



# Machine learning based altitude-dependent empirical LoS probability model for air-to-ground communications\*#

Minghui PANG<sup>1,2</sup>, Qiuming ZHU<sup>†1,2</sup>, Zhipeng LIN<sup>†1</sup>, Fei BAI<sup>1</sup>, Yue TIAN<sup>1</sup>, Zhuo LI<sup>3</sup>, Xiaomin CHEN<sup>1</sup>

<sup>1</sup>Key Laboratory of Dynamic Cognitive System of Electromagnetic Spectrum Space, College of Electronic and Information Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China

<sup>2</sup>State Key Laboratory of Integrated Services Networks, Xidian University, Xi'an 710000, China

<sup>3</sup>Key Laboratory of Radar Imaging and Microwave Photonics, Ministry of Education, College of Electronic and Information Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China

E-mail: {pangminghui, zhuqiuming, linlzp, baifei, tian\_yue, lizhuo, chenxm402}@nuaa.edu.cn

Received Feb. 1, 2022; Revision accepted July 15, 2022; Crosschecked Aug. 23, 2022

**Abstract:** Line-of-sight (LoS) probability prediction is critical to the performance optimization of wireless communication systems. However, it is challenging to predict the LoS probability of air-to-ground (A2G) communication scenarios, because the altitude of unmanned aerial vehicles (UAVs) or other aircraft varies from dozens of meters to several kilometers. This paper presents an altitude-dependent empirical LoS probability model for A2G scenarios. Before estimating the model parameters, we design a  $K$ -nearest neighbor (KNN) based strategy to classify LoS and non-LoS (NLoS) paths. Then, a two-layer back propagation neural network (BPNN) based parameter estimation method is developed to build the relationship between every model parameter and the UAV altitude. Simulation results show that the results obtained using our proposed model has good consistency with the ray tracing (RT) data, the measurement data, and the results obtained using the standard models. Our model can also provide wider applicable altitudes than other LoS probability models, and thus can be applied to different altitudes under various A2G scenarios.

**Key words:** Line-of-sight probability model; Air-to-ground channel; Machine learning; Ray tracing

<https://doi.org/10.1631/FITEE.2200041>

**CLC number:** TN928

## 1 Introduction

The unmanned aerial vehicle (UAV) technique is expected to be widely used in sixth-generation (6G) wireless communication due to its compactness and flexibility (Zhang XF et al., 2010; Alladi et al.,

2020; Xiao et al., 2020; You et al., 2021). Different from traditional ground base station (BS) communication, air-to-ground (A2G) communication involves three-dimensional random scatterers. Therefore, UAV altitude information needs to be considered in A2G scenarios (Fan et al., 2016; Zhu et al., 2019, 2021b; Vitucci et al., 2021). However, the line-of-sight (LoS) path, which is the most reliable A2G link, is time-varying and difficult to predict. Therefore, constructing altitude-dependent LoS probability models is essential for design and optimization of A2G communication systems.

LoS probability models are generally divided into two classifications, i.e., geometry-based analytical models and measure-based empirical

<sup>†</sup> Corresponding authors

\* Project supported by the National Key Scientific Instrument and Equipment Development Project, China (No. 61827801) and the Open Research Fund of the State Key Laboratory of Integrated Services Networks, China (No. ISN22-11)

# Electronic supplementary materials: The online version of this article (<https://doi.org/10.1631/FITEE.2200041>) contains supplementary materials, which are available to authorized users

ORCID: Minghui PANG, <https://orcid.org/0000-0001-7798-3154>; Qiuming ZHU, <https://orcid.org/0000-0002-4995-5970>; Zhipeng LIN, <https://orcid.org/0000-0001-5941-2163>

© Zhejiang University Press 2022

models. The geometry-based analytical models predict the LoS probability according to the electromagnetic wave propagation theory and the geometry information. The geometry information is usually obtained from the given digital maps or statistical properties of environments. Digital maps analyze the signal propagation in a deterministic way, i.e., via the ray tracing (RT) (Holis and Pechac, 2008; Zhu et al., 2022) and point cloud (Järveläinen et al., 2016) methods, but these methods require an accurate digital map with detailed material information. Statistical properties describe the built-up environment using statistical characteristics of buildings (ITU-R, 2003; Liu et al., 2018; Al-Hourani, 2020; Cui et al., 2020; Gapeyenko et al., 2021). For example, ITU-R (2003) used three parameters to describe the geometrical statistics of urban areas. However, it is complex and difficult for the analytical methods to reflect the impact of the scenarios accurately.

Measure-based standard empirical models, e.g., the Third Generation Partnership Project (3GPP) TR 38.901 (3GPP, 2016b), 5G Channel Model (5GCM) (3GPP, 2016a), and WINNER II (WINNER, 2008), are usually established based on the measurement data. Since field measurement is complex and costly, many researchers use RT simulation data for channel modeling (Holis and Pechac, 2008; Samimi et al., 2015; Khawaja et al., 2018; Lee et al., 2018; Mao et al., 2020; Zhu et al., 2021b). Specifically, by analyzing the RT simulation data of New York City, Samimi et al. (2015) proposed an RT-based empirical model, where the square index was added to improve the descent speed. Lee et al. (2018) proposed a model that had a low mean square error in the high-rise urban scenario. Holis and Pechac (2008) built an empirical prediction LoS probability model with respect to the elevation angle for scenarios with altitudes of platforms more than 10 km. However, these models cannot be employed in UAV communication scenarios, because they are applicable only for scenarios where altitudes of the platforms are below dozens of meters or over several kilometers.

The aforementioned empirical methods are designed based on traditional parameter estimation methods, e.g., the minimum mean square error (MMSE) and least square (LS) (Lin et al., 2018). These traditional methods usually require specific functional relationships between variables. There-

fore, they are ineffective when the relationship is uncertain. Recently, machine learning (ML) based parameter estimation methods have received a lot of attention. This is because they can accurately determine the internal connection between parameters (Li et al., 2019; Yang WF et al., 2019; Huang C et al., 2020; Huang J et al., 2020; Zheng et al., 2020; Yang M et al., 2021). Moreover, as the first step of parameter estimation, the LoS and non-LoS (NLoS) classifications of channel data were studied in Huang C et al. (2020) and Zheng et al. (2020). Huang C et al. (2020) used the support vector machine (SVM) and artificial neural network (ANN) to perform LoS/NLoS recognition on vehicle-to-vehicle network measurement data. In Zheng et al. (2020), a new LoS/NLoS identification method based on the convolutional neural network (CNN) was proposed, achieving an error rate below 1%. Some researchers focused on using ML to predict the LoS probability directly. For example, Yang WF et al. (2019) used ANN,  $K$ -nearest neighbor (KNN), and gradient boosting decision tree (GBDT) methods to predict the LoS probability directly under indoor scenarios. However, none of these prediction models can be applied to different scenarios without modification.

In this paper, a new altitude-dependent empirical LoS probability model is presented for A2G communication scenarios. The main contributions are summarized as follows:

1. We propose a new altitude-dependent empirical LoS probability model based on massive RT simulation data for A2G communications. Considering the scenario effects, distance effects, and the altitude factor, this model is more suitable for A2G scenarios than other models.

