MACHINE LEARNING-BASED SMALL-SCALE PARAMETER EXTRACTION FOR IMPROVED WIRELESS CHANNEL MODEL FIDELITY

Ben Earle[a], Ala'a Al-Habashna[ab], Gabriel Wainer[a], Xingliang Li[c], and Guoqiang Xue[c]

[a]Systems and Computer Engineering, Carleton University, Canada
benearle@cmail.carleton.ca, {alaaalhabashna, gwainer}@sce.carleton.ca
[b]Al Hussein Technical University, Amman, Jordan
alaa.alhabashna@htu.edu.jo
[c]Engineering, Ericsson, Canada
{xingliang.li, guoqiang.xue}@ericsson.com

ABSTRACT

This paper introduces a methodology to improve simulated wireless channel model fidelity. The methodology involves developing machine learning models using synthetic data to extract channel characteristics. Cluster Delay Line (CDL) channel models, which are state-of-the-art models defined by 3GPP, are commonly used in industry for MIMO link-level simulations, since they support a wide variety of environments and frequency ranges. However, configuring the parameters for the simulation of such models is non-trivial. As such, many studies only use the pre-set model configurations. This research investigates how to extract CDL channel parameters from over-the-air channel information to provide more diverse and accurate simulation models. The research focuses on the extraction of several small scale CDL channel parameters; specifically, the cluster angles and gains. The model parameters are estimated from Channel State Information logs using machine learning models trained on synthetic data.

Keywords: wireless channel reconstruction, direction of arrival, Cluster Delay Line, machine learning

1 INTRODUCTION

Wireless network Research and Development (R&D) is a massive industry that relies heavily on simulation [1]. Proper Over-The-Air (OTA) testing of such wireless networks is very difficult as it requires expensive hardware, significant time investment and large and dynamic deployment areas. This makes simulation-based testing an important stage in the R&D process. The overall fidelity of the model is important to accurately test systems prior to investing in OTA testing. The higher the simulation accuracy, the fewer the discrepancy between simulated and actual performance. Improved wireless simulations will reduce the amount of time spent conducting OTA tests which reduces the time and cost spent on R&D. This research proposes improving the simulation accuracy through studying advanced channel models and their parameter generation processes. Existing works have studied analytical techniques to estimate model parameters using ray tracing [2] [3]. These analytical solutions require a 3D map and complete knowledge of the physical environment, which is not widely available and is subject to change over time. Instead, this study will use Machine Learning (ML) to extract the simulation parameters from data that is easily collected from any deployed wireless network. As far as the authors know, they are the only people exploring machine-learning based data-driven channel model tuning. Other studies working on improving CDL model parameters use expensive laboratory equipment to measure one specific scenario, as is done in [4].

Proc. of the 2024 Annual Simulation Conference (ANNSIM'24), May 20-23, 2024, American University, DC, USA
C. Ruiz-Martin, B. Oakes and R. Cárdenas, eds.

This work is a part of a collaborative effort with Ericsson Canada, with the objective of improving wireless channel model fidelity by using simulation models and Machine Learning (ML) to tune the simulated model parameters. In particular, the research presented in this paper builds on the Cluster Delay Line (CDL) channel model, which is a state-of-the-art model developed by 3GPP for 5G and beyond (5GB) networks [5]. CDL models offer realistic multi-path representation through modeling each path between the transmitter and receiver as a cluster of rays. They are adaptable to many propagation environments and frequency ranges. The models are a hybrid between stochastic and geometric channel models. Instead of using ray tracing like normal geometric models, they assign the system geometry randomly. This way the model maintains the advantages of geometric channel models while drastically reducing the computational complexity. CDL models are commonly used in industry for MIMO link level system simulation due to their balance between efficiency and accuracy. However, configuring and generating the parameters for these models is non-trivial, and as such, many studies test their systems using only the pre-set configurations. The five pre-sets do not cover all potential wireless scenarios, and hence, should only be used as a benchmark. This research is thus focusing on creating physically accurate CDL profiles by recreating channels experienced during OTA tests. A training dataset is required to recreate the CDL profiles using ML. This dataset is generated using a simulation tool that spans the space of potential CDL profiles. The tool then simulates each of the CDL profiles to generate labeled data for a supervised learning study.

