

Scedasticity descriptor of terrestrial wireless communications channels for multipath clustering datasets

Jojo Blanza¹, Emmanuel Trinidad², Lawrence Materum^{3,4}

¹Electronics Engineering Department, Faculty of Engineering, University of Santo Tomas, Manila, Philippines

²Electronics Engineering Department, College of Engineering and Architecture, Don Honorio Ventura State University, Bacolor, Philippines

³Electronics and Computer Engineering Department, De La Salle University, Manila, Philippines

⁴International Centre, Tokyo City University, Tokyo, Japan

Article Info

Article history:

Received Apr 2, 2023

Revised Jun 28, 2023

Accepted Jul 3, 2023

Keywords:

5G

Channel model

Data handling

Heteroscedasticity

Homoscedasticity

Multipaths

Multiple-input multiple-output

ABSTRACT

Fifth-generation (5G) wireless systems increased the bandwidth, improved the speed, and shortened the latency of communications systems. Various channel models are developed to study 5G. These channel models reproduce the stochastic properties of multiple-input multiple-output (MIMO) antennas by generating wireless multipath components (MPCs). The MPCs with similar properties in delay, angles of departure, and angles of arrival form clusters. The multipaths and multipath clusters serve as datasets to understand the properties of 5G. These datasets generated by the Cooperation in Science and Technology 2100 (COST 2100), International Mobile Telecommunications-2020 (IMT-2020), Quasi Deterministic Radio Channel Generator (QuaDRiGa), and Wireless World Initiative New Radio II (WINNER II) channel models are tested for their homoscedasticity based on Johansen's procedure. Results show that the COST 2100, QuaDRiGa, and WINNER II datasets are heteroscedastic, while the IMT-2020 dataset is homoscedastic.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

Jojo Blanza

Electronics Engineering Department, Faculty of Engineering, University of Santo Tomas

Manila, Philippines

Email: jfblanza@ust.edu.ph

1. INTRODUCTION

Fifth-generation (5G) wireless systems improved cellular communications due to increased bandwidth, faster data rates, and shorter latency times. The properties of 5G are studied using channel models that reproduce the stochastic properties of multiple-input multiple-output (MIMO) antennas. Channel models like the European Cooperation in Science and Technology 2100 (COST 2100) [1], International Mobile Telecommunications-2020 (IMT-2020) [2], Quasi Deterministic Radio Channel Generator (QuaDRiGa) [3], and Wireless World Initiative New Radio II (WINNER II) [4] generate multipath components (MPCs) that are grouped into clusters when they have similar characteristics in delay, angles of departure, and angles of arrival. The multipaths and multipath clusters serve as datasets and are clustered using clustering algorithms to study the effectiveness of clustering approaches.

Previous researches investigated the effect of heteroscedasticity on datasets. The performance of feed forward neural network and multiple regression are compared in the presence of heteroscedasticity in simulated data in [5] while heteroscedasticity was accommodated in allometric models to predict the forest biomass in [6]. The interval fusion with preference aggregation procedure is applied to process the heteroscedastic measured direct control (DC) voltage and resistance data in [7]. Also, the effect on clustering

is examined. There are differences in robust clustering when the assumption of homogeneity fails [8], [9] and if so, an unsatisfactory accuracy can result [10].

Earlier works [11], [12] on clustering multipaths did not consider the scedasticity of the dataset. This work tests the scedasticity of datasets generated by COST 2100, IMT-2020, QuaDRiGa, and WINNER II channel models. The test is based on Johansen's multivariate analysis of variance procedure under heteroscedasticity. Hence, this study is the first to conduct a homoscedasticity test on datasets generated by 5G channel models. The paper is organized as: section 2 presents the methodology of the study, while section 3 discusses the results. Lastly, section 4 concludes the study.

2. METHOD

The methodology of the study is shown in Figure 1. The MATLAB implementations of the channel models [13]–[16] are used to generate the channel scenarios. The different channel scenarios depicting the propagation environments of the wireless communications system contain various multipath components and multipath clusters. The data generated are subjected to the directional cosine transform (DCT). The results of DCT become the datasets used in the Homoscedasticity Test. Figure 1 details the parts of the study.