2. We design an ML-based parameter estimation algorithm that introduces the altitude factor to every model parameter. We establish virtual city scenarios based on the statistical characteristics of buildings, and perform numerous RT simulations on the virtual scenarios to obtain the training data for the parameter estimation algorithm.

3. We propose a KNN-based LoS/NLoS identification solution to recognize the LoS path and calculate the LoS probability. We also construct a two-layer back propagation neural network (BPNN) for model parameter estimation. As a result, the adaptability and accuracy of the proposed empirical LoS probability model can be further improved.

## 2 New empirical model for the LoS probability

### 2.1 Background

The LoS probability is defined as the probability that the signal propagating from the transmitter to the receiver along the geometrically shortest route is not blocked by any object available in the propagation environment. The LoS probability is typically modeled as an exponential function against the distance with two undetermined parameters (Holis and Pechac, 2008; 3GPP, 2016b; Zhu et al., 2021a). Specifically, the 3GPP LoS probability model can be expressed as

$$P_{\text{LoS}}^{3\text{GPP}}(d_{\text{TR}}) = \min\left(\frac{D_1}{d_{\text{TR}}}, 1\right) \left[1 - \exp\left(-\frac{d_{\text{TR}}}{D_2}\right)\right] + \exp\left(-\frac{d_{\text{TR}}}{D_2}\right), \quad (1)$$

where  $d_{\text{TR}}$  is the distance between the transmitter and the receiver,  $D_1 = 18$  m is the breakpoint distance where the LoS probability is no longer equal to 1, and  $D_2 = 36$  m is a decay parameter that controls the decreasing rate of the LoS probability with distance. Based on Eq. (1), Samimi et al. (2015) proposed the New York University (NYU) LoS probability model and improved the accuracy of the model by squaring the above formula as a whole. In particular, this model was constructed using a higher resolution intersection test, compared with the 3GPP model based on the real database of New York City. The NYU LoS probability model can be expressed as Eq. (2) (see the bottom of this page), where  $D_1$  and  $D_2$  are 27 m and 71 m, respectively.

### 2.2 New multi-height LoS probability model

In this subsection, we propose an altitude-dependent empirical LoS probability model with three parameters. Current LoS probability models fit only the measurement data well when the altitude of the platform is about 10 m. Thus, the scope of application of these models is limited. The ex-

pression of our model is given by Eq. (3) (see the bottom of this page), where  $h_{\text{TR}}$  is the height from the transmitter to the receiver, and  $D_3$  is a new auxiliary parameter that balances the influence of  $D_1$  and  $D_2$ . Considering the flexible model architecture and extensive data, ML is employed in this study for LoS probability prediction because it shows excellent ability to improve the performance on parameter estimation (Li et al., 2019; Yang WF et al., 2019; Huang C et al., 2020; Huang J et al., 2020; Zheng et al., 2020; Yang M et al., 2021). In our method, we introduce the altitude factor to the parameters of the LoS probability model.

The process of our proposed ML-based parameter estimation is shown in Fig. 1. Since field measurements for A2G channels are difficult and costly, the RT simulation data is adopted as the training data. We first carry out scenario reconstruction and RT simulation for four typical urban scenarios to obtain the required data for parameter estimation. Then, we classify the simulated or measured input data as LoS and NLoS data. Note that in this study, we develop a modified KNN algorithm to perform LoS/NLoS classification. The ratio of the LoS path number to the total path number with respect to the  $n^{\text{th}}$  height and  $m^{\text{th}}$  distance is  $P_{\text{LoS}}^{m,n}$ . The LoS probability at a specific height with different distances can be denoted as  $(P_{\text{LoS}}^m; d_{\text{TR}}^m)$ ,  $m = 1, 2, \dots, 40$ . Moreover, the LS method is adopted to fit  $D_1$ ,  $D_2$ , and  $D_3$  at  $n$  different heights ( $h_{\text{TR}}^n$ ) ( $n = 1, 2, \dots, 50$ ) as the BPNN training data sets  $(D_1^n; h_{\text{TR}}^n)$ ,  $(D_2^n; h_{\text{TR}}^n)$ , and  $(D_3^n; h_{\text{TR}}^n)$ . Finally, a part of the data sets is used to perform BPNN training and obtain the parameters of  $D_1(h_{\text{TR}})$ ,  $D_2(h_{\text{TR}})$ , and  $D_3(h_{\text{TR}})$ , which are altitude dependent.

## 3 ML-based estimation for model parameters

### 3.1 RT-based channel data

The RT technique has been widely used in channel modeling and verification to deal with the

$$P_{\text{LoS}}^{\text{NYU}}(d_{\text{TR}}) = \left\{ \min\left(\frac{D_1}{d_{\text{TR}}}, 1\right) \left[1 - \exp\left(-\frac{d_{\text{TR}}}{D_2}\right)\right] + \exp\left(-\frac{d_{\text{TR}}}{D_2}\right) \right\}^2. \quad (2)$$

$$P_{\text{LoS}}(d_{\text{TR}}, h_{\text{TR}}) = \left\{ \min\left(\frac{D_1(h_{\text{TR}})}{d_{\text{TR}}}, 1\right) \left[1 - \exp\left(-\frac{d_{\text{TR}}}{D_2(h_{\text{TR}})}\right)\right] + \exp\left(-\frac{d_{\text{TR}}}{D_2(h_{\text{TR}})}\right) \right\}^{D_3(h_{\text{TR}})}. \quad (3)$$

