IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received June 21, 2020, accepted July 12, 2020, date of publication July 29, 2020, date of current version August 20, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3012661

Radio Propagation Models Based on Machine Learning Using Geometric Parameters for a Mixed City-River Path

ALLAN DOS S. BRAGA[1], HUGO A. O. DA CRUZ[1], LESLYE E. C. ERAS[2], JASMINE P. L. ARAÚJO[1], MIÉRCIO C. A. NETO[1], DIEGO K. N. SILVA[2], AND GERVÁSIO P. S. CAVALCANTE[1]

[1]Institute of Technology, Federal University of Pará, Belém 66075-110, Brazil
[2]Institute of Geoscience and Engineering, Federal University of Southern and Southeastern of Pará, Marabá 68505-080, Brazil

Corresponding author: Allan Dos S. Braga (allan.braga@itec.ufpa.br)

This work was supported in part by the CNPq and in part by the INCT.

ABSTRACT This work presents and evaluates the use of geometric parameters of the environment in the prediction of the electric field in mixed city-river type environments, employing two techniques of Machine Learning (ML) as Artificial Neural Networks (ANN) and Neuro-Fuzzy System (NFS). For its development, measurements were carried out in Amazon Region, Belém city, in the 521 MHz band. The input parameters for an ANN and NFS are the distance between transmitter and receiver, the distance only over the river, the height of the ground, the radius of the first Fresnel ellipsoid, and the electric field of free space. The ANN is a Multilayer Perceptron Network (MLP) that uses the Levenberg-Marquardt training algorithm and cross-validation method. The NFS is an Adaptive Neuro-Fuzzy Inference System (ANFIS) that uses the model Sugeno. The results obtained compared with the classic literature models (ITU-R 1546 and Okumura-Hata) in the city for distances up 20 km and over the river for distances up 5 km. A quantitative analysis is performed between the measured and predicted data through the Standard deviation (SD), Root Mean Square Error (RMSE), and the Grey Relational Grade, combined with the Mean Absolute Percentage Error (GRG-MAPE). For ANN, the SD is 2.13, the RMSE is 2.11 dB, and the GRG-MAPE is 0.96. Also, for the NFS, the SD is 1.99, the RMSE is 2.06 dB, and the GRG-MAPE is 0.97. It should be noted that the transition zone between the city and the river was characterized by the proposed ANN and NFS in contrast with the classic literature models, which did not demonstrate coherence in the transition zone.

INDEX TERMS Artificial neural networks, geometric parameters, neuro-fuzzy systems, mixed city-river path, radio propagation model.

I. INTRODUCTION

Several countries around the World use rivers for commerce, transportation, and tourism. The use of rivers is an economical and ecological alternative compared to other means of transport such as road and rail. This work takes as a case of study, a scenario in the Amazon Region of Brazil in Belém city. Rivers in the Amazon Region have an essential role, especially in the daily routine for people who live on the border of the islands [1]. Then, telecommunications systems need coverage over the river even when the transmitter is in the city, as Digital Television, which reaches long distances.

In the propagation of radio waves, several factors can influence the electric field level in the path from the transmitter to the receiver, such as the transition through different types of environments, such as the height of buildings, as well as the types of vegetation and if the waves travel over areas without vegetation or over areas covered with water [2].

In the literature, there are works considering mixed land-sea paths. Most of these studies are for low frequencies, in [3] it was for LF, MF and HF (Low, Medium and High Frequency) and vertical polarization using the Finite Element Method (FEM) which was compared with the Millington

The associate editor coordinating the review of this manuscript and approving it for publication was Wei Zhang.

method for mixed paths and obtained a good agreement. In [4], a mixed land-sea path and a sea path for HF were compared using Millington and measured data. Results for the mixed land-sea path presented higher attenuation than results for only sea path.

For higher frequencies as the case of the UHF (Ultra High Frequency) band, the classical models of literature which study mixed land-water path are Okumura [5] and ITU-R. P.1546 [6]. Both models do not consider obstacles when the signal passes from city to the water; as a consequence, the transition zone city-water is not well predicted. Furthermore, in [7] ITU-R. P.1546 presents a difference of 20 dB about measured data.

A recent work presents a radio propagation model called Inland for a mixed land-river path [8], and it based in the Round Earth Loss model (REL) with a modification of the free space equation, which replaced with the equations of Okumura-Hata. In order to calculate path loss for the mixed path, was used the ITU-R P.1546 methodology. Additionally, knife diffraction was used for calculating the attenuation caused by constructions. The Inland model does not is applied for distances lower than 1km. The Inland model is not applied for distances of less than 1 km.

Another technique different from those mentioned above that considers the electric field's calculation as a regression problem is Artificial Neural Networks (ANN). The ANN allows to design radio propagation models more accurate and computationally efficient for different environments, for example in [9] for the city of Tripoli, capital of Libya, were collected Received Signal Strength (RSS) in three different frequencies 900 MHz, 1800 MHz and 2100 MHz in five different types of areas: Dense Urban, Urban, Dense Suburban, Suburban, and Rural. The ANN-based model presented results that improved from 7.1 dB to 28.8 dB in the accuracy of the Hata model.

The work in [10] analyzes the neural network parameters more relevant to predict the propagation loss in the VHF band, as input types, activation function, number of hidden neurons, and algorithm training.

Furthermore, in [11], an ANN to obtain the path loss for the mixed urban-suburban-rural path was used for different frequencies in the UHF band and has good agreement with measurement data. The inputs of the ANN are from Standard Propagation Model and also environment type, land use distribution, and diffraction loss. To know the environment type, through high-resolution maps.

The work developed in [12] describes the propagation of radio in mixed environments and uses genetic algorithms (GA) to determine the ideal height of the receiver's antenna to obtain the maximum level of Digital TV signal. The work [13] presents a propagation model for mixed environments with city-river-forest paths using Dyadic Green's Functions (DGF). The model obtained acceptable results in the UHF range.

In [14] presents a radio propagation model for the UHF band in an Amazonian environment. The model analyzes parameters such as the transition zone between the city-river, the water level, the angle of incidence, and the forest's electrical parameters.

In work in [15], were developed models based on propagation ANN, ANFIS, and kriging techniques for band VHF and UHF. The models based on artificial intelligence techniques had minor errors compared to literature models, showing the effectiveness of using these techniques in the propagation of radio waves.

The present work proposes two radio propagation models using ANN type Multiple Layer Perceptron (MLP) and NFS type Adaptive Neuro-Fuzzy Inference System (ANFIS), both for a mixed city-river path. These proposed models have as one of its objectives to implement inputs quickly to be calculated, in contrast with [11], which use as one of its inputs the uniform theory of diffraction that involves complex calculations. Another aspect is the type of environment characterized by urban and suburban areas, both densely wooded, a common scenario in Amazonian cities. Economically important rivers surround these cities. Thus, the models presented characterize the transition zones between the media satisfactorily, a characteristic not present in the other models studied.

The ANN and NFS inputs use geometric data of the environment, taking as a reference to the scenario considered by Ikegami in [16]. Ikegami represents the constructions as rectangles with the same size. Otherwise, the proposed model can be used for distances lower than 1 km and can predict the electric field's values in the transition zone city-river. Considering the characteristic of these models and their application will be called a Geometric Model for Mixed Paths using Neural Networks (GMMP-NN) and Neuro-Fuzzy (GMMP-NF).

II. THEORETICAL BACKGROUND

This subsection will discuss some important concepts for the understanding of the work. Empirical radio propagation models are used to compare the results obtained. The Fresnel Zone theory is relevant since the first radius's length is one of the geometric parameters used in the modeling. Finally, understanding the concepts of machine learning techniques, such as Artificial Neural Networks and Neuro-Fuzzy Systems, are essential, as they are the two techniques used in the creation of propagation models in this work.