This paper builds on previous work from [6] and [7] which used a similar approach to estimate the pre-set CDL profile used and the Line-of-Sight (LoS) status of the channel, respectively. The profile estimator proved that it was possible to learn about the CDL model simulated using the limited set of Channel State Information (CSI) data. The LoS status estimator is an essential first step toward recovering the CDL profile, because LoS is one of the most important parameters assigned to a CDL channel model. This research differs from our previous work in that it focuses on recovering Small Scale Parameters (SSP), such as cluster angles and gains experienced during OTA testing. This is done using Random Forests (RF), Deep Neural Networks (DNN) and Meta-Learning with limited CSI data. In deployed systems, CSI data is measured regularly by the cellular device, known as User Equipment (UE), and sent back to the cell tower, or Base Station (BS), to optimize the transmissions. The BS could easily log the UE's CSI data and save it for further analysis. In this work, it will be used to extract parameters to reconstruct the channel in simulation. However, as the true conditions of the channel are unknown, the physical system cannot provide labelled data to train the ML models. Therefore, we needed to develop a custom simulator to generate synthetic training data. This simulator was made in MATLAB and it leverages the phased array and 5G toolboxes. The simulator closely follows the CDL parameter generation process from the 3GPP technical specification [5]. Two scenarios were considered: the first used channels with a single cluster, and the second used channels with two clusters. In the single-cluster scenario, the only relevant features to estimate were the cluster's angles. In the two-cluster scenario, the angles were estimated along with the relative gains. The ML models showed success in estimating most of the parameters. The trained ML models will be used to estimate parameters from OTA CSI logs to recreate the real channel that was experienced and increase future simulation accuracy.

This study aids in choosing correct model parameters to ensure the simulated environment matches the deployed conditions. Furthermore, this research could contribute to the Minimization of Drive Test (MDT), which was introduced in 5G systems to reduce the continual maintenance operators must do on a cell after deployment [8]. Extracting the physical condition of the channel based on logs regularly collected by the cell provider could reduce the required frequency of operator testing. This ultimately increases the model fidelity which reduces the amount of OTA testing required in wireless systems saves time and money for stakeholders.

The remainder of this paper is split into the following sections: Section 2 will provide an overview of CDL models and CSI in which this study is using as features to train the models. Section 3 covers the design of the simulator which generates the synthetic training data. Section 4 will define the different ML models used, along with their hypermarkets and architectures. Section 5 presents the performance of the models when predicting various SSPs. Section 6 provides a summary of the research along with plans for future works.

2 WIRELESS CHANNEL MODEL OVERVIEW

This section provides a brief overview on wireless simulation models and the CSI data that will be used for channel reconstruction. In practice, the BS optimizes its communication with the UE through channel sounding [9]. This process is a part of the regular system operation. The BS periodically transmits pilot signals, which are signals that are known to the UE. The UE compares the received signal to the known sequence, and can deduce the effects of the channel. The UE estimates the channel and calculates a variety of CSI metrics which are sent back to the BS. The BS then uses the CSI data to optimize the communication link with the UE. It does not require much extra effort to store the CSI received at the BS, which makes it an ideal set of data to use when analyzing OTA channels. As such, it was chosen as the input for our channel reconstruction ML models. Specifically, they will use a limited set of CSI that are most commonly used in deployed systems: Received Signal Strength Indicator (RSSI), Reference Signal Received Power (RSRP), Reference Signal Received Quality (RSRQ), Channel Quality Indicator (CQI), Precoding Matrix Indicator (PMI), and Rank Indicator (RI). The RSSI, RSRP, and RSRQ are power readings in dB or dBm. The CQI is an integer index referencing into an array of predetermined signal to noise ratios. The PMI is another integer index into an array of potential beams available for the BS to transmit on. Each index in this array is a matrix of complex numbers which translates to the gain and phase shift that needs to be applied to each antenna to transmit on the desired beam. Lastly, the RI is the number of unique paths that are detected between BS and the UE [7].

The CDL channel model has several configurable parameters[5], they are split into three categories: System Configuration Parameters (SCP), Large Scale Parameters (LSP), and SSP. The aforementioned five pre-sets are example CDL model configurations and they offer a complete set of SCP, LSP, and SSP which show a specific environment. There is also a custom CDL profile generation algorithm which provides a workflow with a series of stochastic equations to generate a new set of random parameters. The algorithm starts from the SCP to generate the LSP, and then uses the LSP to generate the SSP. The SCP consist of generic wireless system configurations, such as, the carrier frequency, bandwidth, the distance between BS and UE, the UE's physical position relative to the BS (in 3D), the desired propagation environment (urban vs rural, macro vs micro cell, etc.), user environment (indoors/outdoors), and the presence of a line of sight cluster. All these parameters will influence the probabilistic functions used to generate the LSP. There are seven LSP: shadow fading, K-factor, delay spread, and azimuth and zenith angle spread for arrival and departure angles. These traits will then influence the generation of the SSP, which are the individual cluster delays, powers, azimuth and zenith angles of arrival and departure, and the cross power ratios. The algorithm then uses the SSP to calculate the channel matrix as a function of time for the simulation [7].

