

## Article

# Clustering-Based Nonstationary Massive MIMO Channel Modeling at 1.4725 GHz

Yanping Lu <sup>1,\*</sup>, Liu Liu <sup>2</sup> and Liqin Fu <sup>1</sup>

<sup>1</sup> School of Engineering and Technology, Beijing Institute of Economics and Management, Beijing 100102, China; fuliqin@biem.edu.cn

<sup>2</sup> Institute of Broadband Wireless Mobile Communications, Beijing Jiaotong University, Beijing 100044, China; liuliu@bjtu.edu.cn

\* Correspondence: yanpinglu@biem.edu.cn

**Abstract:** The paper reports on the radio propagation characteristics of Massive MIMO. The realistic measurements are conducted in typical outdoor LOS and NLOS scenarios with the bandwidth of 100 MHz at the carrier frequency of 1.4725 GHz. In this paper the channel propagation in spectrum and space domains are investigated by employing the high-precision parameter estimation algorithm. Based on big data technology, we propose the multipath clustering algorithm and subinterval programming to bring deeper insight into the cluster evolution over the antenna array axis. The works focus on the correlation, and the result is in accordance with the theory of the cluster's visible region. Furthermore, a non-WSSUS (non-wide sense stationary uncorrelated scattering) channel analytical model is established. The whole research work aims to contribute the radio channel modeling of the 5G Massive MIMO communication system.

**Keywords:** channel modeling; clustering algorithm; Massive MIMO; multipath cluster; non-WSSUS; radio channel



**Citation:** Lu, Y.; Liu, L.; Fu, L. Clustering-Based Nonstationary Massive MIMO Channel Modeling at 1.4725 GHz. *Appl. Sci.* **2021**, *11*, 5083. <https://doi.org/10.3390/app11115083>

Academic Editors: Stefano Tomasin and George C. Alexandropoulos

Received: 14 April 2021

Accepted: 11 May 2021

Published: 30 May 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Massive multiple input and multiple output (Massive MIMO), as a key technique for the 5th generation (5G) wireless communications systems, presents a highly promising solution to meet the demanding requirements of spectrum efficiency and the energy efficiency [1–3]. The full capacity gain is mostly based on the assumption that asymptotic orthogonality of the subchannels, which requires that the channel fading coefficients satisfy independent and identical distribution. However, huge numbers of element antennas have proven to be double-edged, a great opportunity but also challenging the next generation of communication systems. New propagation characteristics are inevitably brought into wireless channel modeling due to the dense deployment of so many antenna elements. Both element number and physical size of the antenna array lead to the sub-channel correlation, which deeply influences the cooperative work of the Massive MIMO antenna element [4].

In recent years, research has spotlighted the state-of-the-art channel models. Among them, non-stationary modelling draws more attention. Lund University conducted measurements in 2.6 GHz, and the received power showed some discontinuous jump points in the dimensions of the antenna array, indicating the inconsistency of the elements [5]. In the outdoor macro-cell, three-dimensional spatial parameters of azimuth angle and elevation angle are extracted based on the realistic field data, and from the viewpoint of scatterer clusters, the correlation between clusters is fully proven [6]. In high-speed railway channel modeling, nonstationarity, as a significant requirement, is pointed out, and an angular spread-dependent model is proposed to describe the relationship between the angular dispersion and the spatial correlation [7,8]. Then, in the outdoor measurement of 64-array antenna in 3.3 GHz, the results of the angle power spectrum and the power delay profile show that the channel does not meet the stationary condition in the spatial

domain nor in the frequency domain [9]. In another measurement, with 128 antenna arrays, the non-stationary characteristic of the channel is further proven due to the influence of multipath distribution [10]. Then, in a series of measurements at 2, 4 and 6 GHz bandwidth, the non-stationary characteristic of the channel is also confirmed by the correlation bandwidth [11]. Some works cast light on visible clusters and visible regions, based on which a novel Massive MIMO model is proposed. Even more, the contribution of non-stationarity to the sum and rate of upstream channels is proved by numerical analysis [12]. In these measurements most multipaths behave in clusters. Here, a cluster is considered as a set of multipath components with similar power, delay and spatial characteristics. So far, many traditional models have been based on clusters, such as Saleh-Valenzuela [13], COST2100 [14], 3GPP SCM [15] and WINNER [16].

The traditional geometric model is mathematical based, which would go some way to weaken the universality. The dependence of the channel parameters on the measurement data result in that the channel model can only be applied to the cases with high statistical similarity with the measurement environment. Once the changes happen, the original model will degrade or even fail to work. Therefore, a multi-path clustering algorithm based on big data technology is introduced into the Massive MIMO wireless channel modeling with the development of information technology, which can help us to determine the random channel parameters through model training, which means, based on the real measurement data, we can quickly and effectively construct the channel model.

It is an inevitable trend that statistics-based big data technology is widely accepted in wireless channel modeling. With the enlargement of the antenna array scale, the improvement in channel bandwidth, the subdivision of transmission scenarios and the expansion of information scale, the in-field data volume shows an “exponential growth”. It is impossible to mine huge amounts of channel measurement data solely by human labor. Big data technology is a new technology that deals with massive and complex data and extracts valuable information quickly. With the development of large data technology, the application of mathematical algorithms on massive data to classify data types, analyze data regularity, and extract data relationships can help people explore correlations in complex data quickly and effectively. Machine learning, as a branch of artificial intelligence, is a key method of large data clustering. In recent years, the theory of artificial intelligence has been developed rapidly. The method of machine learning is helpful to enhance the adaptability of the channel model. Suppose that we have the framework of the model, the only work we need to do is to train the model with mass in-field data. Before that, we can develop the training criteria, such as MMSE (minimum mean square error). Good performance and practicability are important in channel modeling, which is meaningful to reduce the cost of re-modeling.

Massive MIMO wireless channel modeling takes a step forward by applying the methods of big data and machine learning. Compared with the logical ways, such as deduction and reasoning, the method of big data pays more attention to interconnections of data instead of causal analysis. It is also different from the traditional modeling method using geometric statistics. The geometric mode is usually based on a certain physical framework, then the model parameters are estimated using a large number of measurement data. Big data technology is a serial operation of searching, comparison and clustering, aiming to reveal the inner connections among massive data sets. It makes the model more complex to take the spatial dimension into consideration during channel modeling. However, if we catch the key item of the correlation, which is verified to exist in entries of the channel matrix, it is possible to lower the dimensions. Usually we use statistical and theoretical deduction methods to determine model parameters, while the big data method estimates these variable parameters through model training. It means a large number of in-field measurement data are employed to train the model based on a given model framework. Then, the optimal parameter set can be calculated, and the model can be determined. The innovative application of big data technology in channel modeling has attracted wide attention in academia.

The main contributions of this paper include realistic channel measurement, parameter estimate based on big data technology and state-of-the-art modelling. We focus on correlation and non-stationary characteristics. First, an account is given to describe the measurements campaign. The platform is equipped with a scalable virtual 128-element antenna array, the raw measurement data are recorded, and the channel impulse response (CIR) is illustrated. Next, a series of high-resolution estimation algorithms are employed. We carry out the data analysis based on big data technology. Then, a state-of-the-art channel model is presented based on the antenna correlations over the array. Last, we draw the conclusions.

