Hindawi
Wireless Communications and Mobile Computing
Volume 2018, Article ID 8489326, 9 pages
https://doi.org/10.1155/2018/8489326

Research Article

Air-to-Air Path Loss Prediction Based on Machine Learning Methods in Urban Environments

Yan Zhang, Jinxiao Wen, Guanshu Yang, Zunwen He, and Xinran Luo

School of Information and Electronics, Beijing Institute of Technology, Beijing 100081, China

Correspondence should be addressed to Zunwen He; hezunwen@bit.edu.cn

Received 6 April 2018; Accepted 14 May 2018; Published 13 June 2018

Academic Editor: Liu Liu

Recently, unmanned aerial vehicle (UAV) plays an important role in many applications because of its high flexibility and low cost. To realize reliable UAV communications, a fundamental work is to investigate the propagation characteristics of the channels. In this paper, we propose path loss models for the UAV air-to-air (AA) scenario based on machine learning. A ray-tracing software is employed to generate samples for multiple routes in a typical urban environment, and different altitudes of Tx and Rx UAVs are taken into consideration. Two machine-learning algorithms, Random Forest and KNN, are exploited to build prediction models on the basis of the training data. The prediction performance of trained models is assessed on the test set according to the metrics including the mean absolute error (MAE) and root mean square error (RMSE). Meanwhile, two empirical models are presented for comparison. It is shown that the machine-learning-based models are able to provide high prediction accuracy and acceptable computational efficiency in the AA scenario. Moreover, Random Forest outperforms other models and has the smallest prediction errors. Further investigation is made to evaluate the impacts of five different parameters on the path loss. It is demonstrated that the path visibility is crucial for the path loss.

1. Introduction

In recent years, unmanned aerial vehicles (UAVs), as aircraft without pilots on board, have shown great promise due to their high mobility and deployment flexibility. With the development of UAV manufacturing, its cost is reduced while its performance continuously increases. As a result, there are more and more attractive applications for UAV, such as traffic monitoring, emergency rescue, forest fire detection, cargo transport, and so on [1, 2]. Stable and efficient wireless communication links are indispensable in most UAV applications. Therefore, the UAV communications play an important role in the future fifth-generation wireless networks (5G), providing vast coverage and reliable relaying.

Meanwhile, the propagation environment of UAV-aided communication systems differs from that of traditional ones, which brings enormous challenges. An accurate understanding of the UAV wireless channels is crucial for the design and deployment of these communication systems. The wireless signals from/to UAVs may be obstructed and may encounter different propagation conditions along the path. The attenuation of the electromagnetic wave, which is usually described by the path loss, is of great significance for the link budget analysis and network planning for UAV communications. Therefore, many works have been finished to develop flexible and precise models for the path loss in the UAV communication scenarios.

The high altitude range of 500 m to 2000 m was considered in [3], and the air-to-ground (AG) channel based on the curved-earth two-ray model was investigated in various scenarios. In [4], a statistical propagation model was proposed for the UAV channel at low altitude in the urban environment, and it was shown that the prediction results were dependent on the elevation angle between the airborne transmitter (Tx) and the receiver (Rx) on the ground. In [5], the impact of the UAV altitude on path loss exponent and shadow fading in the rural scenario was studied. In [6], statistical path loss models were established by modifying the current 3GPP terrestrial channel models for urban macrocell and rural macrocell scenarios. In [7, 8], measurements were carried out in suburban scenarios, and large-scale parameters and multipath components were extracted and analyzed. In [9], measurement campaigns were conducted in an urban scenario and a distance-dependent model was proposed for the UAV path loss prediction.

Most of these aforementioned works are focused on the AG communication. An alternative application approach of UAVs is to use them as both sides of the communication, i.e., air-to-air (AA) communication. Until now, only a few papers have investigated the channel models in the AA scenario. In [10], Rice model was extended to derive the AA channel parameters and it was reported that the attenuation caused by the distance effect followed a free-space model. In [11], with data generated by a ray-tracing software, the close-in free-space model and excess fading loss model were adopted to characterize the path loss in the AA channel.

Furthermore, the existing works are mainly based on empirical models, such as free-space model and log-distance model, which rely on data collected in specific propagation scenarios. Statistical analysis is performed to build the mapping relationship between path loss and parameters such as propagation distance and flight altitude. The empirical models are computationally efficient and easy to implement. They can describe the statistical characteristics of the path loss at a given distance in the measured scenario. However, the actual path loss at a specific location cannot be obtained. Besides, the accuracy of these models decreases when they are applied to more general environments [12].

