Modified Fuzzy C-means Algorithm for Clustering Multipath Components in Radio Channel Modeling

Zhiming Huang
Department of Communication Engineering
Northwestern Polytechnical University, China
Email: huangzhiming@mail.nwpu.edu.cn

Jianping Pan
Department of Computer Science
University of Victoria, Canada
Email: pan@uvic.ca

Ruonan Zhang
Department of Communication Engineering
Northwestern Polytechnical University, China
Email: rzhang@nwpu.edu.cn

***Abstract*—Multipath Components (MPCs) are proved to be distributed as different groups known as clusters in radio channel. Since many advanced radio channel model have adopted the cluster concept, clustering MPCs becomes a critical process in radio channel modeling.Thus it is of great importance to cluster the large amount of MPCs automatically and accurately. In this paper, Fuzzy C-means (FCM) algorithm, one of the methods that have once been introduced to group MPCs, is modified with a mapping function, which renders the algorithm converge to a better solution. Meanwhile, a new framework based on the modified FCM is proposed. Compared with the former FCM and KPowerMeans, the proposed framework is proved to have a better clustering performance and provides basis for further research.**

I. INTRODUCTION

Massive results from channel measurements justify that multipath components (MPCs) are not independent with each other, but are distributed in groups [1]. Besides, with the development of radio channel model, the concept of cluster has been adopted by many advanced models to maintain accuracy, such as WINNER channel model, COST2100 [2] and Saleh-Valenzuela (SV) [3] model. MPCs in the same cluster share similar parameters like delay, azimuth and elevation of arrival [4], and power. Therefore, Clustering MPCs efficiently and accurately is of great importance in establishing such models. Many contributions have been made to cluster MPCs to meet the demand of research and engineering. Among those contributions, there are two main techniques. One is applying Multipath Components Distance (MCD) [5] to measure the similarity among MPCs and using cluster algorithm to group MPCs. The typical representative is KPowerMeans, which is proposed by Czink $et\ al.$ [6]. It is based on KMeans algorithm and taking account of power factor. Following KPowerMeans, a statistical technique to cluster CIRs by dividing the data into multidimensional analysis regions was presented by Xiao $et\ al.$ [7]. And a density-based cluster algorithm that do not need prior knowledge about cluster number is proposed by Ruisi $et\ al.$ [8]. The other technique is to exploit the property that the power of a MPC declines exponentienly with the MPC's delay increasing. By fitting the power-delay curve, the MPCs can be clustered automatically. For instance, James $et\ al.$ proposed an automated identification of clusters in Channel Impulse reponses(CIRs), where the objective is to fit a series of exponential curves to the measured CIRs by minimizing the root-mean-squared error (RMSE) [9]. Simillarly, Gentile presented a region competition method to fit curves so as to cluster MPCs [10]. And a new sparsity-based clustering method was adopted by Ruisi $et\ al.$ [11], [12], which fits the curve very well. Each technique has its own merits. This paper focuses on the first, which considers more parameters of MPCs while the other only takes account of power and delay factors. Among the first kind of methods, schneider $et\ al.$ introduced Fuzzy C-means (FCM) into clustering MPCs [13], which is proved that with random initialization, FCM algorithm outperforms KPowerMeans. It is a good start of introducing the soft cluster algorithm but sometimes the algorithm seems to be unable to converge to an optimum solution. Therefore, this paper proposes a map function to modify FCM algorithm and presents a novel framework based on Modified FCM, which incorporates the initialization process, clustering performance indices and pruning process. The comparison between KPowerMeans and Modified FCM framework shows the latter not only clusters the MPCs effectively but also reduces noise efficiently.

The rest of paper is organized as follow, Section II will demonstrate the preliminaries about channel model, the classic FCM algorithm and cluster performance indices. The novel framework and comparison with KPowerMeans will be presented in Section III and Section IV. The final section concludes the paper.

