Characterization of base station deployment distribution and coverage in heterogeneous networks

Hui Zhang¹ | Liang Feng² | Tianyu Zheng³ | Dan Keun Sung⁴

¹ Tianjin Key Laboratory of Optoelectronic Sensor and Sensing Network Technology, Nankai University, Tianjin, P. R. China
² Luoyang Electronic Equipment Center, Luoyang, P. R. China
³ Department of Communication Engineering, Nankai University, Tianjin, P. R. China
⁴ School of Electrical Engineering, Korea Advanced Institute of Science and Technology, Daejeon, Republic of Korea

Abstract
Considering different types of base stations (BSs) in future cellular networks are overlapping deployment with the status of dense, multi-tier and heterogeneous in general, how to optimize the real BS deployment becomes a complicated problem. Based on it, repulsive BS dataset and clustering BS dataset are statically characterized with various types of spatial point processes. It shows that the improvement of coverage probability between 1-tier and 2-tier network fluctuates as the signal-to-interference ratio (SIR) threshold increases for different datasets. The authors’ proposed hybrid model fits well with repulsive BS dataset, while the log-Gaussian Cox point process (LGCP) and Cauchy model are reasonable models for clustering BS dataset. Besides, in order to dynamically analyze the coverage problem affected by adding new BSs, a cell boundary constructed by an irregular circle is introduced under an equal SIR constraint, and a BS placement scheme is proposed to place new BSs at the points of minimum interference. Numerical results show that the coverage probability may increase after adding BSs in the target area of heterogeneous network by using the proposed scheme. However, as the density of femto BS reaches a certain value, its coverage may remain unchanged even after adding more femto BSs.

1 | INTRODUCTION

With the commercialization of the fifth generation (5G) system, current base station (BS) deployments in cellular networks are generally overlapping and coexisted with the previous systems, showing the status of dense, multi-tier and heterogeneous in general. Considering the real BS deployments are usually from different operators, it becomes a complicated problem for optimizing the BS deployment, which is related closely with intensity of interference and quality of service (QoS). What is more serious, due to a rapid growth with mobile Internet services, customers dramatically demand faster wireless data traffic, such as high-rate downloading at hot spots [1]. Driven by this demand, it needs to deploy more BSs to enhance signal coverage and improve the QoS [2], even by adding unmanned aerial vehicle BSs [3, 4]. Under this background, it is critical to find a way to optimize BS coverage problems by means of the statistical modelling and analysis of BS deployment.

In the past years, several approaches in geometry theory were used to characterize these coverage probabilities in BS deployment [5]. Many previous works showed that the spatial point processes were often taken to model the BS deployment [6, 7], such as Poisson point process (PPP) [8–10] etc. However, it usually existed the gaps between theory and real results for homogeneous BS distribution and its corresponding system-level coverage had been generally analyzed statically according to the real database. Due to the gaps of theoretical model and real BS deployment, the deployment distance of a real BS was usually repulsive with each other for homogeneous networks and 1-tier heterogeneous networks, and hence, it exhibited more regular and more uniform than the realization of a PPP model, which was shown that the PPP model was not be an accurate model for real BS deployment in homogeneous networks.

Considering this problem, a series of non-PPP models were proposed to model different scenarios in the real BS deployment. For example, Strauss process, Matern cluster process (MCP) [11], Poisson cluster process [12], Poisson hardcore hole process [13], Poisson hole process [14] etc. Besides, the pairwise interactions among BSs were also considered in [15], which theoretically illustrated that the Strauss hardcore process was...
suitable for inhibitive BS deployment patterns, and the Geyer saturation process was suitable for clustered BS deployment patterns. Specially, a Strauss process was tested to provide a better fit than the PPP [16]. As an effective tool, determinantal point processes (DPP), such as Gauss DPP, Cauchy DPP and generalized Gamma DPP had been characterized in homogeneous networks [17], in which the system-level coverage results for each model were compared, respectively. In particular, a Gini-bre point process as a special DPP was well known to model repulsion among points [18, 19]. By means of a goodness-of-fit test, it showed that the DPP models may be more suitable for the datasets in homogeneous networks because real macro BS deployments exhibited repulsion among the BSs.

Although the non-PPP models could bring a better fit for real BS deployment in some cases, it was still hard to quantitatively derive the interference statistical characteristics in theory. Aiming to this problem, Haenggi et al. [20–23] proposed a simple approximate framework, written as approximate signal to interference ratio (SIR) analysis based on PPP (ASAPP). In this framework, it approximated the SIR distribution in a non-PPP model by scaling that in the corresponding PPP model, where the scaling coefficient was called SIR gain [24]. By means of the ASAPP method, we could get an approximate SIR distribution based on non-PPP models for networks, but the drawback of this method was that it needed to use lots of Monte Carlo simulations to obtain the SIR gain.

In addition to theoretical modelling, some other works studied BS distribution characteristics from the statistical analysis of massive BS datasets. For example, the real BS deployment datasets in European homogeneous networks were analyzed statistically in [25], in which BSs were classified into the following two types: global system for mobile communications (GSM) and universal mobile telecommunication system (UMTS), and its system-level coverage results related to the European cities showed that a log-Gaussian Cox process provided a higher accuracy in system-level coverage, compared with the PPP, Strauss process and MCP.

However, for K-tier heterogeneous networks (K > 1), since BSs belonging to different tiers, such as macro BS, pico BS and femto BS, were always deployed in a hybrid way, a practical BS deployment always exhibited clustering among the BSs [26], which indicated that it was hard to apply the repulsive property of real homogeneous BSs into multiple tier heterogeneous networks. Considering this problem, Gauss-Poisson process [27] had been recently used to model spatial distribution with attraction for clustering nodes. Besides, the Poisson cluster process [28, 29], Poisson hardcore process [30] and Matern hardcore process [31] were, respectively, used to model the real BS deployment in two-tier heterogeneous cellular networks, and the coverage problem was analyzed by using the ASAPP method. It showed that the ASAPP method may provide simple approximations of the SIR distribution in general cellular networks [32]. But until now, most of the existing works were usually major in taking a single model to fit the BS deployment dataset and analyze its SIR distribution. It still lacked the effective algorithms that was from the statistical method to establish the precise models to fit both repulsive and clustering BS deployment dataset.

