# Predictive Modeling of Millimeter-Wave Vegetation Scattering Effect Using Hybrid Physics-Based and Data-Driven Approach

Peize Zhang, *Student Member, IEEE*, Cheng Yi, *Student Member, IEEE*, Bensheng Yang, *Student Member, IEEE*, Haiming Wang, *Member, IEEE*, Claude Oestges, *Fellow, IEEE*, and Xiaohu You, *Fellow, IEEE*

***Abstract*—System design and deployment of millimeter-wave (mmWave) wireless communication systems for outdoor coverage need sophisticated channel models to describe the wave interaction with illuminated physical objects. However, the conventional physical-statistical model cannot accurately predict the site-specific mmWave channel characteristics when involving complex system configurations and geometric information of transceiver relative locations. A framework of machine learning assisted channel modeling approach is proposed, in which the statistical models are leveraged for inter-cluster level channel characterization and the propagation properties within each kind of clusters are predicted using a hybrid physics-based and data-driven approach. In particular, with a focus on the mmWave through-vegetation scattering effect, a set of dedicated directional channel measurements and ray-tracing simulations is performed in an identical vegetated street canyon environment at 28 GHz for the performance evaluation of the proposed approach. Moreover, the training results and model validation in different environments show that comparing with the physical-statistical model, the proposed hybrid model, which adds the environment features to the artificial neural network as inputs, has higher prediction accuracy and greater generalization ability in terms of the site-specific through-vegetation cluster parameters, such as vegetation attenuation, delay spread, and angular spread.**

***Index Terms*—Millimeter-wave propagation, channel prediction, vegetation scattering, physical-statistical model, artificial neural network**

Manuscript received April 10, 2021; revised July 3, 2021; accepted July 27 2021. Date of publication ; date of current . This work was supported in part by the National Key R&D Program of China under Grant 2020YFB1804901, the National Natural Science Foundation of China under Grants 61960206006 and 62001101, the Science Foundation of Jiangsu Province of China under Grant BK20200349. *(Corresponding author: Haiming Wang.)*

P. Z. Zhang is with the State Key Laboratory of Millimeter Waves, Southeast University, Nanjing 210096, China and also with the Institute of Information and Communication Technologies, Electronics and Applied Mathematics, Université Catholique de Louvain, 1348 Louvain-la-Neuve, Belgium (e-mail: pzzhang@seu.edu.cn).

C. Yi, B. S. Yang, and H. M. Wang are with the State Key Laboratory of Millimeter Waves, Southeast University, Nanjing 210096, China and also with Purple Mountain Laboratories, Nanjing 211111, China (e-mail: chengyi@seu.edu.cn, yangbensheng@seu.edu.cn, hmwang@seu.edu.cn).

C. Oestges is with the Institute of Information and Communication Technologies, Electronics and Applied Mathematics, Université Catholique de Louvain, 1348 Louvain-la-Neuve, Belgium (e-mail: claude.oestges@uclouvain.be).

X. H. You is with the National Mobile Communications Research Laboratory, Southeast University, Nanjing 210096, China and also with Purple Mountain Laboratories, Nanjing 211111, China (e-mail: xhyu@seu.edu.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier

## I. Introduction

THE global rollout of the fifth-generation (5G) mobile communication system is fully underway, while the commercial deployment of the millimeter-wave (mmWave) technology is still in its early stages. This is partly because mmWave signals suffer from severe path loss and high sensitivity to physical objects with their sizes comparable to the wavelengths [1], [2]. Accordingly, the coverage of mmWave networks strongly depends on the channel characteristics, especially in obstacle-line-of-sight (OLoS) and non-line-of-sight (NLoS) scenarios. Existing channel measurement and simulation results reveal that the clustering nature of mmWave propagation channels is significant for the multipath combination of direct, reflected, and diffracted paths [3]. Mapping the virtual clusters with surrounding scatterers, e.g., building corners, vehicles, human body, and trees [4], [5], indicates that propagating through these physical obstacles has an impact on both large-scale and small-scale channel characteristics. To progress cost-efficient real-world deployment of mmWave cellular networks, an investigation into the scattering effect of physical objects on channel characterization is in turn practically necessary.

With a focus on the vegetation scattering effect in the mmWave band, existing cellular-type channel measurements in suburban environments [5]–[8] and site-specific foliage propagation measurements [9], [10] mainly studied the directional and omnidirectional signal attenuations due to the presence of trees or bushes in the forward path. Meanwhile, the sounding data were fitted with the standard vegetation attenuation models recommended in ITU-R P.833-9 [11], as well as their modified models (e.g., the Weissberger model [12] and the fitted ITU-R model [13]). However, the above-mentioned models are generally characterized as functions of vegetation depth and frequency, which ignore the impact of system configurations on vegetation attenuation for mmWave terrestrial communications. Given the decrease of the cell size (e.g., on the order of 200 m for 28 GHz NLoS link), mmWave signals are featured with wider spatial distribution in the vertical dimension especially when base stations (BSs) are placed in elevated positions. Thus, if the BS height reduces or antenna downtilt increases, the likelihood of the increase of excess attenuation becomes much larger regarding the through-vegetation paths. Moreover, despite the empirical-based vegetated channel model contains both delay informa-

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAP.2021.3118463, IEEE Transactions on Antennas and Propagation



tion and angular information about the multipath components (MPCs), much remains to be analyzed referring to the forward vegetation scattering effect on the statistics of space-time channel parameters.

The advantage of the physical-statistical model is that the consistency of the cluster-based channel representation and physical propagation paths can be experimentally verified; and in turn, the statistical behavior of inter- and intra-cluster parameters in a specific environment is studied for channel prediction. However, this method generally fails in precise channel prediction with the knowledge of the geometric information of transmitter (TX) and receiver (RX) antennas' locations, as well as their relative orientations. Additionally, in the absence of standard criteria for multipath clustering, it is challenging to accurately estimate the statistics of these cluster-level channel parameters [14], [15]. Such uncertain intra-cluster parameters leveraged for channel prediction would also lead to an imprecise characterization of different clusters corresponding to different physical scatterers and propagation mechanisms. For example, the forward through-vegetation cluster tends to have a larger angular spread [5], whereas it follows the same distribution (the mean and standard deviation) as other direct and reflected clusters.

To address the aforementioned problems, machine learning (ML)-based approaches were proposed to complement the physical-statistical model [16]–[21]. In particular, artificial neural networks (ANNs) were commonly introduced to find the best functional fit of path loss and transceiver settings which include geometric information, antenna patterns, and environmental features [22]–[24]. Only a handful of reported works simultaneously fulfilled the prediction of path loss and spatial-temporal characteristics, along with more complicated ML algorithms, such as random forest [18], convolutional neural network [16], and feed-forward and radial basis function neural networks [20]. Owning to the change in the environment geography, most of these works were still unable to capture the variations in channel characteristics based on limited channel datasets collected in specific environments. Therefore, the data-driven approach needs to combine the advantages of the physics-based channel model, which is expected to provide scalable and robust channel predicting results.

The aims of this paper are twofold. The first is to develop a framework of the ML-assisted predictive channel modeling approach. The inter-cluster parameters, such as cluster number and type, Ricean K-factor, and inter-cluster delay and angular spreads, are identified via the physical-statistical channel model; and for each cluster, the intra-cluster characteristics are predicted through an ANN-based predictor with less number of neurons in the hidden layer. It is expected that the same kinds of physical objects in the propagation channel exhibit similar propagation characteristics. These ANNs with relatively simple architectures are thus adequate for the approximation of different kinds of multipath clusters. The second is to provide valuable insights into the prediction of mmWave vegetation scattering effect on both signal attenuation and dispersion. The statistical behavior of through-vegetation scattering clusters is studied versus TX antenna orientations and RX locations, including vegetation attenuation, delay spread, and angular spread. Consequently, the hybrid physics-based and data-driven modeling approach is employed, combining with the essential environmental features, to predict the mmWave propagation characteristics of through-vegetation scattering clusters. The collections of channel data are produced from directional measurement and ray-tracing simulation at 28 GHz in the identical vegetated street canyon environment.

The remainder of this paper is organized as follows. Section II describes the framework of our proposed ML-assisted channel modeling approach. Section III presents the details of mmWave channel measurement and simulation setups. Section IV discusses the hybrid physics-based and data-driven modeling results for the prediction of the mmWave vegetation scattering effect. Finally, Section V concludes the paper.

## II. Framework of the ML-Assisted Channel Modeling Approach

The cluster-based statistical channel models reveal the mapping results between virtual clusters and physical scatterers, which group the MPCs in both time and angular domains. Consequently, the propagation channels are characterized by several inter- and intra-cluster parameters following specific distributions, while the existing models disregard the differences in the statistics of intra-cluster parameters among various cluster types. Moreover, the universality and accuracy of the physical-statistical channel models are usually established based on sufficient channel data collected in multiple environments, coupled with the high-resolution channel parameter estimation algorithm. Early studies on ML-based modeling of both large-scale and small-scale channel characteristics [16], [20] show great potential for the improvement of channel prediction accuracy when using the geometric information of TX and RX locations as input data. However, the greatest drawback of these models lies in that they can only be used in the cases having been trained for specific environments. To remedy this situation, as shown in Fig. 1, an ML-assisted

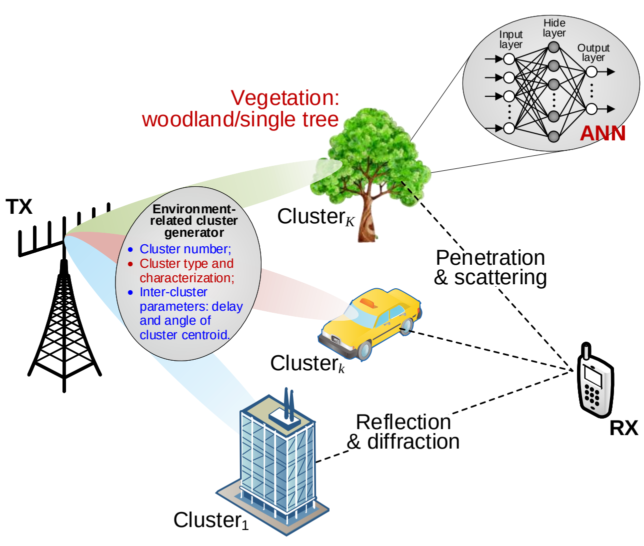



Fig. 1. The framework of the ML-assisted channel modeling approach.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAP.2021.3118463, IEEE Transactions on Antennas and Propagation



