Point Process Discrimination According to Repulsion

Contents

In numerous applications, cloud of points do seem to exhibit *repulsion* in the intuitive sense that there is no local cluster as in a Poisson process. Motivated by data coming from cellular networks, we devise a classification algorithm based on the form of the Voronoi cells. We show that, in the particular set of data we are given, we can retrieve some repulsiveness between antennas, which was expected for engineering reasons.

In the performance analysis of cellular systems, the locations of antennas (or base stations) play a major role (see [1]). It is usually admitted that they can be modeled by a Poisson process. But the data which can be gathered from the Web site of the French National Agency of Radio Frequencies, Cartoradio, see [2], tend to prove that this may not be the case. More precisely, if we look at the global picture of all antennas in Paris, we see features reminiscent of a Poisson process (local clusters for instance), see Fig. 1 (left). However, if we look closer and finer, by specifying a region and a frequency band, we see that the antennas locations do seem to exhibit some repulsion (see Fig. 1, right picture).
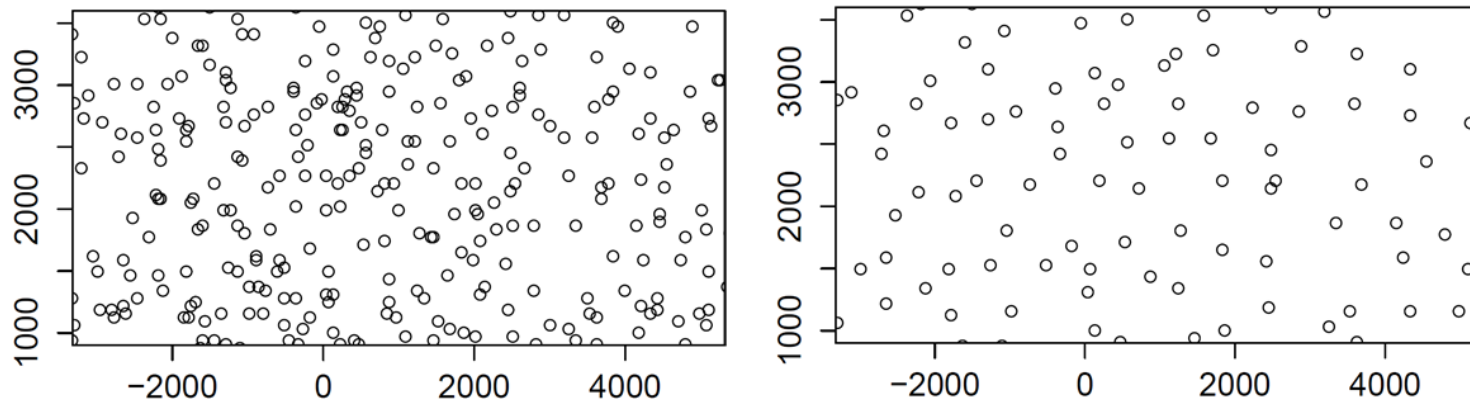
Back to top

Skip to main content

Fig. 1 Left: Antennas in Paris. Right: Antennas in one frequency band only.

In previous papers, point processes with repulsion have been used to model such systems [3], [4], [5] for no reason but a mere resemblance between the pictures like the right picture in Fig. 1 and those obtained by simulating a point process with repulsion. The question is then to decide, given one sample of positions of base stations in a bounded domain, whether it is more likely to be modeled by a point process with repulsion or by a *neutral* point process, i.e. where the locations could be considered as coming from independent drawings of some identically distributed random variables. As we only have a single realization, we cannot use frequency methods. Since the observation window is finite, we cannot either resort to estimates based on stationarity or ergodicity and we must take care from the side effects.

The rationale behind our work comes from [6]. It is shown there that the Voronoi cells of the Ginibre point process (a particular point process with repulsion, see below for the exact definition) are in some sense more regular (closer to a circle) than those of a Poisson process (see (3) in Theorem 1.). By simulation, this feature seems to persist for other point processes with repulsion, like Gibbs processes. In [7], the surface of Voronoi cells is claimed to be a good discrepancy indicator between Poisson process and several processes with repulsion (Gibbs processes, Strauss processes with repulsion and the Geyer saturation model). For any of these models, we do not have any closed formula on the surface of the Voronoi cells so the procedure proposed in this paper is to simulate a large number of realizations of each of these processes and compute the empirical mean and variance of the Voronoi cells area. They obtain mixed conclusions as this sole indicator does not enable to rule out the Poisson hypothesis for many situations.

Our contribution is to consider the ratio of the surface by the squared perimeter instead of the surface of the Voronoi cells alone. Actually, we can interpret the result of [6] by saying that the Voronoi cells of a Ginibre point process are more circular than those of a Poisson point process. The isoperimetric inequality stands for any regular enough domain in the plane, $R = \frac{4\pi S}{P^2}$ is less than 1 and the equality is obtained for disks. It is thus sensible to think that the ratio R will be closer to 1 for repulsive processes than for neutral point processes. Following the procedure of [7], we show that we get a much better

However, for the application we have in mind, which is to decide for one single map which model is the most pertinent, we cannot use this criterion based on probability. That is why we resort to an ML model. After several tries, we concluded that the most efficient algorithm was to use Logistic Regression. In a first step, we trained it on simulations of Ginibre and Poisson point processes. The advantage of the Ginibre process is that we have efficient algorithm to simulate it [8] and it does not seem to alter the accuracy of our algorithm to use one single class of repulsive point process. We remarked that we obtain a much better discrimination by considering the mean value of R for the five most central cells instead of just the most central one. We can even improve our discrimination rate by adding to the input vector the value of each of the five ratios.

Furthermore, the repulsion in the Ginibre class of point processes can be also modulated by making a β-thinning (to weaken the repulsion) and then a $\sqrt{\beta}$-dilation (to keep the same intensity of points per surface unit) to obtain what is called a β-Ginibre. For $\beta = 1$, we have the original Ginibre process and when β goes to 0, it tends in law to a Poisson process (see [9]) so that we have a full scale of point processes with intermediate repulsion between 0 and 1. We show that our logistic regression algorithm can still accurately discriminate between Poisson and β-repulsive point processes for β up to 0.7.

