Accuracy and Cluster Analysis of 5.3 GHz Indoor and 285 MHz Semi-urban MIMO LOS and NLOS Propagation Multipaths

Antipas T. Teologo Jr.[1,2]* and Lawrence Materum[1,3]

[1]De La Salle University, Philippines; Email: antipas_jr_teologo@dlsu.edu.ph
[2]FEU Institute of Technology, Philippines
[3]Tokyo City University, Japan
*Correspondence: antipas_jr_teologo@dlsu.edu.ph

***Abstract*—Over the past decade, several studies have been conducted to discover a better-performing multipath clustering technique. Developing a multipath clustering technique with better accuracy performance is a big challenge considering the varying properties of the multipath propagations that change over time. In this study, several clustering techniques have been investigated and compared to the newly-developed technique for performance analysis. Using the Jaccard score as a metric for the accuracy of grouping correctly the wireless multipaths, the performance of the different clustering techniques has been determined and compared to the newly-developed technique. The proposed clustering algorithm shows improved performance in the indoor channel scenarios but needs further investigation in the semi-urban environment.**

***Keywords*—radiowave propagation, wireless multipath, Jaccard score, clustering analysis**

I. Introduction

In this modern world, high-speed data transfer and high data rate wireless communications channels are necessary. There is always an increasing demand for high data rates and faster wireless communication, which requires developing an accurate channel model. Any wireless communications system's design and evaluation greatly depend on an accurate channel model.

Channel models play an important role in wireless communications systems, and having an accurate model can ensure the attainment of the target performance. However, achieving an accurate channel model is difficult as it needs the correct grouping of the various wireless multipaths. It is challenging to accurately group these multipaths because of their time-varying properties that need to be considered in the design. There is always a tradeoff between the complexity and accuracy of performance, and finding the best clustering technique is still needed.

Several studies have been conducted to develop an improved clustering technique. The studies of [1-3] utilized K-Power Means (KPM) framework to develop an improved clustering algorithm. Efforts have been made to optimize the various parameters of a multipath component (MPC) to develop an improved algorithm. These newly developed clustering approaches have provided some improvement but are still not enough for better performance. This state paved the way for more novel algorithms introduced in the studies of [4-10]. The study of [6] presented the sparsity-based channel impulse response (CIR) clustering algorithm that exploits the Saleh-Valenzuela (SV) model's feature. Compared with KMeans and KPM, this proposed algorithm has obtained the best result. Meanwhile, the study of [5] used the kernel density and the power of MPCs in clustering. This new technique only considers the K nearest MPCs and the relative density. Its performance was compared with KPM and density-based spatial clustering for applications with noise (DBSCAN) and showed the best performance.

On the other hand, the algorithm of [7] focused on the channel dynamics in the time-domain and clusters MPC at every snapshot, and finds the relationship between adjacent snapshots. Another novel algorithm is from the study of [8] which introduced a tracking-based MPC clustering technique that clusters MPCs based on their trajectories. Adding to the list of novel algorithms, also a machine learning-based clustering technique, that was presented in the study of [9] which investigated the grouping of MPCs using machine learning and also analyzed the characteristics of a cluster in a conventional scenario of a high-speed railway. In the study of [10], a general framework of the Mahalanobis-distance metric is used for the grouping of MPCs in MIMO channels.

These different proposed methods have their own advantages and limitations, especially concerning complexity. Most of these studies evaluated their performance by comparing only one or two clustering techniques. Choosing which among these clustering techniques is the best is still a challenge since the comparison is just limited. Moreover, many improvements are based only on a specific scenario or

Manuscript received July 1, 2022; revised November 21, 2022; accepted December 30, 2022.

few parameters. None of these studies have proved that it could be the best choice.

This study presents the accuracy performance evaluation and analysis of the new multipath clustering technique. Its accuracy is compared with the recent clustering techniques in their clustering category.

The rest of the paper is organized as follows. Section II discusses the methodology of the study, while Section III presents the data and results. Conclusion and recommendations are given in Section IV.