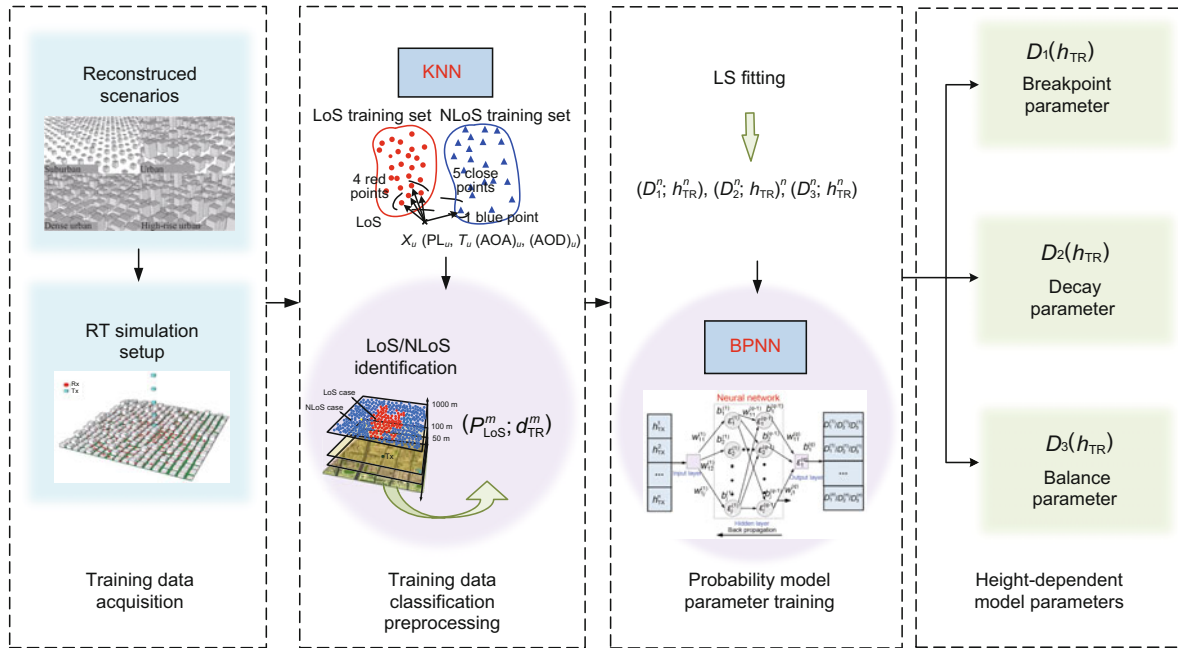


Fig. 1 A flowchart of parameter estimation

inconvenience and high cost of field measurement (Holis and Pechac, 2008; Samimi et al., 2015; Khawaja et al., 2018; Lee et al., 2018; Mao et al., 2020; Zhu et al., 2021b). In RT simulations, the electromagnetic wave radiating from the source is considered as a bunch of rays using a ray-optic approximation, and thus a geometric solution can be obtained based on the uniform theory of diffraction and geometric optics. By tracking all rays with the forward or reverse technique, the propagation parameters can be calculated.

When the RT technique is applied, accurate and detailed geometric and electromagnetic descriptions of the scattering environment are required. In this study, we reconstruct four typical urban scenarios according to the statistical characteristics of buildings (ITU-R, 2003). The locations of buildings and streets follow a uniform distribution and the heights of buildings follow a Rayleigh distribution. The environment-dependent statistical parameters are shown in Table 1, where  $\alpha$  is the percentage

Table 1 Environment-dependent parameters of four scenarios

Scenario	$\alpha$	$\beta$	$\gamma$	$W$	$S$
Suburban	0.1	750	8	11.55	24.97
Urban	0.3	500	15	24.49	20.23
Dense urban	0.5	300	20	40.82	16.91
High-rise urban	0.5	300	50	40.82	16.91

of the land area covered by buildings,  $\beta$  represents the mean number of buildings in the unit area,  $\gamma$  is a random variable denoting the random building height with the probability density function (PDF) as

$$P(h) = \frac{h}{\gamma^2} \exp\left(-\frac{h^2}{2\gamma^2}\right), \quad (4)$$

$W$  is the width of the building, and  $S$  is the width of the street (ITU-R, 2003; Holis and Pechac, 2008; Al-Hourani et al., 2014). Fig. 2 shows the four typical reconstructed scenarios, i.e., suburban, urban, dense urban, and high-rise urban, and the area of each scenario is 4 km<sup>2</sup>.

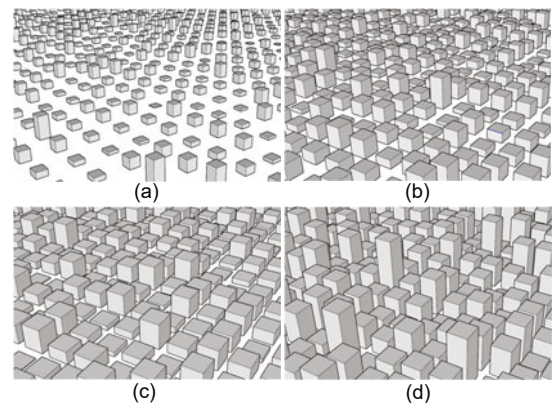


Fig. 2 An illustration of the reconstructed scenarios: (a) suburban; (b) urban; (c) dense urban; (d) high-rise urban

We perform the RT technique on the reconstructed scenarios and obtain the characteristic parameters, including the path loss (PL), delay, angle of arrival (AoA), and angle of departure (AoD). The simulation scenario setting is illustrated in Fig. 3. The blue squares represent the locations of transmitters, whose heights are from 5 to 1005 m with an interval of 20 m. The red points are the positions of receivers distributed in concentric circles with the transmitter as the center. The radii of concentric circles are from 5 to 1000 m with an interval of 25 m, and the number of receivers on each concentric circle is around 200. To avoid contingency, we repeat the simulation five times and calculate the average values of the four characteristic parameters. The total number in the data set is about 10 000, and the simulation parameters are shown in Table 2.

### 3.2 KNN-based LoS/NLoS classification

To calculate the LoS probability, we develop a KNN-based LoS/NLoS classifier to carry out the classification of the massive data obtained from the RT simulation. Based on the raw data set of PL, delay, AoA, and AoD, the classifier can identify the new input data as the LoS or NLoS condition.

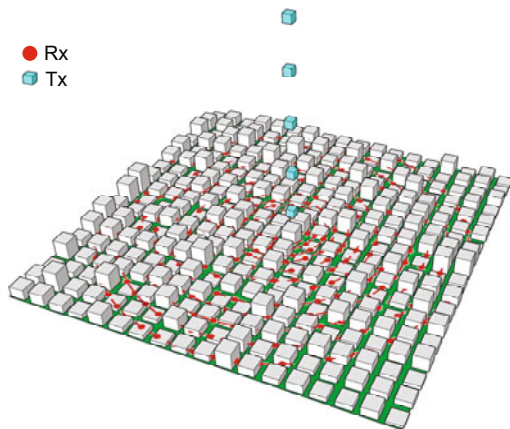


Fig. 3 An illustration of data acquisition setup (References to color refer to the online version of this figure)

Table 2 Simulation parameters

Parameter	Value
Carrier frequency	28 GHz
Tx height (interval)	5–1005 m (20 m)
Rx height	2 m
Rx distance	0–1000 m
Bandwidth	500 MHz
Antenna type	Omnidirectional

In the KNN algorithm, the distance refers to the difference between two samples. Typical distance metrics include the Euclidean distance, Minkowski distance, Manhattan distance, and Chebyshev distance (Zhang Y et al., 2018; Yang GS et al., 2019). The KNN-based classifier in this study is based on the Euclidean distance. As shown in Fig. 4,  $X_u$  denotes the new input data with four characteristic elements, the red dots represent the pre-labeled LoS data set, and the blue dots represent the pre-labeled NLoS data set. The KNN network obtains the LoS or NLoS status of the five points closest to  $X_u$  by calculating the Euclidean distance between  $X_u$  and the pre-labeled data. If the proportion of the LoS status in the five points is larger than that of the NLoS status, the new input data will be judged as LoS, and vice versa.

When we calculate the Euclidean distance, there is a problem that different characteristic elements have significantly different orders of magnitude. For fairness, the samples are normalized before model training. In this study, the linear normalization method is adopted for sample normalization, given by

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)}, \quad (5)$$

where  $x$  is the input value of each feature and  $x'$  is the normalized value. The distance calculation formula can be expressed as Eq. (6) (see the bottom of the next page), where  $PL_u$ ,  $\tau_u$ ,  $(AoA)_u$ , and  $(AoD)_u \in X_u$  are the four characteristic element values of the new input data, and  $PL_v$ ,  $\tau_v$ ,  $(AoA)_v$ , and  $(AoD)_v$  are the four characteristic element values of the  $v^{\text{th}}$  labeled data in the KNN network.