The paper is organized as follows. We first remind what is a Ginibre point process and the property of its Voronoi cells which motivates the sequel. Then two approaches are employed, one based on statistics and the other on machine learning, to classify the processes and compare their efficiencies and outcomes. Finally, tests are conducted on the Cartoradio data. We consider finite point processes on a bounded window E. The law of a such a point process N can be characterized by its correlation functions (for details we refer to [10][Chapter 5]). These are symmetric functions $(\rho_k, k \geq 1)$ such that for any bounded function f, we can write:

$$\mathbb{E}\left[\sum_{\alpha \subset N} f(\alpha)\right] = \sum_{k=1}^{+\infty} \frac{1}{k!} \int_{E^k} f(\{x_1, \ldots, x_k\}) \rho_k(x_1, \ldots, x_k) \, dx_1 \ldots dx_k.$$

Intuitively speaking, $\rho_k(x_1, \ldots, x_k) \, dx_1 \ldots dx_k$ represents the probability to observe in N, at least k points located around the point x_j. For a Poisson point process of control measure $m(x) \, dx$, we have

$$\rho_k(x_1, \ldots, x_k) = \prod_{j=1}^{k} m(x_j).$$

The **Ginibre point process**, restricted to $E = B(0, r)$, with intensity $\rho = \frac{\lambda}{\pi}$ (with $\lambda > 0$) has correlation functions (see [11])

$$\rho_k(x_1, \ldots, x_k) = \det(K(x_i, x_j), \ 1 \leq i, j \leq k) \tag{1}$$

where K is given by

$$K_r(x,y) = \sum_{j=1}^{\infty} \frac{\gamma(j+1,r^2)}{j!} \phi_j(x)\phi_j(\bar{y}) \tag{2}$$

with

$$\phi_j(x) = \sqrt{\frac{\rho}{\gamma(j+1,r^2)}} \left(\sqrt{\lambda}x\right)^j e^{-\frac{\lambda}{2}|x|^2}$$

and $\gamma(n,x)$ is the lower incomplete Gamma function. The simulation of such a point process is a delicate matter, first solved in [12]. It remains costly because the algorithm contains complex calculations and some rejections. In order to fasten the procedure, an approximate algorithm, with error estimates, has been given in [8] (see the bibliography therein to get the URL of the Python code).

For an at most denumerable set of points $\{x_n, n \geq 1\}$, the Voronoi cells are defined as the convex sets

$$\mathcal{C}(x_i) = \{z \in \mathbb{C}, |z - x_i| \leq |z - x_j| \text{ for all } j \neq i\}.$$

When the points are drawn from a point process, we thus have a collection of random closed sets. When the process under consideration is stationary with respect to translations, it is customary to define the typical law of a Voronoi cell as the law of the cell containing the origin of $\mathbb{R}^2$ when the point process is taken under its Palm distribution [6], [13]. It turns out that we know the Palm distribution of the Poisson process (which is itself) and of the Ginibre point process (the correlation functions are of the form (1) with K being K_R with the first term removed). We denote by $\mathcal{C}_p$ (respectively $\mathcal{C}_G$) the typical cell of the Voronoi tessellation associated to a stationary Poisson process in $\mathbb{C}$ with intensity λ (respectively to the Ginibre point process of intensity ρ). One of the main theorems of [6] is the following.

Theorem 1. When $r \to 0$,

$$\mathbb{E}\left[V(\mathcal{C}_G \cap B(0,r))\right] = \mathbb{E}\left[V(\mathcal{C}_p \cap B(0,r))\right](1 + r^2 W + \circ(r^2)) \tag{3}$$

where W is a positive random variable.

This theorem shows that near the germs of the cells a more important part of the area is captured in the Ginibre–Voronoi tessellation than in the Poisson–Voronoi tessellation. This is an indication that the Voronoi cells of the Ginibre point process are more circular than those given by the Poisson process. This can be corroborated by simulation as shows the Figure Fig. 2
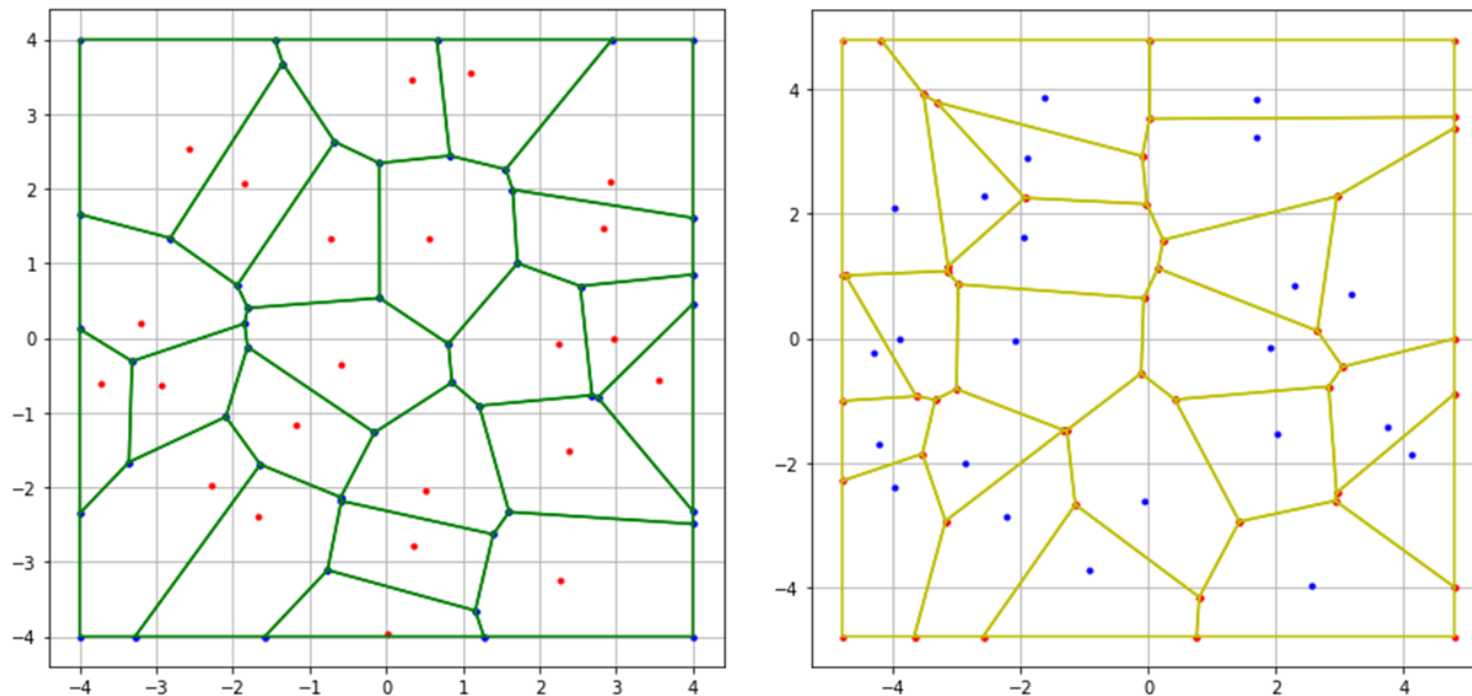
Fig. 2 On the left, Voronoi cells associated to a realization of a Ginibre process. On the right, Voronoi cells associated to a realization of a Poisson process.

