Hindawi
International Journal of Antennas and Propagation
Volume 2020, Article ID 8494185, 12 pages
https://doi.org/10.1155/2020/8494185

Research Article

Application of Hybrid ARIMA and Artificial Neural Network Modelling for Electromagnetic Propagation: An Alternative to the Least Squares Method and ITU Recommendation P.1546-5 for Amazon Urbanized Cities

Ramz L. Fraiha Lopes, Simone G. C. Fraiha, Herminio S. Gomes, Vinicius D. Lima, and Gervasio P. S. Cavalcante

Post Graduation Program in Eletrical Engineering (PPGEE), Federal University of Pará (UFPA), Belém 66075-110, Brazil

Correspondence should be addressed to Ramz L. Fraiha Lopes; fraiha.ramz@gmail.com

Received 13 November 2019; Revised 27 January 2020; Accepted 25 February 2020; Published 18 March 2020

Academic Editor: Lorenzo Luini

This study sets out an empirical hybrid autoregressive integrated moving average (ARIMA) and artificial neural network (ANN) model designed to estimate electromagnetic wave propagation in densely forested urban areas. Received signal power intensity data was acquired through measurement campaigns carried out in the Metropolitan Area of Belém (MAB), in the Brazilian Amazon. Comparisons were made between estimates from classical least squares (LS) fitting and ITU (International Telecommunication Union) recommendation P. 1546-5. The results indicate the model is, at least, 44% more precise than every ITU estimate and, in some situations, is at least 11% better than an LS estimate, depending on the respective values of the relative error (RE).

1. Introduction

This study examines a hybrid ARIMA-ANN model inspired by [1] a model to predict received signal power intensity at a receiver (Rx) location as a function of the distance to the transmitter (Tx). This study is based on the Brazilian digital television (DTV) frequency range and looks at the special case of a densely forested and urbanized city in the Amazon region.

Television (TV) still is one of the most significant means of communication and, in view of this, is of crucial importance as a source of entertainment and information. Since it includes DTV transmission, which operates in a different frequency range from analogic TV transmission, a performance analysis of received signal power is required for both frequency ranges. Nonetheless, it should be taken into account that there is a scarcity of wave propagation models in the literature adapted for towns and cities in the Amazon region or those near the Equator line. Weather itself is a key factor in the effectiveness of telecommunication services in this kind of region, as shown in [2].

Some other studies related to what we propose in this work can be seen in [3–11]. In the study by Liangping and Sternberg, two approaches are proposed to predict the Peak Signal-to-Noise Ratio (PSNR) in video transmissions. Both rely on time series modelling and both can achieve satisfactory results, compared with the performance of the usual mean or median algorithms. The work in [4] shows an ARIMA model used to address an electromagnetic propagation problem. It is one of the few works that use this type of modelling to tackle a problem of this kind. In [5], a hybrid ARIMA-ANN (where the ANN works as a generalized regressor) is proposed to predict the incidence of hepatitis in Heng County, China. The results were compared with the single ARIMA and single ANN estimates. The authors of [6] propose a hybrid SARIMA (Seasonal ARIMA) and nonlinear autoregressive neural network (NARNN) for forecasting the incidence of hand-foot-and-mouth disease in Chenzen, China.

In study [7], the authors propose a hybrid ARIMA and support vector machines (SVM) neural networks for forecasting stock prices. In [8], the authors propose a technique for time series forecasting where models from state space (ETS) modelling for exponential smoothing are combined with a neural network. The aim is to enable the authors to obtain different combinations of linear or nonlinear patterns in a time series more easily. Comparisons were made between a single ARIMA, a single ETS, a multilayer perceptron neural network, and some ARIMA-ANN, and the planned modelling achieved good results.

The authors of [9] put forward a hybrid ARIMA-ANN model which, before being fitted, takes note of the volatility of the studied series. The results obtained outperform those of the ARIMA, ANN, and ARIMA-ANN models. The work in [10] devises a hybrid evolutionary system comprising a simple exponential filter for smoothing, ARIMA, autoregressive (AR) linear models and a support vector regression (SVR) model. The authors employ a particle swarm optimization method to select the order of the AR model, SVR parameters, and the number of lags in the time series. The authors claim their results are promising in the domain of forecasting. Finally, the study includes a review of various hybrid modelling techniques applied to time series forecasting [11].

The studies outlined above show the wide range of applications of both time series models, neural networks and hybrid approaches. However, only one of these works directly tackles the problem of electromagnetic propagation modelling by means of any kind of time series models.

This work aims to illustrate an alternative strategy for addressing electromagnetic propagation problems to achieve satisfactorily results. The results of this work show the feasibility of the proposed model. Comparisons with classical LS fitting and ITU recommendation P. 1546-5, which treats on wave propagation in frequencies from 30 MHz to 3 GHz, were performed, using relative error (RE) and root-mean-square error (RMSE) errors values as benchmarks.

2. The Proposed Model

Time series is a sequence of observations taken sequentially in time [12], or, in other words, an outcome of a stochastic process. An intrinsic feature of a time series is that, typically, adjacent observations are dependent. The same concept can be extended to any kind of observations that follow a sequential pattern, not necessarily in time. One example is the datasets for the predictions in this study, where the "distance" variable is used to replace the "time" variable.

Models of the ARIMA type are linear. This means that they are able to give a satisfactory description of a series in which the main information is represented in linear terms. However, there are some limitations to the range of problems that can be tackled using ARIMA models. One way to get around this problem is to use a hybrid modelling technique, as in this study. The hybrid model proposed here was influenced by [1] and consists of a hybrid ARIMA-ANN technique.