Another candidate solution is to utilize deterministic approaches, such as ray tracing and finite-difference time-domain (FDTD), within which the path loss values are calculated by applying radio wave propagation mechanisms and numerical analysis techniques to model computational electromagnetics. With detailed geographic information and dielectric properties of materials, these methods are very accurate and reliable for predicting the spatial distribution of electromagnetic fields. Due to the high cost of carrying out measurement campaigns, the deterministic approaches have been widely used for wireless network planning. The only disadvantage is that the computation procedure consumes huge time and memory resources and thus it is inappropriate to use these approaches for real-time applications. Moreover, the complicated calculation has to be run again once the propagation environment changes.

Actually, path loss modeling is a supervised regression problem and can be solved by machine learning [13]. It has been proved that machine-learning-based models are able to provide more accurate path loss prediction results than the empirical ones and are more computationally efficient than the deterministic approaches [12]. Different algorithms have been adopted to train prediction models in traditional terrestrial communication scenarios. For example, artificial neural networks (ANNs) were used for path loss prediction in urban [14], suburban [15], rural [16], and railway [17] scenarios. Support vector regression (SVR) was applied for the prediction of path loss in suburban environment in [18]. In order to build a connectivity model for an environmental wireless sensor network, several methods, including Random Forest, Adaboost, ANNs, and K-Nearest-Neighbors (KNN), were analyzed and compared in [13]. It was reported that Random Forest performed better than others for the considered complex terrain environments.

In this paper, we build the prediction models for path loss in the AA scenario based on machine learning. Two algorithms, Random Forest and KNN, are taken into consideration. To evaluate the feasibility of the proposed models, the ray-tracing approach is used to generate data for training and testing purposes. In addition, the prediction accuracies of machine-learning-based models are compared with those of the empirical ones, such as the Stanford University Interim (SUI) model [19] and the COST231-W-I model [20]. It is shown that the proposed models outperform the empirical ones. Furthermore, we analyze the commonly used parameters related to the path loss in the AA scenario, including the propagation distance, Tx UAV altitude, Rx UAV altitude, path visibility, and elevation angle. Meanwhile, the importance of these parameters is discussed.

We summarize the major contributions and novelties of this paper as follows.

(1) The path loss for the UAV communication in the AA scenario is modeled based on machine learning methods, including Random Forest and KNN algorithms.

(2) The prediction results are evaluated with the data generated by a ray-tracing software. It is proved that the machine-learning-based models are able to provide better accuracy than the empirical ones.

(3) We analyze the impacts of different parameters on the AA path loss and sort these parameters by their importance.

The remainder of this paper is organized as follows. The considered AA propagation environment and the ray-tracing-based data generation are described in Section 2. Section 3 presents the machine-learning-based methods for path loss prediction. The model training procedure is introduced in Section 4. In Section 5, the performance of machine-learning-based models is evaluated and the importance ranking of different parameters is discussed. At last, conclusions are drawn in Section 6.

2. Propagation Environment Description

In order to investigate the path loss model in the AA scenario, we consider a typical urban environment in which two UAVs are employed as Tx and Rx. As a new emerging scenario, measurements for AA communications are still in a very preliminary stage. Since the machine-learning-based models require a large amount of data for training purpose, a ray-tracing software is employed to generate data for model building and performance evaluation. It has been proved that the channel data obtained by the ray-tracing software are in good agreement with the actual measured values [21]. As illustrated in Figure 1, the considered environment is a region in Helsinki, with dimensions of 1000 m by 600 m. Gray areas and green areas indicate buildings and grounds, respectively. The maximum height of the buildings is 50 m. All the buildings are assumed to be made of concrete with the following dielectric half-space properties: permittivity 6, conductivity 0.02, and thickness 0.3 m. The material of the

Figure 1: Urban environment for AA communications.

ground surface is assigned as asphalt whose permittivity is 10 and conductivity is 0.01.

The simulations were performed at the central frequency of 2.4 GHz, with a bandwidth of 100 MHz. The red square in Figure 1 represented the position of the Tx UAV, which was equipped with a directional antenna. Rx UAV was moved at a spacing of 2 m along six different routes. Different flight altitudes were taken into account for both Tx and Rx UAVs. The altitudes of the Tx UAV included 60 m, 70 m, and 80 m. Meanwhile, Rx UAV were assumed to fly at heights of 10 m, 20 m, 30 m, and 40 m, lower than the maximum height of the buildings. It should be noted that the direct, reflected, and diffracted paths were considered, whereas the penetration paths were neglected because of the high attenuation through building. Details of the parameter setting can be found in Table 1.

Through calculations, we obtain the spatial distribution of received powers. Then, path loss values at different locations can be extracted. In practice, the path loss in the AA scenario is related to many environmental parameters. The goal of the machine learning method is to find the optimal function describing the relationship between these parameters and the path loss. In the following analysis, five parameters which have impacts on the path loss are selected as the input features of the machine-learning-based models and they are listed as follows.

(1) **Propagation distance** (d, in meter): the distance between the Tx and Rx UAVs calculated from their coordinates.