As we know that circles saturate the isoperimetric inequality, it is sensible to consider classification algorithms based on area and squared perimeter of Voronoi cells. In order to avoid side effects, we concentrate on the innermost cells of the observation window.

The Cartoradio web site contains the locations (in GPS coordinates) and other informations about all the antennas (or base stations) in metropolitan France for any operator, any frequency band and all generation of wireless systems (2G to 5G). The capacity of an antenna depends on its power and on the traffic demand it has to serve. Outside metropolitan areas, the antennas are relatively scarce and located along the main roads to guarantee a large surface coverage (around 30 km^2). Hence there is no need to construct models for these regions. On the contrary, in big towns, the density of base stations is much higher to handle the traffic demand: An antenna covers around half a squared kilometer. This is where the dimensioning problem do appear. One should have a sufficient number of antennas per unit of surface to transport all the traffic, on the other hand, base stations operating in a given frequency band cannot be to close to mitigate interference. This explains the right picture of Figure Fig. 1.

When it comes to assess the type of point process we should consider in this situation, we cannot consider the city as a whole:

railway stations or touristic sites, for instance) which entails a higher density of antennas, ruin any kind of invariance a statistician could hope for. That means, we should restrict our expectations to local models of the size of a district or a bit more. Since interference, which are the main annoyance to be dealt with, are a local phenomenon, working on a partial part of the whole domain is sufficient to predict the behavior and dimension a wireless network.

In the following sections, we will use Python code that assumes that the following packages have been loaded:

▶ Show code cell source

Statistical approach

Given a circular domain with N points, we want to decide whether the points exhibit repulsion or not. To do so, we will begin with a statistical approach, where we will first calculate, for Poisson processes as well as for Ginibre and β-Ginibre processes, the probability that the ratio $R = \frac{4\pi S}{P^2}$ of the central cell is less than or equal to r, for values of r ranging from 0 to 1. And then we apply the same approach using the mean ratio of the five central cells. Finally, we will calculate 95% confidence intervals for each of these processes.

The following code illustrates the generation of various point samples and the calculation of the surface to squared perimeter ratios given the number of points N and the parameter β for β-Ginibre processes. The Ginibre and β-Ginibre processes are generated using the "sample" function given in the Python code of [8].

▶ Show code cell source

```
Importing libraries ...
Compiling functions ...
```

The simulation algorithm provides a method for computing the quantity $\mathbb{P}\left(\frac{4\pi S}{P^2} \leq r\right)$ as a function of r for the Ginibre processes (the same algorithm is applied to other processes as well). The Algorithm takes as input the number of points N, the number of experiences for the simulation N_{exp} and the range of the variable r as a list of values. Since the simulations require a lot of time to run, we are not going to attach the associated Python code, the latter is based on the algorithm described in the following Python code using the functions defined previously.

▶ Show code cell source

Figure Fig. 3 shows the results of the simulations, where we compare the confidence intervals of the poisson process with the Ginibre process and the 0.7-Ginibre process, using first the central cell and then the five central cells.