A. RADIO PROPAGATION MODELS

In this work are used and analyzed, two classical propagation models to compare with this proposal. These models are Okumura-Hata and ITU-R. P.1546-5, which are detailed next:

1) OKUMURA-HATA

This model is valid for distances between 1 and 100 km for transmission antenna heights of 30 to 200 m and considers an E.R.P (Effective Radiated Power) of 1 kW. Okumura obtained curves based on an extensive measurement campaign carried out in Japan. Hata [17] developed numerical expressions for

Okumura's standard propagation curves, including the most commonly used corrections in mobile radio communications.

The basic formula for the urban environment, E_{urb}, is as follows [5]:

$$E_{urb}(dB\mu) = 69.82 - 6.16log(f) + 13.82log(h_t) + a(h_r) - (44.9 - 6.55log(h_t))(log(d))^b \quad (1)$$

where:

- f: Frequency: 150 MHz $\leq$ f $\leq$ 1500 MHz.
- h_t: Effective transmitting antenna height, 30 m $\leq h_r \leq$ 200 m.
- h_r: Height above ground of receiving antenna, 1 m $\leq h_r \leq$ 10 m
- d: Distance (Km)
- b: 1 if d $\leq$ 20 Km
- a(h_r): Receiver antenna height correction factor [3]

In order to calculate b, when distances are greater than 20 km, is used the following equation:

$$b = 1 + (0.14 + 0.000187f + 0.00107h_t)[log(0.05d)]^{0.8} \quad (2)$$

The following equation gives the electric field for the suburban environment:

$$E_{sub} = E_{urb}(dB\mu) - 2\left(log\left(\frac{f}{28}\right)\right)^2 - 5.4 \quad (3)$$

Finally, for distances, up to 30 km between a transmitter located on the ground and the receiver located over the river, the correction factor, K_{mp}, for a mixed path is:

$$\beta = \frac{d_1}{d} \quad (4)$$

$$\beta = -8\beta^2 + 19\beta \quad (5)$$

where:

- d_1: Distance only over water.
- β: The fraction of water in the total path between transmitter and receiver.

2) ITU-R P.1546

This recommendation presents a set of normalized propagation curves to predict the electric field values in point-to-multipoint terrestrial links. These curve patterns are for E.R.P. of 1 kW, at the nominal frequencies of 100, 600, and 2000 MHz for terrestrial, hot, and cold seas [6]. The methodology to calculate the electric field in a mixed land-sea path uses the following expressions:

$$E = (1 - A) \cdot E_{land}(d_{total}) + A \cdot E_{sea}(d_{total}) \quad (6)$$

The interpolation factor of the mixed path, A, is given by:

$$A = A_0(F_{sea})^V \quad (7)$$

Analytically this factor is defined by:

$$A_0(F_{sea}) = 1 - (1 - F_{sea})^{\frac{2}{3}} \quad (8)$$

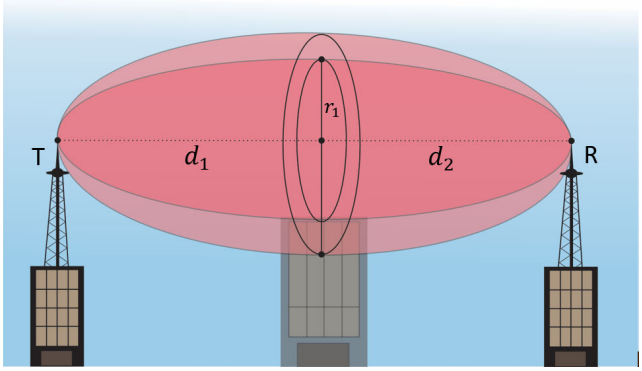

FIGURE 1. Successive fresnel zones.

V is calculated with the following expression:

$$V = max\left[1.0, 1.0 + \frac{\Delta}{40.0}\right] \quad (9)$$

$$\Delta = E_{sea}(d_{total}) - E_{land}(d_{total}) \quad (10)$$

$$F_{sea} = \frac{d_{ST}}{d_{total}} \quad (11)$$

where:

- d_{ST}: Distance only over the water.
- d: Distance between transmitter and receiver.
- $Eland$ (dBμV/m): Electric field on land.
- $Esea$ (dBμV/m): Electric field over the sea.

E_{land} and E_{sea} is described at [6].

B. FRESNEL ZONES

Fresnel zones allow determining the diffraction loss as a function of the path difference around an obstruction. Fresnel zones represent successive regions where secondary waves have a path length from the transmitter to receiver, which are $\frac{n\lambda}{2}$ greater than the total path length of a line of sight path illustrated in Fig.1. [18].

The radius of n^{th} Fresnel zone circle, can be expressed in terms of λ, d_1 and d_2:

$$r_n = \sqrt{\frac{n\lambda d_2}{d_1 + d_2}} \quad (12)$$

- r_n: Number of radius of Fresnel.
- d_1: Distance between the transmitter and the obstacles.
- d_2: Distance between the obstacle and the receiver.

This approximation is valid for $d_1, d_2 \gg r_n$.

C. MACHINE LEARNING TECHNIQUES IMPLEMENTED

Machine learning techniques are applied in problems related to pattern recognition, such as Artificial Neural Networks (ANN) and Neuro-Fuzzy System (NFS). Generally, these problems are based on previously known data sets.

1) ARTIFICIAL NEURAL NETWORK

ANNs models systems using synaptic connections that simulate the human nervous system. In other words, they can

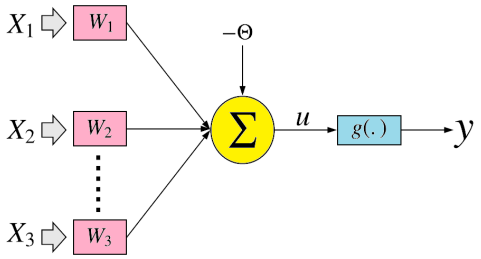

FIGURE 2. Artificial neuron [19].

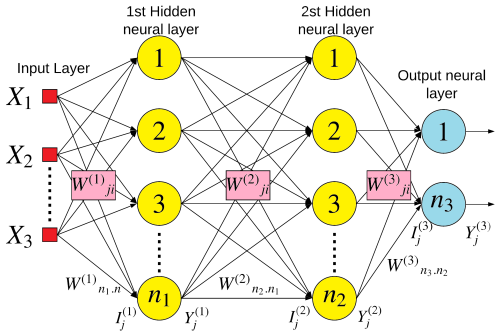

FIGURE 3. Generic multilayer perceptron neural network [19].

process information [19]. The ANNs consist of a large number of simple computational units, called artificial neurons, shown in Fig. 2, interconnected by a large number of interconnections. Mathematically, ANN's are represented by vectors / matrixes of artificial synaptic weights, which tend to reproduce human characteristics (learning, association, generalization, abstraction, etc.).

$$\mu = \sum_{i=1}^{n} W_i \cdot X_i - \Theta \tag{13}$$

- Activation function, g, limits the neuron's output within a range of reasonable values to be assumed by its functional image.
- The output signal, y, is the final value produced by the neuron concerning a given set of input signals, given by [19].

$$y = g(\mu) \tag{14}$$

The arrangement of the neurons is defined for the architecture of an ANN. Generally, the classification of ANNs is about the number of layers or types of connections. Both of them consider the direction of the neurons' synaptic connections, and they can be feed-forward or feedback. This article uses a Multiple Layer Perceptron (MLP) characterized by at least one intermediate layer of neurons [20] with a feed-forward architecture. Thus, the output of the neurons from the first layer will represent the input of the neurons of the second layer and so on, as shown in Fig.3

Each neuron processes the input signals from the network, which are controlled by synaptic weights that adapt, by a

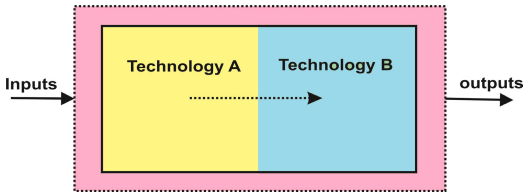

FIGURE 4. A sequential hybrid system [21].

