ISSN 2347 - 3983
Volume 7, No. 8 August 2019
International Journal of Emerging Trends in Engineering Research
Available Online at http://www.warse.org/IJETER/static/pdf/file/ijeter16782019.pdf
https://doi.org/10.30534/ijeter/2019/16782019

Cluster-Wise Jaccard Accuracy of KPower Means on Multipath Datasets

Antipas T. Teologo Jr., Lawrence Materum
Electronics and Communications Engineering Department, De La Salle University, Philippines, antipas_jr_teologo@dlsu.edu.ph

ABSTRACT

This paper presents the accuracy performance of the KPower Means (KPM) algorithm in clustering wireless multipaths using the generated datasets from COST2100 channel model (C2CM). KPM is one of the popular techniques used to cluster wireless multipath components (MPCs) and has been a basis of other complex multipath clustering approaches. KPM is implemented in Matlab using eight different channel scenarios obtained from C2CM representing various indoor and semi-urban environments at 2.85 MHz and 5.3 GHz bands, respectively. Results show that KPM performs well in an indoor environment than in a semi-urban due to the presence of numerous scatterers in a semi-urban environment yielding more multipaths. Jaccard similarity index is used to validate the accuracy performance of the KPM.

Key words : channel models, clustering methods, COST 2100, Jaccard index, KPower Means, radiowave propagation

1. INTRODUCTION

Channel modeling plays a significant role in achieving a reliable wireless communication system. The cluster-based channel model is now gaining much attention than its non-clustered counterpart as it offers more potential in obtaining higher data rate which is an important consideration in many wireless communication systems such as the multiple-input multiple-output (MIMO), fourth-generation (4G), and fifth-generation (5G). Characterizing the multipath components (MPCs) in a wireless environment is the primary objective of channel modeling. To date, many widely used channel models such as the European Wireless World Initiative New Radio (WINNER), 3GPP Spatial Channel Model, European Cooperation in Science and Technology (COST) 259, and COST 2100 are mainly based on how MPCs are being clustered.

Clustering MPCs involves grouping MPCs with the same delay and angular parameters. Parameterization of cluster's number, position, delay, and angular spreads is vital if a cluster-based channel model is being considered. Either manual or automatic clustering techniques have been used to group MPCs. In the past, the manual method has been widely used [1], [2], [3], [4], [5], [6] through visual inspection but has posed limitations when high-dimensional data is involved. For this reason, automatic clustering has replaced the manual method. Various automatic clustering algorithms have been developed with improved performance in the clustering of high-dimensional data. However, because of the unpredictable nature of a wireless channel, automatic methods are still facing challenges in terms of accurately grouping MPCs. Some of these proposed automatic algorithms are the K-Means [7], [8], Fuzzy C-means [9], the density-based spatial clustering for applications with noise (DBSCAN) [10], hierarchical agglomerative clustering algorithm [11], ant colony clustering (ACC) [12], [13], kernel power density-based (KPD) algorithm [14], variational gaussian mixture model (VGMM) [15], and the improved version of K-Means which is the KPower Means (KPM) [16]. Most of these techniques, if not all, used datasets from measurements

This study will determine the performance of KPM using the datasets generated from the COST 2100 channel model (C2CM) with representations of indoor and semi-urban environments. Since KPM is a well-known algorithm in multipath clustering, it will be utilized in this study. Section II of this paper will talk about KPM, its framework, and its Matlab implementation. Section III gives an overview of the datasets. Section IV discusses the Jaccard index used in the validation and the obtained results are presented in Section V. Section VI gives the conclusion.

2. THE KPM ALGORITHM

KPM's main difference from its predecessor, K-Means, is basically the addition of MPCs power in the processing. Simply, KPM is identical to a K-Means technique with the inclusion of MPCs power and is one of the well-known clustering techniques. K-means is typically combined with the Euclidean metric in obtaining the distance between points and the centers of the clusters which is very helpful in obtaining the ball or spherical-shaped clusters as it makes the task a lot easier. However, the initialization with the number of clusters is required in K-Means which is a priori unknown, therefore K-Power Means has been introduced in several studies to solve this problem.

