

Fitting Determinantal Point Processes to Macro Base Station Deployments

Yingzhe Li, François Baccelli, Harpreet S. Dhillon, Jeffrey G. Andrews

Abstract—The macro base station (BS) deployments in modern cellular networks are neither regular nor completely random. We use determinantal point process (DPP) models to study the repulsiveness among macro base stations observed in cellular networks. Three DPP models are fitted to base station location data sets from two major US cities. Hypothesis testing is used to validate the goodness-of-fit for these DPP models. Based on performance metrics including the K-function, the L-function and coverage probability, DPP models are shown to be accurate in modeling real BS deployments. On the contrary, the Poisson point process and perturbed hexagonal grid model are shown to be less realistic. Different DPP models are compared, and several computational properties of these models are also discussed.

I. INTRODUCTION

Traditionally, cellular base stations have been modeled by grid-based models. However, real base station deployments are irregular and semi-random, due to network planning, opportunistic BS placement, and other topological and demographic factors. In addition to the grid-based model, random spatial point process models have become increasingly popular, both because they can describe highly irregular placements, and because stochastic geometry provides a tractable tool to analyze them [1]. In most prior works, starting with [2] [3], cellular base station locations are modeled as a homogeneous Poisson point process (PPP), and stochastic geometry is used to analytically derive key performance metrics such as the coverage probability and the mean Shannon rate.

Although tractable results can be derived under PPP, it is an idealized model since cellular base stations are assumed to be located in a totally independent way from each other. Real (macro) BS deployments exhibit “repulsion” between the BSs, because it generally does not make sense to install macro BSs very close to each other. Therefore, macro BSs are typically distributed more regularly than the PPP. Therefore, Gibbs processes which are able to model inter-point interactions were investigated in [4] [5] [6]. In terms of the SIR distribution and Voronoi cell area distribution, Gibbs models were validated to be statistically similar to real BS deployment in [4]. The Strauss process, which is an important class of Gibbs processes, is shown to provide an accurate statistical fit to some real BS deployments [5] [6]. By contrast, the PPP and the grid

Y. Li, F. Baccelli and J. G. Andrews are with the Wireless Networking and Communications Group (WNCG), The University of Texas at Austin (email: yli@utexas.edu, francois.baccelli@austin.utexas.edu, jandrews@ece.utexas.edu). H. S. Dhillon is with the Wireless@VT, the Department of Electrical and Computer Engineering, Virginia Tech (email: hdhillon@vt.edu).

models are demonstrated to be less accurate models for real BS deployments [4] [5].

A significant limitation of using Gibbs processes is their lack of tractability, since the probability generating functionals of Gibbs models are unknown in general [5]. Therefore, more tractable point processes are desirable, ideally preserving the excellent accuracy given by the Gibbs process. Determinantal point processes (DPPs) form a promising class of such models for cellular BS distributions for several reasons. First, DPPs have soft and adaptable repulsiveness [7]. Second, statistical inference for DPP models can be effectively evaluated [8] [9]. Third, many stationary DPP models can be easily simulated [8]. Fourth, DPPs have many nice mathematical properties, which lend them to the analysis of cellular network performance [10] [11].

Although DPPs have many potential advantages, the accuracy of using DPP models for macro BS distributions is still unexplored. In this work, we fit three DPP models (Gauss, Cauchy, Generalized Gamma) proposed in [8] to two real base station location data sets of major US cities. Based on metrics including the K-function and the L-function [12, p. 142], as well as coverage probability [3], the goodness-of-fit of these DPP models is evaluated through hypothesis testing procedures. The main observations of this paper are: (1) DPPs are accurate models for real macro BS deployments in terms of the above summary statistics, especially for sprawling and flat areas where repulsiveness among BSs are expected; (2) the fitted Generalized Gamma model is found to have the strongest repulsiveness, followed by the Gauss model, while the Cauchy DPP model is the least repulsive; and (3) the Gauss DPP model has the best balance between modeling accuracy and mathematical tractability, which makes it a promising model to use in future analysis.

II. DETERMINANTAL POINT PROCESS

In this section, we provide the definition, some important properties and examples of DPP models.