(2) **Tx altitude** (h_t, in meter): the height of Tx UAV from the ground, with three values of 50 m, 60 m, and 70 m.

(3) **Rx altitude** (h_r, in meter): the height of Rx UAV from the ground, with four values of 10 m, 20 m, 30 m, and 40 m.

(4) **Path visibility** (I_v, 0 or 1): parameter indicating whether there exists line-of-sight (LOS) path between the Tx and Rx UAVs. $I_v = 1$ for the LOS case and $I_v = 0$ for the non-line-of-sight (NLOS) case.

(5) **Elevation angle** (θ, $-\pi/2$ to $\pi/2$): the angle between the LOS path and the horizontal.

We collected all the samples when the Tx UAV was in different altitudes and the Rx UAV flied along six routes at

Table 1: Parameter configuration.

Parameter	Value
Environment	Helsinki urban scenario
Area	1000 m × 600 m
Max. building height	50 m
Carrier frequency	2.4 GHz
Bandwidth	100 MHz
Transmit power	15 dBm
Tx altitude	60, 70, 80 m
Rx altitude	10, 20, 30, 40 m
Distance between adjacent Rx positions	2 m
Number of Tx locations	1 × 3
Number of Rx locations	405 × 4
Max. number of reflection	10
Max. number of diffraction	1
Max. penetration	Not simulated

Input:

Training set $X = \{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_N\}$ with responses $Y = \{PL_1, PL_2, \cdots, PL_N\}$, where $\mathbf{x}_i = (d_i, h_{ti}, h_{ri}, I_{vi}, \theta_i)$, $i = 1, ..., N$.
Number of ensemble members T.

Training Process:

For $t = 1$ to T:

(1) Take a bootstrap sample $\{X_t, Y_t\}$ of size N from $\{X, Y\}$.
(2) Use $\{X_t, Y_t\}$ as the training data to train the tth ensemble member by using binary recursive partitioning.
(3) Repeat the following steps recursively for each unsplit node until the stopping criterion is met:
(i) Select m features randomly from the f available features ($f = 5$ in this study).
(ii) Calculate the square error for each possible splitting point of each feature, and find the best binary split among all binary splits on the m features.
(iii) Split the node into two descendant nodes using the best split.

Prediction:

Given a new $\mathbf{x} = (d, h_t, h_r, I_v, \theta)$, the predicted path loss value is obtained by $PL' = (1/T)\sum_{t=1}^{T} \widehat{h}_t(\mathbf{x})$, where $\widehat{h}_t(\mathbf{x})$ is the prediction of the tth ensemble member.

Algorithm 1: Random Forest algorithm for path loss prediction in the AA scenario.

different heights. Each sample was with an output (path loss value) and five input features. Removing the locations where the received signals are too weak to detect, we obtained 5508 samples in total. Then, these samples were separated into two set, training set and test set. The former were utilized to train the models while the latter were employed to evaluate the performance of the trained models.

3. Machine-Learning-Based Models for AA Path Loss Prediction

Machine learning is a method to improve performance on a specific task based on extensive data and a flexible model architecture. In recent years, it has been widely used in many fields like computer vision, speech recognition, autonomous driving, and so on. Machine learning tasks can be broadly classified into supervised learning and unsupervised learning, depending on whether data samples have labels or not. For supervised learning, tasks can be further divided into classification problems and regression problems based on whether the predicted values are discrete or continuous. The AA path loss prediction is a typical regression task, which can be solved by many algorithms, such as Random Forest, ANN, and SVR. We aim to build the path loss prediction model in the AA scenario based on machine learning. With given path loss values and corresponding input features, the model can be trained and then the path loss values in new conditions can be predicted with the various inputs.

In this study, two typical supervised learning algorithms, Random Forest and KNN, are chosen to build prediction models for the AA path loss. Their performance evaluation results will be compared in Section 5 and Random Forest will be proved to have a better agreement with the test data compared with KNN. The major principles of these two algorithms are introduced as follows.

3.1. Random Forest. Ensemble learning, which uses multiple individual learners to solve classification and regression problems, can achieve a significantly superior generalization performance [22]. Random Forest is a commonly used ensemble learning algorithm and employs decision tree as ensemble member. It applies bootstrap aggregating to select training samples for each ensemble member. Ensemble members are trained based on these samples and then the final result is obtained by averaging the results of all the ensemble members.

Besides, Random Forest further introduces random selection of features in the decision tree training process. Usually, the traditional decision tree selects an optimal feature from the feature set of current node for split. For Random Forest, a subset is randomly chosen from the feature set of each node, and then the optimal feature is selected from this subset.

In ensemble learning, the greater the diversity of ensemble members, the better the prediction performance. By introducing sample perturbations and feature perturbations, the diversity of ensemble members in Random Forest is increased. It can improve the generalization performance of the model. It is worth noting that the decision trees grow without pruning, due to the randomness of samples and features. Algorithm 1 shows the method we used for path loss prediction in the AA scenario. The detailed descriptions of Random Forest can be found in [23].