The SSPs are the parameters that are going to be estimated in this study. As a proof of concept, this study will focus on simplified scenarios, specifically channels with a one or two clusters. The cluster powers are in dB, their measurement is taken relative to the highest power cluster. The cluster delay is taken relative to the first cluster that arrives at the BS. Note that in the one-cluster scenario, the delay and gain will both be zero, as it will have the highest power and be the first signal to arrive at the BS. The only parameter estimated for the single cluster will be the angles. The angle of departure is the angle in which the signal leaves the BS to arrive at UE via the given cluster path. The angle of arrival is the direction that the signal arrives at the UE. Since this is a 3D spatial model, the arrival and departure are represented by two angles. The first is the azimuth angle, which is the horizontal angle, i.e., the direction in a plane parallel to the ground. The second is the zenith angle, which is the vertical direction, i.e., the direction in a plane perpendicular to the ground.

Through this study we tried to estimate both arrival and departure angles. However, there was insufficient information present in the CSI recorded at the BS to estimate the angles of arrival. This is expected as these

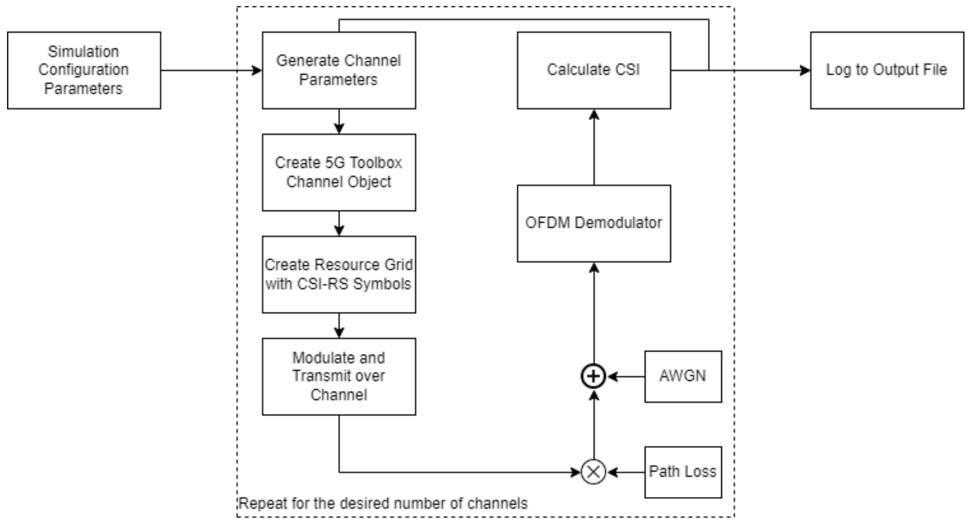

Figure 1: Simulator architecture [7].

are the angles in which the cluster paths hit the UE. This information is not relevant for the BS's decision making, therefore not included in the CSI fed back from the UE. Future works will either need to generate the arrival angles at random or acquire more data to estimate them accurately. Additionally, there was also an attempt to recover the second cluster delay, which was unsuccessful for the same reasons.

3 SYNTHETIC DATA GENERATION

This section focuses on the simulation used to generate the synthetic training data. The simulator for this study is an updated version of the simulator used in [7], using a hybrid of MATLAB's 5G toolbox and custom functions for channel generation, following the process described in 3GPP TS 38.901[5], and adapted from [10]. The simulator architecture is shown in Figure 1. The SCP are configured and passed as input to the simulation, which will use them to generate LSP and SSP for the CDL channels. These parameters are used to generate custom CDL model objects for the MATLAB 5G Toolbox. Each of the channels is simulated to calculate the CSI. The 5G toolbox functions were used due to their ability to assists in building the OFDM frames, configuring the channel sounding sequences, generating the actual channel response from the CDL model parameters, modulating the signal, and calculating most of the CSI metrics. Each channel's CDL parameters and the resulting CSI metrics from the simulations are logged to an output file for future analysis.