Skip to main content

Fig. 3 Simulation results using the central cell (up) and the five central cells (down).
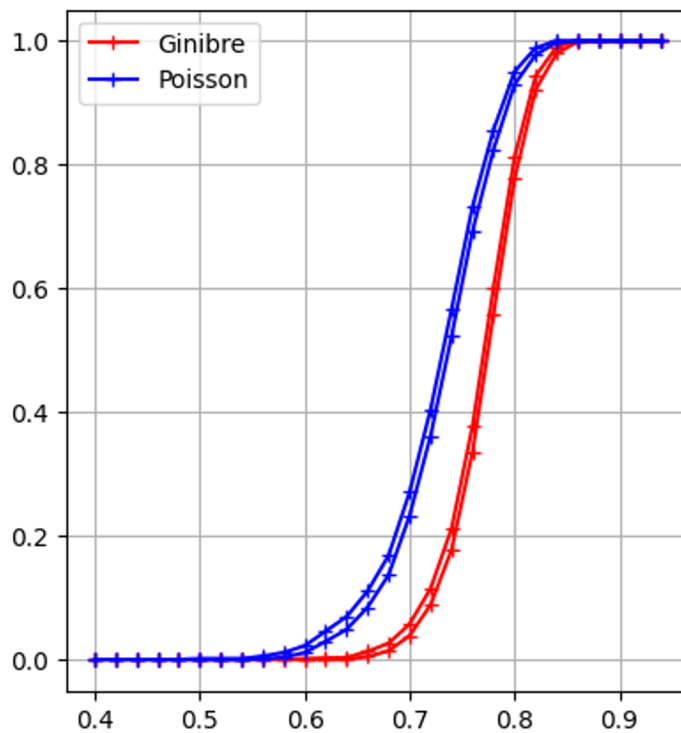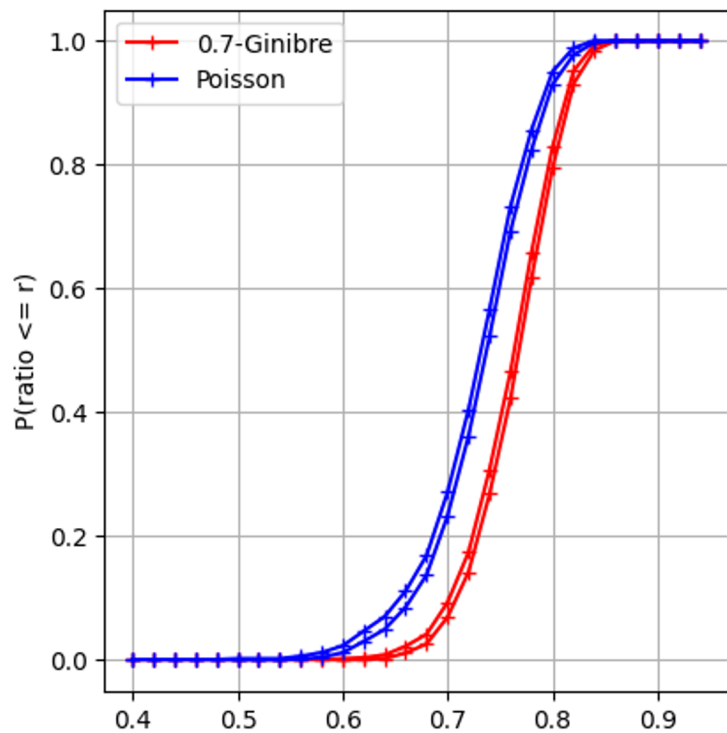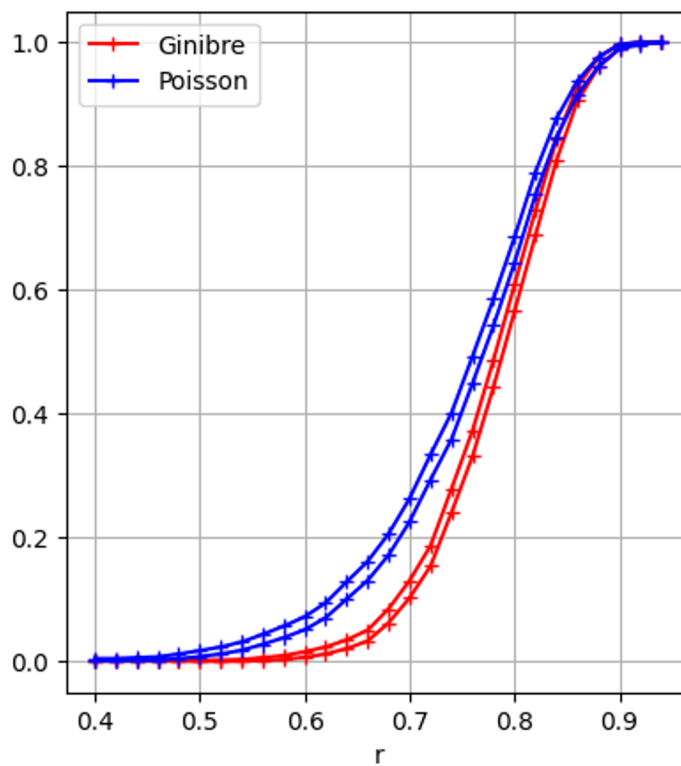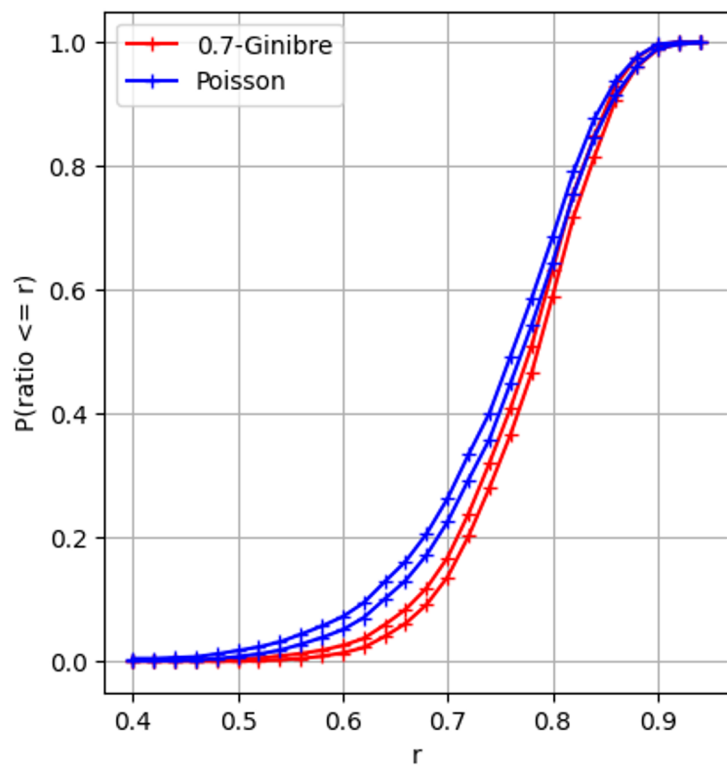
The limitation of the statistical approach using only the central cell is made visible by the presence of some overlap between the confidence intervals of the Poisson process and that of the 0.7-Ginibre process. Consequently, in specific cases, it may not be possible to determine the true nature of some processes based on this statistical test. On the other hand, if we average the ratio of the five most central cells (the cells whose centers are the closest to the origin), there is no longer an overlap among the various curves.

This approach shows that the chosen ratio variable represents a good repulsion criterion. On the other hand, our objective is to decide for a single map which model is the most pertinent, and that cannot be done by a frequentist approach. This is what motivated us to use a ML method.

Machine Learning approach

In this approach, we will use the same circular domain with N points as in the statistical approach. Since the repulsion is not sensitive to scaling, we normalize the radius to $R = \sqrt{N}$. This is due to the fact that a cloud drawn from a Ginibre point process of intensity 1 with N points occupies roughly a disk with this radius. We begin by generating the data of the Ginibre process, the 0.7-Ginibre process and the poisson process on which we will train the classification model, which is a Logistic Regression Classifier. Using only the central cell (respectively the five most central cells), the initial variables in our database consist of the surface and perimeter of the central cell (respectively surfaces and perimeters of the five central cells) of each generated sample, along with a binary variable that takes the value 1 if the process is repulsive and 0 otherwise. Subsequently, we add the ratio variable $\frac{4\pi S}{P^2}$ of the central cell (respectively the five ratios of the five central cells) to provide the classification model with additional information on which to base its predictions.

▶ Show code cell source

As the generation of data requires considerable execution time, we will prepare the data locally (using the previous code), by generating, for each classification, a sample of 5000 observations (2500 repulsive and 2500 non-repulsive) of $N = 50$ points. Then we read them directly as follows:

```
beta_ginibre_data_1 = pd.read_csv('data/beta_ginibre_1cell.csv', sep=',')
ginibre_data_1 = pd.read_csv('data/ginibre_1cell.csv', sep=',')

beta_ginibre_data_2 = pd.read_csv('data/beta_ginibre_5cells.csv', sep=',')
ginibre_data_2 = pd.read_csv('data/ginibre_5cells.csv', sep=',')
```

Skip to main content

Here is an example of the data created of configurations of 0.7-Ginibre and poisson processes with one cell:

▶ Show code cell source

	S1	P1	R1	process
0	3.282609	6.851499	0.878735	1
1	3.229371	7.052515	0.815905	0
2	5.542461	9.335383	0.799187	1
3	3.409986	8.318448	0.619267	0
4	4.336329	8.748695	0.711944	1

Before starting the model's training using the data generated, we will scale the data in order to avoir the concern of overfitting. Here is an example of the scaling of the previous data:

▶ Show code cell source

	S1	P1	R1	process
0	-0.290349	-0.584119	1.447490	1
1	-0.326183	-0.439422	0.661112	0
2	1.230698	1.203856	0.451857	1
3	-0.204615	0.471835	-1.800031	0
4	0.418882	0.781540	-0.640083	1

