

Application of Spatial Correlation Functions to Radio Propagation Modelling

Panwei Hu

Bachelor Thesis

July 9, 2020

Examiners

Prof. Dr. Petri Mähönen
Prof. Dr.-Ing. Mariana Petrova

Supervisors

Prof. Dr. Petri Mähönen
Dr. Ljiljana Simić

Institute for Networked Systems
RWTH Aachen University



The present work was submitted to the Institute for Networked Systems

Application of Spatial Correlation Functions to Radio Propagation Modelling

Bachelor Thesis

presented by
Panwei Hu

Prof. Dr. Petri Mähönen
Prof. Dr.-Ing. Mariana Petrova

Aachen, July 9, 2020

(Panwei Hu)

ACKNOWLEDGEMENTS

I am extremely grateful to Prof. Petri Mähönen for his constructive advice and help on my thesis. He recommended that I use the *Healpix* software, which is the building block of the implementation of the calculation. He has answered several important questions regarding the two-point correlation function and the angular power spectrum for me during my work on the thesis. His amiability and kindness and tolerance of my ignorance has motivated me a lot to keep working hard.

I would also like to extend my deep gratitude to Dr. Ljiljana Simić, who introduced me this interesting topic. I want to thank her for her advice on the analysis of the experiment data, which brought me light to better understand the problem and analyze it correctly. I also want to thank her for the excellent coordination of my study condition.

I would like to express my appreciation to Aleksandar Ichkov M.Sc., who carried out the experiment and collected the data. He has explained to me the scheme of the experiment clearly and soundly. He has also preprocessed the data neatly which eased a lot of my work for the analysis.

I am obliged to Gero Schmidt-Kärst for his technical support. His configuration of the working environment has saved me a considerable amount of time and enabled me to work on my thesis directly.

I would like to thank Juan Bolanos and Adel Kanaan, with whom I studied together before the pandemic lockdown. Their concentration and hardworking has also given a boost to my study.

I am deeply indebted to my parents for their consistent support and unconditional love. I would especially like to thank my mother. Her love and care have always been a great consolation for me. Her persistence and fortitude during a hard time have also taught me to have a strong will and never give up. I would also like to give special thanks to my cousin Jingjing Ren, who takes great care of my parents when I cannot due to my studies abroad.

I must also thank my friend Ningyuan Chen for his companionship along with my study in Germany. His expertise in computer science has broadened my mind in problem-solving and his passion for studying and his strive for achieving the best has always propelled me to keep moving.

I am also grateful to Jason Jin, whose general advice on study and life, and whose help for the texting in \LaTeX has been of enormous importance to me.

CONTENTS

ACKNOWLEDGEMENTS	II
CONTENTS	III
ABSTRACT	V
1 INTRODUCTION	1
1.1 ANGULAR TWO POINT CORRELATION FUNCTION	1
1.2 APPLICATION TO RADIO PROPAGATION	2
2 IMPLEMENTATION OF ESTIMATOR	3
2.1 SETTING UP	3
2.2 RUN-TIME ANALYSIS	6
2.3 BIAS OF THE ESTIMATORS	9
3 TESTING ESTIMATOR WITH GENERATED DATA	10
3.1 RANDOMLY PICKING PIXEL CELLS	10
3.2 GENERATED HOMOGENEOUS <i>Poisson</i> POINT PROCESS	17
3.2.1 MATHEMATICAL DERIVATION OF SIMULATION OF POISSON POINT PROCESS ON A SPHERE	17
3.2.2 TESTING GENERATED POISSON POINT PROCESS ON <i>Healpix</i> MAP	18
3.3 GENERATED <i>Matérn</i> PROCESS	19
3.3.1 DERIVATION OF THE TRANSFORMATION MATRIX	20
3.3.2 TESTING GENERATED MATÉRN POINT PROCESS ON <i>Healpix</i> MAP	21
4 APPLICATION OF 2PCF ON EXPERIMENTAL DATA	23
5 INTERPRETATION USING ANGULAR POWER SPECTRUM	30
6 CONCLUSION	35
6.1 LIMITATIONS	35
6.2 OUTLOOK	35
A ABBREVIATIONS	36
BIBLIOGRAPHY	37

CONTENTS

IV

DECLARATION

37

ABSTRACT

This thesis is about applying the principle of angular two-point correlation function (2pCF) introduced in the cosmology to analyze the large scale structure of the universe to study the data from an experiment channel propagation studies. The goal is to find some optimal angles of achieving large signal strength.

The thesis will first illustrate the building blocks of calculating the 2pCF and analyze the running time. Next, the thesis will first propose methods to simulate point processes on a sphere and then present test results of several generated data. Then the thesis will show the result of 2pCF applied to the experiment data. Finally the thesis will demonstrate briefly how the test results could be interpreted with the help of the angular power spectrum.

INTRODUCTION

1.1 ANGULAR TWO POINT CORRELATION FUNCTION

The two point angular correlation function $w(\theta)$, introduced in [1], describes the excess probability of finding a neighbor in a region $d\Omega_2$ at an angular distance θ away from an element in $d\Omega_1$, compared with a randomly distributed point process.

Mathematically, the function can be defined as following:

$$dP = \zeta_0 d\Omega_1 \zeta_0 d\Omega_2 (1 + w(\theta)), \quad (1.1)$$

where ζ_0 denotes the mean points density and $d\Omega_i$, $i \in \{1, 2\}$ denotes the small area on a sphere. In particular, if $w(\theta) = 0$, the probability degenerates to that of a *Poisson* point process. The equivalent formulation of 2pCF is the normalized autocorrelation of the density function

$$w(\theta) = \frac{\mathbb{E}[(\zeta(x + \theta) - \zeta_0)(\zeta(x) - \zeta_0)]}{\zeta_0^2}, \quad (1.2)$$

where $\mathbb{E}[X]$ stands for the expectation value of the random variable X , $X \in \mathbb{R}^3$. This formulation tends to be useful when we try to use angular power spectrum to interpret test results, as will be illustrated in Section 5.

In this thesis, four well-known estimators will be used, which were introduced by Peebles & Hauser

$$w_1(\theta) = \frac{2DD(\theta)\Omega}{N^2\mathbb{E}[d\Omega]} - 1, \quad (1.3)$$

Davis & Peebles (1983):

$$w_2(\theta) = \frac{N_r}{N-1} \frac{DD(\theta)}{DR(\theta)} - 1, \quad (1.4)$$

Hamilton (1993):

$$w_3(\theta) = \frac{N_r(N_r-1)}{N(N-1)} \frac{DD(\theta)}{RR(\theta)} - \frac{N_r-1}{N} \frac{DR(\theta)}{RR(\theta)}, \quad (1.5)$$

and Landy & Szalay (1993)

$$w_4(\theta) = \frac{4N_r}{(N-1)(N_r-1)} \frac{DD(\theta) \cdot RR(\theta)}{(DR(\theta))^2} - 1. \quad (1.6)$$

Here $DD(\theta)$ stands for the number of pairs with angular distance θ , $RR(\theta)$, the corresponding pairs in a randomly generated set, $DR(\theta)$ the cross number of pairs with

angular distance θ , N the number of points in the given point process and N_r the number of points of a randomly generated point process.

For the following section, the thesis will denote these estimators as PH, DP, Hamilton, and LS. As we shall see, the behaviors of these four estimators on the tested data are almost the same, but in general, the LS estimator tends to have the least variance. For completeness, I have used all the four estimators to illustrate results. In practice, it may suffice simply to choose one estimator.

Before discussing further the analysis, it would be remiss not to mention the notation used in this thesis. As the conventional notation of the variable in the 2pCF is θ , I will not introduce unnecessary complications by using another variable. However, since the usual spherical coordinate also uses θ to denote the polar angle, I will use ϑ to denote the polar angle and θ to represent the variable meant in the 2pCF when the two variables appear in the same formula, see for example (3.4). When there is no coexistence of these two variables, I will use the θ to denote either the variable in the 2pCF or the polar angle, whose meaning will be clear from the context.

1.2 APPLICATION TO RADIO PROPAGATION

While the details of the actual data used and how it is generated are given in Chapter 4, we will give in this section general justification and view on the topic.

It is well known that radio signals propagate following Maxwells laws. The reflection, refraction and diffraction affect received signal power in distinct location from the source. Thus it is not signal attenuation related issue. However, many empirical models treat received signal power as a function of distance between transmitter and receiver. This approximation has been sufficient for lower frequencies. However, very high frequency so called millimeter-wave systems enforce strong directionality and signal can be received only from certian angle of arrival(AoA) directions. This leads to a natural question: if there are correlations among thes AoA-directions. Thus, the application of angular correlation functions for radio propagation modelling becomes an interesting research topic.

IMPLEMENTATION OF ESTIMATOR

2.1 SETTING UP

The building block of the calculation of the 2pCF is based on the software *Healpix* [2]. One advantage of the software *Healpix* is that the sphere is divided into equal areas, which makes it ideal for being the bin to calculate 2pCF. Another advantage is the provided APIs for calculating the angular power spectrum, which will be used later for the analysis. Detailed information can be found in [2, 3].

I have derived a subclass called *Healpix_Correlation* from the provided class to suits the calculation of 2pCF, which is called *Healpix_Map*. The idea is to interpret each pixel value as the count of points in that cell, or later the binary value 0 or 1 to indicate that at this region the signal strength is above a certain threshold or not. There are two enumerating schemes of the pixels. Since we do not need to complicate the issues unnecessarily, the calculation will use the RING scheme, which is counting sequentially from north to south.

There are several technical issues about this class that need to be considered: The first issue relates to the question about how to calculate the bin width. Since each pixel cell is of equal area, the area of each small region is thus approximated as $Rd\theta Rd\varphi$. As the southern hemisphere is symmetric to the northern hemisphere, only the northern hemisphere needs to be studied.

According to the set up of *Healpix_Map*, there are $2N_{side}$ rings from north pole to the equator, where the parameter N_{side} is a natural number, indicating the number of division along the side of a base pixel cell. , so I approximate the bin width as $\frac{\pi}{2 \cdot 2N_{side}}$. It may be a suboptimal estimate of the bin width when the pixel is large, but the error is of order of square root of each pixel cell area, so as the resolution increases, the error becomes negligible.

The next issue relates to the counting of the pixels. For simplicity, we assume the points lie on a unit sphere. The idea is that given a pixel p_0 with cartesian coordinate (x_0, y_0, z_0) , each point lying at an angle of θ to p_0 , satisfies the equation:

$$x \cdot x_0 + y \cdot y_0 + z \cdot z_0 = \cos \theta. \quad (2.1)$$

There are several rings that the θ -disc of p_0 intersects. To retrieve all the points lying in the disc, we only need to run through all the possible rings. Here comes another property of the *Healpix* in to play, which is the property of isolatitude, simplifying the calculation of z coordinate.

Setting the azimuthal angle of p_0 to be 0 degrees, (2.1) simplifies to

$$x \cdot \sin \theta_0 + z \cdot \cos \theta_0 = \cos \theta. \quad (2.2)$$

The value of z could be given by the function of *pix2ang* of the *Healpix_Map*, so solving for x :

$$x = \frac{\cos \theta - z \cdot \cos \theta_0}{\sin \theta_0}.$$

We then add the azimuthal angle ϕ_0 to the point to obtain the correct coordinate.

Third issue relates to the calculating of the neighbors of the random generated points. The naive way is to calculate the $RR(\theta)$ and $DR(\theta)$ for each realization of a random map, and then calculate the average. However, simply doing the calculation in this way will lead to dramatic bias, as shown by the following Figure 2.1.

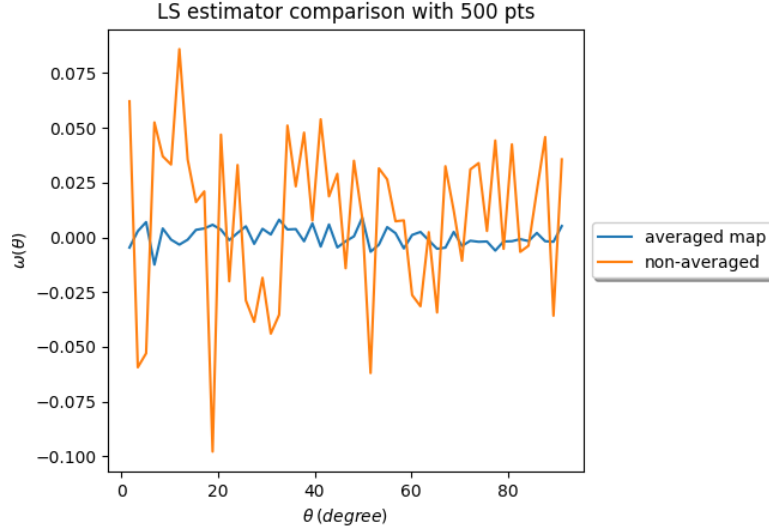


FIGURE 2.1: Comparison between averaged and non-averaged map.