Basically, we use the ARIMA model to make a first adjustment on the analysed series to represent its linear information. Then, we adjust the residuals of the ARIMA fitting with the nonlinear technique (in this work, it is a generalized regressor neural network). The necessary calculations and programs were carried out on MATLAB software, by means of internal functions, both for ARIMA and the ANN.

Owing to the empirical nature of the proposed model, data regarding scenarios different than the one studied in this work are crucial to generalize the possibilities of applications for this modelling.

3. Measurement Campaign

The data used in this work was acquired through measurement campaigns in the surrounding area of the city of Belém (north of Brazil). Data were obtained from a single transmitter, which operates in the frequency range of 518.14 to 524.14 MHz. These points were divided into three groups called radials, namely, radial 1 (angle of 30°), radial 2 (angle of 45°), and radial 3 (angle of 80°). They are shown in Figure 1. The first point of each radial is located at a minimum distance of 1 km from the origin.

The DTV transmitter is located at a height of 114.58 m from the ground at a central, inhabited, and urbanized neighbourhood of Belém. Receiver antenna was situated at a height of 1.5 m from the ground, on the roof of a car (in order to simulate the scenario of a DTV service user), properly isolated from its body. Measurements were carried out in the morning, when there was clear weather and the temperature was approx. 30°C. Traffic was normal as well, that is, there were no traffic jams around the measurement points.

4. Data Handling

The results of this study were obtained by following the series of steps shown in Figure 2. As shown in the diagram, there is an interpolation branch in the testing process. We did this for two reasons: first, to increase the number of samples for each measured dataset, which allows the ARIMA model to work with more samples and, thus, refine its adjustments. We used a shape-preserving piecewise cubic interpolation (here abbreviated as SPPCI) to increase the number of samples of each dataset to 200 (two hundred).

In addition, the interpolated group of datasets is able to simulate a "no stop" measurement campaign scenario, which is usually more desirable than a "stop-and-go" campaign scenario, where it is necessary to stop at every measured point to acquire data. Our measurement campaigns were of the "stop-and-go" type. Since there are no stops on a "no stop" campaign, ideally, the receiver antenna operates at a constant speed through the measured radial, where it is continuously acquiring data. This is a desirable measurement scenario, since it is faster and, usually, less expensive than a "stop-and-go" measurement campaign. In this type of measurement scenario, the number of samples acquired is naturally higher, since the receiver is always acquiring information.

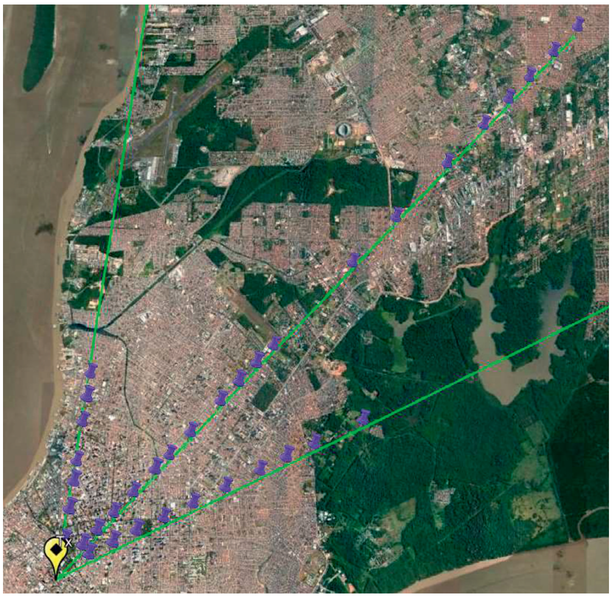


Figure 1: Measured points. Radials 1, 2, and 3 (adapted from [2]).