2.1 Related Works

Being a parameter-based automatic clustering algorithm [17], [18], KPM tries to minimize the various parameters of the MPCs that belong to the same cluster such as the propagation delay time (DT), the angle-of-departure (AoD) and the angle-of-arrival (AoA). Many studies have used KPM as their clustering technique and being compared with other clustering algorithms. In the study of [19], KPM was used to group the generated data from Saleh-Valenzuela (SV) model with the aid of the different initialization procedures. Meanwhile, KPM algorithm was utilized in the 60 GHz band channel model in the study of [20]. It was found out that a log-normal distribution can be used to model the changes on the peak power of cluster around the mean. Adding also in the list of related studies, [21] applied the KPM framework to group data generated from an indoor environment at 11 GHz MIMO channel. Clustering of data was done by measuring in the ray tracer and exploiting the scattering points (SPs) geometry. Recently, in the study of [15], a hybrid clustering approach was introduced. KPM combined with Space Alternating Generalized Expectation-Maximization (SAGE) algorithm was also utilized to cluster MPCs with an added tracking capability. This novel technique was able to group MPCs successfully and have captured as well all the clusters' characteristics. This novel approach was tested in a subway station scenario using the MIMO channel model.

2.2 The KPM Framework

KPM works on a principle of iteratively locating the centroids of the cluster. The distance of each MPC to its respective centroid is determined and by minimizing the total sum of these distances will determine the corresponding centroid of that given cluster to be its center of mass. Figure 1 illustrates the main concept of KPM. Below are the main steps of the framework [14]:

1. Randomly initialize K cluster centroids $\mu_1, \mu_2,\ldots,\mu_K$, i.e., the positions of the K centroids are chosen independently from the dataset $\boldsymbol{\Phi}$ as events of equal probability (without replacement).

2. Each MPC sample x is being assigned to the reasonable cluster centroid μ_j: for each set x, set
$$c^{(k)} := \arg\min_j \left\{ \alpha_x . d_{MPC}\left(x, \mu_j^k\right) \right\} \quad (1)$$
where superscript (k) represents the k-th iteration and c gives the store indices in the k-th iteration of the MPC clustering.

3. Update the cluster centroids: for each j, set
$$\mu_j^{(k+1)} := \frac{\sum_{x \in \Phi} 1\{c^{(k)} = j\} \alpha_x . x}{\sum_{x \in \Phi} 1\{c^{(k)} = j\} \alpha_x} \quad (2)$$

4. Repeat steps 2 and 3 until convergence.

KPM differs from the standard K-Means due to the power weighting in the determination of MPC distance, d_{MPC}.

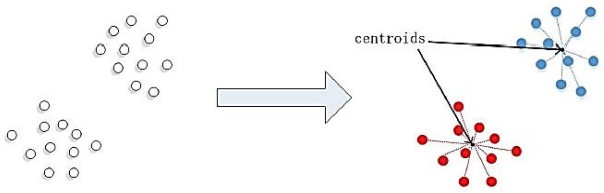

Figure 1: The main concept of the KPM algorithm [14].

2.3 Matlab Implementation

The first step of KPM requires the value of K, which is the maximum number of multipath clusters. In this study, it was set to half of the total number of clusters in order to avoid having multiple clusters with one or less member. Next, K number of centroids were generated containing the random value of MPC parameters but the same count (N) of parameters with multipath. To identify the groupings, the multipath component distance (MCD) between each path and the generated centroids was determined. The cluster IDs were calculated by the KPM by finding the least MCD of each path to a centroid. These generated cluster IDs determine the group or cluster of each path.

3. C2CM DATASETS

The datasets used in this study were taken from the IEEE Dataport [22]. It is composed of eight xlsx files, with each file representing different channel scenarios. Below are the eight datasets used:

1) Indoor, B1, line-of-sight, single link (Scenario 1)
2) Indoor, B2, line-of-sight, single link (Scenario 2)
3) Semi-urban, B1, line-of-sight, multiple links (Scenario 3)
4) Semi-urban, B1, line-of-sight, single link (Scenario 4)
5) Semi-urban, B1, non-line-of-sight, single link (Scenario 5)
6) Semi-urban, B2, line-of-sight. multiple links (Scenario 6)
7) Semi-urban, B2, line-of-sight, single link (Scenario 7)
8) Semi-urban, B2, non-line-of-sight, single link (Scenario 8)

These datasets were generated using the C2CM and were already pre-processed to eliminate ambiguities in the data. Pre-processing methods done include line-of-sight (LOS) removal, directional cosine transform, clusterability test, and whitening transform. Each file consists of 30 sets of data coming from the 30 trials. These data will also serve as the reference or ground truth data which is important in evaluating the performance of a clustering algorithm. One of the information in these datasets are the assigned cluster IDs for each multipath.

