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 channels. Clustering MPCs becomes a critical process in many advanced channel models. In this paper, the Fuzzy C-means (FCM) algorithm, one of the methods that have once been introduced to group MPCs automatically, is modified with a map function to cluster MPCs more effectively. Meanwhile, a new framework based on the modified FCM is proposed. Compared with the former FCM and KPowerMeans framework, the proposed framework is tested to have a better clustering performance and provides basis for further research.**

I. Introduction

Results from lots of channel measurements have proven that multipath components (MPCs) are not independent to each other, but they are distributed in groups [1], known as clusters. MPCs in the same cluster share similar parameters like Power, Delay, Angle of Arrival(AoA), and Angle of Departure [2]. Besides, with the development of radio channel model, the concept of cluster has been adopted by many advanced channel models, such as Saleh-Valenzuela (SV) [3] model, WINNER channel model, COST2100 [4]. Establishing those models requires the prior knowledge of the clusters of MPCs in a radio channel.

Many contributions have been made to cluster MPCs automatically and effectively. 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 framework, which was proposed by Czink $et\ al.$ [6]. It is based on KMeans algorithm and taking account of power factor. Following KPowerMeans, a statistical technique to determine clusters in Channel Impulse responses(CIRs) by dividing the data into multidimensional analysis regions was presented by Xiao $et\ al.$ [7]. 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 exponentially 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 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]. Similarly, 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.

Schneider $et\ al.$ introduced Fuzzy C-means (FCM) into clustering MPCs [13] in radio channels. It is proved that the FCM algorithm outperforms KPowerMeans with random initialization. It is a good start of introducing the soft cluster algorithm into this field, but sometimes the algorithm seems hard converging to an optimum solution. In this paper, the failure situation of the FCM algorithm is presented and analyzed. In addition, a map function is proposed to modify FCM algorithm, greatly improving the clustering performance. At last, a novel framework based on the modified FCM is given, which incorporates the initialization process, clustering performance indices and pruning process. The comparison between KPowerMeans and the 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. Section III will give information about the modifications on FCM algorithm and the novel framework. The comparison with KPowerMeans will be presented in Section IV. The final section concludes the paper.

II. Preliminaries

A. Channel Model

A double-directional channel model [14] is considered in this work. It is a geometric based model which takes the clusters of MPCs into consideration. According to its definition, the MPCs in a cluster share similar parameters such as Delays, Powers, Angle of Departure(AoD), and Angle of Arrival(AoA). The Channel Impulse Response (CIR) of the model is established on the concept of clusters of MPCs, 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. 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. 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 that 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, 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 as follows,

1) Initialize the membership matrix u_{ij} or the centroids vectors c_j
2) Iterating Eq.(6) and Eq.(7) to converge to an optimal solution

[13] introduces FCM to cluster MPCs. It replaces the term $\|x_i-c_j\|$ with $MCD(x_i,c_j)$ in Eq.(11) and Eq.(12). MCD is a distance metric, measuring the similarities between MPCs. Greater the MCD, less the smilarity, as shown in Definition 1.

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

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

However, the replacement makes the FCM algorithm theoretically challenging to converge to an optimum solution. The existence of trigonometric function in MCD makes it impossible to derive Eq.(12), for Eq.(5) will be different when the distance metric in Eq.(1) is replaced by MCD. Therefore, the FCM introduced in [13] sometimes cannot obtain an optimum result due to the mathematical problem. Modifications on the FCM is needed to cluster MPCs in a proper way.

C. Cluster Validity Indices

Cluster validity indices are often used to evaluate the performance of clustering algorithm. This paper adopts Xie-beni (XB) index and Generalized Dunn (GD) index to assess the clustering algorithm, which are found most appropriate to evaluate MPCs clustering algorithms [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 results the algorithm obtains.

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

$$MCD_{ij} = \sqrt{\|MCD_{AoA,ij}\|^2 + \|MCD_{AoD,ij}\|^2 + MCD_{\tau,ij}^2} \tag{8}$$

$$MCD_{\tau,ij}^2 = \frac{\tau_{std}}{\Delta\tau_{\max}{}^2} \cdot \|\tau_i - \tau_j\| \tag{9}$$

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

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

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

On the contrary to GD index, the most desirable solution is obtained when XB index is minimum.

3) Decision Rank Fusion Method(Kr): Decision Rank Fusion Method(Kr) is proposed by [15] to select the best cluster number. Because the information of cluster number k is always unknown, the typical way is to obtain a series of solutions for different cluster numbers in a set K then select the optimal number k according to the Kr method. Its main idea is to combine the rank generated by the above two indexes, shown as follows.

