

# Exploiting Channel Knowledge Mapping for Path Loss Prediction in Aerial Networks

Najmeh Alibabae\*, Antonello Calabro\*, Pietro Cassarà<sup>\*†</sup>, Alberto Gotta<sup>\*†</sup>, Eda Marchetti\*, and Arcangela Rago<sup>†‡</sup>

<sup>\*</sup>ISTI-CNR, Pisa, Italy

<sup>†</sup>CNIT, Consorzio Nazionale Interuniversitario per le Telecomunicazioni, Italy

<sup>‡</sup>Dept. of Electrical and Information Engineering, Politecnico di Bari, Bari, Italy  
na89.alibabae@gmail.com; {name.surname}@isti.cnr.it; arcangela.rago@poliba.it

**Abstract**—Accurate wireless channel modeling is essential for reliable communication in complex aerial networks, particularly in UAV-based infrastructures operating in urban environments. Although deterministic ray tracing (RT) provides high-fidelity channel estimates, its computational cost limits scalability. Machine learning (ML) offers a more efficient alternative, but many existing models rely either on coarse communication-centric features or on raw geometric inputs that may not adequately represent the underlying propagation mechanisms. This paper investigates propagation-aware descriptors extracted from RT interaction paths and associated with identifiable physical and geometric properties, including reflection, diffraction, path curvature, spatial orientation, and vertical variation. Their global relevance to path-loss prediction is assessed using Random Forest (RF) feature importance, while the structure and class separability of the resulting feature space are analyzed through UMAP-based dimensionality reduction. RF and Light Gradient Boosting Machine (LightGBM) regressors are evaluated across six aerial-network scenarios and for uplink, downlink, and joint uplink–downlink configurations. The proposed descriptors achieve competitive or improved predictive performance in most evaluated settings. Although currently derived from RT outputs, they provide a physically grounded intermediate representation that can support explainable channel-prediction models and future hybrid approaches aimed at learning propagation-aware information directly from environmental geometry.

**Index Terms**—Aerial Networks, Channel Knowledge Map (CKM), Machine Learning (ML), ML-based ray tracing (RT), Path loss prediction, Propagation-Aware Features

## I. INTRODUCTION

Accurate wireless channel modeling is essential for the design of modern communication systems based on UAV [1] and satellite technologies [2], [3]. Predicting parameters such as path loss, delay spread, and received power is crucial to ensuring reliable and efficient performance across diverse environments [4]. In 5G and emerging 6G systems, *Channel Knowledge Maps* (CKMs) are becoming central to spatially aware radio optimization [5]. CKMs provide georeferenced maps of key channel parameters—such as path loss, multipath spread, and angular profiles—which are used for location-based beamforming, intelligent resource allocation, and autonomous network operation. However, constructing CKMs in dense urban environments remains challenging due to the complexity of three-dimensional (3D) geometries and the high cost of acquiring measurements. Traditional channel

modeling approaches rely either on analytical formulations or on deterministic ray tracing (RT) simulations [6]. While RT provides high-fidelity insights, it remains computationally prohibitive for large-scale or real-time applications [7]. As a result, data-driven machine learning (ML) methods have gained increasing popularity as an efficient alternative for channel characterization [8]. Most ML-based models employ simple geometric descriptors—such as transmitter–receiver distance or obstacle count—because they are easy to extract and correlate with channel quality. However, these features are often inadequate in capturing the underlying propagation phenomena in complex environments. Parameters like angles of arrival or departure and delay can describe certain aspects of channel dynamics, but they remain only weakly linked to the physical geometry of the environment. In this work, we introduce a class of *propagation-aware* descriptors derived from RT simulations, specifically designed to characterize ray-surface interactions. Unlike traditional parameters such as path loss, these descriptors preserve latent spatial and directional structures that can reveal deeper relationships between the channel and the environment. Such representations can help ML models generalize more effectively across different deployment scenarios without relying directly on raw 3D geometry. By encoding physically meaningful propagation properties, the proposed descriptors provide a physically grounded feature space for Explainable Artificial Intelligence (XAI), supporting global analyses of model behavior and, through feature-attribution methods, local explanations of individual path-loss predictions [9], [10].

The main contributions of this paper are as follows: (i) a unified representation of propagation-aware features encapsulating geometric and environmental attributes; (ii) a statistical and ML-based assessment of their relevance to signal behavior; and (iii) an evaluation of their predictive performance across multiple environments and link directions.

The remainder of the paper is organized as follows. Section II reviews related work, while Section III describes the dataset generation and feature-extraction process. Section IV assesses feature relevance and class separability, and Section V presents the predictive evaluation. Finally, Section VI concludes the paper and outlines future research directions.