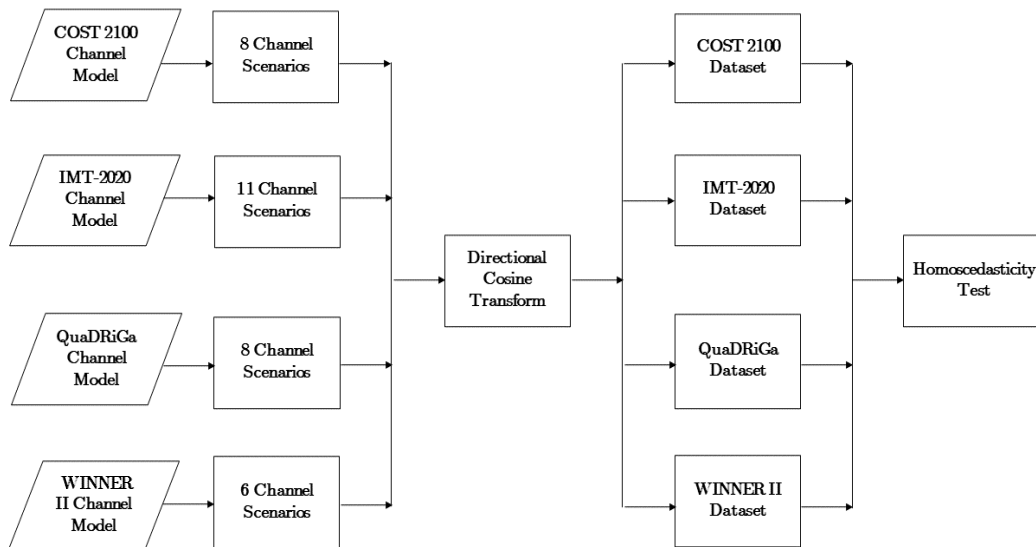


Figure 1. Methodology of the study

2.1. Channel models

This section presents the standard channel models (CM) COST 2100, IMT-2020, WINNER II, and QuaDRiGa. The global parameters line-of-sight (LOS) and non-line-of-sight (NLOS) propagation conditions are configured, extracted the MPCs, and arranged the data in an excel file. Channel models based on the clustering of the azimuth, elevation, and delay domain parameters at both link ends lead to the double-directional channel models [17]. The use of clusters in channel models provides a good trade-off between the complexity and accuracy.

2.1.1. COST 2100

The COST 2100 channel model (C2CM) is a geometry-based stochastic channel model and a part of the COST family of channel models. The model can replicate the MIMO behavior in time, frequency, and space [13] through simulation and generating channel coefficients and the small-scale parameters (SSPs). The C2CM adopts the visibility region concept of its predecessor where the cluster can be seen by the mobile station (MS) and at maximum as the MS approaches the center. Hence, this visibility region (VR) concept allows the simulation of a non-stationary channel. Furthermore, clusters can be identified as single-bounce cluster, local cluster, and twin clusters. Local clusters have omnidirectional spread in the azimuth, single-bounce cluster positions are rotated with a Gaussian distributed angle, and both can be treated as special twin clusters with zero cluster-link delays. The C2CM supports the 285 MHz for semi-urban and 5.3 GHz for indoor carrier frequencies.

2.1.2. IMT-2020

The IMT-2020 channel model specifies the use cases for 5G which supports the 3D MIMO by extending the elevation domain. The model supports a center frequency of 0.5-100 GHz in its MATLAB implementation. The use cases of 5G have corresponding scenarios based on the model; for the indoor hall (InH) enhanced mobile broadband (eMBB), the dense urban eMBB is the urban macro (UMa) and urban micro (UMi) layers, and rural eMBB is represented by Rural Macro (RMa). Additionally, the channel model supports the massive machine type communications (mMTC) and ultra-reliable and low latency communications (URLLC) use cases of 5G, which are excluded from the data generation. The IMT-2020 is also considered a geometry-based stochastic channel model (GBSCM) for its primary module, an additional module below 6 GHz is a map-based hybrid channel module based on the ray-tracing model. For its SSPs, the delay and cluster power are modeled from exponential delay distribution and exponential power delay profile, respectively, while a wrapped Gaussian or Laplacian for the power angle spectrum in the azimuth of all its clusters. The number of clusters generated by the IMT-2020 follows the Poisson distribution. It also takes into consideration the vegetation effects on mmWave bands.

2.1.3. WINNER II

The WINNER II channel model has a module of 2 to 6 GHz, 100 MHz bandwidth, and 19 propagation scenarios. WINNER II is based on ray-tracing, which also replicates the double-directional feature. Furthermore, the chosen scenarios Indoor A1 LOS and NLOS, UMi LOS and NLOS indicated by B1, and UMa LOS and NLOS indicated by C1 [4]. The A1 is defined as the indoor office or residential scenario, the B1 is a typical urban microcell, and the C1 is the urban microcell for wide area networks.

2.1.4. QuaDRiGa

The QuaDRiGa channel model is an extension of the WINNER SCM model, which supports the three dimensions (3D) MIMO modeling, continuous time evolution, and transitions of propagation scenarios, and also provides terrestrial and satellite scenarios [3]. The approach used by the authors to develop the QuaDRiGa channel model is a statistical ray-tracing model which differs from other channel models and has the spatial consistency of both large-scale parameters (LSPs) and SSPs. The QuaDRiGa model supports the carrier frequency range of 0.45 to 100 GHz and is compatible with the 3rd generation partnership project (3GPP) channel model. Its propagation scenarios are urban macrocellular UMa or urban microcellular UMi, validated from measurements in downtown Berlin, Germany.

2.2. Directional cosine transform

A time snapshot of the channel model can be characterized by an $L \times D$ matrix,

$$X_{\text{RAW}} = [x_1, \quad x_2, \quad \dots \quad x_\ell, \quad \dots \quad x_{L-1}, \quad x_L]^T \quad (1)$$

where $[\cdot]^T$ is the transpose operator and the ℓ -th multipath vector,

$$x_\ell = [\phi_{\ell,\text{AOD}} \quad \theta_{\ell,\text{AOD}} \quad \phi_{\ell,\text{AOA}} \quad \theta_{\ell,\text{AOA}} \quad \tau_\ell]^T \quad (2)$$

describes the azimuth angle of departure (AOD) $\phi_{\ell,\text{AOD}}$, elevation AOD $\theta_{\ell,\text{AOD}}$, azimuth angle of arrival (AOA) $\phi_{\ell,\text{AOA}}$, elevation AOA $\theta_{\ell,\text{AOA}}$, and ℓ -th multipath delay τ_ℓ that is illustrated in Figure 2. The angular ambiguity in the circular data, θ in X_{RAW} , can be avoided in the process of clustering by transforming each spherical coordinate in x_ℓ into its equivalent rectangular coordinate using the directional cosines.

$$x_x = \sin \theta \cos \phi \quad (3)$$

$$x_y = \sin \theta \sin \phi \quad (4)$$

$$x_z = \cos \theta \quad (5)$$

The transformation results in an X_{DCT} with seven dimensions where,

$$x_\ell = [x_{\ell,x,\text{AOD}} \quad x_{\ell,y,\text{AOD}} \quad x_{\ell,z,\text{AOD}} \quad x_{\ell,x,\text{AOA}} \quad x_{\ell,y,\text{AOA}} \quad x_{\ell,z,\text{AOA}} \quad \tau_\ell]^T \quad (6)$$

X_{DCT} serves as the dataset for the homoscedasticity test.