4. VALIDATION INDEX

To evaluate the accuracy of KPM, the Jaccard index, η_{jac}, was utilized. This Jaccard score or index reflects the "intersection over union" between the clustering results of KPM and generated clusters of the C2CM. Output values range from 0 (which means a void match) and 1 (which means a perfect match). This is to provide an objective way of showing the similarity and dissimilarity between the experimental and the simulated classification vectors. Metrics used are given as follows:

$$\eta_{jac} = \frac{n_{11}}{n_{11} + n_{10} + n_{01}} \quad (3)$$

Where

n_{11} is the number of pairs that are classified together correctly

n_{01} is the number of pairs which are not classified together correctly by the algorithm

n_{10} is the number of pairs which are incorrectly classified together when they are not supposed to

5. RESULTS

Figures 2 and 3 demonstrate the representation of the generated data in C2CM for the indoor and semi-urban scenario, respectively. Each group or cluster is assigned a certain color and it can be observed that an indoor environment produces a lesser number of clusters compared to a semi-urban environment. KPM takes these data as its input and Table 1 presents the computed cluster-wise Jaccard similarity index or score for each channel scenario using KPM clustering technique. These indices or scores provide information on how accurate the KPM algorithm in determining the correct cluster or group for each MPC. Description of each channel scenario is given in Section III. The Jaccard index shown for each scenario is the mean of the Jaccard indices for 30 trials. It can be seen from the table that KPM performed best in scenario 1 followed closely by scenario 2. Both scenarios represent the indoor environment. On the other hand, channel scenarios 3 to 8, which are the semi-urban environments, obtained low values of Jaccard index signifying poor performance of KPM.

Meanwhile, Figures 4 to 11 present the histogram plots of the eight channel scenarios. For scenario 1, 19 out 30 trials (63.33%) got a 1.0 Jaccard score and 28 out of 30 trials (93.33%) have Jaccard scores greater than 0.5. For scenario 2, 18 out of 30 trials (60.00%) obtained a Jaccard score of 1.0 and 27 out of 30 trials (90.00%) have Jaccard scores more than 0.5. The two indoor scenarios showed almost the same results. For scenarios 3 to 8, on the other hand, all calculated Jaccard indices are less than 0.5 with a range of values from 0.0893 to 0.1775.

Clusters can be detected easily in an indoor scenario which makes it easier for the KPM to find the centroid of each group or cluster unlike in the semi-urban setup wherein there is an overlapping of some clusters, making it difficult for the algorithm to detect the correct assigned cluster for each multipath. Overlapping of clusters in a semi-urban setting is expected as more scatterers or obstacles are present along its path.

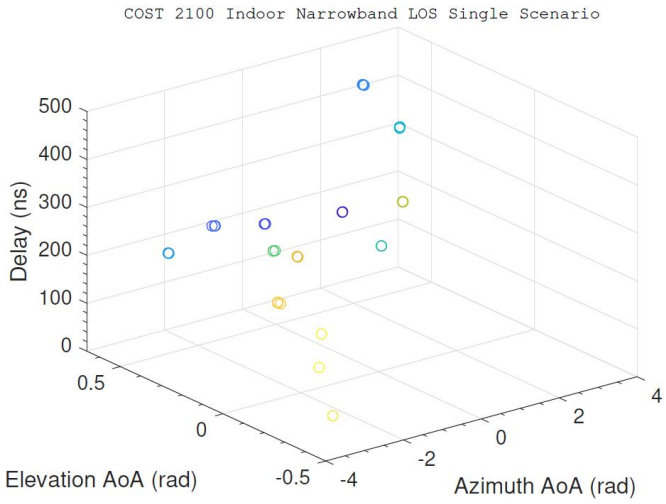

Figure 2: COST 2100 indoor channel scenario [22].