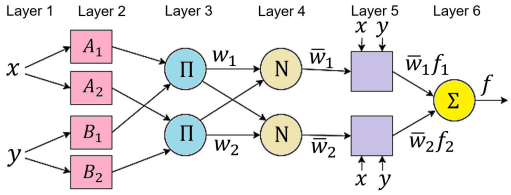

FIGURE 5. Basic architecture of an ANFIS.

learning algorithm, during training. Because of their structure, ANNs are quite effective in learning patterns from data: non-linear, incomplete, with noise, and even compounded by contradictory examples.

2) NEURO-FUZZY SYSTEMS

This type of system uses two types of complementary tools, a part that uses a low-level data structure to form the neural network. On the other hand, the other fuzzy logic (FL), which uses high-level knowledge and linguistic information provided by the specifications of a domain.

It is a type of hybrid system formed by the union of two modeling techniques, neural network, and fuzzy logic. Through FL, knowledge is inferred, and learning from artificial neural networks is added [21]. An outline of a sequential type hybrid system can be identified in Fig. 4. These systems use technologies in a pipeline-like structure where one technology's output becomes the input to another technology.

The use of a neuro-diffuse system has the advantages of each improved technique and the disadvantages of each are reduced. In this work a type of NFS called Adaptive Neuro-Fuzzy Inference System (ANFIS) was employed, the representation of knowledge is encoded in a format similar to the logic used by humans, while in ANN, knowledge is encoded in weights, which can be difficult to interpret. ANFIS is applied in several areas of knowledge, such as the works [15], [22], applicable to the area of radio propagation. In Fig. 5 shows the structure ANFIS layers:

- Layer 1: Represents the input values normalized at intervals such as [0, 1] or [−1, 1]. The output of this layer is the degrees of pertinence based on the premise of each rule;
- Layer 2: The execution stage of the "IF" part of the Sugeno Fuzzy system rules e the function parameters will be adjusted by some training algorithm or by an optimization method;

FIGURE 6. Points measured in Belém of Pará, both in the city and on the river.

- Layer 3: Defined by FIS rules and normalizes the degrees of activation of the rules. The neurons have fixed activation functions, and they cannot be adjustable. Each node of this layer performs the normalization function, which is a preprocessing for defuzzification;
- Layer 4: The calculate each neuron output normalized by the product of the previous layer with the consequent activation value. The neurons have a Sugeno-type activation function and are n-order combinations of the input signals. These functions can have the parameters adjusted by the optimization method or training algorithm;
- Layer 5: The sum of the outputs of the previous layer, and so the numerical value of the ANFIS output.

III. MEASUREMENT CAMPAIGN

Two measurement campaigns were carried out in Belém of Pará for a transmitter located in the city for a central frequency of 521 MHz, to collect receiver power, in order to elaborate and validate the proposed radio propagation model. The first measurement campaign was over the city for distances from 0.5 km to 20 km. The second measurement campaign was over the river for distances up 5 km.

A. SCENARIO MEASUREMENTS

Belém is a city formed of urban and suburban environments with the presence of forest. The weather is equatorial, and the number of inhabitants is 1.452 million [23]. In Fig. 6 the measurement points in the city are the red ones, and the measurement points over the river are the yellow ones.

B. MEASUREMENT METHODOLOGY

At the central frequency of 521 MHz was performed receiver power collection, operated by a Digital TV station, which shared the data used for the transmission. The transmitting antenna's height on the ground is 114.58 m, the transmission power 6 kW, the gain of the transmitting antenna 11.1 dB, and the cable and connector losses of 1.72 dB.

For the signal's reception, in both measurement campaigns, the spectrum analyzer site master ANRITSU S332E collected the receiver power. In the city was used an omnidirectional antenna with a gain of 5 dB and a height of 5 m. Over the water, the receiver antenna was a dipole with a beamwidth of 78 degrees, a gain of 2.15 dB, and a height of 4 m.

FIGURE 7. City-river scenery in Belém of Pará.

FIGURE 8. "Curupira" boat.

Data over the river were collected continuously at an average speed of 11 km/h. The surface of the river is smooth because it does not present considerable waves, as shown in Fig.7 where is illustrated the scenario city-river and the Fig.8 the boat used, called Curupira.

IV. METHODOLOGY

The methodology is divided into three stages, according to Fig. 9, for designing the ANN-based and NFS-based empirical model for electric field prediction in the mixed city-river path, at 521 MHz.

A. PRE-PROCESSING

In both measurement campaigns, a Garmin GPS was used to collect the geographic coordinates. In order to obtain the distance in kilometers, the sexagesimal geographical coordinates were transformed into decimal coordinates, then was

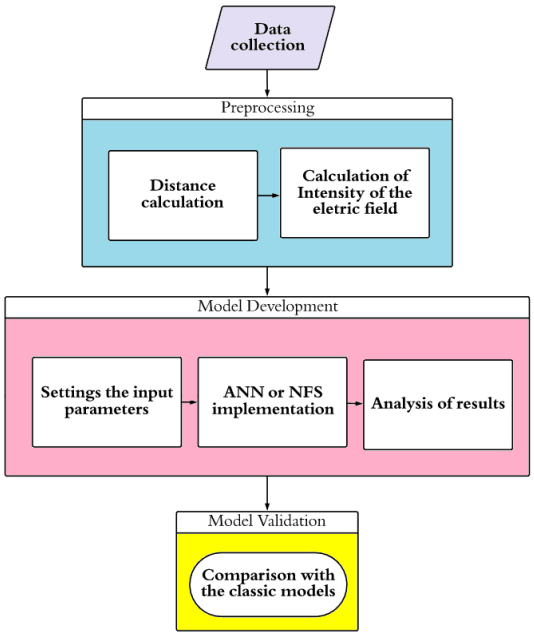

FIGURE 9. Flowchart of the methodology of the proposed models.

used the Haversine's formula [24] and was applied spherical law cosine.

Furthermore, receiver power is converted to the electric field using the following formula [25]:

$$E(dB\mu) = Pr + 20logf + 77.2 - G_r \tag{15}$$

where,

- P_r: Received power.
- f: Frequency.
- G_r: Receiver antenna gain.

B. SCENARIO OF THE MODELS

The scenario of the proposed models uses uniform constructions, as Ikegami in [16], where his model has good coherence with measurement data in Europe. However, this scenario is for a mixed city-river path, where it includes the height of the ground because it affects the transmitter's height when the receiver is over the river.

The geometrical parameters illustrated in Fig. 10 are important because they are used to calculate the inputs of the ANN and NFS. The height of the receiver antenna, h_R; the height of the building, h_B; the height of the ground, h_G; distance between transmitter and receiver, d_{TR}, distance only over the water, d_W; the distance of the direct ray between transmitter and receiver, R_{TR}, and angle of the direct ray, θ_D.

The width of the street is calculated using a top view as in [11], where between transmitter and receiver is traced a straight line, which forms an angle with the street, α, this angle used to calculate W_s' and the radius of the first Fresnel zone for city, r_{FC}, and for water, r_{FW}.

C. DESCRIPTION OF GEOMETRIC PARAMETERS FOR ANN AND NFS

The methodology consists of regression of data collected in the measurement campaign, using the machine learning techniques, ANN and NFS. The purpose of the techniques is to predict the electric field intensity. Then, are used five input parameters that influence the propagation of the electromagnetic signal; these parameters are the following:

1) Distance between transmitter and receiver, D_{TR}, illustrated in Fig. 10.

The loss of propagation has a direct relationship with the distance between the transmitter and the receiver. As the distance increases, the signal strength decreases, both for the receiver in the city or on the river.

2) Electric field of direct ray.

The direct ray has the maximum electric field since there is a line of sight between transmitter and receiver. The equation of free space is the base of many radio propagation models as the case of log-normal path loss models using for urban environments [18] and seaport environments [26]. The equation of free space is the following:

$$E_{los}(R_{TR}) = \frac{\sqrt{30P_TG_T}}{R_{TR}\sqrt{L}}e^{-jR_{TR}} \tag{16}$$

where:

- P_T: Power transmission.
- G_T: Gain of transmitter antenna.
- L: Cable Loss.
- k: Propagating constant in free space [18].

3) Height of ground, h_G, illustrated in Fig.10.