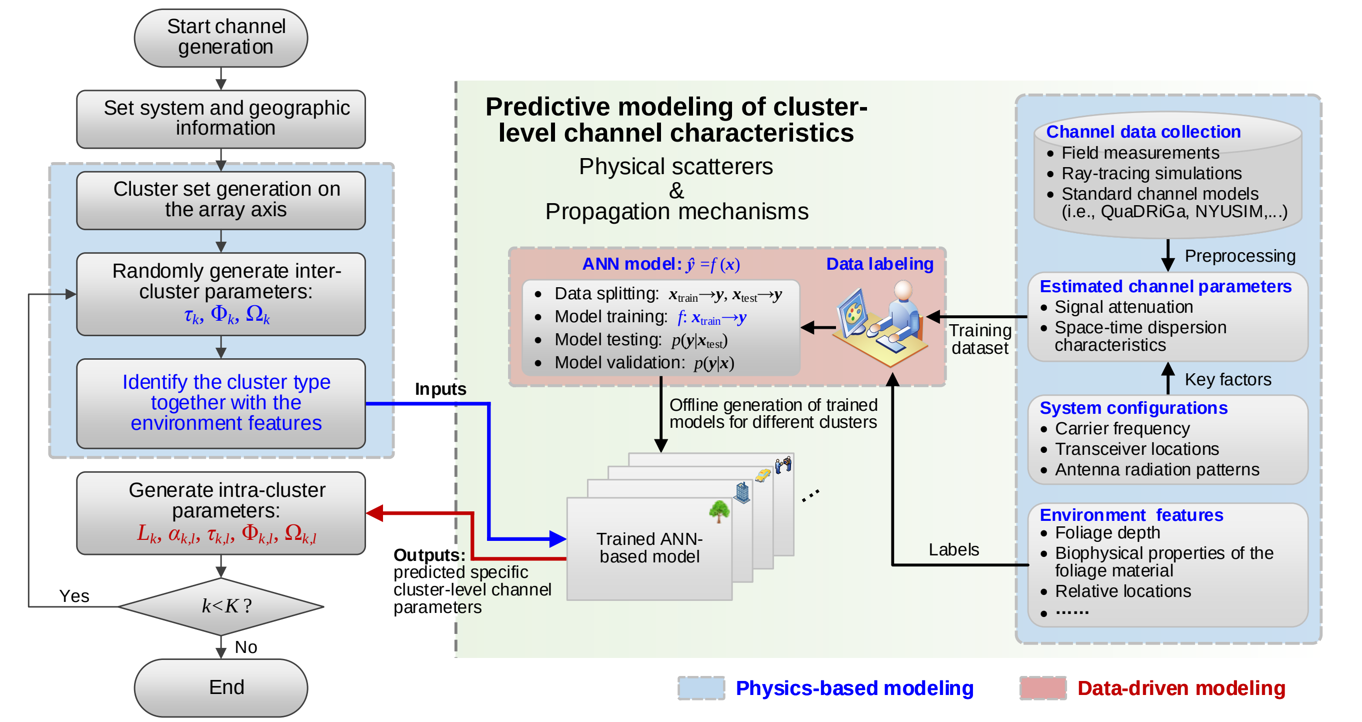



Fig. 2. The flowchart of the ML-assisted channel modeling, as well as a case study using the hybrid physics-based and data-driven approach to predict the mmWave channel characteristics of the through-vegetation scattering cluster.

channel modeling approach is proposed, which integrates the advantages of both physical models and ML-based models. It is expected to improve the generalization properties by leveraging the prior knowledge of the physical structure of concerned environments.

Multipath channels are normally modeled by a sum of $K$ clusters and $\sum_{k=1}^{K} L_k$ subpaths, written as [2]:

$$h(t,\Phi,\Omega) = \sum_{k=1}^{K}\sum_{l=1}^{L_k}\alpha_{k,l}e^{j\varphi_{k,l}}\delta(t-\tau_k-\tau_{k,l}) \cdot \delta(\Phi-\Phi_k-\Phi_{k,l})\,\delta(\Omega-\Omega_k-\Omega_{k,l}) \quad (1)$$

where $\Phi$ is the angle of departure (AoD) and $\Omega$ is the angle of arrival (AOA); $L_k$ is the number of subpath within the $k$th cluster; $\alpha_{k,l}$ and $\varphi_{k,l}$ are the amplitude and random phase of the $l$th subpath within the $k$th cluster; $\tau_k$, $\Phi_k$, and $\Omega_k$ are the delay, AoD, and AoA of the cluster centroid, respectively; and $\tau_{k,l}$, $\Phi_{k,l}$, and $\Omega_{k,l}$ are the delay, AoD, and AoA of the $l$th subpath relative to the cluster centroid, respectively. In such a model, the cluster parameters consist not only of the inter-cluster parameters to identify clusters' location, but also of the intra-cluster parameters to characterize the relationship between MPCs and their corresponding clusters. As depicted in Fig. 2, we first need to set system configurations (e.g., carrier frequency and antenna field patterns) and geographic information (e.g., scenarios, LoS/OLoS/NLoS, and transceiver locations). Following the standard procedure recommended in [25], the inter-cluster parameters (e.g., $K$, $\tau_k$, $\Phi_k$, and $\Omega_k$) are generated based on the statistical channel model, and in turn, clusters are placed in parameter space. In particular, the cluster type is assigned to each cluster together with its physical characteristics. This step enables the differences of spatio-temporal dispersion parameters between clusters to be distinguished. Due to the fact that each cluster could be mapped to the objects in physical space, it is reasonable to separately characterize the cluster-level channel interacting with the corresponding scatterers, such as building reflection, human body diffraction, and vegetation scattering. By leveraging a set of ML-based predictors, the intra-cluster space-time channel parameters (e.g., $\alpha_{k,l}$, $\varphi_{k,l}$, $\tau_{k,l}$, $\Phi_{k,l}$, and $\Omega_{k,l}$) can be sequentially generated for different cluster types. Such ML-assisted modeling approach takes the advantages of cluster-based representation of physical channels and forecasts the propagation channels with low level of uncertainty to the variation in environment geography in comparison with conventional ML-based approaches [16], [17], [26].

To provide scalable and robust cluster-level channel predicting results, the hybrid physics-based and data-driven modeling approach is promoted. Based on the channel data collected from field measurements, ray-tracing simulation, or standard channel simulators, several channel parameters can be estimated together with the key impact factors of system configurations, such as frequency, geometric information, and antenna downtilt. Consequently, these channel data are labeled with environment features and used for training the data-driven model. For instance, the ANN model is employed to predict vegetation attenuation and spatio-temporal spreads of the forward through-vegetation cluster. The network is trained with directional channel sounding and simulation data, combining with the inputs of several key impact factors extracted from physics-based observations.

This framework tries to show that, when predicting

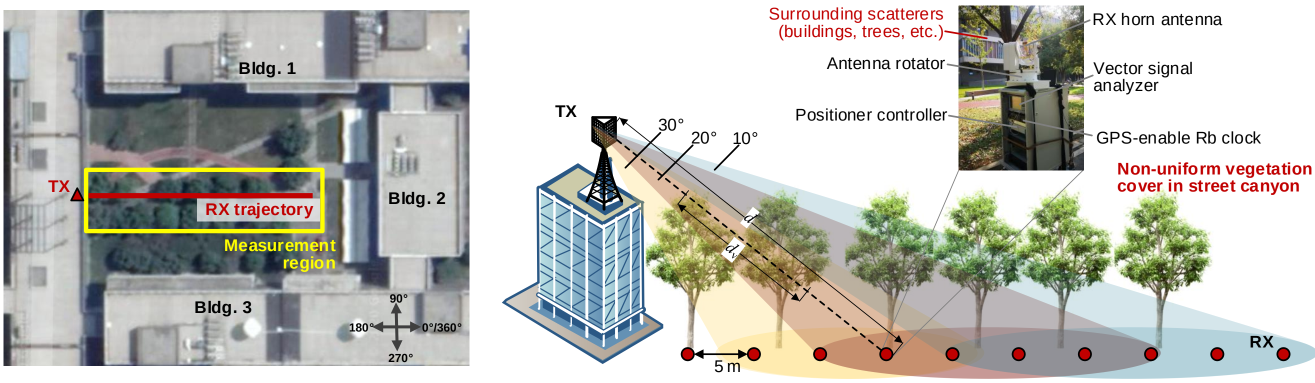



Fig. 3. Illustration of the foliage measurement sites with different TX antenna downtilts in urban street canyon.

mmWave channel characteristics in multiple environments, limited site-specific channel data can be used to train the networks for the predictive modeling of cluster-level channel characteristics along with the environmental features. We could call the corresponding models, which are connected via statistical-based inter-cluster parameters, to rebuild the propagation channels. Another advantage of this framework is that it naturally integrates the blockage effect in cluster generation rather than adding more steps for the application of the blockage model [25]. To enable this framework, we first need to promote a novel set of cluster predictors, that can predict the complex intra-cluster level propagation characteristics over different kinds of physical objects. In particular, the mmWave propagation characteristics of the forward vegetation scattering clusters are investigated in this paper using the hybrid physics-based and data-driven modeling approach.

## III. MmWave Through-Vegetation Channel Data Collection

### A. Directional Channel Measurement Setup

The 28 GHz directional channel measurement in a vegetated street canyon environment was carried out using the directional scanning sounding (DSS) method. As depicted in the overview of the measurement system in Fig. 3, the TX the antenna was mounted on a tripod with a height of 11 m above the ground level and fixed to a particular orientation and three downtilt angles (i.e., 10°, 20°, and 30°). The baseband Golay complementary sequence with a length of 4096 was up-converted to the center frequency of 28 GHz with a radio frequency (RF) null-to-null bandwidth of 600 MHz. A wideband wide-beam horn antenna with a gain of 11.4 dBi and half-power beamwidth (HPBW) of 60° at 28 GHz was employed at the TX side, resulting in the total equivalent isotropically radiated power (EIRP) of 40.5 dBm. The RX antenna was mounted on a custom-designed positioner at 1.9 m height, which was automatically rotated in the entire $2\pi$ azimuth plane and six fixed elevation angles (from 60° to 110°) with an increment of 10°. The RX was equipped with a narrow-beam horn antenna with a gain of 25.6 dBi, HPBW of 8.8° in azimuth and 9° in elevation at 28 GHz. The received signals were recorded via a vector signal analyzer with a multipath delay resolution of 3.3 ns. By using this setup and assuming a minimum signal-to-noise ratio (SNR) of 10 dB, the maximum measurable directional path loss was estimated to be approximately 162 dB. The details on our recently developed channel sounder can be found in [27], [28].