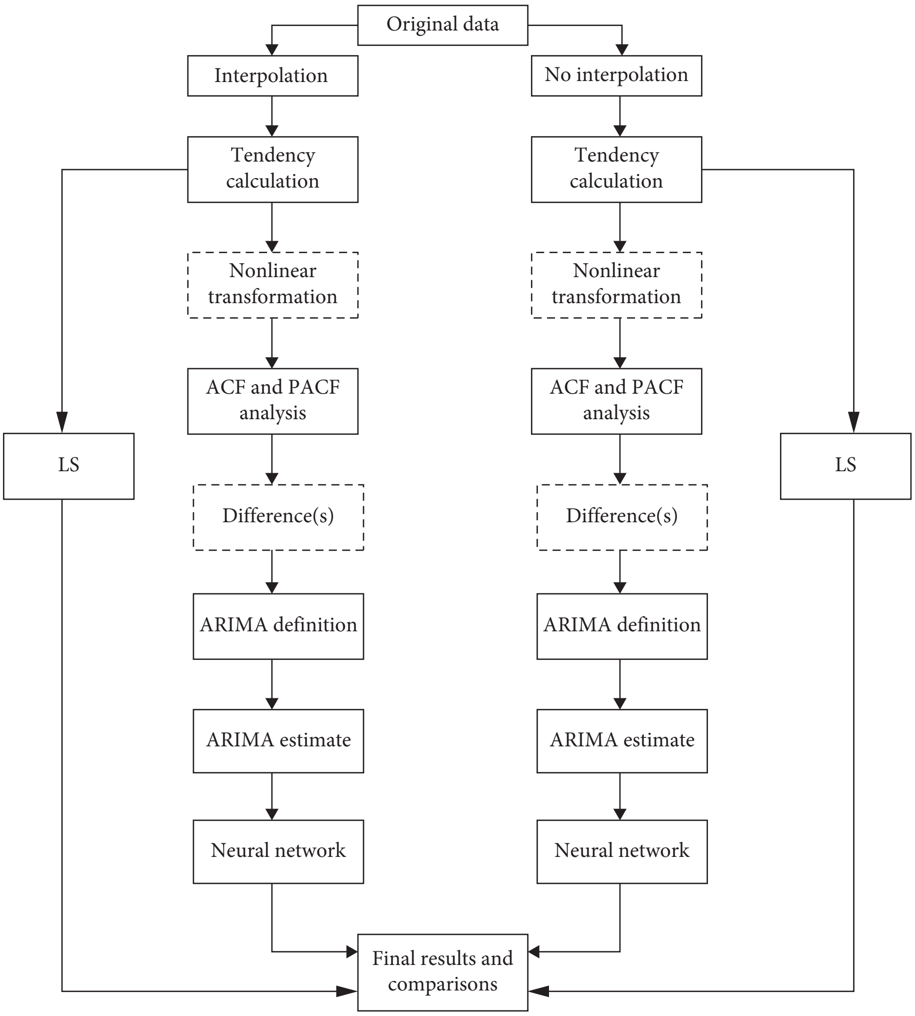


Figure 2: All the steps followed in the fitting process. Optional steps in dashed lines.

In this work, we divided the procedures in two groups: "original datasets," whose number of samples for each series are not increased, and "interpolated datasets," that are the interpolated versions of the original series. They will be addressed in this manner from now on when exposing results and making comparisons.

The two groups of datasets (original and interpolated) undergo the same procedures, in order to obtain the results. In addition, after analysing the studied datasets, we decided to isolate the trend in every dataset before both the LS and ARIMA fitting. In the case of the ARIMA fitting, it is a predicted measure to make the studied series stationary [12]. With regard to the LS fitting, we proceeded in this way as well so that a fair comparison could be made with the proposed modelling. We calculated a linear tendency for each dataset, carried out the adjustments for the series and, before comparing the results, these tendencies were reintegrated to the estimated curves.

5. ARIMA Fitting Methodology

The ARIMA fitting methodology was based on the autocorrelation function (ACF) and partial autocorrelation function (PACF) analysis of the studied series [12]. It should be noted that, when using the ARIMA model, the (usual) variable "time" is replaced with the "distance" variable. In other words, it is assumed that the received signal power intensity in one point depends on the previous values, according to the chosen metric (in this case, the distance to from Rx to Tx).

Before an ARIMA model can be fitted, the series that needs to be adjusted must be stationary. Manipulations of the series such as nonlinear transformations (e.g., logarithmic transformations), differences and attempts to isolate its tendency are the usual ways of turning a nonstationary series into a stationary series [12]. After ensuring that the analysed series is stationary, we proceed to an analysis of ACF and PACF. When these functions behave like that of a stationary process, we can define the order of the ARIMA model [12]. As indicated by the diagram in Figure 2, the "nonlinear transformation" and the "differences" steps in the series are optional steps for the data of this work. When analysing other datasets or fitting this modelling on another problem, these steps may become mandatory.

6. LS Fitting Methodology

Aiming a fair comparison between the proposed hybrid model and the LS method, we chose to represent the studied scenario, an equation similar as the one of an ARIMA model, that is, a recursive polynomial. In this work, we chose a second-order polynomial when applying the LS method (see equation (1)).

$$f_h = a_1 f_{k-1} + a_2 f_{k-2} + a_3. \tag{1}$$

The coefficients a_i, $i = 1, 2, 3$, of equation (1) were determined by using an LS method solved by means of a Levenberg–Marquardt [13] algorithm. However, the LS method has some limitations, especially if the analysed dataset contains a large number of samples. In these situations, the LS methods may not be able to find, directly, an optimal solution (or may take a long time finding it), owing to the huge size of the search area. As a means of overcoming these problems, the authors recommend fitting an ARIMA (or the proposed hybrid ARIMA-ANN) model as an alternative to the LS method (it is the objective of this work, after all).

7. Neural Network Fitting Methodology

When refining the results obtained from the ARIMA model, it is possible to complement the ARIMA adjustment with a nonlinear methodology (in this case, an ANN) to fit the nonlinear part of the datasets, which are not fitted in an ARIMA model. When complemented with the ARIMA fitting, we call it a "combined model" (CM).

In this study, we employ a radial basis function (RBF) ANN with two layers with a Gaussian activation function. This ANN works as a generalized regressor. A theoretical diagram of a generalized regressor is shown in Figure 3. The neurons of the first layer make an element-wise product between the biases and the weights and each neuron correspond to a training point. The neurons of the second layer normalize the values previously found (see MATLAB documentation on *newgrnn* neural network [14]). In the original training dataset, eight of the twenty-five original samples were used to train the network (as in Figure 4). With regard to the interpolated datasets, we used 24 of the 200 available samples. We proceeded in this way to avoid overfitting the ANN, since its adjustment must be used in other datasets as well. The boundaries and the central samples are always used as fitting points. The other points are chosen at random. We used 1 as the spread value of the neural network. The output of the network is, thus, interpolated (SCCIP) to ensure that the final output vector has the same number of elements as the measured data and the ARIMA vector.

In Figure 3, $h(x)$ is the activation function, w are the weights, x are the inputs, n is the number of inputs, and $f(x)$ is the exit function. The diagram of the architecture of the ANN used in the original datasets fitting is shown in Figure 4.