II. PRELIMINARIES

A. Channel Model

In this paper, a double-directional channel model [14] is considered. It is a geometric based model which takes the cluster of MPCs into consideration. According to its definition, the MPCs in a cluster have similar parameters such as Delays, Powers, Angle of Departure(AoD) and Angle of Arrival(AoA). Thus the clustering algorithm could be applied to classify the MPCs automatically. The model can be represented by Channel Impulse Response function, as illustrated by Eq.(1), where M represents for the number of clusters, N_m for the number of MPCs in the $m-$th cluster, $P_{n,m}$ for the Power of an MPC and $\tau_m, \Omega_{T,m}, \Omega_{R,m}$ for the Delay, AoD and AoA of an MPC respectively. Therefore, it is clear that clustering MPCs is a prerequisite for the establishment of such a model.

$$h(t,\tau,\Omega_T,\Omega_R)=\sum_{m=1}^{M}\{\sum_{n=1}^{N_m}P_{n,m}e^{j\Phi_{m,n}}\delta(\tau-\tau_m)\times\delta(\Omega_T-\Omega_{T,m})\times\delta(\Omega_R-\Omega_{R,m})\} \tag{1}$$

B. Fuzzy-C means Algorithm

This section will introduce Fuzzy-C means (FCM) algorithm and its application in clustering MPCs [13]. FCM is a soft clustering algorithm, which allows every sample belonging to different clusters at the same time. The object function of FCM is Eq.(1)

$$J_m=\sum_{i=1}^{N}\sum_{j=1}^{C}u_{ij}^m\cdot\|x_i-c_j\|^2 \tag{1}$$

where m is the superscript which usually takes value 2. u_{ij} is the membership belief of i-th sample to j-th cluster which should satisfy $\sum\limits_{j=1}^{C}u_{ij}=1$, N and C is the number of samples and clusters respectively, x_i and c_j is i-th sample and j-th centroid. In order to minimize the value of J_m, Lagrange Multiplier method is adopted as below:

$$L_{J_m}=\sum_{i=1}^{N}\sum_{j=1}^{C}u_{ij}^m\cdot\|x_i-c_j\|^2+\lambda(\sum_{j=1}^{C}u_{ic}-1) \tag{2}$$

And the following conditions should be satisfied to minimize J_m:

$$\frac{\partial L_{J_m}}{\partial\lambda}=(\sum_{j=1}^{C}u_{ij}-1)=0 \tag{3}$$

$$\frac{\partial L_{J_m}}{\partial u_{ij}}=(\sum_{i=1}^{N}\sum_{j=1}^{C}m(u_{ij})^{m-1}\|x_i-c_j\|^2-\lambda)=0 \tag{4}$$

$$\frac{\partial L_{J_m}}{\partial c_j}=(\sum_{i=1}^{N}\sum_{j=1}^{C}{u_{ij}}^m x_i-c_j\sum_{i=1}^{n}{u_{ij}}^m)=0 \tag{5}$$

From Eq.(3), Eq.(4) and Eq.(5), the calculation for membership belief u_{ij} and centroid c_j can be obtained as follows:

$$u_{ij}=\frac{1}{\sum\limits_{k=1}^{C}(\frac{\|x_i-c_j\|}{\|x_i-c_j\|})^{\frac{2}{m-1}}} \tag{6}$$

$$c_j=\frac{\sum\limits_{i=1}^{N}{u_{ij}}^m\cdot x_i}{\sum\limits_{i=1}^{N}{u_{ij}}^m} \tag{7}$$

The main idea of FCM algorithm is to initialize the membership matrix u_{ij} or the centroids vectors c_j, then iterating with Eq.(6) and Eq.(7) to converge to an optimal solution.

Work [13] introduces FCM to cluster MPCs. It replaces the term $\|x_i-c_j\|$ with $MCD(x_i,c_j)$ and proposes the calculation for membership belief of every MPC and centroids as Eq.(11) and Eq.(12). MCD is a MPCs similarity measure function, greater the MCD, less the similarity, as shown in Definition 1.

$$u_{ij}=\frac{1}{\sum\limits_{k=1}^{C}(\frac{MCD(x_i,c_j)}{MCD(x_i,c_j)})^{\frac{2}{m-1}}} \tag{11}$$

$$c_j=\frac{\sum\limits_{i=1}^{N}{u_{ij}}^m\cdot x_i}{\sum\limits_{i=1}^{N}{u_{ij}}^m} \tag{12}$$

As mentioned above, Eq.(12) is derived directly from Eq.(5). The existence of trigonometric function in MCD makes it impossible to deduce Eq.(12), for Eq.(5) will be different when distance metric is not the linear transformation of Euclidean metric. Therefore, the FCM introduced in work [13] sometimes cannot converge to a optimum result. And a modified FCM is needed to cluster MPCs in a proper way.