The ground's height is related to the shadow effect caused by the city on the first reception points over the river. Then, the higher the ground is the higher attenuation for the reception points in the transition zone city-river. Furthermore, for a receiver over the river, the transmitter's height is higher than for a receiver over the city. As shown in [16], the variation of transmitter antenna height affects the received signal.

4) Distance only over the water, d_W, illustrated in Fig.10.

The distance over the water for a mixed path is essential because Millington [27] showed when the signal passes from a less conductivity medium to a more conductivity medium, signal experiments the recovery effect, that is, an increase of the signal over the water. Furthermore, in ITU-R 1546 [6], the distance only over the water is used to calculate the interpolation factor, and Okumura [5] uses the distance only over the water in the correction factor for mixed paths.

5) Radius of the first Fresnel ellipsoid for the city, r_{FC}, and for the river, r_{FW}.

The radius of the first Fresnel zone indicates the minimum distance between the propagated wave and the obstacle (building) to avoid diffraction. Since when 60% of the first Fresnel zone is free of obstacles, it is considered a line of sight between transmitter and receiver.

The radius of the first Fresnel zone for city, as in (12), considers d_1 the distance between the transmitter and the

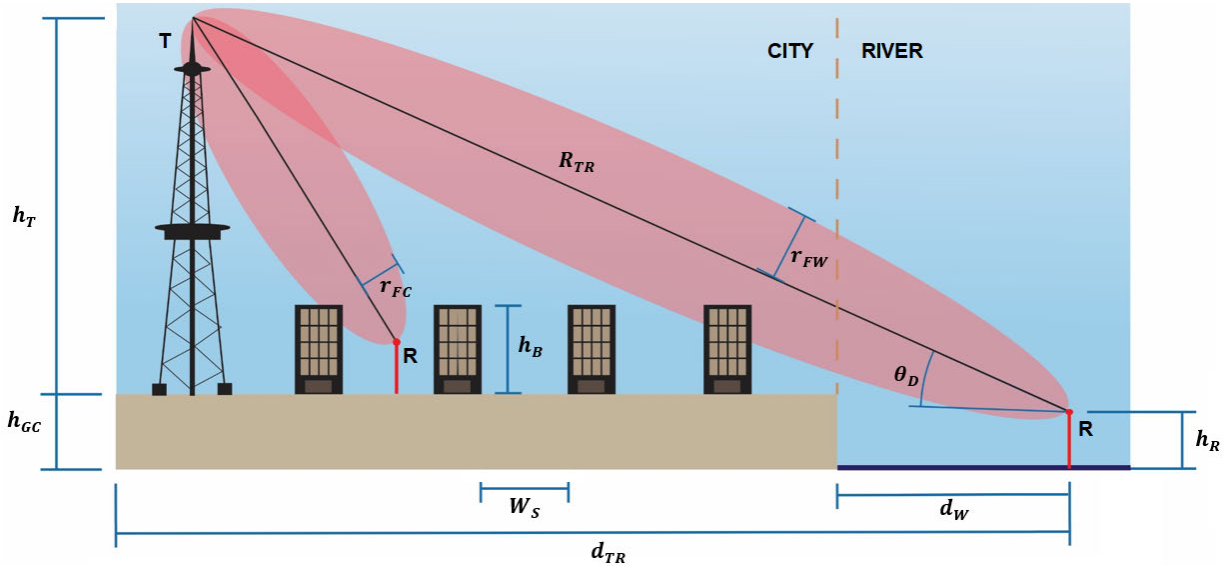

FIGURE 10. Mixed city-river scenario.

building, and d_2 the distance between the building and the receiver. The equations for Fresnel radius for the city, r_{FC}, are:

$$W_S' = \frac{W_S}{sin(\alpha)} \tag{17}$$

$$h_c = cos\theta_D \left(h_b - h_r - \left(\frac{W_s'}{2} \right) tan\theta_D \right) \tag{18}$$

$$r_{FC} = \sqrt{\frac{\lambda \left(d_{TR} - \frac{W_s'}{2} - h_c sin\theta_D \right) \left(\frac{W_s'}{2} + h_c sin\theta_D \right)}{d_{TR} cos\theta_D}} \tag{19}$$

where,

- W_s: Width of street.
- λ: Wavelength.

For the radius of the first Fresnel zone for water, d_1 is the distance between transmitter and the last building in the city, and d_2 is the distance between the building the receiver over the river. The equations for Fresnel radius over the river, r_{FW}, are:

$$h_w = cos\theta_D(h_B + h_G - h_R - d_W tan\theta_D) \tag{20}$$

$$r_{FW} = \sqrt{\frac{\lambda(d_{TR} - d_W - h_w sin\theta_D)(d_{w1} + h_w sin\theta_D)}{d_{TR} cos\theta_D}} \tag{21}$$

the city, the radius of Fresnel, presents similar values when the distance between the building and the receiver are similar. For the river, the first Fresnel zone's radius increases when the distance between the transmitter and the receiver located over the water increases.

D. DESCRIPTION OF MACHINE LEARNING TECHNIQUES IMPLEMENTED

For the development of the models, relevant parameters were taken into account, capable of describing the signal's attenuation in the mixed city-river path, and it is, therefore, possible to use it in other similar environments.

In order to perform ANN and ANFIS training, 420 samples were distributed in three sets of data randomly: 294 points in the training set, with 70% of the data about the total and the validation and test sets, which contained 63 points each, being 30% of the total samples, as shown in Table 1.

TABLE 1. Set of samples.

Set of Samples	Number of samples	Percentage
Training	294	70%
Validation	63	15%
Test	63	15%

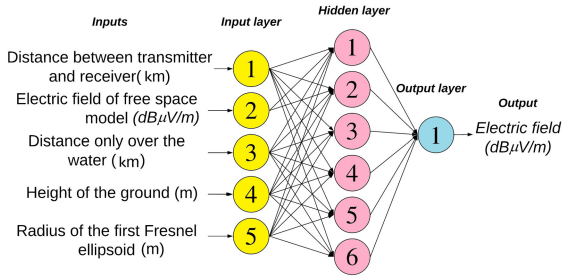

FIGURE 11. Generic multilayer perceptron neural network [19].

1) DESCRIPTION OF ANN IMPLEMENTED

An ANN of type Multi-Layer Perceptron (MLP), Fig. 11, was implemented in MATLAB©software. There are 5 neurons in the input layer (one for each model input), with linear activation function; 6 neurons in the hidden layer, with sigmoid activation function; and in the output layer with 1 neuron (the model output is the electric field) and linear activation function.

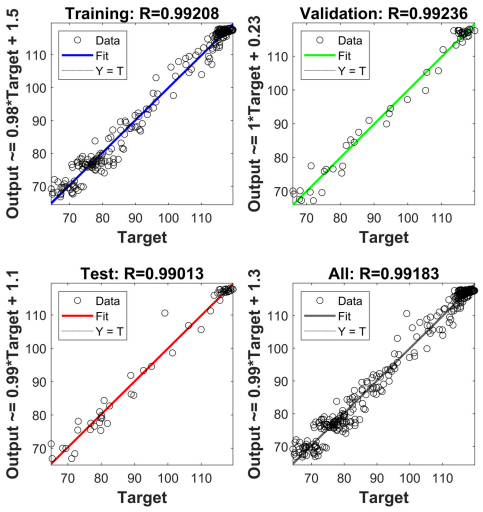

FIGURE 12. Regression of data generated by ANN and its respective targets.

This architecture was chosen through the trial and error strategy and the number of neurons in the hidden layer varied up to 100 and registered the lowest RMSE for 6 neurons. Only 1 hidden layer was used because the implementation uses little data and has few entries. For these types of implementations, just one layer finds the solution [20]. The supervised training algorithm for ANN was Levenberg-Marquardt.

Training is to present a pattern to the units in the input layer. Then the units calculate a response that is produced by the output layer. Then the value of the error is calculated and propagated from the output layer to the input layer. Additionally, the weights of the units of the hidden layers are being modified.

