Joint Identification of the Clustering and Cardinality of Wireless Propagation Multipaths

Jojo F. Blanza[1], Lawrence Materum[2]
[1, 2] Department of Electronics and Communications Engineering, De La Salle University, Manila, Philippines,
jojo_f_blanza@dlsu.edu.ph
[1] Electronics Engineering Department, University of Santo Tomas, Manila, Philippines, jfblanza@ust.edu.ph

ABSTRACT

COST 2100 channel model (C2CM) wireless propagation multipaths taken from IEEE DataPort are grouped using simultaneous clustering and model selection (SCAMS). SCAMS solves simultaneously the membership and the number of clusters in contrast to just the number of clusters popular clustering approaches provide. The membership and the number of clusters rely on λ and γ, the parameters that avoid the trivial solution of the affinity matrix. Jaccard index is used to assess the accuracy of SCAMS by comparing the clustered multipaths with the reference multipath datasets from IEEE DataPort.

Key words: channel models, clustering methods, data handling, data models, data preprocessing, multipath channels, radiowave propagation

1. INTRODUCTION

The stochastic properties of multiple-input multiple-output (MIMO) wireless propagation channels can be reproduced using the European Cooperation in Science and Technology (COST) 2100 channel model. Groups of multipath components with similar delay and angles called multipath clusters characterize C2CM. A multipath component (MPC) is distinguished according to its delay, angle of departure (Azimuth of Departure (AoD), Elevation of Departure (EoD)), and angle of arrival (Azimuth of Arrival (AoA), Elevation of Arrival (EoA)). The properties of wireless communications systems can be studied with the use of channel modeling. The generated channel model is essential in analyzing and improving the performance of the communications system. The clustering of multipaths is indicated in many channel measurements and models. The accuracy of clustering wireless propagation multipaths influences the correctness of the channel models. Many clustering approaches [1]–[5] use only the number of clusters as the validity measurement. The drawback of giving just the number of clusters is that there is no guarantee that the membership of the multipath clusters is correct despite the accuracy of the number of clusters. The inaccurate clustering of the wireless propagation multipaths would result in an erroneous channel model. This problem can be solved by jointly identifying the membership and the number of multipath clusters using SCAMS.

This study presents the results of clustering C2CM datasets obtained from IEEE DataPort using SCAMS. It describes the clustering of wireless propagation multipaths by jointly identifying the membership and the number of clusters. The main contributions of this paper are (1) SCAMS is used to cluster C2CM wireless propagation multipaths which give both the membership and the number of clusters; and (2) the results confirm the possibility of applying the clustering approach as another option in studying channel models.

The paper is organized in the following way. Section 2 describes the multipath clustering datasets. Section 3 discusses SCAMS. Section 4 presents the clustering results provided by SCAMS. Section 5 concludes the study.

2. WIRELESS PROPAGATION MULTIPATH CLUSTER DATASETS

The combination of MPCs from all the multipath clusters based on the position of the mobile station results in a channel impulse response that is varying with time (designated by t). Based on the delay and direction domain, the channel impulse response is given as

$$h(t, \tau, \Theta^{\text{BS}}, \Theta^{\text{MS}}) = \sum_{k=1}^{K} \sum_{p} \alpha_{n,p} \delta(\tau - \tau_{n,p}) \delta(\Theta^{\text{BS}} - \Theta_{n,p}^{\text{BS}}) \delta(\Theta^{\text{MS}} - \Theta_{n,p}^{\text{MS}}) \quad (1)$$

where K is the set of visible cluster indexes, $\alpha_{n,p}$ is the complex amplitude of the pth MPC in the nth cluster, $\Theta_{n,p}^{BS}$ is the direction of departure (AoD, EoD), and $\Theta_{n,p}^{MS}$ is the direction of arrival (AoA, EoA) of the MPC. C2CM is discussed in detail in [6] while the concise description can be found in [7].

