Urban Mobile Radio Propagation Prediction at 750 MHz using Symbolic Regression Techniques

Franz M. E. Camilo, Rayner A. S. e Silva, Gustavo F. Rodrigues, Nuno R. Leonor, Rafael F. S. Caldeirinha, Luiz da Silva Mello, Glaucio L. Ramos.

Abstract— **Radio propagation prediction is a fundamental stage in mobile communication systems design and improvement. That prediction is based on propagation models often developed from empirical measurements. That is a complex process, which has recently been boosted by several machine learning and artificial intelligence tools. Among those techniques, symbolic regression is an interpretable model that provides analytical equations based on a database. This work presents a preliminary work that applies symbolic regression techniques to find analytical equations to model mobile radio propagation channels based on measurements made in the urban area of Rio de Janeiro city in Brazil.**

Keywords— **Mobile communications, radio propagation prediction, channel modeling and characterization, artificial intelligence, machine learning, symbolic regression.**

I. Introduction

Mobile communications have become one of the most important tools for human activities in our times. They allow us to be connected everywhere we go and expand our personal experiences, social interactions and business. Consequently, this technology is developing really fast with expressive changes after each generation, which has sought better data rates and low latency at the expense of higher frequencies and wider bandwidths. The 5G is the current mobile generation and utilizes FR1 and FR2 as two different spectrum bands. The FR1 (410 MHz-7.125 GHz) is also called sub-6 GHz spectrum and allows a broader coverge area and the operation of the conventional cellular lines. The FR2 (24.25 GHz-52.6 GHz) is known as millimeter wave and offers high data rates in small areas. On the other hand, the future next generation is expected to reach even higher frequencies with broader bandwidths and higher communication speed up to 100-1000 times the one reached by the 5G [1].

A move up in the spectrum after every generation makes the new technologies to face new propagation characteristics and challenges, and new analyses are required to characterize mobile radio propagation channels towards future applications.

Franz Camilo, PPGEL, UFSJ, São João del Rei/MG, Brazil, e-mail: franz.camilo@hotmail.com; Rayner Silva, PPGEL, UFSJ, São João del Rei/MG, Brazil, e-mail: raynераugustosilva18@gmail.com; Gustavo Rodrigues, UFSJ, PPGEL/DETEM, Ouro Branco/MG, Brazil, e-mail: gfernandes@ufsj.edu.br; Nuno Leonor, Telecommunications Institute/Polytechnic Institute of Leiria, Leiria, Portugal, e-mail: nuno.leonor@ipleiria.pt; Rafael Caldeirinha, Telecommunications Institute/Polytechnic Institute of Leiria, Leiria, Portugal, e-mail: rafael.caldeirinha@ipleiria.pt; Luiz da Silva Mello, CETUC/PUC-Rio, Rio de Janeiro/RJ, Brazil, e-mail: larsmello@puc-rio.br; Glaucio Ramos, UFSJ, PPGEL/DETEM, Ouro Branco/MG, Brazil, e-mail: glopesr@gmail.com.

However, those channels are complex to analyze as a consequence of the movement, infinity of obstacles, and many unknown environmental aspects especially in urban areas. Then, propagation is often observed as a stochastic process, and the development of channel models for predictions becomes very useful. That modeling results in an easier description of the propagation phenomenon and simplifies the understanding of those mobile communication medium. Specifically, empirical models are based on several measurements made in the place of interest and represents how the radio wave statistically behaves in that specific scenario [2].

Free Space, Okumura, Hata, COST 231, and Longley-Rice are some well known models. Recently, the 3rd generation partnership project (3GPP) has worked for the standardization of channel models regarding the development of the future 6G mobile communication systems. For instance, the 3GPP technical report (TR) 38.901 contains the 5G channel model for frequencies between 0.5 and 100 GHz. WINNER is also a standard channel model widely used for mobile communication systems design. No specific model is suitable for all situations and might demand adjustments. Lately, machine learning (ML)-based artificial intelligence (AI) has become an important tool for radio propagation modeling and the next mobile geration development. Besides it is flexible and accurate, it can learn and offer insights from measured data that is not possible with conventional methods, especially for complex environments [3].