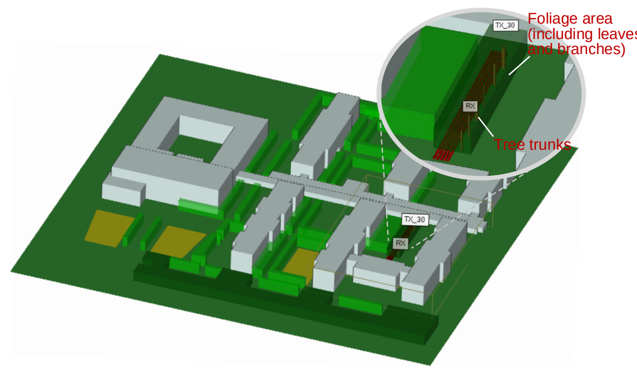



Fig. 4. The perspective view of the 3D ray-tracing model of the concerned environment.

As shown in Fig. 3, the street canyon area is conformed by camphor trees and high-rise (three- or four-storey) buildings with 12-15 m tall, which resembles a typical 3GPP urban microcell (UMi) scenario [25]. Ten RX locations along the street were selected inside the TX illuminated area and 30 total transmitter-receiver (T-R) radio links were tested considering three TX antenna downtilts. The measurement campaign was performed in the OLoS scenario, where the forward paths were obstructed by foliage, branches, and trunks. 216 (36×6) directional power delay profiles (PDPs) across various antenna pointing angles for each T-R combination were recorded.

### B. Ray-Tracing Simulation Setup

In addition to the field measurement, the commercial 3D ray-tracing simulator Wireless InSite® is used by this work to collect sufficient channel data for training the neural network. As shown in Fig. 4, the same street canyon environment is modeled in the simulator, consisting of buildings, floor, foliage, and trucks, while a simplified description is adopted for lacking details, such as small windows, glass doors, and lawn. In terms of vegetation area, new groups of foliage and trunks are respectively created, where the foliage is modeled as a dielectric slab containing randomly oriented leaves (discs) and branches (cylinders) which act to scatter and attenuate the

TABLE I
THE BIOPHYSICAL PROPERTIES OF THE FOLIAGE MATERIAL

| Parameter | Value |
|---|---|
| Base to top heights [m] | 3–10 |
| Leaf radius [m] | 0.020 |
| Leaf thickness [mm] | 0.500 |
| Leaf density [items/m$^3$] | 350 |
| Branch radius [m] | 0.019 |
| Branch length [m] | 0.500 |
| Branch density /vol | 5 |

propagating field [29]. The material of conformal foliage is configured as dense deciduous forest and its properties are detailed in Table I, and the material of cylindrical trunks with the radius of 0.13 m is configured as wood with relative permittivity of 1.99 and conductivity of 0.167 S/m at 28 GHz. The corresponding parameters are estimated from an environment survey.

As for the simulation setup, the maximum numbers of reflections, diffractions, and transmissions are set to 4, 2, and 2, respectively. The maximum number of paths to be computed and stored for each T-R link is set to 100 with the received power threshold of -200 dBm. At the TX side, a directional antenna is employed and the same features as field measurement are configured including height, pointing angles, radiation pattern, and polarization. At the RX side, an omnidirectional antenna with larger HPBW and first-null beamwidth in E-plane is used to cover the same area as the DSS method-based measurement. Only vertical-to-vertical polarization combination is considered. A total of 275 RX locations is selected along the RX trajectory (the red line in Fig. 3) and the spacing of two adjacent RXs is specified as 0.25 m. The simulation results are firstly calibrated against the field channel sounding results.

### C. Channel Data Calibration

To calibrate the measurement and simulation results, directional and omnidirectional path losses are compared based on the close-in (CI) free-space reference distance model. The CI model provided in (2) is parameterized by the path loss exponent (PLE) $n_{\text{CI}}$ as

$$L_{\text{CI}}(d) = L_o(d_0) + 10 n_{\text{CI}} \log_{10}(d/d_0) + X_{\text{CI}} \tag{2}$$

where $L_o(d) = 20\log_{10}(4\pi f d/c)$ is the free space path loss with the speed of light $c$ ($3\times10^8$ m/s) and frequency $f$ in hertz, the free space reference distance $d_0 = 1\,\text{m}$ is used in this work, $n_{\text{CI}}$ is computed via minimum mean square error fitting, and $X_{\text{CI}}$ is a Gaussian random variable $\sim \mathcal{N}(0, \sigma_{\text{CI}}^2)$ with the standard deviation $\sigma_{\text{CI}}$ in dB.

Following the procedure for the calculation of directional and omnidirectional path loss, as recommended in [1] and [5], path loss fitting results are depicted in Fig. 5, together with the parameters summarized in Table II. Note that in Fig. 5 pink crosses represent the measured directional path loss values of the top 50 strongest paths at each RX location; and the blue diamonds, black squares, and pink squares represent the path loss values of measured minimum directional, measured omnidirectional, and simulated omnidirectional channels, respectively, and the blue, black, and pink solid lines represent their corresponding fitting results. The directional best path loss model refers to the fitting results over the lowest directional path loss for each RX location, corresponding to the single antenna pointing direction of the strongest directional received power [1]. It can be observed that the PLEs in OLoS scenario are significantly larger than the PLE of 2 for free-space propagation. The excess loss is expected due to the presence of vegetation blockage. The path loss in the OLoS scenario is mainly related to the tree type, leaf density, and fraction of vegetation cover, so the PLEs are in line with previous studies at 28 GHz in vegetated urban and suburban environments. This observation also proves that propagation losses are very alike when mmWave signals propagate through the same kind of trees with similar features. Fig. 5 also shows the very closed fitting results of measured and simulated omnidirectional path losses, indicating that both channel data can be further used for ANN training and testing. In addition, we observe slight path loss differences concerning TX antenna downtilt for the single strongest power sample, and the omnidirectional PLEs generally increase with the antenna downtilt. The angle-dependent relation between these two components is not significant probably because the height of the TX antenna is relatively low for the illuminated area.

TABLE II
THE MEASUREMENT- AND SIMULATION-BASED PATH LOSS MODEL PARAMETERS IN VEGETATED STREET CANYON AT 28 GHZ

| Data | | Measurement | | | Simulation | | |
|---|---|---|---|---|---|---|---|
| TX ant. downtilt | | 10° | 20° | 30° | 10° | 20° | 30° |
| Dir. best | $n_{\text{CI}}$ | 3.99 | 3.94 | 3.96 | / | / | / |
| | $\sigma_{\text{CI}}$ [dB] | 5.01 | 5.63 | 6.50 | / | / | / |
| Omni. | $n_{\text{CI}}$ | 3.42 | 3.43 | 3.44 | 3.35 | 3.38 | 3.40 |
| | $\sigma_{\text{CI}}$ [dB] | 2.94 | 3.76 | 4.88 | 4.53 | 4.04 | 3.76 |

## IV. RESULTS AND DISCUSSION

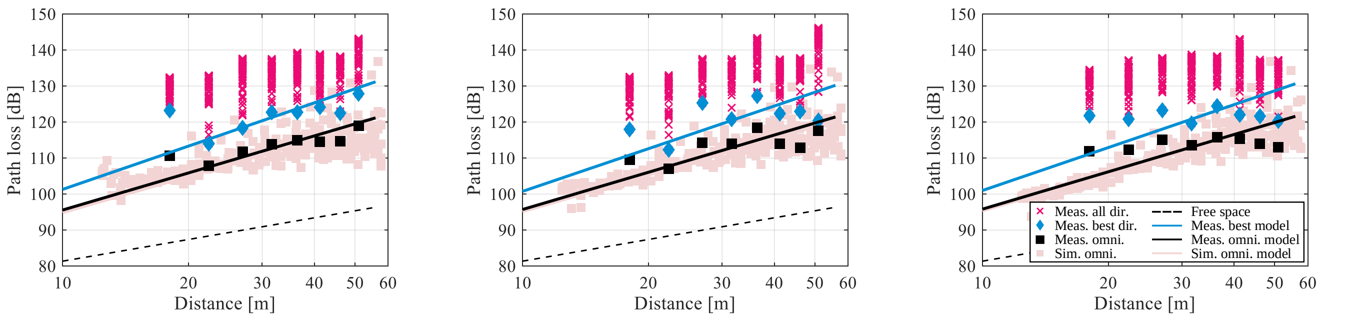
### A. Attenuation in Vegetation

Following the standard recommendation in [11], attenuation in vegetation is defined as the excess transmission loss to all other propagation mechanisms, including free-space, reflection, and diffraction loss.

*1) Data Preprocessing:* The forward received signals propagating through vegetation suffer from the attenuation only in excess of free-space path loss. Accordingly, measurement-based vegetation attenuation is computed as

$$L_{\text{forward,v}}^{\text{Meas}} = \text{EIRP} + G_{\text{R}} + G_{\text{S}} - 10\log_{10}\left(\sum_{\Omega\in\boldsymbol{\Omega}_{\text{v}}} P(\Omega)\right) - L_o(d) \tag{3}$$

where $G_{\text{R}}$ is the RX antenna gain in dB, $G_{\text{S}}$ is the system gain in dB which can be obtained via back-to-back calibration, $P(\Omega) = P(\theta, \phi)$ is the 3D power angular spectrum (PAS) of arrival with $\theta$ in elevation and $\phi$ in azimuth, and $\boldsymbol{\Omega}_{\text{v}}$ is the angle range of through-vegetation spatial lobe. Here,




(a) (b) (c)

Fig. 5. Comparison of the measurement and ray-tracing simulation results in terms of directional and omnidirectional path loss at 28 GHz in vegetated street canyon with TX antenna downtilts of (a) 10°, (b) 20°, and (c) 30°. Three subfigures share the same legend as depicted in (c).