Wireless propagation multipaths in indoor and semi-urban scenarios are taken from the multipath datasets in IEEE DataPort [8]. This common set of data can be used in channel modeling, evaluating clustering approaches, and assessing clustering accuracies. The generation of the datasets is detailed in [9]. The reference data used in multipath clustering are as follows:

1. Indoor, band 1 (B1), line-of-sight (LOS), single link (SL)
2. Indoor, band 2 (B2), line-of-sight, single link
3. Semi-Urban, band 1, line-of-sight, single link

4. Semi-Urban, band 2, line-of-sight, single link
5. Semi-Urban, band 1, non line-of-sight, single link
6. Semi-Urban, band 2, non line-of-sight, single link
7. Semi-Urban, band 1, line-of-sight, multiple links
8. Semi-Urban, band 2, line-of-sight, multiple links (ML)

For each channel scenario, there are thirty trials with different number of clusters and multipaths. Column 8 power component is removed as it is not needed in the clustering process. Column 9 cluster identifications or IDs are also eliminated in the clustering process as they are only used as reference IDs to compare with the calculated IDs.

Clustering results are greatly affected by the values of the affinities. To improve clustering accuracy, columns 1 to 7 whitened data are normalized [0, 1] using

$$X_{\text{norm}} = (X_{\text{white}} - X_{\text{s}}) \circ (X_{\text{l}} - X_{\text{s}})^{-1} \qquad (2)$$

where X_{norm} is the whitened data's normalized value, X_{white} is the columns 1 to 7 whitened data, X_{l} is each column's maximum data, X_{s} is each column's minimum data, and $\circ$ is the Hadamard product. Jaccard index η compares the similarity between the reference data and the computed data. The similarity measure is computed as

$$\eta = \frac{|C_{\text{data}} \cap C_{\text{clust}}|}{|C_{\text{data}} \cup C_{\text{clust}}|} = \frac{M_a}{M_a + M_b + M_c} \qquad \in [0,1] \qquad (3)$$

where $|\cdot|$ refers to cardinality, $C_k \in C$, $K = |C|$ is the number of multipath clusters, C_{data} is the reference clusters, C_{clust} is the calculated clusters, M_a is the total number of multipath clusters for the accuracy on the number of clusters or total number of multipaths for the accuracy on the membership of the clusters in C_{data} that are the same as in C_{clust}, M_b is the total number of multipath clusters for the accuracy on the number of clusters or total number of multipaths for the accuracy on the membership of the clusters in C_{data} that are not in C_{clust}, and M_c is the total number of multipath clusters for the accuracy on the number of clusters or total number of multipaths for the accuracy on the membership of the clusters in C_{clust} that are not in C_{data}. A Jaccard index of 1 means that the calculated multipath clusters are the same as the reference multipath clusters or the membership of the calculated multipath clusters is the same as the membership of the reference multipath clusters.

3. SIMULTANEOUS CLUSTERING AND MODEL SELECTION (SCAMS)

SCAMS [10]–[11] solves the clustering problem by computing the number of clusters and determining the membership of the clusters. SCAMS represents a dataset D with DC_{clust} where C_{clust} is an affinity matrix. Self-expression method [12] can be used to formulate C_{clust} which corresponds to the solution of

$$\min \| C_{\text{clust}} \|_1 \quad \text{s.t.} \quad D = DC_{\text{clust}}, \quad \text{diag}(C_{\text{clust}}) = 0. \qquad (4)$$

With the introduction of an ideal affinity matrix C_{ideal} and by designating $V = -C_{\text{clust}}$, the clustering problem can be represented as

$$\min \ \langle V, C_{\text{optimal}} \rangle,$$
$$\text{s.t.} \ \ z_k\{0,1\}^R, \quad \textstyle\sum_{k=1}^{K} z_k = e_M,$$
$$C_{\text{optimal}} = \sum_{k=1}^{K} z_k \circ z_k, \quad (C_{\text{optimal}}) = K \qquad (5)$$

where $\langle \cdot, \cdot \rangle$ is the Frobenius inner product and e_M is an all one vector of size R. An augmented Lagrange function can be used to express (5) as

$$\mathcal{L} = \text{tr}(V^{\text{T}} C_{\text{optimal}}) + \lambda \text{rank}(C_{\text{optimal}}) + \gamma \|H\|_0 + g(H)$$
$$\quad + \text{tr}(X^{\text{T}}(C_{\text{optimal}} - H + \text{diag}(H) - I))$$
$$\quad + \frac{1}{2\mu} \|C_{\text{optimal}} - H + \text{diag}(H) - I\|_{\text{F}}^2, \qquad (6)$$
$$\text{s.t.} \ \ C_{\text{optimal}} \in S_+$$