DPPs are defined based on their n -th order product densities. Consider a locally finite spatial point process Φ defined on a Borel set $B \subseteq \mathbb{R}^2$. Then Φ has n -th order product density function $\rho^{(n)} : B^n \rightarrow [0, \infty)$ if for any Borel function $h : B^n \rightarrow [0, \infty)$:

$$E \sum_{X_1, \dots, X_n \in \Phi}^{\neq} h(X_1, \dots, X_n) = \int_B \dots \int_B \rho^{(n)}(x_1, \dots, x_n) \times h(x_1, \dots, x_n) dx_1 \dots dx_n, \quad (1)$$

where \neq means X_1, \dots, X_n are pair-wise different.

Definition 1: The spatial point process Φ defined on the Borel set B is called a determinantal point process with kernel K , if its n -th order product density has the following form:

$$\rho^{(n)}(x_1, \dots, x_n) = \det(K(x_i, x_j))_{1 \leq i, j \leq n}, \quad (2)$$

where $\det(A)$ denotes the determinant of the square matrix A .

Throughout this paper, the kernel $K(x, y)$ is assumed to be a continuous, Hermitian, locally square integrable and non-negative definite function. The repulsive nature of DPP can be explained based on the fact that $\rho^{(n)}(x_1, \dots, x_n) \approx 0$ when $x_i \approx x_j$ for $i \neq j$.

A DPP defined on \mathbb{R}^2 is stationary if its n -th order product density is invariant under translations. A natural way to guarantee the stationarity of a DPP model Φ is that its kernel K be of the form: $K(x, y) = K_0(x - y)$ for $x, y \in \mathbb{R}^2$. In this case, $K_0(\cdot)$ is also referred to as the covariance function of Φ .

For stationary DPPs, the intensity measure (i.e., first order product density) is constant over \mathbb{R}^2 , and the pair correlation function of x, y only depends on the difference $y - x$.

Another important feature of stationary DPPs is their spectral density:

Definition 2: (Spectral Density [8]) The spectral density φ of a stationary DPP model Φ with covariance function $K_0(t)$ is defined as:

$$\varphi(x) = \int_{\mathbb{R}^2} K_0(t) e^{-2\pi i x \cdot t} dt, \quad x, t \in \mathbb{R}^2. \quad (3)$$

The spectral density is extremely useful for simulating stationary DPP models, and it also serves for checking the existence of a DPP with a given kernel. Specifically, from Proposition 5.1 in [8], the existence of a DPP model is equivalent to its spectral density φ belonging to $[0, 1]$.

Below, we study several examples of stationary DPP models proposed in [8]:

- 1) (Gauss DPP Model): A stationary point process Φ is a Gauss DPP if it has covariance function:

$$K_0(x) = \lambda \exp(-\|x/\alpha\|^2), \quad x \in \mathbb{R}^2. \quad (4)$$

Equivalently, the spectral density of a Gauss DPP is defined as:

$$\varphi(x) = \lambda(\sqrt{\pi}\alpha)^2 \exp(-\|\pi\alpha x\|^2), \quad x \in \mathbb{R}^2.$$

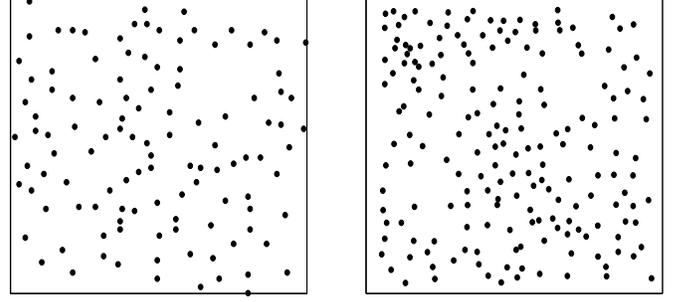
In the above definition, λ denotes the spatial intensity of the Gauss DPP, while α is a measure of its repulsiveness. In order to guarantee the existence of Gauss DPP model, the parameter pair (λ, α) needs to satisfy:

$$\lambda \leq (\sqrt{\pi}\alpha)^{-2}.$$

- 2) (Cauchy DPP Model): The Cauchy DPP model has a covariance function $K_0(\cdot)$ of the form:

$$K_0(x) = \frac{\lambda}{(1 + \|x/\alpha\|^2)^{\nu+1}}, \quad x \in \mathbb{R}^2. \quad (5)$$

In the Cauchy DPP model, λ describes the intensity, while



(a) Houston BS deployment

(b) LA BS deployment

Fig. 1: Actual macro BS deployments of Houston and LA.