The synthetic dataset used in this study was generated using the SCP shown in Table 1. 5,000 unique random channels were generated for each of the scenarios to provide sufficient training data for the ML models. Other dataset sizes were tested; it was found that performance of models trained on 5,000 samples was noticeably better than those trained on 2,500. Increasing to 10,000 samples did not result in a significant improvement over models trained using 5,000 samples. The simulator took approximately one day to complete 5000 channel sounding simulations. The propagation scenario for each channel was Urban Macro (UMA), which has a suggested maximum cell size of around 500m [5]. The maximum speed is used to calculate the maximum Doppler shift within the cell. A max speed of 16 m/s was chosen as this equates to 60km/h, which is a reasonable maximum driving speed within a densely populated urban environment. The carrier frequency is 2GHz with a 15kHz subcarrier spacing. The BS antenna array was configured to be a uniform rectangular array with 32x4 antennas, with spacing of half the wavelength. The UEs were distributed uniformly throughout a 120° segment of a hexagonal cell between 15 and 500m. Only one segment is considered. In a deployed system, there would be an antenna array for each segment (adding up to 360° coverage). The paths generated were filtered to only contain paths to would arrive at the BS from within the cell segment. Paths arriving from outside the segment would be serviced by a different antenna array, therefore they were not considered in this study. Each of the unique CDL channel parameters are used to generate a channel matrix and simulate a channel sounding sequence. The CDL parameters along with the resulting CSI metrics are all collected for the synthetic dataset. Each line in the output represents one single CDL channel and contains the CSI measured by the simulation and the SSP that went into creating the channel matrix. This simulation experiment is designed to provide the ML models with the features (i.e., CSI) and the labels (i.e., SSP) needed for supervised learning.

Table 1: Simulation parameters.

Input	Value
Number of Channels	5000
Scenario	UMa
Cell Size	500m
Max Delay Spread	1000ns
Max Speed	16m/s
Carrier Frequency	2GHz
Subcarrier Spacing	15kHz
Transmitter Antennas	32
Receiver Antennas	4

4 MACHINE LEARNING MODELS

Prior to training any models the data is split between the features (CSI metrics) and the labels (SSP). The features are passed through a MinMaxScaler transform; this is essential because they are all in different scales. For instance, the RSSI can be above 100, while the RI is going to be a single digit. If the features are used without scaling the larger features will have more impact on the network, increasing the amount of learning the model will require. Scale all the input features will improve the performance, while leaving the values being predicted untouched. Next the data is split into testing and training datasets. 30% of the data (1500 channels) are selected at random to serve as the testing set, while the remaining 70% (3500 channels) are used for training and validation of the models. The hyper parameters were tuned using k-fold cross validation on the training and validation set. The machine learning models are trained and tested on Google Collab, a free to use online ML compute tool for researchers [11]. This specialized hardware is designed to improve the training time of DNN made in TensorFlow (TF). TF is a ML library designed by Google that is widely used to build DNNs [12]. There are three machine learning model types used in this study: an RF, several DNN, and a meta learning model that combines all the estimation results. An RF model is built using many decision trees, each decision tree is made of a series of nodes and branches. The nodes in the tree will do some kind of comparison, then select a branch based on the feature value. Each tree is trained independently with different hyper parameters and starting conditions [13]. When estimating a value, the RF will average the result of all the decision trees. The size of the individual decision trees (width and depth), along with the number of trees are hyper parameters for this model. This method combines many weak learners to make an accurate prediction. It was chosen for this study because it is quick to train and effective. The SKLearn Python library [14] was used to implement and optimize the RF hyper parameters. Although slower to train, the DNN offer improved performance in certain scenarios. The DNN's are implemented using TF and their hyper parameters are the number of layers, the number of nodes per layer, the learning rate, and the activation function used in each layer. A potential method to improve the results is to use a hyper parameter optimization library. Instead, four DNN architectures were chosen labeled N1-N4 to span the different DNN configurations. The model parameters for these architects are shown in Table 2. Models

Table 2: DNN architectures.