where g is the indicator function of the convex set $[0,1]^{R \times R}$, H is a variable introduced to make the problem tractable, I is an identity matrix, X is a Lagrange parameter, $\|\cdot\|_0$ is the ℓ_0 norm which counts the number of nonzero elements, $\mu > 0$ is a penalty parameter, λ and γ are the parameters that weigh the respective penalty terms, S_+ is the positive semi-definite cone, and $\|\cdot\|_{\text{F}}$ is the Frobenius norm. The Lagrange function can be minimized alternatingly with respect to H and C_{optimal}, by fixing the other variable, and then updating X. Algorithm 1 shows the solution of C_{optimal} using Alternating Direction Method of Multipliers (ADMM) [13]. C_{optimal} can then be factorized as YY^{T} where Y is an indicator matrix whose rows indicate to which cluster a point belongs. Algorithm 2 illustrates the solution of Y using Asso Constrained Boolean Matrix Factorization (AssoCBMF) [10]. Superscript B is a "Boolean" matrix containing only 0's and 1's, $|\cdot|$ is the cardinality of the Boolean matrix and defined as the number of 1's in it, $\oplus$ is the exclusive-or operation applied element-wise, $D(i, j)$ is the association accuracy as for rule $C_{\text{optimal}}{}^B(j, :) \Rightarrow C_{\text{optimal}}{}^B(i, :)$, υ is a threshold for constructing D^B, and r_{thresh} is a threshold for deleting the j-th columns. AssoCBMF algorithm gives the number of clusters and the membership of the clusters.

4. RESULTS

SCAMS solves the correct number of clusters by using the calculated value of λ from [12] and selecting the appropriate value of γ. The number of clusters is inaccurate when the value of γ is not the correct value. It follows that the membership of the clusters is also inaccurate when the number of clusters is incorrect.

Table 1 shows the Jaccard indices of the eight-channel scenarios. The results are the completion of the previous works [14], [15]. Indoor channel scenarios have better accuracy both for the number of clusters and membership of clusters due to lesser number of multipaths and multipath clusters generated by the enclosed space where reflections of

signals are limited. Semi-Urban scenarios have lower accuracy for both the number of clusters and membership of clusters due to higher number of multipaths and multipath clusters generated by the wider surroundings where more interacting objects reflect the signals.

Algorithm 1: Alternating Direction Method of Multipliers

Input: Negative affinity matrix V, parameters λ and γ

Initialize: $C_{\text{optimal}} = H = X = 0_{N \times N}, \mu = 10^6, \rho = 1.1,$
$\mu_{\min} = 10^{-10}$ and $\varepsilon = 10^{-8}$.
while not converged **do**
Step 1 Fix the others and update C_{optimal} as

$C_{\text{optimal}} = \text{argmin}_{C_{\text{optimal}}} \left\| C_{\text{optimal}} - H + \mu(V + X) \right\|_F^2 +$
$2\mu\lambda\text{rank}(C_{\text{optimal}}), \text{s.t. } C_{\text{optimal}} \in S_+.$
Step 2 Fix the others and update H as

$H' = \text{argmin}_H \left\| H - C_{\text{optimal}} - \mu X \right\|_F^2 + 2\mu\gamma\|H\|_0 + g(H),$
$H = H' - \text{diag}(H') + I.$
Step 3 Update the multipliers

$X = X + \frac{1}{\mu}(C_{\text{optimal}} - H).$
Step4 Update the parameter μ by

$\mu = \max\left(\frac{\mu}{\rho}, \mu_{\min}\right).$
Step 5 Check the convergence conditions:

$\left\| C_{\text{optimal}} - H \right\|_\infty \leq \varepsilon.$
endwhile

Table 1: Jaccard Indices for the Number of Clusters and Membership of Clusters

Channel Scenario	Number of Clusters	Membership of Clusters
Indoor Band 1	0.6034	0.7305
Indoor Band 2	0.6487	0.7582
Semi-Urban Band 1 Line-of-Sight Single Link	0.0186	0.1875
Semi-Urban Band 2 Line-of-Sight Single Link	0.0159	0.1818
Semi-Urban Band 1 Non Line-of-Sight Single Link	0.0052	0.1597
Semi-Urban Band 2 Non Line-of-Sight Single Link	0.0108	0.1505
Semi-Urban Band 1 Line-of-Sight Multiple Links	0.0080	0.1459
Semi-Urban Band 2 Line-of-Sight Multiple Links	0.0084	0.1436