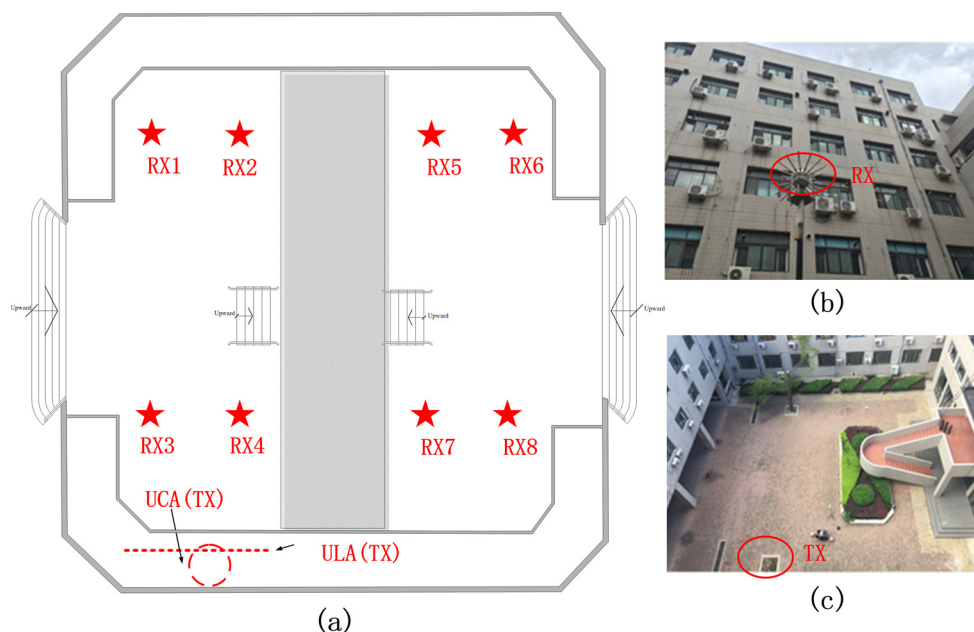
The paper is organized as follows: in Section 2 the Massive MIMO channel measurement campaign and primary signal processing are introduced. Then, in Section 3 the clustering algorithm is outlined, and the channel parameters extracted from the measured snapshots is demonstrated. We also establish the non-stationary Massive MIMO channel model. Finally, in Section 5, we summarize our contributions and draw conclusions.

## 2. Channel Measurement

### 2.1. Measurement System Setup

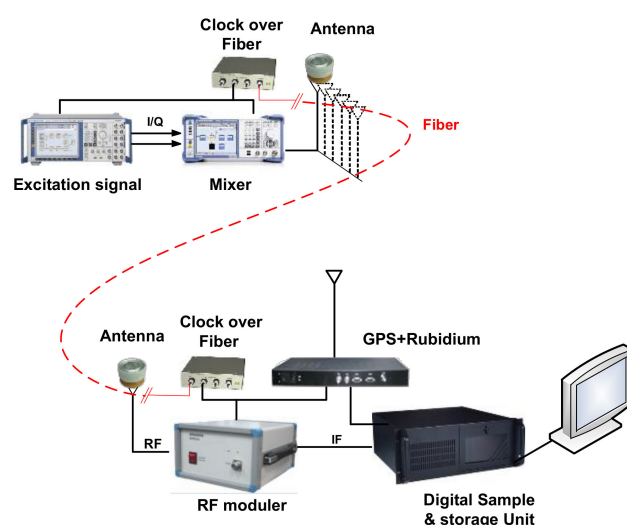
The investigations were based on channel measurements in an open-air hall at Beijing Jiaotong University. In the campaign we employed a 128-element virtual uniform linear antenna array (ULA) and uniform cylindrical antenna array (UCA) at the base station (BS), which was realized by moving an omni-directional antenna in 128 fixed positions of a rail track. A 2-element virtual antenna array was employed at the MS (mobile station). Element spacing was chosen to be a half of a wavelength in order to reduce the mutual coupling effect, which deteriorates the performance of the angle of departure (AOD) estimation algorithm.

We placed the BS antenna array on the 6th floor platform, with a height about 30 m from the ground, and set the receive equipment in the hall, at eight different MS positions. Figure 1 shows the measurement scenario. In Figure 1a, TX means the BS, where the antenna array lies in, and the RX1 to RX8 represent the eight spots of MS pointed out with eight red stars. Figure 1b,c provides views of the measurement scenario in-field.



**Figure 1.** Massive MIMO channel measurement system. (a) The measurement scenario sketch map; (b) sight from the receiver in-field; (c) sight from the transmitter in-field.

The complex-valued Zad-Off Chu (ZC) sequence worked as the excitation signal, which performs well in orthogonality and constant amplitude. A vector signal generator (R&S SMBV100A) was used as the transmitter. A pair of bi-conical antennas worked as the elements of the virtual antenna array at the BS and MS side. The antenna was assumed to provide an ideally omnidirectional radiation pattern in the azimuth plane, so idealized equal gain can be obtained in the H-plane. The data collector was composed of an RF down-converter, a high-speed digitizer card and a data storage unit. The rubidium clock trained by the GPS was used to maintain the time synchronism of the measurement system. At remote end the accurate synchronization performance was promised by a fiber unit to transmit the frequency standard. The measurement system is shown in Figure 2.



**Figure 2.** The measurement system setup.

The measurement environment was a typical outdoor micro-cell scenario. This configuration can represent a mobile station surrounded by large obstacles. Only a few operators were allowed in field, so we can assume that the channel is time-invariant. The signal bandwidth was 100 MHz, which was up-converted to radio frequency of 1.4725 GHz. The measurement parameters are listed in Table 1.

**Table 1.** Measurement system specifications.

Parameter	Quantity
Central frequency	1.4725 GHz
Bandwidth	100 MHz
Transmit power	25 dBm
Excitation signal	Zad-Off Chu sequence
Code length	2047
Tx antenna number	128
Rx antenna number	2
Tx antenna height	30 m
Rx antenna height	2 m

## 2.2. Signal Model

Describing a wireless communication system, the received signal can be decomposed into a large number of multipath signals. Furthermore, multipath can be classified and merged to form clusters according to the correlation of parameters. For a given multipath, the propagation can be characterized by power, delay, polarization, Doppler velocity, departure angle, arrival angle and so on. Considering a 3D (3 dimension) model, the angle of elevation should also be included. Therefore, we can treat the cluster as a multipath set that the multipath parameters in the set are similar.

The different locations of the antenna element determine the inconsistency of the scattering environment. As a result, the observed incident waves were different from the antennas, whether in quantity, in energy, or in angle of radio waves, as shown in Figure 3. In this case, the correlation coefficient between different antenna elements in the correlation matrix will be inconsistent [17].