II. METHODOLOGY

A. The COST2100 Datasets

The Cooperation in Science and Technology (COST) 2100 channel model (C2CM) was utilized to create datasets for this investigation and implemented in Matlab. Supported in this C2CM are the eight channel scenarios generated using the various conditions. For bandwidth, it is either band 1 (B1) or band 2 (B2). The dataset is also classified as either single link or multiple links. For the network, indoor at 5.3 GHz or semi-urban at 285 MHz are the options. The datasets are also identified as either line-of-sight (LOS) or non-line-of-sight (NLOS).

From the eight different indoor and semi-urban channel environments, eight datasets were used. These datasets are available at the IEEE DataPort [11] and before being used by the various clustering approaches, they underwent preprocessing. The various pre-processing methods applied are clusterability, whitening transform (WT), and directional cosine transform (DCT).

The datasets are in Excel file format with 30 trials for each dataset. Each trial is recorded in an Excel sheet and has a size of N × D where N represents the number of propagation multipaths and D = 9 which consists of the triple cartesian components of the angle-of-arrival (AoA), triple cartesian components of angle-of-departure (AoD), relative power level, delay, and cluster affinity label/ID.

B. Multipath Clustering Techniques

Four novel clustering techniques were selected as the basis of comparison for the accuracy of performance evaluation. These are the Gaussian Mixture Model (GMM), K-Power Means (KPM), Ant Colony Optimization (ACO), and Kernel-Power Density-based Estimation (KPD). These clustering algorithms are the latest improvement in their respective clustering category and the reason why they are chosen to be the basis of comparison.

C. The Jaccard Score

To objectively evaluate how accurate the different algorithms cluster the multipaths, the Jaccard score, η_{jac}, is used as the validation index. Jaccard score is one of the several external comparison indices but in this study, only the Jaccard score is considered. The Jaccard score is 0 to 1, with 1 as the highest accuracy. The metric is given as follows [12]:

$$\eta_{\text{jac}} = \frac{n_{11}}{n_{11}+n_{10}+n_{01}} \tag{1}$$

where

n_{11} is the correct number of pairs both in the algorithm and in the reference

n_{01} is the correctly classified number of pairs in the algorithm but classified incorrectly in the reference

n_{10} is the incorrectly classified number of pairs in the algorithm but correctly classified in the reference

D. The Proposed Clustering Technique

The proposed and developed clustering technique is based on the KPM framework. The steps are as follows:

1. Initialize random K cluster centroids $\mu_1, \mu_2, \dots, \mu_K$, wherein from the data set Φ, the positions of K centroids are selected to be independent events.
2. Assign a particular weight to every feature or dimension of sample x of MPC. The weight can be obtained by running the dataset in principal component analysis (PCA).
3. Designate every weighted sample x of MPC to a particular cluster centroid μ_j: for every set x, define
$$c^{(k)} := \underset{j}{\operatorname{argmin}}\left\{\alpha_x \cdot d_{\text{MPC}}(x, \mu_j^{(k)})\right\} \tag{2}$$
where c acts as the store indices and superscript (k) denotes the iteration's number. For the d_{MPC} instead of the Euclidean distance, the Minkowski distance in Eq. (3) with the optimum p-value is used.
$$MD = \sqrt[p]{(x1-y1)^p + (x2-y2)^p + \ldots + (xN-yN)^p} \tag{3}$$
4. Change the centroids of the cluster: for each j, set
$$\mu_j^{(k+1)} := \frac{\sum_{x\in\Phi} 1\{c^{(k)}=j\}\alpha_x \cdot x}{\sum_{x\in\Phi} 1\{c^{(k)}=j\}\alpha_x} \tag{4}$$
5. Perform again steps 3 and 4 until the data has converged.

This proposed clustering technique is called the PCA-based Weighted KPM (PW-KPM).

E. Weight Determination Using PCA

PCA is one of the methods in analyzing the underlying information in a dataset. The important information of a multivariate data is being extracted using PCA. This essential information of the dataset is being converted into principal components with new set of variables. In this study, PCA is used as a basis in determining the weight that can be assigned to each feature of the original dataset to improve the accuracy performance of the clustering technique.