Random Forest is easy to implement and can realize parallel computing. It is also insensitive to input data and can handle thousands of input features. In addition, an important advantage of Random Forest is that it can sort the importance of features. In the following, we will use this property to analyze the significance of different input features for the AA path loss.

3.2. KNN. KNN is a classical machine learning algorithm that is often used to solve classification problems. It has no explicit training process and its implementation is simple. The mechanism of KNN is to find the k training samples closest to the sample to be predicted based on a distance metric and then to perform prediction based on the information of these k neighbors. It is also suitable for regression tasks by averaging the values of k neighbors to get the final prediction result.

The distance metric plays a very important role in KNN. The distance reflects the difference between two samples. Commonly used distance metrics include Manhattan distance, Euclidean distance, and so on. In this study, Euclidean distance is chosen for analysis. In general, features have different ranges of values and their influences on the distance calculation are not the same. As a result, KNN algorithm is more sensitive to the input data compared with Random Forest. For the sake of fairness, the samples need to be normalized before model training. In this study, Z-score normalization method is adopted and it can be expressed as

$$x_N = \frac{x_i - \mu}{\sigma} \tag{1}$$

where x_i is the input value of the feature, x_N is the normalized value, μ is the mean, and σ is the standard deviation.

4. Model Training and Accuracy Metrics

The procedure of machine-learning-based path loss predictors for AA channel is introduced as follows. Firstly, we collect enough data samples for analysis, each with path loss record and corresponding input features. As mentioned above, for KNN the features need to be scaled by the normalization process, while for Random Forest it is not necessary. Secondly, these samples can be divided into two categories, training set and test set, which are used for model training and evaluation purposes, respectively. Thirdly, based on the training data and selected algorithms, we train the model and tune its parameters. Finally, some metrics are employed to assess the prediction accuracy of the trained model, and then in view of the evaluated results we can further improve the machine-learning-based predictor for the path loss in the AA scenario.

In this section, we will introduce the division of training set and test set. Then, the model training process is explained in detail. In addition, accuracy metrics for model validation are presented.

4.1. Data Division. The performance of the machine-learning-based models strongly depends on the amount and quality of training data. In general, more training samples lead to more accurate reflections of the inherent laws. Thus, we must try to obtain enough samples in order to get accurate models for path loss prediction. In addition, the rules extracted from the model training are hidden in the samples, so the training samples must be representative. Different from the training data, the test set is used to assess and further improve the trained models.

As aforementioned, 5508 samples were collected in the considered AA scenario, including samples from six routes at all different Tx/Rx altitudes. In this study, the samples from the third route with the Rx altitudes of 20 m and 30 m were used for test purpose, and they did not participate in the training process. The remaining samples were included in the training set. Then, the proportions of the training samples and test ones were 84% and 16%.

4.2. Model Training. Model training aims at acquiring parameters for the model to optimize the performance and effectiveness of the path loss prediction. Some machine learning algorithms, such as ANN and SVR, have many parameters whose values need to be set before the learning process begins. In contrast, Random Forest and KNN have only a few parameters that need to be tuned and thus they are both efficient for implementation.

For Random Forest, the model accuracy is affected by the parameters including maximum tree depth and the number of ensemble members. The former controls the maximum split number of the decision tree and the latter determines the size of the ensemble. Generally, a small ensemble with deep decision trees has a greater tendency to overfit than a shallow ensemble of many decision trees [13].

For KNN, the number of neighbors, k, is deterministic for the prediction performance. If k is too small, the model becomes complicated and may overlearn when the neighboring points are noises. Meanwhile, large k makes the model structure simple but neighboring samples with large differences will affect the prediction result.

These aforementioned parameters cannot be learned directly from the data. The optimization methods for tuning parameters mainly include grid search, random search, and Bayesian optimization. In this study, grid search is used to find the optimal combination of parameters by searching all possible points in the given range. For Random Forest, we evaluate the following parameters: the number of ensemble members between 10 and 200 (at 10-unit intervals) and the maximum tree depths between 5 and 50 (at 5-unit intervals). For KNN, the value of k is set between 2 to 10 with the interval of 1. The obtained parameters are as follows. For Random Forest, the depth of trees and the number of ensemble members are finally set as 30 and 140, respectively. The number of neighbors is equal to 5 in KNN.

4.3. Metrics for Evaluating Prediction Accuracy. To evaluate the performance of different models, two statistical properties, mean absolute error (MAE) and root mean square error (RMSE) [24], are chosen as metrics. They can be calculated by comparing the predicted path loss with the data in the test set as

$$\mathrm{MAE} = \frac{1}{I}\sum_{i=1}^{I}\left|PL_i - PL_i{}'\right|, \tag{2}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{I}\sum_{i=1}^{I}\left(PL_i - PL_i{}'\right)^2} \tag{3}$$

where I is the total number of test samples, PL_i is the path loss value of the ith sample in the test set, and $PL_i{}'$ is the predicted value.