Figure 2.1 shows the LS estimator on a randomly generated *Healpix_Correlation* map, with 500 points. We could see how the result shown in Figure 2.1 is biased from zero. The reason lies in the analysis of the variance. Using the notation introduced by Landy and Szalay [4]. The second moment of $DR(\theta)$ is given by:

$$\mathbb{E}[DR(\theta) \cdot DR(\theta)] = \mathbb{E}\left[\sum_{i \neq j} v_i \rho_j \Theta_{ij}^\theta \sum_{k \neq l} v_k \rho_l \Theta_{kl}^\theta\right]. \quad (2.3)$$

Multiplying out (2.3) leads to several different terms.

$$\begin{aligned}
\sum_{i \neq j} v_i \rho_j \Theta_{ij}^\theta \sum_{k \neq l} v_k \rho_l \Theta_{kl}^\theta = & \underbrace{\sum_{i,j,k,l \text{ distinct}} v_i \rho_j v_k \rho_l \Theta_{ij}^\theta \Theta_{kl}^\theta}_1 + \\
& \underbrace{\sum_{i,j,l \text{ distinct}} v_i \rho_j v_i \rho_l \Theta_{ij}^\theta \Theta_{il}^\theta}_2 + \underbrace{\sum_{i,j,l \text{ distinct}} v_i \rho_j v_l \rho_j \Theta_{ij}^\theta \Theta_{il}^\theta}_3 + \\
& \underbrace{\sum_{i,j \text{ distinct}} v_i \rho_j \Theta_{i,j}^\theta}_4 + \underbrace{\sum_{i,j \text{ distinct}} v_i \rho_j v_j \rho_i \Theta_{ij}^\theta}_5 + \\
& \underbrace{\sum_{i,j,l \text{ distinct}} v_i \rho_j v_j \rho_l \Theta_{ij}^\theta \Theta_{jl}^\theta}_6 + \underbrace{\sum_{i,j,l \text{ distinct}} v_i \rho_j v_l \rho_i \Theta_{ij}^\theta \Theta_{il}^\theta}_7 \quad (2.4)
\end{aligned}$$

1 :

$$\frac{N}{K} \cdot \frac{N-1}{K-1} \cdot \frac{N_r}{K} \cdot \frac{N_r}{K} \cdot K(K-1)(K-2)(K-3) \cdot G_q(\theta)$$

2 :

$$\frac{N}{K} \cdot \frac{N_r}{K} \cdot \frac{N-1}{K} \cdot \frac{N_r}{K} \cdot K(K-1)(K-2) \cdot G_t(\theta) \xrightarrow{K \rightarrow \infty} 0$$

3 :

$$\frac{N}{K} \cdot \frac{N-1}{K-1} \cdot \frac{N_r}{K} \cdot \frac{N_r}{K} \cdot K(K-1)(K-2) \cdot G_t(\theta) \xrightarrow{K \rightarrow \infty} 0$$

4:

$$\frac{N}{K} \cdot \frac{N_r}{K} \cdot K(K-1) \cdot G_p(\theta)$$

5:

$$\frac{N}{K} \cdot \frac{N_r}{K} \cdot \frac{N-1}{K-1} \cdot \frac{N_r}{K} \cdot K(K-1) \cdot G_p(\theta) \xrightarrow{K \rightarrow \infty} 0$$

6:

$$\frac{N}{K} \cdot \frac{N_r}{K} \cdot \frac{N-1}{K-1} \cdot \frac{N_r}{K} \cdot K(K-1)(K-2) \cdot G_t(\theta) \xrightarrow{K \rightarrow \infty} 0$$

7:

$$\frac{N}{K} \cdot \frac{N_r}{K} \cdot \frac{N-1}{K-1} \cdot \frac{N_r}{K} \cdot K(K-1)(K-2) G_t(\theta) \xrightarrow{K \rightarrow \infty} 0$$

We see that to make terms go to zero, as we enhance the resolution (refine the bins), it is essential to achieve a higher order of K in the denominator than that of the numerator. The higher order of K in the denominator in the 2,3,5,6,7 is achieved because we have extra $\frac{N}{K}$ term in the factor, which is only possible if we average in the first step the generated map so that the points tend to lie more uniformly in the

region. Failing to do so, the extra terms $\frac{N_r}{K}$ will not appear. As a consequence, more terms contribute to the variance, leading to severe oscillation.

2.2 RUN-TIME ANALYSIS

The run-time of the calculation of 2pCF depends mainly on four variables, the number of points N , the bin width $d\theta$ the the chosen point counts of a randomly generated map N_r and the spanned region, which is characterized by a span angle.

I first set the span angle to be a constant $\frac{\pi}{2}$, which corresponds to a half-sphere¹. At the end of this section, I will talk about the effect of the span angle briefly.

Since the number of points is the primary variable in practice, it is worth analyzing the running time by setting the bin width as constant. The point count of the randomly generated map N_r tends to of less importance, as can be seen from Table 2.1, where the number of points in the map is set to 2000, and bin width 0.049, the parameter *deviation* denotes the maximum absolute value of the estimator from zero. We could observe that the number of points in a randomly generated map does not significantly change the running time or the precision of the calculation very much. Generally, random points are set to be equal with the points of the studied point process [i.e., actual data points]: a too large number will bring unnecessary work for the extra calculation, a too small number will lead to statistical uncertainty.

We can easily see that the running time is a linear function of N , since for each point, the calculation of the neighbors in the disc is mainly determined by the bin width of the map, i.e., the number of bins lie within that θ disc. The assumption is justified by the Figure 2.2, where the bin width is set to be 0.0061, and the random count is the same as N .

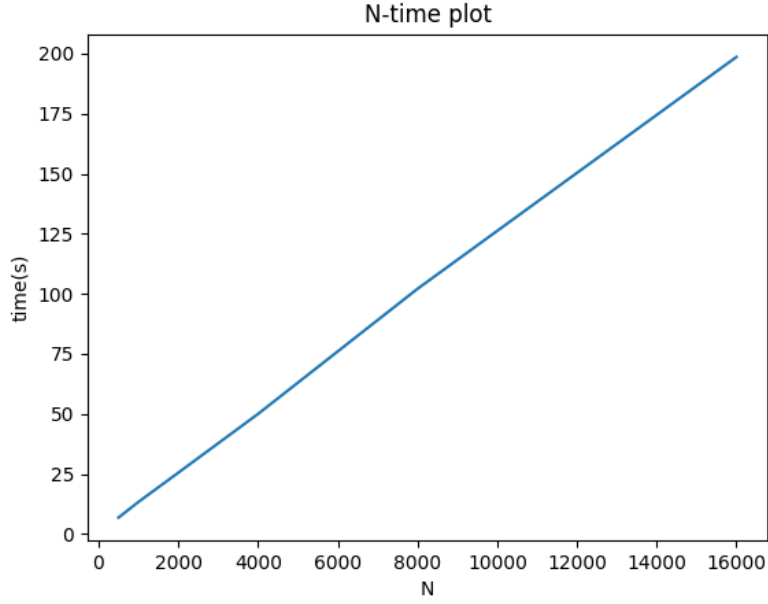
N	Ω	time(ms)	deviation	N_r
2000	0.0490874	78.4477	0.00549775	200
2000	0.0490874	83.6017	0.00411683	1000
2000	0.0490874	78.8419	0.00473122	2000
2000	0.0490874	75.706	0.00372029	4000
2000	0.0490874	86.3993	0.00401143	8000
2000	0.0490874	92.3003	0.00431473	16000

TABLE 2.1: Analysis of the effect of points of a randomly generated map on 2pCF.

Next step is to consider the impact of the bin width on the computational time. As indicated above, the running time is predominated by the counting process. As the number of bins is inversely proportional to the square of the bin width, we would expect the running time is inverse proportional to the square of the bin width. The assumption is justified by Figure 2.3, where the number of points in the map is set to be $N = 1000$.

The final step is to set both variables N and $d\theta$ free and also treat the span angle as a variable. I have chosen several combination and plot the resultant calculation time, as shown in Figure 2.4 The linearity of the dependency of calculation time to N and

¹A half-sphere is often well justified in the case of terrestrial wireless communications.

FIGURE 2.2: Run time with N as variable.

$\frac{1}{(d\theta)^2}$ is clear to be seen. We could also see that increasing the span angle will increase the calculation time, as expected. Since each latitude may contain different amount of bins, the direct relation between the span angle and the running time is hard to retrieve.

time(ms)	deviation	span angle
17.9229	0.00773891	0.785398
64.2565	4.44763e-05	1.5708
130.925	0.0101783	2.0944
245.183	0.124442	3.14159

TABLE 2.2: Impact of the span angle on the calculation time and precision, with $N = 1000, d\theta = 0.05$.

As for the impact of the span angle on the precision, the problem tends to be more complicated. There are two counteracting effects caused by an increasing span angle: increasing span angle with the same number of points lower the mean density, making the ensemble seem more discrete, leading to imprecision, on the other hand, the distribution of points tend to be more random, and the bias from zero should be smaller. The combination of the two counteracting effects can be seen from Figure 2.5 In conclusion, more detailed analysis needs to be performed to study the impact of the span angle on the calculation and precision.

In this section, we have seen how the number of points in a point process, the bin width, and the points of the randomly generated point process could influence the

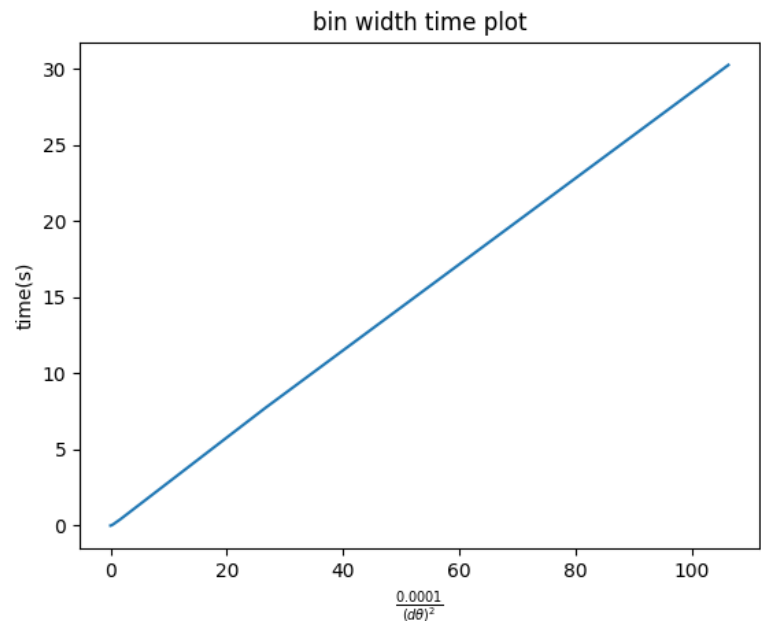


FIGURE 2.3: Run time analysis with $d\theta$ as variable, $N = 1000$.

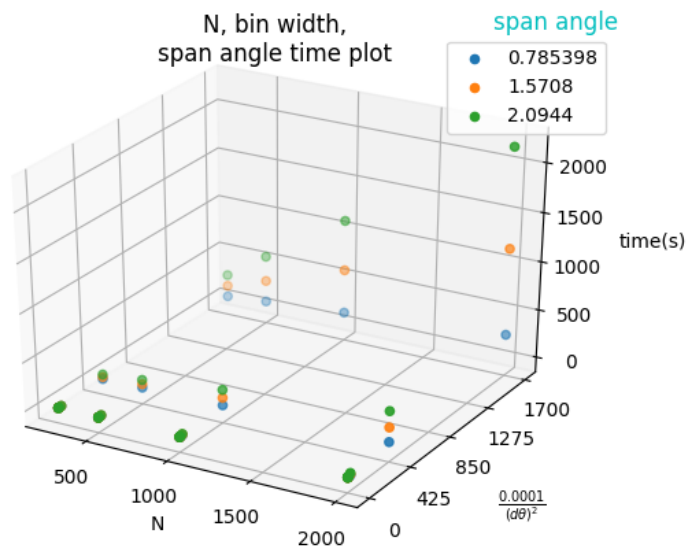
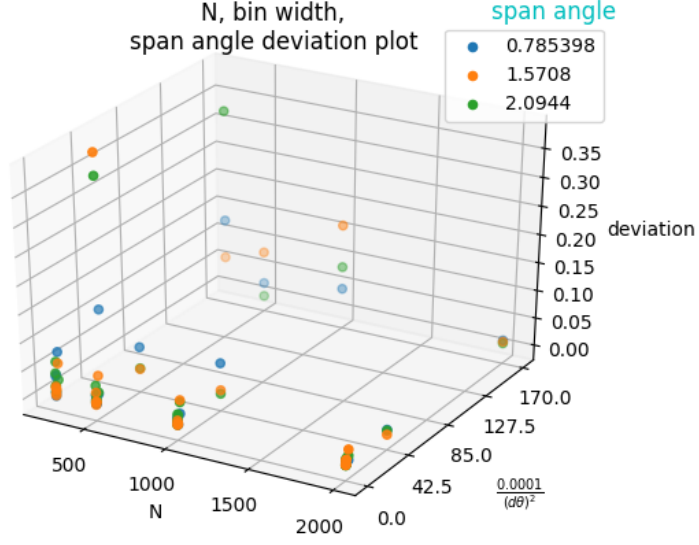


FIGURE 2.4: Run time with N , $d\theta$ and span angle as variable.

FIGURE 2.5: Deviation with N , $d\theta$ and span angle as variable.

run time and the precision of the calculation. In practice, the combination of the three variables needs to be adjusted to achieve the desired speed and precision.

2.3 BIAS OF THE ESTIMATORS

The expected values of each estimator have already been given by Landy and Szalay [4]. In particular, all the estimators are biased, at least by a factor $\frac{1}{1+w_\Omega}$, where $w_\Omega = \int w(\theta)G_p(\theta)d\theta$. However, this factor is uniform for all the angles, in particular, the relative order of the correlation function value at two different angles are not changed. Since we are interested in moving optimal angles to receive higher signal strength from given positions, the exact value of the correlation function is less important. In conclusion, we may be satisfied with the biased estimators to retrieve the useful information we need.

TESTING ESTIMATOR WITH GENERATED DATA

In this chapter, I will mainly discuss the test results of randomly generated *Healpix* map, the homogeneous *Poisson* point process, and the *Matérn* process. I will also illustrate the methods of how to simulate the three processes. The importance of these tests is to validate the code and method for actual research use. It is also worth noting here that despite the common usage of degrees in the engineering convention, I will use the radians for the mathematical derivation due to the simplicity of the usage of trigonometric functions. For the other part of this thesis, I will comply with the convention of using degrees.(e.g. the angle of the studied spanning region)

3.1 RANDOMLY PICKING PIXEL CELLS

The first test to perform is to calculate the 2pCF of a randomly generated *Healpix* map that spans a certain region of the sphere. The simulation is quite straightforward, since the pixel cells are enumerated sequentially, we can use a random number generator to generate a sequence of numbers. Each time a number is chosen, the corresponding value of that pixel cell will be increased by 1.