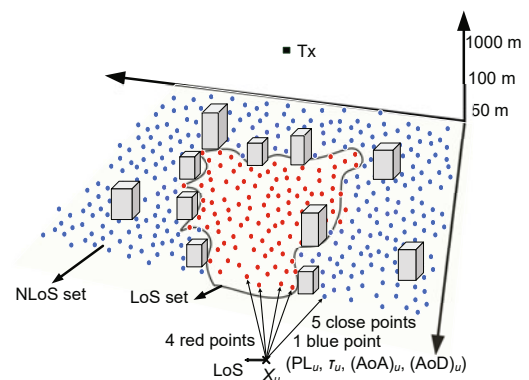


Fig. 4 Diagram of the KNN network (References to color refer to the online version of this figure)

In this study, 740 sets of simulation data at different heights and distances are used to train the KNN network. Note that there exists a data matrix for each height with a certain distance, and there are 200–700 collections of four different characteristic elements in each data matrix. Subsequently, 80% of the characteristic elements in each matrix are randomly selected as the training set, whereas the remaining 20% are selected as the validation set.

To evaluate the network performance on classification, we can obtain the judgement error as

$$A_{\text{error}} = N_i / N_i', \quad (7)$$

where  $N_i$  is the number of data points with incorrect judgments in the validation set, and  $N_i'$  is the total data point number in the validation set.

After using the KNN classifier to identify the new input data, we can obtain the LoS probability  $P_{\text{LoS}}^{m,n}$  corresponding to a specific height and distance, denoted as  $(P_{\text{LoS}}^{m,n}; d_{\text{TR}}^m, h_{\text{TR}}^n)$ .

### 3.3 BPNN-based parameter estimation

BPNN is developed in this study to perform parameter estimation based on the RT data. Note that we use 2000 sets of simulation data at different heights and distances to train the BPNN network. We first keep the altitude variable unchanged and analyze the data set  $(P_{\text{LoS}}^m; d_{\text{TR}}^m), m = 1, 2, \dots, 40$ . At each specific height, the data set  $(P_{\text{LoS}}^m; d_{\text{TR}}^m)$  and Eq. (8) (see the bottom of this page) are used to perform LS fitting, and then we can obtain the data sets  $(D_1^n; h_{\text{TR}}^n), (D_2^n; h_{\text{TR}}^n)$ , and  $(D_3^n; h_{\text{TR}}^n), n = 1, 2, \dots, 50$ . The data set is divided into two parts by a proportion of 8:2 randomly, where 80% of the data is used for training while the rest for validation. The proposed BPNN-based training framework is given in Fig. 5. The neural network, middle part of the BPNN, consists of an input layer, two hidden layers,

and an output layer. The data set  $(h_{\text{TR}}^n)$  is the input layer, and the data sets  $(D_1^n), (D_2^n)$ , and  $(D_3^n)$  make up the output layer. Moreover, the neural network uses two hidden layers, each of which has four neurons, significantly improving the estimation accuracy.

To improve the accuracy of the training process, Eq. (4) is used to normalize the data sets  $(D_1^n; h_{\text{TR}}^n), (D_2^n; h_{\text{TR}}^n)$ , and  $(D_3^n; h_{\text{TR}}^n)$ . After inputting the data sets into BPNN, the altitude-dependent model parameters  $D_1(h_{\text{TR}}), D_2(h_{\text{TR}})$ , and  $D_3(h_{\text{TR}})$  can be expressed as Eq. (9) (see the bottom of this page), where  $j$  is the number of neurons in the first hidden layer,  $i$  is the number of neurons in the second hidden layer,  $k$  is the number of neurons in the output layer,  $w_{ji}^{(q)}$  is the weight between the  $i^{\text{th}}$  neuron in the  $q^{\text{th}}$  layer and the  $j^{\text{th}}$  neuron in the former layer,  $b_p^{(q)}$  is the bias of the  $p^{\text{th}}$  neuron in the  $q^{\text{th}}$  layer, and  $\varepsilon_p^{(q)}$  is the activation function of the  $p^{\text{th}}$  neuron in the  $q^{\text{th}}$  layer. Here,  $p = 1, 2, 3, 4$ , and  $q = 1, 2, 3$ . In this study, the activation sigmoid function is introduced into the hidden layers of BPNN, expressed as

$$\varepsilon_{\text{sig}}(x) = \frac{1}{1 + e^{-x}}. \quad (10)$$

The output loss function is important in BPNN. When the network parameters are adjusted to minimize the value of the loss function, we can obtain the best fitting result. The loss function is expressed as

$$L_F = \sqrt{\frac{1}{N} \sum_{n=1}^N (D_k(h_{\text{TR}}^n) - D_k^n)^2}. \quad (11)$$

The learning rate used by the algorithm is 0.1, the momentum parameter is 0.9, and the number of training times is 40 000. After training and improvement, the neuron weights for four scenarios can be found in the supplementary materials (Tables S1–S4).

$$\rho = \sqrt{(\text{PL}_u - \text{PL}_v)^2 + (\tau_u - \tau_v)^2 + [(\text{AoA})_u - (\text{AoA})_v]^2 + [(\text{AoD})_u - (\text{AoD})_v]^2}. \quad (6)$$

$$P_{\text{LoS}}(d_{\text{TR}}, h_{\text{TR}} = h_{\text{TR}}^n) = \left\{ \min \left( \frac{D_1(h_{\text{TR}})}{d_{\text{TR}}}, 1 \right) \left[ 1 - \exp \left( -\frac{d_{\text{TR}}}{D_2(h_{\text{TR}})} \right) \right] + \exp \left( -\frac{d_{\text{TR}}}{D_2(h_{\text{TR}})} \right) \right\}^{D_3(h_{\text{TR}})}, n = 1, 2, \dots, 50. \quad (8)$$

$$D_k(h_{\text{TR}}; \mathbf{w}, \mathbf{b}, \boldsymbol{\varepsilon}) = \varepsilon_k^{(3)} \left\{ \sum_{i=1}^4 w_{ik}^{(3)} \varepsilon_i^{(2)} \left[ \sum_{j=1}^4 w_{ji}^{(2)} \varepsilon_j^{(1)} \left( w_{lj}^{(1)} h_{\text{TR}} + b_j^{(1)} \right) + b_i^{(2)} \right] + b_k^{(3)} \right\}, k = 1, 2, 3. \quad (9)$$