In that perspective, the paper [4] presents a ML path loss prediction improvement and impulsive noise removal for measurements in 2.4 GHz in a public square with vegetation. The least RMSE found was 0.39 dB. On the other hand, deep learning was used to predict radio propagation behavior from images of the receiver area in [5]. The research in [6] was able to find an accurate model to predict radio power in urban area regarding its buildings, terrain and other obstacles with graph neural network. The work [7] proposes to improve prediction accuracy in path loss models and apply Probabilistic ML by means of images or tabular data. Finally, a generative adversarial network is utilized to produce a high resolution received signal strength map from one of lower resolution obtained from ray-tracing in [8].

Although all of those papers present a promising result, they count on non-interpretable models. Otherwise, a model is interpretable if the connection between its input and output can be logically expressed in a concise way with mathematical equations. One example of interpretable model is the symbolic regression: a machine learning-based regression method, which is able to generate an analytical equation to express an input data set. The work in [9] applied symbolic regression to find new equations to fit Okumura model curves. Similarly, the paper [10] expressed the ITU-R P.1546-6 propagation curves into analytical equations also provided by symbolic regression. Despite the next generation of mobile communication systems is expected to be structered in millimeter-wave pico-cells at 24-28 GHz bands for high data rates, micro-cells at sub-6-GHz frequency bands (700 MHz, 2.5 GHz, and 3.5 GHz) may overlap the smaller cells for wide coverage. Based on those approaches and frequency bands, this paper presents an analytical equation by symbolic regression to model urban mobile radio propagation in 750 MHz. The reference database in this work is no longer graphical curves as in the mentioned papers, but it is now made of real measurements in the urban area of Rio de Janeiro city.

II. Symbolic Regression

Symbolic regression (SR) is an effective way of determining mathematical equations in data sets accurately through machine learning methods using analytical regression. Unlike other conventional methods such as Machine Learning (ML) and Deep Learning (DL), with SR it is possible to find mathematical expressions that can efficiently describe data sets. Conventional methods such as linear or quadratic regression have limitations regarding the necessary manual adjustments of the expression coefficients, while symbolic regression automatically performs the adjustments [11].

The construction of symbolic regression models, which is commonly used in evolutionary codes utilizing artificial intelligence, uses the creation of tree-like structures using primitive functions such as ('+', '-', '*' '/') and terminal nodes. This process aims to identify the appropriate number of nodes and terminals to provide the best fit for that specific data set, resulting in an equation. All possible combinations between primitive functions and nodes are tested using geometric progression to assign the best values. Note that this search takes place in a domain that is, in theory, unlimited and includes all available mathematical operators and constants.

SR starts the process by randomly generating equations with a predefined number of individuals and population size. The first equations tend to have a lower potential than the next ones. As the algorithm evolves, new equations with better performance will emerge as the average error tends to decrease. The genetic programming process is completed at a predetermined point and generates the final equation. Completion can occur for various reasons such as reaching the maximum number of iterations, obtaining an equation with an error lower than a predefined value (e.g. 0.01). It is important to mention that equations of low complexity can fail, while those of high complexity can suffer from overfitting [12].

III. Settings and Procedures

A. Database

Different from similar works based on curves, this research work is based on mobile radio measurements at 750 MHz made in the urban area of Rio de Janeiro city [13]. A continuous-wave (CW) signal was transmitted in vertical polarization by an Arronia HyperLOG 60100 antenna, and a RFS I-ATO1-380/6000 H-plane omnidirectional antenna was used in the receptor. A low-noise amplifier was applied. The system parameters can be seen in Table I. That experimental database contains received signal level captured in two different routes georeferenced by GPS. The route 1 was made around Rodrigo de Freitas lagoon and is represented by the green line in Fig. 1. The route 2 was made in Leblon neighborhood and is represented by the blue line in Fig. 1. The red line in Fig. 1 represents stretches affected by surrounding vegetation in both routes. More details about that radio measurement campaign can be found in [14].

