Region 3 of Beijing and three fitted models. Right axis: the difference between the K function of the observed point pattern and the other curves of three models.

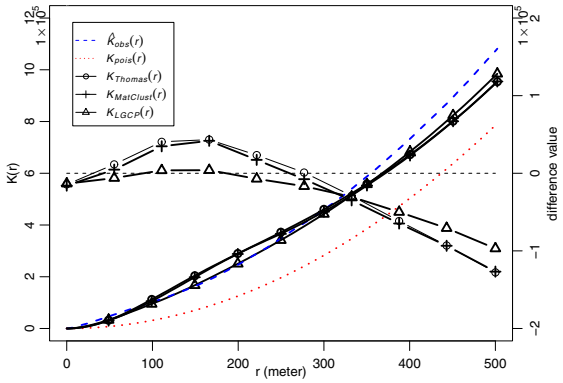

Fig. 12. Left axis:K function of the observed point pattern in Porto and three fitted models. Right axis: the difference between the K function of the observed point pattern and the other curves of three models.

PMF is the least, making it a good fit for the model. The same conclusion can also be derived in the other regions of Beijing. The LGCP model is uniquely determined by parameters $\theta = (\alpha, \mu, \beta)$. The values of the parameters of fitted LGCP model for Region 3 of Beijing and Porto are $(177.20, -13.30, 2.62)$, and $(105.56, 10.57, 2.22)$, respectively.

We deduce that the LGCP is an accurate model for the spatial point patterns of vehicle locations. Additionally, we perform Monte Carlo test to determine the goodness-of-fit between the empirical and fitted models. In the Monte Carlo test, we regard the LGCP model as the null hypothesis, and the fitted LGCP model as alternative hypothesis. To determine whether the spatial point patterns of random vehicle location follow LGCP, we compare the K functions of the observed data and the fitted LGCP. Naturally, it is not possible to obtain exactly matching K function curve of LGCP because of random variability. Thus, we simulate realizations from the null hypothesis and compute the K function for each realization. The K function curves of the realizations form an envelope for a large number of simulations. The Monte Carlo test rejects the null hypothesis, if the K function of the original spatial point pattern lies outside the envelope. As shown in Figs. 13 and 14, the observed curve stays inside the envelope of the LGCP fitted to the original point pattern, ascertaining that the original point pattern can be characterized

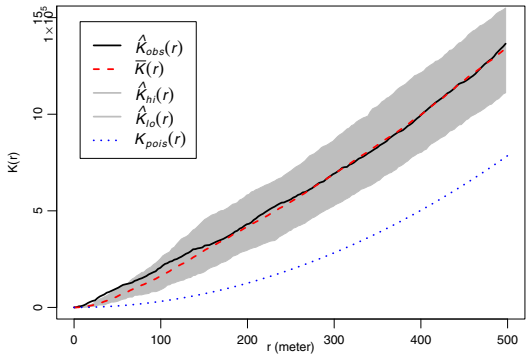

Fig. 13. Monte Carlo test on K function of taxis in Region 3 of Beijing, and the envelope of 39 realizations of the LGCP model. The red dashed line is the average value of the K function of 39 realizations of the fitted LGCP model.

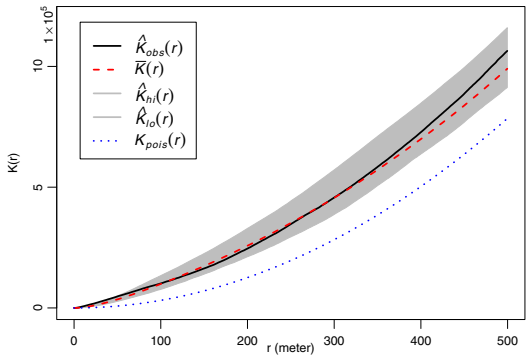

Fig. 14. Monte Carlo test on K function of taxis of Porto, and the envelope of 39 realizations of the LGCP model. The red dashed line is the average value of the K function of 39 realizations of the fitted LGCP model.

by LGCP. Therefore, the LGCP model whose empirical point count PMF nicely fits a NB distribution can be used to model the random vehicle locations.

V. Conclusion

We adopted various sampling methods to study the distribution of vehicles in Beijing, with large-scale and regular road systems, and Porto, with small-scale and irregular road system. We showed that the number of vehicles in both the cities follow a NB distribution through probability mass function analysis. To find a suitable point process for the vehicular network, we fitted the Thomas process, Matérn cluster process, and the LGCP with the PMF following a certain NB distribution to the given point set using the minimum contrast method. It is shown that the LGCP model whose empirical PMF nicely fits the NB distribution can generally characterize the spatial point pattern of random vehicle locations in both large and small cities. We also use a Monte Carlo test on the point count in test sets and classical statistics as the criteria for goodness-of-fit, which proves that the LGCP model performs better in fitting than other models. We can use the LGCP model to analyze performance metrics (i.e., connectivity, coverage, and capacity) and optimize practical deployment for VANET.