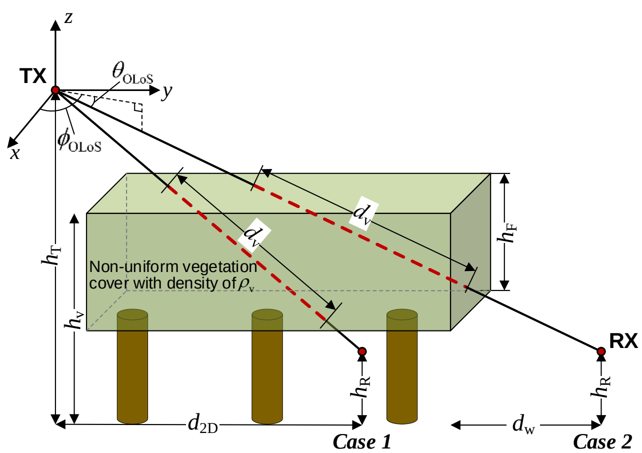



Fig. 6. The geometry definition of vegetation depth in two cases. *Case* 1: TX and RX are located outside and within vegetation, respectively. *Case* 2: both TX and RX are located outside the vegetation.

the forward scattering lobe is detected via a power threshold (typically 20 dB) below the peak power. As for ray-tracing simulations, $L_{\mathrm{v}}$ is computed as

$$L_{\mathrm{omni,v}}^{\mathrm{Sim}} = L_{\mathrm{omni}} - \tilde{L}_{\mathrm{omni}} \tag{4}$$

where $L_{\mathrm{omni}}$ ($\tilde{L}_{\mathrm{omni}}$) denotes simulated omnidirectional path loss with (without) vegetation area. Meanwhile, when beam combination is considered, the directional path loss differences within the same angle range $\boldsymbol{\Omega}_{\mathrm{v}}$ are computed as

$$L_{\mathrm{forward,v}}^{\mathrm{Sim}} = -10\log_{10}\left(\sum_{\Omega_i \in \boldsymbol{\Omega}_{\mathrm{v}}} p_i - \sum_{\tilde{\Omega}_j \in \boldsymbol{\Omega}_{\mathrm{v}}} \tilde{p}_j\right) \tag{5}$$

where $p_i$ ($\tilde{p}_j$) denotes the received power of the $i$th ($j$th) path in ray-tracing simulation with (without) vegetation area and $\Omega_i$ ($\tilde{\Omega}_j$) denotes the AOA.

*2) Modified ITU-R Woodland Obstruction Model:* The modified ITU-R vegetation attenuation model considering non-uniform vegetation cover is expressed as

$$L_{\mathrm{v}}(s) = L_1\left[1 - \exp(-d_{\mathrm{v}}(s)\rho_{\mathrm{v}}\gamma/L_1)\right] \tag{6}$$

where $d_{\mathrm{v}}(s)$ is the length of path within vegetation at RX location $s$, $L_1$ is the maximum attenuation (the upper limit) depending on the type and density of trees, $\rho_{\mathrm{v}}$ is the ratio of effective vegetation region over the total distance in vegetation $d_{\mathrm{v}}$ within $[0,1]$ due to the non-uniform vegetation cover, and $\gamma$ is the specific attenuation for very short vegetative paths in dB/m. Note that $d_{\mathrm{v}}\rho_{\mathrm{v}}$ denotes the sum total distance for the intersections of the OLoS path and the vegetation region. Here, the vegetation depth is estimated following the geometry definition in Fig. 6, written as

$$d_{\mathrm{v}} = \begin{cases} \dfrac{h_{\mathrm{F}}\sqrt{d_{\mathrm{2D}}^2 + (h_{\mathrm{T}} - h_{\mathrm{R}})^2}}{h_{\mathrm{T}} - h_{\mathrm{R}}}, & \text{for } \textit{case } 1 \\ \dfrac{d_{\mathrm{2D}} - d_{\mathrm{w}}}{\cos\theta_{\mathrm{OLoS}}} - \dfrac{h_{\mathrm{T}} - h_{\mathrm{v}}}{\sin\theta_{\mathrm{OLoS}}}, & \text{for } \textit{case } 2 \end{cases} \tag{7}$$

where $h_{\mathrm{T}}$ and $h_{\mathrm{R}}$ are the TX and RX antenna heights, $h_{\mathrm{F}}$ is the difference of top and bottom heights of foliage, $d_{\mathrm{w}}$ is the distance of the RX antenna to the roadside vegetation, $h_{\mathrm{v}}$ is average tree height, and $\theta_{\mathrm{OLoS}} = \arctan\frac{h_{\mathrm{T}} - h_{\mathrm{R}}}{d_{\mathrm{2D}}}$.

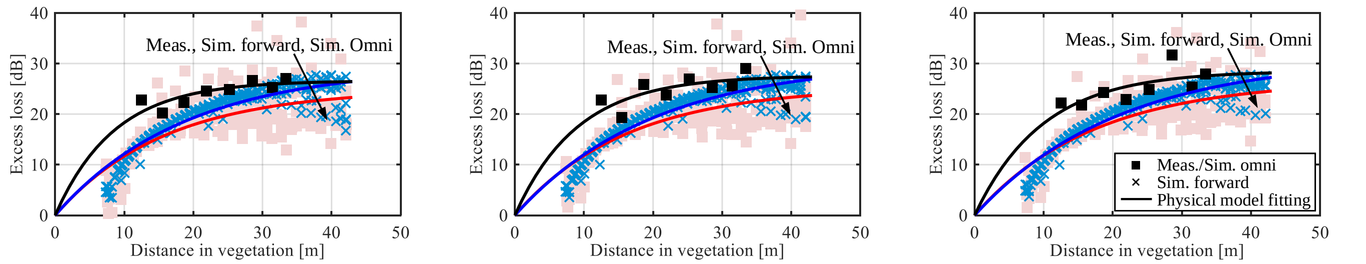
The scatter plots in Fig. 7 show the $L_{\mathrm{v}}(s)$ values with respect to the estimated foliage depths, as well as the fitting results using (6). Table III summarizes the estimated physical-statistical model parameters along with the 95% confidence intervals and the root mean square error (RMSE) of the estimation. When the T-R separation distance is relatively small, the forward through-vegetation propagation is dominant and thus the $L_{\mathrm{forward,v}}^{\mathrm{Sim}}$ is very close to the $L_{\mathrm{omni,v}}^{\mathrm{Sim}}$. With the RX moving away from the TX, the upper limit of forward vegetation attenuation is 5 dB higher than the omnidirectional result along with smaller RMSE. This observation indicates that the strength of the reflected signal from Bldg. 2 would also reduce with the presence of vegetation. Therefore, in the subsequent discussion, only (3) and (5) are used to compute measured and simulated foliage losses, respectively, which have filtered the secondary impact of foliage blockage.

The maximum foliage losses are similar at a specific TX antenna downtilt, while the measured $\gamma$ is generally 1.5 dB/m larger than the simulated result due to the significant loss differences when the $d_{\mathrm{v}}$ is relatively small. Considering practical beam pattern, we underestimated the simulated excess loss when RX moves into the dead zone[1] of TX, corresponding to the smaller foliage depth. Hence, vegetation attenuation also depends on the geometric information of TX and RX relative

[1]Here, the area where RX is very close to the elevated TX antenna is known as dead zone considering practical TX antenna radiation pattern and pointing angle. When RXs are within the dead zone, only weak rays can be observed in the direction of forward propagation (i.e., $\boldsymbol{\Omega}_{\mathrm{v}}$).

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAP.2021.3118463, IEEE Transactions on Antennas and Propagation






(a) (b) (c)

Fig. 7. Predictions from the modified ITU-R vegetation attenuation model based on field measurement and ray-tracing simulation data with TX antenna downtilts of (a) 10°, (b) 20°, and (c) 30°.

TABLE III
THE PARAMETERS FOR THE PHYSICAL-STATISTICAL FOLIAGE ATTENUATION MODEL WITH THE 95% CONFIDENCE INTERNAL

| TX ant. downtilt | | 10° | | | 20° | | | 30° | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Parameter | | Value | 95% conf. | RMSE | Value | 95% conf. | RMSE | Value | 95% conf. | RMSE |
| Meas. forward | $L_1$ [dB] | 26.60 | (23.68, 29.53) | 18.29 | 27.59 | (23.15, 32.02) | 21.54 | 28.51 | (22.81, 34.20) | 25.56 |
| $\rho_\mathrm{v} = 0.95$ | $\gamma$ [dB/m] | 3.10 | (1.92, 4.29) | | 3.05 | (1.48, 4.62) | | 2.89 | (1.28, 4.51) | |
| Sim. forward | $L_1$ [dB] | 28.41 | (27.15, 29.68) | 19.62 | 29.48 | (28.27, 30.69) | 19.20 | 30.22 | (29.03, 31.41) | 18.92 |
| $\rho_\mathrm{v} = 1$ | $\gamma$ [dB/m] | 1.56 | (1.48, 1.65) | | 1.52 | (1.46, 1.59) | | 1.50 | (1.44, 1.56) | |
| Sim. omni | $L_1$ [dB] | 24.28 | (22.68, 25.88) | 29.28 | 24.91 | (23.30, 26.52) | 30.33 | 26.22 | (24.53, 27.91) | 30.97 |
| $\rho_\mathrm{v} = 1$ | $\gamma$ [dB/m] | 1.68 | (1.49, 1.87) | | 1.66 | (1.48, 1.83) | | 1.62 | (1.47, 1.78) | |

locations. To understand the impact of TX antenna downtilt, fitting parameters for measurement and simulation data listed in Table III show that the upper limits of the vegetation loss model increase with the TX antenna downtilt, but the $\gamma$ follow the opposite trend. Although the model in (6) characterizes the impact of TX antenna downtilt with different best-fit values of $L_1$ and $\gamma$, the angle-dependent vegetation attenuation behavior need to be parameterized which jointly consider the impact of TX setups, such as antenna height, downtilt, and radiation pattern.

*3) Angle-Dependent Partition Vegetation Attenuation Model:* The impact of the BS antenna height and mechanical downtilt together with its vertical beam pattern on the coverage of mmWave downlink cellular networks is significant in outdoor high-rise environments [30]. In particular, previous analysis also shows that vegetation attenuation not only depends heavily on the vegetation path length, but also slowly increases with the antenna downtilt. As a result, an angle-dependent partition model is constructed:

$$L_\mathrm{v}(s) = \begin{cases} 0, & \text{if } d_\mathrm{2D}(s) \le \overline{d}_\mathrm{2D} \\ L_1 \left[1 - e^{-d_\mathrm{v}(s)\rho_\mathrm{v}\gamma/L_1}\right] \theta_\mathrm{T}^{\lambda}, & \text{if } d_\mathrm{2D}(s) > \overline{d}_\mathrm{2D} \end{cases} \tag{8}$$

where $\theta_\mathrm{T}$ denotes the TX antenna mechanical downtilt in degrees, $\lambda$ is an empirical found parameter, RX location $s_o$ denotes the boundary of the dead zone, $\overline{d}_\mathrm{2D} = d_\mathrm{2D}(s_o)$ is computed as

$$\overline{d}_\mathrm{2D} = \frac{h_\mathrm{T} - h_\mathrm{R}}{\tan(\theta_\mathrm{T} + 0.6\theta_\mathrm{3dB})} \tag{9}$$

and $\theta_\mathrm{3dB}$ denotes 3 dB E-plane beamwidth of the TX antenna.