8. Results

The results are divided into two groups, depending on what type of dataset was used (whether original or interpolated). The best results were obtained by using the "radial 2" dataset as a training set. The "radial 1" and "radial 3" datasets were used for purposes of comparison.

8.1. Least Squares Fitting: Original Datasets.

The LS fitting on the problem originated from equation (1), for the "radial 2" dataset, resulted in $a_1 = 0.2893$, $a_2 = 0.4927$ and $\alpha_3 = -19.5619$. The Euclidean norm of residuals was 17.9024. The graphs with the LS estimated curves are shown along with other results of this work in Figure 5. Table 1

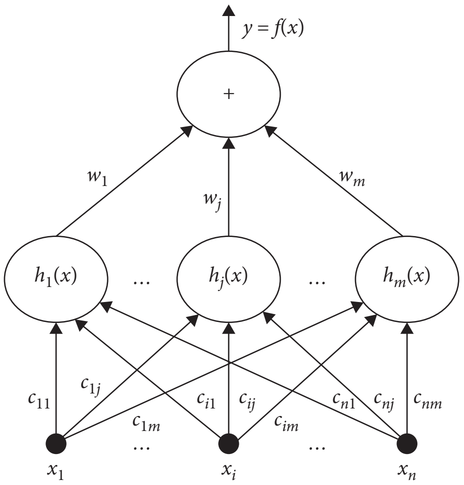


FIGURE 3: Radial basis neural network general diagram.

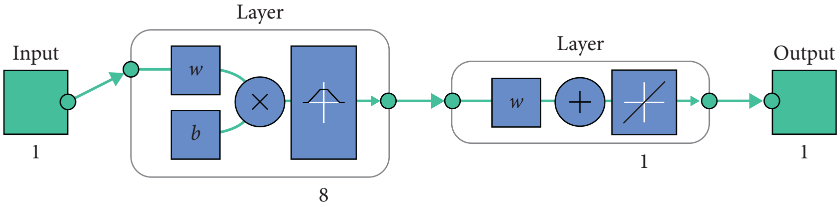


FIGURE 4: Neural network diagram for original training dataset [14].

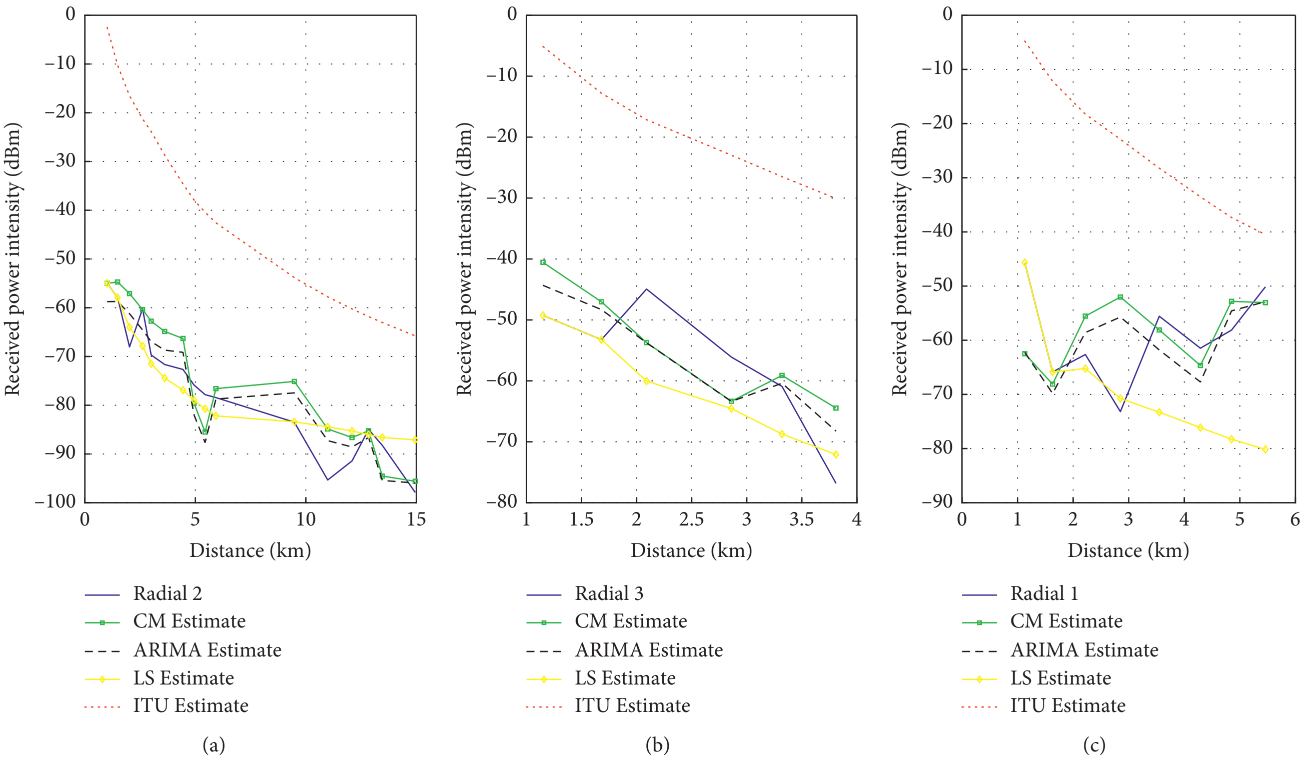


FIGURE 5: Graphs of the combined ARIMA model adjust for radial 2 (a) and estimates for radial 3 (b) and 1 (c).

Table 1: RE and RMSE values for original datasets.