## II. RELATED WORK

To overcome the limitations of deterministic models, ML techniques have emerged as effective tools for channel modeling and path loss prediction. These data-driven methods learn nonlinear mappings directly from measurements, avoiding handcrafted propagation equations. However, their accuracy strongly depends on the selected input features and on the quality and diversity of the training data.

ML-based approaches can be broadly categorized by the type of input features used: *communication-centric* or *geometry-aware*.

**Communication-centric features** rely on easily measurable parameters such as distance, angles, and frequency. They are lightweight and scalable but often overlook spatial context, leading to reduced accuracy in complex environments. Several works have followed this strategy: path delay, reflection angle, and carrier frequency were employed in [7] and [11] using artificial neural network (ANN) and back-propagation neural network (BPNN) models for path loss prediction under Line-of-Sight (LOS) and No-Line-of-Sight (NLOS) conditions. Other studies modeled path loss mainly as a function of distance [12], or jointly predicted path loss and delay spread versus frequency [13]. ANN-based models have also analyzed the influence of distance and height in air-to-ground and air-to-air channels [14], confirming the practical simplicity but limited generalization of communication-only inputs.

**Geometry-aware features**, derived from 3D maps or LiDAR data, integrate environmental context into ML models, improving robustness and generalization [15]. Examples include the use of distance, height, terrain, and building shadowing in [16] for signal strength prediction, and the multi-feature framework in [17], which combined digital terrain models, land-use classes, building heights, and user traces to train multiple regression models. Height-dependent models, such as [18], further demonstrated how altitude variations impact multipath effects across diverse environments.

Recently, several studies have combined both feature types. Works such as [19] and [20] merged manually selected and deep learning-extracted features, incorporating environmental descriptors like average building height and street width as lightweight proxies for full 3D geometry. This hybrid approach enables improved path loss modeling while maintaining practical data requirements. Physically meaningful input features may also provide a suitable basis for XAI methods, as they enable model outputs to be interpreted in relation to identifiable environmental and propagation properties [9], [10].

Existing approaches typically rely on coarse communication parameters, detailed 3D representations, or latent features with limited physical interpretability. This work instead investigates propagation-aware descriptors derived from ray-surface interactions and evaluates their relevance, interpretability, and predictive performance across aerial and terrestrial scenarios.

## III. DATASET GENERATION AND FEATURE ENGINEERING

To evaluate the predictive capability of the proposed propagation-aware features, we utilized channel data gener-

ated with the SIFTER framework [21], a modular ray-based simulation pipeline designed for integrated terrestrial and non-terrestrial communication systems. SIFTER reconstructs 3D environments from geographic data, configures transmitter and receiver parameters, and employs an RT engine to simulate end-to-end signal propagation.

### A. Simulation Setup and Path Data

The scenarios cover urban and suburban regions in Pisa and an urban region in London. Ground nodes are placed between 0.5 and 10 m above ground level, representing devices from handheld terminals to rooftop installations. UAVs operate in two altitude ranges: 40–60 m and 90–110 m. Grid layouts provide regular spatial coverage, whereas random layouts introduce geometric diversity. Table I summarizes the six configurations.

The Pisa environment contains areas with different building densities and heights, including narrow historical streets and wider modern roads. The London environment contains 30 ground nodes arranged on a  $6 \times 5$  grid and 10 additional randomly placed nodes. Forty high-altitude UAVs are distributed across this area.

The framework models complete radio wave interaction paths between transmitters and receivers, where each path is represented as a sequence of interaction points corresponding to reflection, diffraction, transmission, or foliage penetration in 3D space. Each simulated propagation path is represented as an ordered sequence of 3D interaction points, as summarized in Table II, where Tx stands for transmitter and Rx for receiver, respectively, and  $x$ ,  $y$ , and  $z$  represent the Cartesian locations. Integer codes identify the physical event associated with each point: transmitter (1), reflection (2), transmission (3), diffraction (4), foliage entry or exit (5), diffuse scattering (6), and receiver (9). For example, the code 12429 denotes a path that leaves the transmitter, reflects, diffracts, reflects again, and reaches the receiver. Figure 1 illustrates representative interactions.

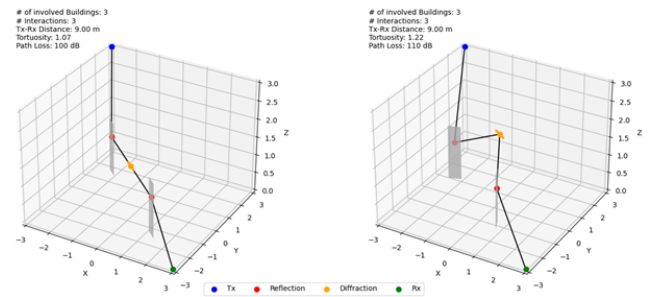


Fig. 1. Examples of path interactions with buildings and the corresponding labeling according to the pattern code.