What is important, it is not sufficient for us to analyze the system-level coverage probability from the static viewpoint, which lacks adaptability for deploying new BSs. For network operators, changes in local BS deployment may directly affect its coverage results. Thus, characterizing the effect of a new BS deployment under a given-cell coverage is more important to improve the BS deployment than system-level coverage results. When we consider two types of cells include macro and femto cells in a multi-tier heterogeneous network (Hetnet), deploying new BSs may increase the interference on their neighbouring BSs in the system. Therefore, it is still hard to reflect the coverage effect of adding new BSs for existing statical analysis of real BS distribution.

The contributions of this paper are summarized as follows:

**Static characteristic analysis with a repulsive BS dataset**

The repulsive BS deployment dataset in multi-tier Hetnet is collected in Smithfield. Compared with a 1-tier network scenario, the coverage probability is improved under a 2-tier network scenario because of small cell deployment. It shows that the coverage probability improves more at a high SIR region for this repulsive BS dataset, which illustrates that the efficiency of small cells is much better for repulsive BS deployment at high SIR regions. In order to find an accurate point distribution model, the repulsive BS deployment dataset is statistically modelled according to the spatial point process theory, in which the good-of-fit is verified by means of the Ripley’s K function, the L function, the empty space function and the nearest neighbour distance function. Considering it is hard to find a single saturation process to accurately fit the real dataset, it is feasible to combine multiple saturation processes in a hybrid way to approach the dataset. The results show that our proposed hybrid model is well fitted with the collected BS dataset in Smithfield, on the basis of the proposed searching algorithm that could effectively find the parameters to yield smaller gaps for hybrid model. Our work gives a novel way for the modelling of BS dataset fitting in the future.

**Static characteristic analysis with a clustering BS dataset**

The clustering BS deployment dataset in multi-tier Hetnets is collected in Glasgow. It shows that the coverage probability also improves under a 2-tier network scenario because of small cell deployment. The difference is that the coverage probability improves more at low SIR regions for this clustering BS dataset. It illustrates that the efficiency of small cells is much better for clustering BS deployment at low SIR regions, and this practice helps to guide small cell deployment in future. Furthermore, the clustering BS deployment dataset is statistically modelled to find the accurate point distribution model. By means of testing a good-of-fit law for real dataset and clustering point processes, a searching algorithm is proposed to find the proper clustering point processes for its dataset, such as MCP, log-Gaussian Cox point process (LGCP), Vargamma model and Cauchy model. The error results illustrate that the LGCP and Cauchy model
become the reasonable models with smaller gaps between real dataset and theoretical value.

**Dynamic coverage probability analysis**
Based on the statical analysis of BS deployment, we further consider the dynamic coverage analysis by adding a new BS, and analyze the change of coverage after adding a different type of a new BS. We propose a BS placement searching scheme for deploying a new BS, in which a new BS can be deployed at the minimum interference location in a region of interest, so that the interference caused by the new BS can be minimized. Numerical results show that the coverage probability may increase after adding BSs in the target area by using the proposed scheme, but as the density of the femto BS reaches a certain value, its coverage may remain unchanged even after adding more femto BSs. Our work could provide a reference for the site selection of new BSs in the future.

The remainder of this paper is organized as follows: BS deployment is statically modelled in Section 2 using spatial point processes. The static spatial distribution of BS deployment is characterized in Section 3. The dynamic effect of a new BS deployment on coverage probability is analyzed in Section 4. The dynamic coverage probability results by adding BSs are compared in Section 5. Finally, conclusions are drawn in Section 6.

## 2 | STATIC MODELLING OF BS DEPLOYMENT USING SPATIAL POINT PROCESSES

The conventional analysis of BS deployment is always taking an approach in which we choose a specific BS dataset from the whole city or one country, and then give system-level results that reflect user distribution. However, it is still hard to provide the guidance to actual BS deployment. In order to characterize the BS deployment problem, we choose some typical real BS distribution datasets from Ofcom database1, such as repulsive BS dataset at a urban area in Smithfield and clustering BS dataset at a rural area in Glasgow, as shown in Figures 1 and 2, respectively, which are scaled according to the corresponding map. First of all, we take spatial point processes to statically model BS deployment [33], and consider the following theoretical concepts in stochastic geometry [6].

### 2.1 | Point process model

#### 2.1.1 | Repulsive point process

Generally, the Gibbs point process is one of classic repulsive spatial point process, in which different points typically inhibit each other. Based on Gibbs point process [34], some other repulsive spatial point processes are proposed for actual point distribution modelling, such as Fiksel model, Geyer model,

![FIGURE 1](image1.png)
**FIGURE 1** Real BS deployment in Smithfield dataset (51 BSs)

![FIGURE 2](image2.png)
**FIGURE 2** Real BS deployment in Glasgow dataset (52 BSs)

Hardcore model, Strauss model and Hybrid model [33, 35]. The parameters in each model are listed in Table 1. Especially, the Geyer model and Strauss model can be classified into saturation point process model, also written as pairwise interaction model, in which the radius of a point from its pairwise interaction with other points is trimmed to a certain value. As its interaction radius is fixed, the saturation model belongs to the Geyer model when its saturation threshold is limited, but this model may reduce to the Strauss model when its saturation threshold is approaching to infinity. As a special case, the hybrid model is combined with the Geyer process and the Strauss process, and

| Model     | Parameter  | Comment            |
|-----------|------------|--------------------|
| Geyer     | $r$, $s$   | Interaction radius | Saturation threshold |
| Strauss   | $r$        | Interaction radius | –                  |
| Hardcore  | $r$        | Hard core distance | –                  |
| Fiksel    | $r$, $\lambda$ | Interaction radius | Rate parameter     |

| Table 1 | Model parameters |
|---------|------------------|
| Model   | $r$    | $s$  |
| Geyer   | Interaction radius | Saturation threshold |
| Strauss | Interaction radius | –                  |
| Hardcore| Hard core distance | –                  |
| Fiksel  | Interaction radius | Rate parameter     |

1 [http://sitefinder.ofcom.org.uk/search](http://sitefinder.ofcom.org.uk/search)
the probability density function of hybrid model is defined as follows

\[ n(\nu) = f(\nu)g(\nu), \]  

(1)

where \( f(\nu) \) and \( g(\nu) \) are the probability density function of Geyer process and the Strauss process, respectively.