Fig. 8 shows the resulting predictions of vegetation losses using the proposed model in (8), where the constant values of $h_\mathrm{T}$, $h_\mathrm{R}$, and $\theta_\mathrm{3dB}$ are selected according to the setups detailed in Section III. The default parameters are fitted based on the measurement and simulation data, respectively. After

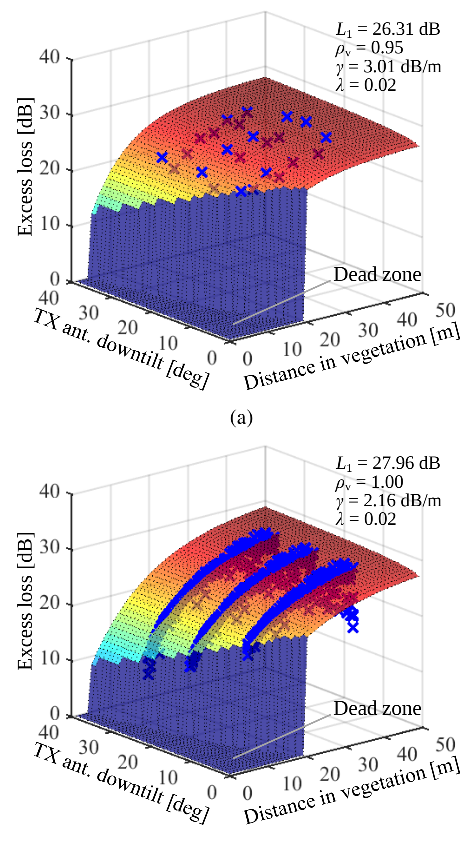



Fig. 8. Fitting results of the proposed height- and angle-dependent partition vegetation attenuation model. (a) Measurement. (b) Simulation.

pruning the outliers, corresponding to the RX located in the dead zone for the ray-tracing model, we have $\gamma = 2.16$ dB/m which is larger than the values in Table III. As for measurement-based results, $\gamma = 3.01$ dB/m is close to the downtilt-specific parameters varying from 3.10 dB/m to 2.89 dB/m. The downtilt-related parameters $\lambda$ for the measured and simulated cases are both 0.02, which is smaller than the value of 0.05 recommended in [11]. However, the ITU-R power function-based vegetation attenuation model $L_{\mathrm{v}} = Af^a d_{\mathrm{v}}^b \theta_{\mathrm{T}}^{\lambda}$ is developed for satellite slant paths, where the TX antenna height is relatively high resulting in large $\theta_{\mathrm{T}}$. Overall, the proposed angle-dependent partition model increases the RMSE performance to 21.63 dB and 16.74 dB for the measurement and simulation data collected over three TX antenna pointing angles, respectively.

### B. Space-Time Propagation Characteristics

To study the mmWave vegetation scattering effect on space-time propagation characteristics, we compare the delay spread (DS) and azimuth angular spread of arrival (ASA) of the through-vegetation OLoS link with the LoS link. Note that only the forward PDPs and PASs within the angle range of $\mathbf{\Omega}_{\mathrm{v}}$ are used for computing channel dispersion parameters.

*1) Delay Spread:* The directional or cluster root-mean-square (RMS) DS is computed as follows

$$\sigma_\tau = \sqrt{\frac{\sum_i \tau_i^2 p_i}{\sum_i p_i} - \left(\frac{\sum_i \tau_i p_i}{\sum_i p_i}\right)^2} \tag{10}$$

where for simulation setup, $p_i$ is the received power of the $i$th path within the angle range of $\mathbf{\Omega}_{\mathrm{v}}$; and for measurement setup, $\{\tau_i = \tau | P(\tau) > N_o + 10\}$ and $N_o$ is the noise floor in dB estimated by the variance of last hundred ns of directional synthesized channel impulse response (CIR). The synthesized PDP $P(\tau)$ is calculated as follows

$$P(\tau) = \sum_{\Omega \in \mathbf{\Omega}_{\mathrm{v}}} P(\tau, \Omega) \tag{11}$$

where $P(\tau, \Omega)$ is the averaged directional PDP.

Fig. 9 depicts the DS samples with respect to RX locations (characterized by $d_{\mathrm{2D}}$) for the LoS and OLoS links, together with their cumulative distribution functions (CDFs). As summarized in Table IV, the mean RMS DSs for the measured and simulated OLoS links blocked by vegetation are on the order of 60 ns. They are significantly larger than the results for the LoS links without vegetation in the simulation model but are smaller than the omnidirectional DS (omni-DS), e.g., the averaged empirical omni-DS of 106.16 ns, 100.87 ns, and 96.37 ns for $\theta_{\mathrm{T}}$ ranging from 10° to 30°. Since the vegetation penetration not only reduces the power of the direct LoS path but also introduces more scattering paths illuminated by the wide-beamwidth antenna, the through-vegetation clusters would eventually promote larger spreads in the delay domain. In addition, the mean RMS DSs reduce with the increasing antenna downtilt for all three cases. This is reasonable as the excess loss caused by forward foliage blockage increases with $\theta_{\mathrm{T}}$, and they contribute with less detectable MPCs at large downtilt which results in smaller DS. The significantly large

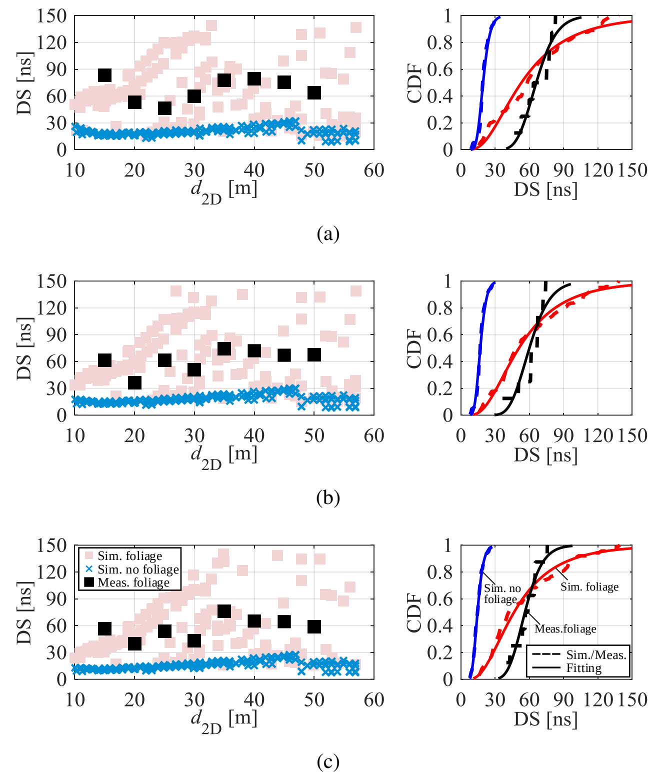



Fig. 9. The forward DS propagating through vegetation versus 2D T-R separation distance, as well as the measured and simulated CDFs, with TX antenna downtilts of (a) 10°, (b) 20°, and (c) 30°.

TABLE IV
THE STATISTICS OF THE CLUSTER-LEVEL DS DUE TO EXISTING FORWARD VEGETATION SCATTERING

| Data | Meas. | | Sim. foliage | | Sim. no foliage | |
|---|---|---|---|---|---|---|
| $\sim \mathcal{N}(\mu, \sigma^2)$ | $\mu$ | $\sigma$ | $\mu$ | $\sigma$ | $\mu$ | $\sigma$ |
| 10° | 67.31 | 13.50 | 62.46 | 33.35 | 20.72 | 4.77 |
| 20° | 61.36 | 12.57 | 59.45 | 32.55 | 18.37 | 4.60 |
| 30° | 57.25 | 11.87 | 55.13 | 32.41 | 16.02 | 4.57 |

standard deviations of simulated RMS DS also indicate that the delay dispersion of foliage shadowed cluster is site-specific, whereas the physical-statistical model seems powerless to characterize this behavior.

*2) Angular Spread:* The cluster RMS ASA is computed as [31]

$$\sigma_\phi = \sqrt{\frac{\sum_i \left|\exp\left(j\phi_i\right) - \mu_\phi\right|^2 p_i}{\sum_i p_i}} \tag{12}$$

and

$$\mu_\phi = \frac{\sum_i \exp\left(j\phi_i\right) p_i}{\sum_i p_i} \tag{13}$$

where $\phi_i \in \mathbf{\Omega}_{\mathrm{v}}$ and for measurement setup $p_i = \sum_\theta P(\theta, \phi_i)$.

Fig. 10 depicts the measured 2D PASs versus RX locations (i.e., $d_{\mathrm{2D}}$) along the trajectory showed in Fig. 3. It is observed that received signals mainly concentrate in the direction of 180°, corresponding to the forward path penetrating the foliage. As the RX moving away from the TX, received signals within the angle range of $[120°, 240°]$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAP.2021.3118463, IEEE Transactions on Antennas and Propagation



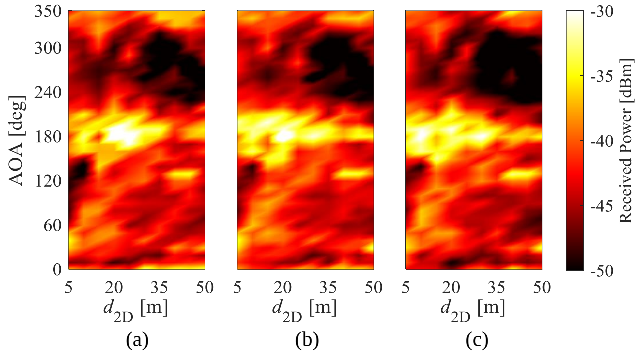



Fig. 10. The empirical PAS variation with the movement of the RX (i.e., $d_{2D}$) with different $\theta_T$: (a) 10°, (b) 20°, and (c) 30°.