C. Cluster Validity Indices

Cluster validity indices are used to evaluate the performance of clustering algorithm. This paper adopts XB and GD to assess the clustering algorithm which are found most appropriate to evaluate performance in MPCs clustering [15].

1) Generalized Dunn's Index: Generalized Dunn's Index (GD) is defined as the quotient between a minimum distance involving two clusters and a maximum distance involving one cluster.

$$V_{D_{ij}}=\frac{\min\limits_{k1,k2}\delta_i(k1,k2)}{\max\limits_{k}\Delta_j(k)} \tag{13}$$

Where i and j represent different functional forms. This work uses $i=5$, $j=3$ by

$$\delta_5=\frac{1}{L_{k1}+L_{k2}}(\sum_{\ell=1}^{L_{k1}}MCD(x_\ell,c_{k1})+\sum_{m=1}^{L_{k2}}MCD(x_m,c_{k2})) \tag{14}$$

$$\Delta_3=2(\sum_{\ell=1}^{L_k}MCD(x_\ell,c_k)) \tag{15}$$

The higher value GD is , the better clustering the algorithm obtains.

2) Xie-beni Index: Xie-beni (XB) Index represents the clusters' compactness to separation ratio, which is defined by

$$V_{XB}=\frac{\sum\limits_{k=1}^{K}\sum\limits_{\ell=1}^{L_k}MCD(x_\ell,c_k)^2}{L\times\left[\min\limits_{k1,k2}(MCD(x_\ell,c_k)^2)\right]} \tag{16}$$

On the contrary to GD index, the most desirable solution is obtained by minimizing the XB index.

Definition 1. *The multi-path components distance (MCD) is definitied by:*

$$MCD_{ij} = \sqrt{\|MCD_{AoA,ij}\|^2 + \|MCD_{AoD,ij}\|^2 + MCD^2_{\tau,ij}} \quad (8)$$

$$MCD^2_{\tau,ij} = \frac{\tau_{std}}{\Delta\tau_{\max}{}^2} \cdot \|\tau_i - \tau_j\| \quad (9)$$

$$MCD_{AOA,ij/AoD,ij} = \frac{1}{2}\left|\begin{pmatrix} \sin\varphi_i \cdot \cos\theta_i \\ \sin\varphi_i \cdot \sin\theta_i \\ \cos\varphi_i \end{pmatrix} - \begin{pmatrix} \sin\varphi_j \cdot \cos\theta_j \\ \sin\varphi_j \cdot \sin\theta_j \\ \cos\varphi_j \end{pmatrix}\right| \quad (10)$$

3) Decision Rank Fusion Method(Kr): Because the information of cluster number k is always unknown, the typical way is to obtain solutions under a range of k values then select an optimal k value according to the indices. Decision Rank Fusion Method(Kr) is proposed by work [15] to select a best cluster number, which is proved to be robust and effective. Its main idea is to combine the rank generated by the above two indexes. Each index rank different k values, assigning scores from 1 to the total number of k values. Adding all scores up, the best k value is whose score is highest.

III. Proposed Methodology

In this section, modified FCM algorithm and a new framework for clustering MPCs will be presented.

A. Modified Fuzzy-C means Algorithm

Consider an MPC is represented by its Power P and a parameter vector $x = [\tau, \varphi_{AoA}, \theta_{AoA}, \varphi_{DoA}, \theta_{DoA}]$, where τ is the delay of the MPC, and φ is the elevation angle, θ is the Azimuth angle.

Definition 2. *Let M be a map function, which maps the Multi-path Components (MPCs) from nolinear domain to linear domain*

$$x \to M(x) \quad (17)$$

where x is the vector represents a MPC and M (x) is

$$M(x) = \begin{bmatrix} \tau \\ \sin\varphi_{AoA} \cdot \cos\theta_{AoA} \\ \sin\varphi_{AoA} \cdot \sin\theta_{AoA} \\ \cos\varphi_{AoA} \\ \sin\varphi_{DoA} \cdot \cos\theta_{DoA} \\ \sin\varphi_{DoA} \cdot \sin\theta_{DoA} \\ \cos\varphi_{DoA} \end{bmatrix} \quad (18)$$