Thus, the error is progressively reduced to identify the learning stop point and obtain the best ANN generalization rate. An alternative is to use a stopping rule based on the cross-validation technique, which consists of validating the data during ANN training, ensuring no over-fitting in training [23].

From this data set were generated the weights that best represent the studied problem. Therefore, with a trained ANN and the error at a satisfactory level and with a good generalization capacity, it is possible to observe the linear regression between the outputs and the targets in Fig 12.

The dotted line in black means the best linear correlation coefficient between outputs and targets. The values of R are an indication of the correlation relationship between the outputs and the targets. If R = 1, it indicates that there is an exact linear relationship between outputs and targets. If R is close to zero, there is no linear relationship between outputs and targets. The topology presents good results showing that the ANN generated can be used as a prediction tool. The results will be discussed in the next section V.

Fig. 13 shows a histogram of the distribution of statistical errors for the ANN test set. Gaussian behavior is observed, which demonstrates the model's accuracy.

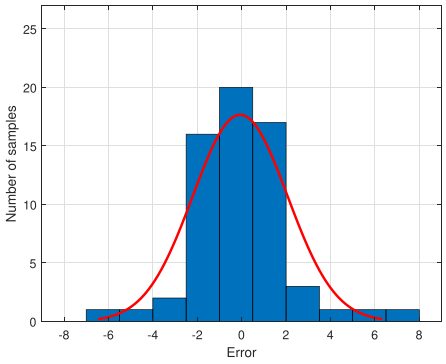

FIGURE 13. Histograms of the distribution of statistical errors (GMMP-NN).

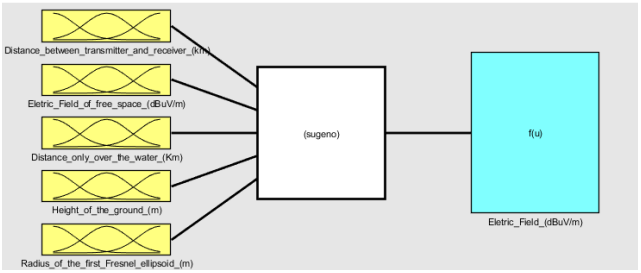

FIGURE 14. ANFIS structure implemented.

2) DESCRIPTION OF NFS IMPLEMENTED

The ANFIS system was developed with MATLAB© software using the ANFIS edit function. The configuration used was the Fuzzy model Sugeno and the defuzzification method weighted average.

The training method chosen for the neural network is a hybrid model, a combination of Least Squares (LS) and back-propagation with Downward Gradient (DG), which provides better accuracy for the system with few entries. Soon after, it was defined as stopping criteria for training, which are two, the desired error, and the number of iterations (300 times).

Other important items to be determined are the membership function and the type of clustering, which respectively were defined as Gaussian and subtractive cluster. The partition type of grouping via input space (subtractive cluster) specify the radius the influence of Gaussian function (value = 1.0), multiplying factor (1.5), acceptance radius in the group (value = 0.75) and rejection radius in the group (0.25). Below are the general characteristics of the implementation of ANFIS. The implemented ANFIS structure is presented in Fig.14.

From the ANN learning, the membership functions for each of the five inputs are generated. From the membership functions, it was generated knowledge in the form of rules governing the fuzzy inference system.

Fig. 15 shows a histogram of the distribution of statistical errors for the ANFIS test set. Gaussian behavior is observed, which demonstrates the model's accuracy.

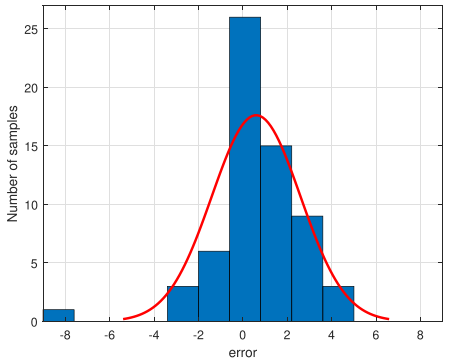

FIGURE 15. Histograms of the distribution of statistical errors (GMMP-NF).

V. RESULTS AND DISCUSSION

In this work, two radio propagation models based on machine learning (GMMP-NN and GMMP-NF) were proposed, which use geometric parameters obtained from the study environment as inputs. The analyzed scenario is of the mixed type, which is, formed by urban areas and rivers. The validation of the models is made with a test set consisting of random points measured from the environment is used and is not used in the algorithms' training phase. The results are shown in Figs. 16 to 21 and in table 2.

These figures compare the data predicted by the models with the data from the test set. Thus, in Figs. 16 and 17, the results of the urban route generated by GMMP-NN and GMMP-NF are presented respectively. In the same way, in Figs. 18 and 19, the results for the river route are presented for both models. Also, in Figs. 20 and 21, the predictions generated by the classic models (Okumura-Hata and ITU-R P.1546) are included.

The proposed models analyze the electric field level's prediction capacity in dBμV/m and are valid for a mixed city-river route of 0.5 km to 20 km in the city and from 1.77 km to 5 km in the river.

Table 2 presents the results for the three metrics of performance evaluation SD, RMSE, and GRG-MAPE. The SD indicates the dispersion of the data with the sample mean, that is, the model's precision. A low SD value means that the model is more accurate and that the data is grouped close to the average. In contrast, a high value indicates that the data is distributed over a larger range of values making the model inaccurate.

The second metric used is the RMSE, which quantifies how much the model's result differs from the expected real value. The model with the lowest RMSE value will be the model with the best accuracy. Finally, the third parameter is the GRG-MAPE, which describes the degree of correlation between two sets of data, which has values between 0 and 1. The closer to 1, the higher the value of the correlation between the predicted results and the expected data. The works developed in [8] and [28] consider the GRG-MAPE more accurate than the RMSE for evaluating propagation models. The equations for calculating the GRG-MAPE are described in [29].

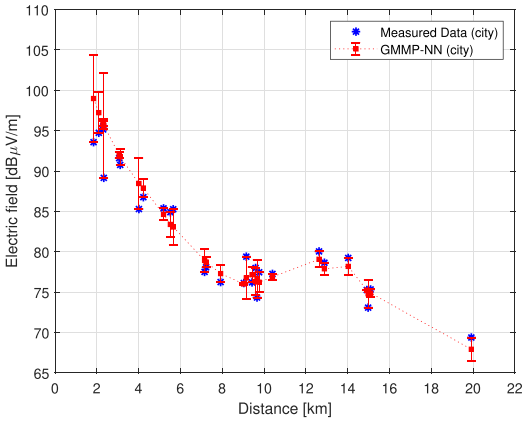

FIGURE 16. Comparison between measured and predicted data of GMMP-NN Model and their errors calculated point to point for city.

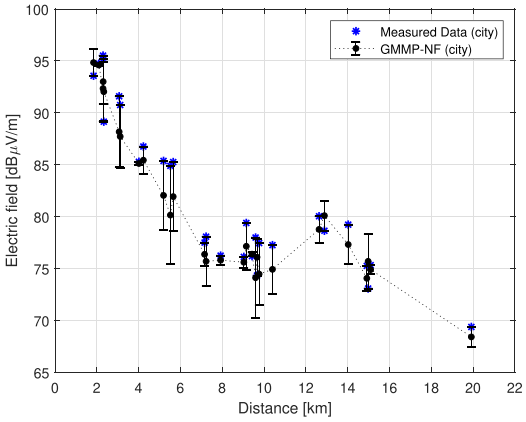

FIGURE 17. Comparison between measured and predicted data of GMMP-NF Model and their errors calculated point to point for city.

A. URBAN ENVIRONMENT

Fig. 16 shows the result of the GMMP-NN model (points in red) for different measured points of the city (points in blue). The result corresponds well to the measured data, and the values of the performance metrics are, according to Table 2, SD of 2.0790, RMSE of 2.09 dB, and GRG-MAPE of 0.9560.

In Fig. 17, the GMMP-NF model (black dots) are observed, and the predicted data have a reasonable correlation with expected data. For this scenario, SD is 1.9664, RMSE is 2.30 dB and GRG-MAPE 0.9427.