From each interaction path, we compute a set of propagation-aware descriptors that quantify its geometric, topological, and physical complexity. These features serve as the foundation for the predictive modeling of communication parameters such as received power and path loss.

TABLE I  
SUMMARY OF SIMULATION SCENARIOS.

| ID | Scenario                | Area Size (km)   | Ground Nodes        | UAV Nodes | UAV Altitude | Placement                        |
|----|-------------------------|------------------|---------------------|-----------|--------------|----------------------------------|
| 1  | Pisa Urban              | $2.7 \times 2.4$ | 40 random           | 15        | High         | random                           |
| 2  | Pisa Urban              | $2.7 \times 2.4$ | 40 random           | 15        | Low          | random                           |
| 3  | Pisa Suburban           | $1.5 \times 2.4$ | 40 random           | 13        | High         | random                           |
| 4  | Pisa Suburban           | $1.5 \times 2.4$ | 40 random           | 17        | Low          | random                           |
| 5  | London Urban            | $1.0 \times 1.5$ | 30 grid + 10 random | 40        | High         | random                           |
| 6  | Pisa Urban and Suburban | $4.0 \times 2.4$ | 40 grid             | 14        | Low          | UAVs clustered near ground nodes |

TABLE II  
EXAMPLE SUBSET OF PATH INTERACTIONS FROM THE DATA.

| Tx ID | Rx ID | Pattern | $x$      | $y$      | $z$     |
|-------|-------|---------|----------|----------|---------|
| 1     | 1     | 19      | -2097.75 | -1192.27 | 67.9891 |
| 1     | 1     | 19      | -2104.35 | -1201.92 | 13.5000 |
| 1     | 1     | 129     | -2097.75 | -1192.27 | 67.9891 |
| 1     | 1     | 129     | -2104.29 | -1201.83 | 13.0000 |
| 1     | 1     | 129     | -2104.35 | -1201.92 | 13.5000 |

Pattern codes: 1 = Transmitter; 2 = Reflection; 3 = Transmission; 4 = Diffraction; 5 = Foliage entry/exit; 6 = Diffuse scattering; 9 = Receiver.

### B. Propagation-Aware Descriptors

More than 30 numerical features, i.e., descriptors, were extracted for each propagation path, capturing its geometric structure, interaction statistics, and physical characteristics. Although all scenarios share the same operating frequency, their specific environmental constraints introduce variations that may affect the feature distributions. To ensure comparability and generalization, all datasets were standardized using z-score normalization. The main feature categories are summarized below.

**Geometry and distance:** number of interactions, total path length, direct Tx-Rx distance, and frequency-normalized distance. The quantity currently referred to as Delay\_Spread is path specific and is computed relative to the direct propagation time; throughout this paper it should therefore be interpreted as *excess delay*, computed as the excess propagation time with respect to the LOS distance. These parameters jointly capture the spatial and temporal dispersion of the propagation paths.

**Segment Statistics:** Average, minimum, and maximum segment lengths summarize the granularity of intermediate path segments, while variance and coefficient of variation (CV) quantify their heterogeneity.

**Angular and Curvature Features:** Directional variations are characterized through the mean interaction angle, total path curvature (sum of turning angles), and their standard deviation, reflecting the overall irregularity and bending of the path.

**Vertical and Height Metrics:** The height difference between Tx and Rx, the mean and standard deviation of interaction heights, and the maximum elevation change per segment characterize the vertical structure of the path.

**Spatial Spread and Shape:** The bounding-box volume (and its frequency-scaled variant) quantifies spatial extent. Horizontal spread and tortuosity indicate path indirectness, while the straightness index and interaction density provide

complementary compactness measures.

**Interaction Diversity:** The counts of reflection, diffraction, transmission, and scattering events describe the physical composition of the path. The interaction diversity metric measures the number of unique propagation phenomena, while the first and last interaction types indicate boundary events.

**Directionality and Planarity:** Azimuth and elevation angles capture path orientation; the planarity index, derived from the Principal Component Analysis (PCA) eigenvalue ratio, quantifies flatness; and segment direction variability measures the angular dispersion along the trajectory.