2.3. Datasets

The The IMT-2020, COST 2100, QuaDRiGa, and WINNER II datasets are available online in IEEE DataPort [18], [19]. They consist of various channel scenarios, multipath clusters, and multipath components. However, each channel scenario has the same thirty Excel sheets of data.

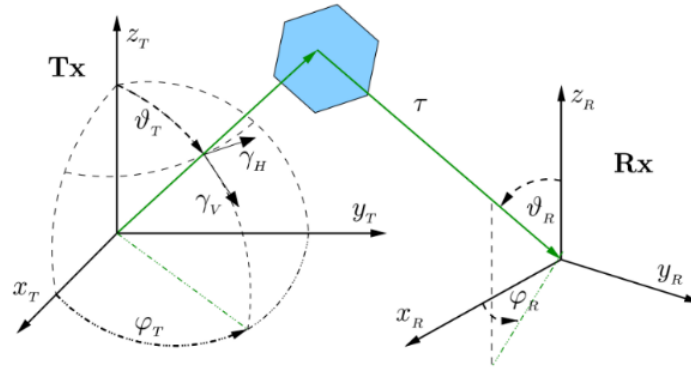


Figure 2. COST 2100 double-direction propagation path parameters [1]

Table 1 shows the channel scenarios for each dataset. IMT-2020 has eleven channel scenarios. The number of clusters are all equal for the thirty sheets of data for each channel scenario. On the other hand, the COST 2100 dataset consists of eight channel scenarios. The number of clusters varies for the thirty sheets of data for each channel scenario. The values shown pertain to the maximum number of clusters per channel scenario. Moreover, QuaDRiGa has eight channel scenarios. The number of clusters is equal for all thirty sheets of data per channel scenario. Lastly, WINNER II has six channel scenarios. The number of clusters is also the same for the thirty sheets of data for each channel scenario.

The number of multipaths per channel scenario is shown in Table 2. The number of multipaths is the same for the thirty sheets of data for each channel scenario for IMT-2020, QuaDRiGa, and WINNER II datasets. Whereas, the COST 2100 dataset has varying number of multipaths per sheet of data for all channel scenario. The values shown refer to the maximum number of multipaths for each channel scenario.

2.4. Homoscedastic test

Channel models have been developed with specific application scenarios in place. However, despite the seeming similarity among the application scenarios of the models discussed in this paper, there is some degree of scenario variability. The authors, thus, introduce a description to measure the variation of each channel model type. The variance of the parameters (delay, angle, and the like) of the channel model type is determined using statistical tests.

Testing for the scedasticity, either heteroscedasticity or homoscedasticity, enables a quantification of the variation. Either homoscedastic or heteroscedastic tests can be done since they are complementary notions of each other. Let a trivariate data matrix be,

$$D_{31} = [d_{f1} \ d_{g1} \ d_{h1}] \quad \in L \times 3 \quad (7)$$

where L is the number of samples under the 1st group of the same dataset with three (3) features $f1$, $g1$, and $h1$. d_f, d_g, d_h are each an $L \times 1$ vector. Similarly, let a second trivariate data matrix of the same dataset be defined as (8),

$$D_{32} = [d_{f2} \ d_{g2} \ d_{h2}] \quad \in L \times 3 \quad (8)$$

The homoscedastic test is based on the covariance matrices of D_{31} and D_{32} . So, if the null hypothesis holds, i.e., D_{31} and D_{32} have equal covariance matrices, then the dataset is considered to be homoscedastic. The homoscedastic test is expected to be done per L samples. The datasets generated from the channel models do not always have equal L . Thus, one based on multivariate analysis of variance (MANOVA), which compares the mean vectors of several multivariate normal populations, was adapted [20]. It is based on Johansen's procedure [21].

Table 1. Number of clusters generated for each channel scenario

Channel model	Channel scenario	Result
IMT-2020	InH A LOS	15
	InH A NLOS	19
	RMa A LOS	11
	RMa A NLOS	10
	RMa A O2I	10
	UMa A LOS	12
	UMa A NLOS	20
	UMa A O2I	12
	UMi A LOS	12
	UMi A NLOS	19
	UMi A O2I	12
	Indoor B1 LOS Single Link	27
	Indoor B2 LOS Single Link	26
	Semi-Urban B1 LOS Single Link	35
COST 2100	Semi-Urban B2 LOS Single Link	38
	Semi-Urban B1 NLOS Single Link	34
	Semi-Urban B2 NLOS Single Link	33
	Semi-Urban B1 LOS Multiple Links	63
	Semi-Urban B2 LOS Multiple Links	66
	BERLIN UMa LOS	15
QuaDRiGa	BERLIN UMa NLOS	25
	BERLIN UMi Campus LOS	12
	BERLIN UMi Campus NLOS	20
	BERLIN UMi Square LOS	12
	BERLIN UMi Square NLOS	20
	Industrial LOS	25
WINNER II	Industrial NLOS	26
	Indoor A1 LOS	12
	Indoor A1 NLOS	16
	UMa C2 LOS	8
	UMa C2 NLOS	20
	UMi B1 LOS	8
UMi B1 NLOS	16	