5. Model Validation and Results

In this section, we will evaluate the performance of these machine-learning-based models in the AA scenario. Two empirical models are also considered for comparison. In

Table 2: Statistics analysis for different predictors.

Evaluation indicators	Random Forest	KNN	SUI	COST231-W-I
MAE (dB)	2.27	4.56	7.54	26.67
RMSE (dB)	3.06	8.90	13.40	28.53

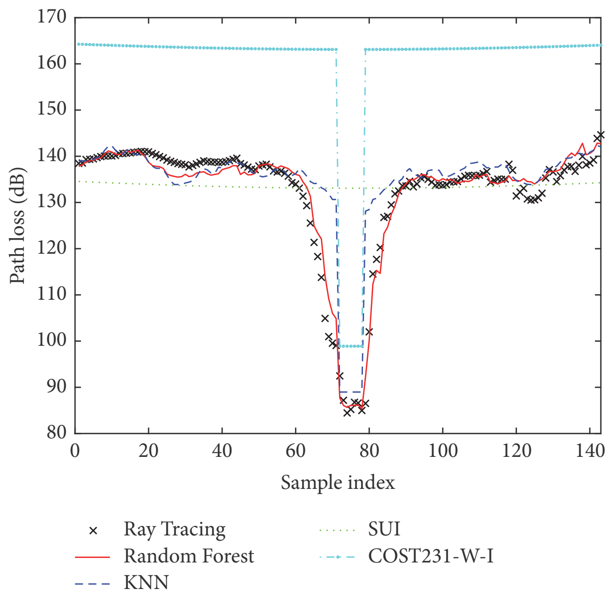

Figure 2: Prediction performance of different models when Rx UAV moves in the third route ($h_t = 60$ m and $h_r = 30$ m).

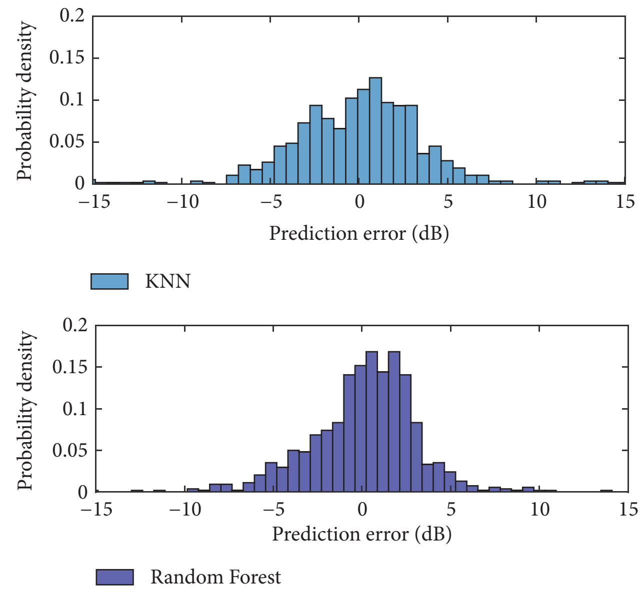

Figure 3: Prediction error distributions for the machine-learning-based models.

addition, the impacts of different features on the path loss are analyzed.

5.1. Comparisons between Empirical Models and Machine-Learning-Based Models. As an example, we consider the samples in the test set gathered when the Tx and Rx altitudes are 60 m and 30 m, respectively. The predicted path loss results from different models are shown in Figure 2. Sample indexes are corresponding to different positions of Rx UAV in the third route from up to down in Figure 1. As mentioned in Section 2, the distance between two adjacent Rx positions is 2 m. As shown in Figure 1, LOS path exists when the Rx UAV moves in the middle, corresponding to the sample index from 72 to 79. It can be found that the path loss values are quite small in this area. In Figure 2, it is illustrated that the machine-learning-based models can accurately approximate the realistic path loss values generated by the ray-tracing software. Two empirical models, SUI model and COST231-W-I model, are chosen for comparison. The description of whether the LOS path exists is not included in the SUI model. Thus, there are large gaps from the path loss results predicted by the SUI model to the true values under the LOS condition. The COST231-W-I model can describe the path loss variations in both LOS and NLOS conditions. Apart from the path visibility, this model also involves the distance, Tx altitude, Rx altitude, and elevation angle into the model parameters. However, its predicted results are much larger than the true values. The major reason may be the fact that the application scenarios of the COST231-W-I model differ from what we used during the analysis. It reflects the poor generalization performance of the empirical model; i.e., its accuracy decreases when it is applied to a different environment. Furthermore, as shown in Figure 2, these empirical models neglect some details of the path loss fluctuations and it is difficult to use them for characterizing the path loss value at a specific location.