With the M function, the distance function between an MPC and a Centroids becomes

$$D = u^2_{ij} \cdot P_i \cdot \|M(x_i) - M(c_j)\|^2 \quad (19)$$

which is equivalent to the FCM distance using MPC in work [13], which means the following Equation is satified:

$$\|M(x_i) - M(c_j)\| = MCD(x_i, c_j) \quad (20)$$

Thus, the object function of FCM becomes:

$$J_m = \sum_{i=1}^{N}\sum_{j=1}^{C} u^m_{ij} \cdot P_i \cdot \|M(x_i) - M(c_j)\|^2 \quad (21)$$

Where m is the superscript which usually takes value 2, u_{ij} is the membership belief of MPC x_i to centroid c_j, which should satisfy $\sum_{j=1}^{C} u_{ij} = 1$

Hence, the map function converges the distance between MPCs to be calculated under Euclidean metric. The calculation for u_{ij} and c_j can be obtained by:

$$u_{ij} = \frac{1}{\sum_{k=1}^{C} \left(\frac{MCD(x_i,c_j)}{MCD(x_i,c_j)}\right)^{\frac{2}{m-1}}} \quad (22)$$

$$M(c_j) = \frac{\sum_{i=1}^{N} u^m_{ij} \cdot P_i \cdot M(x_i)}{\sum_{i=1}^{N} u^m_{ij} \cdot P_i} \quad (23)$$

The algorithm is shown as Algorithm 1.

Algorithm 1 Modified FCM

Input: MPCs x and their Power P, Superscript m, initial membership matrix u or Centroids matrix c

Output: Membership matrix u initial Centroids c

```
while u is changing or c is changing do
    recalculate c according to Eq.(23)
    recalculate u according to Eq.(22)
end while
```

B. The new Framework based on Modified FCM

The framework based on Modified FCM is presented as Algorithm 2. The framework includes three parts, initializing centroids c, clustering MPCs and Censor Process.

1) Main Process: k is iterated from 2 to the half of the number of MPCs. For each k, firstly obtain initial centroids through InitialGuess process and then cluster MPCs using modified FCM and use validate index to assess the clustering performance. This paper adopts Decision Rank Fusion Method (Kr) [15] as validate index. The main idea of Kr method is to rank the cluster number$k$ according to Generalized Dunn's index and Xie-beni(XB) index, which is detailed described in

Algorithm 2 the framework of clustering MPCs using modified FCM

Input: MPCs x and their Power, Superscript m, number of MPCs N

Output: the optimum cluster number k, Membership matrix u, Centroids c

for $k = 2$ to $N/2$ **do**
  c=InitialCentroidsGusse(x,k)
  x,u=Modified FCM(x,p,c,m)
  Censor Process(x,u)
  save k and corresponding solution
end for
applying Decision Rank Fusion Method(Kr) to rank every k value
return the optimum k value and its corresponding solution

Algorithm 3 Censor Process

Input: MPCs x, membership matrix u,threshold

Output: crisp membership information

repeat
  assign the MPCs to the centroids if the corresponding u > t
until go through all the MPCs

work [15] . The final number of clusters is selected from the k with highest rank.

2) Initial Centroids Guess Process: The centroids initializing process is detailed described in work [13]. And work [16] proposes modification on it. The main idea is firstly selecting a MPC with the strongest power as first centroid. Then choosing MPCs with maximum MCD to all the centroids as the next centroids until k centroids has been selected.

3) Censor Process: Censor Process is a defuzzy process. The main idea is to cluster the MPCs according to their membership belief. t is a threshold parameter. Only membership belief of a MPC to a centroid bigger than t, can this MPC be allocated to the centroid. If a centroid whose membership belief to all centroids are lower than t, the MPC will be seemed as noise path which will be eliminated.

IV. Comparison and Validation

In this section, Modefied FCM and its new framework will be compared with former FCM [13] and KPowerMeans framework respectively.