Table 2. Number of multipaths per channel scenario

Channel model	Channel scenario	Result
IMT-2020	InH A LOS	1,425
	InH A NLOS	3,895
	RMa A LOS	31,768
	RMa A NLOS	26,600
	RMa A O2I	58,520
	UMa A LOS	11,172
	UMa A NLOS	77,140
	UMa A O2I	216,144
	UMi A LOS	12,084
	UMi A NLOS	24,187
	UMi A O2I	109,440
	Indoor B1 LOS Single Link	80
	Indoor B2 LOS Single Link	77
	Semi-Urban B1 LOS Single Link	944
COST 2100	Semi-Urban B2 LOS Single Link	1,025
	Semi-Urban B1 NLOS Single Link	1,631
	Semi-Urban B2 NLOS Single Link	1,583
	Semi-Urban B1 LOS Multiple Links	1,700
	Semi-Urban B2 LOS Multiple Links	1,781
	BERLIN UMa LOS	18,000
QuaDRiGa	BERLIN UMa NLOS	300,00
	BERLIN UMi Campus LOS	7,200
	BERLIN UMi Campus NLOS	12,000
	BERLIN UMi Square LOS	7,200
	BERLIN UMi Square NLOS	12,000
	Industrial LOS	7,500
WINNER II	Industrial NLOS	7,800
	Indoor A1 LOS	3,600
	Indoor A1 NLOS	4,800
	UMa C2 LOS	9,600
	UMa C2 NLOS	24,000
	UMi B1 LOS	4,800
UMi B1 NLOS	9,600	

3. RESULTS AND DISCUSSION

Using study [22], the Johansen's test (JT) values and the corresponding p-values of COST 2100, IMT-2020, QuaDRiGa, and WINNER II are shown in Table 3. COST 2100, QuaDRiGa, and WINNER II across all channel scenarios have JT values greater than one while IMT-2020 for all channel scenarios has JT values less than one. Moreover, COST 2100, QuaDRiGa, and WINNER II have p-values less than 0.05 which show that the mean vectors of the different channel scenarios are significantly different. In contrast, IMT-2020 has p-values greater than 0.05 which indicates that the mean vectors are not significantly different.

Table 3. Johansen's test and p-value of the different channel models

Channel model	Channel scenario	Johansen's test value	p-value	
IMT-2020	InH A LOS	0.9848	0.5480	
	InH A NLOS	1.1304	0.0984	
	RMa A LOS	0.8664	0.9158	
	RMa A NLOS	0.9774	0.5777	
	RMa A O2I	0.9079	0.8224	
	UMa A LOS	0.7929	0.9867	
	UMa A NLOS	0.9526	0.6743	
	UMa A O2I	0.9908	0.5239	
	UMi A LOS	0.7203	0.9990	
	UMi A NLOS	0.8058	0.9805	
	UMi A O2I	0.9696	0.6089	
	COST 2100	Indoor B1 LOS Single Link	35.6103	0.0000
		Indoor B2 LOS Single Link	18.6615	0.0000
Semi-Urban B1 LOS Single Link		9.1981	0.0000	
Semi-Urban B2 LOS Single Link		8.3984	0.0000	
Semi-Urban B1 NLOS Single Link		13.2897	0.0000	
Semi-Urban B2 NLOS Single Link		12.4116	0.0000	
Semi-Urban B1 LOS Multiple Links		15.0703	0.0000	
Semi-Urban B2 LOS Multiple Links		7.9604	0.0000	
QuaDRiGa		BERLIN UMa LOS	363.6940	0.0000
		BERLIN UMa NLOS	8.6673	0.0000
	BERLIN UMi Campus LOS	3.3483	0.0000	
	BERLIN UMi Campus NLOS	59.9939	0.0000	
	BERLIN UMi Square LOS	26.3466	0.0000	
	BERLIN UMi Square NLOS	118.6593	0.0000	
WINNER II	Industrial LOS	13.5040	0.0000	
	Industrial NLOS	8.5911	0.0000	
	Indoor A1 LOS	1.8454	0.0000	
	Indoor A1 NLOS	2.5174	0.0000	
	UMa C2 LOS	3.3454	0.0000	
	UMa C2 NLOS	11.4821	0.0000	
	UMi B1 LOS	2.3058	0.0000	
	UMi B1 NLOS	5.0519	0.0000	

The homoscedasticity test results of the four CMs are shown in Table 4. COST 2100, QuaDRiGa, and WINNER II are heteroscedastic, while IMT-2020 is homoscedastic. The heteroscedastic CMs have lesser number of multipaths per cluster in Semi-Urban, UMa, and UMi channel scenarios resulting to a more varied distribution of multipaths across all channel scenarios. On the other hand, the homoscedastic CM has a higher number of clusters for the abovementioned channel scenarios which give a more homogeneous distribution of multipaths for all channel scenarios. The dense distribution of multipaths per cluster resulted to mean vectors that are not significantly different as shown in Table 4.