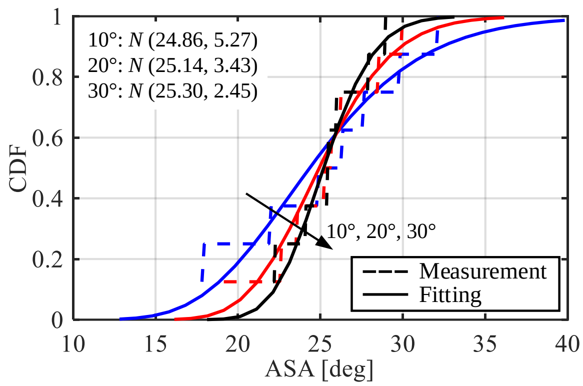



Fig. 11. Empirical and normal fit CDFs of the ASA for the forward foliage scattering spatial cluster with different $\theta_T$.

decay rapidly due to suffering from the distance-dependent foliage attenuation, while the strength of other reflected paths does not change significantly. Fig. 11 shows the empirical and fitted CDFs of the cluster ASA for different antenna downtilt. The CDFs of normal distributions fit well to the empirical data under Kolmogorov-Smirnov (KS) test. The averaged $\sigma_\phi$ are similar with respect to different TX antenna downtilts, e.g., 24.86°, 25.14°, and 25.30° for the $\theta_T$ of 10°, 20°, and 30°, respectively. However, the standard deviations of the $\sigma_\phi$ generally reduce with the increasing of $\theta_T$, which is consistent with the results shown in Fig. 10 that the angular dispersion tends to be very close to the mean at large downtilt.

### *C. ANN-Based Modeling of Vegetation Scattering Effect*

Fig. 2 also depicts a brief overview of the simple adaptive ANN training algorithm, where the ANN-based cluster-level channel predictor can be implemented to characterize mmWave through-vegetation scattering cluster. The ANN consists of a large amount of connected artificial neurons, where each connection is assigned a weight and the weighted sum is then passed through an activation function. The network is trained to be able to learn the nonlinear relationship between the inputs and outputs.

Based on our measurement and simulation data, the physical-statistical prediction model highly reveals the impact of system configurations and environment features on the forward vegetated channel characteristics. Therefore, apart

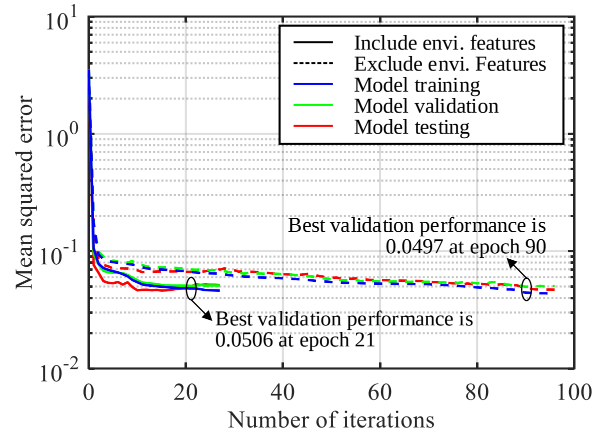



Fig. 12. ANN training performance including and excluding environmental features to identify the RX relative locations (i.e., $c_1$ and $c_2$).

from the basic geometric information of transceivers (TX and RX coordinates), several propagation environment features are converted as inputs to the ANN, which could greatly enhance its predicting performance to the variation of propagation environments. The outputs of the ANN include the estimated channel parameters, such as vegetation attenuation, DS, and ASA. Accordingly, the relationship between the input vector $\boldsymbol{x}$ and output vector $\hat{\boldsymbol{y}}$ is

$$\hat{y}\left(L_{\mathrm{v}}, \sigma_\tau, \sigma_\phi\right) = f_{\mathrm{v}}\left(\boldsymbol{r}_{\mathrm{T}}, \boldsymbol{r}_{\mathrm{R}}, d_{\mathrm{v}}, \theta_{\mathrm{T}} | c_1, c_2\right) \tag{14}$$

where $\boldsymbol{r}_{\mathrm{T}} = (x_{\mathrm{T}}, y_{\mathrm{T}}, z_{\mathrm{T}})$ and $\boldsymbol{r}_{\mathrm{R}} = (x_{\mathrm{R}}, y_{\mathrm{R}}, z_{\mathrm{R}})$ denote the position vectors of TX and RX in the Cartesian coordinate system, respectively. Here, the labels $c_1$ and $c_1$ identify the RX relative location in order to make the channel data more understandable, where $c_1$ labels the RX inside or outside vegetation and $c_2$ indicates whether the RX location is directly below the foliage. These labels are converted into the numeric form (e.g., $0, 1, 2, \ldots$) that ANN training algorithm can then operate the main features of vegetation obstacle in the machine-readable form. Note that the values of input and output parameters should be normalized and mapped within the range of $[-1, 1]$ due to varying in different levels and units.

The MATLAB® neural network fitting tool is adopted using a two-layer feed-forward neural network architecture, which consists of one input layer, one hidden layer with the tansig activation function, and one output layer. The number of hidden neurons $N_{\mathrm{h}}$ is an empirical value with a tradeoff between the rate of convergence and computational complexity. Moreover, increasing the number of hidden neurons may lead to overfitting the training data. Thus, $N_{\mathrm{h}}$ is selected varying from 4 to 20, which would be finally determined by the empirical formula. The network is trained with the Levenberg-Marquardt algorithm, and prediction performance is assessed via the mean square error (MSE) with the maximum iterations of 1000. The total 759 effective channel data samples, collected from both field measurement and ray-tracing simulation after pruning outliers, are split into three sub-datasets, where 70% of the data is for training, 15% for testing, and 15% for validation.

The weight vectors of the connections are trained to achieve the convergence error of 0.05, while the optimal ANN model with 10 hidden neurons performs well with respect to the

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAP.2021.3118463, IEEE Transactions on Antennas and Propagation



overall regression R values. The R value corresponds to the correlation between the outputs of the trained network and the targets of the original estimated channel parameters. To evaluate the performance of adding environmental features during network training, a comparative analysis of the ANN-based channel prediction results including and excluding data labels $c_1$ and $c_2$ is presented in Fig. 12, where more hidden neurons (i.e., $N_{\rm h} = 11$) is required to achieve the same goal of convergence error and R values without data labeling. Fig. 12 also shows that a fast rate of convergence in the case of containing RX location identification would result in a smaller number of iterations (e.g., the maximum iterations of 27 and 96 with and without environmental features, respectively). This observation is in line with the physical-statistical modeling results. For instance, as depicted in Fig. 9, there are significant differences in terms of the RMS DS even if the RX locations share similar T-R separation distances. Accordingly, adding environmental features is indispensable for improving the ANN-based prediction performance of the through-vegetation scattering effect.

### D. Performance Comparison

Once the ANN is trained, the predicting outputs can be compared with original training targets to evaluate the prediction performance. In the case of vegetation attenuation, Fig. 13(a) shows the raw dataset (including training, testing, and validating data) and predicting results overall measurement and simulation data points. The $x$-axis denotes the data sample index. Since there are three TX antenna downtilts of 10°, 20°, and 30° in the ray-tracing setup, the outputs for the simulation part are periodic-like. In general, the trend and value of predicted excess loss fit well with measurement and simulation data, especially for the data samples with shorter foliage depth. The RMSE of the proposed ANN-based model for predicting vegetation attenuation is 7.25 dB, which is significantly smaller than the results fitted with the physical-statistical models in (8). However, relatively large deviations exist when the vegetation attenuation is at a high level. This uncertainty can be also observed in [16] and [20], where the prediction performance gets worse at the data points with either too small or too large real values.

Figs. 13(b) and (c) show the DS and ASA predicted results, respectively. Compared with the training targets, the outcomes of ANN can track the space-time dispersion characteristic with respect to different T-R location combinations. Note that the data samples in one "period" are sorted from largest to smallest T-R separation distances. It can be observed that the DS would rapidly reduce when the RX in the illuminated area moving far away from the TX. Moreover, the ASA also reduces with the increasing of T-R separation distance, which is consistent with the measurement results shown in Fig. 10. The means and standard deviations $(\mu, \sigma)$ of predicted DS and ASA are $(59.97, 16.43)$ in ns and $(24.98, 1.82)$ in deg, respectively. The mean values for the whole dataset are similar to the targets, while the standard deviations of the output space-time channel parameters are much smaller. Compared with the classic physical-statistical channel models, where these

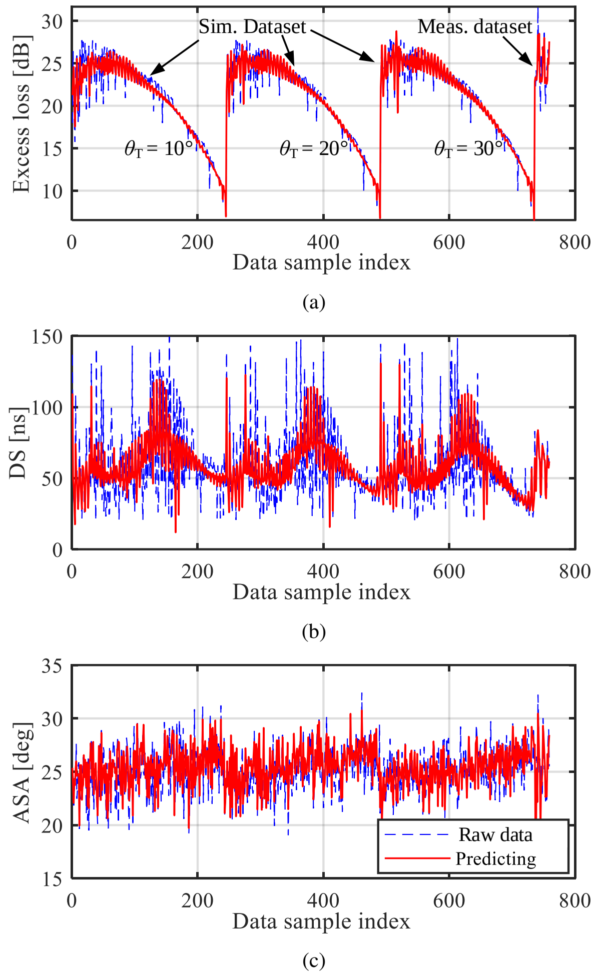



Fig. 13. Comparison of the raw data (including training, testing, and validating data) and predicting results for the forward through-vegetation scattering cluster. (a) vegetation attenuation. (b) DS. (c) ASA.

channel dispersion parameters follow the specific distributions and share the same statistics at each RX location, the ANN-based model has a comparative advantage in the prediction of location-related propagation characteristics combining with the impact of system configurations and environmental features. The RMSE values for the predicted DS and ASA are 12.92 ns and 1.24°, respectively. The prediction performance for the through-vegetation DS is not as accurate as of the other two outputs especially in the course of severe fluctuation with respect to the adjacent RX locations.