1) Comparison between Modified FCM and the former FCM: The former FCM is found incomplete mathematically in Section II-B. The experiment also shows the former FCM have trouble converging to the optimum solution after several iteration. In this experiment, we randomly select a snapshot from measurement data and cluster it with two methods. Centroids are generated randomly for both methods and the membership belief matrices generated by both methods in each iteration are recorded. Both methods converge to a final solution after 6 iterations, and the membership belief matrices of iteration 1, 2 and 6 are selected to show the difference of the two method, as shown in Fig.1. The 3 matrices in the first row are membership belief matrices of the former FCM and the three matrices below are membership belief matrices of Modified FCM. It is clear that the former FCM fails to cluster MPCs, all membership being the same value $\frac{1}{3}$ after 6 iterations while Modified FCM converges to a better solution, the first MPC belonging to the 3rd cluster and the rest MPCs belonging to the 2nd cluster.

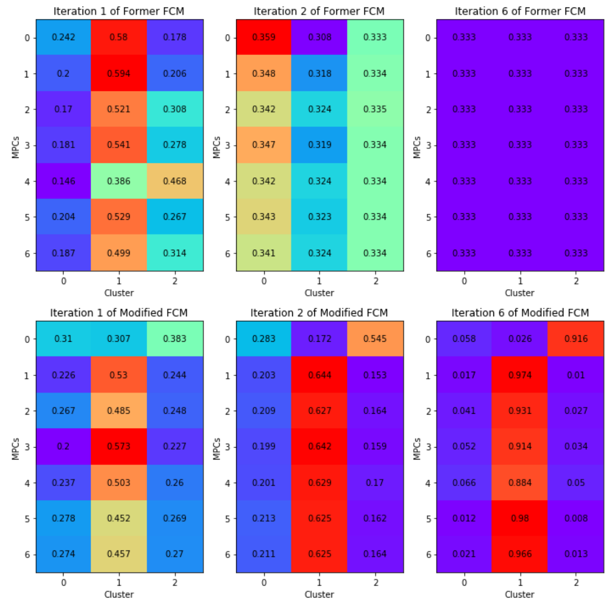

Fig. 1. The evolution of membership belief matrix (the first row is the membership belief matrix generated by the former FCM and the second is generated by the modified FCM)

2) Comparison between New Framework and KPowerMeans: In this part, the new framework based on Modified FCM is compared with KPowerMeans. GD index and XB index are proved to be the most accurate index tools to evaluate the the performance of clustering MPCs [15]. As described in II-C1 and II-C2, with higher GD and lower XB, the clustering performance is better. A snapshot was randomly chosen to test the performance of two Frameworks and the results are shown in Fig.2 and Fig.3. No matter what the number of clusters is, the GD value of Modified FCM is always above the GD value of KPowerMeans and the XB value of Modified FCM is always below the XB value of KPowerMeans. Moreover, Fig.4 shows the average XB and GD value of all the snapshots clustered by Modified FCM and KPowerMeans. From these results, it can be concluded that Modified FCM has better cluster performance than KPowerMeans did.

Also, the new framework based on Modified FCM shows good effectiveness on delineating noise MPCs. As Fig.5 and fig.6 show, the result generated by the new framework is clearer with fewer overlaps than the result generated by KPowerMeans.