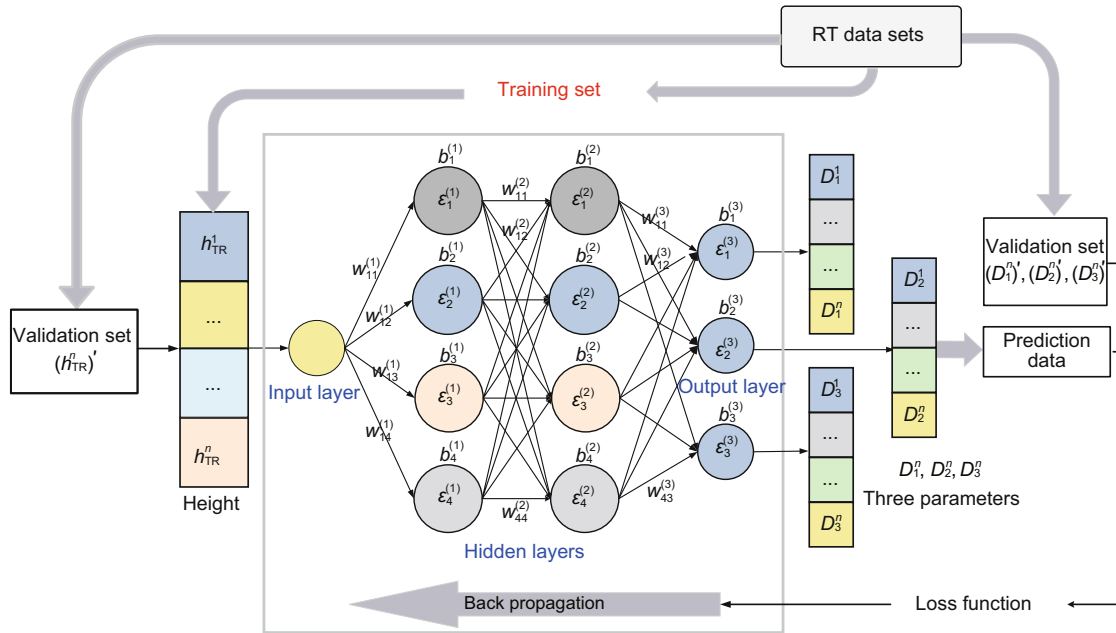


Fig. 5 An illustration of the BPNN-based framework

Most urban scenarios can be approximately represented by one of these four typical scenarios, so the trained parameters in Tables S1–S4 can be used directly for prediction. For the high-precision application, we can calculate  $\psi \in \{\alpha, \beta, \gamma\}$ ,  $W$ , and  $S$  of the new scenario, and retrain the neuron parameters using the BPNN-based estimation method.

## 4 Simulations and validation

### 4.1 Path classification validation

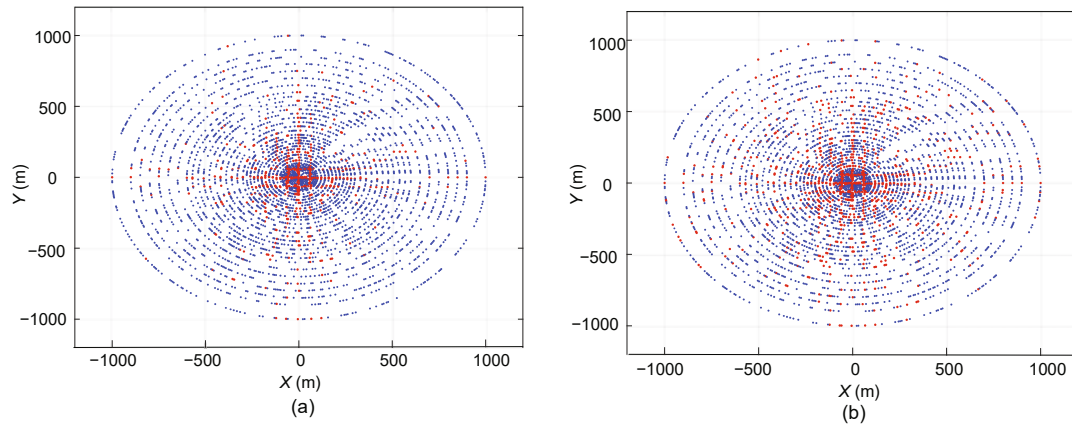
Taking the dense urban scenario as an example, we perform numerical simulations and comparisons to verify the proposed classifier. For the UAV altitudes of 200 m and 400 m, we can obtain the communication statuses between the receiver and the transmitter at different distances. As shown in Fig. 6, the red points are the LoS paths, and the blue points are NLoS paths. We can see that the number of LoS paths decreases as the distance increases, while the number of LoS paths increases when the UAV altitude increases. Note that some square areas close to the center are still NLoS, because they are blocked by buildings in the scenario. Fig. 6 also demonstrates the scientific nature of the KNN-based classification method.

The judgment error calculated using Eq. (7) is

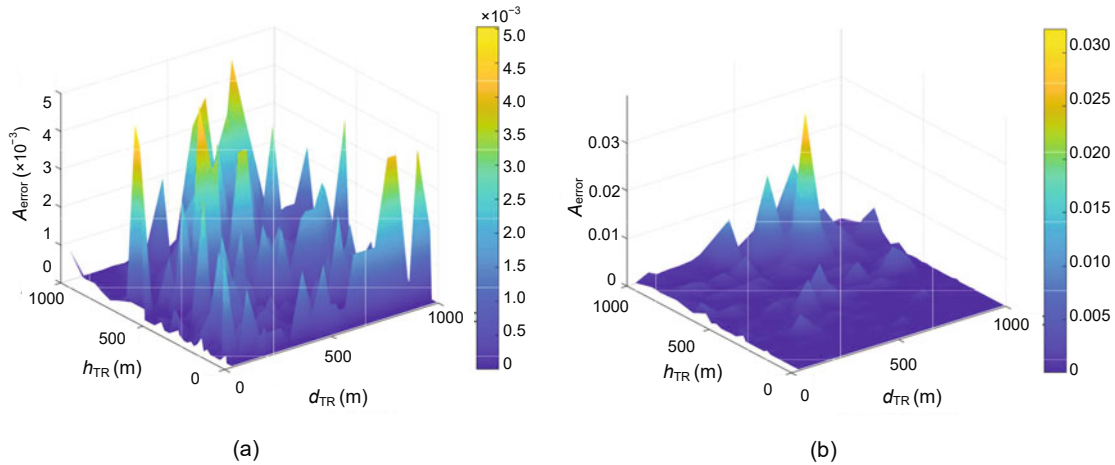
shown in Fig. 7a. As we can see, the judgment error of the KNN-based classifier is below 0.0005. In other words, the accuracy of LoS/NLoS classification is over 0.995, which is practical for further processing, and the accuracy of our classifier is higher than that of the classification method in Huang C et al. (2020). In addition, we use other ML classification methods, such as decision trees (DT), to build classifiers. Specifically, we use the classification and regression tree (CART) method, in which the Gini index is adopted to select the optimal division point of the optimal feature, each feature is divided recursively, and the feature space is divided into finite units. The error of the DT method is shown in Fig. 7b. Comparing the error results in Fig. 7, it can be seen that the performance of the KNN-based classification method is better than that of the DT-based method. This is because the KNN-based classification method can best reveal the data characteristics and balance the weight of each element.