	Radial 2		Radial 3		Radial 1	
	RE	RMSE	RE	RMSE	RE	RMSE
ARIMA	0.0459	5.0659	0.0484	3.7714	0.1391	8.8801
CM	0.0469	5.2547	0.0713	5.5169	0.1593	10.232
LS	0.0474	5.0775	0.1127	7.9867	0.1971	15.199
ITU	0.5237	38.882	0.6731	38.346	0.5753	37.256

shows the relative and RMS errors for the LS fitting in the three radials.

8.2. Least Squares Fitting: Interpolated Datasets. By analogy, in the case of the interpolated datasets, we have $a_1 = -0.9876$, $a_2 = 1.9863$ and $\alpha_3 = -0.1023$. The Euclidean norm of residuals was 1.1923. Figure 6 shows the graphs of the measured interpolated datasets and the LS estimations. Table 2 shows both the relative and RMS errors for the LS fitting in the three interpolated datasets.

We also tested the LS fitting using higher-order polynomials. The second-order LS fitting obtained good results for both the original and interpolated dataset curves. However, when the order of f was increased for the interpolated datasets, the LS could not find an optimal solution, no matter the choice of initial point.

8.3. ARIMA Fitting: Original Datasets. Let Z_2 be the mathematical notation for the original measured series of the "radial 2" dataset. We examined the measured data without seasonal components. Since we also isolated its trend, as described above, we concluded that $Z_2 = L_2 + N_2$, with L_2 representing the linear term of Z_2 and N_2 its nonlinear term. The ARIMA adjustment is made for L_2. That said, we also have $L_2 = T_2 + \alpha_2$, where T_2 is the tendency for Z_2 and α_2 its white noise (in which may be some nonlinear information). Since the linear trend was calculated before, by means of an LS method, we have $\widehat{T}_2$ as an estimate for this tendency. Therefore, the series that must be estimated by the ARIMA model is represented by Y_2 in $Y_2 = L_2 - \widehat{T}_2$. The estimated series will be called $\widehat{Y}_2$ and is represented by $\widehat{Y}_2 = \widehat{L}_2 - \widehat{T}_2$. The mathematical representations for the "radial 3" and "radial 1" series were obtained in an analogous way, and their estimated series are called $\widehat{Y}_3$ and $\widehat{Y}_1$, respectively. After analyzing the ACF and PACF graphs for the series under study, we decided to employ an ARIMA (2, 0, 0) model to fit the training data. This is represented by equation (2).

$$Y_h = \phi_1 Y_{h-1} + \phi_2 Y_{h-2} + c, \tag{2}$$

where $\phi_1 = -0.0310998$, $\phi_2 = -0.543184$, and $c = -0.35095$.

The graph with the best adjustment for the "radial 2" dataset is shown in Figure 5. This is the graph that originated from the estimate of (2) when applied to its own adjustment dataset, i.e., Z_2, as in $\widehat{L}_2 = \widehat{Y}_2 + \widehat{T}_2$. By analogy with $\widehat{L}_2$, we can write $\widehat{L}_3$ and $\widehat{L}_1$. In addition, Figure 5 shows, as well, the graphs of the estimates of the ARIMA model for the "radial 3" and "radial 1" datasets as well. All these graphs also show the estimates of the ITU and LS method for each radial. Table 1 shows the relative and RMS errors of the ARIMA, LS, and ITU estimates for the three radials studied in this work.

8.4. ARIMA Fitting: Interpolated Datasets. We represent variables that are originated from interpolated datasets with the symbol "~" above the variable letter, as seen when comparing equation (3) with equation (2). When proceeding in an analogous way to the original datasets group, the fitting process for the interpolated series gave, as its best result, an ARIMA (4, 0, 0) model expressed as in equation (4).

$$\widetilde{Y}_h = \widetilde{\phi}_1 \widetilde{Y}_{h-1} + \widetilde{\phi}_2 \widetilde{Y}_{h-2} + \widetilde{\phi}_3 \widetilde{Y}_{h-3} + \widetilde{\phi}_4 \widetilde{Y}_{h-4} + \widetilde{c}, \tag{3}$$

where $\widetilde{\phi}_1 = 3.30286$, $\widetilde{\phi}_2 = -4.08633$, $\widetilde{\phi}_3 = 2.24252$, $\widetilde{\phi}_4 = -0.461087$, and $\widetilde{c} = -0.000971798$. The graphs with the ARIMA results for interpolated data are shown in Figure 6. The relative and RMS error for the interpolated datasets are displayed in Table 2.

8.5. Neural Network Fitting: Original Samples. We fitted a neural network for the difference between the ARIMA estimate and the original data. That is, let $\widehat{L}$ be the ARIMA estimate of one measured dataset Z. This can be written as in equation (4).

$$Z = \widehat{L} + N. \tag{4}$$

In equation (4), the nonlinear term of Z, which will be fitted by the neural network, is represented by N.