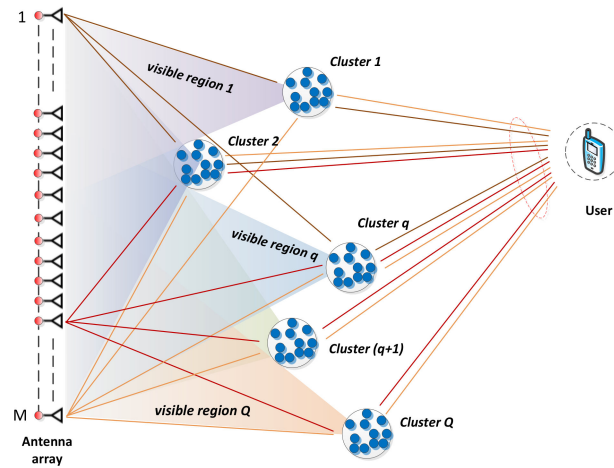


Figure 3. Nonstationary of Multipath over the Antenna Array.

Although the Massive MIMO system's complexity and difficulty will be increased because of the non-stationarity in the array domain, some methods still can be employed to reduce the complexity. In the far-field conditions, we can divide the antenna elements into several adjacent subsets according to their spatial positions, so that the antenna arrays in each subset are compact enough; that is, we can assume that the large-scale fading of all elements in the subset is the same as that of the target user [18]. With this assumption, the channel in each subset satisfies the generalized stationary non-correlation. In this way, we can use the traditional modeling theory to carry out channel modeling based on stationary characteristics in a subset window, which satisfies the stationary modeling conditions. Then the subset combines to the whole channel model, which can accurately reflect the characteristics of Massive MIMO channel.

Considering a system with transmit antennas of  $M$  and single antenna at the receiver, for the traditional MIMO channel, the small-scale propagation matrix can be expressed as

$$H = \begin{bmatrix} h_{11} & h_{12} & \cdots & h_{1M} \\ h_{21} & h_{22} & \cdots & h_{2M} \\ \vdots & \vdots & & \vdots \\ h_{K1} & h_{K2} & \cdots & h_{KM} \end{bmatrix} \quad (1)$$

where  $h_{km}$  presents the sub-channel impulse response between the  $m$ -th antenna element and the  $k$ -th user.

Given the transmit signal  $x_m(t) = a_0 e^{j\omega t} \delta(t)$ , then we have

$$h_{km}(t) = \sum_{l=1}^{L_{km}} a_l e^{j[\omega(t-\tau_l)]} \quad (2)$$

where  $a_l$  is the amplitude of the  $l$ -th incident signals, and  $L_{km}$  is the total number of multipath.  $\tau_l$  is the time-delay. Here, we assume that the antenna has a uniform radiation pattern and ignore the influence of Doppler shift.



Usually, we introduce correlation matrix  $R$  to research the correlation, which can be obtained by the long-term observations of channel matrix  $H$ .

$$R = \text{ivec}\left(E\left(\text{vec}(H) \cdot \text{vec}(H)^H\right)\right) \quad (3)$$

where  $\text{vec}(\cdot)$  presents the operator of converting Matrix to Vector, and  $\text{ivec}(\cdot)$  presents the inverse operation.

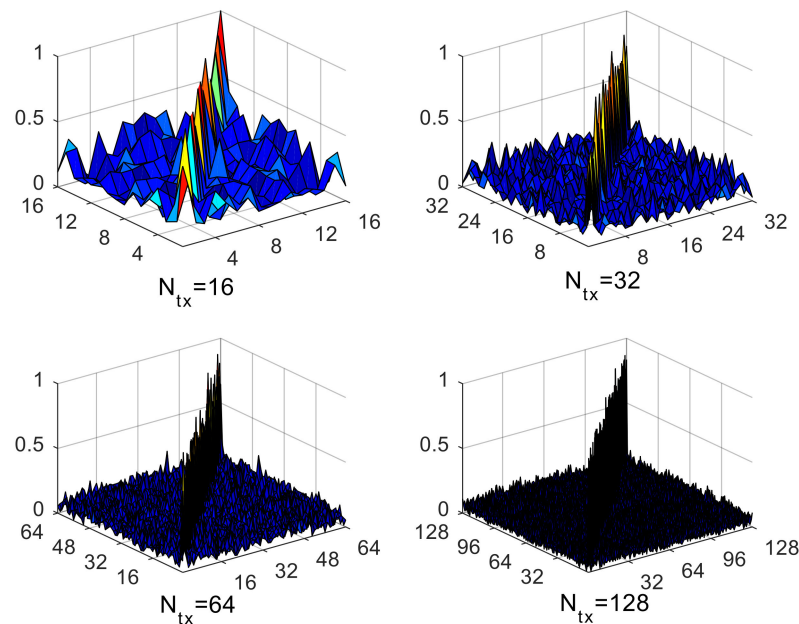
According to Kronecker's channel model, we have

$$H = R_{Rx}^{1/2} \cdot W \cdot R_{Tx}^{1/2} \quad (4)$$

where  $W$  presents the matrix that the entry obeys independent identical distribution complex Gaussian distribution with zero-mean.  $R_{Rx}$  and  $R_{Tx}$  are the correlation matrixes on BS and MS, respectively. We only consider the correlation characteristic on the base station; thus, we have the correlation matrix on user ends as  $R_{Rx} = I_{N \times N}$ . Here  $I_{N \times N}$  represents the unit diagonal array. Thus, the  $K \times M$ -dimensional channel matrix can be expressed as

$$H = W \cdot R_{Tx}^{1/2} \quad (5)$$

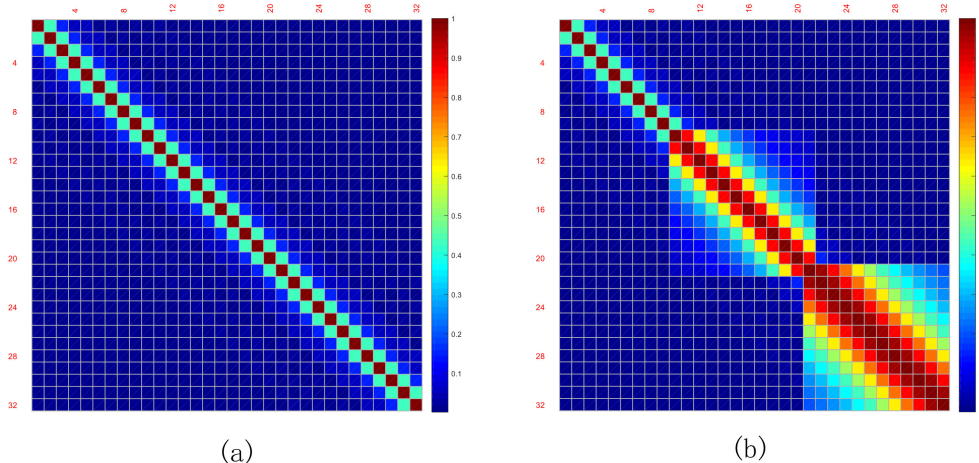
Assuming that the system model of Massive MIMO is frequency-flat and slow fading, the channel matrix  $H$  is usually modelled as independent and identically distributed with mean value of zero and variance of one. According to the Marchenko–Pastur Law, each row of the  $N_{rx} \times N_{tx}$  dimensional channel matrix  $H$  is zero mean, the variance is  $1/N_{rx}$ , and the eigenvalue distribution of the matrix  $H^H H$  and the matrix  $H H^H$  is a deterministic distribution of convergence. At that time, the channel gradually “hardened”, and the channel information was mainly concentrated on the main diagonal. That is to say, the value of the diagonal elements tends to one gradually, while the value of the non-diagonal elements tends to zero gradually. The phenomenon of channel hardening comes from the fact that the variance of the mutual information grows very slowly relative to its mean, or even shrinks as the number of antennas grows, which has been proven true [19,20] and illustrated in Figure 4.