### 4.2 Parameter estimation validation

To evaluate the BPNN-based parameter estimation method, the training results of the training set and validation set on the three parameters are shown in Fig. 8. It can be seen that parameter  $D_1$  increases as the height increases. This means that the distance range continues to increase when the LoS



**Fig. 6 Judgment results under the dense urban scenario: (a)  $h_{TR} = 200$  m; (b)  $h_{TR} = 400$  m (References to color refer to the online version of this figure)**



**Fig. 7 Judgment errors at different heights and distances under the dense urban scenario: (a) KNN; (b) DT**

probability is 1. Since the two parameters  $D_2$  and  $D_3$  jointly control the variation curve trend, they present a state of undulations within a certain range. Note that the results are in line with the theoretical derivation and objective laws. As shown in Fig. 8, the neural network is well trained and the prediction values are in good agreement with the original values in the validation set. It also demonstrates that the prediction framework has good performance and describes the relationship between each parameter and UAV height.

We also use other ML methods for parameter estimation, and the mean square errors (MSEs) of altitude-dependent parameters in the prediction set and validation set are used to evaluate the training performance. Fig. 9 shows the MSEs of DT, BPNN, support vector regression (SVR), and Gaussian process regression (GPR). Among them, DT

and BPNN have good performance on parameter estimation. For the DT method, the MSEs under the high-rise urban and suburban scenarios are larger than those under other scenarios, showing poor robustness in our application. The BPNN method achieves low MSE under each scenario and shows good robustness. For the SVR and GPR algorithms, due to the complex network parameters, it is difficult to achieve a good fitting state in a short period of time. When the DT method is used for regression, only one factor can be considered at each node, and the contribution of each element cannot be balanced according to the weight of multiple elements. Considering performance and robustness, we choose the BPNN method to estimate the model parameters in this study. Note that it is convenient for BPNN to further modify the network parameters with new data sets from measurement and simulation.



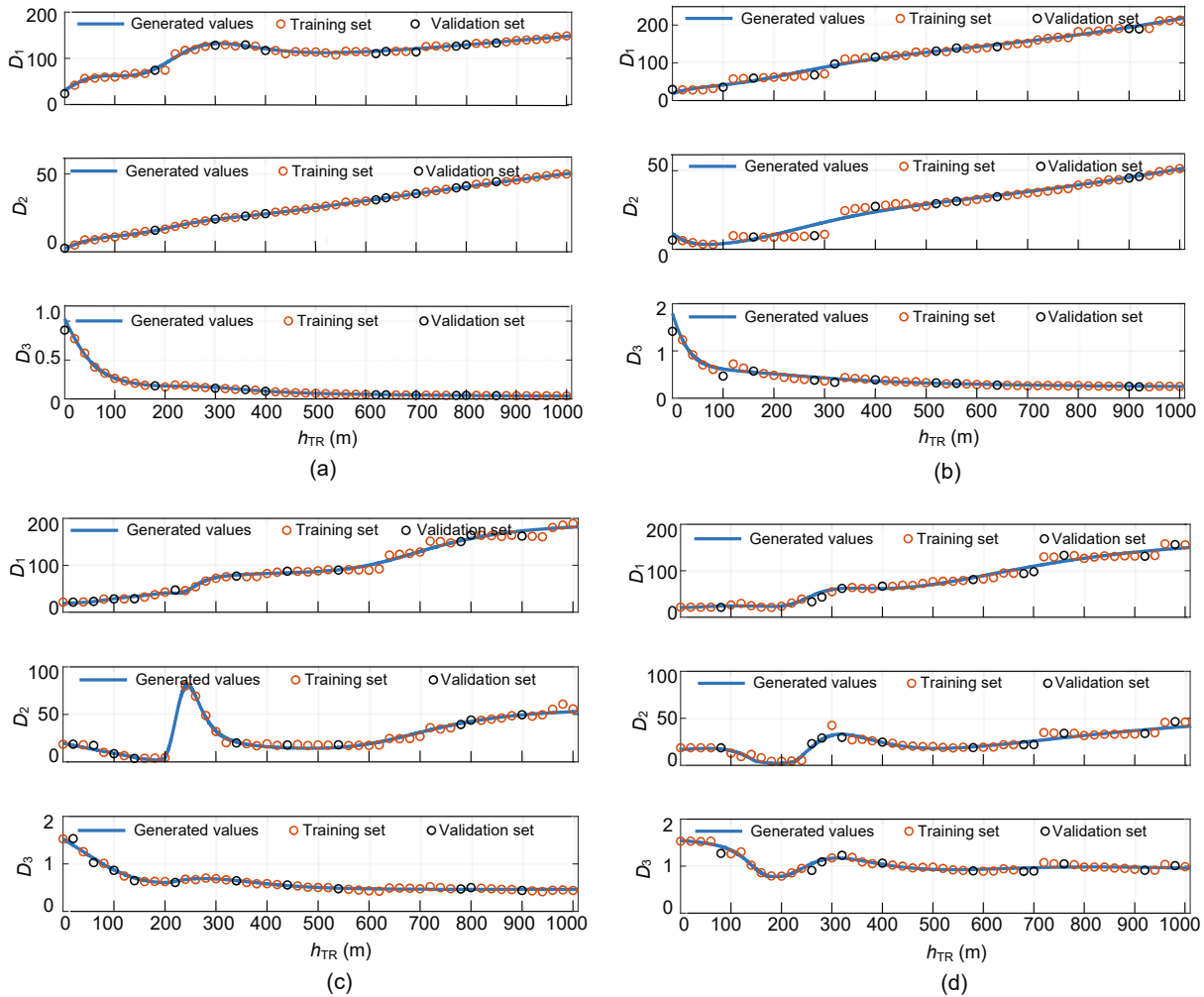


Fig. 8 Training results for height-dependent parameters under different scenarios: (a) suburban; (b) urban; (c) dense urban; (d) high-rise urban

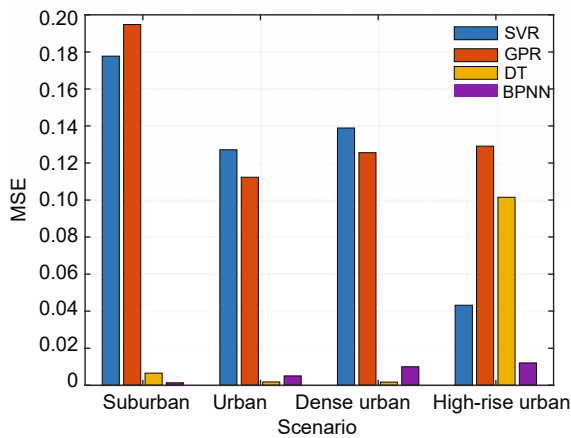
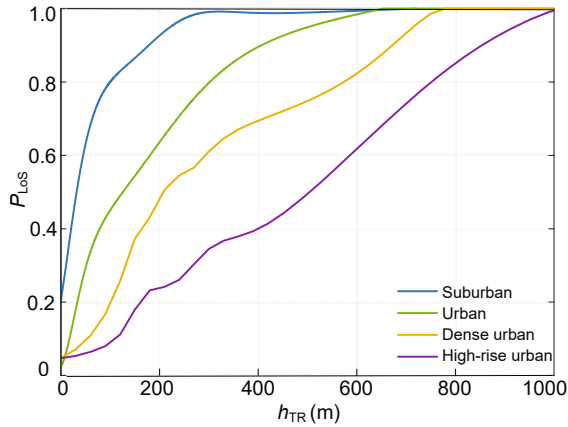