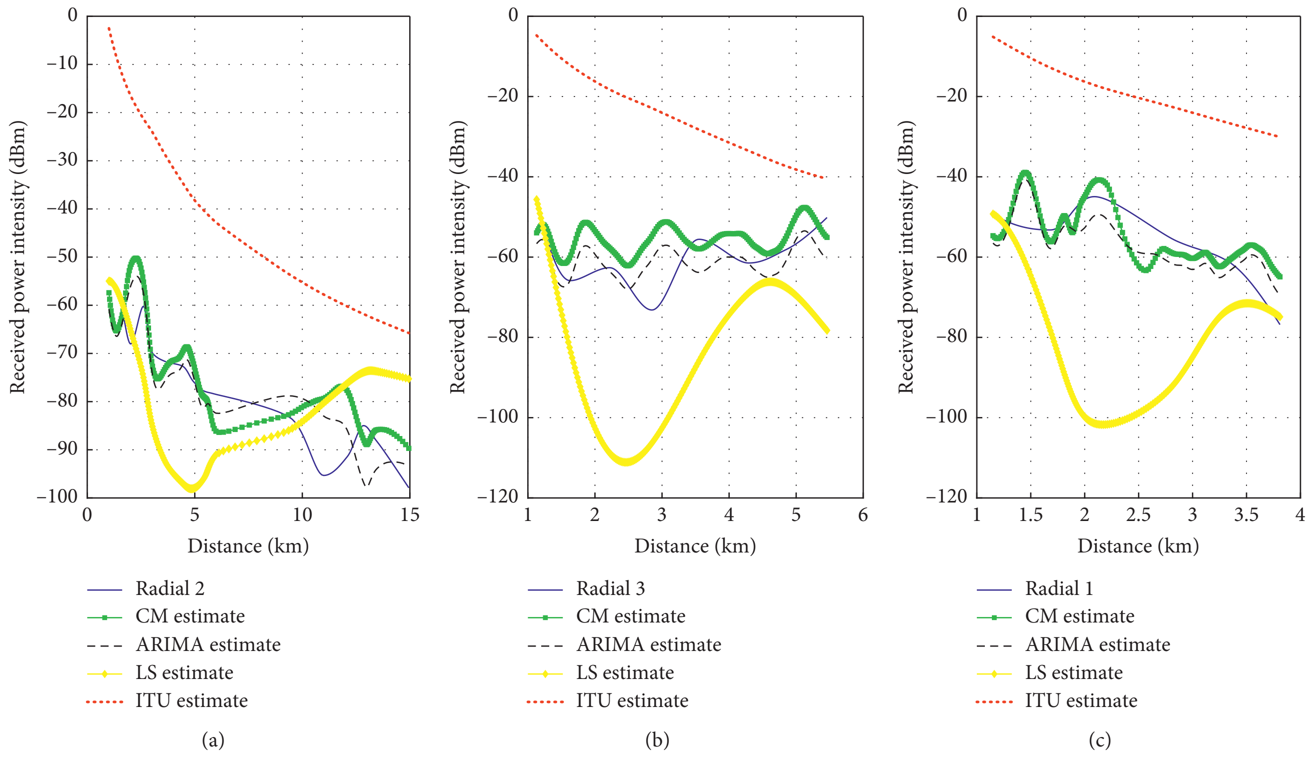
In this study, we use RBF with two layers from an internal generalized regressor function of MATLAB (*newgrnn*). It has a Gaussian activation function, and the network estimate $\widehat{N}$ is given by equation (5).

$$\widehat{N} = \sum_{j=1}^{n} w_j \exp\left(-\frac{\left(x - c_j\right)^2}{2\sigma_j^2}\right), \quad j = 1, \ldots, n. \tag{5}$$

In the original training dataset, we used four of the twelve original samples to train the network, so that the N_2 (non-linear part of Z_2 dataset, which is the training set) group could be fitted. Finally, the estimated values from the neural network $\widehat{N}$ are then added to the estimated ARIMA values, so we have the final estimation model for $\widehat{Z}$, which is given by equation (6).

$$\widehat{Z} = \widehat{L} + \widehat{N}. \tag{6}$$

In the case of the original datasets, the adjustment for "radial 2" and the estimates for "radial 3" and "radial 1" datasets are shown in Figure 5. Table 1 shows the relative and RMS errors values of every type of modelling compared in this work (for the original datasets).



(a) (b) (c)

Figure 6: Graphs of combined ARIMA model adjust for radial 2 (a) and estimates for radial 3 (b) and 1 (c).

Table 2: RE and RMSE values for interpolated datasets.

	Radial 2		Radial 3		Radial 1	
	RE	RMSE	RE	RMSE	RE	RMSE
ARIMA	0.0635	5.8653	0.1019	6.2009	0.0624	4.3491
CM	0.0650	6.3012	0.1183	7.6129	0.0982	7.4572
LS	0.1701	15.088	0.5020	31.829	0.3564	26.320
ITU	0.5155	38.436	0.6544	36.402	0.5820	38.393

8.6. Neural Network Fitting: Interpolated Samples. Analogously, Figure 6 shows the results for the fitting and estimations on the interpolated datasets, and their respective relative and RMS errors are shown in Table 2.

8.7. Complementary Results. The abovementioned results show that the proposed hybrid modelling has a slight worse result than the single ARIMA fitting. There is evidence in the literature that this is, in fact, possible and even expected, sometimes, such as exposed in [15, 16]. Aiming to solve this problem, we can tackle the problem differently. In this work, we apply three other possibilities of combinations and calculations in the nonlinear fitting stage. The first possibility consists on using an algorithm to find the best value for the spread variable of the ANN already used here, as this is the only variable that can be changed in the original architecture of the ANN used so far. We chose a search method of the "for" type between values from 0.3 to 1, with a step of 0.1. We expect that the best value found is 0.3, since this makes the ANN fitting closer to the training points. The second possibility consists in developing another ANN, but using an architecture inspired in [15]. The third possibility is to test other combinations for $\hat{L}$ and $\hat{N}$ using the ANN of the second alternative to calculate $\hat{Z}$. We test the sum combination $(\hat{L} + \hat{N})$, the element-wise product (array element) combination $(\hat{L} \times \hat{N})$ and the exponential combination $(\hat{L}^{(\hat{N}/\max(\hat{N}))})$. The terms $\hat{L}$ and $\hat{N}$ are vectors. The linear calculation is not modified in any way, as well as the hybrid nature of the proposed modelling. All the tests executed in this subsection involve changes in the nonlinear calculation stage, since the first round of tests was not as good as expected.