**Figure 4.** Power distribution of channel correlation matrix.

Based on the above analysis, the correlation matrix  $R_{Tx}$  can be expressed in the form of local stationary autocorrelation. As an example, we assumed that the base station had

32 antenna array elements. If the channel is stationary, the correlation matrix can be shown in Figure 5a. If the channel is non-stationary, the correlation matrix can be divided into several stationary regions. The Massive MIMO non-stationary channel can be composed by several stationary sub-channels. As shown in Figure 5b.



**Figure 5.** Stationary and nonstationary channel correlation matrix. (a) The WSSUS correlation matrix; (b) the non-WSSUS correlation matrix.

### 3. Channel Model

The following sections demonstrate the results of data measured on Massive MIMO radio channels. These investigations aim to unfold the critical correlation features from the experimental data.

According to the measured results, the correlation fading characteristics of antenna elements were different. Therefore, it was globally non-stationary on the Massive MIMO antenna array. As to the closely placed elements, it is easily understood that the fadings were similar due to the same shadow regions they were located in. Thus, the sub-channel can be considered as stationary covered by the same shadow regions.

Generally, there are two methods to map out the stationary intervals. The first is the averaging method. According to the empirical model, the antenna elements can be divided into several subsets. Here, every subset is composed of a fixed number of elements. This method operates easily, which ignores the impact of specific propagation scenarios, but as a trade-off, it's at low accuracy. The second is the method of inspection. That is to say, we can estimate the parameters of the multi-path cluster according to the measurement results [21], then we extract the visible region of the multi-path cluster, furthermore grouping the antenna into a subset to ensure the channels are stationary within the groups. Here, a group of antenna elements is covered by a visible region in sight of radio propagation. As a result, the boundaries of the visible region are formed, and we should note that groups are irregularly divided. The method of inspection is based on the division of the antenna stationary subinterval, which greatly depends on the operator's observation and judgement. In addition, the larger the physical size of the array is, the more difficult the actual operation.

In order to overcome the limitations of these methods, we introduce the automatic clustering algorithm, named ECD (element channel distance) algorithm based on the traditional MCD (multipath component distance) algorithm applied in mathematics [22–24] and clustering algorithm applied in pattern recognition [25,26]. Clustering is to classify data objects into multiple subsets according to certain rules. Considering the special characteristics of radio channels, we develop the traditional algorithms by introducing the continuous coverage of the visible area of the cluster on the dimension of the extended antenna array. Objects in the same subset have higher similarity, while objects in different subsets have lower coupling. Distance algorithm is the most popular criterion in multi-path clustering

to model the radio channel. The multi-path with similar distances is divided into the same cluster. In practice, the distance refers to the compound distance of multi-path parameters, at the same time, the contributions of those parameters are fully considered in the way of the weighting coefficient. This is a scientific and highly efficient clustering method based on machine learning and artificial intelligence.

First of all, we should understand that the goal of clustering is to divide all measurement data into several subsets, which required high correlation within them and low coupling between them. As a result, subsets or so-called clusters are composed of multipaths, which have similar elements, including energy, delay, angle, etc.

Secondly, we should make clear that the foundation of clustering has a rich supply of data, which is not only raw data in the measurement but also the channel parameters of each antenna estimated from certain algorithms, SAGE algorithm as example, and the parameters including power, delay, angle and others. All these are called preprocessed data.

Finally, the stationary subinterval programming is carried out, which is to divide the original set of preprocessed data into some classes with similar attributes by using certain criteria. During this process, the traditional observation by a human is replaced by the machine searching. Then, some potential links between data are now uncovered with the aid of the precise algorithm.

The steps of the programming algorithm are shown in Figure 6.

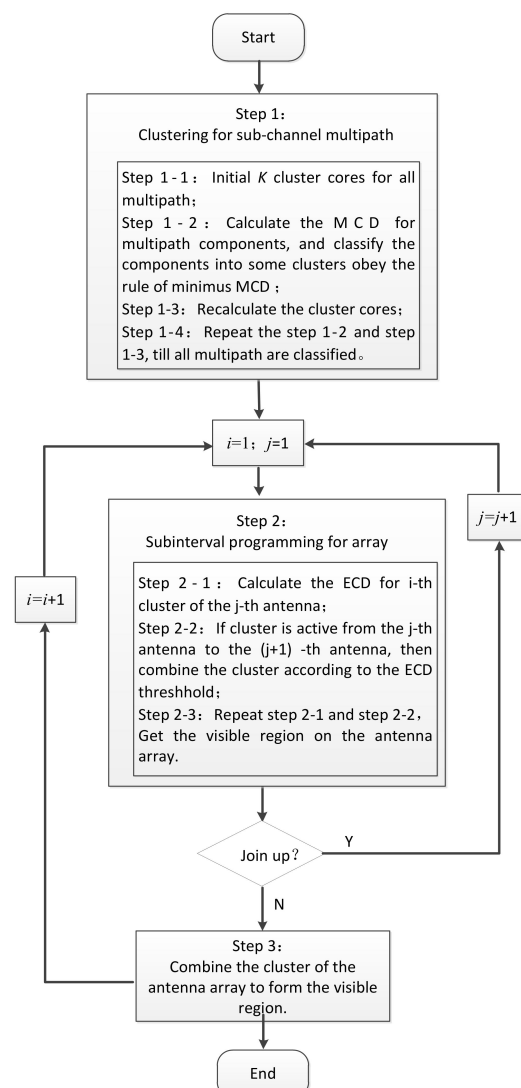


Figure 6. Flow chart of clustering and subinterval programming algorithm.



In the first step, KPowerMeans algorithm is used to cluster the multipath in the sub-channel of the antenna array. This algorithm is one of the most commonly used clustering algorithms in the field of machine learning. It belongs to the iterative clustering algorithm, which is simple and scalable.

The algorithm takes the PDP of the multipath as the initial object, searches for the peaks on the curve as the initial cluster core, and determines the number of clusters  $\{\rho_i | i = 1, \dots, k\}$ . Then the MCD can be calculated as

$$d_{MCD}(i, j) = |\rho_i - m_j| \quad (6)$$

where  $\{m_i | i = 1, \dots, L\}$  is the parameter set of multipath components. The parameters of multipath components and cluster cores include power, relative delay, departure angle, arrival angle, etc. Considering three-dimensional space, the angle information should include azimuth angle and elevation angle, so we have

$$\rho_i \in [\tau_i, \varphi_{AOA,i}, \theta_{AOA,i}, P_i] \quad (7)$$

Then, the distance in the time-delay domain can be expressed as

$$MCD_{\tau,ij} = \zeta \frac{|\tau_i - \tau_j|}{\Delta\tau_{\max}} \quad (8)$$

where  $\Delta\tau_{\max} = \max_{ij} \{|\tau_i - \tau_j|\}$ , and  $\zeta$  means the weight coefficient of time-delay.  $\tau_i$  and  $\tau_j$  are the delay of the current cluster core and the current multipath component, respectively.