Now that the data is gathered, we will train the Logistic Regression model using the baseline model, i.e. all the hyperparameters' values are taken as defaults, (a grid search can be used later in order to select the optimal hyperparameters). Other classification models (Random Forest, Support Vector Machine and XGBoost) have been tested but did not yield more significant results than the chosen classifier. We split each data to a train data and test data in order to see the model's accuracy before testing it on the cartoradio data.
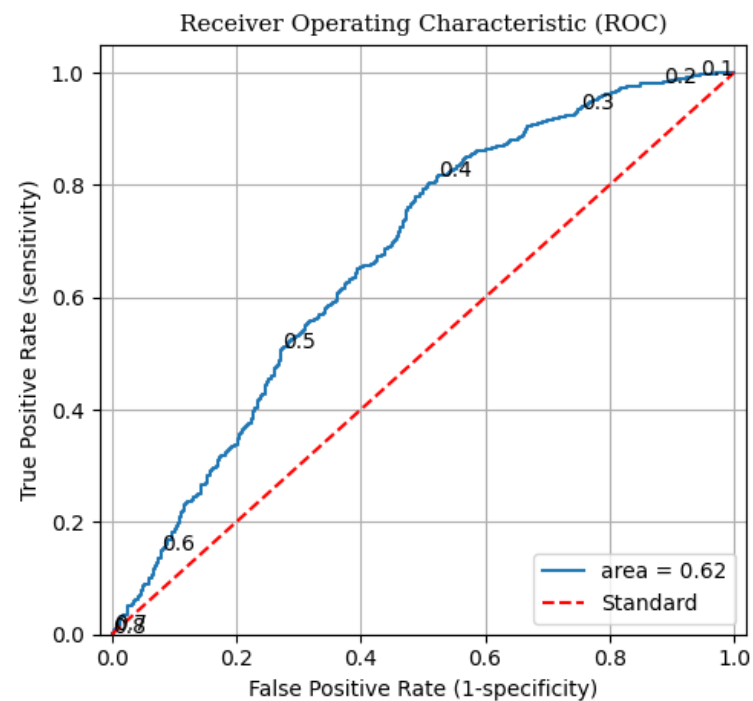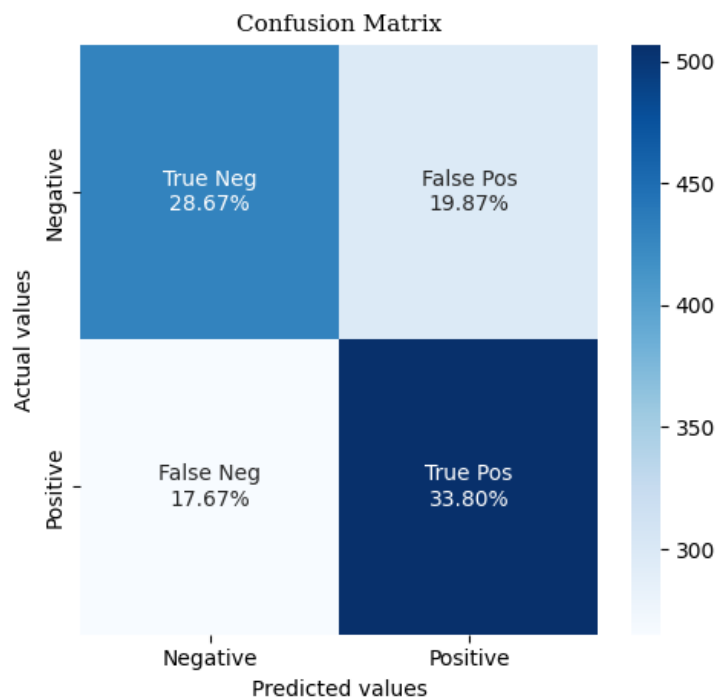
▶ Show code cell source

Here are the results of the classification using each data:

```
            precision    recall  f1-score   support

        0       0.62      0.59      0.60       728
        1       0.63      0.66      0.64       772

 accuracy                          0.62      1500
macro avg       0.62      0.62      0.62      1500
weighted avg    0.62      0.62      0.62      1500
```
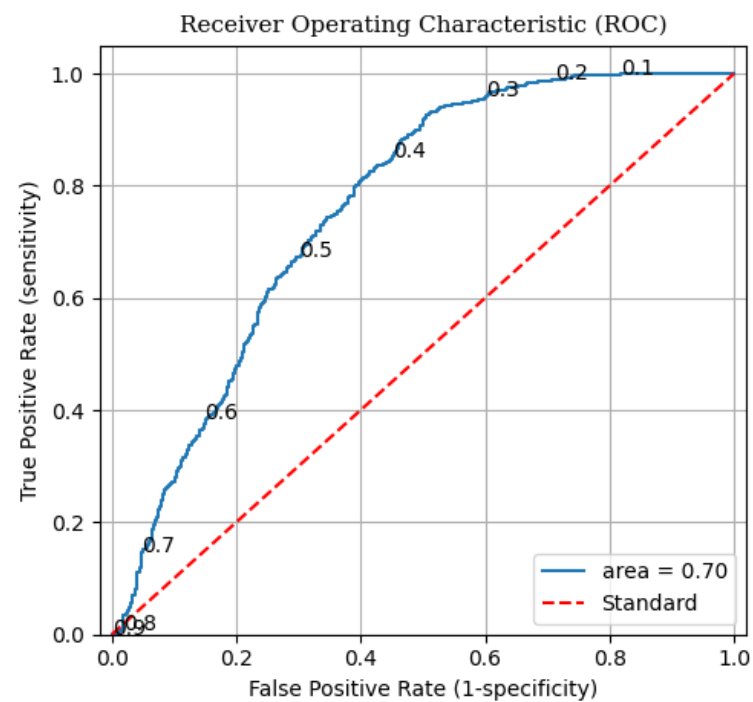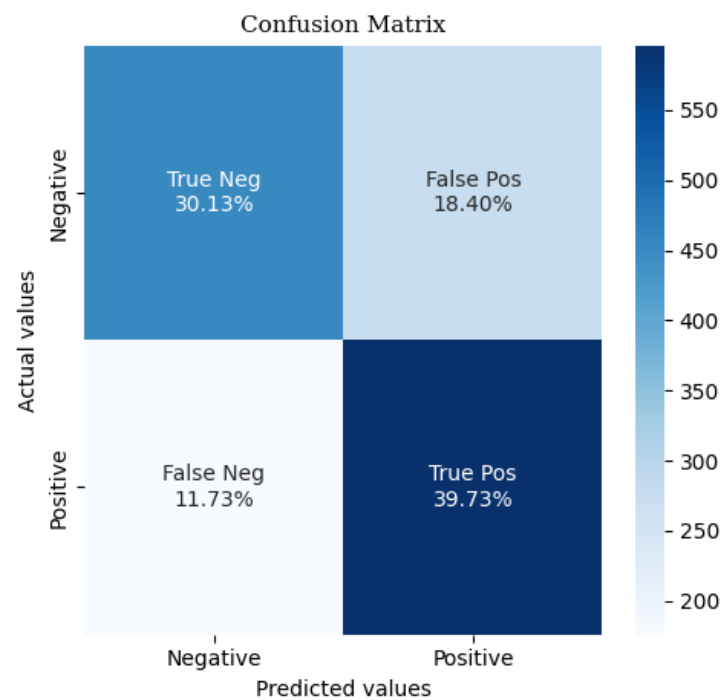