The graphs of the estimates for the spread searched and the ANN inspired by [ARTIGO MAIS FÁCIL], called "ANN #2," are shown in Figure 7. The architecture of ANN #2 is shown in Figure 8. The RMSE values for both sets of estimates as shown in Table 3.

The architecture of ANN#2 is very different from the generalized regressor used on the first set of calculations of the last subsections. It now has the sigmoid function (see equation (7)) as its hidden layer activation function (there are two hidden layers) and the training method is the Levenberg–Marquardt. The last layer has a *purelin* function, which normalizes the values of the ANN aiming to make the output value range the same of the input, since, internally, the ANN may work with a different range of values. The inputs and outputs are now matrices of two columns and n lines, with n being the quantity of samples of each vector. The first column of the input matrix is a vector with elements valued from 1 to n, that is, the "X axis." The second column is the target, i.e., the values which the ANN needs to fit. This setup provided significant better results than the standardized

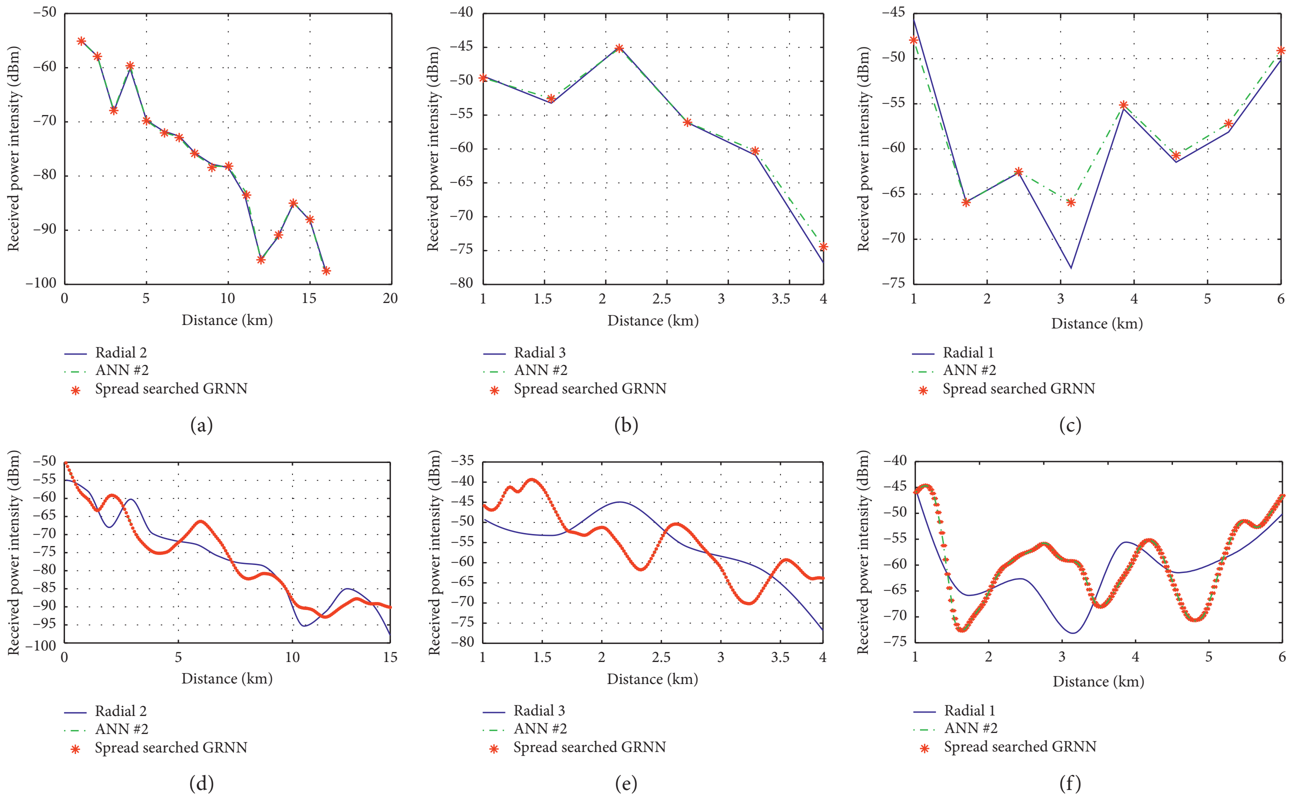


Figure 7: Graphs of ANN #2 and spread searched fittings for original radial 2 (a); original radial 3 (b); original radial 1 (c); interpolated radial 2 (d); interpolated radial 3 (e);and interpolated radial 1 (f).

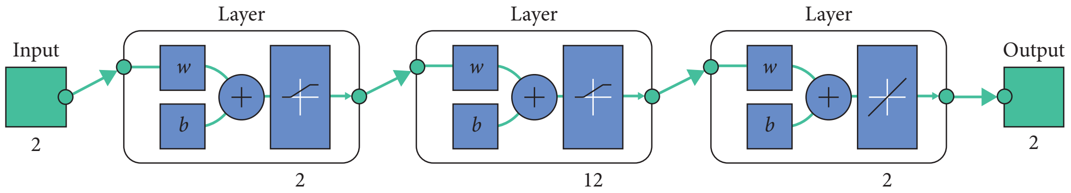


Figure 8: ANN#2 diagram [14].