Fig. 1. Map with the mobile measurement routes in the urban area of Rio de Janeiro: route 1 in green line, route 2 in blue line, and red line for stretches affected by surrounding vegetation in both routes [14].

TABLE I

Link Budget of the Measurement System in 750 MHz [14].

Parameter	Value
Transmitted power	10 dBm
TX antenna gain	5 dBi
RX antenna gain	1 dBi
Combined amplification	40 dB
Losses (splitter, cables, etc.)	-7 dB
Noise floor	-120 dBm
Overall dynamic range	169 dB

B. Symbolic Regression Settings

The symbolic regression was used and configured to manipulate the data from the Rio de Janeiro measurement campaign for routes 1 and 2. It was trained to identify and extract from the database equations capable of representing path loss (PL) as a function of H_{Tx} (height of the transmitting antenna) and d (distance between transmitter and receiver).

The algorithm was configured so that the network was trained with 70% of the data randomly selected, and the other 30% were used for validation and testing. The SR was set to begin its search with 100 populations, each consisting of a maximum of 50 individuals, and a maximum limit of 1,000 interactions between populations. This initial objective aims to balance the exploration of the search space with the computational demands, which is an important step in optimizing the hyper parameters and identifying the parameters with the greatest impact on the model's performance.

It was observed that for the model to converge and reduce the Mean Squared Error (MSE) while avoiding overfitting (i.e., when the model excessively adapts to the training dataset and fails to generalize to new data), the population size of the genetic algorithm should exceed 25 but not 50, with less than 1,500 iterations. The equation of the proposed model was defined based on the Root Mean Squared Error (RMSE) values and minimum possible complexity.

IV. Results

The database used in [13] had to be adjusted due to the fast fading found in all the routes and their samples. In order to smooth the curves under these effects and improve the modeling of the signals, the moving average (MA) technique with a window of 50 samples was used. This filter is able to reduce the rapid variations in the signals, preserving long-term trends, and making easier to recognize more consistent patterns for the symbolic regression stage.

Among all the equations provided by the symbolic regression, the Equation 1 was chosen according to what is expected from propagation models such as log-distance behavior and RMSE (root mean square error) acceptable values, once large-scale path loss models often have terms such as $log_{10}(d)$, $log_{10}(f)$, (where d is the propagation distance and f the signal frequency), transmission (H_{Tx}) and reception (H_{Rx}) heights, and mean obstruction height. Thus, the Equation 1 provides physical insights about the radio channel in those scenarios in a simple way. Figure 2 displays the comparison among symbolic regression, Hata and Alpha-Beta models for route 1, and Figure 3 displays the same comparison for route 2. The measured path loss and its moving average with a 50 samples window can be seen on the background in both Figures.

The symbolic regression and the Alpha-Beta models presented a similar RMSE approximately 9 dB lower than Hata model error. In other words, symbolic regression method was significantly more accurate than Hata as it can be seen in Table II for both route 1 and 2. The SR model was visually an offset of Hata diagram in both routes towards better precision. Therefore, those preliminary results on symbolic regression with real measurement data showed the promising potential of this technique in radio propagation prediction.

$$PL(d) = (54.276 - 0.37872H_{Tx}) \times log_{10}(d), \qquad (1)$$

where d is the propagation distance and H_{Tx} is the TX antenna height.