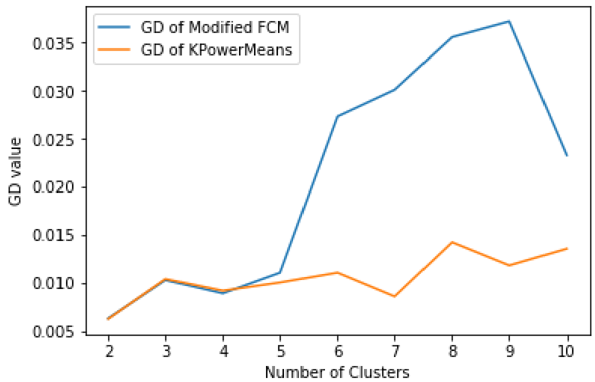

Fig. 2. GD index value of Modified FCM and KPowerMeans

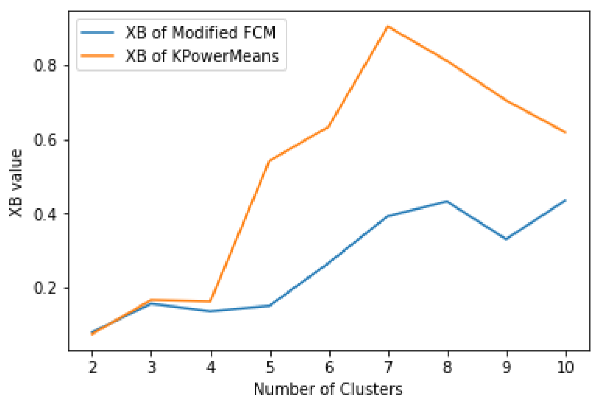

Fig. 3. XB index value of Modified FCM and KPowerMeans

V. Conclusion

Clustering MPCs is an essential process in many modern radio channel models. Appropriately clustered MPCs enable the model to be more accurate and effective. This paper introduces map functions to modify the former FCM [13] used in clustering MPCs, which enables FCM algorithm to converge to an optimum solution. In addition, a new framework based on Modified MPC is presented. In the new framework, InitialGuess technique [13], [15] is adopted to generate initial centroids, Decision Rank Fusion method [15] is used to select the optimum number of clusters, and CensorProcess is proposed to censor noise MPCs and allocate MPCs to a cluster according to membership belief matrix. Moreover, the comparison between Modified FCM and the former FCM algorithm shows an apparent improvement in Modified FCM to converge to an optimum solution. The new framework is tested with 300 snapshots collected from measurement campaign to show its effectiveness. The comparison results prove the proposed algorithm and framework are eligible to handle the clustering process in radio channel modelling. Because Modified FCM generates the membership belief to each cluster, this special property offers advantages for further research. For instance, the noise MPCs can be detected through the membership belief matrix by a more accurate method. And several snapshots containing similar MPCs can be fused to generate more desirable clusters with less noise. Therefore, Modified FCM is a prospective algorithm to cluster MPCs in radio channel modelling.

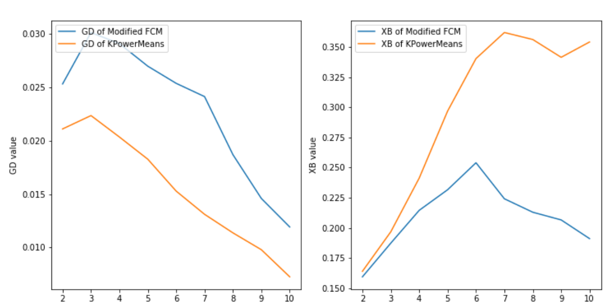

Fig. 4. Average GD and XB value of Modified FCM and KPowerMeans

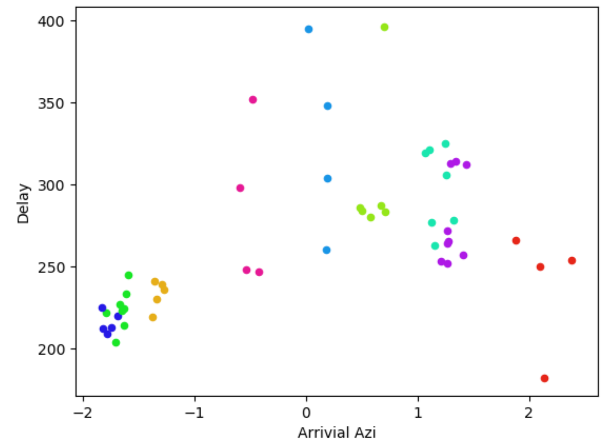

Fig. 5. Clusters generated by the new framework based on Modified FCM

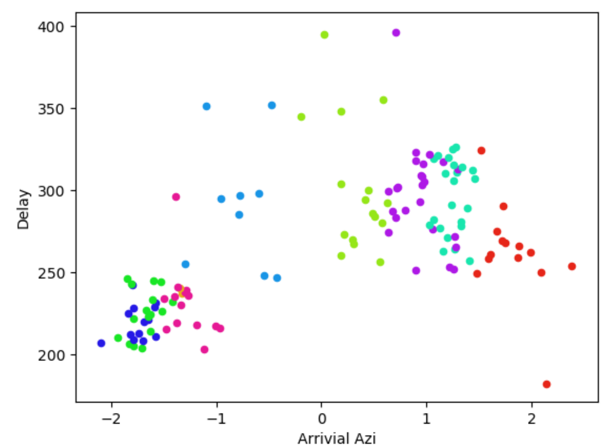

Fig. 6. Clusters generated by KPowerMeans

REFERENCES

[1] A. F. Molisch and F. Tufvesson, “Propagation channel models for next-generation wireless communications systems,” *IEICE Transactions on Communications*, vol. 97, no. 10, pp. 2022–2034, 2014.