## IV. FEATURE RELEVANCE AND CLASS-SEPARABILITY ASSESSMENT

### A. Random Forest Feature Relevance

A Random Forest (RF) regressor [22] is used to estimate global feature relevance for path-loss prediction. Separate models are fitted for uplink and downlink data. The reported score is the normalized Mean Decrease in Impurity (MDI).

For an internal node  $n$ , let  $\mathcal{S}_n$  be the samples reaching that node,  $W_n = \sum_{i \in \mathcal{S}_n} w_i$  their total weight, and  $\bar{y}_n$  their weighted target mean. For squared-error regression, the node impurity is

$$I(n) = \frac{1}{W_n} \sum_{i \in \mathcal{S}_n} w_i (y_i - \bar{y}_n)^2. \quad (1)$$

If node  $n$  is split into left and right children  $n_L$  and  $n_R$ , the impurity decrease is

$$\Delta I(n) = I(n) - \frac{W_{n_L}}{W_n} I(n_L) - \frac{W_{n_R}}{W_n} I(n_R). \quad (2)$$

A node is said to split on feature  $X_j$  when its decision rule compares  $X_j$  with a threshold  $\tau_n$ , assigning samples satisfying  $X_j \leq \tau_n$  to one child node and the remaining samples to the other.

Let  $\mathcal{N}_t(j)$  denote the set of internal nodes of tree  $t$  whose split rule uses feature  $X_j$ . The unnormalized importance of  $X_j$  is

$$\widetilde{\text{FI}}_j = \frac{1}{T} \sum_{t=1}^T \sum_{n \in \mathcal{N}_t(j)} \frac{W_n}{W_{\text{root},t}} \Delta I(n), \quad (3)$$

where  $T$  is the number of trees and  $W_{\text{root},t}$  is the total sample weight at the root node of tree  $t$ . The normalized feature importance is then

$$\text{FI}_j = \frac{\widetilde{\text{FI}}_j}{\sum_{k=1}^p \widetilde{\text{FI}}_k}, \quad \sum_{j=1}^p \text{FI}_j = 1. \quad (4)$$

where  $p$  is the total number of input features. When no explicit sample weights are used,  $w_i = 1$ , and  $W_n$  reduces to the number of training samples reaching node  $n$ .

Because decision-tree splits depend primarily on the ordering of feature values, strictly monotonic transformations generally do not alter the fitted partitions, provided that all other conditions remain unchanged. However, MDI should not be interpreted as an unbiased measure of feature relevance. Correlated variables may share or redistribute importance, while continuous or high-cardinality features may receive inflated scores because they offer more potential split points. Therefore, MDI is used here as a model-specific indication of feature contribution rather than as evidence of causal or physical relevance.

### B. UMAP-based Class-Separability Analysis

Uniform Manifold Approximation and Projection (UMAP) [23] was employed to obtain a two-dimensional visualization of the standardized feature representations. It was not used to estimate individual feature relevance. UMAP constructs a neighborhood graph in the original feature space and optimizes a low-dimensional graph with similar local connectivity.

Feature vectors were standardized and projected using 15 neighbors and a minimum distance of 0.1 to ensure local continuity. Four labels are considered: uplink-LOS, uplink-NLOS, downlink-LOS, and downlink-NLOS. Label-based separation is summarized through the **Silhouette score** and **Davies-Bouldin index (DBI)**. For sample  $i$ , the Silhouette value is:

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}}, \quad (5)$$

where  $a(i)$  is the average distance between sample (i) and the other samples belonging to the same class. For each different class, the average distance between sample (i) and the samples in that class is computed, and  $b(i)$  is the minimum of these average distances. Values closer to 1 indicate greater within-class cohesion and between-class separation. For  $K$  labels, DBI is:

$$\text{DBI} = \frac{1}{K} \sum_{k=1}^K \max_{\ell \neq k} \frac{S_k + S_\ell}{M_{k\ell}}, \quad (6)$$

where  $S_k$  is the within-label dispersion of class  $k$  and  $M_{k\ell}$  is the distance between label centroids (i.e., centroids of classes  $k$  and  $\ell$ ). Higher Silhouette and lower DBI indicate better separation.

The baseline feature set, adopted from [8] as a reference representation, includes the Cartesian Tx-Rx separations  $d_x$ ,  $d_y$ , and  $d_z$ , the three-dimensional distance  $d_{3D}$ , the propagation state  $s$ , and the angles of departure and arrival. Table III reports mixed results. The proposed representation achieves a lower DBI in five of the six scenarios, suggesting more compact and separated class groups according to this metric. However, the Silhouette Score is higher in only three scenarios and lower in the remaining three. Overall, the propagation-aware descriptors improve class separability in several configurations.