Based on studies [23]–[25], the standard deviations (σ) of the parameters AOD, EOD, AOA, EOA, and τ for all channel scenarios of the four CMs are computed as (9):

$$\sigma_{\Omega} \doteq \sqrt{\int |\Omega - \mu_{\Omega}|^2 P(\Omega) d\Omega} \quad (9)$$

where Ω is the parameter, μ is the mean, and $P(\Omega)$ is the power spectrum of the parameter. IMT-2020 has the least delay spread for all (combination of LOS and NLOS) indoor channel scenarios while COST 2100 has the largest delay spread as shown in Figure 3. Also, Figure 4 indicates that IMT-2020 has the least spread in outdoor LOS channel scenarios for the parameters AOD in Figure 4(a), EOD in Figure 4(b), AOA in Figure 4(c), and EOA in Figure 4(d). COST 2100 has the largest AOD and EOA spreads while QuaDRiGa has the largest EOD and AOA spreads. As for the delay spread, COST 2100 has the greatest value while

QuaDRiGa has the least value. For the COST 2100 CM, the dataset generated is heteroscedastic due to varying number of clusters and number of multipaths. For this reason, the means are significantly different and the variances are more pronounced across all channel scenarios.

Figure 5 shows the AOD in Figure 5(a), EOD in Figure 5(b), AOA in Figure 5(c), and EOA in Figure 5(d) spreads for all indoor channel scenarios. It indicates that COST 2100 has the largest spread while QuaDRiGa has the least spread for AOD, EOD, and EOA. WINNER II has the least AOA spread.

Figure 6 shows that WINNER II has the least spread for the parameters AOD in Figure 6(a) and AOA in Figure 6(b) for all outdoor scenarios. All outdoor scenarios are combination of LOS, NLOS and O2I outdoor channel scenarios. It also shows that COST 2100 has the least spread in EOD in Figure 6(c) and EOA in Figure 6(d). As for the delay parameter, WINNER II has the least spread while COST 2100 has the largest spread.

Table 4. Homoscedasticity test results of the different channel models

Channel model	Channel scenario	Result
IMT-2020	InH A LOS	Homoscedastic
	InH A NLOS	Homoscedastic
	RMa A LOS	Homoscedastic
	RMa A NLOS	Homoscedastic
	RMa A O2I	Homoscedastic
	UMa A LOS	Homoscedastic
	UMa A NLOS	Homoscedastic
	UMa A O2I	Homoscedastic
	UMi A LOS	Homoscedastic
	UMi A NLOS	Homoscedastic
COST 2100	UMi A O2I	Homoscedastic
	Indoor B1 LOS Single Link	Heteroscedastic
	Indoor B2 LOS Single Link	Heteroscedastic
	Semi-Urban B1 LOS Single Link	Heteroscedastic
	Semi-Urban B2 LOS Single Link	Heteroscedastic
	Semi-Urban B1 NLOS Single Link	Heteroscedastic
	Semi-Urban B2 NLOS Single Link	Heteroscedastic
	Semi-Urban B1 LOS Multiple Links	Heteroscedastic
Semi-Urban B2 LOS Multiple Links	Heteroscedastic	
QuaDRiGa	BERLIN UMa LOS	Heteroscedastic
	BERLIN UMa NLOS	Heteroscedastic
	BERLIN UMi Campus LOS	Heteroscedastic
	BERLIN UMi Campus NLOS	Heteroscedastic
	BERLIN UMi Square LOS	Heteroscedastic
WINNER II	BERLIN UMi Square NLOS	Heteroscedastic
	Industrial LOS	Heteroscedastic
	Industrial NLOS	Heteroscedastic
	Indoor A1 LOS	Heteroscedastic
	Indoor A1 NLOS	Heteroscedastic
	UMa C2 LOS	Heteroscedastic
	UMa C2 NLOS	Heteroscedastic
	UMi B1 LOS	Heteroscedastic
UMi B1 NLOS	Heteroscedastic	