2.1.2 Clustering point process

The typical clustering spatial point process models generally include a Thomas point process (Thomas), a Neyman–Scott process with variance Gamma cluster kernel (VarGamma), MCP, LGCP and Neyman–Scott process with Cauchy cluster kernel (Cauchy) [35, 37]. For these clustering point processes, we can take the method of minimum contrast to fit a theoretical model to the given dataset [36]. Take the Ripley’s K function as an example, the K function can be computed from the dataset, and its theoretical value of K function under certain point process model is estimated from simulations of this model. Then this model is fitted by finding the optimal parameter values, aiming to give the closest match between the theoretical and empirical curves. Specially, the function of \texttt{thomastestK}, \texttt{lgp.estK} and \texttt{matchlist.estK} in the spatstat software can be used to fit Thomas, LGCP and MCP, respectively.

2.2 Distribution analysis function

2.2.1 G function

The G function is a nearest neighbour distance function, where the estimation of G function is a statistic summarizing one aspect of the clustering point process, \( G(r) = P(\text{dist}(x, X) \leq r) \). As a special case of Poisson point processes, \( G(r) = 1 - e^{-\lambda \pi r^2} \), where \( r \) is the distance among events, and \( \lambda \) is the density of events. It can reflect the spatial clustering or spatial regularity of dataset according to the deviations between the empirical and theoretical G curves.

2.2.2 Ripley’s K function

The Ripley’s K function is used as a measure of clustering and repulsiveness of spatial point processes, and \( K(r) = \Omega \lambda / \pi \), where \( \Omega \) denotes the expectation of the number of extra events within distance \( r \) of a randomly chosen event, and \( \lambda \) is the density (number per unit area) of events [37]. For a stationary Poisson process, the theoretical value \( K(r) \) is expressed as \( K(r) = \pi r^2 \).

2.2.3 L function

The L function is a transformation of the Ripley’s K-function, where \( L(r) = \sqrt{K(r)} / \pi \). The theoretical value of the L function is \( L(r) = r \) for a uniform Poisson point pattern. \( L(r) > r \) corresponds to a clustering point process, while \( L(r) < r \) corresponds to a repulsive point process.

2.2.4 Envelope function

By means of the envelope function, we can compute envelopes of a summary statistic based on the simulations, and the results include lower envelope and upper envelope, which can be used to assess the goodness-of-fit of a point process model to point pattern data.

3 NUMERICAL RESULTS OF STATIC SPATIAL DISTRIBUTION

On the basis of current BS distribution in both Smithfield dataset and Glasgow dataset, we numerically obtain the coverage probability. Furthermore, we make a model fitting analysis for both of the two datasets using spatstat software [37], in which the G function, K function, L function and Envelope function can be obtained conveniently. For one assumed theoretical distribution model (among PPP, MCP, LGCP, Cauchy etc.), it may provide a confidence envelope, where the lower/upper envelope denotes the lower/upper bound, respectively. For real data, if its function is within the lower and upper bound, then the theoretical distribution model can be considered to be effective and its error depends on the gap with the theoretical function. Thus, we can verify the validity of the theoretical distribution model and the real dataset using spatstat software and R language [37].

3.1 Coverage area of BS distribution with equivalent SIR

Combined with various types of BSs in the real BSs dataset, we consider a new BS located at a point \( z \) of the \( k \)-th tier of a Hetnet, which provides users with services within its coverage region. We assume that the signal-to-interference ratio (SIR) threshold for guaranteeing the required quality of service (QoS) is \( \tau_k \) for the \( k \)-th tier BSs. The transmitted signals in this Hetnet experience large-scale fading of propagation path-loss and shadowing, as well as small-scale fading modelled as a Rayleigh distribution with the power of \( \text{E}\{|h|^2\} = 1 \). Then, for a user equipment (UE) located at a point \( z \) and served by the BS at \( z_i \), the average received power is given by

\[ s_i(z) = P_i \text{E}\{|h|^2\} \|z - z_i\|^{-\alpha} = P_i \|z - z_i\|^{-\alpha}, \]  

(2)

where \( \alpha \) is the propagation path-loss exponent, and \( P_i \) denotes the transmit power of the \( i \)-th BS in the \( k \)-th tier. On the other hand, the interference imposed on the user at the point \( z \) by all
the other BSs can be expressed as

$$I_i(z) = \sum_{j=1}^{K} \sum_{l \in \Phi_j, j \neq i} P_l \mathbb{E}\{ |b_j|^2 \} \| z - z_l \|^{-\alpha} \sum_{l \in \Phi \setminus i} P_l \| z - z_l \|^{-\alpha},$$

where $\Phi_j$ is a set containing the BSs belonging to the $j$-th tier.

For notational convenience, we introduce $\Phi = \bigcup_{k=1}^{K} \Phi_k$ in (3).

Given the SIR threshold $\tau_k$, the coverage of the $i$-th BS in the $k$-th tier is given by the enclosed boundary consisting of the points, which yield the SIR equalling to the threshold $\tau_k$. Outside this region, the SIR is below the threshold $\tau_k$. Typically in a multi-tier HetNet, the noise power is much smaller than the interference power. Therefore, the coverage area of a BS can be usually determined by the equivalent SIR according to the condition of

$$\text{SIR}_i(z) = \frac{S_i(z)}{I_i(z)} = \frac{P_i \| z - z_i \|^{-\alpha} \sum_{l \in \Phi \setminus i} P_l \| z - z_l \|^{-\alpha}}{\sum_{l \in \Phi_j, j \neq i} P_l \mathbb{E}\{ |b_j|^2 \} \| z - z_l \|^{-\alpha}} \geq \tau_k. \quad (4)$$

For a two-tier HetNet consisting of macro BSs and micro BSs, the coverage areas can be observed in Figure 3, when actual BS distribution is taken as an example. The coverage areas of BSs show irregular circles. Furthermore, there is a discrepancy between the coverages generated by the grid or Voronoi cells, and the real equal-SIR cell coverage results. Although both the Voronoi cell and grid cell are generally constructed for cell boundary analysis in cellular networks, we show an irregular circle boundary of given multi-tier cells by simulation under an equal SIR constraint, in which its irregular circles are obviously different from the Voronoi cells and grid cells. On this basis, a given-cell could be constructed as a closure line having an identical SIR, and the coverage areas in multi-tier BSs can be constructed as irregular circles.