The distance in the angle domain is

$$MCD_{\theta,ij} = \frac{1}{2} \left\| \begin{pmatrix} \sin \phi_i \cos \varphi_i \\ \sin \phi_i \sin \varphi_i \\ \cos \phi_i \end{pmatrix} - \begin{pmatrix} \sin \phi_j \cos \varphi_j \\ \sin \phi_j \sin \varphi_j \\ \cos \phi_j \end{pmatrix} \right\|. \quad (9)$$

In order to reduce the complexity, (9) can be simplified as

$$MCD_{\theta,ij} = \frac{1}{2} \left\| \begin{pmatrix} \sin \phi_i \\ \cos \phi_i \end{pmatrix} - \begin{pmatrix} \sin \phi_j \\ \cos \phi_j \end{pmatrix} \right\| \quad (10)$$

It should be noted that the distance in the angle domain should be calculated for the departure angle and arrival angle, respectively.

The final MCD distance is the RMS model of the weighted sum of time-delay and angle domain distance as follows:

$$d_{MCD}(i, j) = P_j \sqrt{\|MCD_{AOA,ij}\|^2 + \|MCD_{AOD,ij}\|^2 + \|MCD_{\tau,ij}\|^2} \quad (11)$$

At the same time, we can calculate the matrix of distance as

$$D_{k \times L} = \begin{bmatrix} d(1,1) & d(1,2) & \dots & d(1,L) \\ d(2,1) & d(2,2) & \dots & d(2,L) \\ \dots & \dots & \dots & \dots \\ d(k,1) & d(k,2) & \dots & d(k,L) \end{bmatrix} \quad (12)$$

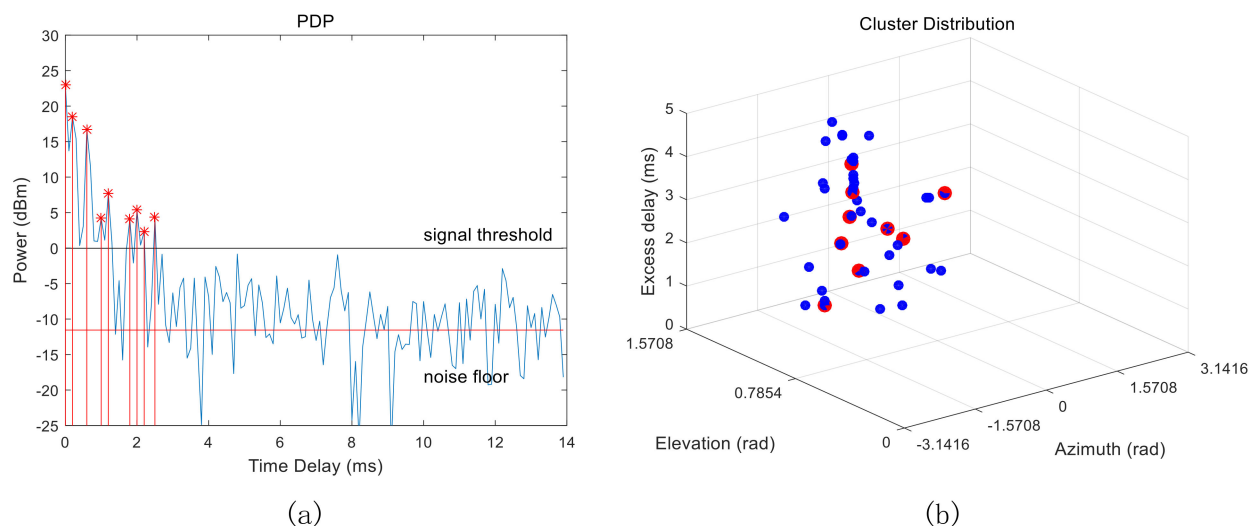
The minimum distance method is used to search for a most optimal group. The process can be summed up as a linear programming problem expressed as

$$\begin{cases} D = \min \sum_{i=1}^k \sum_{j=1}^L p_{ij} d(i, j) \\ s.t. \sum_{i=1}^k p_{ij} = 1, \sum_{j=1}^L p_{ij} = 1 \end{cases} \quad (13)$$

where  $\{p_{ij}\}_{k \times L}$  is a two-dimensional power matrix.

According to the programming results, the multipath is assigned to the nearest cluster, and the cluster cores are recalculated again. An iteration procedure is used to make sure that all the multipath components are divided into corresponding clusters.

Taking the outdoor LOS scenario as an example, the measured PDP of the antenna sub-channel is plotted as shown in the curve in Figure 7a. The average noise power and signal threshold are counted according to the estimation algorithm of bottom noise, which is identified by the horizontal real line segment in the figure. The signal implicates fluctuating nature. Peaks and valleys can be seen on the curve. According to the theory of radio wave propagation, the top points represent the arrival of multipath components with strong energy. Among all the multipath components, due to spatial scattering, we assume that the LOS path exists. It is obvious that the time-delay is the smallest and the power is the strongest as to the LOS component. Therefore, based on the traditional multi-path peak search algorithm, we can reasonably set the number of clusters as the number of peak points on the PDP, and we define the cluster core as the corresponding multipath component of the peak points. Please refer the marked points in Figure 7a for the core location.



**Figure 7.** Cluster distribution of antenna element channel. (a) The PDP of the channel between No. 1 transmit and No. 1 receive element; (b) multipath spatial distribution.

The inherent flaw of KPowerMeans algorithm is that the random hypothesis of the initial cluster core might cause the classification result to converge to the local optima rather than the global optimal solution. This problem can be overcome by using the method of peak searching on PDP. Determining the number of clusters and the initial cluster core by measurement data can greatly improve the effectiveness and reliability of the results. However, the peak search algorithm only considers the distribution of multipath in the time-delay domain, so the classification result is necessary rather than sufficient. On the other hand, the best advantage of the KpowerMeans algorithm is that the contributions of the spatial parameters are fully considered. By analyzing the results of multipath in the spatial domain, the AOA and AOD parameter pairs are obtained, and the MCD distance in the delay domain and angle domain are calculated synthetically to cluster the multipath.

Taking the outdoor LOS scenario as example, the results of the final multipath clustering are shown in Figure 7b. In the figure, the blue solid dot represents the spatial distribution of multipath in the sub-channel of array element, and the red solid dot represents the clustering result, namely the statistical distribution of cluster core. The implementation of the MCD algorithm is simple, and the clustering performance is optimized compared with the traditional algorithm.