The expected result of the estimators should be a flat line near zero. The reason is that if points are from the *Poisson* process, then by allocating them into cells, the data points should also produce *Poisson*-like behaviors.

The first test is on the map which spans the whole sphere, i.e., there are no masks or constraints on where the points may lie. This is the first test of the correctness of the estimators, or we could also view it as a test on whether the assumption that randomly picked pixel cells from the *Healpix* map will have *Poisson* process like behaviors.

For the display of the map, I use the provided *Mollweide* view¹ from the *healpy* package. The visualization is shown in Figure 3.1. The four estimators of the 2pCF are shown in Figure 3.2.

The title of Figure 3.2 consists of several variables: *thetaBegin* denotes the starting polar angle of the region, *thetaEnd* the end polar angle, 1000 data points indicate the total number of points of the point process, random 1000 points indicate an averaged randomly generated map with 1000 points, the number is usually chosen to be equal to that of the studied point process, see Chapter 2. 10 times average denotes the randomly generated map will be averaged 10 times before doing further calculation, *dT* denotes the bin width and *data* indicates the underlying point process the estimators are estimating.

We could see all the four estimators are flat curve around 0, which complies with our expectations. Another interesting effect is that at the endpoints 0° and 180° , larger

¹The *Mollweide* projection is a projection scheme generally used for global maps and night sky. This projection is an equal-area, pseudocylindrical map projection.

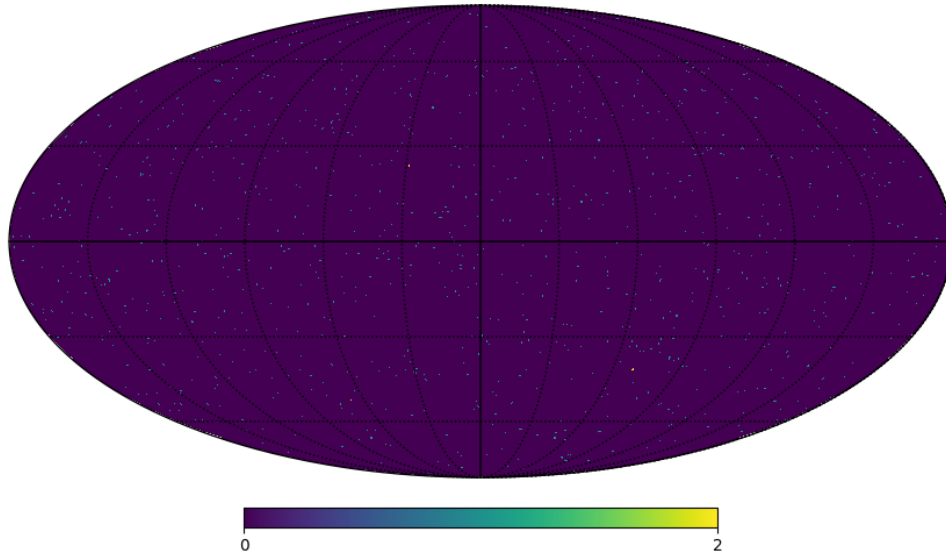


FIGURE 3.1: Randomly generated map with no constraints.

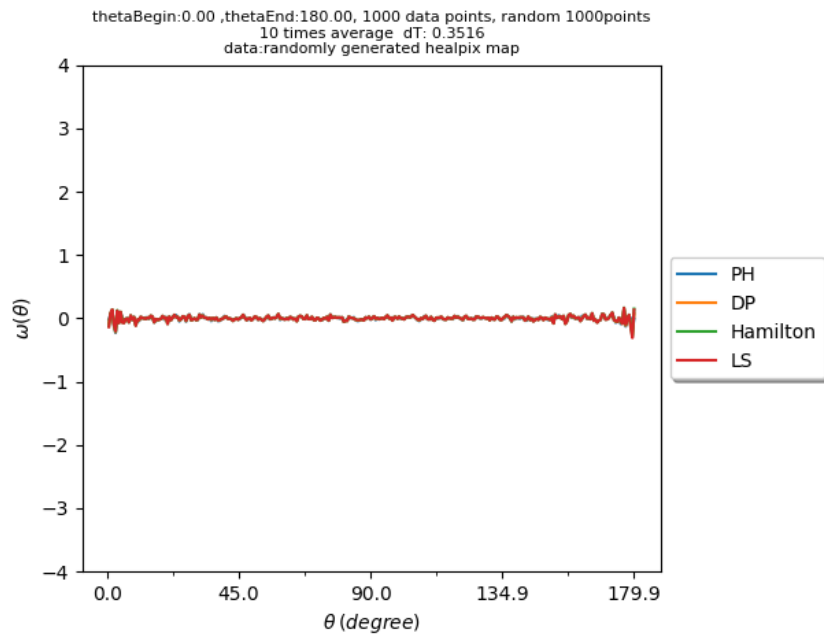


FIGURE 3.2: Four estimators of the 2pCF on randomly generated map on the whole sphere.

oscillation appears, this phenomenon will appear in all the cases below, which can be explained by the high variance of the estimators near the small θ and close to 180° . The oscillation at end points will be explained below.

In the case of practical interest where the data only covers part of the sphere, it is necessary to put constraints on the randomly generated map. Since the experiment data have been collected with the whole sweep along the azimuth angle, see Chapter 4, which leads to the assumption of the azimuthal symmetry of the region. As a consequence, only constraints on the polar angle will be put upon the location of point process throughout this thesis. Since the *Healpix* map aligns pixel in the isolatitude manner, the constraint on the polar angle is not hard to implement.

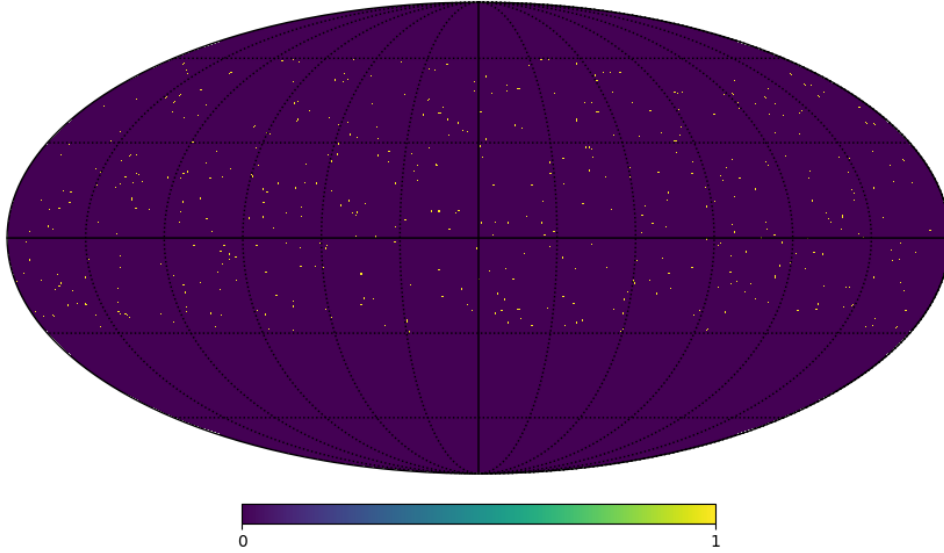


FIGURE 3.3: Healpix random map with mask.

As an example, the test result of a randomly generated map with 500 points in a region which spans from 30° to 120° is shown in Figure 3.4 and the visualisation is shown in Figure 3.3. We could see that the correlation function remains close to zero at large angles. However, at the angles close to zero, the deviation from zero is high. This phenomenon may be explained by the fact that the variance for small value of $w(\theta)$ is approximately $\frac{2}{N(N-1)G_p(\theta)}$, see [4]. In this case, for the region (using radians)

$$S := \{(1, \vartheta, \varphi) | \Theta_B \leq \vartheta \leq \Theta_E, 0 \leq \varphi \leq 2\pi\}$$

the value of $G_p(\theta)$ with $\theta \leq (\Theta_E - \Theta_B)$ can be calculated approximately by the evaluation of the function

$$G_p(\theta) = \frac{\sin \theta}{(\cos \Theta_B - \cos \Theta_E)^2} \cdot (I_1(\theta) + I_2(\theta) + \cos(\Theta_B + \theta) - \cos(\Theta_E - \theta)) d\theta \quad (3.1)$$

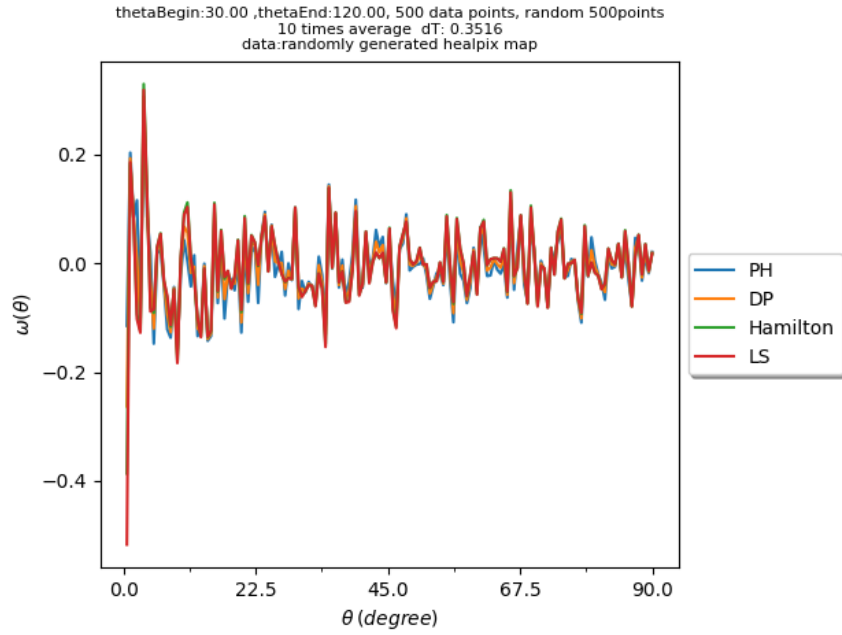


FIGURE 3.4: Four estimators of the 2pCF of randomly generated map.

where the terms $I_1(\theta)$ and $I_2(\theta)$ are given as following:

$$I_1(\theta) = \int_{\Theta_B}^{\Theta_B+\theta} \left\{ 1 - \frac{1}{\pi} \arcsin\left(\frac{\sqrt{1 - \cos^2 \Theta_B - \left(\frac{\cos \theta - \cos \Theta_B \cdot \cos t}{\sin t} \right)^2}}{\sin \theta} \right) \right\} \sin t dt \quad (3.2)$$

$$I_2(\theta) = \int_{\Theta_E-\theta}^{\Theta_E} \left\{ 1 - \frac{1}{\pi} \arcsin\left(\frac{\sqrt{1 - \cos^2 \Theta_E - \left(\frac{\cos \theta - \cos \Theta_E \cdot \cos t}{\sin t} \right)^2}}{\sin \theta} \right) \right\} \sin t dt \quad (3.3)$$

The proof of the (3.1) is based on (2.2). The illustration of the proof is also given in Figure 3.5.

To calculate $G_p(\theta)$, we need to count the number of pairs with distance θ . Due to the structure of S , the set can be divided into three parts:

$$\begin{aligned} S_1 &:= \{(1, \vartheta, \varphi) | \Theta_B \leq \vartheta < \Theta_B + \theta, 0 \leq \varphi \leq 2\pi\}, \\ S_2 &:= \{(1, \vartheta, \varphi) | \Theta_B + \theta \leq \vartheta < \Theta_E - \theta, 0 \leq \varphi \leq 2\pi\}, \\ S_3 &:= \{(1, \vartheta, \varphi) | \Theta_E - \theta \leq \vartheta \leq \Theta_E, 0 \leq \varphi \leq 2\pi\}. \end{aligned}$$

The first part contains points whose θ -disc intersects the latitude ring of polar angle Θ_B , the second part S_2 contains points whose θ -disc intersects neither latitude ring of polar angle Θ_B nor that of Θ_E , and the third part contains points whose θ -disc intersects the latitude ring of polar angle Θ_E .

For the further proof of the (3.1), I will denote the contribution of the three sets to $G_p(\theta)$ as $G_{p,1}(\theta)$, $G_{p,2}(\theta)$ and $G_{p,3}(\theta)$ respectively.

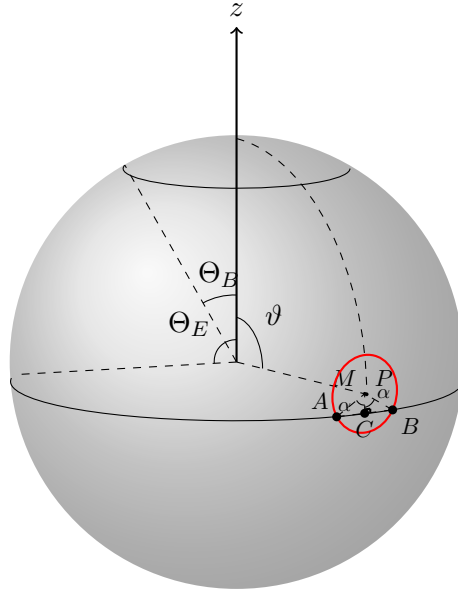


FIGURE 3.5: Illustration of the proof of (3.1).