[2] L. Liu, C. Oestges, J. Poutanen, K. Haneda, P. Vainikainen, F. Quitin, F. Tufvesson, and P. De Doncker, “The cost 2100 mimo channel model,” *IEEE Wireless Communications*, vol. 19, no. 6, pp. 92–99, 2012.

[3] A. A. Saleh and R. Valenzuela, “A statistical model for indoor multipath propagation,” *IEEE Journal on selected areas in communications*, vol. 5, no. 2, pp. 128–137, 1987.

[4] R. Zhang, S. Wang, X. Lu, W. Duan, and L. Cai, “Two-dimensional doa estimation for multipath propagation characterization using the array response of pn-sequences,” *IEEE Transactions on Wireless Communications*, vol. 15, no. 1, pp. 341–356, 2016.

[5] M. Steinbauer, H. Ozcelik, H. Hofstetter, C. F. Mecklenbrauker, and E. Bonek, “How to quantify multipath separation,” *IEICE transactions on electronics*, vol. 85, no. 3, pp. 552–557, 2002.

[6] N. Czink, P. Cera, J. Salo, E. Bonek, J.-P. Nuutinen, and J. Ylitalo, “A framework for automatic clustering of parametric mimo channel data including path powers,” in *Vehicular Technology Conference, 2006. VTC-2006 Fall. 2006 IEEE 64th*. IEEE, 2006, pp. 1–5.

[7] H. Xiao, A. G. Burr, L. Hentila, and P. Kyosti, “Statistical technique to identify clusters from multi-dimensional measurement data,” in *European Conference on Antennas and Propagation*, 2007, pp. 1–7.

[8] R. He, Q. Li, B. Ai, Y. L.-A. Geng, A. F. Molisch, V. Kristem, Z. Zhong, and J. Yu, “An automatic clustering algorithm for multipath components based on kernel-power-density,” in *Wireless Communications and Networking Conference (WCNC), 2017 IEEE*. IEEE, 2017, pp. 1–6.

[9] J. Chuang, S. Bashir, and D. G. Michelson, “Automated identification of clusters in uwb channel impulse responses,” in *Electrical and Computer Engineering, 2007. CCECE 2007. Canadian Conference on*. IEEE, 2007, pp. 761–764.

[10] C. Gentile, “Using the kurtosis measure to identify clusters in wireless channel impulse responses,” *IEEE Transactions on Antennas and Propagation*, vol. 61, no. 6, pp. 3392–3395, 2013.

[11] R. He, W. Chen, B. Ai, A. F. Molisch, W. Wang, Z. Zhong, J. Yu, and S. Sangodoyin, “A sparsity-based clustering framework for radio channel impulse responses,” in *Vehicular Technology Conference (VTC Spring), 2016 IEEE 83rd*. IEEE, 2016, pp. 1–5.

[12] ——, “On the clustering of radio channel impulse responses using sparsity-based methods,” *IEEE Transactions on Antennas and Propagation*, vol. 64, no. 6, pp. 2465–2474, 2016.

[13] C. Schneider, M. Bauer, M. Narandzic, W. T. Kotterman, and R. S. Thoma, “Clustering of mimo channel parameters-performance comparison,” in *Vehicular Technology Conference, 2009. VTC Spring 2009. IEEE 69th*. IEEE, 2009, pp. 1–5.

[14] M. Steinbauer, A. F. Molisch, and E. Bonek, “The double-directional radio channel,” *IEEE Antennas and propagation Magazine*, vol. 43, no. 4, pp. 51–63, 2001.

[15] S. Mota, F. Perez-Fontan, and A. Rocha, “Estimation of the number of clusters in multipath radio channel data sets,” *IEEE Transactions on Antennas and Propagation*, vol. 61, no. 5, pp. 2879–2883, 2013.

[16] S. Mota, M. O. Garcia, A. Rocha, and F. Pérez-Fontán, “Clustering of the multipath radio channel parameters,” in *Antennas and Propagation (EUCAP), Proceedings of the 5th European Conference on*. IEEE, 2011, pp. 3232–3236.