In the second step, the ECD algorithm is used to program the stationary subinterval of the antenna array based on the cluster visible region theory. Because of obstacles, the radio reaches only several elements of the antenna array; that is, only some elements of the array are covered by some scatterer. It can be assumed that the radio wave coming from the same scatterer is coincident. Therefore, the stationary region can be defined. In the process of programming the stationary subinterval in the array dimension, it is key to determine the similarity. The distance criterion is still used here, and the distance of the correlation matrix is introduced to the work. The accuracy of measurements directly impacts the model. The matrix of distance is the metric to the similarity between two correlation matrices. The expression of the distance formula of the correlation matrix is [18]

$$d_{corr}(\Theta_1, \Theta_2) = 1 - \frac{tr\{\Theta_1\Theta_2\}}{\|\Theta_1\|_f\|\Theta_2\|_f} \quad (14)$$

Here  $\{\Theta_i|i = 1, 2, \dots\}$  is the autocorrelation expression of sub-channel matrix.  $tr\{\cdot\}$  is the trace of matrix, and  $\|\cdot\|_f$  is the Frobenius norm.

The search process of the stationary subinterval is actually the process of extracting the similarity of the correlation matrix. The algorithm assumes that there is no overlap between different stationary subintervals. We mainly focus on the influence of local scatterers on the link from the user to the base station. In this way, the signals from the user arrive at some or all elements of the antenna array after scattering. For the case where only some elements are under the coverage of the scatterer, the visible region of the scatterer only covers some elements, as shown in the above Figure 3.

According to Cauchy Schwartz inequality, as to matrices  $A$  and  $B$ , we have

$$|\langle A, B \rangle|^2 \leq \|A\|^2 \|B\|^2 \quad (15)$$

Here  $A = cB$  is the necessary and sufficient condition for the equation to hold.  $A, B$  means the inner product, which can be decomposed by the singular value as below:

$$\Theta_1\Theta_2 = U_1\Lambda_1U_1^H U_2\Lambda_2U_2^H \quad (16)$$

$U_i$  and  $\Lambda_i$  represent the eigenvector matrix and the eigenvalue diagonal matrix of  $\Theta_i$ , respectively. The value  $tr\{\Theta_1\Theta_2\}$  tends to zero, which means that there is only one module factor difference between matrices. Contrarily, if the value tends to 1, this means that the matrices are independent, that is, the matrices are extremely different.

Equation (14) can be rewritten as the following deformation forms

$$d_{corr}(\Theta_1, \Theta_2) = 1 - \frac{\langle vec\{\Theta_1\}, vec\{\Theta_2\} \rangle}{\|vec\{\Theta_1\}\|_2 \|vec\{\Theta_2\}\|_2} \quad (17)$$

Therefore, it can be assumed that the fading of radio clusters from the same scatterer is stationary, so that the stationary region  $q$  can be defined, including the number of antenna elements  $r_q$ . Here, the threshold of the correlation coefficient is 0.75. It is considered that the two correlation matrices are nonstationary if  $d_{corr}(\Theta_1, \Theta_2) > 0.75$ . Note that we suppose there is no overlap between regions, so there are

$$\sum_{q=1}^Q r_q = M \quad (18)$$

Then, the non-stationary Kronecker correlation matrix is

$$R_{Tx} = \text{bdiag}[\Theta_1, \Theta_2, \dots, \Theta_Q]$$

$$= \begin{bmatrix} \Theta_1 & & & 0 \\ & \Theta_2 & & \\ & & \ddots & \\ 0 & & & \Theta_Q \end{bmatrix}_{M \times M} \quad (19)$$

Where  $\text{bdiag}[\cdot]$  is the operator of matrix diagonalization, which is realized by arranging the vector element to the diagonal position of the target matrix.  $\Theta_q$  is the  $r_q \times r_q$  square matrix.

The cluster correlation distance can be calculated by a stepping rate of 1 on the array, and the M-element distance vector can be obtained.

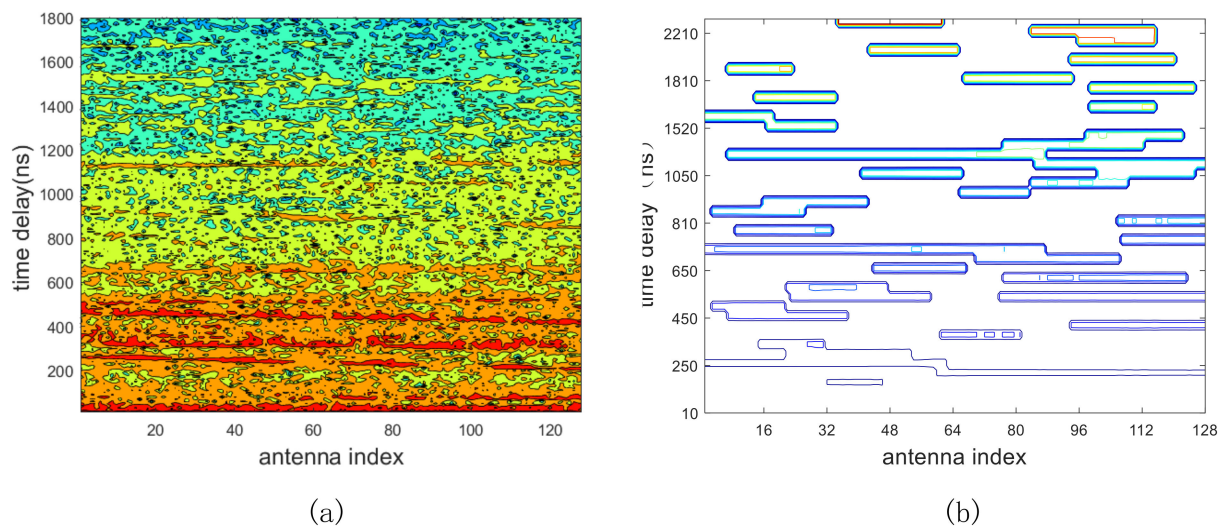
$$d_{EMD} = [d_{corr,1}, d_{corr,2}, \dots, d_{corr,M}] \quad (20)$$

If  $d_{corr,k}$  is larger than the preset threshold, and there are  $(q-1)$  values within  $k = 1, 2, \dots, M$ , then the correlation matrix is divided into matrix blocks with number  $q$ . In each matrix block with dimension  $r_q$ , the cluster correlation distance is recalculated with stepping rate of 2, and a  $(r_q-3)$  element vector is obtained. The entry of matrix block larger than the threshold is searched, and the block is further split. This is repeated until the matrix dimension is smaller than the stepping rate. Finally, every splitting block is the stationary subinterval of the correlation matrix.

In fact, the ECD algorithm is a linear projection algorithm, which aims to reduce dimensions in high-dimensional data space, such as a vector set of power, delay and angle in our application. By summarizing the linear weighted channel parameters, the parameters are converted to a one-dimensional object. Then, employing classification and clustering methods, to maximize the distance between classes, the correlation matrix is constructed. At the same time, the transformation can easily be performed forward and backward. In the mapping space, the algorithm will maximize the variance of the object, so that the projection of all samples can be separated as much as possible to complete the subset programming. Therefore, the ECD algorithm inherits the good performance of every parameter and simplifies the calculation. This method indicates that we can make up the whole through parts.