Confusion Matrix / Receiver Operating Characteristic (ROC)

- Ginibre Vs Poisson using the central cell:

```
            precision    recall  f1-score   support

        0       0.72      0.62      0.67       728
        1       0.68      0.77      0.73       772

 accuracy                          0.70      1500
macro avg       0.70      0.70      0.70      1500
weighted avg    0.70      0.70      0.70      1500
```
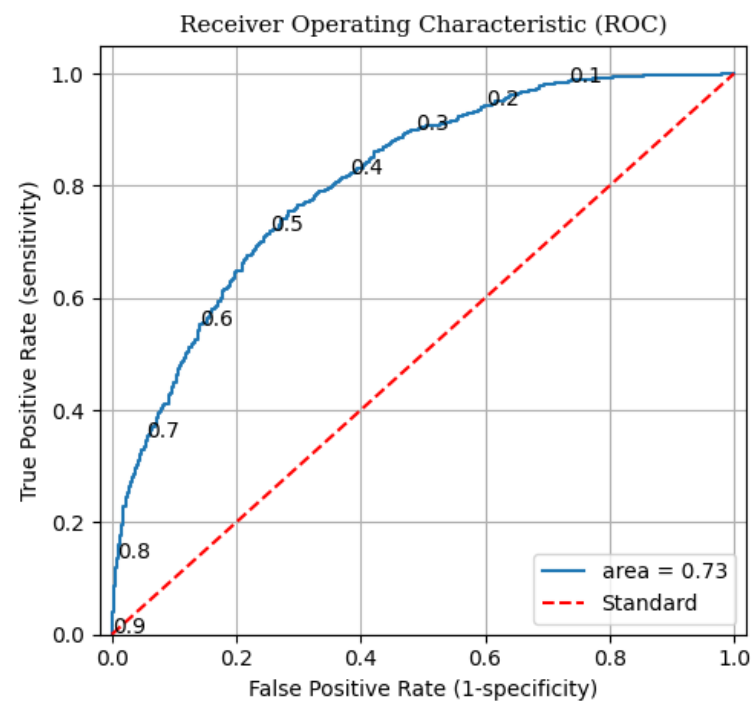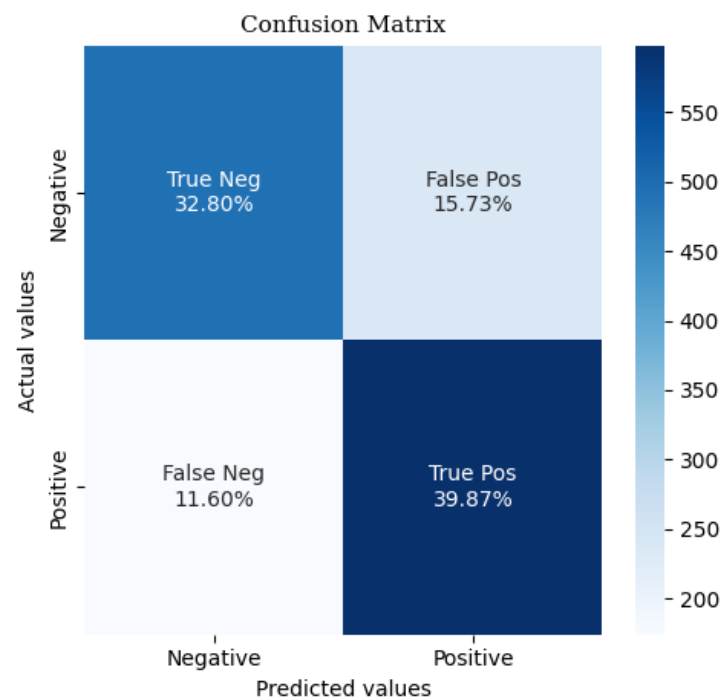
Skip to main content

Confusion Matrix / Receiver Operating Characteristic (ROC)

- 0.7-Ginibre Vs Poisson using the five central cells:

▶ Show code cell source

```
              precision    recall  f1-score   support

           0       0.74      0.68      0.71       728
           1       0.72      0.77      0.74       772

    accuracy                           0.73      1500
   macro avg       0.73      0.73      0.73      1500
weighted avg       0.73      0.73      0.73      1500
```
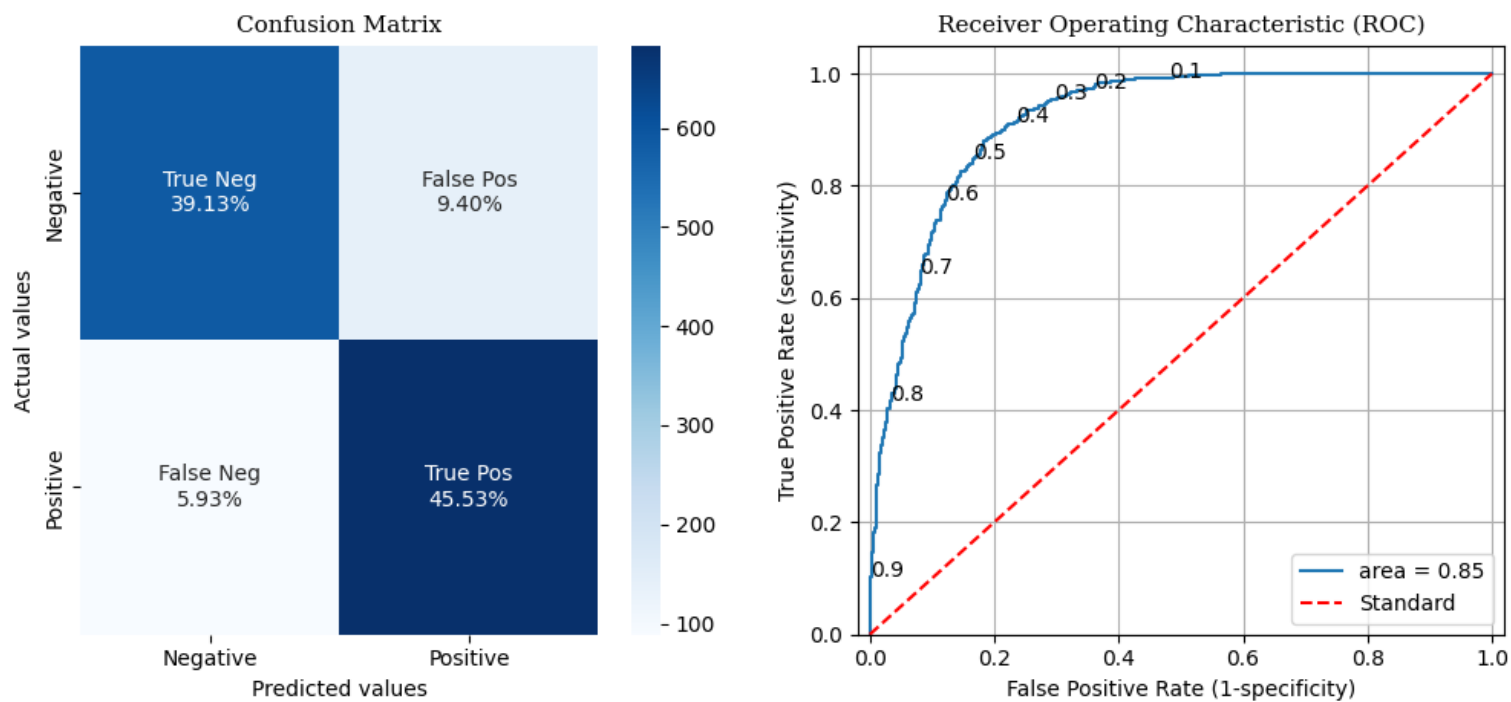
- Ginibre Vs Poisson using the five central cells:

▶ Show code cell source

```
              precision    recall  f1-score   support

           0       0.87      0.81      0.84       728
           1       0.83      0.88      0.86       772

    accuracy                           0.85      1500
   macro avg       0.85      0.85      0.85      1500
weighted avg       0.85      0.85      0.85      1500
```

We can notice that our model's accuracy when using the central cell is approximately 70% for the Ginibre and Poisson processes classification. However, when considering the first five central cells, we achieve an accuracy of 85%, a result consistent with our statistical approach. This is because with the five cells, the model has access to more information about the nature of the sample, increasing the likelihood of successful sample classification by taking into account the surface areas and perimeters of the first five central cells.

Cartoradio data Tests

Cartoradio data is a set of configurations of some mobile phone base stations in Paris. The goal is to decide from the classification model already used, whether the configuration do present some repulsion.

The initial data (positions of the antennas) cover a large area of the city of Paris (see Fig. 4(right)) With a real dataset, we often encounter the problem of heterogeneity between the different parts of the configurations since they depend on the topology of the space in which the antennas are placed.

To cope with this problem, we extract from each configuration a representative sample similar to the type of training data so that the tests make sense. Fig. 4 (left) shows a sample extracted from a given configuration.
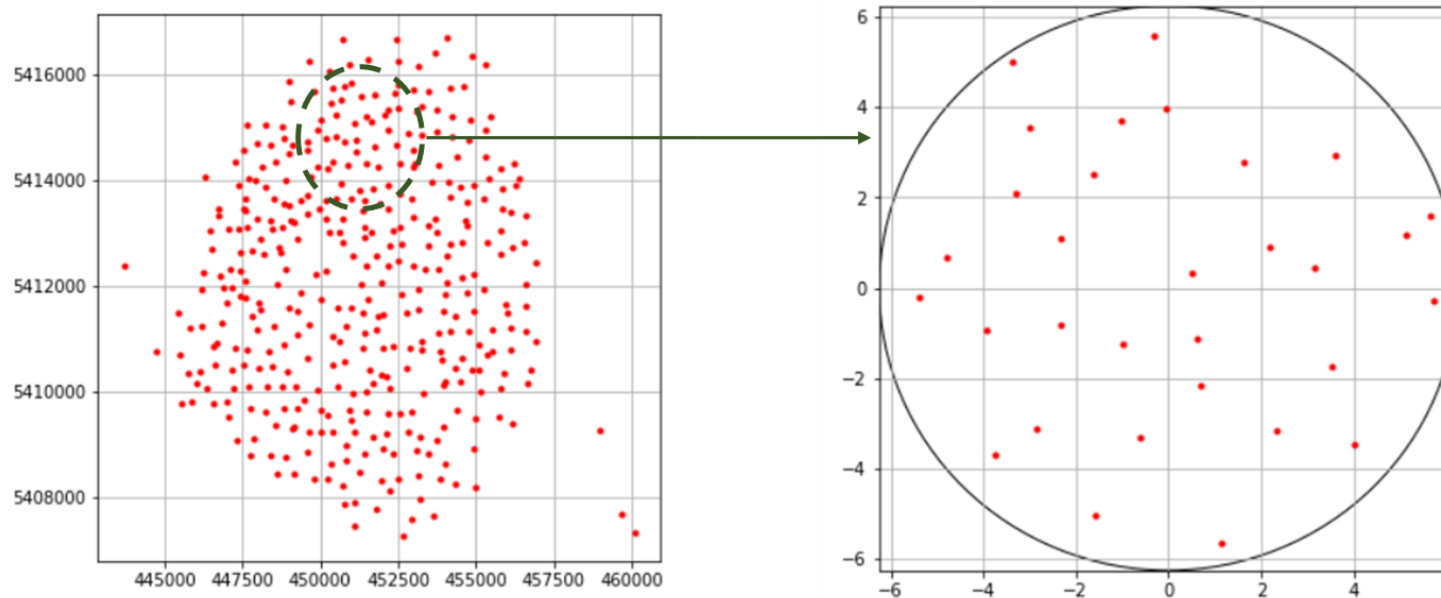
Skip to main content

Fig. 4 On the left, Initial cartoradio configuration. On the right, Sample extracted from it and scaled.

In the following, we read the cartoradio data directly from the "cartoradio_data.csv" file, then we add the variables that represent the ratio of the first five central cell and finally we scale the data. Here is the final (scaled) data on which we will test our model.