### 3.2 Static coverage probability analysis

We assume the equal power ratio among BSs for a 1-tier HetNet, while setting the power ratio between macro BS and femto BS to 1000:1 for a 2-tier HetNet, and the path-loss exponent $\alpha$ to 4. As shown in Figures 4 and 5, the static system-level coverage probability is compared for Smithfield dataset and Glasgow dataset, respectively, in which theoretical results are consistent with simulation results, for both repulsive BS dataset and clustering BS dataset. Moreover, we take the coverage probability results in PPP model as a comparison [8–10]. In Smithfield dataset, it can be observed in Figure 4 that the PPP results are approaching to the 1-tier results, but with a large deviation to the 2-tier results. However, in Glasgow dataset, it can be observed in Figure 5 that the PPP results are only approaching to the 2-tier results in relatively high SIR regions, but still with a deviation to the rest results. On the whole, it illustrates that the PPP model is not a proper model for such two datasets.
Especially, when the characteristics of BS deployment are kept with the same, it can be observed that the coverage probability for both two datasets is improved under the 2-tier network scenario, compared to the 1-tier network scenario. One reason is that BSs with smaller power could effectively reduce interference to the other BSs. Moreover, the comparison of the coverage probability difference for 1-tier and 2-tier scenarios is shown in Figure 6. It can be found that the coverage probability improves more at high SIR regions for a repulsive BS distribution, like Smithfield dataset. On the other side, the coverage probability improves more at low SIR regions for a clustering BS distribution, like Glasgow dataset. As shown in Figure 6, the SIR threshold for high SIR regions and low SIR regions is 6 dB. When the SIR is in the range of [6, 20], the coverage probability improves more at high SIR regions for repulsive BS distribution (Smithfield dataset), where the range of coverage probability improvement is from 0.2620 to 0.0932. When the SIR is in the range of [−10, 6], the coverage probability improves more at low SIR regions for clustering BS distribution (Glasgow dataset), where the maximum value of coverage probability improvement is 0.5858. It illustrates that the efficiency of small cells is much better for the repulsive BS deployment under high SIR regions, while much better for the clustering BS deployment under low SIR regions. This result helps to guide small cell deployment in future.

3.3 Model fitting for smithfield dataset

We further consider finding the proper repulsive spatial point process model that matches the data sample.

3.3.1 Single model fitting

By means of a statistical tool in R language, we attempt to make goodness-of-fit for the Smithfield dataset by using some repulsive spatial point models, which can be called for analysis through spatstat software [37].

It can be observed from Figure 8 that for 0 < r < 88.76, the K function of real dataset approximates to theoretical models, such as Fiksel, Geyer and Strauss, in which the Geyer model yields the smallest gap. But for the hardcore model, the K function of real dataset is outside its theoretical upper envelope, which indicates that it is not suitable for the dataset in this size segment. On the other hand, for r ≥ 88.76, the K function of real dataset deviates from the theoretical value, but is still within its theoretical lower envelope, which is based on Fiksel, Geyer and Strauss models [33]. However, for the hardcore model, the K function of real dataset approximates to its theoretical value in this size segment.

On the other hand, Figure 9 shows the analysis of L function. For 0 < r < 88.76, the L function of real dataset still approximates to the theoretical models, such as Fiksel, Geyer and Strauss, in which the Geyer model yields the smallest gap. For r ≥ 88.76, the L function of real dataset deviates from the theoretical value, but is still within its theoretical lower envelope, which is based on Fiksel, Geyer and Strauss models. But for the hardcore model, its L function of real dataset approximates to its theoretical value in this size segment.
3.3.2 Hybrid model fitting

Since it is difficult to completely match the dataset by using a single model, we consider taking a hybrid model in spatsat software, which generally combines with two types of spatial point processes, and adjusts its theoretical parameters to fit the dataset. The process is shown in Algorithm 1.

In simulation, we attempt to take many different types of hybrid models, and find a hybrid model with the smallest gap, which combines with Strauss(150) and Geyer(50, 2). In such a hybrid model, the interaction radius of the Strauss process is set to 150 m, while the interaction radius of the Geyer process is set to 50 m and its saturation threshold is set to 2. The result of the hybrid model is shown in Figure 10.

The cumulative sums of absolute errors for different models are listed in the Table 2. It is observed that the gap between the real data and the theoretical value is the smallest for hybrid model, where the cumulative sums of absolute errors satisfy $7.6113 \times 10^5$ for K function and $1.9850 \times 10^3$ for L function. From the above analysis of the K function and the L function, it can be found that the saturation process fits well with for the dataset in Smithfield. In particular, the Geyer model is a type of saturation process, which considers its pairwise interaction with all other points. Especially, when its saturation threshold is set to infinity, the Geyer model may approach to the Strauss model. The analysis reveals that even though it is hard to find a single saturation process to fit the real dataset, it is feasible to consider combining multiple saturation processes in a hybrid way to approach the dataset.

3.4 Model fitting for glasgow dataset

As shown in Figure 11(a), the K function of real data in Glasgow is outside the upper envelope boundary of the PPP model, which illustrates that the PPP model could not well match the actual results. It can be observed from Figure 11(b) that the L function of real data in Glasgow satisfies $L(r) > r$, which reveals that it corresponds to a clustering point process.
FIGURE 9  L function analysis for BSs in Smithfield dataset

FIGURE 10  Hybrid model analysis for BSs in Smithfield dataset
ALGORITHM 1 Searching for the precise hybrid model

1: Input: dataset.
2: for \( i = 1, 2, \ldots, X \) do
3:   for \( j = 1, 2, \ldots, Y \) do
4:     \( p[i][j] = 0; q[i][j] = 0 \)
5:   end for
6: end for
7: for \( i = 1, 2, \ldots, X \) do
8:   for \( j = 1, 2, \ldots, Y \) do
9:     fit \( \leftarrow \) ppm(data, \( \sim 1 \), Hybrid(\( A = \text{ Strauss}(i) \), \( B = \text{ Geyer}(j, 2) \)), \( K_{\text{function}} \leftarrow \) envelope(fit, Kest, nsim = 39), \( L_{\text{function}} \leftarrow \) envelope(fit, Lest, nsim = 39))
10:   \( K_{i,j}(r) \leftarrow K_{\text{function}}(r) \leftarrow L_{\text{function}} \)
11:   if \( \sum_{r=0}^{\infty} |K_{i,j}(r) - K_{\text{real}}(r)| \leq \varepsilon_k \) then
12:     \( p[i][j] = 1 \)
13:   end if
14:   if \( \sum_{r=0}^{\infty} |L_{i,j}(r) - L_{\text{real}}(r)| \leq \varepsilon_l \) then
15:     \( q[i][j] = 1 \)
16: end if
17: end for
18: end for
19: Output: \( p, q \).
20: Evaluate the hybrid model.