N1

Layer	Neurons	Activation
Input	38	Relu
Hidden 1	32	Relu
Hidden 2	8	Relu
Output	1	Linear

N2

Layer	Neurons	Activation
Input	38	Relu
Hidden 1	256	Relu
Hidden 2	64	Relu
Output	1	Linear

N3

Layer	Neuron	Activation
Input	38	Relu
Hidden 1-5	32	Relu
Hidden 7-12	16	Relu
Hidden 13-17	8	Relu
output	1	Linear

N4

Layer	Neuron	Activation
Input	38	Relu
Hidden 1-4	256	Relu
Hidden 5-8	128	Relu
Hidden 9-12	64	Relu
Hidden 13-16	32	Relu
Output	1	Linear

N5

Layer	Neuron	Activation
Input	5	Relu
Hidden 1	32	Relu
Hidden 2	8	Relu
Output	1	Sigmoid

1-4 are shallow and narrow, deep and narrow, shallow and wide, and deep and wide, respectively. The meta-learning model is a NN made to combine the result of all the other models described here. Each model will give its prediction, then these will act as the inputs for N5, which will combine them and find an optimal solution. The same models will be trained and tested for all SSP being estimated in this study, for both the one- and two-cluster scenarios. The best performing model for each parameter will be presented in the results section.

5 RESULTS

The synthetic data generated by the channel simulators is used to train ML models, the best performing models for each SSP are presented in this section. The results are split between two subsections, one for the scenario with one-cluster and the other with two-clusters. As mentioned in Section 2, the one-cluster scenario is limited to cluster angle estimations since the delay and gain are normalized relative to the first or dominant power cluster. The performance metrics used in this study were the Mean Absolute Error (MAE) and the R squared (R2) values. The MAE is the absolute value of the errors summed up and divided by the number of testing samples. The R2 value is a metric that measures how much of the variance in the data is captured in the output [15]. The goal is to achieve the MAE value and an R2 value of 1. In this section the model's prediction results on the testing data are shown plotted against the true values in density plots. The lighter colors in the plots indicate a higher density of points. The goal is to have the predicted value equal the true value. This means that a good result will have a high point density (i.e. lighter color) around the y=x line. The title of each plot contains the parameter being estimated, along with the calculated MAE and R2 values. Moving forward, the best performing models will be applied to OTA CSI data to extract the closest possible set of SSP. If it is not possible to extract a given parameter, then its value will need to be generated from the LSP instead of estimated form the CSI data. This set of SSP will be used to recreate those OTA channels as closely as possible in simulation. These new CDL profiles will act as a more accurate representation of true channels for future simulation based testing. This data driven approach to estimate channel parameters in post is not widely studied, as such, there are no other algorithms to compare these results too.

5.1 One Cluster

The one cluster scenario has four labels that can be predicted by the models; Azimuth of Arrival (AoA), Zenith of Arrival (ZoA), Azimuth of Departure (AoD) and Zenith of Departure (ZoD) angles. The best performance achieved is presented in Figure 2. The models were successful in predicting the departure angles (AoD and ZoD). The best model for the AoD was N5, i.e., the meta learning combining model. It achieved an impressive R2 value of 0.81, and an MAE of 10°. The y=x line is clearly highlighted on the density plot, with some variance, particularly towards the edges of the cell. This is likely due to the decrease in beam granularity the further the angle is from perpendicular to the array. The best model for the ZoD plot was the RF, it yielded a very low MAE of 2.6°, with an R2 value of 0.43. The ZoD estimation is more challenging because the cell radius is much larger than the BS height, this causes most clusters to arrive around with a direction that is perpendicular to the array or just below. Since most labels will be in that range, the model does not always predict those clusters that deviate (due to some form of reflected path near the BS). A potential challenge to recovering the signal direction as an angle based on CSI is that the angle is primarily coded as a PMI. Several of the PMI codes correspond to beams that have more than one lobe. This means that the same PMI code could facilitate two different cluster angles which will confuse the ML models. The angles of arrival (signal's direction at the UE) were not well predicted. This is a limitation of the data made available for this study. It was hypothesised in Section 2 that this could happen, but it was tested anyway. The models did manage to identify a trend in some of the points (as can be seen by the non-zero R2 and the slight trend from 90-150°). The trend was not strong enough to make a meaningful prediction, this is seen in the diagram as the predicted and true values are uncorrelated. This is also indicated by the large MAE values and R2 of 0. The CSI data used focuses on the metrics the BS needs to know to optimize the transmission to the UE. The angle of arrival at the UE is not one of those metrics. In a practical system it would be wasted overhead for the UEs to feed this information back to the BS, as it will not be used. This information is included in something known as a Sounding Reference Signal (SRS), which is like the CSI only in the uplink direction (for transmissions from UE to BS). In future works it could be used to predict the AoA and ZoA. In simulation it is easy to generate the SRS and run the models, however in practice this would require UE cooperation (collecting and sharing logs with the network operator). This UE cooperation causes complexities since the UE and BS are often manufactured separately.