Fig. 9 Comparison of other ML methods for parameter estimation

### 4.3 Prediction results and analysis

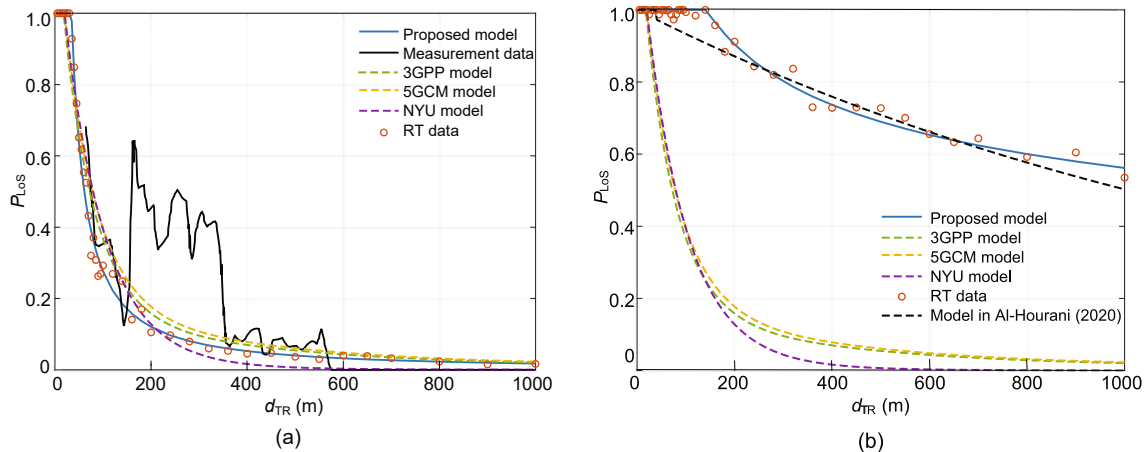
Different scenarios and UAV height factors have a great influence on the LoS probability. Fig. 10 shows the LoS probability when the UAV altitudes range from 0 to 1000 m and the communication distance is 150 m. It can be seen that the LoS probability varies greatly at a certain height under different scenarios. The LoS probability decreases as the density of urban buildings increases, which conforms to the objective laws.

To demonstrate the accuracy of the proposed LoS probability model, we take the urban scenario as an example, and compare the results obtained using our proposed ML-based empirical model (3) with those obtained using representative models, the RT simulation data, and the measurement data. For the



**Fig. 10** LoS probability vs. height in different scenarios (References to color refer to the online version of this figure)

low-altitude cases (Fig. 11a), we set the communication height as 40 m. We can see that the performances of the proposed model and the RT simulation data are very similar to those of other representative standard models (Samimi et al., 2015; 3GPP, 2016a, 2016b). The prediction results of our model also agree with the measurement data in Sun et al. (2015). Moreover, when the communication altitude is high, the standard models are no longer suitable for LoS probability prediction. As shown in Fig. 11b, the standard models deviate and cannot describe the real situation at the altitude of 600 m. Our proposed model is in agreement with the analytical model in Al-Hourani (2020). Moreover, the RT data agrees with the prediction results. Our model can achieve good prediction at both low and high altitudes.



**Fig. 11** Comparisons of different models at different altitudes: (a) low altitude ( $h_{TX} = 40$  m); (b) high altitude ( $h_{TX} = 600$  m)

## 5 Conclusions

In this paper, we have proposed the altitude-dependent empirical LoS probability model based on ML for A2G scenarios. First, we have applied the KNN to classify the LoS and NLoS paths according to the RT simulation data with the recognition accuracy rate as high as 0.995. Then, to introduce the factor of height to the LoS probability, we have developed a two-layer BPNN to estimate the parameters of the proposed LoS probability model, which has better performance than other regression algorithms. Simulation results have demonstrated that the prediction results of our proposed LoS probability model can achieve good versatility at both low and high altitudes and have good agreement with RT simulation and measurement data. In future work, we will improve the accuracy of LoS probability prediction by more measurement data and expand the model to various scenarios.

## Contributors

Minghui PANG and Qiuming ZHU designed the research. Zhipeng LIN and Fei BAI processed the data. Minghui PANG drafted the paper. Qiuming ZHU and Zhipeng LIN helped organize the paper. Yue TIAN helped train the data. Zhuo LI and Xiaomin CHEN revised and finalized the paper.

## Compliance with ethics guidelines

Minghui PANG, Qiuming ZHU, Zhipeng LIN, Fei BAI, Yue TIAN, Zhuo LI, and Xiaomin CHEN declare that they have no conflict of interest.