α is the scale parameter and ν is the shape parameter. Both α and ν affect the repulsiveness of the DPP. To guarantee the existence of a Cauchy DPP, the parameters need to satisfy:

$$\lambda \leq \frac{\nu}{(\sqrt{\pi}\alpha)^2}.$$

- 3) (Generalized Gamma DPP Model): The Generalized Gamma model is defined based on its spectral density, which has the following form:

$$\varphi(x) = \lambda \frac{\nu \alpha^2}{2\pi \Gamma(2/\nu)} \exp(-\|\alpha x\|^\nu), \quad (6)$$

where $\Gamma(\cdot)$ denotes the Euler gamma function.

The existence of a Generalized Gamma DPP can be guaranteed when its parameters satisfy:

$$\lambda \leq \frac{2\pi \Gamma(2/\nu)}{\nu \alpha^2}.$$

III. MODELING BASE STATION DEPLOYMENT AS STATIONARY DPP MODELS

In this section, we fit these stationary DPP models to real base station deployments. Our simulations are based on the publicly available package for DPP models [8] implemented in R, which is used as a supplement to the Spatstat library [13].

BS deployments in two major US cities are investigated¹. Fig. 1a shows the deployment of 115 BSs in a 16 km \times 16 km area of Houston, while Fig. 1b shows the deployment of 184 BSs in a 28 km \times 28 km area of Los Angeles (LA). Both deployments are for sprawling and relatively flat areas, where repulsion among BSs is expected.

A. Summary Statistics

Given these stationary DPP models, an interesting question is how accurate they can model real macro base station deployments. To test the goodness-of-fit of these DPP models, the following metrics are taken into consideration.

¹BS location data was provided by a major tower owner in the USA.

Ripley's K function: Ripley's K function is a second order spatial summary statistic defined for stationary point processes. It counts the mean number of points within distance r of a given point in the point process excluding the point itself. Formally, the K function $\mathbb{K}(r)$ for a stationary and isotropic point process Φ with intensity λ is defined as:

$$\mathbb{K}(r) = \frac{\mathbb{E}^{l_0}(\Phi(B(o,r)))}{\lambda} \quad (7)$$

where $\mathbb{E}^{l_0}(\cdot)$ is the expectation with respect to the reduced Palm distribution of Φ [12, p. 148], and $B(o,r)$ is the ball centered at origin with radius r .

The K-function is used as a measure of repulsiveness/clustering of spatial point processes. Specifically, compared to the PPP which is completely random, a repulsive point process model will have a smaller K function, while a clustered point process model will have a larger K function.

L function: The L function is also a closely related summary statistics which is defined as $L(r) = \sqrt{K(r)}/\pi$. Compared with the L-function of the PPP which is equal to r , $L(r) > r$ corresponds to clustering point processes, while $L(r) < r$ corresponds to repulsive point processes.

Coverage Probability: We focus on the downlink coverage probability of a typical user. The following assumptions are used to evaluate the coverage probability: (1) all BSs are transmitting at fixed power P ; (2) i.i.d. Rayleigh fading channels with mean 1 are assumed; (3) a power-law path loss model with path loss exponent β is used, i.e., $l(x) = \|x\|^{-\beta}$; (4) shadowing effects are neglected; (5) noise power is negligible compared to interference, and specifically we set noise power equal to 0; (6) BS locations form a realization of a stationary and isotropic DPP Φ ; (7) mobile users are uniformly distributed, independent of the BSs; and (8) each mobile user communicates with its nearest BS.

Based on these assumptions, the coverage probability for a typical user located at z with threshold T is defined as:

$$P_c(z, T) = \mathbb{P}\left(\frac{Ph_{x_0}\|x_0 - z\|^{-\beta}}{\sum_{x_i \in \Phi \setminus x_0} Ph_{x_i}\|x_i - z\|^{-\beta}} > T\right), \quad (8)$$

where x_0 denotes the serving BS of z , and $h_{x_i} \sim \exp(1)$ denotes the Rayleigh fading channel from BS x_i to the typical user. A path loss exponent $\beta = 4$ is used in all our simulations.