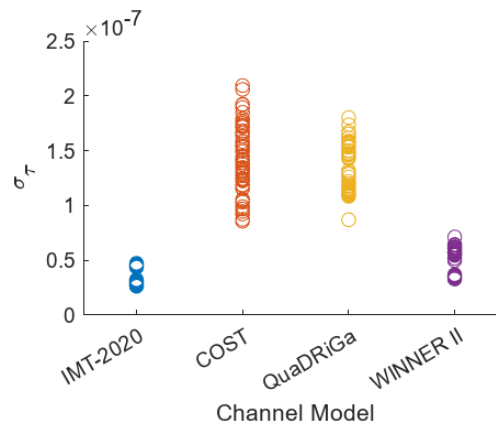


Figure 3. Delay spread of all indoor channel scenarios

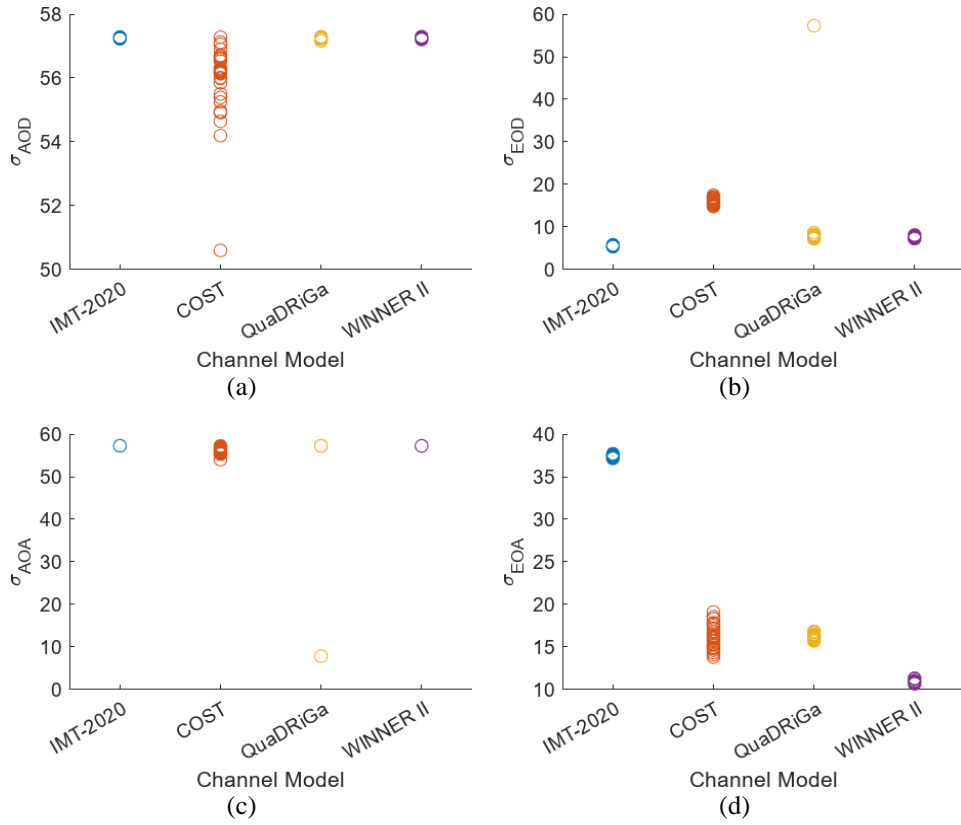


Figure 4. Outdoor LOS channel scenarios spread of (a) AOD, (b) EOD, (c) AOA, and (d) EOA

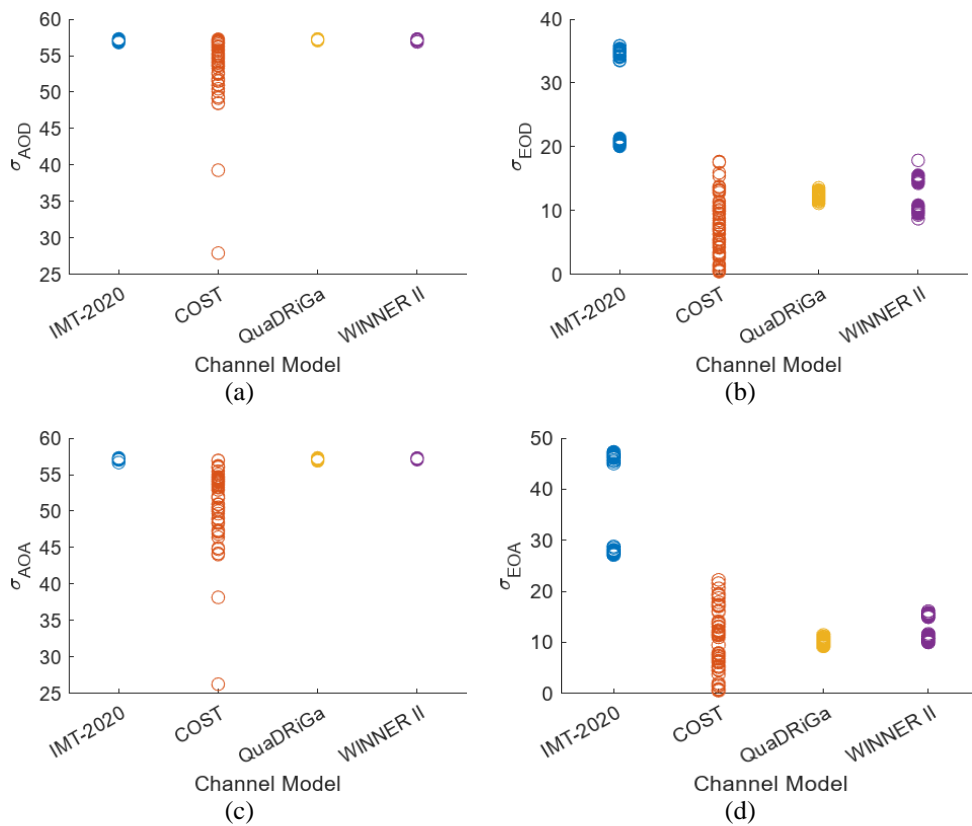


Figure 5. All indoor channel scenarios spread of (a) AOD, (b) EOD, (c) AOA, and (d) EOA