## References

- 3GPP, 2016a. 5G Channel Model for Bands up to 100 GHz. 3<sup>rd</sup> Generation Partnership Project (3GPP).
- 3GPP, 2016b. Technical Specification Group Radio Access Network; Channel Model for Frequency Spectrum above 6 GHz (Release 14). TR 38.900 V14.2.0. 3<sup>rd</sup> Generation Partnership Project (3GPP).
- Al-Hourani A, 2020. On the probability of line-of-sight in urban environments. *IEEE Wirel Commun Lett*, 9(8):1178-1181. <https://doi.org/10.1109/LWC.2020.2984497>
- Al-Hourani A, Kandeepan S, Jamalipour A, 2014. Modeling air-to-ground path loss for low altitude platforms in urban environments. Proc IEEE Global Communications Conf, p.2898-2904 <https://doi.org/10.1109/GLOCOM.2014.7037248>
- Alladi T, Naren, Bansal G, et al., 2020. SecAuthUAV: a novel authentication scheme for UAV-ground station and UAV-UAV communication. *IEEE Trans Veh Technol*, 69(12):15068-15077. <https://doi.org/10.1109/TVT.2020.3033060>
- Cui ZZ, Guan K, Briso-Rodríguez C, et al., 2020. Frequency-dependent line-of-sight probability modeling in built-up environments. *IEEE Int Things J*, 7(1):699-709. <https://doi.org/10.1109/JIOT.2019.2947782>
- Fan W, Carton I, Nielsen J, et al., 2016. Measured wideband characteristics of indoor channels at centimetric and millimetric bands. *EURASIP J Wirel Commun Netw*, 2016:58. <https://doi.org/10.1186/s13638-016-0548-x>
- Gapeyenko M, Moltchanov D, Andreev S, et al., 2021. Line-of-sight probability for mmwave-based UAV communications in 3D urban grid deployments. *IEEE Trans Wirel Commun*, 20(10):6566-6579. <https://doi.org/10.1109/TWC.2021.3075099>
- Holis J, Pechac P, 2008. Elevation dependent shadowing model for mobile communications via high altitude platforms in built-up areas. *IEEE Trans Antenn Propag*, 56(4):1078-1084. <https://doi.org/10.1109/TAP.2008.919209>
- Huang C, Molisch AF, He RS, et al., 2020. Machine learning-enabled LOS/NLOS identification for MIMO systems in dynamic environments. *IEEE Trans Wirel Commun*, 19(6):3643-3657. <https://doi.org/10.1109/TWC.2020.2967726>
- Huang J, Wang CX, Bai L, et al., 2020. A big data enabled channel model for 5G wireless communication systems. *IEEE Trans Big Data*, 6(2):211-222. <https://doi.org/10.1109/TBDATA.2018.2884489>
- ITU-R, 2003. P.1410-2: Propagation Data and Prediction Methods Required for the Design of Terrestrial Broadband Millimetric Radio Access Systems Operating in a Frequency Range of About 20–50 GHz.
- Järveläinen J, Nguyen SLH, Haneda K, et al., 2016. Evaluation of millimeter-wave line-of-sight probability with point cloud data. *IEEE Wirel Commun Lett*, 5(3):228-231. <https://doi.org/10.1109/LWC.2016.2521656>
- Khawaja W, Ozdemir O, Guvenc I, 2018. Temporal and spatial characteristics of mm wave propagation channels for UAVs. Proc 11<sup>th</sup> Global Symp on Millimeter Waves, p.1-6. <https://doi.org/10.1109/GSMM.2018.8439182>
- Lee JH, Choi JS, Kim SC, 2018. Cell coverage analysis of 28 GHz millimeter wave in urban microcell environment using 3-D ray tracing. *IEEE Trans Antenn Propag*, 66(3):1479-1487. <https://doi.org/10.1109/TAP.2018.2797531>
- Li W, Zhang JH, Ma XC, et al., 2019. The way to apply machine learning to IoT driven wireless network from channel perspective. *China Commun*, 16(1):148-164. <https://doi.org/10.12676/j.cc.2019.01.014>
- Lin ZP, Lv TJ, Mathiopoulos PT, 2018. 3-D indoor positioning for millimeter-wave massive MIMO systems. *IEEE Trans Commun*, 66(6):2472-2486. <https://doi.org/10.1109/TCOMM.2018.2797993>
- Liu X, Xu J, Tang HY, 2018. Analysis of frequency-dependent line-of-sight probability in 3-D environment. *IEEE Commun Lett*, 22(8):1732-1735. <https://doi.org/10.1109/LCOMM.2018.2842763>
- Mao K, Zhu QM, Song MZ, et al., 2020. A geometry-based beamforming channel model for UAV mmWave communications. *Sensors*, 20(23):6957. <https://doi.org/10.3390/s20236957>
- Samimi MK, Rappaport TS, MacCartney GR, 2015. Probabilistic omnidirectional path loss models for millimeter-wave outdoor communications. *IEEE Wirel Commun Lett*, 4(4):357-360. <https://doi.org/10.1109/LWC.2015.2417559>
- Sun S, Thomas TA, Rappaport TS, et al., 2015. Path loss, shadow fading, and line-of-sight probability models for 5G urban macro-cellular scenarios. Proc IEEE Globecom Workshops, p.1-7. <https://doi.org/10.1109/GLOCOMW.2015.7414036>
- Vitucci EM, Semkin V, Arpaio MJ, et al., 2021. Experimental characterization of air-to-ground propagation at mm-wave frequencies in dense urban environment. Proc 15<sup>th</sup> European Conf on Antennas and Propagation, p.1-5. <https://doi.org/10.23919/EuCAP51087.2021.9411249>
- WINNER, 2008. WINNER II Channel Models. IST-4-027756, WINNER II D1.1.2 V1.2.
- Xiao ZY, Dong H, Bai L, et al., 2020. Unmanned aerial vehicle base station (UAV-BS) deployment with millimeter-wave beamforming. *IEEE Int Things J*, 7(2):1336-1349. <https://doi.org/10.1109/JIOT.2019.2954620>
- Yang GS, Zhang Y, He ZW, et al., 2019. Machine-learning-based prediction methods for path loss and delay spread in air-to-ground millimetre-wave channels. *IET Microw Antenn Propag*, 13(8):1113-1211. <https://doi.org/10.1049/iet-map.2018.6187>
- Yang M, Ai B, He RS, et al., 2021. Machine-learning-based fast angle-of-arrival recognition for vehicular communications. *IEEE Trans Veh Technol*, 70(2):1592-1605. <https://doi.org/10.1109/TVT.2021.3054757>
- Yang WF, Zhang JL, Zhang J, 2019. Machine learning based in-door line-of-sight probability prediction. Int Symp on Antennas and Propagation, p.1-3.
- You XH, Wang CX, Huang J, et al., 2021. Towards 6G wireless communication networks: vision, enabling technologies, and new paradigm shifts. *Sci China Inform Sci*, 64(1):110301. <https://doi.org/10.1007/s11432-020-2955-6>
- Zhang XF, Xu LY, Xu L, et al., 2010. Direction of departure (DOD) and direction of arrival (DOA) estimation in

- MIMO radar with reduced-dimension MUSIC. *IEEE Commun Lett*, 14(12):1161-1163. <https://doi.org/10.1109/LCOMM.2010.102610.101581>
- Zhang Y, Wen JX, Yang GS, et al., 2018. Air-to-air path loss prediction based on machine learning methods in urban environments. *Wirel Commun Mob Comput*, Article 8489326. <https://doi.org/10.1155/2018/8489326>
- Zheng QB, He RS, Ai B, et al., 2020. Channel non-line-of-sight identification based on convolutional neural networks. *IEEE Wirel Commun Lett*, 9(9):1500-1504. <https://doi.org/10.1109/LWC.2020.2994945>
- Zhu QM, Wang YW, Jiang KL, et al., 2019. 3D non-stationary geometry-based multi-input multi-output channel model for UAV-ground communication systems. *IET Microw Antenn Propag*, 13(8):1104-1112. <https://doi.org/10.1049/iet-map.2018.6129>
- Zhu QM, Wang CX, Hua BY, et al., 2021a. 3GPP TR 38.901 Channel Model. In: Wiley 5G REF: the Essential 5G Reference Online. Wiley Press. <https://doi.org/10.1002/9781119471509.w5GRef048>
- Zhu QM, Yao MT, Bai F, et al., 2021b. A general altitude-dependent path loss model for UAV-to-ground millimeter-wave communications. *Front Inform Technol Electron Eng*, 22(6):767-776. <https://doi.org/10.1631/FITEE.2000497>
- Zhu QM, Mao K, Song MZ, et al., 2022. Map-based channel modeling and generation for U2V mmWave communication. *IEEE Trans Veh Technol*, 71(8):8004-8015. <https://doi.org/10.1109/TVT.2022.3174404>

## List of supplementary materials

- Table S1 Trained parameters of the suburban scenario
- Table S2 Trained parameters of the urban scenario
- Table S3 Trained parameters of the dense urban scenario
- Table S4 Trained parameters of the high-rise urban scenario