Considering all samples in the test set, we can get the statistical assessment of these different models. The MAEs and RMSEs of prediction results are illustrated in Table 2. It is shown that both Random Forest and KNN outperform the empirical models. These machine-learning-based models can also depict the fluctuations of path loss in detail.

5.2. Comparisons between Random Forest and KNN. As listed in Table 2, Random Forest provides the best fit to the test data, with 2.27 dB MAE and 3.06 dB RMSE. KNN also offers acceptable results whereas its predicted values are almost unchanged under the LOS condition. The reason is that, within the KNN model, the path loss is predicted by averaging the values of the nearest k neighbors. Due to the limited number of collected LOS samples, a similar path loss value is probably estimated.

Figure 3 shows the distributions of prediction errors for the two machine-learning-based models. It is shown that most errors concentrate in the range of −5 dB to 5 dB and

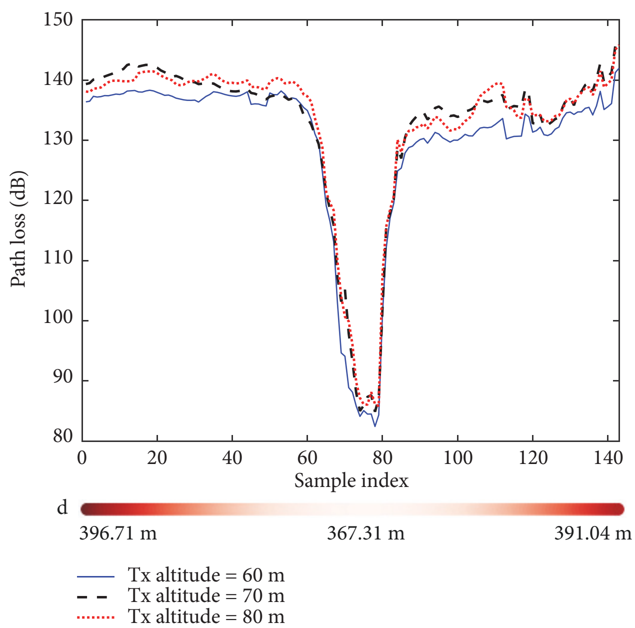

FIGURE 4: Path loss values with different Tx altitudes (h_r = 10 m).

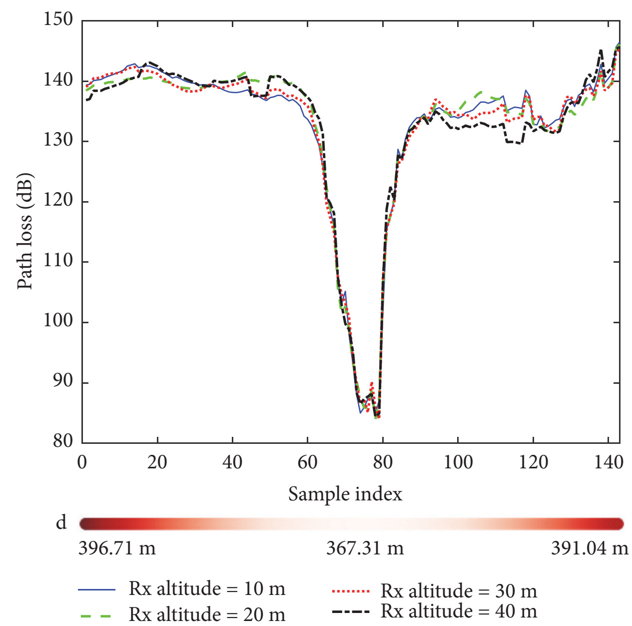

FIGURE 5: Path loss values with different Rx altitudes (h_t = 70 m).

Random Forest shows a higher prediction accuracy than KNN.

5.3. Computational Efficiency. Another aspect to be evaluated is the computational efficiency. The path loss values should be generated in a short time so that the spatial distribution of electromagnetic fields can be quickly updated when the propagation environment changes. The generation durations of our machine-learning-based models are recorded. The computer we used to run the programs has an AMD A8-4500M processor and 4 GB of memory. The required times of Random Forest and KNN predictors are 8.71 s and 5.95 s, respectively. In contrast, running the ray-tracing software would take more than 10 minutes to generate all the samples in the test set. This comparison result is preliminary but it still reflects that the machine-learning-based model can provide higher computational efficiency to the network planning than the deterministic approaches.