For the classic models, they are observed in Fig. 20, it is noted that the ITU-R P.1546 model behaves better than the Okumura-Hata model for distances up to 10 km. The Okumura-Hata model is based on measures developed in Tokyo-Japan, and the correction for the suburban environment was not suitable for the city of Belém-Pará. For distances over 10 km, the Okumura-Hata model performed better than the ITU-R P.1546 model.

The performance metrics for the Okumura-Hata model are SD of 6.2040, RMSE of 7.38 dB, and GRG-MAPE of 0.8957 while the ITU-R P.1546 model obtained SD of 6.2340, RMSE of 5.32 dB and GRG-MAPE of 0.9176.

TABLE 2. Comparison between models.

	ITU-R P.1546			Okumura-Hata			GMMP-NN			GMMP-NF		
Scenario	SD	RMSE (dB)	GRG-MAPE	SD	RMSE (dB)	GRG-MAPE	SD	RMSE (dB)	GRG-MAPE	SD	RMSE (dB)	GRG-MAPE
City	6.2340	5.32	0.9176	6.2040	7.38	0.8957	2.0790	2.09	0.9560	1.9664	2.30	0.9427
River	3.4982	9.15	0.9061	3.9338	16.42	0.8513	2.1470	2.13	0.9636	1.8737	1.84	0.9760
City-River	5.4775	8.38	0.8901	12.0746	12.59	0.8529	2.1346	2.11	0.9628	1.9967	2.06	0.9640

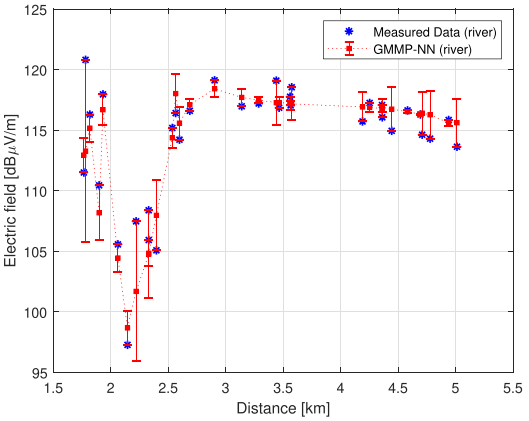

FIGURE 18. Comparison between measured and predicted data of GMMP-NN Model and their errors calculated point to point for river.

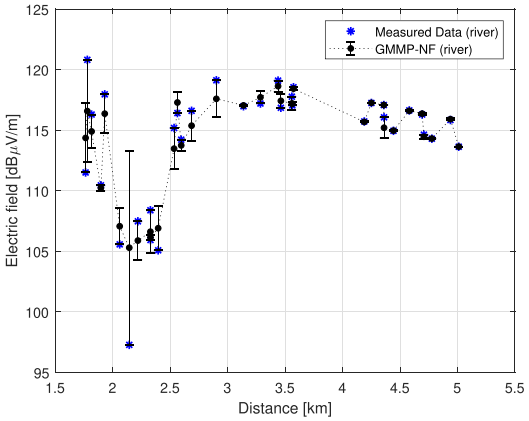

FIGURE 19. Comparison between measured and predicted data of GMMP-NF Model and their errors calculated point to point for river.

Thus, it is observed that for the data on the urban route, the GMMP-NN model obtained a better performance because it managed to extract more knowledge from the test set than the GMMP-NF model. However, both models perform better than the classic models and can be applied to the measured data. Also, the points predicted by the proposed models are consistent with the data measured on the urban route, even after 10 km, where the signal has an increase in amplitude, this is due to the height of the ground, which is higher than in other places where the other measurements were made. In addition, the buildings are predominantly made up of houses with a maximum of three floors.

B. RIVER ENVIRONMENT

As seen in Fig. 18, the GMMP-NN model can be applied to the data measured on the river route, and the following performance metrics were obtained, SD of 2.1470, RMSE of 2.13 dB, and GRG-MAPE of 0.9636.

In Fig. 19, the GMMP-NF model corresponds well to the measured data and SD performance metrics of 1.8737, RMSE of 1.84 dB, and the GRG-MAPE of 0.9640. It is noted that for the river route, the GMMP-NF model showed better performance than the GMMP-NN. However, both models have superior performance compared to classic models.

In Fig. 21, the Okumura-Hata model and the ITU-R P.1546 model have lower values than the measured data, and the ITU-R P.1546 model obtained better results than the Okumura-Hata model, as this model was developed with data obtained from measurements that contemplate the sea environment. However, the ITU-R P.1546 model obtained the RMSE at around 9 dB, a similar error observed in work developed in [7], which was 10 dB for an environment similar to that used in this work.

The Okumura-Hata model's performance metrics on the river route were: SD of 3.9338, RMSE of 16.42 dB, and GRG-MAPE of 0.8513, while for the ITU-R P.1546 model, they were: SD of 3.4982, RMSE of 9.15 dB, and GRG-MAPE of 0.9061.

C. CITY-RIVER ENVIRONMENT

In Fig. 21, from 2 km to 2.5 km, the signal presents an attenuation, which is caused by the transition from the city to the river, in addition to the effect of recovering the signal amplitude, observed by Millington [27], which is when the signal moves from a medium with less conductivity to a medium with greater conductivity. These effects are not predicted by classic models in the literature, which causes significant errors in prediction. In contrast, the proposed models can contemplate and predict these variations with minor errors.

Therefore, for a mixed city-river route, the proposed models represent the general trend of the data measured with high precision, as shown in Figs. 20 and 21 and performance metrics. For the GMMP-NN model, the standard deviation of 2.1346, RMSE of 2.11 dB, and GRG-MAPE of 0.9628 were obtained. For the GMMP-NF, the standard deviation of 1.9967, RMSE is 2.06 dB, and GRG-MAPE of 0.9640. It is noted that the GMMP-NF has a better overall behavior than the GMMP-NN for the mixed city-river route, this is due to the structure of the ANFIS, which is characterized as a hybrid system, as it adds the advantages of networks neural

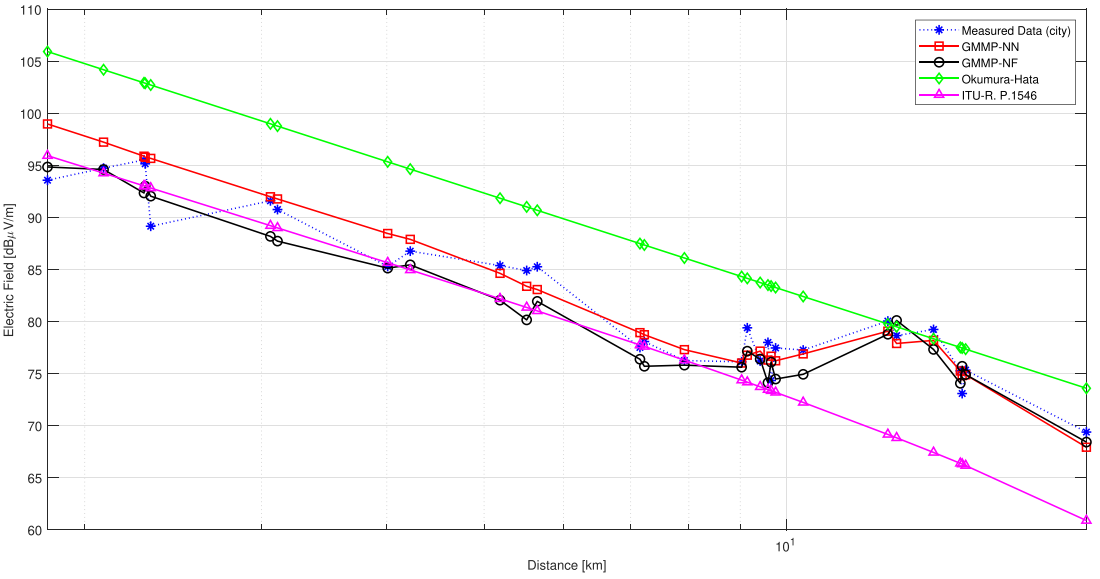

FIGURE 20. Results for data in the city.