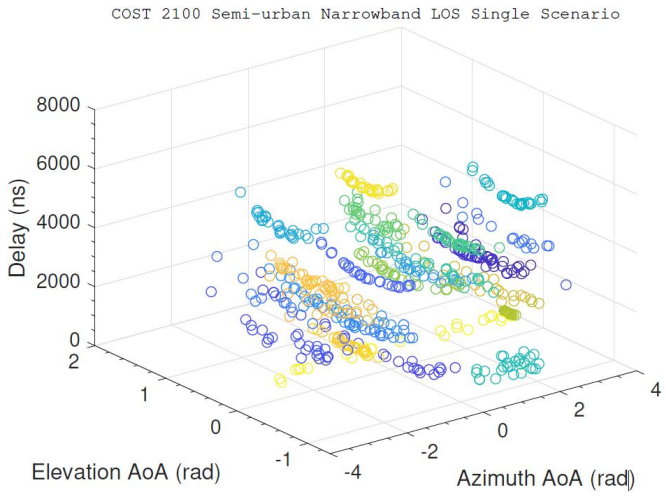

Figure 3: COST 2100 semi-urban channel scenario [22].

Table 1. The mean cluster-wise Jaccard score for each channel scenario using KPM.

Channel Scenario	η_{jac}
Scenario 1	0.8915
Scenario 2	0.8446
Scenario 3	0.1206
Scenario 4	0.1190
Scenario 5	0.1170
Scenario 6	0.1206
Scenario 7	0.1168
Scenario 8	0.1162