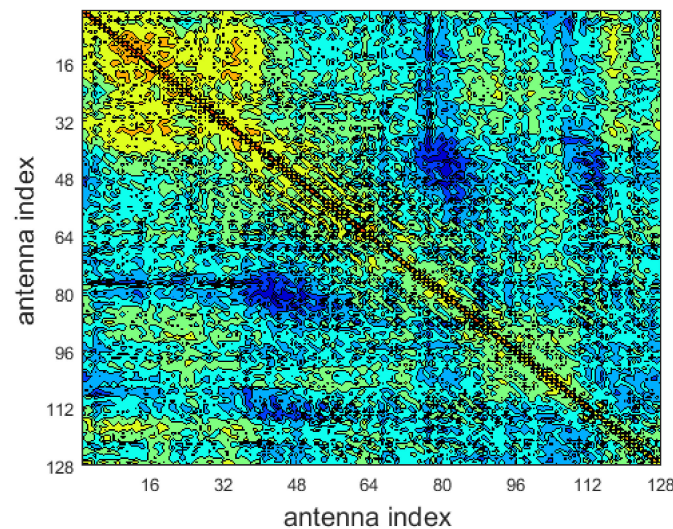
In the above example, the multipath distribution on the array domain is shown in Figure 8a. By employing the ECD algorithm, the visible regions mapped on the array domain are shown in Figure 8b.

In measurements, there are many multipaths within a cluster, and the processing of clustering will consume computing resources. As to the channel matrix, these correlated multipath components in the cluster will consume array degrees of freedom in some degree. Due to the high correlation of multipaths in a cluster, the harm to the channel gain caused by the clustering method is limited and acceptable. Therefore, modeling the Massive MIMO channel, we can consider the multi-path within a cluster as a single path to simplify the process. Here, the statistical method is replaced by searching technology to initialize the cluster core, ignoring the influence of dispersion. This method greatly simplifies the extraction of channel parameters, and it does not lose the accuracy of the model. The most important thing is that it greatly improves the efficiency of modeling. Although the correlation-based channel model ignores the small-scale dispersion characteristics of the channel in time, space and frequency domain, it can help us grasp the key items in the process of modeling. In practical applications, such as wireless communication system simulations at both the link level and system level, the Monte Carlo method is used to randomly generate multiple groups of model parameter samples on the user end. The randomness of the sample distribution can be used to simulate the small-scale dispersion characteristics and overcome the shortcomings.



**Figure 8.** Cluster distribution over the antenna array axis. (a) The PDP over the antenna array; (b) multipath cluster distribution over the antenna array.

We also take the outdoor LOS scenario as an example. According to the above steps, we can extract the multipath cluster parameters to model the channel. The finished channel correlation matrix is shown in Figure 9. In the graph, the block effect of correlation matrix can be clearly observed.



**Figure 9.** Channel correlation matrix modelling.

In the above analysis process, we only consider the situation that the transmitter is equipped with multi-antenna and the receiver holds a single antenna. Thus, we can only obtain the correlation matrix of the transmitter. However, because of the independence of the transmitter and the receiver, the correlation matrix of the receiver can also be obtained by the same method. Finally, according to the Kronecher modeling method, we can obtain the system matrix of Massive MIMO wireless channel as  $H = R_{Rx}^{1/2} \cdot W R_{Tx}^{1/2}$

#### 4. Discussion

Compared with the mostly traditional model, the channel model based on the antenna stationary subinterval has the following advantages:

- (1) The channel satisfies the linear characteristic. The clusters have a linear superposition characteristic just the same as the multipath components.



- (2) The cluster parameters are independent. Time delay, power and angle of different clusters can be described by random distribution.
- (3) The intra-cluster multipath components can be simplified as a single synthetic signal.

Then, what are the basic conditions? Firstly, the channel is non-stationary. This lies in the non-point source radiation, and so sub-channel parameters of the antenna elements are random. The extraction of channel parameters discussed earlier has fully proven it.

Secondly, theories prove that similar elements of the multi-antenna system are correlative and shift invariantly in far-field propagation [27]. The condition of far-field is quite easy to satisfy, due to the similar fading, and then the stationary interval in the antenna array domain is supposed to be held.

## 5. Conclusions

In this paper, the correlation effects of antenna elements on a Massive MIMO system in 3-D space are investigated. The measurement is conducted in an outdoor scenario, and a ULA and UCA antenna topology are employed with bandwidth of 100 MHz at 1.4725 GHz. We focus on the in-field data analysis and modeling method of Massive MIMO non-stationary channel. In the method, KPowerMeans algorithm is used to achieve efficient clustering of sub-channels. The ECD distance measure algorithm is used to program the stationary interval of clusters in an antenna array. Note that the KPowerMeans algorithm is a primary clustering of multipath components in the antenna channel, and the ECD distance measure algorithm is a secondary clustering based on the distance matrix, which increases the constraint requirements of continuous distribution in some or all antenna intervals.

All in all, it is of great significance to accurately recognize the channel propagation characteristics and to properly establish the channel propagation model as to the framework design of 5G and the future large-scale multiantenna system. As we know, propagation characteristics of typical scenarios can be obtained by channel measurements, which helps in the establishment of the channel model in the given scenario. On the contrary, once we have a good grasp of the channel model, this can guide the construction of wireless communication systems, such as antenna layout design, transmit power allocation and network optimization. Therefore, it is necessary to measure and study the Massive MIMO communication system to establish the channel model in typical scenarios. It is important to research and develop communication systems to meet the requirements of 5G communication.

**Author Contributions:** Conceptualization, methodology and funding acquisition, L.L.; validation, formal analysis and writing Y.L.; investigation and supervision, L.F. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by Beijing Natural Science Foundation (NO. L172030) and NSFC Project under Grants No. 61931001.

**Data Availability Statement:** The data used to support the findings of this study have been deposited in the figshare repository (<https://doi.org/10.6084/m9.figshare.11436522.v1> (accessed 14 April 2021)).

**Acknowledgments:** The authors would like to thank Tao Chen and Zhou Tao from Beijing jiaotong university for their help on channel measurements and useful discussion of Massive MIMO antenna topology and configuration. The data related to the results shown in the paper may be obtained by contacting the corresponding author.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Marzetta, T. Noncooperative Cellular Wireless with Unlimited Numbers of Base Station Antennas. *IEEE Trans. Wirel. Commun.* **2010**, *9*, 3590–3600. [CrossRef]
2. Larsson, E.; Edfors, O.; Tufvesson, F.; Marzetta, T. Massive MIMO for next generation wireless systems. *Commun. Mag. IEEE* **2014**, *52*, 186–195. [CrossRef]
3. Rappaport, T.S.; MacCartney, G.R.; Samimi, M.K.; Sun, S. Wideband Millimeter-Wave Propagation Measurements and Channel Models for Future Wireless Communication System Design. *IEEE Trans. Commun.* **2015**, *63*, 3029–3056. [CrossRef]