B. Hypothesis Testing using Summary Statistics

In this part, we will evaluate the goodness-of-fit of stationary DPP models using the summary statistics discussed above. Particularly, we fit the real BS deployments in Fig. 1 to the Gauss, Cauchy and Generalized Gamma DPP models. An important step for fitting real data points to some stationary DPP model is to estimate the parameters that define its kernel or spectral density. This parameter estimation is based on a maximum likelihood (ML) estimate of the density function, which is implemented in the software package provided in [8]. Based on the ML estimate method, we have summarized the estimated parameters for different DPPs in Table I and Table II.

TABLE I: DPP Parameters for the Houston Data Set

Model	λ	α	ν
Gauss DPP	0.4492	0.8417	—
Cauchy DPP	0.4492	1.558	3.424
Generalized Gamma DPP	0.4492	2.539	2.63

TABLE II: DPP Parameters for the LA Data Set

Model	λ	α	ν
Gauss DPP	0.2347	1.165	—
Cauchy DPP	0.2347	2.13	3.344
Generalized Gamma DPP	0.2347	3.446	2.505

To evaluate the goodness-of-fit for these DPP models, we generate 1000 realizations of each DPP model and examine whether the simulated DPPs fit with the behavior of real BS deployments in terms of the summary statistics. Specifically, we verify whether the K-function or the L-function of the data set lies within the envelope of the simulated DPP models. We use similar testing method for coverage probability, a 95% confidence interval is used for evaluation.

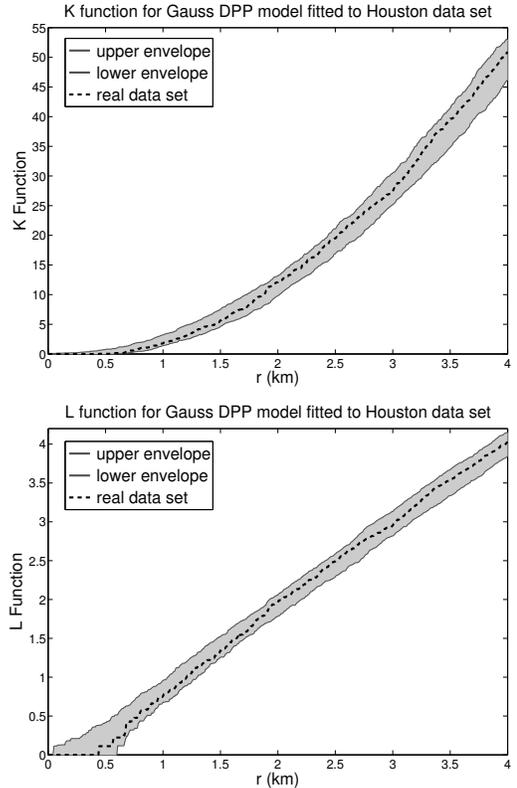


Fig. 2: The K function and the L function for the Gauss DPP model fitted to the Houston data set.

Goodness-of-fit for the Gauss DPP Model: The testing results for the K function and the L function of the fitted Gauss DPP model are given in Fig. 2 for the Houston data set, and in Fig. 3 for the LA data set. Both figures clearly show that the

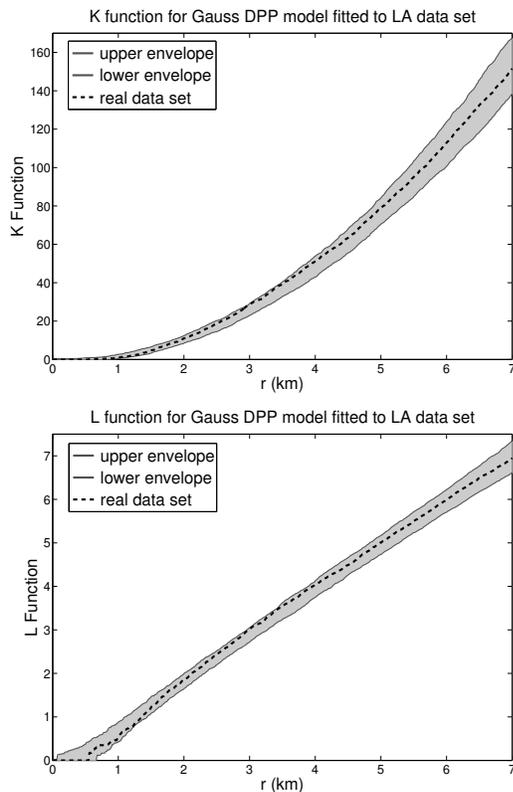


Fig. 3: The K function and the L function for the Gauss DPP model fitted to the LA data set.