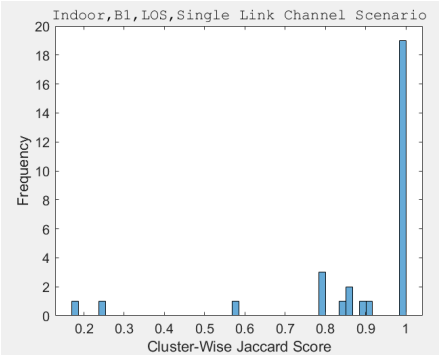

Figure 4. Histogram plot of Indoor, B1, LOS, Single Link Channel Scenario

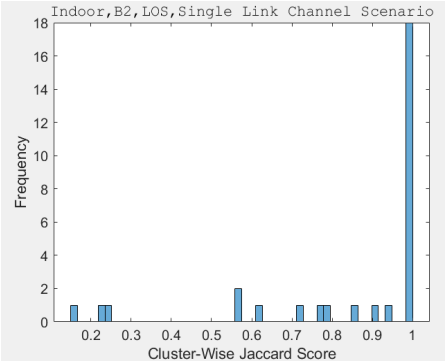

Figure 5. Histogram plot of Indoor, B2, LOS, Single Link Channel Scenario

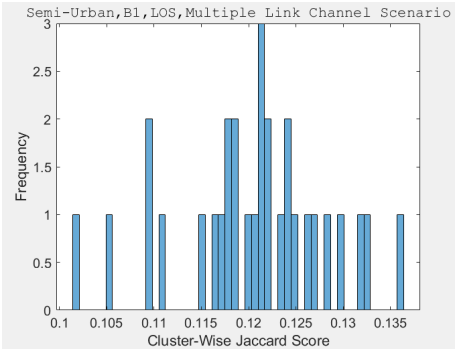

Figure 6. Histogram plot of Semi-Urban, B1, LOS, Multiple Link Channel Scenario

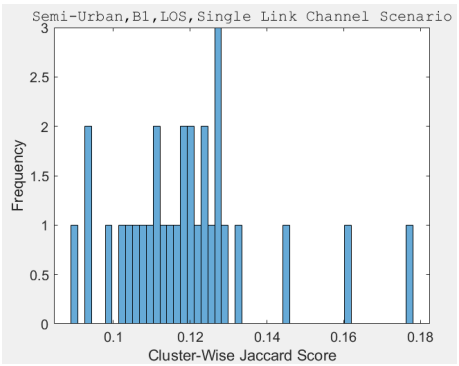

Figure 7. Histogram plot of Semi-Urban, B1, LOS, Single Link Channel Scenario

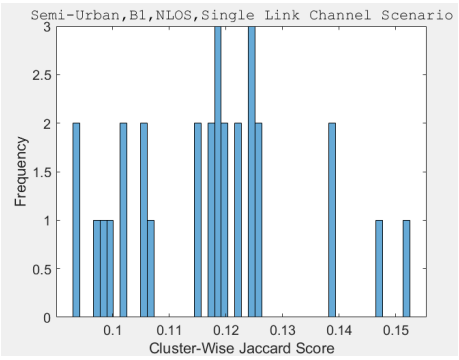

Figure 8. Histogram plot of Semi-Urban, B1, NLOS, Single Link Channel Scenario

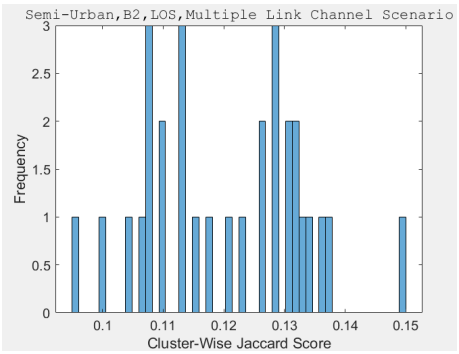

Figure 9. Histogram plot of Semi-Urban, B2, LOS, Multiple Link Channel Scenario

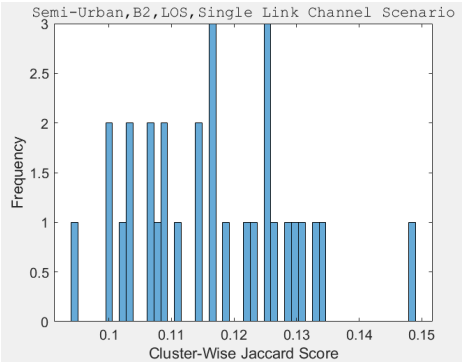

Figure 10. Histogram plot of Semi-Urban, B2, LOS, Single Link Channel Scenario

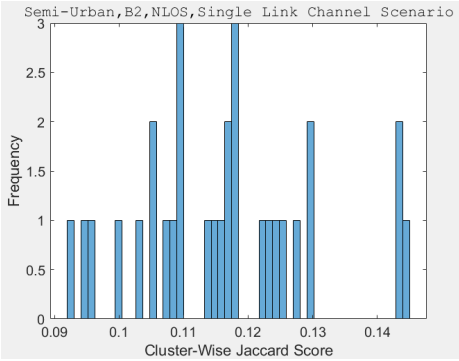

Figure 11. Histogram plot of Semi-Urban, B2, NLOS, Single Link Channel Scenario

6. CONCLUSIONS

Accuracy performance of KPM was evaluated in this study using the datasets generated from C2CM with eight different channel scenarios. KPM offers a promising performance with a mean accuracy of 86.81 percent for the two indoor environments but performs poorly if applied in a semi-urban setup which only achieved a mean accuracy of 11.84 percent for the six semi-urban scenarios. However, the obtained results in this study can be used as a comparison with other multipath clustering techniques using the same datasets in order to find which among of the current clustering techniques will have the best performance in terms of clustering wireless multipaths.

ACKNOWLEDGMENT

The authors would like to thank the Commission on Higher Education and the De La Salle University for the financial support in making this study a possible one.

REFERENCES

[1] M. Toeltsch, J. Laurila, K. Kalliola, A. F. Molisch, P. Vainikainen, and E. Bonek, "Statistical characterization of urban spatial radio channels," *IEEE Journal on selected areas in Communications*, vol. 20, no. 3, pp. 539–549, 2002. https://doi.org/10.1109/49.995513

[2] J. Laurila, K. Kalliola, M. Toeltsch, K. Hugl, P. Vainikainen, and E. Bonek, "Wideband 3D characterization of mobile radio channels in urban environment," *IEEE Transactions on antennas and propagation*, vol. 50, no. 2, pp. 233–243, 2002. https://doi.org/10.1109/8.998000

[3] K. Yu, Q. Li, and M. Ho, "Measurement investigation of tap and cluster angular spreads at 5.2 GHz," *IEEE Transactions on Antennas and Propagation*, vol. 53, no. 7, pp. 2156–2160, 2005. https://doi.org/10.1109/TAP.2005.850721

[4] L. Vuokko, P. Vainikainen, and J. Takada, "Clusters extracted from measured propagation channels in macrocellular environments," *IEEE Trans. Antennas Propag.*, vol. 53, no. 12, pp. 4089–4098, Dec. 2005. https://doi.org/10.1109/TAP.2005.859763