5.4. Analysis of Feature Importance. As mentioned, there are many parameters related to the path loss in the AA scenario and they serve as the input features in our machine-learning-based models. For example, the ray-tracing-based path loss values in the third route at different Tx altitudes are shown in Figure 4. The altitude of the Rx UAV is 10 m and three different Tx altitudes are taken into account, including 60 m, 70 m, and 80 m. It is shown that in the selected low-altitude UAV AA scenario, the path loss values at different Tx altitudes are very close. Besides, when the altitude of the Tx UAV is fixed at 70 m. The path loss values at different Rx altitudes are illustrated in Figure 5. According to Figures 4 and 5, the path visibility is vital for the path loss in the considered AA scenario. The propagation distances corresponding to sample indexes are also shown in these two figures.

TABLE 3: Normalized importance of different features.

Feature	Importance
Path visibility	0.7438
Propagation distance	0.1137
Elevation angle	0.1045
Tx altitude	0.0317
Rx altitude	0.0064

The following task is to investigate the significance of different input parameters. Fortunately, Random Forest can give a natural ranking of the features in the model. In this study, the mean decrease impurity method [23] is employed to analyze the importance of features. As introduced above, Random Forest is composed of multiple decision trees. Each node in the decision tree is a condition about a feature, in order to divide the data into two sets according to different response variables. For regression problems, variance or least-squares fitting is often used as impurity. During the training process, it can be calculated how much impurity of the tree is reduced by a feature. For Random Forest, it is possible to calculate the average reduced impurity of each feature and to use it as the importance of the feature.

Table 3 shows the normalized contribution of each parameter used in the model based on Random Forest. The path visibility has the greatest impact, followed by the propagation distance, elevation angle, Tx altitude, and Rx altitude. Similar to results shown in Figures 4 and 5, the flight altitudes of UAVs have small influences on the path loss. A possible reason is that the selected scenario is a low-altitude UAV AA scenario and the heights of buildings are close to the flight altitudes of UAVs.

6. Conclusions

In this paper, we have proposed a modeling mechanism for AA path loss based on machine learning. A ray-tracing software has been utilized to generate the data for an urban AA scenario, which was subsequently divided into a training set and a test set to be used by the models. The models have been learned by two machine learning algorithms, Random Forest and KNN. The test data have been used to evaluate the accuracy performance of these machine-learning-based models and two empirical models, SUI model and COST231-W-I model. It has been demonstrated that machine learning provides a flexible modeling approach based on the training data for such complex environment and Random Forest has the best prediction performance. In addition, we have analyzed the importance of five input features for the path loss in the AA scenario. Results have confirmed that the path visibility is the dominant factor. Propagation distance and elevation angle have also shown great influences.

Since the UAV AA communication is a newly emerging scenario, the channel modeling and path loss prediction in such a scenario are still in a very preliminary stage. Future work should incorporate introduction of more machine-learning-based models like ANN and SVR. Different scenarios should also be taken into account to verify the generalization property of these models. Last but not least, measurement campaigns should be carried out in the AA scenario. More measured data are expected to further improve the performance and feasibility of the machine-learning-based path loss predictors.

Data Availability

The data that support the findings of this study are available on reasonable request from the first author, Yan Zhang (zhangy@bit.edu.cn), or the corresponding author, Zunwen He (hezunwen@bit.edu.cn). The data are not publicly available since they are generated based on the mentioned parameters by commercial software.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by National Nature Science Foundation of China under Grant no. 61201192 and the National High Technology Research and Development Program of China (863 Program) under Grant no. 2015AA01A706. It was also supported by Ericsson Company.

References

[1] Y. Zeng, R. Zhang, and T. J. Lim, "Wireless communications with unmanned aerial vehicles: Opportunities and challenges," *IEEE Communications Magazine*, vol. 54, no. 5, pp. 36–42, 2016.

[2] Z. Xiao, P. Xia, and X.-G. Xia, "Enabling UAV cellular with millimeter-wave communication: potentials and approaches," *IEEE Communications Magazine*, vol. 54, no. 5, pp. 66–73, 2016.

[3] D. W. Matolak and R. Sun, "Air-ground channels for UAS: Summary of measurements and models for L- and C-bands," in *Proceedings of the 2016 Integrated Communications Navigation and Surveillance (ICNS)*, pp. 8B2-1–8B2-11, Herndon, VA, USA, April 2016.

[4] A. Al-Hourani, S. Kandeepan, and A. Jamalipour, "Modeling air-to-ground path loss for low altitude platforms in urban environments," in *Proceedings of the 2014 IEEE Global Communications Conference, GLOBECOM 2014*, pp. 2898–2904, Austin, TX, USA, December 2014.

[5] R. Amorim, H. Nguyen, P. Mogensen, I. Z. Kovács, J. Wigard, and T. B. Sørensen, "Radio channel modeling for UAV communication over cellular networks," *IEEE Wireless Communications Letters*, vol. 6, no. 4, pp. 514–517, 2017.