### E. Model Validation

Additional ray-tracing simulations in a campus square (see Fig. 14) have been conducted to evaluate the accuracy and generality of the trained model in other environments. The TX and RX antennas share the same configurations as used in Section III-B with the exception of TX antenna height of 26.4 m and pointing angle of $\phi_{\rm OLoS}$ =30° and $\theta_{\rm OLoS}$ =20°. Two kinds of RX regions are considered with a total 180 and 205 RXs in Regions 1 and 2, respectively. Following the procedure

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAP.2021.3118463, IEEE Transactions on Antennas and Propagation



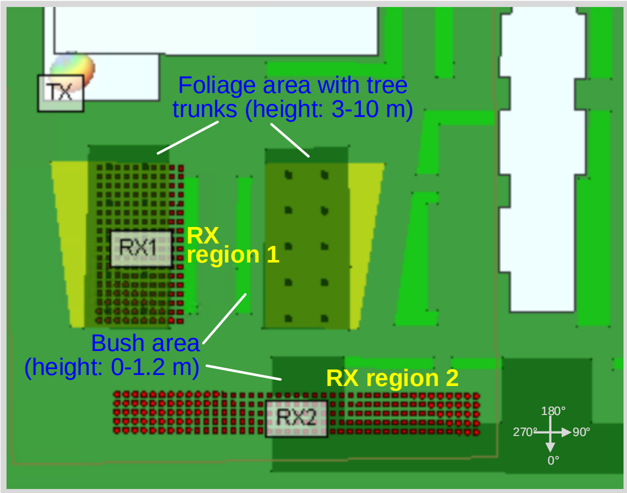



Fig. 14. The 3D model of a campus square environment for model validation.

introduced in Section IV-A and B, the excess loss[2], DS, and ASA for the through-vegetation scattering cluster are obtained. Note that we prune 62 RX data points when the corresponding T-R direct paths do not penetrate through the foliage area. It is worth mentioning that the biophysical properties of the foliage material have a significant impact on the propagation characteristics of through-vegetation cluster [11] in terms of the relationship between the effective permittivity and biophysical parameters [32]. Hence, the same biophysical parameters are used for model validation.

Fig. 15 shows the model validation of simulation and training results. It can be observed that the fluctuations of training results are due to RX locations moving toward TX in rows. For foliage attenuation, Fig. 15(a) compares the ray-tracing simulation results with the predicting results using ANN-based model and physical model in (8). The RMSE values of physical-based predicting results for RX locations in Regions 1 and 2, as well as their combination, are all larger than the values obtained from our proposed ANN-based model as summarized in Table V. The mean values of DS and ASA in campus square are much smaller than the results in the street canyon because no significant scattering effect is considered from bushes with height up to 1.2 m. On the other hand, the ANN-based model performs well in predicting three cluster-level channel parameters for RXs located in region 2, considering the RMSE values in such region are much smaller compared with the results obtained in Region 1. Thus, the trained model is more suitable for RXs far from the foliage area (i.e., $c_1 = 1$) with relatively smaller standard deviations of these channel parameters.

To progress the implementation of the ML-assisted channel modeling approach, a novel set of cluster-level predictors for different kinds of scatterers is urgently needed to rebuild the complex propagation channels. Based on our predicted results, the prediction accuracy and generalization of each predictor can be significantly increased by considering more features.

[2]The corresponding foliage depth is estimated based on the abstract model of all T-R combinations and the relevant foliage following the description given in [4].

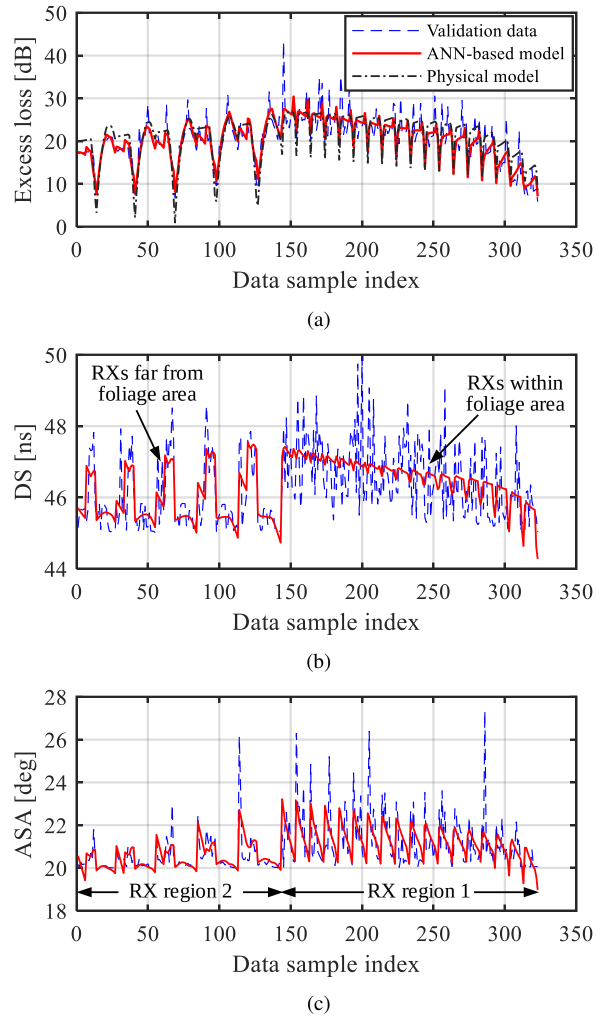



Fig. 15. Comparison of the validation data and predicting results using the trained ANN-based model. (a) vegetation attenuation. (b) DS. (c) ASA.

TABLE V
THE STATISTICS OF THE CLUSTER-LEVEL DS DUE TO EXISTING FORWARD VEGETATION SCATTERING

| Parameters | | Region 1 | Region 2 | All |
|---|---|---|---|---|
| $L_{\mathrm{v}}$ [dB] | RMSE$_{\mathrm{p}}$[a] | 10.67 | 7.36 | 10.25 |
| | RMSE | 10.65 | 7.27 | 10.23 |
| DS [ns] | $(\mu, \sigma)$ | (46.56, 1.03) | (45.91, 0.91) | (46.27, 1.03) |
| | RMSE | 21.16 | 11.02 | 18.17 |
| ASA [deg] | $(\mu, \sigma)$ | (21.26,1.28) | (20.45, 0.73) | (20.90, 1.14) |
| | RMSE | 4.08 | 2.04 | 3.36 |

[a] The RMSE$_{\mathrm{p}}$ is a measure of the differences between raw simulation data and predicting data using the physical model (8) with the parameters shown in Fig. 8(b).

For example, foliage density (high, medium, or low), canopy cover (individual tree or woodland), and carrier frequency can be used as the additional inputs of the ANN to give a complete characterization of vegetation scattering effect across multiple scenarios and frequency bands.

## V. CONCLUSION

The mmWave technology is deemed to be a real solution for future extreme high-throughput wireless communications, where accurate characterization of mmWave propagation channel is a very challenging task due to high sensitivity to physical scatterers. In this study, a framework of the ML-assisted modeling approach has been proposed, which is expected to provide a scalable and robust channel model for multiple environments. To enable this framework, the scattering effect of the forward vegetated cluster is predicted using the hybrid physics-based and data-driven modeling approach. Based on 28 GHz directional measurement and simulation data, physical-statistical channel models have first been developed, including angle-dependent vegetation attenuation, DS, and ASA. The modeling results have demonstrated that the mmWave through-vegetation scattering cluster not only suffers from significant excess loss versus foliage depth but also exhibits larger intra-cluster DS and ASA compared with the LoS path. After data labeling with key environmental features, the ANN is trained for predictive modeling of vegetation scattering effect. Our results have shown that prediction accuracy and the number of iterations can be significantly improved with the identification of RX relative locations. Moreover, the ANN-based model, which takes account of both system configurations and environmental features, provides a remarkable performance improvement in tracking site-specific channel characteristics with validation in different environments.

## Acknowledgment

The authors would like to thank the editors and the anonymous reviewers for their valuable comments and suggestions. They would also like to thank the faculty members from the School of Information Science and Engineering, Southeast University, Nanjing, China, for their kind assistance and fruitful discussions during this study. Additionally, Peize Zhang acknowledges the support from the China Scholarship Council (CSC) program.

## References


[1] T. S. Rappaport, G. R. MacCartney, M. K. Samimi, and S. Sun, "Wideband millimeter-wave propagation measurements and channel models for future wireless communication system design," *IEEE Tran. Commun.*, vol. 63, no. 9, pp. 3029–3056, Sept. 2015.

[2] H. M. Wang, P. Z. Zhang, J. Li, and X. H. You, "Radio propagation and wireless coverage of LSAA-based 5G millimeter-wave mobile communication systems," *China Commun.*, vol. 16, no. 5, pp. 1–18, May 2019.

[3] R. He, B. Ai, G. Wang, M. Yang, C. Huang, and Z. Zhong, "Wireless channel sparsity: Measurement, analysis, and exploitation in estimation," *IEEE Wireless Commun.*, pp. 1–7, 2021. (*Early access*).

[4] C. U. Bas *et al.*, "28 GHz microcell measurement campaign for residential environment," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Singapore, Singapore, Dec. 2017, pp. 1–6.

[5] P. Z. Zhang, B. S. Yang, C. Yi, H. M. Wang, and X. H. You, "Measurement-based 5G millimeter-wave propagation characterization in vegetated suburban macrocell environments," *IEEE Trans. Antennas Propag.*, vol. 68, no. 7, pp. 5556–5567, July 2020.

[6] Y. Zhang, C. R. Anderson, N. Michelusi, D. J. Love, K. R. Baker, and J. V. Krogmeier, "Propagation modeling through foliage in a coniferous forest at 28 GHz," *IEEE Wireless Commun. Lett.*, vol. 8, no. 3, pp. 901–904, June 2019.

[7] Y. Lv, X. Yin, C. Zhang, and H. Wang, "Measurement-based characterization of 39 GHz millimeter-wave dual-polarized channel under foliage loss impact," *IEEE Access*, vol. 7, pp. 151 558–151 568, Oct. 2019.

[8] J. Du, D. Chizhik, R. Feick, M. Rodríguez, G. Castro, and R. A. Valenzuela, "Suburban fixed wireless access channel measurements and models at 28 GHz for 90% outdoor coverage," *IEEE Trans. Antennas Propag.*, vol. 68, no. 1, pp. 411–420, Jan. 2020.