Figure 1 shows the Jaccard index of the number of clusters in the indoor band 1 line-of-sight single link versus the corresponding values of λ and γ. The minimum index is 0.1176 while the maximum index is 1. Figure 2 illustrates the Jaccard index of the membership of the clusters of indoor band 1 line-of-sight single link. The minimum index is 0.3649

while the maximum index is 1. Figure 3 displays the Jaccard index of the number of clusters in the indoor band 2 line-of-sight single link. The minimum index is 0 while the maximum index is 1. Figure 4 presents the Jaccard index of the membership of the clusters of indoor band 2 line-of-sight

Algorithm 2: AssoConstrained Boolean Matrix Factorization

Input: C_{optimal}, K_0
Initialize: Construct the Boolean matrix $C_{\text{optimal}}{}^B$ from C_{optimal} with rounding threshold $t_B = 0.5, Y^B \leftarrow [\ \ ],$
$e = \infty, r_{\text{thresh}} = 0.1.$
for $v = 0.1, 0.2, \ldots, 1$ **do**
Construct D^B with

$$D^B(i,j) = \frac{\langle C_{\text{optimal}}{}^B(i,:), C_{\text{optimal}}{}^B(j,:) \rangle}{\langle C_{\text{optimal}}{}^B(j,:), C_{\text{optimal}}{}^B(j,:) \rangle} > v.$$

for $k = 1, 2, \ldots, K_0$ **do**
$i = \text{argmin}_i \left| C_{\text{optimal}}{}^B \oplus ([Y^B D^B(:,i)] \circ [Y^B D^B(:,i)]^T) \right|.$
$Y^B \leftarrow [Y^B D^B(:,i)].$
Delete all $j - th$ columns with

$\frac{\langle D^B(:,i), D^B(:,j) \rangle}{\|D^B(:,i)\|\|D^B(:,j)\|} > r_{\text{thresh}}$ from D^B

if D^B is empty or min. $\left| C_{\text{optimal}}{}^B \oplus (Y^B \circ Y^{B^T}) \right|,$
s.t. $Y^{B^T} \circ Y^B = I_{K \times K}$ is not reduced in this loop
break
end if
if $\left\| C_{\text{optimal}} - Y^B Y^{B^T} \right\|_F^2 < e$
$Y^{B^*} = Y^B.$
$e = \left\| C_{\text{optimal}} - Y^B Y^{B^T} \right\|_F^2.$
end if
end for
end for
return Y^{B^*}

single link. The minimum index is 0.3684 while the maximum index is 1. Figure 5 reveals the Jaccard index of the number of clusters of semi-urban band 1 line-of-sight single link. The minimum index is 0 while the maximum index is 0.0455. Figure 6 demonstrates the Jaccard index of the membership of clusters of semi-urban band 1 line-of-sight single link. The minimum index is 0.1364 while the maximum index is 0.2452. Figure 7 indicates the Jaccard index of the number of clusters of semi-urban band 2 line-of-sight multiple links. The minimum index is 0 while the maximum index is 0.0141. Figure 8 exhibits the Jaccard index of the membership of clusters of semi-urban band 2 line-of-sight multiple links. The minimum index is 0.1138 while the maximum index is 0.1699.

The relative frequency of the Jaccard indices on the number of clusters and the membership of the clusters for indoor band 1 line-of-sight single link is shown in Figure 9. For the number of clusters, 13% of the indices occurred from 0.9 to 1 while 20% for the membership of the clusters. Figure 10 displays the relative frequency of the Jaccard indices on the number of clusters and the membership of the clusters for

indoor band 2 line-of-sight single link. For the number of clusters, 27% of the indices occurred from 0.9 to 1 while 37% for the membership of the clusters. Figure 11 presents the relative frequency of the Jaccard indices for semi-urban band
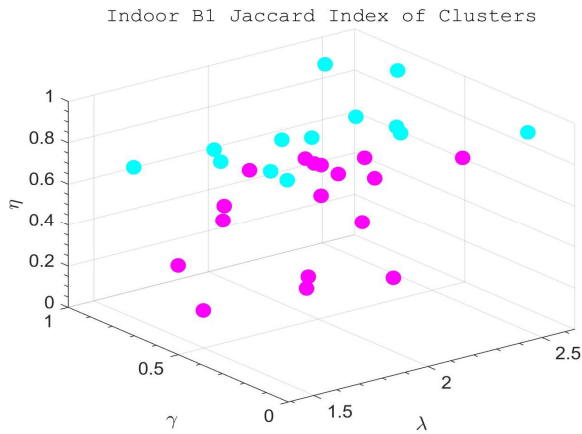
are indices higher than the mean of 0.6487