3.4.1 Model fitting algorithm

On this basis, we also choose a series of clustering point process models to fit the real data and compare its results, respectively, with the Thomas model, the VarGamma model, LGCP and the Cauchy model. Furthermore, we attempt to use these clustering spatial point process models to fit the dataset, and the algorithm for searching reasonable models is shown in Algorithm 2.

3.4.2 Fitting results

We consider the clustering spatial point process models to make goodness-of-fit for the Glasgow dataset sample, such as a Thomas model, a VarGamma model, an LGCP model and a Cauchy model. By using the spatstat software [37], we further test the goodness-of-fit for the Thomas model, the VarGamma model, LGCP and the Cauchy model. Figure 12 shows the K function fitting based on four different models. It can be observed that the K function theoretical values of these models approximate to the actual values of real data, and it is also within the upper and lower boundaries of the envelope, which verify a validity of these models. Besides, the LGCP, Thomas, Cauchy and VarGamma models yield smaller gap among these models.

We further test the goodness-of-fit for the Thomas model, the VarGamma model, LGCP and the Cauchy model by using L function. Figure 13 shows the L function fitting based on the above models, respectively. It can be observed from Figure 11(b) that the L function value of real data lies outside the upper and lower boundaries of the envelope, which verifies an invalidity of PPP model. But for the Thomas model, the VarGamma model, LGCP and the Cauchy model, the L function theoretical values of these models approximate to the actual values of real data, which lie within the upper and lower boundaries of the envelope, which verify a validity of these models.

3.4.3 Error test

The cumulative sums of absolute errors for different models are listed in the Table 3, it can be found that, among all the test models in Glasgow dataset, the minimum cumulative sum of absolute errors for K function is equal to 1.5585 \times 10^6 in the LGCP model, while the minimum cumulative sum of absolute errors for L function is equal to 2.4005 \times 10^3 in the Cauchy model. As a whole, the LGCP model is the most reasonable model
from the perspective of K function, and the Cauchy model is the most reasonable model from the perspective of L function. Especially, Figure 14 shows the relative errors of K function and L function respectively, in which the expression of relative error is defined as follows

$$\delta = \left| \Theta - \Re \right| / \Re,$$

where $\delta$ denotes the relative error, $\Theta$ denotes the theoretical value and $\Re$ denotes the real value. Combined with Table 3 and Figure 14, it can be found that the errors in both LGCP model and Cauchy model are very small and close with each other, so both LGCP model and Cauchy model are reasonable models for real BS deployment in Glasgow dataset.

4 | DYNAMIC COVERAGE ANALYSIS BY ADDING BSs

Generally, we could provide a system-level result as a whole by using the static analysis of BS deployment. However, the process of network optimization is dynamic according to the needs of operators. It is necessary to add additional micro cells or femto cells to accommodate service demand at hot spots, and
FIGURE 13  L function analysis for BSs in Glasgow dataset

FIGURE 14  Error Test for different models in Glasgow dataset
to deploy new BSs at appropriate locations. Therefore, it is critical to analyze the coverage dynamically from the viewpoint of given-cell deployment. In other words, how to add macro BS or femto BS, and its effects on the coverage problem.

### 4.1 BS location searching scheme

The existing studies usually pay attention to the characteristic of real BS deployment in the static coverage analysis. However, the effects of adding BS are neglected. Extending the static coverage analysis, we here consider the dynamic coverage analysis for BS deployment. In order to meet the service demand at hot-spots, it is necessary to add additional micro cells or femto cells, and to deploy new BSs at appropriate locations. In addition, it is highly desirable to find the near optimal positions for deploying the new BSs within a bounded region of interest. Intuitively, if we can place a new BS at the point in a region $S$ of interest, where the BS receives minimum interference from the existing BSs. The transmit power of this new BS, which requires to meet the coverage, may be minimized.

Based on irregular circle cells, we adopt the following strategy in order to find the minimum interference location in a target region $S$, and divide the target region $S$ into a number of sub-regions. After we find the minimum interference point in each of the sub-regions, we can then obtain the minimum interference point in the whole region $S$, simply by comparing the minimum interference points of the sub-regions. According to [8], the coverage results derived from the delaunay triangulation are usually closer to the practice than that derived from the grid-based method [8]. Therefore, as shown in Figure 15, we take a delaunay triangulation approach to partition the region $S$ into $N$ triangular sub-regions, where the value of $N$ depends on the number of existing BSs contained in $S$. Consequently, the site selection problem is to find the minimum interference point within each of the delaunay triangles.

Assume that there are $M$ existing BSs in the region of interest $S \subseteq \mathbb{R}^2$. The BS indices are listed in a set as $\Phi = \{1, 2, \ldots, M\}$, and the corresponding locations of these $M$ BSs are given by the set of $\{z_1, z_2, \ldots, z_M\}$. Since the region $S$ is partitioned into $N$ delaunay triangles, we further assume that the $n$-th delaunay triangle, denoted by $T_n \subseteq S$, has three vertices, $z_n, z_{n+1}$ and $z_{n+2}$, $n_1, n_2, n_3 \in \Phi$, representing that $z_n$, $z_{n+1}$ and $z_{n+2}$ indicate the locations of three existing BSs. Then, at any point $z \in T_n$, the interference received from the existing $M$ BSs is given by

$$I_a(z) = \sum_{m \in \Phi} P_m \|z - z_m\|^{-\alpha}, \quad z \in T_n,$$  

where $P_m$ is the transmit power of the BS located at $z_m$. The minimum interference point $z_{a,\text{min}}$ in $T_n$ is, therefore, given by the solution of the optimization problem

$$z_{a,\text{min}} = \arg\min_{z \in T_n} I_a(z).$$  

In order to find $z_{a,\text{min}}$, the initial iteration is given as follows

$$z_a^{(0)} = \frac{\sum_{m \in \Phi} P_m z - z_m}{\sum_{m \in \Phi} P_m},$$

Next, the following iterative gradient expression can be expressed as [38]