TABLE III  
CLASS SEPARABILITY IN THE UMAP EMBEDDING

| Scenario   | Silhouette Score |               | DBI           |               |
|------------|------------------|---------------|---------------|---------------|
|            | Proposed         | Baseline      | Proposed      | Baseline      |
| Scenario 1 | <b>0.4107</b>    | 0.3289        | <b>0.7209</b> | 0.7655        |
| Scenario 2 | <b>0.4446</b>    | 0.3898        | 0.6938        | <b>0.5749</b> |
| Scenario 3 | 0.2485           | <b>0.2766</b> | <b>0.9731</b> | 1.3569        |
| Scenario 4 | 0.2128           | <b>0.3599</b> | <b>1.3826</b> | 1.8172        |
| Scenario 5 | 0.3026           | <b>0.3101</b> | <b>0.8920</b> | 3.3267        |
| Scenario 6 | <b>0.2285</b>    | 0.1527        | <b>1.3554</b> | 1.6221        |

## V. COMPARATIVE EVALUATION

The feature-relevance analysis presented in Section IV was intended to support the interpretation of the proposed representation rather than to perform feature selection. Therefore, all extracted propagation-aware descriptors were retained for the predictive evaluation. This choice allows the regression models to exploit potentially complementary geometric, spatial, and interaction-related information, including nonlinear dependencies among descriptors.

We compare the proposed propagation-aware representation with the baseline feature set described in Section IV. Both feature representations are used to predict path loss across the six scenarios reported in Table I. Two supervised regression algorithms are considered: an RF regressor and a Light Gradient Boosting Machine (LightGBM) regressor [22].

LightGBM is a gradient boosting framework that constructs additive tree-based models by minimizing a loss function in a stage-wise manner. Given a training set  $\{(\mathbf{x}_i, y_i)\}_{i=1}^n$ , LightGBM learns a function  $F(\mathbf{x})$  composed of  $K$  regression trees:

$$F(\mathbf{x}) = \sum_{k=1}^K f_k(\mathbf{x}), \quad f_k \in \mathcal{F} \quad (7)$$

where each  $f_k$  is a decision tree from the functional space  $\mathcal{F}$ . At each iteration, the model adds a new tree  $f_k$  to minimize the regularized loss:

$$L^{(k)} = \sum_{i=1}^n \ell(y_i, F^{(k-1)}(\mathbf{x}_i) + f_k(\mathbf{x}_i)) + \Omega(f_k) \quad (8)$$

where  $\ell(\cdot)$  denotes the loss function and  $\Omega(f_k)$  is a regularization term controlling the complexity of  $f_k$ .

Although the RF regressor was introduced in Section IV for feature-relevance analysis, it is also evaluated here as a predictive model and compared with LightGBM. The comparison considers two distinct tree-based learning strategies: bagging in RF and gradient boosting in LightGBM [22].

### A. Model Configuration

Each dataset corresponds to a distinct simulation scenario and contains approximately 15,000 samples (propagation paths) for both uplink and downlink. For each path, the geometric, angular, spatial, and interaction-related descriptors described in Section III were extracted and used as model inputs.

TABLE IV

WITHIN-SCENARIO PATH-LOSS PREDICTION USING THE PROPOSED AND BASELINE [8] FEATURE SETS. LOWER RMSE AND HIGHER  $R^2$  ARE BETTER.