The counting in S_2 is easy to perform, since no boundary condition needs to be considered, all the points within the θ -disc of the reference point will be counted. Since each cell has an area of $(d\theta)^2$, the total number of cells $N_{cell} = \frac{2\pi(\cos \Theta_B - \cos \Theta_E)}{(d\theta)^2}$. For a given point $P \in S_2$, its number of neighbors are $\frac{2\pi \sin \theta}{d\theta}$, at each polar angle ϑ , there are in all $\frac{2\pi \sin \vartheta d\vartheta}{(d\theta)^2}$. Note the difference of variables in $d\vartheta$ and $d\theta$, as described in Chapter 1. The former describes the height of the infinitesimally thin latitude disc, whereas the latter denotes the actual bin width.

Combining the factor mentioned before and integrating the terms with respect to ϑ , the $G_{p,2}(\theta)$ is given as following:

$$G_{p,2}(\theta) = \frac{2}{N_{cell}(N_{cell} - 1)} \cdot \int_{\Theta_B + \theta}^{\Theta_E - \theta} \frac{2\pi \sin \theta}{d\theta} \cdot \frac{2\pi \sin \vartheta}{(d\theta)^2} d\vartheta \quad (3.4)$$

$$= \frac{\sin \theta}{(\cos \Theta_B - \cos \Theta_E)^2} \cdot (\cos(\Theta_B + \theta) - \cos(\Theta_E - \theta)) d\theta \quad (3.5)$$

Since the calculation for $G_{p,1}(\theta)$ and $G_{p,3}(\theta)$ is symmetric, I will only illustrate how $G_{p,3}(\theta)$ is calculated, the case for $G_{p,1}(\theta)$ is analogous.

As noted in the case of (2.2), we could solve the x coordinate of the intersection of the θ -disc ring of a given point $p_0 \in S_3$, which is given by

$$x = \frac{\cos \theta - \cos \vartheta \cos \Theta_E}{\sin \vartheta}. \quad (3.6)$$

θ	$G_p(\theta)$	$\frac{2}{N(N-1)G_p(\theta)}$
0.0010	0.0007	0.1796
0.0190	0.0138	0.0095
0.0370	0.0268	0.0049
0.0551	0.0396	0.0033
0.0731	0.0522	0.0025
0.0911	0.0647	0.0020
0.1091	0.0770	0.0017
0.1271	0.0891	0.0015
0.1452	0.1010	0.0013
0.1632	0.1128	0.0012
0.1812	0.1243	0.0011
0.1992	0.1357	0.0010
0.2172	0.1469	0.0009
0.2353	0.1579	0.0008

TABLE 3.1: Variance of estimator.

The portion of the neighbor of p_0 lying in S_3 is given by the fraction $\frac{2\pi-2\alpha}{2\pi} = 1 - \frac{\alpha}{\pi}$, where 2α denotes the acute angle formed by the two intersection points in the ring satisfying (2.2)(given in radians). The angle α satisfies the equation:

$$\sin \alpha = \frac{y}{\sin \theta}. \quad (3.7)$$

Solving the value of y by using the property that the point lies on the unit sphere, we thus have the following equation:

$$1 - \frac{\alpha}{\pi} = 1 - \frac{1}{\pi} \arcsin\left(\frac{\sqrt{1 - \cos^2 \Theta_E - \frac{\cos \theta - \cos \Theta_E \cdot \cos t^2}{\sin t}}}{\sin \theta}\right) \quad (3.8)$$

using the same technique in the calculation of $G_{p,2}(\theta)$, we obtain the final form as in (3.3)

As for the example here, with $N = 500$, $dt = 0.3516^\circ$, some values of $G_p(\theta)$ are given in Table 3.1. We could see from Table 3.1 at small angles, the variance is quite high in comparison to that of larger angles, which explains the fluctuation near the origin.

For the point process spanning the wholesphere, $G_p(\theta)$ has the form of $\frac{\sin \theta}{2}$, which is 0 at $\theta = 0^\circ, 180^\circ$, which explains the large oscillation at the ends. In this case, we could also easily check that the integral of $G_p(\theta)$ with respect to θ from 0 to π is 1, which complies with the normalization of the probability density function.

Though it seems undesirable that the value of the correlation function at smaller angles are non-negligible, this noise tends to be of smaller amplitude than those point process with non-zero angular correlation function at the smaller region, as we shall see in Section 3.3.

As we could also see from the expression of the variance, which is inversely proportional to the square of the number of points in a given realization. So we would expect less bias from zero if more data points appear in the map, which is shown in Figure 3.6, we see the amplitude of bias is around 0.01, which is 40 times lower than in the previous case. The reason we lose another 10 factor might be the fact that we have the same bin width in the two examples and the formula for the variance is derived under the assumption that the expected number of count in a cell is approximately the probability that there is exactly one point in that cell, i.e.

$$\mathbb{E}[n] \approx \mathbb{P}(n = 1) = \frac{N}{N_{cell}}, \quad (3.9)$$

which is only true if the ensemble points are sparse or the bins are fine enough. So as the number of points increases while the bin width remain the same, the assumption above may be unfounded. This has also given us a practical guidance: to achieve low variance, we need to consider both the mean density of points and the bin width collectively.

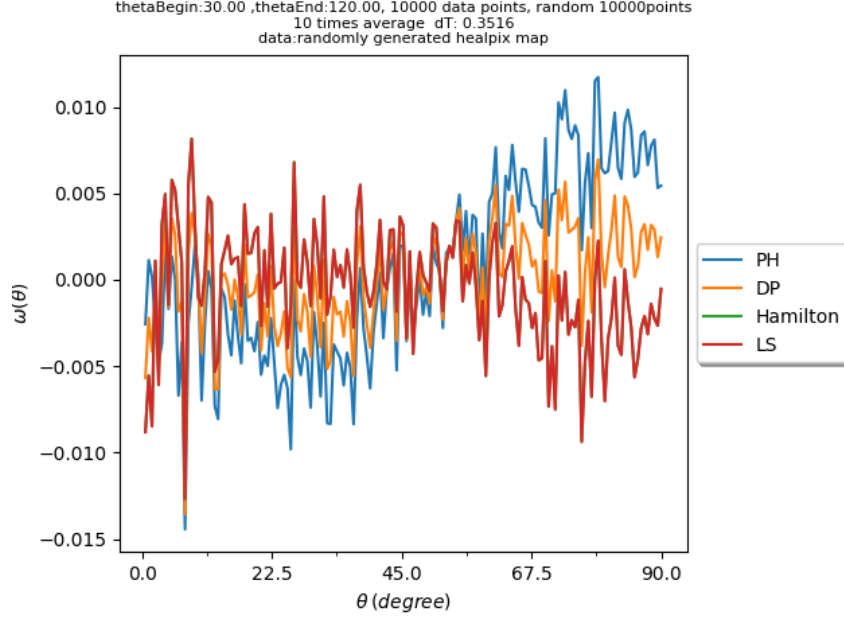


FIGURE 3.6: Four estimators of randomly generated map with 10000 points.

3.2 GENERATED HOMOGENEOUS *Poisson* POINT PROCESS

In this section I will first show the method of generating a homogenous *Poisson* point process on a sphere. Next I will show the test results on the generated *Poisson* point process.

3.2.1 *Mathematical derivation of simulation of Poisson point process on a sphere*

In this subsection I will derive the method of simulating the homogeneous *Poisson* point process on a region of a sphere with azimuthal symmetry, i.e $S := \{(r, \theta, \phi) | r = R, \theta_1 \leq \theta \leq \theta_2, 0 \leq \phi \leq 2\pi\}$, for the case where there is constraint on azimuthal angle, the result can be easily generalized, but we will focus on the case of azimuthal symmetry, since the window applied on the experiment data shares this property, see Chapter 4. Throughout this thesis, the *Poisson* point process is always meant to be homogeneous. For a general scheme about simulating inhomogeneous *Poisson* point process, see [5].

As shown in [5], the homogeneous *Poisson* point process falls into two stages. The first stage is to simulate the number of points in a region, which is a *Poisson* random variable, and the second stage is to simulate the positions of the points in the region as a uniformly distributed point process.

Many programming languages support generating *Poisson* random variables, so I will skip the first stage and show how to simulate uniformly distributed point process on a sphere.

The area of S is given by $2\pi R^2(\cos \theta_1 - \cos \theta_2)$. Denote the probability density function as $f(\vartheta, \varphi)$. As we are dealing with uniformly distributed point process, the value of $f(\vartheta, \varphi)$ at any given region with equal area is the same, so due to the normalization of the probability density function (pdf):

$$\begin{aligned} \iint f(\vartheta, \varphi) d\vartheta d\varphi &= \iint f(x) dA = 1, \quad \text{with} \quad \iint dA = 2\pi R^2(\cos \theta_1 - \cos \theta_2) \\ \Rightarrow f(x) dA &= \frac{1}{2\pi(\cos \theta_1 - \cos \theta_2) R^2} dA \stackrel{!}{=} f(\vartheta, \varphi) d\vartheta d\varphi \\ \Rightarrow \frac{1}{2\pi(\cos \theta_1 - \cos \theta_2) R^2} R^2 \sin \vartheta d\vartheta d\varphi &= f(\vartheta, \varphi) d\vartheta d\varphi \\ \Rightarrow f(\vartheta, \varphi) &= \frac{\sin \vartheta}{2\pi(\cos \theta_1 - \cos \theta_2)} \end{aligned} \quad (3.10)$$

To obtain the marginal distribution, we can integrate with respect to ϑ or φ :

$$p_\phi(\varphi) = \int_{\theta_1}^{\theta_2} f(\vartheta, \varphi) d\vartheta = \frac{\cos \theta_1 - \cos \theta_2}{2\pi(\theta_1 - \theta_2)} = \frac{1}{2\pi} \cdot \mathbb{1}_{0 \leq \varphi < 2\pi} \quad (3.11)$$

$$p_\theta(\vartheta) = \int_0^{2\pi} f(\vartheta, \varphi) d\varphi = \frac{\sin \vartheta}{\cos \theta_1 - \cos \theta_2} \cdot \mathbb{1}_{\theta_1 \leq \vartheta \leq \theta_2} \quad (3.12)$$

In particular, ϕ is uniformly distributed in $(0, 2\pi)$ as expected. To simulate θ under the given pdf. I use the fact from the probability theory that the random variable $u :=$

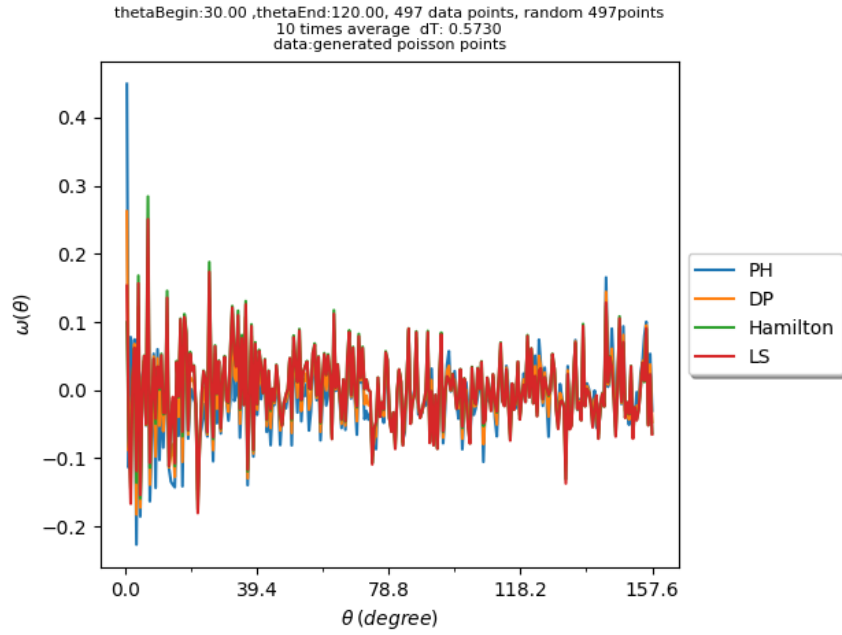


FIGURE 3.7: Four estimators of 2pCF from the *Poisson* point process, with $\lambda = 60$.