1) Initialize the cluster number set K
2) Obtain solutions for each cluster number k in the set K
3) Rank each k by GD index
4) Rank each k by XB index
5) Add two ranks up, the best cluster number k is the one who ranks first

III. Proposed Methodology

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

A. Modified Fuzzy-C means Algorithm

The existing FCM algorithm for clustering MPCs directly replaces the Euclidean distance metric with MCD metric, making the algorithm hard to converge to an optimum solution, as analyzed in Section II-B.

This work proposed a map function, transforming the distance between MPCs from MCD metric to Euclidean metric. In this way, the modified FCM algorithm is mathematically correct while the distance between MPCs remains unchanged.

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 vector x to the vector $M(x)$.*

$$x \to M(x) \tag{17}$$

where x is the parameter vector of an MPC and M (x) is

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

The distance function between an MPC and a Centroids becomes

$$D = u_{ij}^2 \cdot P_i \cdot \|M(x_i) - M(c_j)\|^2 \tag{19}$$

which is equivalent to the MCD metric in [13],

$$\|M(x_i) - M(c_j)\| = MCD(x_i, c_j) \tag{20}$$

Thus, the object function of the FCM becomes:

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

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

Hence, the map function converts the calculation of distance between MPCs from MCD metric to Euclidean distance metric. The u_{ij} and c_j can be obtained by Lagrange Multiplier method, shown as follows:

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

$$M(c_j) = \frac{\sum_{i=1}^{N} u_{ij}^m \cdot P_i \cdot M(x_i)}{\sum_{i=1}^{N} u_{ij}^m \cdot P_i} \tag{23}$$

The modified FCM algorithm is shown as Algorithm 1.

B. The new Framework based on Modified FCM

The framework based on the modified FCM is shown as Fig. 1. The framework mainly incorporates four parts, the initialization of Cluster Number Set K, Modified FCM, Censor Process, and the selection of cluster number. The algorithm for the whole framework is presented as Algorithm 2.

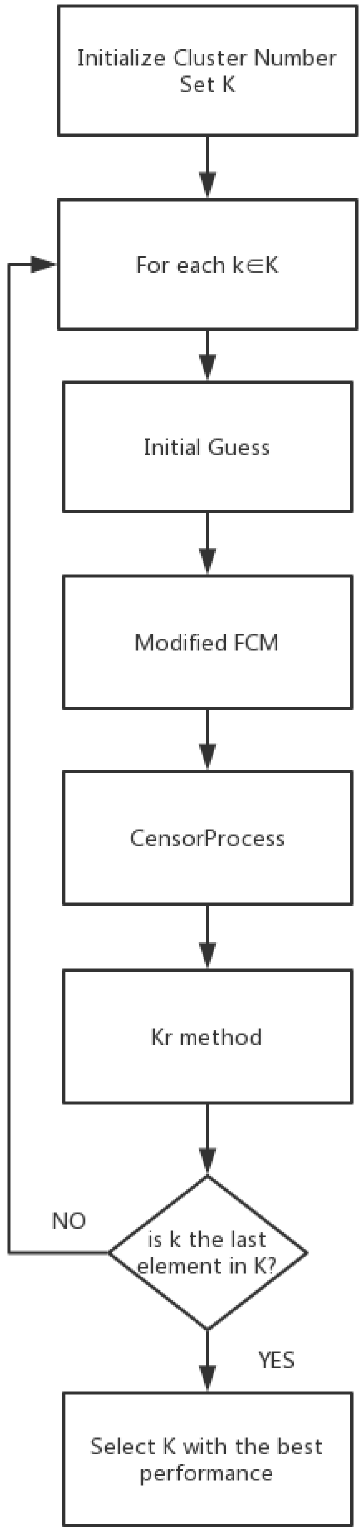

Fig. 1. The Modified FCM Framework

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

Algorithm 2 the framework of clustering MPCs using the modified FCM

Input: MPCs x and their Power, Superscript m, number of MPCs N, Cluster Number Set K

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

for k in K **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

1) Main Process: The user should specify the Cluster Number Set K. Usually, k is iterated from 2 to the half of the number of MPCs. For each k, InitialGuess generates the initial centroids matrix c for the modified FCM algorithm. After the modified FCM converges to an optimum solution, the Censor Process eliminates the noise components. Then the Decision Rank Fusion Method (Kr) [15] is adopted as validate index to rank the MPCs. The main idea of Kr method is to rank the cluster number k according to Generalized Dunn's index and Xie-beni(XB) index. The final number of clusters is selected from the k with the best rank.