| Scenario | Model    | Uplink   |              |          |              | Downlink |             |          |       | Uplink + Downlink |              |          |       |
|----------|----------|----------|--------------|----------|--------------|----------|-------------|----------|-------|-------------------|--------------|----------|-------|
|          |          | Proposed |              | Baseline |              | Proposed |             | Baseline |       | Proposed          |              | Baseline |       |
|          |          | $R^2$    | RMSE         | $R^2$    | RMSE         | $R^2$    | RMSE        | $R^2$    | RMSE  | $R^2$             | RMSE         | $R^2$    | RMSE  |
| 1        | RF       | 0.8199   | 7.73         | 0.8227   | <b>7.67</b>  | 0.7974   | <b>8.06</b> | 0.6620   | 10.42 | 0.7743            | <b>8.69</b>  | 0.7436   | 9.26  |
|          | LightGBM | 0.8751   | <b>6.44</b>  | 0.8221   | 7.68         | 0.8109   | <b>7.79</b> | 0.6614   | 10.43 | 0.8370            | <b>7.38</b>  | 0.7516   | 9.11  |
| 2        | RF       | 0.8263   | <b>10.26</b> | 0.7766   | 11.63        | 0.8067   | <b>9.96</b> | 0.7180   | 12.03 | 0.8088            | <b>10.53</b> | 0.7763   | 11.38 |
|          | LightGBM | 0.8367   | <b>9.95</b>  | 0.7966   | 11.10        | 0.8093   | <b>9.89</b> | 0.7213   | 11.96 | 0.8388            | <b>9.66</b>  | 0.7923   | 10.97 |
| 3        | RF       | 0.8558   | 11.65        | 0.9271   | <b>10.42</b> | 0.8579   | <b>7.28</b> | 0.6666   | 11.14 | 0.8690            | <b>11.44</b> | 0.8663   | 11.56 |
|          | LightGBM | 0.9317   | 10.08        | 0.9397   | <b>9.48</b>  | 0.8740   | <b>6.85</b> | 0.7099   | 10.39 | 0.9202            | <b>8.93</b>  | 0.8843   | 10.75 |
| 4        | RF       | 0.8882   | 9.17         | 0.9579   | <b>8.08</b>  | 0.9208   | <b>8.12</b> | 0.7984   | 12.95 | 0.9140            | <b>10.32</b> | 0.9123   | 10.42 |
|          | LightGBM | 0.9379   | 9.82         | 0.9553   | <b>8.33</b>  | 0.9308   | <b>7.58</b> | 0.8332   | 11.78 | 0.9308            | <b>9.26</b>  | 0.9138   | 10.33 |
| 5        | RF       | 0.7665   | <b>7.16</b>  | 0.6923   | 8.23         | 0.7494   | <b>8.30</b> | 0.6254   | 10.15 | 0.7826            | <b>7.39</b>  | 0.6619   | 9.21  |
|          | LightGBM | 0.8172   | <b>6.34</b>  | 0.7433   | 7.51         | 0.7881   | <b>7.63</b> | 0.6921   | 9.20  | 0.8134            | <b>6.84</b>  | 0.7068   | 8.58  |
| 6        | RF       | 0.8710   | <b>7.96</b>  | 0.8069   | 8.86         | 0.8853   | <b>6.43</b> | 0.7416   | 8.30  | 0.9361            | <b>6.73</b>  | 0.8914   | 8.22  |
|          | LightGBM | 0.9158   | <b>6.43</b>  | 0.8674   | 7.34         | 0.9215   | <b>5.32</b> | 0.7937   | 7.41  | 0.9530            | <b>5.78</b>  | 0.9174   | 7.17  |

The datasets were divided into 80% training and 20% testing subsets, using a fixed random seed to ensure reproducibility. Model selection was performed through 3-fold cross-validation on the training set, while hyperparameters were independently tuned for each configuration (scenario, link direction, and feature set) using the Optuna framework<sup>1</sup>, with 100 trials and 3-fold validation. For the RF regressor, the tuned parameters included the number of estimators, maximum depth, and split thresholds; for the LightGBM model, learning rate, number of leaves, and subsampling ratios were also optimized.

Each sample represents an individual propagation path. However, a random path-level split does not guarantee independence between the training and test sets, because paths associated with the same transmission link or with closely related geometries may occur in both subsets. Therefore, the proposed evaluation primarily measures performance on randomly held-out paths. Link-disjoint or spatially disjoint splits would provide a stronger assessment of generalization of the unobservable propagation configurations.

### B. Predictive Performance

In this analysis, predictive performance is evaluated using two standard metrics: the coefficient of determination ( $R^2$ ) and the Root Mean Square Error (RMSE). Higher  $R^2$  and lower RMSE values indicate better predictive performance.

As reported in Table IV, the propagation-aware feature set generally provides competitive or improved path-loss prediction compared with the baseline representation. The advantage is particularly evident for downlink and joint uplink-downlink prediction, where the proposed features achieve higher  $R^2$  and lower RMSE in all reported configurations. Uplink results are more mixed: the proposed representation performs better in some scenarios, whereas the baseline remains more accurate in others. Therefore, the results support the effectiveness of the propagation-aware descriptors in most configurations, although the improvement is not uniform across all propagation conditions.

<sup>1</sup><https://optuna.readthedocs.io/>

The two regression algorithms exhibit similar overall trends. LightGBM generally achieves higher  $R^2$  and lower RMSE than RF, although RF is marginally more accurate in a limited number of configurations. Accordingly, LightGBM should be regarded as the better-performing model in most cases, rather than as consistently superior across all scenarios and feature sets. This result suggests that gradient boosting may capture nonlinear interactions among the descriptors more effectively; however, a dedicated analysis would be required to identify the mechanisms responsible for the observed differences.