III. RESULTS AND DISCUSSION

A. Overall Comparison of the Mean Jaccard Indices

Table I presents the overall comparison of the average Jaccard indices of the five algorithms used in multipath clustering. The Jaccard scores of KPM, ACO, KPD, and

GMM were taken from the results of the study of [13], while the Jaccard scores for the indoor channel scenarios of PW-KPM were taken from the study of [14]. Fig. 1 presents the combined histogram plots of the various clustering techniques in each channel scenario.

TABLE I. OVERALL COMPARISON OF THE AVERAGE JACCARD INDICES

Channel Scenario	KPM	ACO	KPD	GMM	PW-KPM
Indoor, B1, LOS, Single Link	0.8915	0.2776	0.7459	0.2981	0.9471
Indoor, B2, LOS, Single Link	0.8446	0.2906	0.6389	0.2530	0.9481
Semi-urban, B1, LOS, Multiple Links	0.1206	0.0296	0.0547	0.0830	0.1146
Semi-urban, B1, LOS, Single Link	0.1190	0.0387	0.0882	0.0926	0.1174
Semi-urban, B1, NLOS, Single Link	0.1170	0.0176	0.0619	0.0932	0.1136
Semi-urban, B2, LOS, Multiple Links	0.1206	0.0290	0.0517	0.0856	0.1161
Semi-urban, B2, LOS, Single Link	0.1168	0.0373	0.0861	0.0834	0.1143
Semi-urban, B2, NLOS, Single Link	0.1162	0.0394	0.0598	0.0997	0.1133

It can be observed that PW-KPM leads indoor channel scenarios as can be seen in Fig. 1, with most of its Jaccard scores almost perfect. Out of the 30 trials in channel scenario 1, 26 of them obtained a Jaccard score of 1.0. In channel scenario 2, 27 out of 30 trials got a perfect Jaccard score. This shows that PW-KPM is consistent in its results for indoor environments. Other clustering techniques such as KPM and KPD also show high Jaccard scores in the indoor channels but are not consistently producing high scores in all the trials. In channel scenario 1, only 19 trials out of 30 obtained the perfect score in KPM while for the KPD, no trial has obtained a perfect Jaccard score. In channel scenario 2, 18 out of the 30 trials of KPM achieved a perfect score while KPD still did not obtain a perfect score in any of its trials. However, the majority of the trials in KPM and KPD obtained a Jaccard score of above 0.5 with only a few trials below 0.5 for both channel scenarios 1 and 2. Meanwhile, GMM and ACO Jaccard scores are below 0.5 most of the time with only a few trials that are above 0.5. In channel scenario 1, GMM only produced 1 trial with a 0.5 score and the rest are below 0.5. ACO, on the other hand, generated 2 trials with an above 0.5 score. Almost the same conditions happen in channel scenario 2. For GMM, only 2 trials are above 0.5 while for ACO, no trial is above 0.5 This makes either GMM or ACO not a good option for indoor channel scenarios.

For the semi-urban scenarios shown in Fig. 1, it can be noticed that KPM and PW-KPM have an almost similar distribution of Jaccard scores with most of the scores above 0.1. However, there is a noticeable decrease in the performance of KPM and PW-KPM when semi-urban scenarios are concerned. The only highest score obtained by KPM in a semi-urban environment is 0.1775 in channel scenario 4 while for PW-KPM, it is only 0.1543 also in channel scenario 4.

In the case of KPD and GMM, the same trend in the data distribution can be seen as they also exhibit low scores in the semi-urban channel environments with most of the scores below 0.1. On the other hand, ACO consistently obtains the lowest Jaccard scores in the semi-urban scenarios and produces low Jaccard scores in most trials. The majority of the trials produced by ACO in the semi-urban environments have values around 0.01 as can be seen in Fig. 1.