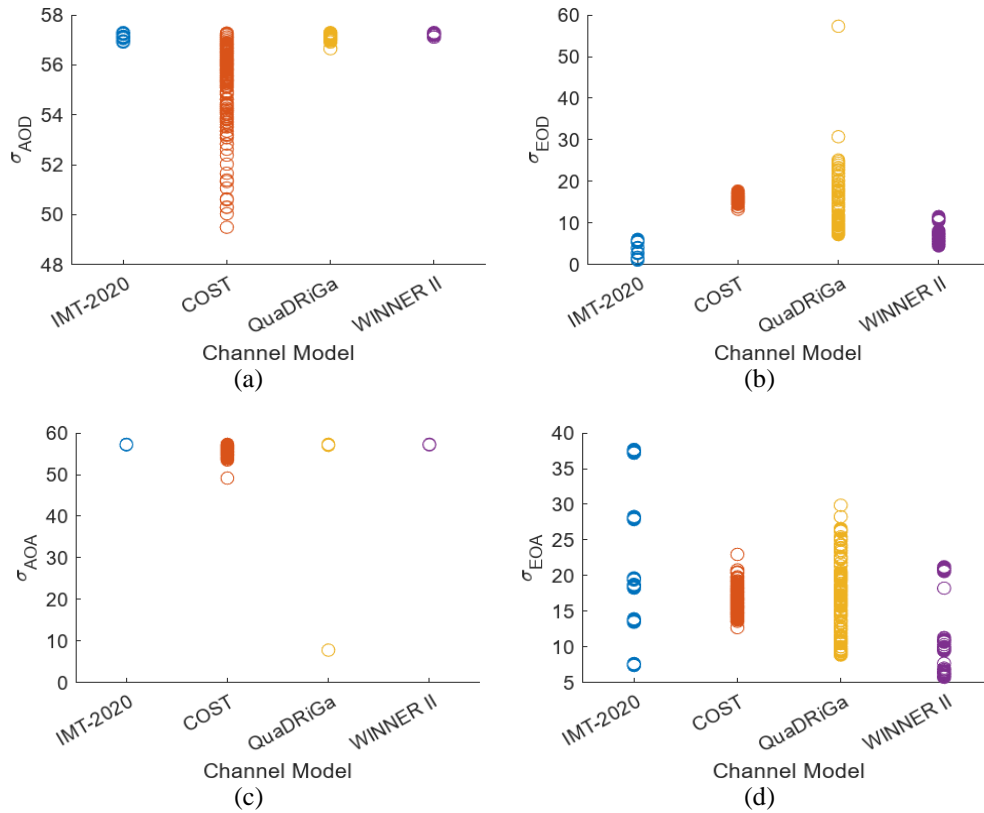


Figure 6. All outdoor channel scenarios spread of (a) AOD, (b) EOD, (c) AOA, and (d) EOA

EOD and EOA spread are directly proportional to the BS-MS antenna height difference. Table 5 shows the relationship between BS antenna and MS antenna height difference, the difference between the maximum and minimum EOD spread, and the difference between the maximum and minimum EOA spread. COST 2100 has the largest antenna height difference in all indoor scenarios, hence, it has the greatest EOD and AOD spread. Figure 5(a) shows the AOD spread for all indoor channel scenarios while Figure 5(b) shows the EOD spread. AOD and AOA spread are almost the same across all channel scenarios for IMT-2020, QuaDRiGa, and WINNER II while COST 2100 has the greatest spread for the majority of the channel scenarios. In general, IMT-2020 has the least spread, thus, the dataset it generates is homoscedastic while the other three CMs are heteroscedastic. It is based on the correlative observation of the relationship between homoscedasticity and the order of statistical values, and the rise and fall of these values but without mathematical proofs or equations.

Table 5. Relationship between BS-MS antenna height difference, difference between maximum and minimum EOD spread, and difference between maximum and minimum EOA spread

Channel model	Channel scenario	Antenna height difference (m)	EOD spread (°)	EOA spread (°)
IMT-2020	All indoor scenarios	1.5	15.77	20.22
	All outdoor scenarios	22.5	5.00	30.22
COST 2100	All indoor scenarios	9	17.28	21.71
	All outdoor scenarios	15	4.37	10.27
QuaDRiGa	All indoor scenarios	3.5	2.42	2.26
	All outdoor scenarios	22.5	50.11	21.01
WINNER II	All indoor scenarios	1	9.15	6.16
	All outdoor scenarios	22.5	7.07	15.51

4. CONCLUSION

The study presents the homoscedasticity test based on Johansen's procedure of the COST 2100, IMT-2020, QuaDRiGa, and WINNER II 5G channel model datasets. Results show that the COST 2100, QuaDRiGa, and WINNER II datasets are heteroscedastic while the IMT-2020 dataset is homoscedastic. Future study will look into the effect of scedasticity on the accuracy of clustering the datasets.

ACKNOWLEDGEMENTS




Authors thank the University of Santo Tomas Research Center for the Natural and Applied Sciences and De La Salle University for the financial support.