2) Initial Centroids Guess Process: The InitialGuess process is firstly proposed by [13] and modified by [16]. The main idea of is as follows,

- Select an MPC with the strongest power as the first centroid
- Choose another MPC with maximum MCD to all the selected centroids as the next centroid
- Iterate with above two steps until the number of centroids is equal to k

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

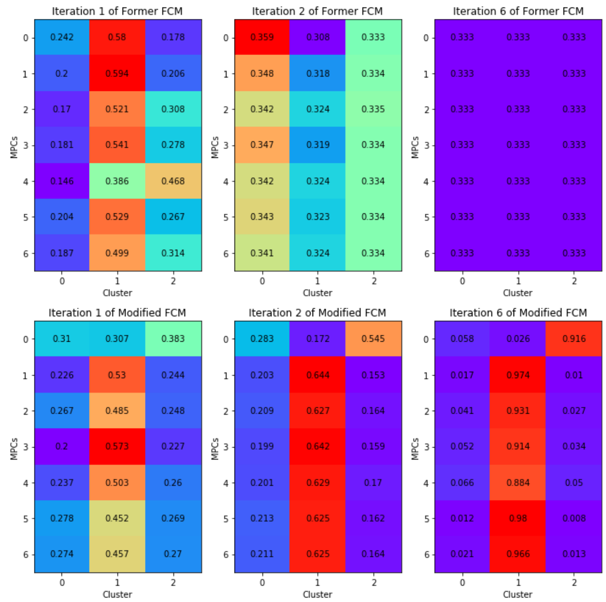

Fig. 2. The evolution of membership belief matrix (the first row is the membership belief matrix generated by the former FCM and the second is generated by the modified FCM)

3) Censor Process: Censor Process is used to diminish the influence of noise components. The solution generated by the modified FCM is the membership belief matrix of MPCs towards each cluster centroid. Let t be a threshold parameter. Only when membership belief of an MPC to a centroid bigger than t, can this MPC be allocated to this centroid. If an MPC whose membership belief to all centroids are lower than t, the MPC will be deemed as noise path and 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. The experimental data sets are generated from a simulated channel based on the 3GPP Spatial Channel Model Extended MIMO channel model [17].

1) Comparison between Modified FCM and the former FCM: The experiment shows the former FCM have trouble converging to the optimum solution after several iteration. In this experiment, we randomly select a snapshot from the data sets and apply 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.2. 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 the modified FCM converging to a better solution, where the first MPC belongs to the 3rd cluster and the rest MPCs belong to the 2nd cluster.

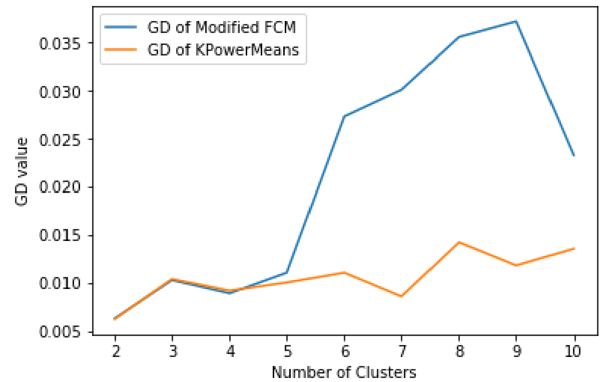

Fig. 3. GD index value of Modified FCM and KPowerMeans

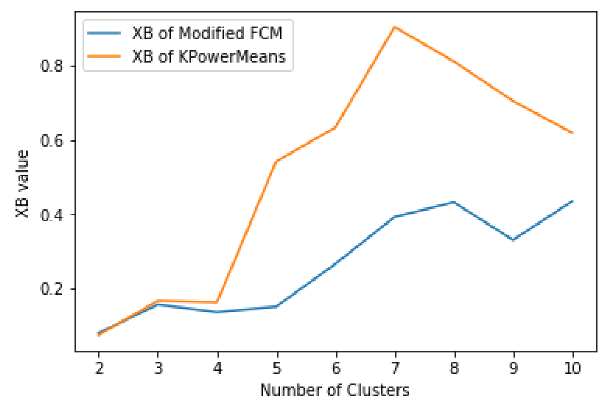

Fig. 4. XB index value of Modified FCM and KPowerMeans

2) Comparison between New Framework and KPowerMeans: In this part, the new framework based on the modified FCM is compared with KPowerMeans. GD index and XB index are proved to be the most accurate index tools evaluating the the performance of clustering MPCs [15]. As described in II-C1 and II-C2, with higher GD or lower XB, the performance of the algorithm is better. A snapshot is randomly chosen to test the performance of two Frameworks and the results are shown in Fig.3 and Fig.4. No matter what the number of clusters is, the GD value of the 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.5 shows the average XB and GD value of all the snapshots clustered by the modified FCM and KPowerMeans. From these results, it can be concluded that the modified FCM has better clustering performance than KPowerMeans did.