4. Ampoma, A.E.; Wen, G.; Huang, Y.; Gyasi, K.O.; Tebe, P.I.; Ntiamoah-Sarpong, K. Spatial Correlation Models of Large-Scale Antenna Topologies Using Maximum Power of Offset Distribution and its Application. *IEEE Access* **2018**, *6*, 36295–36304. [CrossRef]
5. Liu, G.; Hou, X.; Wang, F.; Jin, J.; Tong, H.; Huang, Y. Achieving 3D-MIMO With Massive Antennas from Theory to Practice with Evaluation and Field Trial Results. *IEEE Syst. J.* **2017**, *11*, 62–71. [CrossRef]
6. Sangodoyin, S.; Kristem, V.; Bas, C.; Käske, M.; Lee, J.; Schneider, C.; Zhang, J.; Thomä, R.; Molisch, A.F. Cluster-based analysis of 3D MIMO channel measurement in an urban environment. In Proceedings of the MILCOM 2015—2015 IEEE Military Communications Conference, Tampa, FL, USA, 26–28 October 2015; pp. 744–749.
7. Zhou, T.; Li, H.; Wang, Y.; Liu, L.; Tao, C. Channel Modeling for Future High-Speed Railway Communication Systems: A Survey. *IEEE Access* **2019**, *7*, 52818–52826. [CrossRef]
8. Zhou, T.; Tao, C.; Salous, S.; Liu, L. Joint Channel Characteristics in High-Speed Railway Multi-Link Propagation Scenarios: Measurement, Analysis, and Modeling. *IEEE Trans. Intell. Transp. Syst.* **2019**, *20*, 2367–2377. [CrossRef]
9. Fei, D.; He, R.; Ai, B.; Zhang, B.; Guan, K.; Zhong, Z. Massive MIMO channel measurements and analysis at 3.33 GHz. In Proceedings of the 10th International Conference on Communications and Networking in China (ChinaCom), Shanghai, China, 15–17 August 2015; pp. 194–198.
10. Liu, L.; Tao, C.; Matolak, D.W.; Lu, Y.; Ai, B.; Chen, H. Stationarity Investigation of a LOS Massive MIMO Channel in Stadium Scenarios. In Proceedings of the IEEE 82nd Vehicular Technology Conference (VTC2015-Fall), Boston, MA, USA, 2015; pp. 1–5.
11. Li, J.; Ai, B.; He, R.; Guan, K.; Wang, Q.; Fei, D.; Zhong, Z.; Zhao, Z.; Miao, D.; Guan, H.; et al. Measurement-Based Characterizations of Indoor Massive MIMO Channels at 2 GHz, 4 GHz, and 6 GHz Frequency Bands. In Proceedings of the IEEE 83rd Vehicular Technology Conference (VTC Spring), Nanjing, China, 15–18 May 2016; pp. 1–5.
12. Li, X.; Zhou, S.; Björnson, E.; Wang, J. Capacity Analysis for Spatially Non-Wide Sense Stationary Uplink Massive MIMO Systems. *IEEE Trans. Wirel. Commun.* **2015**, *14*, 7044–7056. [CrossRef]
13. Wu, W.; Lee, C.; Wang, C.; Chao, C. Signal-to-Interference-Plus-Noise Ratio Analysis for Direct-Sequence Ultra-Wideband Systems in Generalized Saleh–Valenzuela Channels. *IEEE J. Sel. Top. Signal. Process.* **2007**, *1*, 483–497. [CrossRef]
14. Liu, L.; Oestges, C.; Poutanen, J.; Haneda, K.; Vainikainen, P.; Quitin, F.; Tufvesson, F.; De Doncker, P. The COST 2100 MIMO channel model. *IEEE Wirel. Commun.* **2012**, *19*, 92–99. [CrossRef]
15. Spatial Channel Model for Multiple Input Multiple Output Simulation, 3GPP TR 25.996 V11.0.0. September 2012. Available online: [https://www.etsi.org/deliver/etsi\\_tr/125900\\_125999/125996/11.00.00\\_60/tr\\_125996v110000p.pdf](https://www.etsi.org/deliver/etsi_tr/125900_125999/125996/11.00.00_60/tr_125996v110000p.pdf) (accessed on 14 April 2021).
16. Baum, D.S.; El-Sallabi, H.; Jämsä, T.; Meinilä, J. IST-2003-507581 WINNER D5.4 v1.4 Final Report on Link Level and System Level Channel Models. 2005. Available online: <http://www.ist-winner.org/DeliverableDocuments/D5.4.pdf> (accessed on 14 April 2021).
17. Weichselberger, W.; Herdin, M.; Ozelik, H.; Bonek, E. A stochastic MIMO channel model with joint correlation of both link ends. *IEEE Trans. Wirel. Commun.* **2006**, *5*, 90–100. [CrossRef]
18. Ngo, H.Q.; Larsson, E.G.; Marzetta, T.L. Energy and Spectral Efficiency of Very Large Multiuser MIMO Systems. *IEEE Trans. Commun.* **2013**, *61*, 1436–1449.
19. Roy, M.; Paquelet, S.; Magoarou, L.L.; Crussière, M. MIMO Channel Hardening for Ray-based Models. In Proceedings of the 14th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob), Limassol, Kipar, 15–17 October 2018; pp. 1–7.
20. Narasimhan, T.L.; Chockalingam, A. Channel Hardening-Exploiting Message Passing (CHEMP) Receiver in Large-Scale MIMO Systems. *IEEE J. Sel. Top. Signal. Process.* **2014**, *8*, 847–860. [CrossRef]
21. Czink, N.; Mecklenbrauker, C.; Galdo, G.d. A Novel Automatic Cluster Tracking Algorithm. In Proceedings of the IEEE 17th International Symposium on Personal, Indoor and Mobile Radio Communications, Helsinki, Finland, 11–14 September 2006; pp. 1–5.
22. Czink, N.; Cera, P.; Salo, J.; Bonek, E.; Nuutinen, J.-P.; Ylitalo, J. Improving clustering performance using multipath component distance. *Electron. Lett.* **2006**, *42*, 33–45. [CrossRef]
23. Herdin, M.; Czink, N.; Ozelik, H.; Bonek, E. Correlation matrix distance, a meaningful measure for evaluation of non-stationary MIMO channels. In Proceedings of the IEEE 61st Vehicular Technology Conference, Stockholm, Sweden; 2005; Volume 1, pp. 136–140.
24. Steinbauer, M.; Ozelik, H.; Hofstetter, H.; Mecklenbrauker, C.F.; Bonek, E. How to Quantify Multipath Separation. *IEICE Trans. Electron.* **2002**, *85*, 552–557.
25. Pi, J.; Shi, Y.; Chen, Z. Similarity and cluster analysis algorithms for microarrays using R\* trees. In Proceedings of the IEEE Computational Systems Bioinformatics Conference—Workshops (CSBW'05), Stanford, CA, USA; 2005; pp. 91–92.
26. Bandyopadhyay, S.; Saha, S. A Point Symmetry-Based Clustering Technique for Automatic Evolution of Clusters. *IEEE Trans. Knowl. Data Eng.* **2008**, *20*, 1441–1457. [CrossRef]
27. Patel, K.M.A.; Thakral, P. The best clustering algorithms in data mining. In Proceedings of the International Conference on Communication and Signal Processing (ICCSP), Melmaruvathur, India, 6–8 April 2016; pp. 2042–2046.