REFERENCES




- [1] R. Verdone and A. Zanella, *Pervasive mobile and ambient wireless communications*. London: Springer London, 2012.
- [2] M. Series, "Guidelines for evaluation of radio interface technologies for IMT-2020." ITU-R M.2412-0, 2020.
- [3] F. Burkhardt, S. Jaeckel, E. Eberlein, and R. Prieto-Cerdeira, "QuaDRiGa: a MIMO channel model for land mobile satellite," in *The 8th European Conference on Antennas and Propagation (EuCAP 2014)*, 2014, pp. 1274–1278, doi: 10.1109/EuCAP.2014.6902008.
- [4] P. Kyosti, "Winner ii channel models," *IST*, Tech. Rep. IST-4-027756 WINNER II D1. 1.2 V1. 2, 2007.
- [5] M. Paliwal and U. A. Kumar, "The predictive accuracy of feed forward neural networks and multiple regression in the case of heteroscedastic data," in *2008 IEEE International Conference on Industrial Engineering and Engineering Management*, Dec. 2008, pp. 430–434, doi: 10.1109/IEEM.2008.4737905.
- [6] I. Dutcá, R. E. McRoberts, E. Næsset, and V. N. B. Blujdea, "Accommodating heteroscedasticity in allometric biomass models," *Forest Ecology and Management*, vol. 505, Feb. 2022, doi: 10.1016/j.foreco.2021.119865.
- [7] S. V. Muravyov, L. I. Khudonogova, and M. D. Ho, "Analysis of heteroscedastic measurement data by the self-refining method of interval fusion with preference aggregation – IF&PA," *Measurement*, vol. 183, Oct. 2021, doi: 10.1016/j.measurement.2021.109851.
- [8] C. H. Lee and D. G. Steigerwald, "Inference for clustered data," *Stata Journal*, vol. 18, no. 2, pp. 447–460, Jun. 2018, doi: 10.1177/1536867x1801800210.
- [9] A. V. Carter, K. T. Schnepel, and D. G. Steigerwald, "Asymptotic behavior of a t-test robust to cluster heterogeneity," *Review of Economics and Statistics*, vol. 99, no. 4, pp. 698–709, Oct. 2017, doi: 10.1162/REST_a_00639.
- [10] M. Mendez, S. Calderon, and P. N. Tyrrell, "Using cluster analysis to assess the impact of dataset heterogeneity on deep convolutional network accuracy: a first glance," in *Communications in Computer and Information Science*, vol. 1087 CCIS, Springer International Publishing, 2020, pp. 307–319.
- [11] J. Blanza, "Wireless propagation multipaths using spectral clustering and three-constraint affinity matrix spectral clustering," *Baghdad Science Journal*, vol. 18, Jun. 2021, doi: 10.21123/bsj.2021.18.2(Suppl.).1001.
- [12] J. F. Blanza and L. Materum, "Three-constraints affinity matrix on simultaneous identification of the clustering and cardinality of wireless propagation multipaths," *Journal of the Franklin Institute*, vol. 359, no. 5, pp. 2359–2376, Mar. 2022, doi: 10.1016/j.jfranklin.2022.01.007.
- [13] L. Liu, "COST 2100 channel model," *GitHub*, Accessed: Jan. 15, 2019. [Online]. Available: <https://github.com/cost2100/cost2100/tree/master/matlab>
- [14] J. Zhang, "ITU-R IMT-2020 channel model platform," *zjhlab.net*, 2018, Accessed: Jul. 03, 2022. [Online]. Available: <http://www.zjhlab.net/publications/imt-2020>.
- [15] F. H. Hertz, "Quasi deterministic radio channel generator," *User Manual and Documentation*. 2014.
- [16] J. Blanza, "Concurrent and spectral clustering of wireless waves," *IEEE Dataport*, 2022, doi: 10.21227/gex6-bs84.
- [17] 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, doi: 10.1109/74.951559.
- [18] J. F. Blanza, A. T. Teologo, and L. Materum, "Datasets for multipath clustering at 285 MHz and 5.3 GHz bands based on COST 2100 MIMO channel model," in *2019 International Symposium on Multimedia and Communication Technology (ISMAT)*, Aug. 2019, pp. 1–5, doi: 10.1109/ISMAT.2019.8836143.
- [19] J. Blanza, "Scedasticity descriptor of terrestrial wireless communications channels for multipath clustering datasets," *IEEE Dataport*. 2023, doi: 10.21227/4khe-gx0.
- [20] C. A. Mertler and R. A. Vannatta, *Advanced and multivariate statistical methods: practical application and interpretation*, Routledge, pp. 1–347, 2002.
- [21] S. Johansen, "The Welch-James approximation to the distribution of the residual sum of squares in a weighted linear regression," *Biometrika*, vol. 67, no. 1, pp. 85–92, 1980, doi: 10.1093/biomet/67.1.85.
- [22] A. Trujillo-Ortiz and R. Hernandez-Walls, "MATLAB central file exchange," *jomaovhet, MATLAB Central File Exchange*, 2023, Accessed: Jan. 31, 2023. [Online]. Available: <https://www.mathworks.com/matlabcentral/fileexchange/37173-jomaovhet>
- [23] B. H. Fleury, "First- and second-order characterization of direction dispersion and space selectivity in the radio channel," *IEEE Transactions on Information Theory*, vol. 46, no. 6, pp. 2027–2044, 2000, doi: 10.1109/18.868476.
- [24] B. H. Fleury, "Direction dispersion and space selectivity in the mobile radio channel," in *IEEE Vehicular Technology Conference*, 2000, vol. 2, no. 52 ND, pp. 805–810, doi: 10.1109/vetecf.2000.887115.
- [25] K. I. Pedersen, P. E. Mogensen, and B. H. Fleury, "A stochastic model of the temporal and azimuthal dispersion seen at the base station in outdoor propagation environments," *IEEE Transactions on Vehicular Technology*, vol. 49, no. 2, pp. 437–447, Mar. 2000, doi: 10.1109/25.832975.

BIOGRAPHIES OF AUTHORS






Jojo Blanza    received the B.Sc. degree in electronics and communications engineering from the University of Santo Tomas, Manila, Philippines, in 1999 and the M.S. and Ph.D. degrees in Electronics and Communications Engineering from De La Salle University, Manila, Philippines, in 2005 and 2020, respectively. Currently, he is an Assistant Professor at the Electronics Engineering Department, University of Santo Tomas. His research interests include wireless communications, radio wave propagation, channel modeling, and multipath clustering. He can be contacted at email: jfblanza@ust.edu.ph.



Emmanuel Trinidad    received the B.Sc. degree in electronics engineering from Angeles University Foundation, Angeles City, Philippines in 2016, and the M.Sc. in electronics and communications engineering from De La Salle University, Manila, Philippines in 2022. Currently, he serves as an instructor at the Electronics Engineering Department, Don Honorio Ventura State University. His research interests include television white space, wireless multipath clustering, and dimensionality reduction techniques. He can be contacted at email: ettrinidad@dhsu.edu.ph.



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. He can be contacted at email: materuml@dlsu.edu.ph.