$\mathbb{P}(\vartheta \leq \theta) = F_\theta(\theta)$ is uniformly distributed in $[0, 1]$, where $\mathbb{P}(A)$ denote the probability of the set A and $F_\theta(\vartheta)$ is the cumulative distribution function of θ . The proof can be found in any books on probability theory, for example [6]

$$u = \mathbb{P}(\vartheta \leq \theta) = \int_{\theta_1}^{\theta} p_\theta(\vartheta) d\vartheta = \frac{\cos \theta_1 - \cos \theta}{\cos \theta_1 - \cos \theta_2} \sim U(0, 1).$$

Since $F_\theta(\vartheta)$ is strictly monotone increasing for $\vartheta \in [0, \pi]$ in particular in $[\theta_1, \theta_2]$ there is an inverse function. In particular, we can solve θ from u :

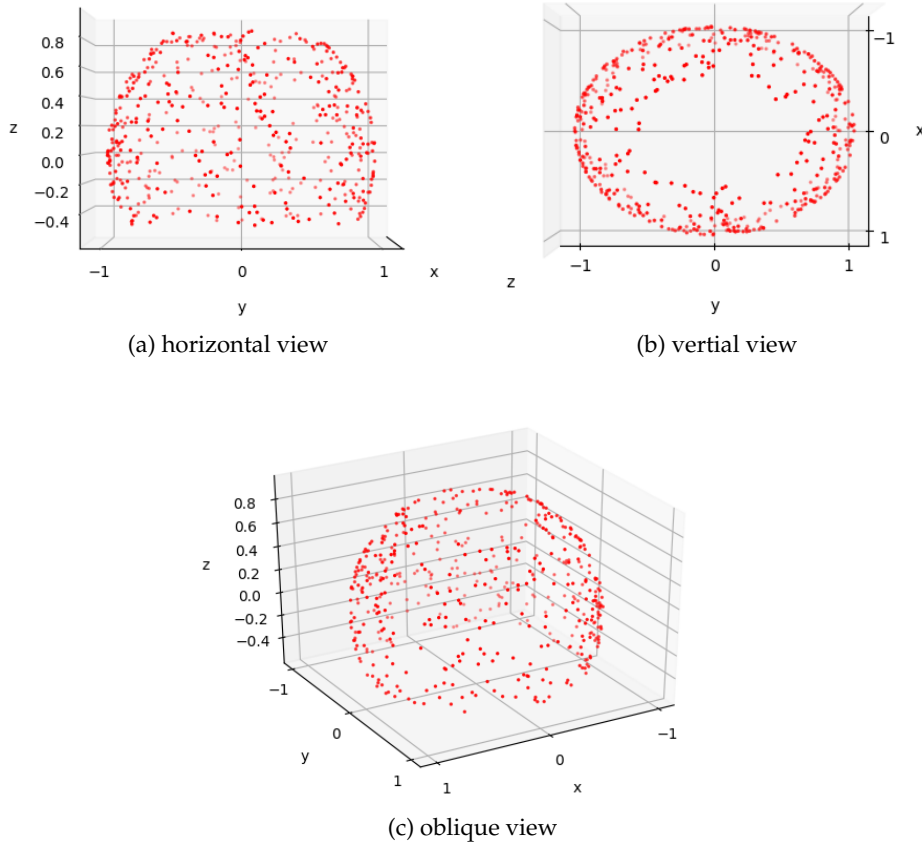
$$\theta = \arccos(\cos \theta_1 - u(\cos \theta_1 - \cos \theta_2)).$$

3.2.2 Testing generated *Poisson* point process on **Healpix** map

To use the structure of the **Healpix** map, I first generated the *Poisson* point process on a sphere, then allocating these points to the cell of the **Healpix** map.

As an example, I have attached the test result of a generated random point process with the same constraint put on the randomly generated map as in Section 3.1 with $\lambda = 60$, 497 points in all, the bin width is 0.01. The test result can be seen from Figure 3.7. I have also attached the visualisation of the points from different angles. The flat line is clear to see, as expected, indicating the *Poisson* property of the generated point process. Just as in the case of randomly generated **Healpix** map, the oscillation at small angles are more dramatic compared with that in larger angles.

However, when we compare the results of the two, we observe that the difference is quite small. Thus the generated *Poisson* data pass another test.

FIGURE 3.8: Visualisation of *Poisson* point process, $\lambda = 60$.

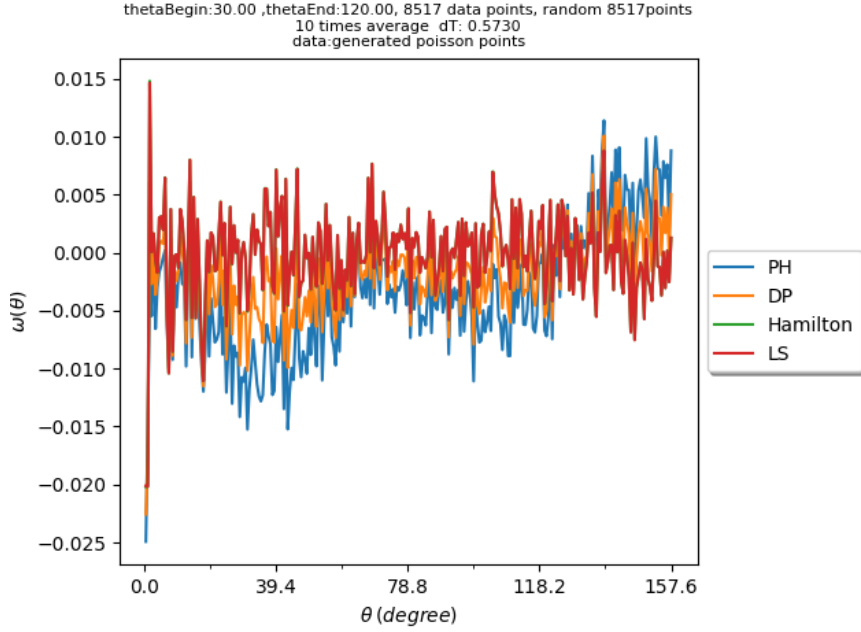
Just as in the case of the randomly generated map, as the number of points increase, we would expect the bias from zero decrease. The assumption is justified in Figure 3.9

Since the experiment data consist of 428 values, the remaining examples in this section will consider the point process with a number of points in the similar magnitude. The sparsity of the point process also makes the visualisation clearer. The previous examples with large number of points are given not only for completeness, but also as a way to show what kind of effect the size of the data set may have on the accuracy of the estimation. Having studied the *Poisson* point process, I will move further to the *Matérn* process, which will be used for the analysis of the experiment data.

3.3 GENERATED *Matérn* PROCESS

In this section, I will first illustrate the details about how to generate a *Matérn* process. Next I will show the test result of simulated *Matérn* process.

The idea of the simulation of the *Matérn* process is the following: first generate a parent *Poisson* process, which are the position of each cluster center. Having settled down each cluster center, generate within each cluster center a *Poisson* point process within a certain θ range. However, directly generating them would be complicated.

FIGURE 3.9: Generated *Poisson* point process with 8517 points, $\lambda = 1000$.

An alternative approach is to first set the cluster center p_0 to point to the z -axis, then calculate the point coordinate with respect to p_0 , then transform the coordinate to the actual position. It can be seen that this transformation is linear, so it only remains to determine the transformation matrix.

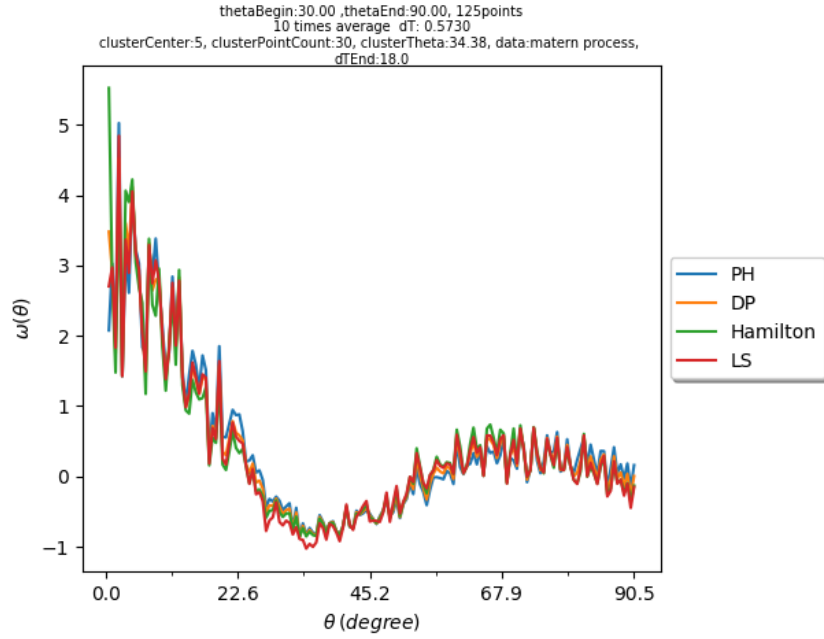
3.3.1 Derivation of the transformation matrix

For simplicity, we assume the points are on a unit sphere, otherwise we could just normalize the distance to the origin to achieve the assumption. Let the spherical coordinate of the center p_0 be $(1, \theta_0, \phi_0)$. We first define three basis vectors $B := (b_1, b_2, b_3)$ with:

$$b_1 := \begin{pmatrix} \cos \phi_0 \\ \sin \phi_0 \\ 0 \end{pmatrix}, b_2 := \begin{pmatrix} -\sin \phi_0 \\ \cos \phi_0 \\ 0 \end{pmatrix}, b_3 := \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

For the derivation below, I use the same notation introduced in [7]. Let $M_B^E(\varphi)$ denotes the matrix of the linear transformation φ with respect to the bases B in the domain and E in the range. The matrix representation of f under the basis B is then given as:

$$M_B^B(f) = \begin{pmatrix} \cos \theta_0 & 0 & \sin \theta_0 \\ 0 & 1 & 0 \\ -\sin \theta_0 & 0 & \cos \theta_0 \end{pmatrix}.$$

FIGURE 3.10: Estimators of 2pCF from the correlation of *Matérn* process.

Let $E := (e_1, e_2, e_3)$ be the canonical standard basis in \mathbb{R}^3 and id denotes the identity transformation. The transformation matrix under E is thus given as

$$\begin{aligned}
 M_E^E(f) &= M_E^B(id) \cdot M_B^B(f) \cdot M_B^E(id) \\
 &= \begin{pmatrix} \cos \phi_0 & -\sin \phi_0 & 0 \\ \sin \phi_0 & \cos \phi_0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \cos \theta_0 & 0 & \sin \theta_0 \\ 0 & 1 & 0 \\ -\sin \theta_0 & 0 & \cos \theta_0 \end{pmatrix} \begin{pmatrix} \cos \phi_0 & \sin \phi_0 & 0 \\ -\sin \phi_0 & \cos \phi_0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \\
 &= \begin{pmatrix} \cos^2 \phi_0 \cos \theta_0 + \sin^2 \phi_0 & \cos \phi_0 \sin \phi_0 (\cos \theta_0 - 1) & \cos \phi_0 \sin \theta_0 \\ \cos \theta_0 \sin \phi_0 (\cos \theta_0 - 1) & \sin^2 \phi_0 (\cos \theta_0 - 1) + 1 & \sin \theta_0 \sin \phi_0 \\ -\cos \phi_0 \sin \theta_0 & -\sin \theta_0 \sin \phi_0 & \cos \theta_0 \end{pmatrix}.
 \end{aligned}$$

3.3.2 Testing generated *Matérn* point process on **Healpix** map

As an example, the test result on a generated *Matérn* process with 325 points in total are shown in Figure 3.10. The generated point process locates in the same region as in the previous tests, spanning from 30° to 120° . The intensity λ is chosen such that the expected number of points in the 34.38° -disc is 60.

We could see that the correlation function shows a high peak at smaller angles, which is expected due to the property of the *Matérn* process.

For the plotted figure, further annotations were added to the title, *clusterCenter* denotes the number of clusters. the *clusterPointCount* denotes the expected number of points within each cluster, *clusterTheta* denotes the minimum angular distance between each cluster center, and *dTEnd* denotes the angular region each cluster has.

We could see from the estimators how the clustering property has blown up the function values near the origin, up to around 22° . As mentioned before, the amplitude

of the peak near origin is much larger than the poisson noise from a generated *Poisson* point process or randomly generated *Healpix* map. So we could distinguish between whether the underlying point process have some cluster properties or not.

The visualisation of the points in three dimension also clearly shows the cluster pattern of the point process.

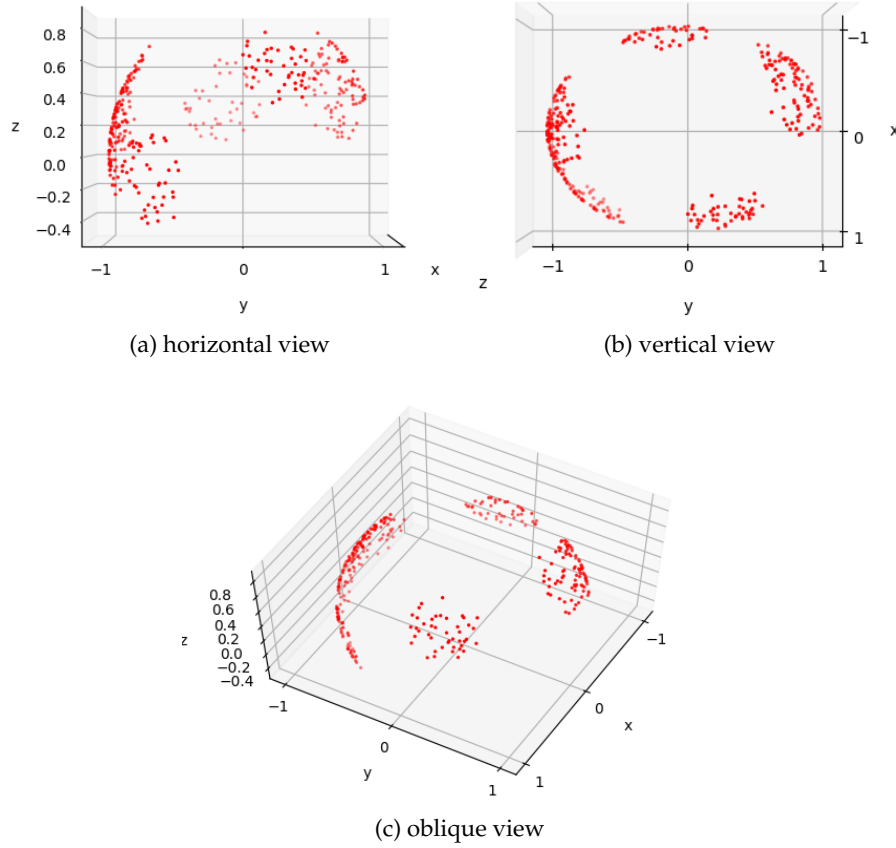


FIGURE 3.11: 3d visualisation of *Matérn* process.

APPLICATION OF 2PCF ON EXPERIMENTAL DATA

Having examined the fundamental point processes and the corresponding estimators, I will discuss the analysis of the real experimental data.

The experiment took place in building blocks in the town Langenfeld to measure the path reflections and attenuation during the signal transmission. The operating frequency was 58.32 GHz. The antenna used was 15-dBi horn antenna. The layout is shown in Figure 4.1.