Figure 1: Jaccard index of the number of clusters as a function of λ and γ in indoor B1 LOS SL where cyan colors are indices higher than the mean of 0.6034
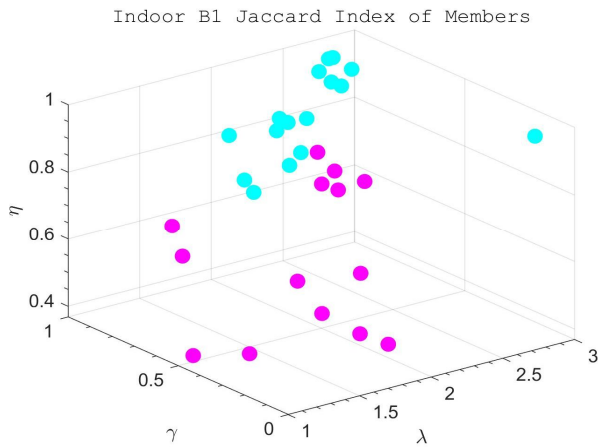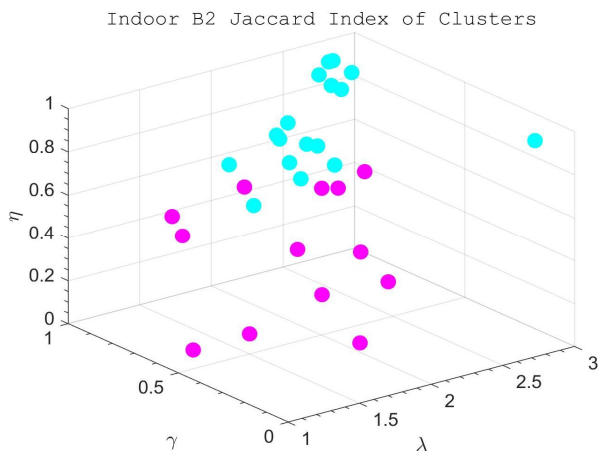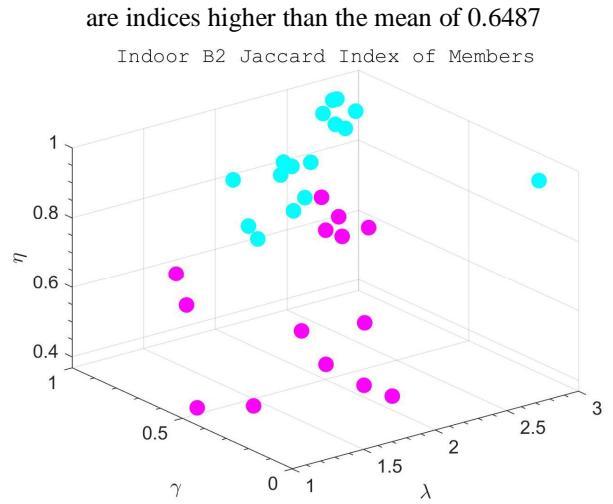