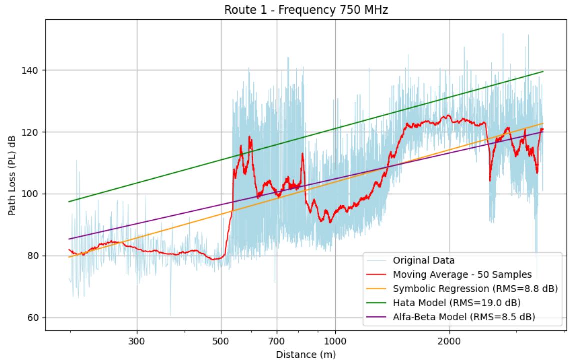

Fig. 2. The figure shows measured path loss for route 1, moving average of measured path loss, and predicted path loss using the SR equation, Hata and Alpha-Beta models.

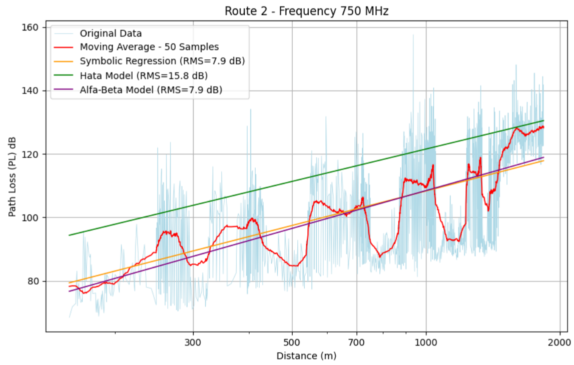

Fig. 3. The figure shows measured path loss for route 2, moving average of measured path loss, and predicted path loss using the SR equation, Hata and Alpha-Beta models..

V. Conclusion

Radio propagation modeling is an important step in the development of wireless and mobile communication systems. No model is universal. Some of them need adjustments and work for specific scenarios and frequencies. There are several models, but the constant move up in the spectrum requires new analyses. Machine learning and artificial intelligence have become important resources for radio prediction. In that perspective, symbolic regression is a powerful tool to provide analytical equations from known diagrams and measured databases. This preliminary work showed that symbolic regression seems to be a promising tool for radio system analysis, improvement, and design. In short, this paper points out that symbolic regression is able to produce results with significant accuracy in comparison to well known models.

VI. Acknowledgments

The authors would like to thank the funding agencies for partially supporting this research work: National Council for Scientific and Technological Development, Brazil (CNPq

TABLE II

RMSE between models and measured path loss.

Model	Route 1	Route 2
Hata	19.0 dB	15.8 dB
Alpha-Beta	8.5 dB	7.9 dB
Symbolic Regression	8.8 dB	7.9 dB

445178/2024-8); FCT - Fundação para a Ciência e Tecnologia, I.P. by project reference 10.54499/UIDB/50008/2020, and DOI identifier https://doi.org/10.54499/UIDB/50008/2020; project ALBATROZ - Altice LaBs Advanced Technology on Radio, Optics and Virtualization (COMPETE2030-FEDER-01335900); and the Foundation for the Support of Research of the State of Minas Gerais (FAPEMIG) for a PhD student scholarship.

REFERENCES

[1] Imam-Fulani, Yusuf Olayinka and Faruk, Nasir and Sowande, Olugbenga A and Abdulkarim, Abubakar and Alozie, Emmanuel and Usman, Aliyu D and Adewole, Kayode S and Oloyede, Abdulkarim A and Chiroma, Haruna and Garba, Salisu and others, “5G frequency standardization, technologies, channel models, and network deployment: Advances, challenges, and future directions," *Sustainability*, v. 15, pp. 5173, 2023.

[2] Salous, S and Haneda, K and Degli-Esposti, Vittorio, “5G to 6G: A paradigm shift in radio channel modeling," *Radio Science*, v. 57, pp. 1–4, 2022.