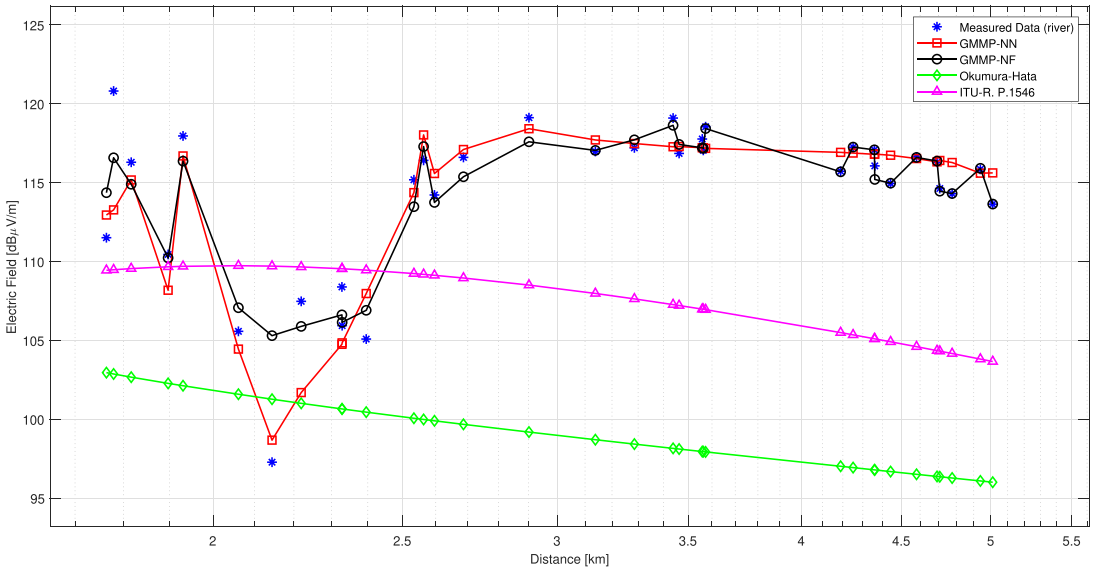

FIGURE 21. Results for data in the river.

and fuzzy logic, as well as reducing the disadvantages of both techniques. However, the GMMP-NN and GMMP-NF models predict the electric field satisfactorily, with values close to the measured data.

The classic models for the mixed city-river route, on the other hand, obtained inferior performances and cannot be used for this type of environment. Performance metrics for the Okumura-Hata model, SD is 12.0746, RMSE is 12.59 dB, and GRG-MAPE is 0.852. While for the ITU-R P.1546 model, the SD is 5.4775, RMSE is 8.38 dB, and GRG-MAPE is 0.8901.

VI. CONCLUSION

Dense forests and wide rivers characterize the Amazon region. The proposal to use geometric parameters to develop radio propagation models based on machine learning techniques, such as the Artificial Neural Network and the Neuro-Fuzzy System, was efficient and promising.

Propagation models (GMMP-NN and GMMP-NF) for mixed city-river environments can be used to plan or improve coverage of radio telecommunication systems for the UHF band. The data were obtained from a campaign to measure the signal's electric field at 521 MHz, held in Belém city, in the Amazon, Brazil.

The approach used five input parameters obtained from the geometry of the city-river scenario: the distance between the transmitter and the receiver, the radius of Fresnel's first ellipsoid, the electric field for free space, the distance only over the river and the elevation of the terrain.

Performance assessments were made using three metrics: SD, RMSE, and GRG-MAPE. The proposed models achieved

higher performance to the classic models and are suitable for predicting the electric field in a mixed city-river route in the Amazon region.

For future work, it is intended to expand the relevance of modeling using different geometric parameters to train the GMMP-NN and GMMP-NF models, with more heterogeneous data from different environments and radio frequencies, making the model more general, in addition to the possibility of using other artificial intelligence.

ACKNOWLEDGMENT

The authors would like to thank PROPESP/UFPA, CAPES, CNPq, and the National Institute of Science and Technology - Wireless (INCT-CSF) to support the research described in this article.

REFERENCES

[1] D. C. dos Reis, M. E. C. Araújo, S. S. L. dos Santos, S. S. D. C. Silva, and F. A. R. Pontes, "Araraiana e combú: Um estudo comparativo de dois contextos ribeirinhos amazônicos," *Temas em Psicologia*, vol. 20, no. 2, pp. 429–438, 2012, doi: 10.9788/tp2012.2-11.

[2] T. K. Sarkar, M. N. Abdallah, and M. Salazar-Palma, "Survey of available experimental data of radio wave propagation for wireless transmission," *IEEE Trans. Antennas Propag.*, vol. 66, no. 12, pp. 6665–6672, Dec. 2018, doi: 10.1109/TAP.2018.2878108.

[3] G. Apaydin and L. Sevgi, "FEM-based surface wave multimixed-path propagator and path loss predictions," *IEEE Antennas Wireless Propag. Lett.*, vol. 8, pp. 1010–1013, 2009, doi: 10.1109/LAWP.2009.2030966.

[4] C. C. Teague, P. M. Lilleboe, and D. E. Barrick, "Estimation of HF radar mixed-media path loss using the millington method," *Waves Turbulence Meas. (CWTM)*, vol. 11, pp. 1–6, Mar. 2015, doi: 10.1109/CWTM.2015.7098111.

[5] Y. Okumura, E. Ohmori, T. Kawano, and K. Fukuda, "Field strength and its variability in VHF and UHF land-mobile radio service," *Rev. Elect. Commun. Lab.*, vol. 16, nos. 9–10, pp. 825–873, 1968.

[6] *Method for Point-to-Area Predictions for Terrestrial Services in the Frequency Range 30 MHz to 3000 MHz*, document Recomendation ITU-R P.1546, 2013.

[7] B. A. Witvliet, P. W. Wijninga, E. van Maanen, B. Smith, M. J. Bentum, R. Schiphorst, and C. H. Slump, "Mixed-path trans-horizon UHF measurements for p.1546 propagation model verification," in *Proc. APWC*, Sep. 2011, pp. 303–306, doi: 10.1109/apwc.2011.6046762.

[8] J. Yu, W. Chen, K. Yang, C. Li, F. Li, and Y. Shui, "Path loss channel model for inland river radio propagation at 1.4 GHz," *Int. J. Antennas Propag.*, vol. 2017, pp. 1–15, Jan. 2017, doi: 10.1155/2017/5853724.

[9] T. A. Benmus, R. Abboud, and M. K. Shatter, "Neural network approach to model the propagation path loss for great tripoli area at 900, 1800, and 2100 MHz bands," in *Proc. 16th Int. Conf. Sci. Techn. Autom. Control Comput. Eng. (STA)*, Dec. 2015, pp. 793–798, doi: 10.1109/STA.2015.7505236.

[10] S. I. Popoola, A. Jefia, A. A. Atayero, O. Kingsley, N. Faruk, O. F. Oseni, and R. O. Abolade, "Determination of neural network parameters for path loss prediction in very high frequency wireless channel," *IEEE Access*, vol. 7, pp. 150462–150483, 2019, doi: 10.1109/ACCESS.2019.2947009.

[11] 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, Jul. 2017, doi: 10.1109/TAP.2017.2705112.

[12] T. da Silva, A. Costa, D. da Silva, L. Castro, J. Araujo, and G. Cavalcante, "Radio propagation for the amazon region considering the river level," in *Proc. Workshop Commun. Netw. Power Syst. (WCNPS)*, Oct. 2019, pp. 1–4, doi: 10.1109/WCNPS.2019.8896281.

[13] D. K. Nakata da Silva, L. E. Castro Eras, A. A. Moreira, L. M. Correia, F. J. Brito Barros, and G. Protasio dos Santos Cavalcante, "A propagation model for mixed paths using dyadic green's functions: A case study over the river for a city–river–forest path," *IEEE Antennas Wireless Propag. Lett.*, vol. 17, no. 12, pp. 2364–2368, Dec. 2018, doi: 10.1109/lawp.2018.2875333.