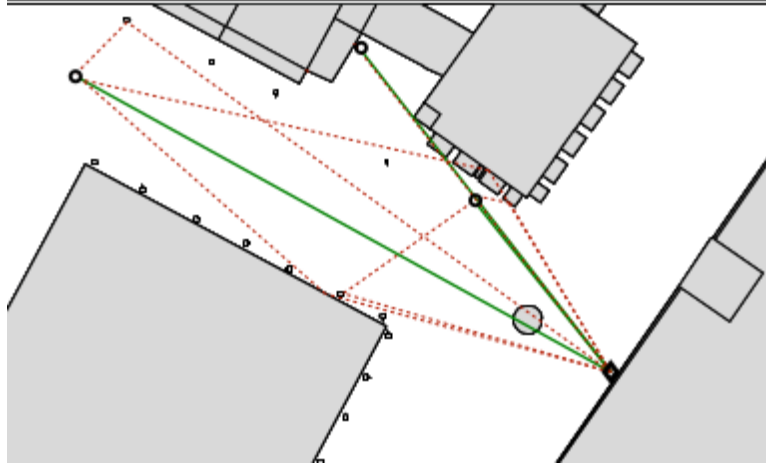


FIGURE 4.1: Experiment layout.

The experiments were carried out in the following way: The antenna was poised at different path directions. The receivers measure the receiver signal strength (RSS) at different angle positions, at each pointing direction of the antenna. The values of maximum RSS were collected at each configuration of the antenna. The resultant heatmap of the maximum RSS is constrained in the region with elevation angle from 0° to 60° with 10° step, azimuth angle from -180° to 180° , with 6° step. The layout of the antenna and the receivers is given in Figure 4.1. The transmitter are positioned at the right bottom corner. The receivers are displayed as dots with thick black rings. The green lines shows the path of line-of-sight propagation. The red dotted lines implicate the possible reflection along the propagation. The resultant heatmap of maximum RSS is shown in Figure 4.2.

Due to the scheme of the experiment, the data are constrained in the set:

$$S := \{(r, \vartheta, \varphi) | r = 1, 0^\circ \leq \vartheta \leq 60^\circ, -180^\circ \leq \varphi \leq 180^\circ\}$$

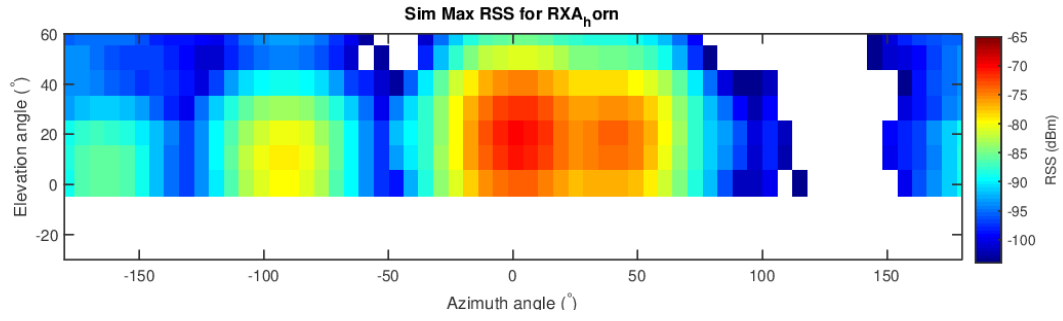


FIGURE 4.2: Maximum RSS heatmap at the receiver side.

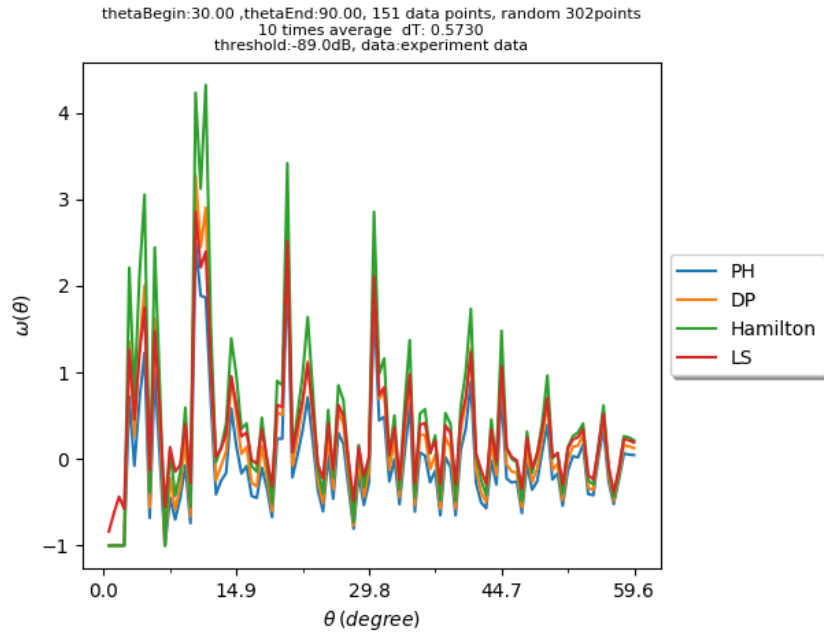


FIGURE 4.3: Max RSS with threshold -89 dB.

so it is natural to define the corresponding region on the *Healpix_Correlation* to analyze the spatial correlation.

By predefining a threshold value, we set the pixel value of each cell to be 1 or 0, depending on whether the signal strength at that point is above the threshold or not. Figure 4.3 shows the test result with the threshold value -89dB , with the bin width 0.006 radians. As mentioned before, we use the same notation as before for the title of the figure. Only a new parameter *threshold* is added to denote the value of the threshold for the analysis.

The visualization of the data uses the following strategy, each pixel with RSS above the threshold will be treated as a data point. The sets of points are thus the set of "opened slots."

We see how the peaks in the estimators of correlation function are distributed. These peaks might be explained with the help of the visualization of the experiment

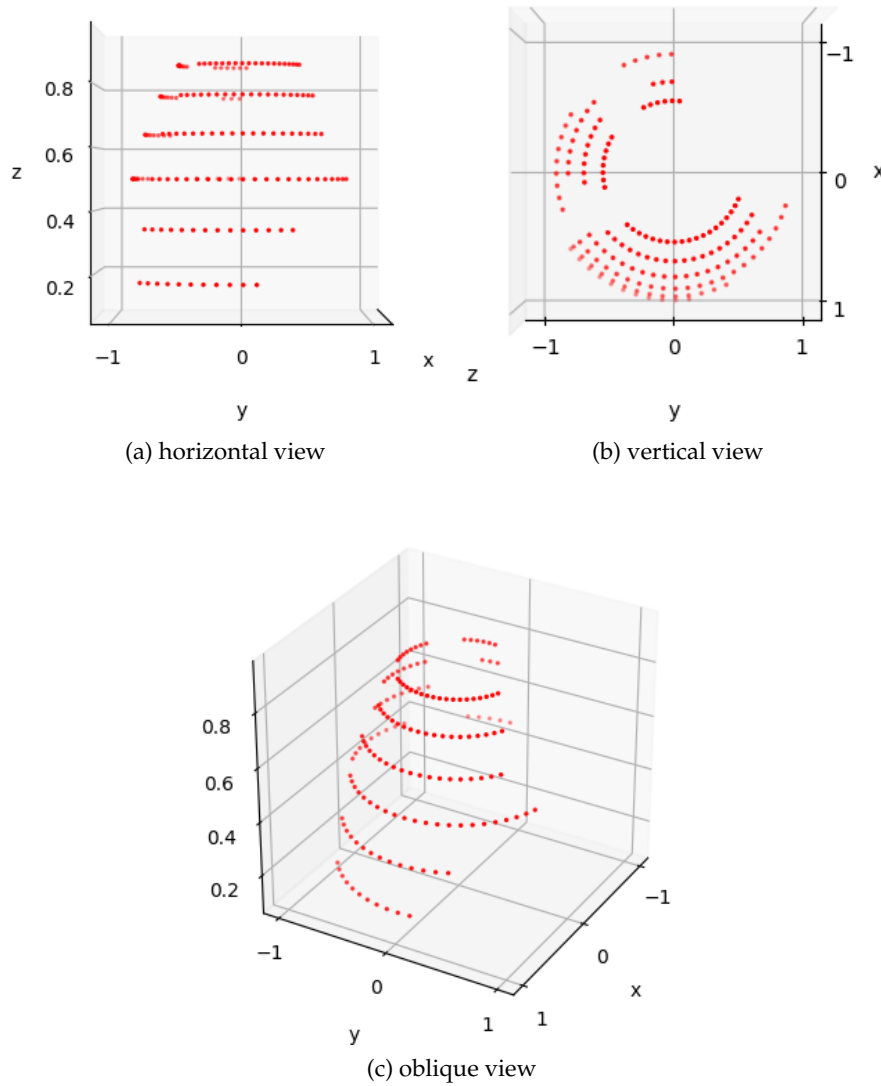


FIGURE 4.4: Visualization of experiment data.

data. The “opened slots” are distributed with 10° in latitude, which is caused by the discreteness of the data. The discreteness may also explain the angle at which the peak appears. As the threshold decreases, more slots will be open, and the peak will tend to have a decreasing amplitude. On the contrary, when the threshold increases, fewer slots would satisfy the required RSS, thus the peak will have an increasing amplitude, selecting the optimal angle more clearly, which can be seen from Figure 4.5.

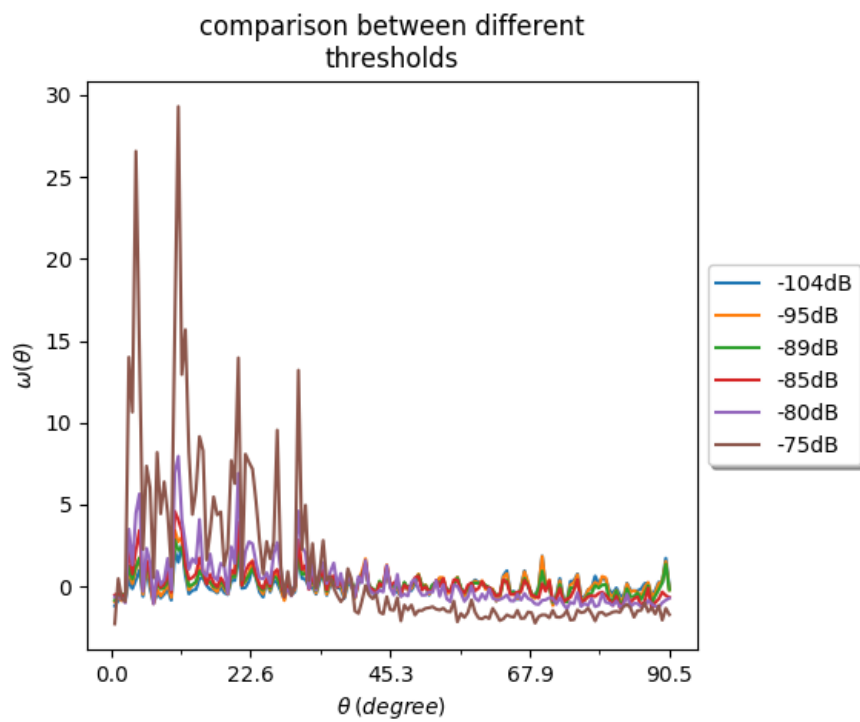
The challenge of the analysis is the small size of the data set, which might lead to considerable statistical uncertainty. In particular, we are not sure up till now, whether the peak shown in the estimators of the correlation function is due to the intrinsic correlative property or merely the result of the discreteness and statistical variance, which may appear in every random point process.

To examine the clustering property, it is thus helpful to compare the result with that of a randomly generated *Poisson* point process and *Matérn* process. The Figures 4.7, 4.8 show the 2pCF estimation of generated *Poisson* point process and *Matérn* process with total points in the same order of magnitude, around 120 points.

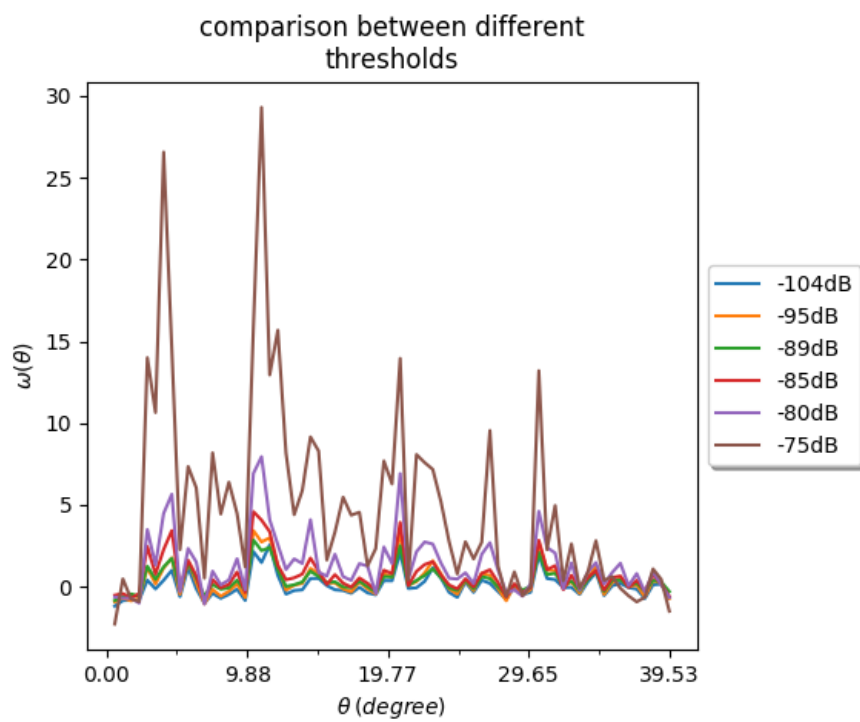
We see from the Figure 4.7 that although the sparsity leads to great bias from zero, no significant peaks can be seen. In particular, we may deduce that the peaks appearing in the estimator of the 2pCF of the experiment data is a consequence of real intercorrelation, not the sparsity of the data set.