5.2 Two Cluster

In this scenario, the models will predict the departure angles of both paths and their gains. The two-cluster scenario is more complex than one-cluster alternative. As such, there is no need to try estimating the arrival paths again without changing the experiment. The results are shown in Figure 3. The best model for AoD was N5 and the performance was lower than expected based on the one-cluster scenario. For the estimated values of AoD, an MAE of 19 and R2 of 0.43 were achieved, and the models could not recover the second cluster AoD. This can likely be improved with more feature data, as we are operating with limited CSI, or more sophisticated models, as this parameter was the best performing in the one cluster scenario. The RF performed best for ZoD, which had the best results in the two-cluster test. The MAE was 3.29 and 2.94, while the R2 was 0.46 and 0.61, for the first and second cluster, respectively. This exceeded the performance of the ZoD estimates in the one-cluster scenario, and was likely because the ZoD is heavily correlated to the UE's distance and LOS status. The increase in cluster count likely helps the model determine the UEs position, and ultimately its ZoD angles. The final parameter was the cluster gains, for which reasonable results were achieved with N5. The MAE was 1.71 and 1.75, with an R2 of 0.32 and 0.68, for the first and second clusters, respectively. The dominant cluster will always have a gain of zero and the other cluster's

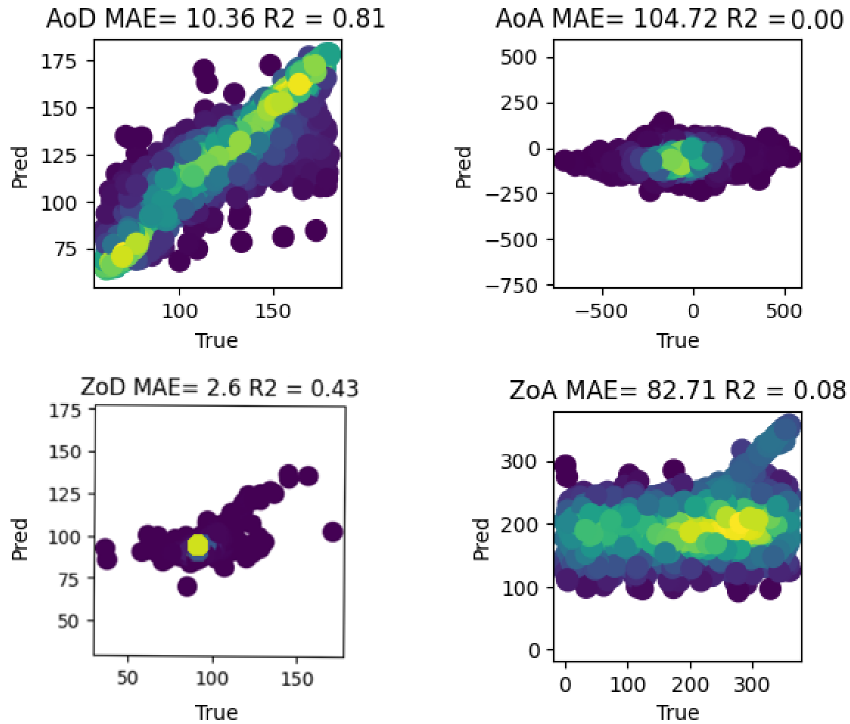

Figure 2: Best performing one cluster estimations.

gain will be with respect to the dominant cluster. This makes zero the most common value to predict and leads to inconsistency in the gains could prove to be a challenge for the ML models long term, especially when scaling the project to include more clusters. In a future work the model will be trained to identify which cluster is the dominant cluster (gain of zero). Then a seperate model will predict the gain of the other cluster(s) with the expected dominant cluster as input. This should make it easier for the ML model than estimating them independently.

6 CONCLUSION

This research proposes a method to improve simulated CDL channel model fidelity through reconstructing wireless channels experienced during OTA testing. The simulation parameters are extracted using ML models trained on synthetic data, generated through a custom channel simulator. The resulting models can extract parameters from real data to create CDL channels that better represent the operating environment for subsequent simulations. This study focused on the extraction of SSP and had success extracting the AoD and ZoD in the one-cluster scenario. The two-cluster scenario is more difficult and yielded mixed results. The AoD estimates were good for the first cluster but not the second. The ZoD was improved on both clusters compared to the one-cluster scenario. Finally, the gain was estimated with better accuracy than expected. Parameters that could not be extracted from the CSI data will need to be generated following the random parameter generation process. This will create a simulated model that is closest to the true environment. These results show progress toward the goal of recreating OTA channels in CDL models.