[14] L. E. C. Eras, D. K. N. D. Silva, F. B. Barros, L. M. Correia, and G. P. S. Cavalcante, "A radio propagation model for mixed paths in amazon environments for the UHF band," *Wireless Commun. Mobile Comput.*, vol. 2018, pp. 1–15, Nov. 2018, doi: 10.1155/2018/2850830.

[15] N. Faruk, S. I. Popoola, N. T. Surajudeen-Bakinde, A. A. Oloyede, A. Abdulkarim, L. A. Olawoyin, M. Ali, C. T. Calafate, and A. A. Atayero, "Path loss predictions in the VHF and UHF bands within urban environments: Experimental investigation of empirical, heuristics and geospatial models," *IEEE Access*, vol. 7, pp. 77293–77307, Jun. 2019, doi: 10.1109/ACCESS.2019.2921411.

[16] F. Ikegami, S. Yoshida, T. Takeuchi, and M. Umehira, "Propagation factors controlling mean field strength on urban streets," *IEEE Trans. Antennas Propag.*, vol. AP-32, no. 8, pp. 822–829, Aug. 1984, doi: 10.1109/TAP.1984.1143419.

[17] M. Hata, "Empirical formula for propagation loss in land mobile radio services," *IEEE Trans. Veh. Technol.*, vol. 29, no. 3, pp. 317–325, Aug. 1980, doi: 10.1109/T-VT.1980.23859.

[18] T. S. Rappaport, *Wireless Communications: Principles and Practice*, 2nd ed. Upper Saddle River, NJ, USA: Prentice-Hall, 2009, pp. 72–100.

[19] I. N. Silva, D. H. Spatti, and R. A. Flauzino, *Artificial Neural Networks for Engineering and Applied Sciences—Practical Course*. São Paul, Brazil: Artliber, 2010, ch. 5, sec. 5.2, pp. 91–94.

[20] S. Haykin, *Neural Networks: A Comprehensive Foundation*, 2nd ed. Upper Saddle River, NJ, USA: Prentice-Hall, 1999, ch. 4, sec. 4.1, pp. 178–180.

[21] J.-S. R. Jang, "ANFIS: Adaptive-network-based fuzzy inference system," *IEEE Trans. Syst., Man, Cybern.*, vol. 23, no. 3, pp. 665–685, Jun. 1993.

[22] H. A. O. Cruz, R. N. A. Nascimento, J. P. L. Araujo, E. G. Pelaes, and G. P. S. Cavalcante, "Methodologies for path loss prediction in LTE-1.8 GHz networks using neuro-fuzzy and ANN," in *IEEE MTT-S Int. Microw. Symp. Dig.*, Aug. 2017, pp. 1–5, doi: 10.1109/IMOC.2017.8121127.

[23] *IBGE*. Accessed: May 9, 2018. [Online]. Available: https://cidades.ibge.gov.br/brasil/pa/belem/panorama

[24] M. K. N. R. Nitin Chopde, "Landmark based shortest path detection by using a* and haversine formula," *Int. J. Innov. Res. Comput. Commun. Eng.*, vol. 1, pp. 298–302, Apr. 2013.

[25] J. M. Hernando, *Radio Transmission*, 7th ed. Madrid, Spain: Univ. Editorial Ramón Areces (in Spanish), 2003, pp. 115–607.

[26] J. C. Reyes-Guerrero and L. A. Mariscal, "5.8 GHz propagation of low-height wireless links in sea port scenario," *Electron. Lett.*, vol. 50, no. 9, pp. 710–712, Apr. 2014, doi: 10.1049/el.2014.0098.

[27] G. Millington, "Ground-wave propagation over an inhomogeneous smooth Earth," *Proc. IEE III, Radio Commun. Eng.*, vol. 96, no. 39, pp. 53–64, Jan. 1949, doi: 10.1049/pi-3.1949.0013.

[28] N. A. Pérez García, A. D. Pinto, J. M. Torres, J. E. Rengel, L. M. Rujano, N. Robles Camargo, and Y. Donoso, "Improved ITU-R model for digital terrestrial television propagation path loss prediction," *Electron. Lett.*, vol. 53, no. 13, pp. 832–834, Jun. 2017, doi: 10.1049/el.2017.1033.

[29] J. Yu, C. Li, K. Yang, and W. Chen, "GRG-MAPE and PCC-MAPE based on uncertainty-mathematical theory for path-loss model selection," in *Proc. VTC Spring*, Nanjing, China, May 2016, pp. 1–5, doi: 10.1109/VTCSpring.2016.7504265.

ALLAN DOS S. BRAGA received the B.Sc. degree in computer engineering and the master of science degree in electrical engineering from the Federal University of Pará, in 2013 and 2015, respectively, where he is currently pursuing the Ph.D. degree in advanced electromagnetism. His current research interests include telecommunications systems, mobile systems, computer simulators, radio propagation, digital TV and radio systems, computational environment, computational intelligence, and video quality.

HUGO A. O. DA CRUZ received the B.Sc. degree in computer engineering and the master of science degree in electrical engineering from the Federal University of Pará, in 2016 and 2018, respectively, where he is currently pursuing the Ph.D. degree in applied computing. His current research interests include computational intelligence, propagation, and video quality.

LESLYE E. C. ERAS received the B.Sc. degree in electronics and telecommunications from the Particular Technical University of Loja, Ecuador, in 2013, and the master's and doctor of science degrees in electrical engineering from the Federal University of Pará, in 2016 and 2019, respectively. She is currently a Professor with the Federal University of South and Southeast of Pará. Her research interests include radio propagation for outdoor environments, electromagnetism and automation, and control.

JASMINE P. L. ARAÚJO received the B.Sc., master's, and doctor of science degrees in electrical engineering from the Federal University of Pará, in 1994, 2002, and 2011, respectively, and the Ph.D. degree from INESC-TEC, Porto, in 2015. She is currently an Adjunct Professor with the Federal University of Pará. Her current research interests include electrical engineering, especially in the areas of telecommunications and applied computing. She is also working on the mobile networks, propagation, electromagnetic compatibility, and bio-inspired algorithms.

MIÉRCIO C. A. NETO received the B.Sc. degree in telecommunications engineering from the Institute of Higher Studies of the Amazon (IESAM), in 2010, and the master's, doctor's, and Ph.D. degrees in electrical engineering with emphasis on telecommunications from the Postgraduate Program in Electrical Engineering (PPGEE / Institute of Technology (ITEC) / Federal University of Pará (UFPA)), in 2013, 2015, and 2018, respectively. He is currently an Adjunct Professor, under Exclusive Dedication (DE), at UFPA, crowded to ITEC and linked to the Faculty of Electrical and Biomedical Engineering (FEEB), also serving as the Director of FEEB. His research interests include bioinspired computing (BIC) - heuristic and meta-heuristic computational optimization techniques, digital signal processing, and electromagnetic theory with applications in the design of resonant devices, such as antennas and frequency selective surfaces (FSS).

DIEGO K. N. SILVA received the B.Sc. degree in computer engineering and the master's and doctor of science degrees in electrical engineering from the Federal University of Pará, in 2011, 2014, and 2018, respectively. He is currently an Adjunct Professor with the Federal University of South and Southeast of Pará. His current research interests include radio propagation for environments with mixed paths, electromagnetism, and mobile networks.

GERVÁSIO P. S. CAVALCANTE received the M.S. degree in electronic engineering from the Federal University of Paraíba, Campina Grande, Brazil, and the Ph.D. degree in telecommunications engineering from UNICAMP - Estate University of Campinas, Brazil, in 1982. From 1974 to 2000, he was a Titular Professor with the Department of Electrical Engineering. Since 2001, he has been a Titular Professor with the Faculty of Computation and Telecommunication, Federal University of Pará, Brazil. His research interests include the development of models for wire antennas in communications, and the frequency selective surface theory using bio-inspired multi-objective algorithms.