$$z_a^{(t)} = z_a^{(t-1)} - \mu \frac{\nabla_z I_a(z_a^{(t-1)})}{\|
abla_z I_a(z_a^{(t-1)})\|},$$  

where $(t)$ denotes the number of iterations, $\nabla_z$ denotes to calculate the gradient about $z_a$, and $\mu$ is a small positive step size. Note that, after iterations, if $z_a^{(t)}$ lies outside $T_n$, then it is projected onto the boundary of $T_n$ along $\nabla_z I_a(z_a^{(t-1)})$. The algorithm is terminated, when $\nabla_z I_a(z_a^{(t)}) \rightarrow 0$ or when $z_a^{(t)}$ is on the boundary of $T_n$. The initial point $z_a^{(0)}$ can be set to the center of gravity of $T_n$, which is given by $z_a^{(0)} = \frac{z_1 + z_2 + z_3}{3}$. In summary, the algorithm for finding the minimum interference point $z_{\text{min}}$ in deploying a new BS is given by Algorithm 3.

Especially, the time complexity of our searching algorithm is $O(m^2)$, which increases as the number of statistical base stations.

**Algorithm 2** Searching model for Glasgow dataset

1. **Input:** dataset.
2. **switch (a)** Case 1: $a \leftarrow \text{VarGamma}$ Case 2: $a \leftarrow \text{Thomas}$ Case 3: $a \leftarrow \text{MatClust}$ Case 4: $a \leftarrow \text{Cauchy}$ Case 5: $a \leftarrow \text{LGCP}$
3. **end switch**
4. $fit < kppm(data, 1)$
5. $K_{\text{fit}} \leftarrow \text{envelope}(fit, Kst, nim = 39)$
6. $L_{\text{fit}} \leftarrow \text{envelope}(fit, Lst, nim = 39)$
7. Evaluate the dataset.

**FIGURE 15** Delaunay triangulation for coverage area of HetNet

![Delaunay triangulation](image)
### ALGORITHM 3 Searching algorithm for deploying a new BS

1: Input: The area of interest $S$, existing BS indices: $\Phi = \{1, 2, \ldots, M\}$ and their locations: $\{z_1, z_2, \ldots, z_M\}$; The transmission power of BSs: $\{P_1, P_2, \ldots, P_M\}$; step size: $\mu$.
2: Delaunay triangulation to partition $S$ into $N$ delaunay triangles $\{T_n\}_n$. Let the vertices of $T_n$ be denoted as $z_{n1}$, $z_{n2}$ and $\varsigma_n$. Compute the centre of gravity of $T_n$ as $z_n^{(0)} = \frac{z_{n1} + z_{n2} + \varsigma_n}{3}$, for $n = 1, \ldots, N$.
3: for $n = 1, 2, \ldots, N$ do
4: Find the minimum interference point $\zeta_{n\min}$ of $T_n$ using the iterative gradient algorithm of (9);
5: Calculate the interference value $I_n(\zeta_{n\min})$ at the point $\zeta_{n\min}$.
6: end for
7: Solve the problem of $\hat{n} = \arg\min_{1 \leq j \leq N} I_n(\zeta_{n\min})$. Then, $\zeta_{\min} = \zeta_{n\min}$ gives the minimum interference point in $S$ for deploying the new BS.

### ALGORITHM 4 Coverage probability statistical analysis

1: Input: The area of interest $S$; BS indices: $\Phi = \{1, 2, \ldots, M\}$; BS locations: $\{z_1, z_2, \ldots, z_M\}$; transmission power of BSs: $\{P_1, P_2, \ldots, P_M\}$; SIR thresholds: $\{\tau_1, \tau_2, \ldots, \tau_M\}$.
2: Uniformly and randomly distribute $N_u$ UEs in $S$, whose indices are given by $U = \{n_1, n_2, \ldots, n_{N_u}\}$, and locations are denoted by $\{\varsigma_{n1}, \varsigma_{n2}, \ldots, \varsigma_{n_{N_u}}\}$.
3: $m = 0$;
4: for $i = 1, 2, \ldots, N_u$ do
5: for $k = 1, 2, \ldots, M$ do
6: Compute $\text{SIR}_k(\varsigma_{ni})$ using Equation (10);
7: end for
8: Find the serving BS of $U_i$ using Equation (11);
9: if $\text{SIR}_k(\varsigma_{ni}) > \tau_k$ then
10: $m \leftarrow m + 1$
11: end if
12: end for
13: Evaluate the coverage probability of the Hetnet by Equation (12).

Then, based on the SIR values caused by the $M$ BSs, the one serving the $u_m$-th UE can be chosen as

$$k_{u_m} = \arg\max_{k \in \Phi} \text{SIR}_k(\varsigma_{u_m}).$$

Furthermore, if $\text{SIR}_k(\varsigma_{u_m}) > \tau_{k_{u_m}}$, the required SIR of the $u_m$-th UE is met. Otherwise, the UE is not efficiently covered. Therefore, the coverage probability can be defined as the percentage of the UEs, which meet their SIR requirements, formulated as

$$P_{cp} = \frac{\sum_{j} 1(\text{SIR}_{k_j}(\varsigma_{u_j}) \geq \tau_{k_j})}{N_u},$$

where $1(\text{SIR}_{k_j}(\varsigma_{u_j}) \geq \tau_{k_j}) = 1$, if $\text{SIR}_{k_j}(\varsigma_{u_j}) \geq \tau_{k_j}$. Otherwise, $1(\text{SIR}_{k_j}(\varsigma_{u_j}) \geq \tau_{k_j}) = 0$.

## 5 DYNAMIC COVERAGE RESULTS BY ADDING BSs

In our simulations, we set the power ratio between macro BS and femto BS to 1000:1, the path-loss exponent $\alpha$ to 4, and the SIR threshold $\tau$ to 4.77dB. Furthermore, we assume that active UEs are uniformly distributed in the areas of interest, with a density satisfying $\lambda_f \gg \lambda$, where $\lambda$ is the density of BSs.
The numerical results obtained by Algorithm 3 are shown in Figures 16–18, for a Poisson distribution model, urban and rural BS distribution models, respectively. In these figures, three new BSs are planned to be placed at the area within the solid red rectangle. A pentagram indicates the minimum interference location, while diamonds indicate the candidate locations with non-minimum interference. It should be noted that, in our simulation, the target area is completely divided into triangles by the delaunay triangulation, in order to avoid the boundary effect.