Also, the new framework based on the modified FCM shows good effectiveness on delineating noise MPCs. As Fig.6 and fig.7 show, the result generated by the new framework is clearer with fewer overlaps than the results generated by KPowerMeans.

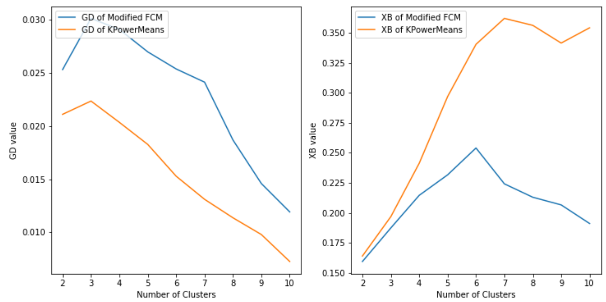

Fig. 5. Average GD and XB value of Modified FCM and KPowerMeans

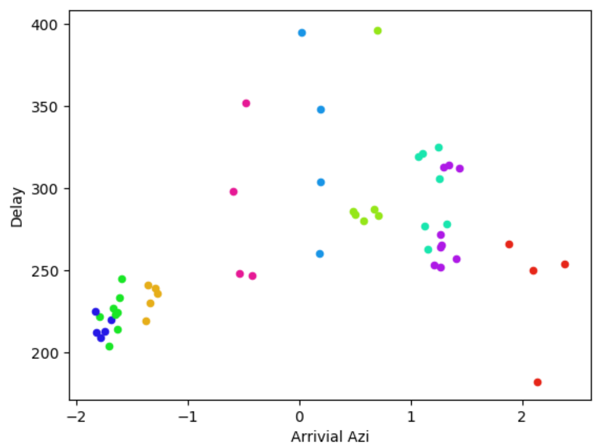

Fig. 6. Clusters generated by the new framework based on Modified FCM

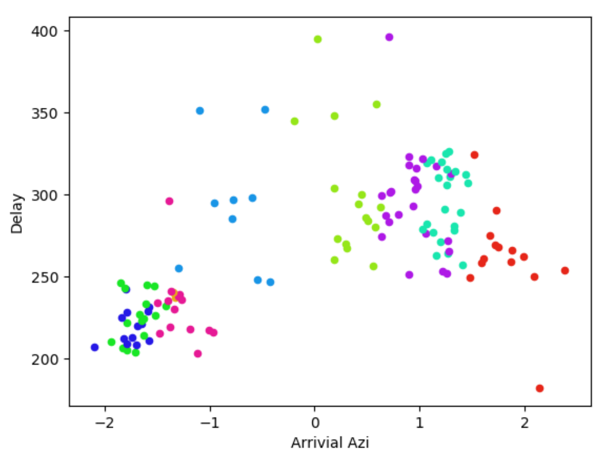

Fig. 7. Clusters generated by KPowerMeans

V. Conclusion

Clustering MPCs is an essential process in many modern radio channel models. Appropriately clustered MPCs establish a reliable and solid channel model. This paper introduces map functions to modify the former FCM [13] in clustering MPCs, enabling the FCM algorithm to acquire an optimum solution. In addition, a new framework based on the modified MPC is presented. The new framework adopts many advanced techniques that proposed recently, such as InitialGuess [13], [15] to generate initial cluster centroids, Decision Rank Fusion method [15] to select the optimum number of clusters. Censor-Process is proposed to censor noise components. Moreover, the comparison between the modified FCM and the former FCM algorithm shows an great improvement in converging to an optimum solution. The experiment results prove the proposed algorithm and framework are eligible to handle the clustering process in radio channel modelling. Because the modified FCM is a soft clustering algorithm, this special property also offers advantages for further research. For instance, Several snapshots containing similar MPCs can be easily fused to generate more desirable clusters with less noise. To sum up, the modified FCM algorithm is an effective tool in clustering MPCs.

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] 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.

[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] 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.

[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.

[17] D. S. Baum, J. Hansen, and J. Salo, "An interim channel model for beyond-3g systems: extending the 3gpp spatial channel model (scm)," in *Vehicular Technology Conference, 2005. Vtc 2005-Spring. 2005 IEEE*, 2005, pp. 3132–3136 Vol. 5.