[6] K. Wang, R. Zhang, L. Wu et al., "Path loss measurement and modeling for low-altitude UAV access channels," in *Proceedings of the 2017 IEEE 86th Vehicular Technology Conference (VTC-Fall)*, pp. 1–5, Toronto, Canada, September 2017.

[7] X. Cai, A. Gonzalez-Plaza, D. Alonso et al., "Low altitude UAV propagation channel modelling," in *Proceedings of the 2017 11th European Conference on Antennas and Propagation (EuCAP)*, pp. 1443–1447, Paris, France, March 2017.

[8] W. Khawaja, I. Guvenc, and D. Matolak, "UWB channel sounding and modeling for UAV air-to-ground propagation channels," in *Proceedings of the 59th IEEE Global Communications Conference, GLOBECOM 2016*, pp. 1–7, Washington, DC, USA, December 2016.

[9] Z. Yang, L. Zhou, G. Zhao, and S. Zhou, "Channel model in the urban environment for unmanned aerial vehicle communications," *Accepted by EuCAP'18*, 2018.

[10] N. Goddemeier and C. Wietfeld, "Investigation of air-to-air channel characteristics and a UAV specific extension to the rice model," in *Proceedings of the IEEE Globecom Workshops, GC Wkshps 2015*, pp. 1–5, San Diego, Cal, USA, December 2015.

[11] L. Zhou, Z. Yang, G. Zhao, L. Xiao, and S. Zhou, "Modeling air-to-air path loss for dense urban environments," in *submitted to IEEE Globecom'18*.

[12] M. Ayadi, A. Ben Zineb, and S. Tabbane, "A UHF path loss model using learning machine for heterogeneous networks," *Institute of Electrical and Electronics Engineers. Transactions on Antennas and Propagation*, vol. 65, no. 7, pp. 3675–3683, 2017.

[13] C. A. Oroza, Z. Zhang, T. Watteyne, and S. D. Glaser, "A machine-learning-based connectivity model for complex terrain large-scale low-power wireless deployments," *IEEE Transactions on Cognitive Communications and Networking*, vol. 3, no. 4, pp. 576–584, 2017.

[14] I. Popescu, D. Nikitopoulos, P. Constantinou, and I. Nafornita, "ANN prediction models for outdoor environment," in *Proceedings of the 2006 IEEE 17th International Symposium on Personal, Indoor and Mobile Radio Communications*, pp. 1–5, Helsinki, Finland.

[15] I. Popescu, I. Nafornita, and P. Constantinou, "Comparison of neural network models for path loss prediction," in *Proceedings of the IEEE International Conference on Wireless and Mobile Computing, Networking and Communications, 2005. (WiMob '2005)*, pp. 44–49, Montreal, Canada.

[16] E. Ostlin, H.-J. Zepernick, and H. Suzuki, "Macrocell path-loss prediction using artificial neural networks," *IEEE Transactions on Vehicular Technology*, vol. 59, no. 6, pp. 2735–2747, 2010.

[17] D. Wu, G. Zhu, and B. Ai, "Application of artificial neural networks for path loss prediction in railway environments,"

in *Proceedings of the 5th International ICST Conference on Communications and Networking in China*, pp. 1–5, Beijing, China, August 2010.

[18] K. Lin, K. Hung, J. Lin, C. Wang, and P. Pai, "Applying least squares support vector regression with genetic algorithms for radio-wave path-loss prediction in suburban environment," in *Advances in Neural Network Research and Applications*, vol. 67 of *Lecture Notes in Electrical Engineering*, pp. 861–868, Springer, Berlin, Germany, 2010.

[19] V. Abhayawardhana, I. Wassell, D. Crosby, M. Sellars, and M. Brown, "Comparison of empirical propagation path loss models for fixed wireless access systems," in *Proceedings of the 2005 IEEE 61st Vehicular Technology Conference VTC'05-Spring*, pp. 73–77, Stockholm, Sweden, May-June 2005.

[20] V. Erceg, "Channel models for fixed wireless applications, tech. rep," in *IEEE 802*, 16 Broadband Wireless Access Working Group, 2001.

[21] P. Mededovic, M. Veletic, and Z. Blagojevic, "Wireless insite software verification via analysis and comparison of simulation and measurement results," in *Proceedings of the 35th International Convention MIPRO*, pp. 776–781, Opatija, Croatia.

[22] W. Hou, D. Shi, Y. Gao, and C. Yao, "A new method for radio wave propagation prediction based on finite integral method and machine learning," in *Proceedings of the 2017 IEEE 5th International Symposium on Electromagnetic Compatibility (EMC-Beijing)*, pp. 1–4, Beijing, China, October 2017.

[23] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[24] J. Isabona and V. M. Srivastava, "Hybrid neural network approach for predicting signal propagation loss in urban microcells," in *Proceedings of the 2016 IEEE Region 10 Humanitarian Technology Conference (R10-HTC)*, pp. 1–5, Agra, India, December 2016.