Figure 2: Jaccard index of membership of clusters as a function of λ and γ in indoor B1 LOS SL where cyan colors are indices higher than the mean of 0.7305
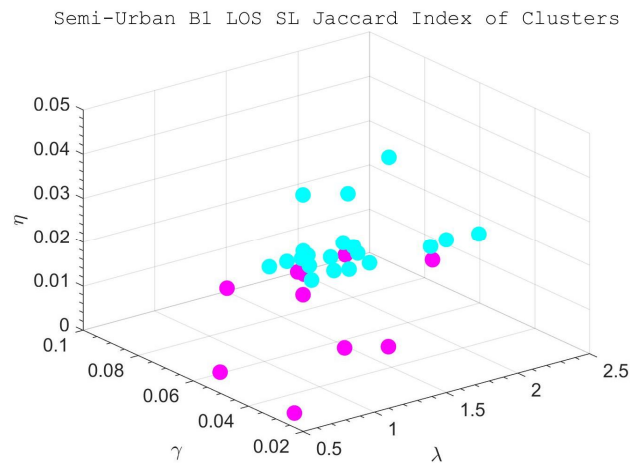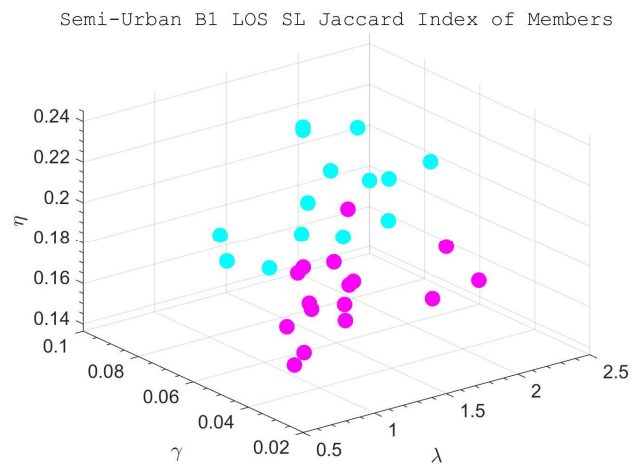

Figure 3: Jaccard index of the number of clusters as a function of λ and γ in indoor B2 LOS SL where cyan colors

Figure 4: Jaccard index of membership of clusters as a function of λ and γ in indoor B2 LOS SL where cyan colors are indices higher than the mean of 0.7582
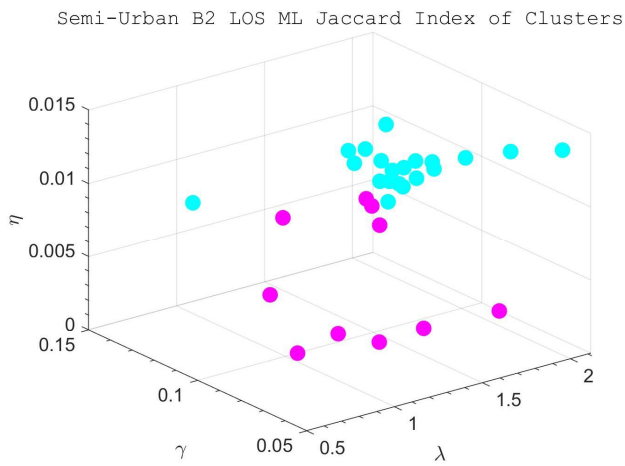

Figure 5: Jaccard index of the number of clusters as a function of λ and γ in semi-urban B1 LOS SL where cyan colors are indices higher than the mean of 0.0186
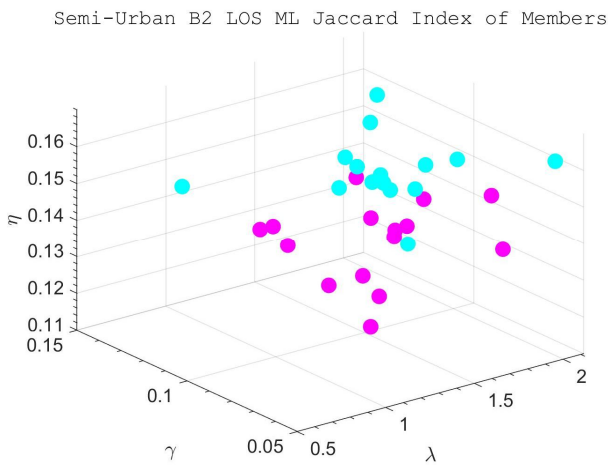

Figure 6: Jaccard index of membership of clusters as a function of λ and γ in semi-urban B1 LOS SL where cyan colors are indices higher than the mean of 0.1875

Figure 7: Jaccard index of the number of clusters as a function of λ and γ in semi-urban B2 LOS ML where cyan colors are indices higher than the mean of 0.0084

Figure 8: Jaccard index of membership of clusters as a function of λ and γ in semi-urban B2 LOS ML where cyan colors are indices higher than the mean of 0.1436