K and the L functions of the real BS deployments lie within the envelope of the fitted Gauss DPP model.

The coverage probability for the fitted Gauss DPP model is provided in Fig. 4 for both data sets, from which it can be observed that the coverage probability of the Houston and LA data set lies within the 95% confidence interval of the simulated Gauss DPP models. In addition, the average coverage probability of the fitted Gauss model is slightly lower than that of real data sets, which means the fitted Gauss model corresponds to a slightly smaller repulsiveness than the real deployments.

Therefore, in terms of the above summary statistics, the Gauss DPP model can be used as a reasonable point process model for real BS deployments. Due to the concise definition of its kernel, a shot noise analysis of the Gauss DPP is possible.

Goodness-of-fit for the Cauchy DPP Model: Based on the same method as for the Gauss DPP model, we tested the goodness-of-fit for the Cauchy DPP model. The fitting results for the Houston data set are shown in Fig. 5 and Fig. 6, from which it can be concluded that the Cauchy DPP model is also a reasonable point process model for real BS deployments. Similar fitting results are also observed for the LA data set, and thus we omit the details. Compared to the Gauss DPP model, the average coverage probability for the Cauchy DPP model in Fig. 6 is slightly lower than that in Fig. 4, which means the Cauchy DPP model corresponds to a smaller repulsiveness than

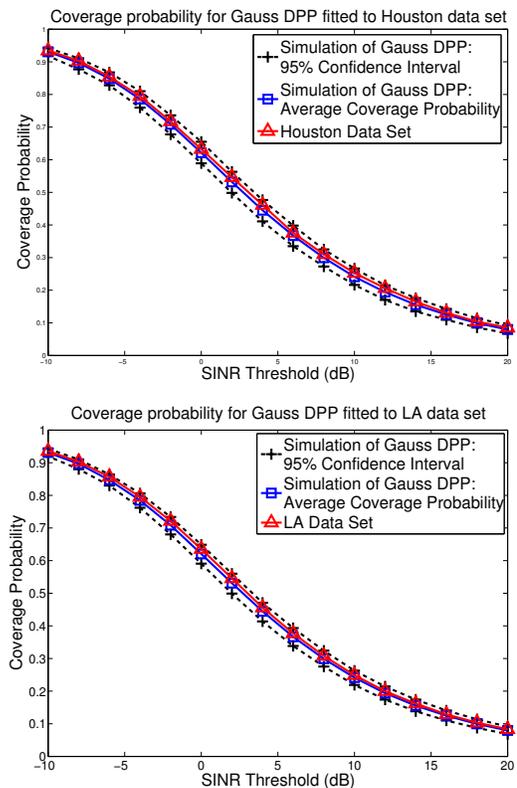


Fig. 4: Coverage probability of the fitted Gauss DPP model.

the Gauss DPP model. In addition, compared to Fig. 2, Fig. 5 shows that the Cauchy DPP model has a wider envelope for the L function and therefore a lower precision in modeling the real BS deployments.

Goodness-of-fit for the Generalized Gamma DPP Model: The goodness-of-fit for the Generalized Gamma DPP model is evaluated for the Houston data set in Fig. 7 and Fig. 8 (the LA data set has similar fitting results). The Generalized Gamma model provides the best fit among all these DPP models, especially in terms of coverage probability. In Fig. 8, the average coverage probability of the Generalized Gamma model almost exactly matches the real BS deployment, while the average coverage probability of the Gauss DPP and the Cauchy DPP all stay below the real data set. This is because the Generalized Gamma model corresponds to a higher repulsiveness (which will be proved in next subsection), from which a larger coverage probability is expected. In addition, Fig. 7 also shows that the Generalized Gamma model has high precision in modeling real BS deployments, since it corresponds to a narrow envelope in terms of the K function and the L function.

