

Wireless propagation modeling in 5G networks utilizing machine learning methods

Vasileios P. Rekkas¹, Leandro dos Santos Coelho², and Sotirios K. Goudos¹

¹Aristotle University of Thessaloniki, {vrekkas, sgoudo}@physics.auth.gr

²Federal University of Parana (UFPR) Department of Electrical Engineering Curitiba, Brazil, leandro.coelho@ufpr.br

The development of precise wireless propagation models is critical for the architectural planning and operational deployment of fifth-generation (5G) network infrastructures and base stations. To satisfy the growing demands and the staged deployment of 5G technologies, highly accurate and effective path loss (PL) predictive models are essential for optimizing network layouts and enhancing deployment strategies. Although traditional propagation models have shown utility, their applicability is frequently constrained by their lack of adaptability to diverse propagation environments and the significant computational investments required by deterministic methodologies. Contemporary advances in data-centric artificial intelligence (AI) have substantially propelled the progress of adaptable and intelligent propagation models, markedly improving the accuracy of PL predictions. This detailed investigation uses an extensive dataset from NYUSIM software, which includes a comprehensive spectrum of atmospheric variables during various seasons in urban microcell configurations in South Asia, enriched with broad empirical data across multiple environmental contexts. The primary objective of the study is to formulate, train, and evaluate four different machine learning (ML) models tailored to predict PL at crucial high frequencies vital to the effective performance of 5G networks. The analysis rigorously scrutinizes each model's ability to accurately predict PL values, focusing on their precision and efficiency in processing large datasets under varying atmospheric conditions.

Index Terms—wireless propagation, path loss, 5G, machine learning.

I. INTRODUCTION

The adoption of wireless communications has significantly intensified, driven primarily by the broadening scope of services being offered. This expansion has effectively augmented the capacity of networks to support an increasing number of users. The formulation of accurate wireless propagation models is crucial for the systematic planning, sophisticated design, and comprehensive evaluation of telecommunications networks. Path loss (PL), a key metric in network performance, can be influenced by a variety of physical and environmental factors including the absorption of electromagnetic waves, reflection phenomena, atmospheric conditions, and diffraction processes. The precision in modeling PL projections is critical, as it ensures the network's optimal performance by supporting rigorous frequency and interference analyses, proactive network coverage management, reliable system configurations, and thorough link budget evaluations [1].

Particularly in the context of fifth-generation (5G) networks, which utilize the millimeter-wave (mm-Wave) spectrum, signal propagation is highly sensitive to environmental factors, which can severely curtail the effective communication range [2]. Addressing this challenge necessitates a deployment strategy where antennas are placed significantly closer to the user base, requiring a density nearly tenfold greater than that of 4G networks, which typically maintain antenna spacings of 100 to 200 meters. Consequently, the establishment of 5G networks involves intricate radio network planning and analysis, employing a comprehensive array of data types and analytical parameters to determine the most advantageous antenna placements. The complexity inherent in these engineering tasks greatly exceeds the capabilities of traditional approaches, necessitating

the adoption of sophisticated data analytics techniques such as machine learning (ML). This strategy aligns with the latest 3GPP technical standards that integrate ML functionalities into advanced network design frameworks, facilitating the progressive advancement of 5G technologies [3].

II. ATMOSPHERIC IMPACT ON 5G SIGNAL ANALYSIS

In this study, a robust dataset is deployed to assess the impact of various atmospheric variables on 5G signal degradation across multiple urban microcell scenarios in South Asia. This dataset encompasses data from simulations across the four distinct meteorological seasons, employing frequencies of 7.125 GHz, 24.25 GHz, 52.60 GHz, and 71 GHz. This extensive frequency range facilitates a comprehensive evaluation of path loss variability in relation to atmospheric conditions such as temperature variability, humidity levels, and rates of precipitation. The results from these simulations demonstrate that while higher frequency bands are prone to greater signal attenuation under less favorable weather conditions, the 7.125 GHz band maintains a higher level of signal integrity, suggesting its effectiveness for sustainable 5G network deployment in urban settings [4].

III. MACHINE LEARNING METHODS

Machine learning (ML) methods have been extensively validated for their utility in addressing quantitative prediction challenges, such as estimating path loss in telecommunication pathways. Regression, a core supervised learning technique, was employed to engage with this continuous outcome variable, leveraging several advanced algorithms to refine prediction fidelity and computational agility. The ensemble algorithms

adopted for enhancing the predictive accuracy of path loss assessments encompass:

- Extreme Gradient Boosting (XGBoost), which leverages sequential model enhancements through gradient boosting to systematically reduce errors in predictive modeling.
- Light Gradient Boosting Machine (LGBM), distinguished by its rapid processing capabilities and adeptness at handling large-scale data sets with minimal computational resource expenditure.
- Random Forest (RF), an ensemble method that aggregates predictions from multiple decision trees to achieve higher accuracy and reduce the likelihood of model overfitting.
- Extra Trees (ETs), which employs a multitude of randomized decision trees to foster greater model diversity and enhance the robustness of the predictions.