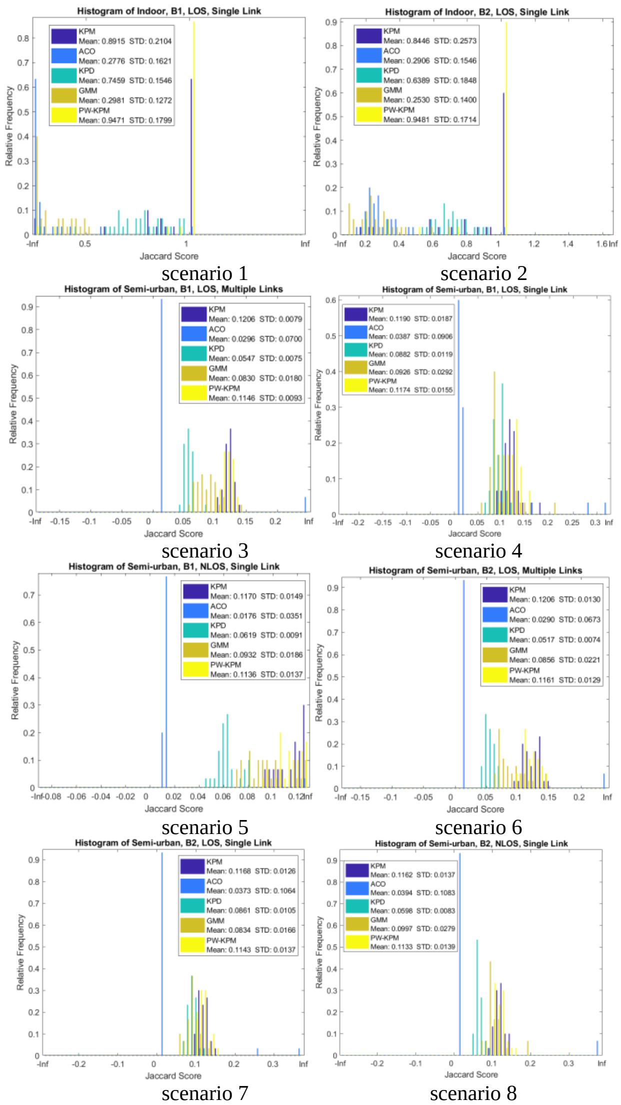

Figure 1. Combined histograms of the clustering techniques in channel scenario 1-8

B. Pairwise Comparison of the Jaccard Scores

Fig. 2 shows the accuracy performance of the five clustering algorithms using pairwise comparison. Group 1 is assigned for KPM, group 2 for ACO, group 3 for KPD, group 4 for GMM, and group 5 for the newly-developed PW-KPM. The pairwise comparison is made using one-way ANOVA.

It can be seen in Fig. 2 that even the PW-KPM obtains higher Jaccard scores than the KPM in the indoor channel scenarios, but statistically, they are equal as the comparison interval of their means overlap with each other. The average Jaccard score of the PW-KPM in channel scenario 1 is higher compared with KPM but the comparison interval for the Jaccard score mean of KPM, in gray bar, overlaps with the comparison interval for the Jaccard mean score of PW-KPM, the blue bar. The same condition can be observed in channel scenario 2. As compared with ACO, KPD, and GMM, PW-KPM is statistically different as can be seen from Fig. 2 where the blue bar does not overlap with the red bars.

The same condition applies in the semi-urban channel scenarios shown in Fig. 2. KPM, in gray bars, obtained slightly higher Jaccard scores compared with PW-KPM but when ANOVA is applied, the comparison interval of the Jaccard score mean of KPM always overlaps with that of PW-KPM, in blue bars. Therefore, statistically, KPM and PW-KPM can be considered equal in the accuracy performance of the semi-urban channel scenarios.

In the case of ACO, KPD, and GMM, they are statistically not equal to PW-KPM as their comparison intervals of mean do not overlap with that of PW-KPM as can be seen in channel scenarios 3, 5, and 6. In channel scenario 4, only the ACO is not statistically equal to the PW-KPM. KPD and GMM fall within the range of comparison interval of mean of PW-KPM thus there is no significant difference among them. The same situation can be observed in channel scenario 7 where only the ACO is not statistically equal with the other three clustering techniques while in channel scenario 8, only the GMM overlaps with PW-KPM.