The plot of the estimators of 2pCF of the *Matérn* process is given in Figure 4.8. In this case, we see a blow-up near the origin due to the clustering pattern and a trough near the angular distance around 34° , which can also be explained by the clustering pattern: the sparsity and cluster pattern lead to the vacancy of particular regions on the sphere. A second peak at around 68° can be viewed as the result of cross-correlation between each cluster.

We see from the comparison another issue: the large scale of the 2pCF of the *Matérn* process can be seen clearly from the estimators, whereas the estimators of 2pCF of the experiment data fail to give a clear view of the large scale structure. The problem lies in the smoothing effect of the calculation: at large angles, the θ -disc of each point tend to overlap with each other, smoothing the structural difference between each point or each cluster. However, this problem can be solved using the angular power spectrum, which offers a fair method to study the large scale structure. I will explain the brief idea and the usage of the angular power spectrum in the next chapter.



(a) large scale



(b) small scale

FIGURE 4.5: LS estimators with different thresholds.

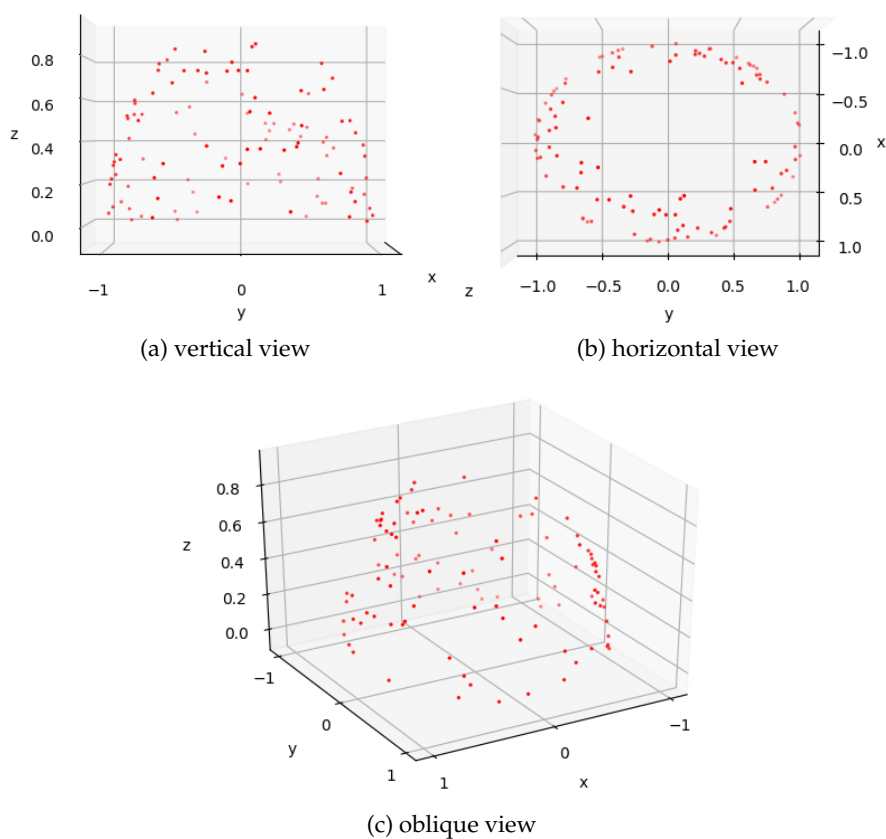


FIGURE 4.6: Visualisation of *Poisson* process with 110 points, intensity $\lambda = 22$.

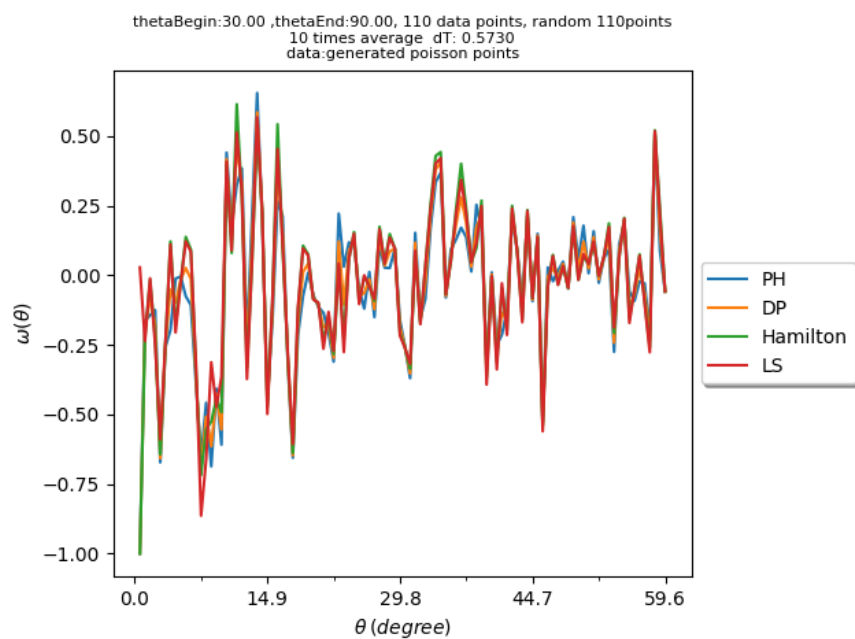


FIGURE 4.7: Four estimators of *Poisson* point process with 110 points, intensity $\lambda = 22$.

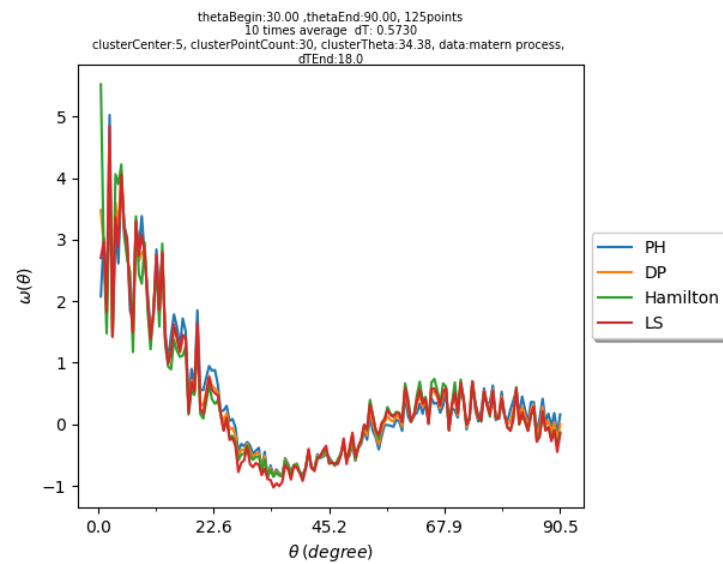


FIGURE 4.8: Correlation of 125 points *Matérn* process.

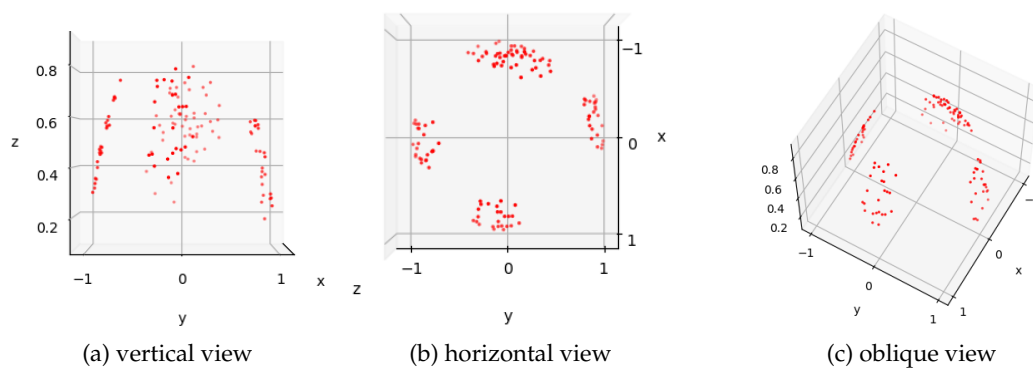


FIGURE 4.9: Visualisation of *Matérn* process with 125 points.

INTERPRETATION USING ANGULAR POWER SPECTRUM

As discussed in Chapter 1, another powerful tool is using the angular power spectrum. We know that a continuous density field $\zeta(\theta, \phi)$ can be expanded as a sum of spherical harmonics:

$$\zeta(\theta, \phi) = \sum_{l=0}^{\infty} \sum_{m=-l}^l a_{l,m} Y_{l,m}(\theta, \phi) \quad (5.1)$$

By using the (1.2), we could derive the following relation between 2pCF and angular power spectrum, detailed derivation is given in [1]

$$1 + w(\theta) = \frac{1}{4\pi\zeta_0^2} \sum_{l,m} (\mathbb{E}[|a_{l,m}^m|^2] - \zeta_0) P_l(\cos \theta) \quad (5.2)$$

Although (5.2) shows clearly the relation of angular power spectrum and the 2pCF, it is hard to directly estimate $w(\theta)$ using (5.2). What is done is actually by using the (5.1) to calculate $\widehat{a_{l,m}} = \sum_j Y_l^m(j)$, where j denotes all the point in a given realization of a point process, as an estimation of $a_{l,m}$, then calculate the power spectrum consequently.

Although the scaling before the c_l is $l(l+1)$ in the cosmology, it has little meaning in the field of what this thesis focuses on. The scaling of $2l+1$ before the coefficient c_l is more interpretable, which indicates the coefficient of the l th *legendre polynomial*.

In general, the value of ζ_0 is estimated as the mean number density within a certain window Ω . However, the geometry of the window may make it inconvenient to calculate the window size directly. Since we only care about the general behavior of the function, as explained in Section 2, we could ignore the constant term and focus on the essential term $(2l+1)c_l$.

As we shall see in the following examples, there are two ways to interpret the result. To illustrate the way to use angular power spectrum for analysis, I will use the angular power spectrum to analyze the examples we have seen in Chapter 3 and 4.

The first example is about generated random *Healpix* map with 1000 points.

The angular power spectrum has a peak at $l=0$. For $l \geq 1$, the coefficient tends to increase linearly. The peak at the origin can be seen as the static error of the estimation, whereas the linearly increasing term can first be interpreted using (5.2). I first ignore the constant to consider the sum $l \cdot P_l(\cos x)$. The cutted sum up to $l=80$ is shown in Figure 5.3. We could see how the value at the two end points tend to explode, while the value at the other angle tends to be of less deviation from zero. The explosion of the sum $l \cdot P_l(\cos(x))$ at $x=0$ is expected since $P_l(1) = 1$ for all l , the deviation at the other end is also expected since $P_l(-1) = (-1)^l$. However since the multipole

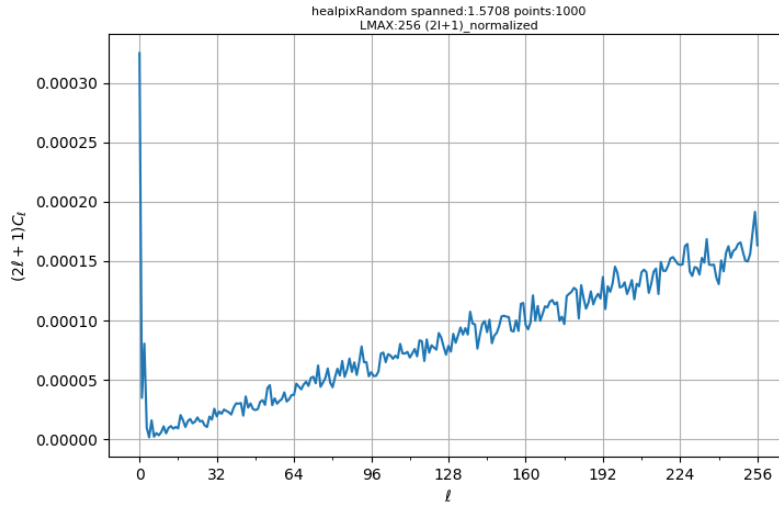


FIGURE 5.1: Scaled angular power spectrum of randomly generated map with 1000 points.

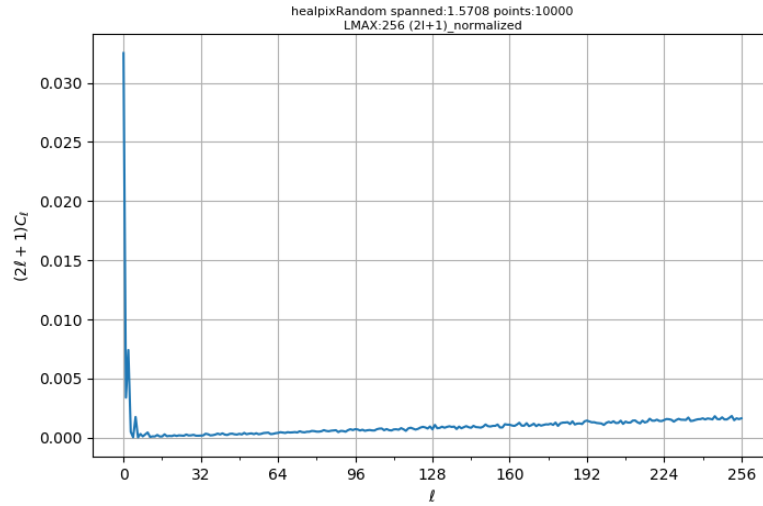


FIGURE 5.2: Angular power spectrum of randomly generated map with 10000 points.

moment l is cutted up to a certain value, this sum may not lead to infinity in reality, but it shows how the value at origin deviates from zero, which we have already seen many times in the previous analysis.