[9] J. Ko *et al.*, "28 GHz millimeter-wave measurements and models for signal attenuation in vegetated areas," in *Proc. Eur. Conf. Antennas Propag. (EuCAP)*, Paris, France, Mar. 2017, pp. 1808–1812.

[10] C. U. Bas *et al.*, "28 GHz foliage propagation channel measurements," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Abu Dhabi, United Arab Emirates, Dec. 2018, pp. 1–6.

[11] ITU, "Attenuation in vegetation," ITU-R, Rec., Tech. Rep. P.833-9, Sep. 2016.

[12] M. A. Weissberger, "An initial critical summary of models for predicting the attenuation of radio waves by trees," Electromagnetic Compat. Anal. Center, Annapolis, MD, USA, Rep. ESD-TR-81-101, 1982.

[13] M. Al-Nuaimi and R. Stephens, "Measurements and prediction model optimisation for signal attenuation in vegetation media at centimetre wave frequencies," *IEE Proc.-Microw. Antennas Propag.*, vol. 145, pp. 201–206, June 1998.

[14] R. He *et al.*, "A kernel-power-density-based algorithm for channel multipath components clustering," *IEEE Trans. Wireless Commun.*, vol. 16, no. 11, pp. 7138–7151, Nov. 2017.

[15] ——, "Clustering enabled wireless channel modeling using big data algorithms," *IEEE Commun. Mag.*, vol. 56, no. 5, pp. 177–183, May 2018.

[16] L. Bai *et al.*, "Predicting wireless mmWave massive MIMO channel characteristics using machine learning algorithms," *Wireless Commun. Mobile Comput.*, vol. 2018, pp. 1–12, Aug. 2018.

[17] S. M. Aldossari and K.-C. Chen, "Machine learning for wireless communication channel modeling: An overview," *Wireless Pers. Commun.*, vol. 106, pp. 41–70, May 2019.

[18] 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 Microw. Antennas Propag.*, vol. 13, no. 8, pp. 1113–1121, July 2019.

[19] U. Masood, H. Farooq, and A. Imran, "A machine learning based 3D propagation model for intelligent future cellular networks," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Waikoloa, HI, USA, Dec. 2019, pp. 1–6.

[20] J. Huang, C.-X. Wang, L. Bai, J. Sun, Y. Yang, J. Li, O. Tirkkonen, and M. Zhou, "A big data enabled channel model for 5G wireless communication systems," *IEEE Trans. Big Data*, vol. 6, no. 2, pp. 211–222, June 2020.

[21] C.-X. Wang, J. Huang, H. Wang, X. Gao, X. You, and Y. Hao, "6G wireless channel measurements and models: Trends and challenges," *IEEE Veh. Technol. Mag.*, vol. 15, no. 4, pp. 22–32, Dec. 2020.

[22] E. Ostlin, H. Zepernick, and H. Suzuki, "Macrocell path-loss prediction using artificial neural networks," *IEEE Trans. Veh. Technol.*, vol. 59, no. 6, pp. 2735–2747, July 2010.

[23] S. P. Sotiroudis, S. K. Goudos, K. A. Gotsis, K. Siakavara, and J. N. Sahalos, "Application of a composite differential evolution algorithm in optimal neural network design for propagation path-loss prediction in mobile communication systems," *IEEE Antennas Wireless Propag. Lett.*, vol. 12, pp. 364–367, Mar. 2013.

[24] M. Ayadi, A. Ben Zineb, and S. Tabbane, "A UHF path loss model using learning machine for heterogeneous networks," *IEEE Trans. Antennas Propag.*, vol. 65, no. 7, pp. 3675–3683, July 2017.

[25] 3GPP, "Study on channel model for frequency from 0.5 to 100 GHz," 3GPP, Tech. Rep. 38.901 (V15.0.0), June 2018.

[26] J. Huang, Y. Liu, C. Wang, J. Sun, and H. Xiao, "5G millimeter wave channel sounders, measurements, and models: Recent developments and future challenges," *IEEE Commun. Mag.*, vol. 57, no. 1, pp. 138–145, Jan. 2019.

[27] J. Li, P. Z. Zhang, H. M. Wang, C. Yu, and W. Hong, "High-efficiency millimeter-wave wideband channel measurement system," in *Proc. Eur. Conf. Antennas Propag. (EuCAP)*, Krakow, Poland, Apr. 2019, pp. 1–5.

[28] P. Z. Zhang, H. M. Wang, and W. Hong, "Radio propagation measurements and cluster-based analysis for 5G millimeter-wave cellular systems in dense urban environments," *Front. Inf. Technol. Electron. Eng.*, vol. 22, no. 4, pp. 471–487, Apr. 2021.

[29] S. A. Torrico, H. L. Bertoni, and R. H. Lang, "Modeling tree effects on path loss in a residential environment," *IEEE Trans. Antennas Propag.*, vol. 46, no. 6, pp. 872–880, June 1998.

[30] J. Yang, M. Ding, G. Mao, Z. Lin, D. Zhang, and T. H. Luan, "Optimal base station antenna downtilt in downlink cellular networks," *IEEE Trans. Wireless Commun.*, vol. 18, no. 3, pp. 1779–1791, Mar. 2019.

[31] B. H. Fleury, "First- and second-order characterization of direction dispersion and space selectivity in the radio channel," *IEEE Trans. Inf. Theory*, vol. 46, no. 6, pp. 2027–2044, Sep. 2000.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAP.2021.3118463, IEEE Transactions on Antennas and Propagation



[32] S. Torrico, H. Bertoni, and R. Lang, "Theoretical investigation of foliage effects on path loss for residential environments," in *Proc. IEEE Veh. Technol. Conf. (VTC)*, vol. 2, Atlanta, GA, USA, May 1996, pp. 854–858.

**Peize Zhang** (Student Member, IEEE) received the B.S. degree from Beijing University of Posts and Telecommunications, Beijing, China, in 2015, and the M.S. degree from China Academy of Telecommunications Technology, Beijing, China, in 2018, both in Electrical Engineering. He is currently pursuing the Ph.D. degree at the State Key Laboratory of Millimeter Waves, Southeast University, Nanjing, China.

From 2016 to 2018, he was also a research assistant with the China Academy of Information and Communications Technology, Beijing, China. Since 2021, he is a visiting Ph.D. student at the Université catholique de Louvain, Louvain-la-Neuve, Belgium. His current research interests focus on propagation measurement, deployment, and performance analysis of millimeter-wave wireless communication systems.

**Cheng Yi** (Student Member, IEEE) received the B.S. degree in information engineering from Southeast University, Nanjing, China, in 2016. He is currently working toward the Ph.D. degree at the State Key Laboratory of Millimeter Waves, Southeast University.

His main research interests include flexible millimeter-wave channel sounder design and channel modeling.

**Bensheng Yang** (Student Member, IEEE) received the B.S. degree from Northeastern University at Qinhuangdao, China, in 2013, and the M.S. degree from Xidian University, Xi'an, China, in 2016. He is currently pursuing the Ph.D. degree at the State Key Laboratory of Millimeter Waves, Southeast University, Nanjing, China.

From 2016 to 2017, he was an Engineer with Huawei Technologies Co., Ltd., Nanjing, China. His current research interests include electromagnetic vector antenna design, radio propagation measurements, and signal processing.

**Haiming Wang** (Member, IEEE) was born in 1975. He received the B.S., M.S., and Ph.D. degrees in Electrical Engineering from Southeast University, Nanjing, China, in 1999, 2002, and 2009, respectively.

He joined the School of Information Science and Engineering, Southeast University, Nanjing, China, in 2002, and is currently a distinguished professor. He is also a part-time professor with the Purple Mountain Laboratories, Nanjing, China. He has authored and co-authored over 50 technical publications in IEEE TRANSACTIONS ON ANTENNAS AND PROPAGATION and other peer-reviewed academic journals. Prof. Wang has authored and co-authored over more than 70 patents and 52 patents have been granted. He was awarded for contributing to the development of IEEE 802.11aj by the IEEE Standards Association in July 2018.

His current research interests include AI-powered antenna and radiofrequency technologies, AI-powered channel measurement and modeling technologies, and millimeter-wave and THz wireless communications.

**Claude Oestges** (Fellow, IEEE) received the M.Sc. and Ph.D. degrees in Electrical Engineering from the Université catholique de Louvain (UCLouvain), Louvain-la-Neuve, Belgium, in 1996 and 2000, respectively.

In January 2001, he joined the Smart Antennas Research Group (Information Systems Laboratory), Stanford University, Stanford, CA, USA, as a Postdoctoral Scholar. From January 2002 to September 2005, he was associated with the Microwave Laboratory, UCLouvain, as a Postdoctoral Fellow of the Belgian Fonds de la Recherche Scientifique (FRS-FNRS). He is currently a Full Professor with the Electrical Engineering Department, Institute for Information and Communication Technologies, Electronics and Applied Mathematics (ICTEAM), UCLouvain. He is the author or coauthor of four books and more than 200 journal papers and conference communications.

Dr. Oestges was the Chair of COST Action CA15104 IRACON during 2016–2020. He was the recipient of the 1999—2000 IET Marconi Premium Award and the IEEE Vehicular Technology Society Neal Shepherd Award in 2004 and 2012.

**Xiaohu You** (Fellow, IEEE) was born in 1962. He received the M.S. and Ph.D. degrees in electrical engineering from Southeast University, Nanjing, China, in 1985 and 1988, respectively.

Since 1990, he has been working with the National Mobile Communications Research Laboratory, Southeast University, where he holds the rank of the Director and a chair Professor. From 1999 to 2002, he was the Principal Expert of the C3G Project, where he was responsible for organizing China's 3G Mobile Communications Research and Development Activities. From 2001 to 2006, he was the Principal Expert of the China National 863 Beyond 3G FuTURE Project. Since 2013, he has been the Principal Investigator of the China National 863 5G Project. He has contributed over 100 IEEE journal articles and two books in the areas of adaptive signal processing, neural networks, and their applications to communication systems. His research interests include mobile communication systems and signal processing and its applications.

Dr. You was selected as an IEEE Fellow due to his contributions to development of mobile communications in China, in 2011. He was a recipient of the National First Class Invention Prize in 2011. He served as the General Chair for the IEEE Wireless Communications and Networking Conference (WCNC) 2013, the IEEE Vehicular Technology Conference (VTC) 2016 Spring, and the IEEE International Conference on Communications (ICC) 2019. He is currently the Secretary General of the FuTURE Forum, and the Vice Chair of the China IMT-2020 Promotion Group and the China National Mega Project on New Generation Mobile Network.