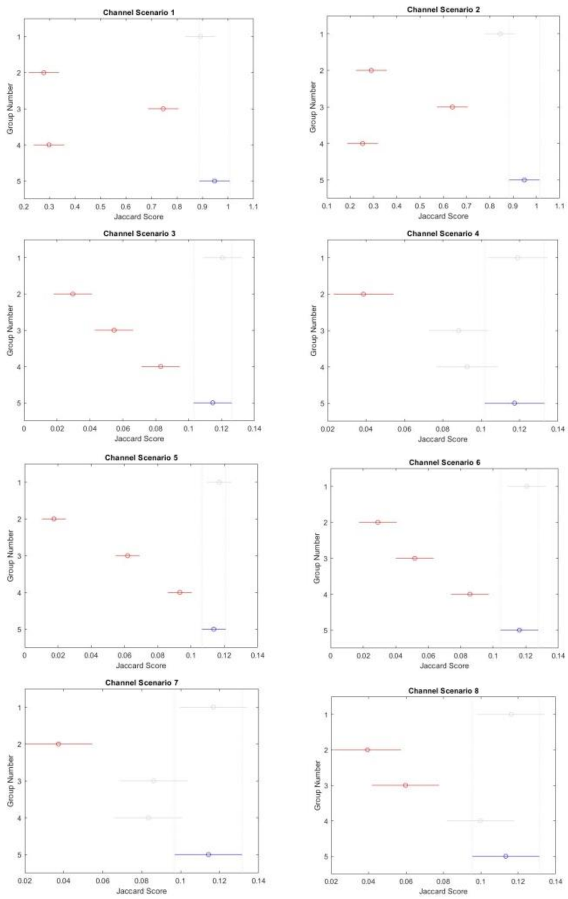

Figure 2. Pairwise comparison of the five clustering techniques in channel scenario 1-8

IV. CONCLUSION AND RECOMMENDATIONS

Clustering wireless multipaths is an essential aspect of channel modeling, and having an accurate channel model could lead to a reliable wireless communication system. This study found that PW-KPM, the new clustering technique, performed well in indoor channel scenarios, with a 5.56% increase in the accuracy in channel scenario 1 and a 10.35% increase in channel scenario 2 compared with the conventional KPM. However, it is worth noting that in semi-urban channel scenarios, there is a slight drop in the accuracy performance. This result is due to the fact that there are more scatterers in the semi-urban setting than in the indoor setup. This can have a major impact on the propagation multipath properties, making it challenging to cluster or group.

Although there is an improvement in the indoor scenarios, a deeper investigation of the semi-urban environment is recommended to develop a new clustering technique that can also improve the accuracy of the semi-urban channel scenarios. Also, it can be considered for future studies the comparison of the performance of the new method to the different deep learning-based algorithms and the use of other validation indices in evaluating the accuracy performance.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

A. Teologo conducted the research through the supervision of L. Materum. A. Teologo gathered the data and wrote the paper while L. Materum did the checking of the paper. All authors had approved the final version.

ACKNOWLEDGMENT

The Commission on Higher Education (CHED) and the De La Salle University (DLSU) are acknowledged for supporting this work. Special acknowledgment also goes to the Computing and Archiving Research Environment (COARE) of the Department of Science and Technology - Advanced Science and Technology Institute (DOST-ASTI) for the computing resources.

REFERENCES

[1] 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.

[2] P. Hanpinitsak, K. Saito, J. I. Takada, M. Kim, and L. Materum, “Multipath clustering and cluster tracking for geometry-based stochastic channel modeling,” *IEEE Transactions on Antennas and Propagation*, pp. 6015-6028, 2017.

[3] Y. Li, J. Zhang, Z. Ma, and Y. Zhang, “Clustering analysis in the wireless propagation channel with a variational gaussian mixture model,” IEEE Transactions on Big Data, 2018.