There are several ways to improve on this work to achieve more accurate parameter extraction. Some planned future works include using an automated hyper-parameter optimization tool for the DNNs. Additionally, each channel could be simulated for more than one channel sounding sequence. The CSI collected over a series of time could be used in an RNN to track the channel and more accurately identify the true

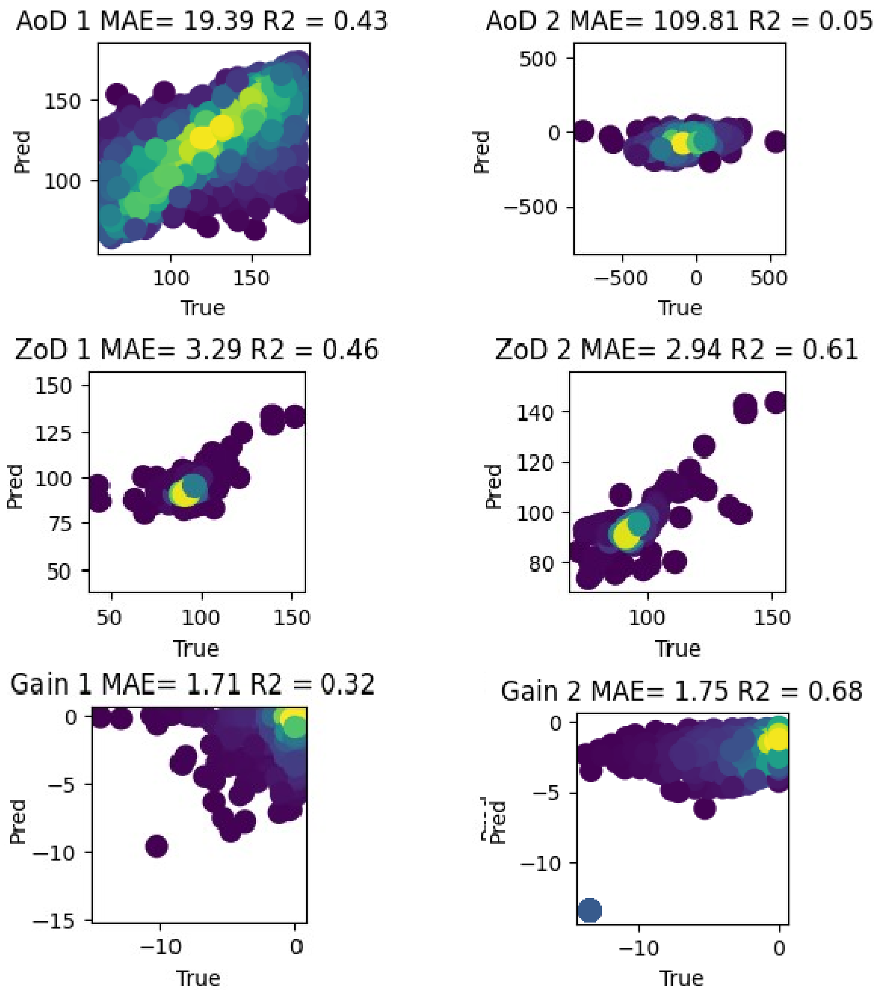

Figure 3: Best performing two cluster estimates.

channel parameters. Alternatively, a matrix of CSI over time could be used in a CNN to identify trends in the data. Additionally, the model can be expanded to support channels with a higher number of cluster paths. In this scenario a ML model could identify the number of relevant clusters present in the channel, then a subsequent model could predict the SSP for each of the clusters. Finally, these types of ML models could be used to predict the LSP, in addition to the SSP. This would give a more complete view of the channel and allow for the generation of SSP parameters that are not able to be estimated from the data available.

REFERENCES

[1] M. S. Obaidat and D. B. Green, *Simulation of Wireless Networks*. Boston, MA: Springer US, 2003, pp. 115–153. [Online]. Available: https://doi.org/10.1007/978-1-4419-9218-5_6

[2] J. Yao, H. Ren, and Q. Liu, "Massive MIMO channel modeling using map-based ray tracing method," in *2017 7th IEEE International Symposium on Microwave, Antenna, Propagation, and EMC Technologies (MAPE)*, 2017, pp. 1–5.