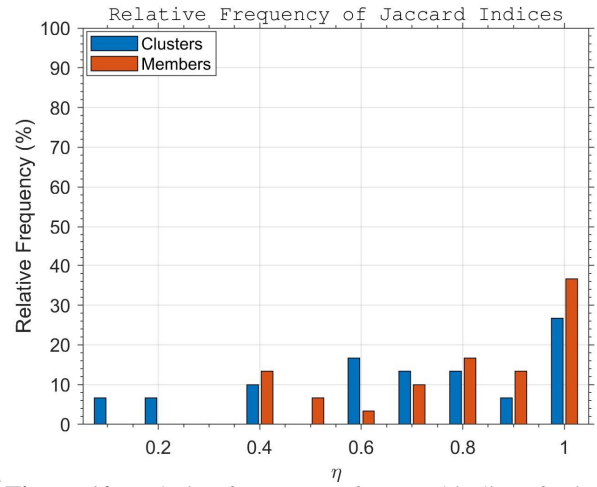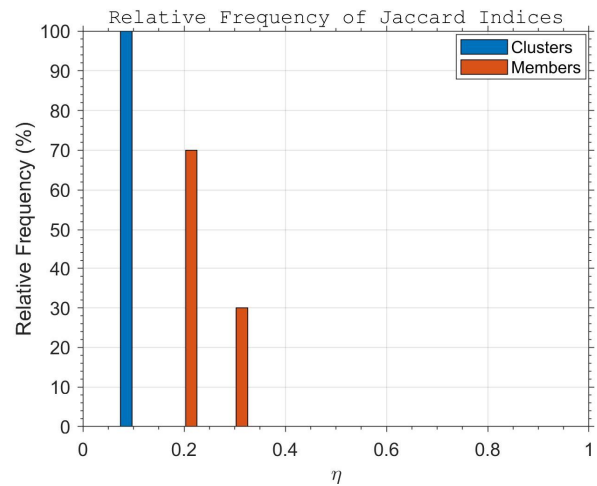**Figure 9:** Relative frequency of Jaccard indices for indoor band 9 line-of-sight single link

Figure 10: Relative frequency of Jaccard indices for indoor band 2 line-of-sight single link

Figure 11: Relative frequency of Jaccard indices for semi-urban band 1 line-of-sight single link
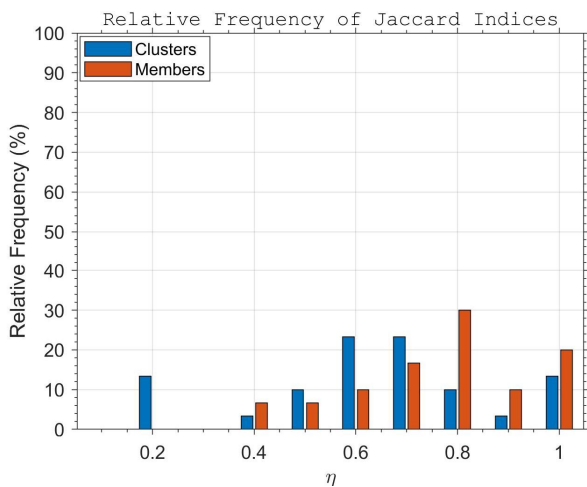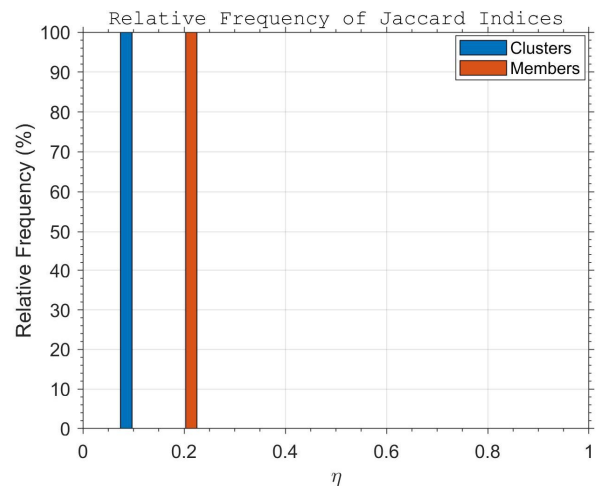

Figure 12: Relative frequency of Jaccard indices for semi-urban band 2 line-of-sight multiple links

1 line-of-sight single link. For the number of clusters, all of the indices fall between 0 and 0.1 while for the membership of the clusters all indices are within the 0.2 to 0.4 range. Figure 12 illustrates the relative frequency of the Jaccard indices for semi-urban band 2 line-of-sight multiple links. All of the indices fall within the 0 to 0.1 range for the number of clusters while 0.2 to 0.3 for the membership of clusters. The results show that SCAMS has good clustering accuracy for the indoor scenarios while improvements can be made for the semi-urban scenarios. Overall, SCAMS can be used as an alternative clustering approach in channel modeling.