As listed in Table 4, it can be observed that the interference received at three different candidate locations, for a Poisson distribution, a real urban area (Smithfield) and a rural area (Glasgow), respectively. Especially, candidate location 1 is reserved for a macro-cell, and candidate location 2 is reserved for a micro-cell, which is considered according to the strength of interference. By means of the proposed scheme, it provides a way to find a series of candidate locations for adding new BSs, according to the interference based on gradient decrement. Once the locations of the existing and added BSs are fixed, the coverage probability can be evaluated by using Algorithm 4, and the result is listed in Table 5 for three different types of distribution: the Poisson distribution, urban area and rural area. It can be found that adding a macro BS at the candidate location improves the coverage.

We assume that the BS density is set to 8.75 atoms/km² for macro BSs and 55.60 atoms/km² for femto BSs. Then, the coverage improves in the range of 2.08%, 3.64%, 0.93%, respectively, for the three models: Poisson, urban, and rural area. By contrast, adding a femto BS at the candidate location can improve the local QoS, but not necessary the overall system performance. As shown in the table, the coverage improves in the range of 0.04%, 0.49%, −1.14%, respectively, for the three models considered too. Hence, depending on the situations, the performance of the entire network may improve or degrade.

The effect of femto BS density on the coverage probability in a two-tier Hetnet with different site selection methods is shown in Figure 19, where the density of macro BS $\lambda_1$ is set to 5 atoms/km², while the femto BS density is initialized to $\lambda_2 = 0$. 

### Table 4: The interference in different locations

| Distribution | New BS Loc. | Candidate Loc. 1 | Candidate Loc. 2 |
|--------------|-------------|------------------|------------------|
| Poisson      | $7.1162 \times 10^{-10}$ | $7.3510 \times 10^{-9}$ | $2.0451 \times 10^{-9}$ |
| Smithfield   | $8.7484 \times 10^{-10}$ | $1.2613 \times 10^{-9}$ | $9.1495 \times 10^{-10}$ |
| Glasgow      | $2.6690 \times 10^{-10}$ | $6.4699 \times 10^{-10}$ | $6.2940 \times 10^{-10}$ |

### Table 5: The coverage probability after adding a new BS

| Pattern               | Poisson | Urban | Rural |
|-----------------------|---------|-------|-------|
| Before add a new BS   | 76.40   | 68.12 | 72.79 |
| Adding a Macro BS     | 77.99   | 70.60 | 73.47 |
| Percentage            | 2.08%   | 3.64% | 0.93% |
| Adding a Femto BS     | 76.43   | 68.46 | 71.96 |
| Percentage            | 0.04%   | 0.49% | −1.14%|
We assume that the SIR threshold for macro BSs is $\beta_1 = 2\text{dB}$, and that for femto BSs is $\beta_2 = 1\text{dB}$, while the size of the simulation area is set to 1000 km$^2$. The results of Figure 19 shows that our proposed scheme yields the highest coverage probability among the three site selection methods: our proposed one, the grid model [8] and the PPP model [11].

The coverage probability remains constant, when the thresholds for macro BSs and femto BSs are kept the same after adding femto BS. The coverage probability can increase if the threshold for macro BSs is greater than that for femto BSs, while it decreases when the threshold for macro BSs is less than that for femto BSs. Although adding a new BS is motivated to improve network coverage, it also results in the interference among nearby BSs. Consequently, the performance of the entire network may become saturated, when the density of BSs reaches a certain value. As shown in Figure 19, the coverage of networks may even decrease due to a resultant increase in the interference as the density of femto BSs increases. When the density of femto BSs reaches a certain value, approximately 30 atoms/km$^2$, the network coverage does not further improve by adding more femto BSs.

6.1 CONCLUSIONS

Considering the complication of optimizing BS deployment problems in future cellular networks, we chose repulsive BS dataset and clustering BS dataset to make statistical modelling, and analyzed the real BS deployment from a static and dynamic perspective respectively. The real datasets were statistically characterized with various types of spatial point processes. Our work could provide a guidance for BS location optimization, such as the modelling of BS dataset fitting and small cell deployment.

It showed that the coverage probability increased more at high SIR regions for the repulsive BS dataset in Smithfield. Since it was hard to find a single saturation process to fit this real dataset, we proposed an algorithm to search for reasonable joint saturation models, which was feasible to combine multiple saturation processes in a hybrid way to approach the dataset. The error results showed that our proposed hybrid model fitted well with the real repulsive BS dataset in Smithfield. The difference was that the coverage probability increased more at low SIR regions for the clustering BS dataset in Glasgow. Especially, we proposed a searching algorithm to find the proper clustering point processes, and made a good-of-fit test for this dataset. The error results illustrated that the LGCP and Cauchy model became reasonable models.

Furthermore, on the basis of statical BS deployment analysis, we investigated a dynamic analysis for the effect of adding new BSs on the coverage problem. In addition, we constructed a cell boundary with irregular circles under an equal SIR constraint. Moreover, we proposed a BS placement searching algorithm to find the location for adding new BSs, in which we placed additional BSs at the points of minimum interference in the region of interest. The numerical results showed that the coverage probability increased after adding BSs in the target area of a heterogeneous network by using the proposed scheme, but as the density of the femto BS reached a certain value, its coverage remained unchanged even after adding more femto BSs.

ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation Project of China (No. 61671254, 61871239).