Goodness-of-fit for the PPP and the perturbed hexagonal model: Finally, the goodness-of-fit for the PPP and the perturbed hexagonal grid model are studied. The perturbed hexagonal grid model is obtained by independently perturbing each point of a hexagonal grid in the random direction by a distance d [4]. This distance is uniformly distributed between 0 and ηr , with r being the radius of the hexagonal cells

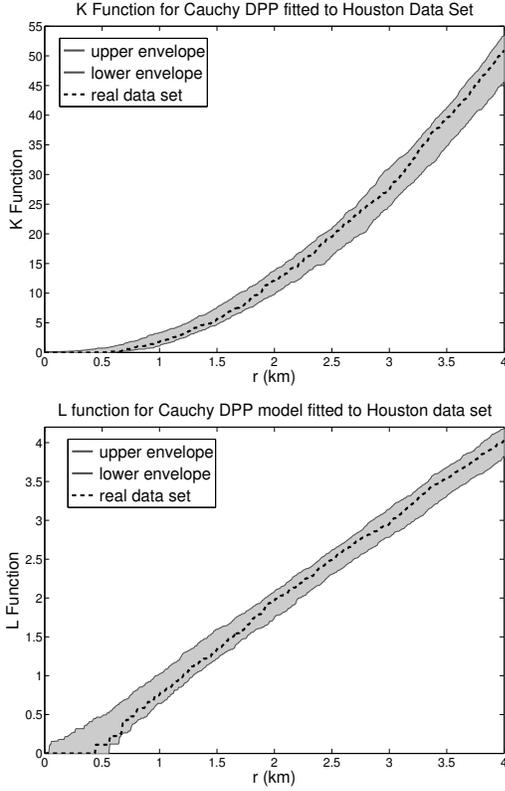


Fig. 5: The K function and the L function for the Cauchy DPP model fitted to the Houston data set.

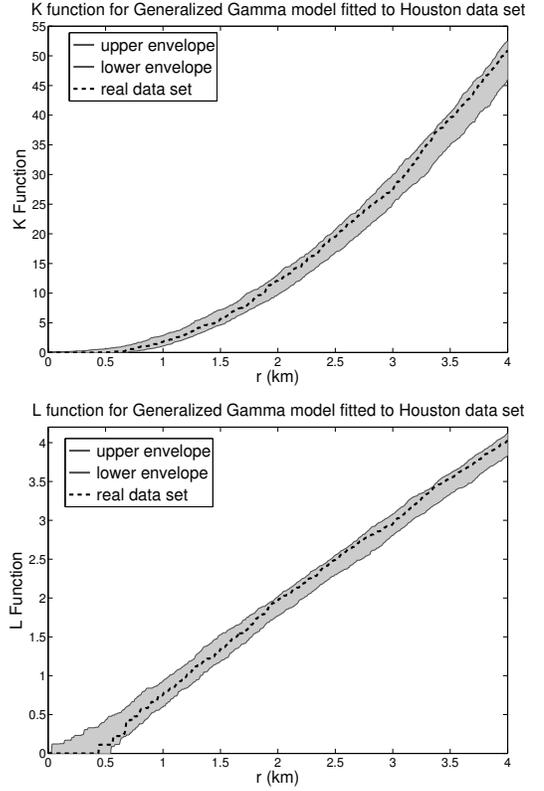


Fig. 7: The K function and the L function for the Generalized Gamma model fitted to the Houston data set.

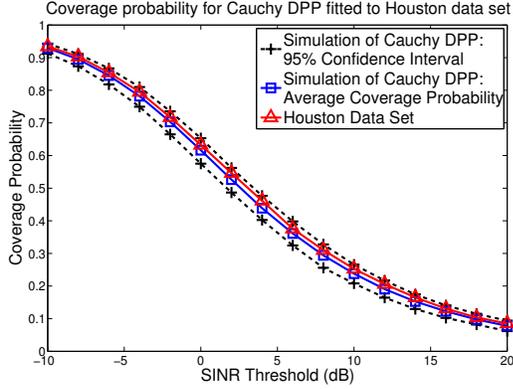


Fig. 6: Coverage probability of the fitted Cauchy DPP model.