Another interpretation of the result is by considering (5.1). Since the scaled coefficient $(2l + 1)c_l$ tends to be linear, the original coefficient c_l tends to be constant for all $l \neq 0$. In fact, the c_l s are close to zero, as can be seen from Figure 5.4. The flat curve indicates the homogenous density field in the whole region, which underlies the zero angular correlation. As more points are included, the randomness will be

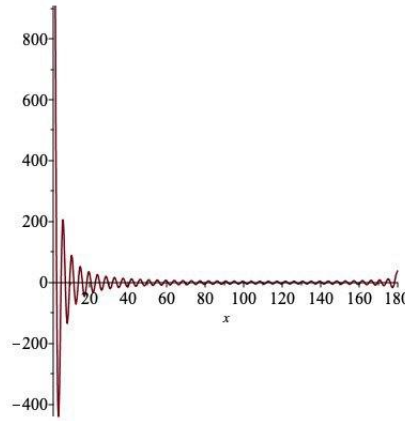


FIGURE 5.3: $\sum_{l=0}^{l=80} l \cdot P_l(\cos x)$.

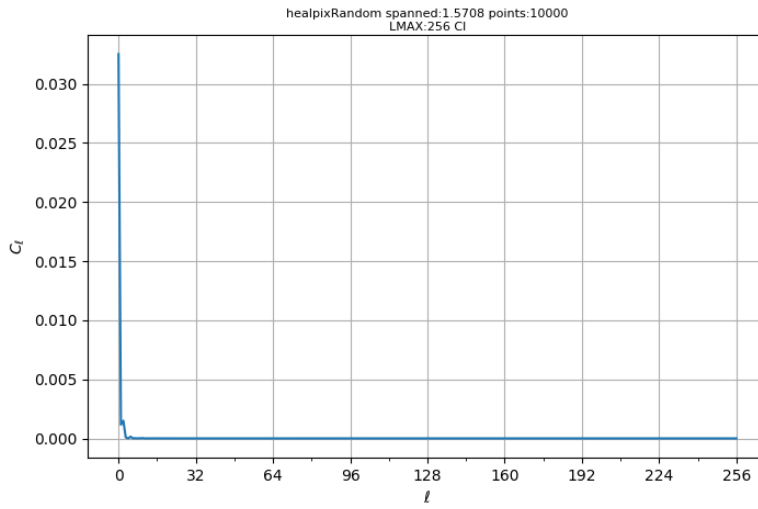


FIGURE 5.4: Angular power spectrum of randomly generated map with 1000 points.

more significant, and this will lead to a more homogeneous density field which means that all the $a_{l,m}$ except $a_{0,0}$ tends to 0, as can be seen from the comparison between Figure 5.1 and Figure 5.2.

Another interesting point worth mentioning from the comparison between Figure 5.1 and Figure 5.2 is that by examining the exact value of the angular power spectrum, we see that the ratio between the coefficient before the l -th multiple are approximately 100, which corresponds to the ζ_0^2 in the (5.2).

The reverse of the growth as mentioned earlier is also true, as can be seen from Figure 5.5, which shows the angular power spectrum of the example in Section 3.2.2.

We can see from Figure 5.5 that as point distribution goes sparse, the linear increment of the $(2l + 1)c_l$ is still to be seen, but having a larger amplitude compared with the 0th multipole. This phenomenon is expected because the fewer points there are, the less homogeneous the density field tends to be, leading to the larger growth of the

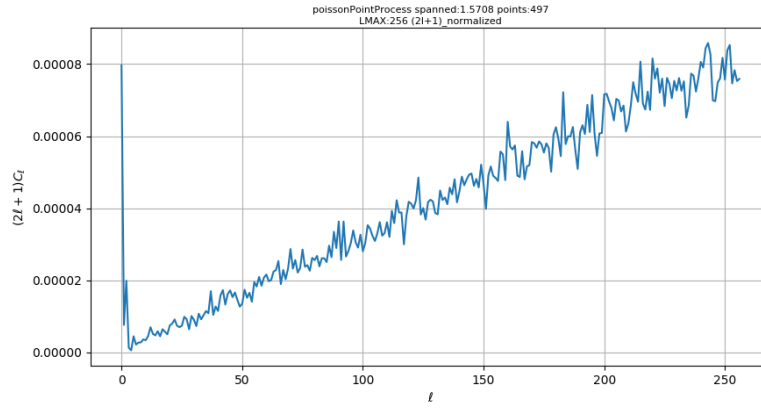


FIGURE 5.5: Angular power spectrum of *Poisson* point process with 497 points.

higher multipole coefficient. However, this inhomogeneity is counteracted by the randomness of the point distribution, thus leading to the linear increasing as what we have seen in the examples before.

The advantage of the angular power spectrum lies in the analysis of the correlation function at large angular scales. As we have seen in the examples before, the four estimators of the 2pCF always have roughly uniform amplitude when moving to a larger angular scale, which might be explained by the smoothing effect. However, due to the linear scaling in the angular power spectrum, the correlation fluctuation will be separated on different angular scales. Usually, the c_l has the highest signal at a lower l , which corresponds to roughly $\frac{\pi}{l}$ fluctuation on angular scales. In the scaled plot $(2l+1)c_l$, we will see significant peaks at small l . This larger angular scale will help us understand how the correlation function behaves at a large angular scale, since the four estimators before usually fail to give us such information, these two methods may be used as a complement to each other.

As an example, the angular power spectrum of the experimen data is shown in Figure 5.6. We see there are peaks at small l s, with $l = 3, 35, 71$. Extracting the value of the coefficient, summing the three corresponding *legendre polynomial*, the results are shown in Figure 5.7. To make the plot clearer, I have rescaled the coefficient to accommodate the plotting. We see from the plot that there is a small peak at around $\frac{5\pi}{8}$, which means that we may expect a higher correlation from the experiment data at a large angle of around 120° .

In conclusion, we see how angular power spectrum can be used to analyze the 2pCF on a large angular scale, complemented with the four estimators, we may obtain a complete view of the 2pCF.

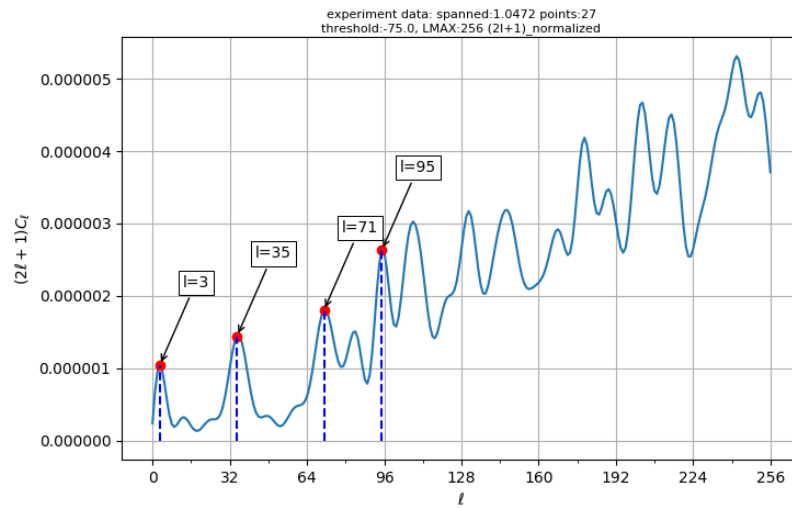


FIGURE 5.6: Scaled angular power spectrum of experimental data with threshold -75 dB.

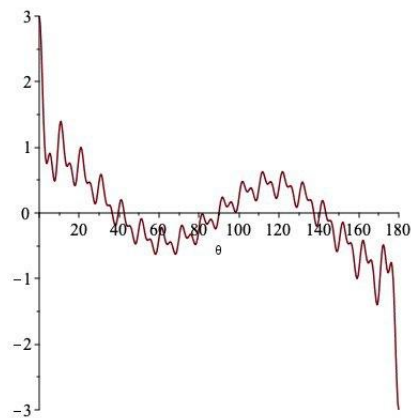


FIGURE 5.7: Scaled sum of the *Legendre polynomials* with $l = 3, 35, 71$.

CONCLUSION

This thesis demonstrates the application of the two-point angular correlation function (2pCF) to the analysis of the RSS in communication technology. Based on the *Healpix* software, this thesis first analyses the four estimators of the 2pCF on the generated data and also proposes methods of generating *Poisson* point process and *Matérn* process on a sphere. The second part of the thesis applies the 2pCF to analyze the data collected from a 60 GHz channel propagation measurements that measured the maximum received signal strength on the receiver side, as the transmitter changed its position. The third part of the thesis applied the usage of the angular power spectrum to complement the four estimators to obtain a complete view of the 2pCF.

6.1 LIMITATIONS

More massive data sets need to be tested to verify the result achieved so far. Besides, there is also a constraint on the window applied to the sphere. For simplicity and practical interest, this thesis assumes azimuthal symmetry to simplify the analysis and implementation. Further modification is required to generalize the method to the region with less spatial symmetry.

6.2 OUTLOOK

The 2pCF may further be generalized to higher dimensions. The n -point correlation could give more information about the data collected. The speed of the algorithms to calculate 2pCF can also be enhanced by further design and modification of the current algorithms. To make the estimation of the correlation more suitable and specific to the property of a given data set, the bin width of each cell could also be designed to be adaptive and of variable sizes.

A

ABBREVIATIONS

AoA angle of arrival

API Application Programming Interface

pdf (probability density function)

RSS Received Signal Strength

2pCF two-point correlation function

BIBLIOGRAPHY

- [1] P. J. E. Peebles, *The large-scale structure of the universe*. Princeton University Press Princeton, N.J, 1980.
- [2] “Healpix-features,” Available at <https://healpix.sourceforge.io/> [Last time visited 09.07.2020].
- [3] K. M. Gorski, E. Hivon, A. J. Banday, B. D. Wandelt, F. K. Hansen, M. Reinecke, and M. Bartelmann, “HEALPix: A framework for high-resolution discretization and fast analysis of data distributed on the sphere,” *The Astrophysical Journal*, vol. 622, no. 2, pp. 759–771, Apr 2005.
- [4] S. D. Landy and A. S. Szalay, “Bias and Variance of Angular Correlation Functions,” *The Astrophysical Journal*, vol. 412, p. 64, Jul. 1993.
- [5] S. N. Chiu, D. Stoyan, W. S. Kendall, and J. Mecke, *Stochastic geometry and its applications*. John Wiley & Sons, 2013.
- [6] A. Papoulis and S. U. Pillai, *Probability, random variables, and stochastic processes*. Tata McGraw-Hill Education, 2002.
- [7] D. S. Dummit and R. M. Foote, *Abstract algebra*, 3rd ed. Wiley Hoboken, 2004.

Eidesstattliche Versicherung

Statutory Declaration in Lieu of an Oath

Name, Vorname/Last Name, First Name

Matrikelnummer (freiwillige Angabe)

Matriculation No. (optional)

Ich versichere hiermit an Eides Statt, dass ich die vorliegende Arbeit/Bachelorarbeit/
Masterarbeit* mit dem Titel

I hereby declare in lieu of an oath that I have completed the present paper/Bachelor thesis/Master thesis* entitled

selbstständig und ohne unzulässige fremde Hilfe (insbes. akademisches Ghostwriting)
erbracht habe. Ich habe keine anderen als die angegebenen Quellen und Hilfsmittel benutzt.
Für den Fall, dass die Arbeit zusätzlich auf einem Datenträger eingereicht wird, erkläre ich,
dass die schriftliche und die elektronische Form vollständig übereinstimmen. Die Arbeit hat in
gleicher oder ähnlicher Form noch keiner Prüfungsbehörde vorgelegen.

independently and without illegitimate assistance from third parties (such as academic ghostwriters). I have used no other than
the specified sources and aids. In case that the thesis is additionally submitted in an electronic format, I declare that the written
and electronic versions are fully identical. The thesis has not been submitted to any examination body in this, or similar, form.

Ort, Datum/City, Date

Unterschrift/Signature

*Nichtzutreffendes bitte streichen

*Please delete as appropriate

Belehrung:

Official Notification:

§ 156 StGB: Falsche Versicherung an Eides Statt

Wer vor einer zur Abnahme einer Versicherung an Eides Statt zuständigen Behörde eine solche Versicherung
falsch abgibt oder unter Berufung auf eine solche Versicherung falsch aussagt, wird mit Freiheitsstrafe bis zu drei
Jahren oder mit Geldstrafe bestraft.

Para. 156 StGB (German Criminal Code): False Statutory Declarations

Whoever before a public authority competent to administer statutory declarations falsely makes such a declaration or falsely
testifies while referring to such a declaration shall be liable to imprisonment not exceeding three years or a fine.

§ 161 StGB: Fahrlässiger Falscheid; fahrlässige falsche Versicherung an Eides Statt

(1) Wenn eine der in den §§ 154 bis 156 bezeichneten Handlungen aus Fahrlässigkeit begangen worden ist, so
tritt Freiheitsstrafe bis zu einem Jahr oder Geldstrafe ein.

(2) Strafflosigkeit tritt ein, wenn der Täter die falsche Angabe rechtzeitig berichtigt. Die Vorschriften des § 158
Abs. 2 und 3 gelten entsprechend.

Para. 161 StGB (German Criminal Code): False Statutory Declarations Due to Negligence

(1) If a person commits one of the offences listed in sections 154 through 156 negligently the penalty shall be imprisonment not
exceeding one year or a fine.

(2) The offender shall be exempt from liability if he or she corrects their false testimony in time. The provisions of section 158 (2)
and (3) shall apply accordingly.

Die vorstehende Belehrung habe ich zur Kenntnis genommen:

I have read and understood the above official notification:

Ort, Datum/City, Date

Unterschrift/Signature