5. CONCLUSION

The results of SCAMS in clustering wireless propagation multipaths are presented in this work. SCAMS can jointly identify the number of clusters and the membership of the clusters. The accuracy of SCAMS in clustering multipaths is dependent on the parameters λ and γ. The reference data for the C2CM multipath clusters are taken from the IEEE DataPort. Results show that SCAMS can be used as an alternative clustering approach in the field of channel modeling. SCAMS gives good clustering accuracy for the indoor scenarios but improvements for the semi-urban scenarios are necessary. Other options can be explored based on the clustering results with the joint identification of the membership and the number of multipath clusters into consideration.

ACKNOWLEDGEMENT

REFERENCES

1. 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 64th Veh. Technol. Conf.*, pp. 1–5, 2006. https://doi.org/10.1109/VTCF.2006.35

2. C. Gentile. **Using the kurtosis measure to identify clusters in wireless channel impulse responses**, *IEEE Trans. Antennas Propag.*, vol. 61, no. 6, pp. 3392-3395, 2013.

3. 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.*, vol. 16, no. 11, pp. 7138-7151, 2017.

https://doi.org/10.1109/TWC.2017.2740206

4. Y. Li, J. Zhang, Z. Ma, and Y. Zhang. **Clustering analysis in the wireless propagation channel with a variational Gaussian mixture model**, *IEEE Trans. Big Data*, 2018.

5. A. Teologo and L. Materum. **Cluster-Wise Jaccard Accuracy of KPower Means on Multipath Datasets**, *Int. J. Emerg. Trends Eng. Res.*, vol. 7, no. 8, pp. 203-208, 2019.

https://doi.org/10.30534/ijeter/2019/16782019

6. R. Verdone and A. Zanella. **Pervasive Mobile and Ambient Wireless Communications: COST Action 2100**, Springer Science & Business Media, 2012, ch. 3.

7. L. Liu, C. Oestges, J. Poutanen, K. Haneda, P. Vainikainen, F. Quitin, F. Tufvesson, and P. De Doncker. **The COST 2100 MIMO channel**, *IEEE Trans. Wireless Commun.*, vol. 19, no. 6, pp. 92-99, 2012.

https://doi.org/10.1109/MWC.2012.6393523

8. 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**, [Online]. Available: http://dx.doi.org/10.21227/4cb9-hf81, 2019.

9. 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**, accepted in 9th *Int. Symp. Multimedia Commun.*, Aug. 2019.

10. Z. Li, L. Cheong, and S. Zhou. **SCAMS: Simultaneous clustering and model selection**, in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, pp. 264-271, 2014.

11. Z. Li, L. Cheong, S. Yang, and K. Toh. **Simultaneous clustering and model selection: Algorithm, theory and applications**, *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 8, pp. 1964–1978, 2018.

https://doi.org/10.1109/TPAMI.2017.2739147

12. E. Elhamifar and R. Vidal. **Sparse subspace clustering: Algorithm, theory, and applications**, *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 11, pp. 2765-2781, 2013.

13. S. Boyd, N. Parikh, E. Chu, B. Peleato, J. Eckstein, et al. **Distributed optimization and statistical learning via the alternating direction method of multipliers**, *Found. Trends Mach. Learn.*, 3(1):1–122, 2011.

14. J. Blanza and L. Materum. **Grouping of COST 2100 indoor multipaths using simultaneous clustering and model selection**, *Int. J. Emerg. Trends Eng. Res.*, vol. 7, no. 8, pp. 197-202, 2019.

https://doi.org/10.30534/ijeter/2019/15782019

15. J. Blanza and L. Materum. **Wireless propagation multipath clustering: On simultaneously solving the membership and the number of clusters**, *Int. J. Adv. Trends Comput. Sci. Eng.*, vol. 8, no. 5, pp. 1914-1919, 2019.

https://doi.org/10.30534/ijatcse/2019/15852019