[5] C. Oestges and B. Clerckx, "Modeling outdoor macrocellular clusters based on 1.9-GHz experimental data," *IEEE Transactions on Vehicular Technology*, vol. 56, no. 5, pp. 2821–2830, 2007. https://doi.org/10.1109/TVT.2007.900391

[6] Q. Pei, B. Kang, L. Zhang, K. R. Choo, Y. Zhang, and Y. Sun, "Secure and privacy-preserving 3D vehicle positioning schemes for vehicular ad hoc network," *EURASIP Journal on Wireless Communications and Networking*, 2018. https://doi.org/10.1186/s13638-018-1289-9

[7] J. MacQueen, "Some methods for classification and analysis of multivariate observations," in *Proc. BSMSP*, 1967, pp. 281–297.

[8] B. Chandirika, N.K. Sakthivel, and S. Subasree, "An energy-efficient K-Means clustering based trust model for wireless sensor networks," *International Journal of Advanced Trends in Computer Science and Engineering*, vol. 8, no. 2, pp. 144-153, March-April 2019. https://doi.org/10.30534/ijatcse/2019/08822019

[9] C. Schneider, M. Bauer, M. Narandži´c, W. A. T.Kotterman, and R. S. Thomä, "Clustering of MIMO channel parameters—Performance comparison," in *Proc. IEEE VTC*, Apr. 2009, pp. 1–5. https://doi.org/10.1109/VETECS.2009.5073445

[10] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise," in *Proc. ACM KDD*, 1996, pp. 226–231.

[11] *A Hierarchical Agglomerative Clustering Algorithm for Channel Modelling*, document 3GPP TSG RAN WG1 R1-163115, 3GPP, 2016, pp. 1–6.

[12] 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 Transactions on Communications*, vol. 63, no. 1, pp. 286–300, 2015. https://doi.org/10.1109/TCOMM.2014.2377120

[13] D. Hema Latha and P. Premchand, "Estimating software reliability using ant colony optimization technique with salesman problem for software process," *International Journal of Advanced Trends in Computer Science and Engineering*, vol. 7, no. 2, pp. 20-29, March-April 2018. https://doi.org/10.30534/ijatcse/2018/04722018

[14] R. He, Q. Li, B. Ai, Y. L.-A. Geng, A. F. Molisch, V. Kristem, Z. Zhong, and J. Yu, "A kernel-power-density-based algorithm for channel multipath components clustering," *IEEE Transactions on Wireless Communications*, vol. 16, no. 11, pp. 7138–7151, 2017. https://doi.org/10.1109/TWC.2017.2740206

[15] 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. https://doi.org/10.1109/TBDATA.2018.2840696

[16] R. Xu and D. Wunsch, "Survey of clustering algorithms," *IEEE Trans. Neural Netw.*, vol. 16, pp. 645–678, May 2005. https://doi.org/10.1109/TNN.2005.845141

[17] 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 *IEEE Veh. Technol. Conf.*, 2006, pp. 1–5. https://doi.org/10.1109/VTCF.2006.35

[18] N. Czink, R. Tian, S. Wyne, F. Tufvesson, J. Nuutinen, J. Ylitalo, E. Bonek, and A. Molisch, "Tracking time-variant cluster parameters in MIMO channel measurements," in *2007 Second International Conference on Communications and Networking in China*, 2007. https://doi.org/10.1109/CHINACOM.2007.4469589

[19] 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. https://doi.org/10.1109/TAP.2013.2242823

[20] C. Gustafson, K. Haneda, S. Wyne, and F. Tufvesson, "On mm-wave multipath clustering and channel modeling," *IEEE Transactions on Antennas and Propagation*, vol. 62, no. 3, pp. 1445–1455, 2014. https://doi.org/10.1109/TAP.2013.2295836

[21] P. Hanpinitsak, K. Saito, J. Takada, M. Kim, and L. Materum, "Multipath Clustering and Cluster Tracking for Geometry-Based Stochastic Channel Modeling," *IEEE Transactions on Antennas and Propagation*, vol. 65, no. 11, pp. 6015–6028, 2017. https://doi.org/10.1109/TAP.2017.2754417

[22] J.F. Blanza, A.T. Teologo, Jr., L. Materum, "Datasets for Multipath Clustering at 285 MHz and 5.3 GHz Bands Based on COST 2100 MIMO Channel Model," *International Symposium on Multimedia and Communications*, Aug.2019.