Overall, the proposed representation provides useful predictive capability and achieves a strong agreement between predicted and reference path-loss values in several configurations. Nevertheless, the remaining prediction errors and the mixed uplink results indicate that further validation is required, particularly under link-disjoint, spatially disjoint, or cross-scenario evaluation settings.

The weaker relative performance observed for uplink prediction in Scenarios 3 and 4 suggests that the usefulness of the propagation-aware descriptors may depend on the propagation environment and link configuration. One possible explanation is the greater heterogeneity of the propagation paths in these scenarios, including variations in ground-level obstruction and geometry. However, this interpretation remains hypothetical and would require a dedicated scenario-specific analysis.

## VI. CONCLUSION AND FUTURE WORK

This study investigated propagation-aware descriptors derived from ray-interaction data for path-loss prediction in aerial-network scenarios. By capturing geometric, angular, topological, and interaction-related properties, the proposed representation achieved competitive or improved performance compared with the baseline feature set in most evaluated configurations, with clearer gains for downlink and joint uplink-downlink prediction. Because the descriptors are associated with identifiable propagation properties, they provide a physically grounded representation that can support the interpretation of model behavior. However, in the current study,

these features are extracted from RT outputs and therefore do not yet remove the computational cost of ray tracing. Rather, they represent an intermediate step toward explainable and computationally efficient surrogate channel models. The extent to which the proposed descriptors can be inferred directly from raw environmental geometry may vary across feature categories.

Future work will therefore investigate modular learning approaches for estimating propagation-aware information from 3D maps without requiring RT at inference time. Graph neural networks, spatial transformers, and attention-based architectures are promising candidates for this task. Additional directions include channel-specific representations, descriptor-level ablation studies, broader comparisons with alternative learning models, repeated evaluations to quantify performance variability, and feature-attribution methods for global and local explanations. More challenging evaluation protocols, including link-disjoint, spatially disjoint, and cross-scenario testing, will also be considered. These extensions will help assess the robustness of the proposed representation across different altitude regimes, deployment configurations, and propagation conditions.

#### ACKNOWLEDGMENTS

The authors thank Tarun Chawla and Remcom Inc. for providing access to the Wireless InSite X3D ray tracer used in the numerical study. The contribution of Arcangela Rago was supported by the PN RIC project "Secure Ubiquitous Networks and Services, SUNS" (code: QIIR112 00048, CUP: B99H26000340007) and the PN RIC project "Centro meridionale per l'innovazione quantistica, Q-SUD" (code: QIIR112 00051, CUP: B99H26000380007), both funded by the Italian MUR, by the project "Artificial Intelligence On-Board Satellite, AIOS" (CUP: B99J25000160005) funded by the Italian MIT, and by the HORIZON MSCA project BRIDGITISE (grant no. 101119554).