[4] B. Li, C. Zhao, H. Zhang, Z. Zhou, and A. Nallanathan, “Efficient and robust cluster identification for ultra-wideband propagations inspired by biological ant colony clustering,” *IEEE Trans. Commun.*, vol. 63, no. 1, pp. 286-300, 2015.

[5] R. He, Q. Li, B. Ai, Y. Geng, A. Molisch, V. Kristem, Z. Zhong, and J. Yu, “A kernel-power density-based algorithm for channel

multipath components clustering," *IEEE Trans. Wireless Commun.*, pp. 7138-7151, 2017.

[6] R. He, W. Chen, B. Ai, A. Molisch, W. Wang, Z. Zhong, J. Yu, and S. Sangodoyin, "On the clustering of radio channel impulse responses using sparsity-based methods," *IEEE Trans. Antennas Propag*, vol. 64, no. 6, pp. 2465-2474, 2016.

[7] Q. Wang, B. Ai, R. He, K. Guan, Y. Li, Z. Zhong, and G. Shi, "A framework of automatic clustering and tracking for time-variant multipath components," *IEEE Communications Letters*, vol. 21, no. 4, pp. 953-956, 2017.

[8] C. Huang, R. He, Z. Zhong, Y. A. Geng, Q. Li, and Z. Zhong, "A novel tracking-based multipath component clustering algorithm," *IEEE Antennas and Wireless Propagation Letters*, vol. 16, pp. 2679-2683, 2017.

[9] T. Zhou, Y. Qiao, S. Salous, L. Liu, and C. Tao, "Machine learning-based multipath components clustering and cluster characteristics analysis in high-speed railway scenarios," *IEEE Transactions on Antennas and Propagation*, vol. 70, no. 6, pp. 4027-4039, 2022.

[10] Y. Chen, C. Han, J. He, and G. Wang, "A framework of Mahalanobis-distance metric with supervised learning for clustering multipath components in MIMO channel analysis," *IEEE Transactions on Antennas and Propagation*, vol. 70, no. 6, pp. 4069-4081, 2022.

[11] J. Blanza, A. Teologo, and L. Materum, "Datasets for multipath clustering at 285 MHz and 5.3 GHz bands based on COST 2100 MIMO channel model," in *Proc. 9th Int. Symp. Multimedia Commun.*, 2019.

[12] R. Varshavsky, M. Linial, and D. Horn, "COMPACT: A comparative package for clustering assessment," in *Proc. Parallel Distrib. Process. Appl.*, 2005.

[13] A. T. Teologo Jr., L. Y. Materum, J. F. Blanza, and T. Hirano, "Comparative study of KPower Means, ant colony optimization, kernel power density-based estimation, and gaussian mixture model for wireless propagation multipath clustering," *International Journal of Emerging Trends in Engineering Research*, vol. 8, no. 7, pp. 3942-3950, 2020.

[14] A. T. Teologo Jr. and L. Y. Materum, "An improved K-Power means technique using Minkowski distance metric and dimension weights for clustering wireless multipaths in indoor channel scenarios," *Journal of Information and Communication Technology*, vol. 20, no. 4, pp. 541-563, 2021.

Antipas Teologo Jr. received his B.Sc. degree in Electronics and Communications Engineering from the Technological Institute of the Philippines – Manila under the Department of Science and Technology – Science Education Institute (DOST-SEI) scholarship grant. He finished his M.Sc. degree in Electronics and Communications Engineering major in communications at De La Salle University (DLSU) through the DOST-PCIEERD scholarship. Currently, he is taking his Ph.D. degree in Electronics and Communications Engineering major in communications at DLSU under the Commission on Higher Education (CHED) K-12 scholarship grant. His research interests include multipath clustering, wireless communications, MIMO systems, source localization, and direction-of-arrival estimation.

Lawrence Materum received the B.Sc. degree in electronics and communications engineering with honors from Mapúa Institute of Technology (now Mapúa University), the M.Sc. degree in electrical engineering major in computers and communications from the University of the Philippines Diliman through an Analog Devices Fellowship, and the Ph.D. degree in international development engineering (major in electrical/electronics engineering) from Tokyo Institute of Technology through a MEXT Scholarship.