REFERENCES

1. Tran, T.X., et al.: Collaborative mobile edge computing in 5G networks: New paradigms, scenarios, and challenges. IEEE Communications Magazine 55(4), 54–61 (2017)
2. Zhang, H., et al.: Network slicing based 5G and future mobile networks: Mobility, resource management, and challenges. IEEE Communications Magazine 55(8), 138–145 (2017)
3. Zhang, H., Li, L., Sung, D.K.: Closed-form expression for coverage and average rate of 2D and 3D cellular networks. Electronics Letters 55(18), 1015–1017 (2019)
4. Zhang, H., Sung, D.K., Wang, J.: Modelling and analysis of coverage for unmanned aerial vehicle base stations. IET Communications 14(17), 2878–2888 (2020)
5. Stefanatos, S., Alexiou, A.: On the competition among small cell wireless operators with large scale deployments. IEEE Transactions on Mobile Computing 16(7), 1981–1993 (2017)
6. Moller, J., Waagepetersen, R.P.: Statistical Inference and Simulation for Spatial Point Processes. CRC Press, Boca Raton (2003)
7. Blaszczyszyn, B., et al.: Stochastic Geometry Analysis of Cellular Networks. Cambridge University Press, Cambridge (2018)
8. Andrews, J.G., Baccelli, F., Ganti, R.K.: A tractable approach to coverage and rate in cellular networks. IEEE Transactions on Communications 59(11), 3122–3134 (2011)
9. Dhillon, H.S., et al.: Modeling and analysis of K-tier downlink heterogeneous cellular networks. IEEE Journal on Selected Areas in Communications 30(3), 550–560 (2012)
10. Zhang, H., et al.: A universal approach to coverage probability and throughput analysis for cellular networks. IEEE Transactions on Vehicular Technology 64(9), 4245–4256 (2015)
11. ElSawy, H., Hossain, E., Haenggi, M.: Stochastic geometry for modeling, analysis, and design of multi-tier and cognitive cellular wireless networks: A survey. IEEE Communications Surveys & Tutorials 15(3), 996–1019 (2013)
12. Saha, C., Afshang, M., Dhillon, H.S.: 3GPP-inspired hetnet model using Poisson cluster process: Sum-product functionals and downlink coverage. IEEE Transactions on Communications 66(5), 2219–2234 (2018)
13. Flint, I., et al.: Analysis of heterogeneous wireless networks using Poisson hard-core hole process. IEEE Transactions on Wireless Communications 16(11), 7152–7167 (2017)
14. Yang, J., et al.: Approximate coverage analysis of heterogeneous cellular networks modeled by Poisson hole process. In: IEEE 18th International Conference on Communication Technology (ICCT), Chongqing (2018)
15. Taylor, D.R. et al.: Pairwise interaction processes for modeling cellular network topology. In: IEEE Global Communications Conference (GLOBECOM), Anaheim (2012)
16. Guo, A., Haenggi, M.: Spatial stochastic models and metrics for the structure of base stations in cellular networks. IEEE Transactions on Wireless Communications 12(11), 5800–5812 (2013)
17. Li, Y., et al.: Statistical modeling and probabilistic analysis of cellular networks with determinantal point processes. IEEE Transactions on Communications 63(9), 3405–3422 (2015)
18. Deng, N., Zhou, W., Haenggi, M.: The Ginibre point process as a model for wireless networks with repulsion. IEEE Transactions on Wireless Communications 14(1), 107–121 (2015)
19. Kong, H.-B., Wang, P., Niyato, D.: Performance analysis of wireless sensor networks with Ginibre point process modeling. In: IEEE International Conference on Communications (ICC), Paris (2017)
20. Haenggi, M.: The mean interference-to-signal ratio and its key role in cellular and amorphous networks. IEEE Wireless Communications Letters 3(6), 597–600 (2014)
21. Guo, A., Haenggi, M.: Asymptotic deployment gain: A simple approach to characterize the SINR distribution in general cellular networks. IEEE Transactions on Communications 63(3), 962–976 (2015)
22. Ganti, R.K., Haenggi, M.: Asymptotics and approximation of the SINR distribution in general cellular networks. IEEE Transactions on Wireless Communications 15(3), 2130–2143 (2015)
23. Wei, H., et al.: Approximate SIR analysis in general heterogeneous cellular networks. IEEE Transactions on Communications 64(3), 1259–1275 (2016)
24. Takahashi, Y., et al.: Simple and fast PPP-based approximation of SIR distributions in downlink cellular networks. IEEE Wireless Communications Letters 7(6), 898–901 (2018)
25. Kibilda, J., Galkin, B., Dasilva, L.A.: Modelling multi-operator base station deployment patterns in cellular networks. IEEE Transactions on Mobile Computing 15(12), 3087–3099 (2016)
26. Liu, Z.-g. Ma, et al.: Research on coverage probability in ultra-dense 5G heterogeneous cellular networks based on Poisson clustered process. Wireless Personal Communications 95(3), 2915–2930 (2017)
27. Zabini, F., Pasolini, G., Conti, A.: On random sampling with nodes attraction: The case of Gauss-Poisson process. In: IEEE International Symposium on Information Theory (ISIT), Aachen (2017)
28. Chun, Y.J., Hasna, M.O., Ghrayeb, A.: Modeling heterogeneous cellular networks interference using Poisson cluster processes. IEEE Journal on Selected Areas in Communications 33(10), 2182–2195 (2015)
29. Miyoshi, N.: Downlink coverage probability in cellular networks with Poisson–Poisson cluster deployed base stations. IEEE Wireless Communications Letters 8(1), 5–8 (2019)
30. Flint, I., et al.: Analysis of heterogeneous wireless networks using Poisson hard-core hole process. IEEE Transactions on Wireless Communications 16(11), 7152–7167 (2017)
31. Yang, J., Pan, Z., Guo, L.: Coverage and energy efficiency analysis for two-tier heterogeneous cellular networks based on Matern hard-core process. Future Internet 12(1), 1–13 (2020)
32. Kalamkar, S.S., Haenggi, M.: Simple approximations of the SINR meta distribution in general cellular networks. IEEE Transactions on Communications 67(6), 4393–4406 (2019)
33. Haenggi, M.: Stochastic geometry for wireless networks. Cambridge University Press, Cambridge (2012)
34. Coeurjolly, J., Moller, J., Waagepetersen, R.: A tutorial on palm distributions for spatial point processes. International Statistical Review 85(3), 404–420 (2017)
35. Chiu, S.N. et al.: Stochastic Geometry and its Applications. John Wiley & Sons, New York (2013)
36. Taylor, B.M. et al.: lgcp: An R package for inference with spatio-temporal Log-Gaussian Cox processes. arXiv:1110.6054 (2011)
37. Baddeley, A., et al.: spatstat: an R package for analyzing spatial point patterns. Journal of Statistical Software 12(6), 1–42 (2005)
38. Bertsekas, D.P.: Nonlinear programming, 2nd ed. Athena Scientific, Nashua (1999)

How to cite this article: Zhang H, Feng L, Zheng T, Sung DK. Characterization of base station deployment distribution and coverage in heterogeneous networks. IET Commun. 2021;15:850–865. https://doi.org/10.1049/cmu2.12126