#### REFERENCES

- [1] M. Bacco, S. Chessa, M. Di Benedetto, D. Fabbri, M. Girolami, A. Gotta, D. Moroni, M. A. Pascali, and V. Pellegrini, "UAVs and UAV swarms for civilian applications: communications and image processing in the SCIADRO project," in *International Conference on Wireless and Satellite Systems*. Springer International Publishing Cham, 2017, pp. 115–124.
- [2] A. Machumilane, P. Cassarà, and A. Gotta, "Toward a fully-observable markov decision process with generative models for integrated 6g-non-terrestrial networks," *IEEE Open Journal of the Communications Society*, vol. 4, pp. 1913–1930, 2023.
- [3] A. Cardaci, L. Caviglione, A. Gotta, and N. Tonello, "Performance evaluation of spdy over high latency satellite channels," in *International Conference on Personal Satellite Services*. Springer International Publishing Cham, 2013, pp. 123–134.
- [4] L. Wu, Y. Cai, Y. Song, Y. Cao, H. Wang, and L. Yan, "Artificial neural network based path loss prediction for wireless communication network," *IEEE Access*, vol. 8, pp. 199 523–199 538, 2020.
- [5] L. Gholami, P. Ducange, A. Rago, P. Cassarà, and A. Gotta, "Explainable machine learning for environment-aware channel state prediction in uav-based 6g networks," in *2025 21th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*, 2025, pp. 1–6.
- [6] S. P. Sotiroidis, G. Athanasiadou, G. Tsoulos, P. Sarigiannidis, C. G. Christodoulou, and S. K. Goudos, "Evolutionary ensemble learning pathloss prediction for 4g and 5g flying base stations with uavs," *IEEE Transactions on Antennas and Propagation*, vol. 71, no. 7, pp. 5994–6005, Jul. 2023.
- [7] H. Li, X. Chen, K. Mao, Q. Zhu, Y. Qiu, X. Ye, W. Zhong, and Z. Lin, "Air-to-Ground Path Loss Prediction Using Ray Tracing and Measurement Data Jointly Driven DNN," *Computer Communications*, vol. 196, pp. 268–276, 2022.
- [8] W. Xia, S. Rangan, M. Mezzavilla, A. Lozano, G. Geraci, V. Semkin, and G. Loianno, "Generative Neural Network Channel Modeling for Millimeter-Wave UAV Communication," *IEEE Transactions on Wireless Communications*, vol. 21, no. 11, pp. 9417–9431, Nov. 2022.
- [9] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Proceedings of the 31st International Conference on Neural Information Processing Systems*, ser. NIPS'17. Red Hook, NY, USA: Curran Associates Inc., 2017, p. 4768–4777.
- [10] A. K. Gizzini, Y. Medjahdi, A. J. Ghandour, and L. Clavier, "Towards Explainable AI for Channel Estimation in Wireless Communications," *IEEE Transactions on Vehicular Technology*, vol. 73, no. 5, pp. 7389–7394, 2024.
- [11] K. Mao, X. Chen, Q. Zhu, W. Zhong, and C.-X. Wang, "MI-based delay-angle-joint path loss prediction for uav mmwave channels," *Wireless Networks*, vol. 30, pp. 3947–3959, Oct. 2021.
- [12] X. Zhao, F. Du, S. Geng, Z. Fu, Z. Wang *et al.*, "Playback of 5g and beyond measured mimo channels by an ann-based modeling and simulation framework," *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 9, pp. 1945–1954, 2020.
- [13] G. Yang, Y. Zhang, Z. He, J. Wen, Z. Ji, and Y. Li, "Machine-learning-based prediction methods for path loss and delay spread in air-to-ground millimetre-wave channels," *IET Microwaves, Antennas & Propagation*, vol. 13, no. 8, pp. 1113–1121, 2019.
- [14] S. Duangsuwan, P. Juengkittikul, and M. Maw, "Path loss characterization using machine learning models for gs-to-uav-enabled communication in smart farming scenarios," *International Journal of Antennas and Propagation*, vol. 2021, pp. 1–10, 2021.
- [15] S. P. Sotiroidis, A. D. Boursianis, S. K. Goudos, and K. Siakavara, "From spatial urban site data to path loss prediction: An ensemble learning approach," *IEEE Transactions on Antennas and Propagation*, vol. 70, no. 7, pp. 6101–6105, Jul. 2022.
- [16] 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, Dec. 2022.
- [17] U. Masood, H. Farooq, M. A. Imran, and A. Abu-Dayya, "Interpretable ai-based large-scale 3d pathloss prediction model for enabling emerging self-driving networks," *IEEE Transactions on Mobile Computing*, vol. 22, no. 7, pp. 3967–3984, Jul. 2023.
- [18] W. Khawaja, O. Ozdemir, and I. Guvenc, "Uav air-to-ground channel characterization for mmwave systems," in *Proceedings of the 2017 IEEE 86th Vehicular Technology Conference (VTC-Fall)*. Toronto, Canada: IEEE, 2017, pp. 1–5.
- [19] S. P. Sotiroidis, S. K. Goudos, and K. Siakavara, "Deep learning for radio propagation: Using image-driven regression to estimate path loss in urban areas," *ICT Express*, vol. 6, no. 3, pp. 160–165, Sep. 2020.
- [20] J. Thrane, B. Sliwa, C. Wietfeld, and H. L. Christiansen, "Deep learning-based signal strength prediction using geographical images and expert knowledge," in *Proceedings of the IEEE Global Communications Conference (GLOBECOM)*. IEEE, Dec. 2020, pp. 1–6.
- [21] N. Alibabae, A. Calabrò, P. Cassarà, A. Gotta, and E. Marchetti, "Advancing the Future of Integrated 5G-Satellite Networks: A Practical Framework for Performance Evaluation, Dataset Generation, and AI-Driven Approaches," in *Proceedings of the 15th International Conference on Simulation and Modeling, Multidimensional Systems and Signal Processing*, 2025.
- [22] A. Choudhury, A. Mondal, and S. Sarkar, "Searches for the bsm scenarios at the llc using decision tree-based machine learning algorithms: a comparative study and review of random forest, adaboost, xgboost and lightgbm frameworks," *The European Physical Journal Special Topics*, vol. 233, no. 15, pp. 2425–2463, 2024.
- [23] L. McInnes, J. Healy, N. Saul, and L. Großberger, "Umap: Uniform manifold approximation and projection," *Journal of Open Source Software*, vol. 3, no. 29, p. 861, 2018.