Acknowledgment

The work was supported by National Nature Science Foundation of China Project (Grant No. 61471058), Key National Science Foundation of China (61461136002, 61631005), the Funds for Creative Research Groups of China under Grant 61421061,Hong Kong, Macao and Taiwan Science and Technology Cooperation Projects (2016YFE0122900) and the 111 Project of China (B16006). The partial support of the U.S. National Science Foundation through grant CCF 1525904 is gratefully acknowledged. We also thank Jeya Pradha for her comments on earlier drafts of the manuscript.

References

[1] F. Baccelli, M. Klein, M. Lebourges, et al., "Stochastic geometry and architecture of communication networks," *Telecommun. Syst.*, vol. 7, no.1-3, pp. 209-227, Jun. 1997.

[2] F. Voss, C. Gloaguen, F. Fleischer, et al., "Distributional properties of Euclidean distances in wireless networks involving road systems," *IEEE J. Sel. Areas in Commun.*, vol. 27, no. 7, pp. 1047-1055, Sep. 2009.

[3] M. Khabazian and M. K. M. Ali, "A performance modeling of connectivity in vehicular ad hoc networks," *IEEE Trans. Veh. Technol.*, vol. 57, no. 4, pp. 2440-2450, Jul. 2008.

[4] V. K. Muhammed Ajeer, P. C. Neelakantan, and A. V. Babu, "Network connectivity of one-dimensional vehicular ad hoc network," in *2011 Int. Conf. Commun. Signal Processing*, Calicut, India, 2011, pp. 241-245.

[5] S. Yousefi, E. Altman, R. El-Azouzi, et al., "Analytical model for connectivity in Vehicular Ad Hoc Networks," *IEEE Trans. Veh. Technol.*, vol. 57, no. 6, pp. 3341-3356, Nov. 2008.

[6] S. Busanelli, G. Ferrari, and R. Gruppini, "Performance analysis of broadcast protocols in VANETs with Poisson vehicle distribution," in *11th Int. Conf. ITS Telecommun.*, St. Petersburg, Russia, 2011, pp. 133-138.

[7] A. Busson, "Analysis and simulation of a message dissemination algorithm for VANET," *Int. J. Commun. Syst.*, vol. 24, no. 9, pp. 1212-1229, Jan. 2011.

[8] Y. Jeong, J. W. Chong, H. Shin, et al., "Intervehicle communication: Cox-Fox modeling," *IEEE J. Sel. Areas in Commun.*, vol. 31, no. 9, pp. 418-433, Sep. 2013.

[9] C. Gloaguen, F. Fleischer, H.Schmidt, et al., "Fitting of stochastic telecommunication network models, via distance measures and Monte-Carlo tests," *Telecommun. Syst.*, vol. 31, is. 4, pp. 353-377, Apr. 2006.

[10] C. Gloaguen, F. Voss, and V. Schmidt, "Parametric distance distributions for fixed access network analysis and planning," in *Proc. 21st Int. Teletraffic Congress*, Paris, France, 2009, pp. 1-8.

[11] I. W. H. Ho, K. K. Leung, and J. W. Polak, "Stochastic model and connectivity dynamics for VANETs in signalized road systems," *IEEE/ACM Trans. Networking*, vol. 19, no. 1, pp. 195-208, Feb. 2011.

[12] A. Guo and M. Haenggi, "Spatial stochastic models and metrics for the structure of base stations in cellular networks," *IEEE Trans. Wireless Commun.*, vol. 12, no. 11, pp. 5800-5812, Nov. 2013.

[13] P. Golmohammadi, P. Mokhtarian, F. Safaei, et al., "An analytical model of network connectivity in vehicular ad hoc networks using spatial point processes," in *Proc. of IEEE Int. Symposium on a World of Wireless, Mobile and Multimedia Networks 2014*, Sydney, New Sales Wales, Australia, 2014, pp. 1-6.

[14] F. L. Mannering, S. S. Washburn, and W. P. Kilareski. *Principles of highway engineering and traffic analysis*, 4th Edition. Wiley, 2008.

[15] A. Y. and Abul-Magd, "Modeling highway-traffic headway distributions using superstatistics," *Phys. Rev. E*, vol. 76, is. 5, pp. 057101.1-057101.4, Nov. 2007.

[16] M. Haenggi, *Stochastic geometry for wireless networks*. Cambridge University Press, 2012.

[17] Diggle, Peter J., and R. J.Gratton, "Monte Carlo methods of inference for implicit statistical models," *J. Royal Statist. Soc.*, vol. series B 46, pp. 193-212, 1984.

[18] J. Møller, A. R. Syversveen, and R. P. Waagepetersen, "Log Gaussian Cox processes". *Scand. J. Statist.*, vol. 25, No.3, pp. 451-482, Sep. 1998.

[19] Yanyan Lu, Qimei Cui, Yanzhao Hou, Zhenguo Gao, Yuhao Zhang, "Graph-based optimal revenue packet scheduling in Vehicle-to-Infrastructure communication", in *ICC Workshops*, Paris, France, 2017, pp. 583-588.