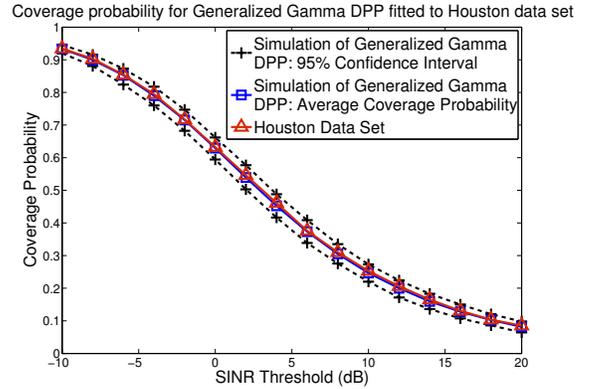


Fig. 8: Coverage probability of the fitted Generalized Gamma model.

and η is chosen as 0.5 in our simulation. Fig. 9 shows the coverage probability of the PPP and the perturbed hexagonal grid model fitted to the Houston data set, which correspond to a lower bound and an upper bound of the actual coverage probability respectively. This is because the PPP has complete spatial randomness while the perturbed grid model maintains good spatial regularity. Similar observations can also be derived for the LA data set. Therefore, compared to DPP models, the PPP model and the perturbed hexagonal grid model should be rejected as reasonable random point process models to study real macro BS deployments.

C. Repulsiveness Comparison of fitted DPP Models

In this part, we use the metric suggested in [8] to measure repulsiveness of fitted DPP models, and explain why the Generalized Gamma model has the highest repulsiveness. Specifically, the intensity measure of a stationary DPP model Φ under its reduced Palm distribution is $\rho_o^{(1)}(x) = \rho^{(2)}(0, x) / \rho^{(1)}(x)$, where $\rho^{(2)}$ and $\rho^{(1)}$ are the second and the first order product density of Φ . By calculating the difference of the total expected number of points under the probability distribution \mathbb{P} and the

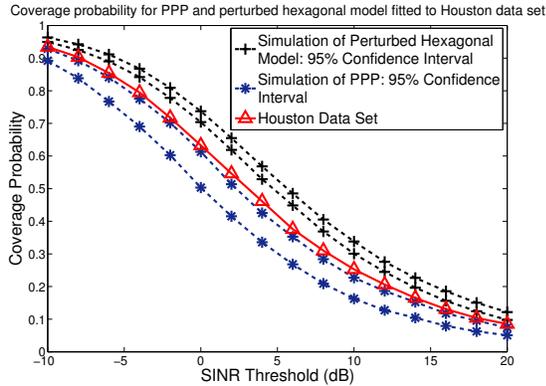


Fig. 9: Coverage probability of the PPP and the perturbed grid model.

reduced Palm distribution \mathbb{P}^{l_0} , the repulsiveness of a stationary DPP model Φ with constant intensity λ can be measured using the following metric [8]:

$$\mu = \int_{\mathbb{R}^2} [\lambda - \rho_o^{(1)}(x)] dx = \frac{1}{\lambda} \int_{\mathbb{R}^2} |K_0(x)|^2 dx = \frac{1}{\lambda} \int_{\mathbb{R}^2} |\varphi(x)|^2 dx. \quad (9)$$

In particular, larger value of μ will correspond to higher repulsiveness.

Based on metric μ , the repulsiveness of the Gauss, Cauchy and Generalized Gamma DPP models can be measured as:

$$\begin{aligned} \mu_{\text{gauss}} &= \frac{\lambda\pi\alpha^2}{2}, \\ \mu_{\text{cauchy}} &= \frac{\lambda\pi\alpha^2}{2\nu + 1}, \\ \mu_{\text{gengamma}} &= \frac{\lambda\nu\alpha^2}{2^{1+2/\nu}\pi\Gamma(2/\nu)}. \end{aligned} \quad (10)$$

By substituting the parameters listed in Table I and Table II to Eq. (10), we can explicitly calculate the repulsiveness measure of each DPP model fitted to the Houston data set as $\mu_{\text{gauss}} = 0.4999$, $\mu_{\text{cauchy}} = 0.4365$ and $\mu_{\text{gengamma}} = 0.5905$. The same procedure is adopted for the LA data set, where the repulsiveness measures are given by $\mu_{\text{gauss}} = 0.5004$, $\mu_{\text{cauchy}} = 0.4351$, $\mu_{\text{gengamma}} = 0.5479$. Therefore, it can be concluded that the fitted Generalized Gamma model has the largest repulsiveness, followed by the Gauss model, while the Cauchy model is the least repulsive. Since higher repulsiveness will result in more regularity for the distribution of the point process, Generalized Gamma models generally correspond to larger average coverage probability.