[3] R. Gentile, D. Garcia-Alis, and H. Chen, "Algorithms to antenna: Clustered-Delay-Line channel model with ray tracing," https://www.mwrf.com/technologies/embedded/systems/article/21155377/mathworks-algorithms-to-antenna-clustered-delay-line-channel-model-with-ray-tracing.

[4] Z. Fu, H. Cui, X. Zhao, Y. Wang, Z. Wang, L. Qiu, and Y. Wang, "An improved CDL model for 5G millimeter wave communication in a substation scenario," in *2020 IEEE/CIC International Conference on Communications in China (ICCC Workshops)*, 2020, pp. 18–22.

[5] T. 3rd Generation Partnership Project, "Study on channel model for frequencies from 0.5 to 100 ghz tr 38.901," 2020.

[6] B. Earle, A. Al-Habashna, G. Wainer, X. Li, and G. Xue, "Prediction of 5G new radio wireless channel path gains and delays using machine learning and csi feedback," in *2021 Annual Modeling and Simulation Conference (ANNSIM)*, 2021, pp. 1–11.

[7] ——, "Machine learning models for channel status classification in m-MIMO systems using limited csi feedback," in *2023 Annual Modeling and Simulation Conference (ANNSIM)*, 2023, pp. 645–656.

[8] D. Baumann, "Minimization of drive tests (mdt) in mobile communication networks," in *2016 IEEE 17th International Symposium on A World of Wireless, Mobile and Multimedia Networks (WoWMoM)*, 2014, pp. 9–15.

[9] E. Dahlman, S. Parkval, and J. Skold, *5G NR: The Next Generation Wireless Access Technology*. Academic Press, 2018.

[10] D. Riviello, F. Di Stasio, and R. Tuninato, "Performance analysis of multi-user MIMO schemes under realistic 3gpp 3-d channel model for 5G mmwave cellular networks," *Electronics*, vol. 11, 01 2022.

[11] "Google Colaboratory," https://colab.google/.

[12] Google Brain team, "Introduction to TensorFlow," https://www.{TensorFlow}.org/learn.

[13] A. Burkov, *The Hundred-Page Machine Learning Book*, 2019.

[14] "Random forests and other randomized tree ensembles," https://scikit-learn.org/stable/modules/ensemble.html#random-forests.

[15] D. Figueiredo, S. Júnior, and E. Rocha, "What is r2 all about?" *Leviathan-Cadernos de Pesquisa Polútica*, vol. 3, pp. 60–68, 11 2011.

AUTHOR BIOGRAPHIES

BEN EARLE is a PhD student in Computer Engineering at Carleton University under the supervision of Dr. Gabriel Wainer and Dr. Ala'a Al-Habashna. His research interests include wireless communication, embedded systems, modeling and simulation, and machine learning. His email address is benearle@cmail.carleton.ca.

ALA'A AL-HABASHNA received his Master of Engineering degree from Memorial University of Newfoundland in 2010, and his PhD degree from Carleton University in 2018, both in Electrical and Computer Engineering. Currently, Dr. Al-Habashna is an Adjunct Research Professor at Carleton University and a senior researcher at Statistics Canada, Ottawa, Canada. His current research interests include 5G wireless networks, IoT applications, multimedia communication over wireless networks, discrete-event modeling and simulation, signal detection and classification, cognitive radio systems, and applied machine learning and computer vision. His email is alaaalhabashna@sce.carleton.ca.

GABRIEL WAINER is a Professor in the Department of Systems and Computer Engineering, Carleton University (Ottawa, ON, Canada). His current research interests are related with modeling methodologies and tools, parallel/distributed simulation, and real-time systems. He is a Fellow of SCS. His e-mail is gwainer@sce.carleton.ca. His website is www.sce.carleton.ca/faculty/wainer.

XINGLIANG LI is a Developer, Team Leader at Ericsson. He received his PhD is in Telecommunications. His expertise is in algorithm development for 5G and LTE systems. His research is focused on performance verification, channel modeling, and beam management. His email is xingliang.li@ericsson.com.

GUOQIANG XUE is a manager of network system verification at Ericsson. His PhD is in Electrical Engineering. His team is specialized in 5G performance and capacity verification. Guoqiang is an expert in MIMO and beamforming testing and verification. He has filed many IPRs on the topics of MIMO testing and CDMA data transmission. His email is guoqiang.xue@ericsson.com.