▶ Show code cell source

	S1	P1	R1	S2	P2	R2	S3	P3	R3	S4
0	0.666867	0.552371	0.771643	-1.320617	-0.826125	-0.776361	-0.096326	-0.168613	0.404993	-0.669259
1	-0.671167	-0.539183	-0.749317	0.190101	0.256283	-0.237297	0.631242	0.372228	1.268942	-1.348825
2	-0.217555	-0.375224	0.929404	0.020793	-0.495321	1.305963	-1.264007	-1.425377	-0.384128	0.778718
3	0.022384	0.160723	-0.473342	0.718430	0.472801	0.225675	-0.191550	-0.088225	-0.237366	0.982122
4	2.098913	1.686911	1.622808	1.381928	1.424882	-0.729521	1.978979	1.785494	0.952267	-0.502795

Here are the results of the tests on the cartoradio data using each model already trained, showing the classification value and its probability for each observation.

Skip to main content

```
Classification results: [0 0 1 0 0 0 0 1 1 1]
Classification probabilities:
 [[0.50899884 0.49100116]
 [0.50682787 0.49317213]
 [0.32945666 0.67054334]
 [0.56881892 0.43118108]
 [0.70288621 0.29711379]
 [0.82576939 0.17423061]
 [0.59513612 0.40486388]
 [0.48556473 0.51443527]
 [0.26548525 0.73451475]
 [0.32048159 0.67951841]]
```

- Ginibre Vs Poisson using the central cell:

```
Classification results: [0 0 1 0 0 0 0 1 1 1]
Classification probabilities:
 [[0.53921238 0.46078762]
 [0.51718021 0.48281979]
 [0.21820399 0.78179601]
 [0.63275798 0.36724202]
 [0.86608324 0.13391676]
 [0.94844677 0.05155323]
 [0.67866702 0.32133298]
 [0.47846419 0.52153581]
 [0.13668011 0.86331989]
 [0.20397696 0.79602304]]
```

It can be noted that the classification results using only the central cell are not significant enough. This is largely due to the low accuracy of the model used with the central cell, which is normal since the data does not contain enough variables for the model's training.

- 0.7-Ginibre Vs Poisson using the five central cell:

```
Classification results: [1 0 1 0 0 0 1 1 1 1]
Classification probabilities:
 [[0.09840518 0.90159482]
 [0.51964363 0.48035637]
```

```
[0.78607822 0.21392178]
[0.84358136 0.15641864]
[0.26659285 0.73340715]
[0.23202252 0.76797748]
[0.17172631 0.82827369]
[0.07726097 0.92273903]]
```

- Ginibre Vs Poisson using the five central cell:

▶ Show code cell source

```
Classification results: [1 0 1 0 0 0 1 1 1 1]
Classification probabilities:
 [[0.02247336 0.97752664]
 [0.51484798 0.48515202]
 [0.22732894 0.77267106]
 [0.8325097  0.1674903 ]
 [0.97662433 0.02337567]
 [0.96091364 0.03908636]
 [0.27231771 0.72768229]
 [0.09138561 0.90861439]
 [0.0856342  0.9143658 ]
 [0.01795407 0.98204593]]
```

In contrast, the results found using the five central cells are much better, the majority of configurations are classified as repulsive. Regarding the configurations classified as non-repulsive by our model, we can say that this comes down to one of the following two reasons:

- As long as we are dealing with real data, these samples may be a non-repulsive ones and the results are actually coherent.
- It is sure that the accuracy of our models is high, but we may have some classification errors, which means that even if the configuration is repulsive, the model decides that it is not.

In this paper it has been shown numerically (based on the theoretical results in [6]) that Voronoi cells represent an effective means for determining the nature of repulsion of a configuration (repulsive or not), and this by creating a database of various configurations and extracting the areas and perimeters of the Voronoi cells in order to use them as input to the classification model described earlier.

Once the model is trained and tested on the data created, it is tested after that on real data, which are the positions of a mobile phone base stations in Paris. Visually, we can easily say that these configurations are repulsive, which we have confirmed for the majority of these configurations by testing them by the previously trained model, especially the one classifying Ginibre and poisson processes using the first five central cells.

[1]   F. Baccelli and B. Błaszczyszyn. Stochastic Geometry and Wireless Networks: Volume I Theory. *Foundations and Trends® in Networking*, 3(3–4):249–449, 2008. doi:10.1561/1300000006.

[2]   ANFR. Https://www.cartoradio.fr. URL: https://www.cartoradio.fr.

[3]   Na Deng, Wuyang Zhou, and Martin Haenggi. The ginibre point process as a model for wireless networks with repulsion. *IEEE Transactions on Wireless Communications*, 2015. arXiv:1401.3677v1.

[4]   Naoto Miyoshi and Tomoyuki Shirai. Spatial modeling and analysis of cellular networks using the Ginibre point process: a tutorial. *IEICE Transactions on Communications*, E99.B(11):2247–2255, 2016. doi:10.1587/transcom.2016nei0001.

[5]   Jean-Sébastien Gomez, Aurélien Vasseur, Anaïs Vergne, Philippe Martins, Laurent Decreusefond, and Wei Chen. A case study on regularity in cellular network deployment. *IEEE Wireless Communications Letters, IEEE, 2015, pp.4*, 2015. arXiv:1505.06073v1.

[6](1,2,3,4,5) André Goldman. The Palm measure and the Voronoi tessellation for the Ginibre process. *The Annals of Applied Probability*, 20(1):90–128, February 2010. arXiv:math/0610243, doi:10.1214/09-AAP620.

[7](1,2)   D. B. Taylor, H. S. Dhillon, T. D. Novlan, and J. G. Andrews. Pairwise interaction processes for modeling cellular network topology. In *Globecom*. 2012. doi:10.1109/glocom.2012.6503831.

[8](1,2,3) Laurent Decreusefond and Guillaume Moroz. Optimal transport between determinantal point processes and application to fast simulation. *Mod. Stoch. Theory Appl.*, 8(2):209–237, 2021.

[9]   L. Decreusefond and A. Vasseur. Asymptotics of superposition of point processes. In *Geometry of Science Information*. Palaiseau, October 2015.

[10]   D. J. Daley and D. Vere-Jones. *An introduction to the theory of point processes. Vol. I*. Probability and its Applications (New York). Springer-Verlag, New York, second edition, 2003. ISBN 0-387-95541-0. Elementary theory and methods.

[11]   L. Decreusefond, I. Flint, and A. Vergne. A note on the simulation of the Ginibre point process. *Journal of Applied Probability*, 52(04):1003–1012, dec 2015. arXiv:1310.0800v2, doi:10.1239/jap/1450802749.

[12]   J. B. Hough, M. Krishnapur, Y. Peres, and B. Virág. Determinantal processes and independence. *Probability Surveys*, 3:206–229 (electronic), 2006.

[13]   F. Baccelli and B. Błaszczyszyn. Stochastic Geometry and Wireless Networks: Volume II Applications. *Foundations and Trends® in Networking*, 4(1–2):1–312, 2009. doi:10.1561/1300000026.