IV. CONCLUSION

In this paper, the accuracy of using determinantal point processes to model cellular base station deployment is investigated. Three DPP models: the Gauss model, the Cauchy model and the Generalized Gamma model are fitted to two real macro base station deployments. By using the K-function, the L-function, and coverage probability as performance metrics,

hypothesis testing procedures are used to evaluate the goodness-of-fit of these DPP models. Due to its higher repulsiveness, the Generalized Gamma model provides the best fit to real BS deployments in terms of coverage probability. However, the Generalized Gamma model is generally less tractable since it is defined based on spectral density. In contrast, the Gauss DPP model also provides reasonable fit to real BS deployments, but with better mathematical tractability due to the simple definition of its kernel. Compared to other DPP models, the fitted Cauchy model has the smallest repulsiveness and also less precise results in terms of the summary statistics such as coverage probability.

Future work will analyze cellular networks using DPP distributed BSs, which is possible since DPPs have several computational properties which facilitate the analysis. First, DPPs have closed-form product density of any order as shown in Eq. (1). Second, DPPs have closed-form Laplace functional for any function $f : \mathbb{R}^2 \rightarrow \mathbb{R}^+$ [10, Theorem 1.2]. Third, under the reduced Palm distribution, the original DPP has the same law as another DPP model whose kernel is given in closed-form [10, Theorem 1.7]. Based on these properties, many important metrics related to cellular networks can be derived, such as the empty space function, the mean/variance of the interference, and the coverage probability, etc.

REFERENCES

- [1] M. Haenggi, J. Andrews, F. Baccelli, O. Dousse, and M. Franceschetti, "Stochastic geometry and random graphs for the analysis and design of wireless networks," *IEEE Journal on Selected Areas in Communications*, vol. 27, pp. 1029–1046, Sep. 2009.
- [2] F. Baccelli, M. Klein, M. Lebourges, and S. Zuyev, "Stochastic geometry and architecture of communication networks," *Telecommunication Systems*, vol. 7, no. 1-3, pp. 209–227, 1997.
- [3] J. Andrews, F. Baccelli, and R. Ganti, "A tractable approach to coverage and rate in cellular networks," *IEEE Transactions on Communications*, vol. 59, pp. 3122–3134, Nov. 2011.
- [4] D. Taylor, H. Dhillon, T. Novlan, and J. Andrews, "Pairwise interaction processes for modeling cellular network topology," in *IEEE Global Communications Conference*, pp. 4524–4529, Dec. 2012.
- [5] A. Guo and M. Haenggi, "Spatial stochastic models and metrics for the structure of base stations in cellular networks," *IEEE Transactions on Wireless Communications*, vol. 12, pp. 5800–5812, Nov. 2013.
- [6] J. Riihijarvi and P. Mahonen, "Modeling spatial structure of wireless communication networks," in *INFOCOM IEEE Conference on Computer Communications Workshops*, pp. 1–6, Mar. 2010.
- [7] R. Ganti, F. Baccelli, and J. Andrews, "Series expansion for interference in wireless networks," *IEEE Transactions on Information Theory*, vol. 58, pp. 2194–2205, Apr. 2012.
- [8] F. Lavancier, J. Møller, and E. Rubak, "Statistical aspects of determinantal point processes," *arXiv preprint arXiv:1205.4818*, 2012.
- [9] J. B. Hough, M. Krishnapur, Y. Peres, and B. Virág, *Zeros of Gaussian analytic functions and determinantal point processes*, vol. 51 of *University Lecture Series*. American Mathematical Society, 2009.
- [10] T. Shirai and Y. Takahashi, "Random point fields associated with certain Fredholm determinants I: fermion, Poisson and boson point processes," *Journal of Functional Analysis*, vol. 205, pp. 414–463, Dec. 2003.
- [11] N. Miyoshi and T. Shirai, "A cellular network model with Ginibre configured base stations," *Research Rep. on Math. and Comp. Sciences (Tokyo Inst. of Tech.)*, Oct. 2012. Available at: <http://www.is.titech.ac.jp/research/research-report/B/B-467.pdf>.
- [12] S. N. Chiu, D. Stoyan, W. S. Kendall, and J. Mecke, *Stochastic geometry and its applications*. John Wiley & Sons, 2013.
- [13] A. Baddeley and R. Turner, "Spatstat: an R package for analyzing spatial point patterns," *Journal of Statistical Software*, vol. 12, pp. 1–42, Jan. 2005.