[3] Huang, Chen and He, Ruisi and Ai, Bo and Molisch, Andreas F and Lau, Buon Kiong and Haneda, Katsuyuki and Liu, Bo and Wang, Cheng-Xiang and Yang, Mi and Oestges, Claude and others, “Artificial intelligence enabled radio propagation for communications—Part II: Scenario identification and channel modeling," *IEEE Transactions on Antennas and Propagation*, v. 70, pp. 3955–3969, 2025.

[4] Gonsioroski, Leonardo and Santos, Amanda and Viana, Jairon and Ferreira, Sandra and Silva, Rogerio and Mello, Luiz da Silva and Matos, Leni and Molina, Marcelo, “Artificial Intelligence Enabled Radio Propagation: Path Loss Improvement and Channel Characterization in Vegetated Environments," *Journal of Microwaves, Optoelectronics and Electromagnetic Applications*, v. 23, pp. e2024277600.

[5] Thrane, Jakob and Sliwa, Benjamin and Wietfeld, Christian and Christiansen, Henrik L, “Deep learning-based signal strength prediction using geographical images and expert knowledge," *GLOBECOM 2020-2020 IEEE Global Communications Conference*, pp. 1–6, 2020.

[6] Bufort, Adrien and Lebocq, Laurent and Cathabard, Stefan, “Data-driven radio propagation modeling using graph neural networks," *arXiv preprint arXiv:2501.06236*, 2025.

[7] Sotiroudis, Sotirios P and Matin, Mohammad A and Wan, Shaohua and Christodoulou, Christos and Goudos, Sotirios K, “A Deep Probabilistic Machine Learning Approach to Ray Tracing Path Loss Prediction at 900 MHz," *IEEE Transactions on Antennas and Propagation*, 2024.

[8] Seretis, Aristeidis and Hashimoto, Takahiro and Sarris, Costas D, “A generative adversarial network approach for indoor propagation modeling with ray-tracing," *2021 IEEE International Symposium on Antennas and Propagation and USNC-URSI Radio Science Meeting (APS/URSI)*, pp. 657–658, 2021.

[9] Ramos, Glaucio and Fernandes, Gustavo and Rego, Cássio and Caldeirinha, Rafael, “Symbolic Regression Applied to Radio Frequency Prediction," *2024 IEEE 1st Latin American Conference on Antennas and Propagation (LACAP)*, pp. 1–2, 2024.

[10] de Oliveira, Guilherme Kneitz and Haddad, Diego Barreto and Giraldi, Gilson Antonio and Dias, Maurício Henrique Costa, “Analytical Expression for Recommendation ITU-R P. l546-6 Propagation Curves of Land Paths Up to 20 km Using Symbolic Regression," *2024 Symposium on Internet of Things (SIoT)*, v. 1, pp. 1–5, 2024.

[11] Tohme, Tony and Liu, Dehong and Youcef-Toumi, Kamal, “GSR: A generalized symbolic regression approach," *arXiv preprint arXiv:2205.15569*, 2022.

[12] Papastamatiou, Konstantinos and Sofos, Filippos and Karakasidis, Theodoros E, “Machine learning symbolic equations for diffusion with physics-based descriptions," *AIP Publishing*, v. 12, 2022.

[13] Leonor, Nuno R and Faria, Stefânia and Ramos, Glaucio and Castellanos, Pedro V Gonzalez and Rodríguez, Carlos and da Silva Mello, Luiz and Caldeirinha, Rafael FS, “Site-specific radio wave measurements for 5G macrocell coverage at 750 MHz, 2.5 GHz and 3.5 GHz signal frequencies," *IEEE Dataport*, Apr. 2020, doi: 10.21227/wmks-4475.

[14] Leonor, Nuno R and Faria, Stefânia and Ramos, Glaucio and Castellanos, Pedro V Gonzalez and Rodríguez, Carlos and da Silva Mello, Luiz and Caldeirinha, Rafael FS, “Site-specific radio propagation model for macrocell coverage at sub-6 GHz frequencies," *IEEE Transactions on Antennas and Propagation*, v. 70, pp. 9706–9715, 2022.