IV. EVALUATION METRICS AND NUMERICAL RESULTS

In assessing the suitability of each model against the experimental data, we applied the following evaluation indices:

- Mean Absolute Error (MAE): MAE is employed to evaluate the average magnitude of errors in predictive models, ignoring the direction of these errors. It effectively quantifies the typical absolute discrepancy between predicted outcomes and observed actual values. Mathematically, MAE is expressed as:

$$MAE = \frac{1}{n_{all}} \sum_{i=1}^{n_{all}} |y_i - \hat{y}_i| \quad (1)$$

- Root Mean Squared Error (RMSE): is a robust statistical metric that determines the standard deviation of prediction errors within a model. It accomplishes this by computing the square root of the average of the squared differences between each predicted value and its corresponding actual value. This method of squaring the deviations prior to averaging them accentuates more significant errors, thereby prioritizing larger discrepancies in the evaluation process. The mathematical expression for RMSE is:

$$RMSE = \sqrt{\frac{1}{n_{all}} \sum_{i=1}^{n_{all}} (y_i - \hat{P}_i)^2} \quad (2)$$

- Mean Absolute Percentage Error (MAPE): is utilized to assess the proportion of prediction inaccuracies relative to the actual observed values, expressing these discrepancies as a percentage. This metric offers a standardized measure of error magnitude, facilitating comparisons across different measurement scales and contexts. The computation of MAPE is articulated through the following formula:

$$MAPE = \frac{100}{n_{all}} \sum_{i=1}^{n_{all}} \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (3)$$

where n_{all} denotes the total number of observations, y_i represents the actual PL readings, and $\hat{y}_i$ indicates the predicted values.

The quantitative values for these metrics are outlined in Table I: The systematic evaluation of the four ML models

TABLE I
PERFORMANCE OF ML MODELS ON PL PREDICTION

Model	MAE	RMSE	MAPE (%)
RF	3.32	4.76	2.07
ET	2.88	4.20	1.78
XGBoost	2.97	4.08	1.84
LGBM	2.78	3.93	1.62

reveals distinct levels of performance efficacy. The LGBM model demonstrates the most favorable outcomes, as evidenced by the lowest error measurements, confirming its capability to adeptly handle complex datasets with optimal efficiency. The ET algorithm also exhibits commendable results, highlighting its robustness against model overfitting. On the other hand, the XGBoost, although slightly less effective than LGBM and ET in terms of error metrics, provides a viable compromise between computational expediency and accuracy, suitable for conducting detailed probabilistic assessments. The RF model, despite registering higher error metrics, presents opportunities for performance enhancement through more meticulous calibration or adaptation to specific dataset attributes. Ultimately, model selection must be carefully aligned with the specific data challenges and computational limitations to ensure maximization of performance efficacy.

V. CONCLUSION

This research thoroughly evaluated various ML models for predicting PL in 5G networks under different environmental conditions in urban microcells in South Asia. Through detailed validation using indicators such as MAE, RMSE, and MAPE, it was confirmed that the LGBM model achieved the highest accuracy in predictions, closely followed by the ET and XGBoost models. These results emphasize the importance of advanced ML techniques in managing the complexities of large-scale data to enhance the design and reliability of emerging 5G telecom networks.

These findings offer practical guidance for the enhancement of the 5G network for reliable urban connectivity. With millimeter-wave (mm-Wave) frequencies highly susceptible to environmental influences, integrating solid predictive models into network design is crucial for preventing potential service disruptions.

REFERENCES

- [1] A. Gupta, J. Du, D. Chizhik, R. A. Valenzuela, and M. Sellathurai, "Machine learning-based urban canyon path loss prediction using 28 ghz manhattan measurements," *IEEE Transactions on Antennas and Propagation*, vol. 70, no. 6, pp. 4096–4111, 2022.
- [2] H. Zhang, J. Dong, X. Liu, J. Liu, and X. Zhang, "An artificial intelligence radio propagation model based on geographical information," *IEEE Transactions on Antennas and Propagation*, vol. 70, no. 12, pp. 12 049–12 060, 2022.
- [3] M. F. Ahmad Fauzi, R. Nordin, N. F. Abdullah, and H. A. H. Alobaidy, "Mobile network coverage prediction based on supervised machine learning algorithms," *IEEE Access*, vol. 10, pp. 55 782–55 793, 2022.
- [4] R. H. Ratul, S. M. Mehedi Zaman, H. A. Chowdhury, M. Z. Hassan Sagor, M. T. Kawser, and M. Muntasir Nishat, "Atmospheric influence on the path loss at high frequencies for deployment of 5g cellular communication networks," in *2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, 2023, pp. 1–6.