Table 3: RMSE values for ANN#2 and spread searched ANN tests.

Hybrid modelling	Radial 2 RMSE (dB)	Radial 3 RMSE (dB)	Radial 1 RMSE (dB)
Original ARIMA-ANN#2	0.7227	1.0602	2.7523
Interpolated ARIMA-ANN#2	3.2126	7.1481	5.1963
Original ARIMA-spread-searched-ANN	0.7227	1.0602	2.7523
Interpolated ARIMA-spread-searched-ANN	3.2126	7.1481	5.1963

generalized regressor ANN used before, as confirmed by the RMSE values on Table 3. We applied the same new type of input (two-column matrices, instead of a single vector) to the RBF ANN used with the spread search technique as follows:

$$\mathrm{Sig}(x) = \frac{1}{1+e^{-x}}. \tag{7}$$

We want to stress that, in Figures 9(d)–9(f), the spread searched ANN and the ANN#2 networks obtained the same result, since the red dots are exactly on the green curve, which is almost invisible. We can conclude either the ANN#2 or the spread searched generalized regressor were able to improve the first results (Figures 5 and 6 and Tables 1 and 2).

Curiously, the best spread value obtained from the "for" technique was 1, differently from what we expected. This value was the same used before and the standard value from MATLAB. Thus, we conclude that the new inputs in the generalized regressor were the key

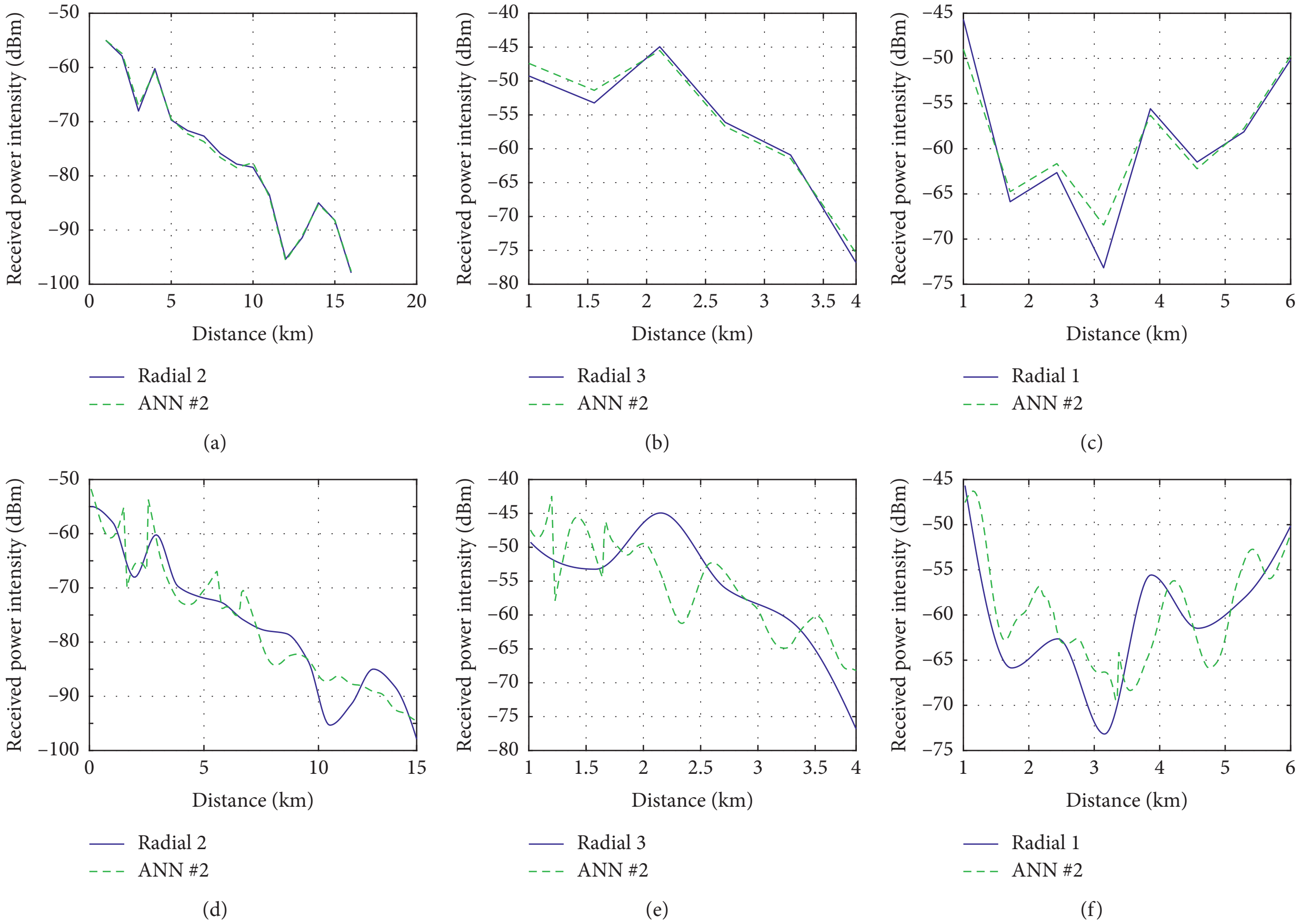


FIGURE 9: Sum combination for original radial 2 (a); original radial 3 (b); original radial 1 (c); interpolated radial 2 (d); interpolated radial 3 (e);and interpolated radial 1 (f).

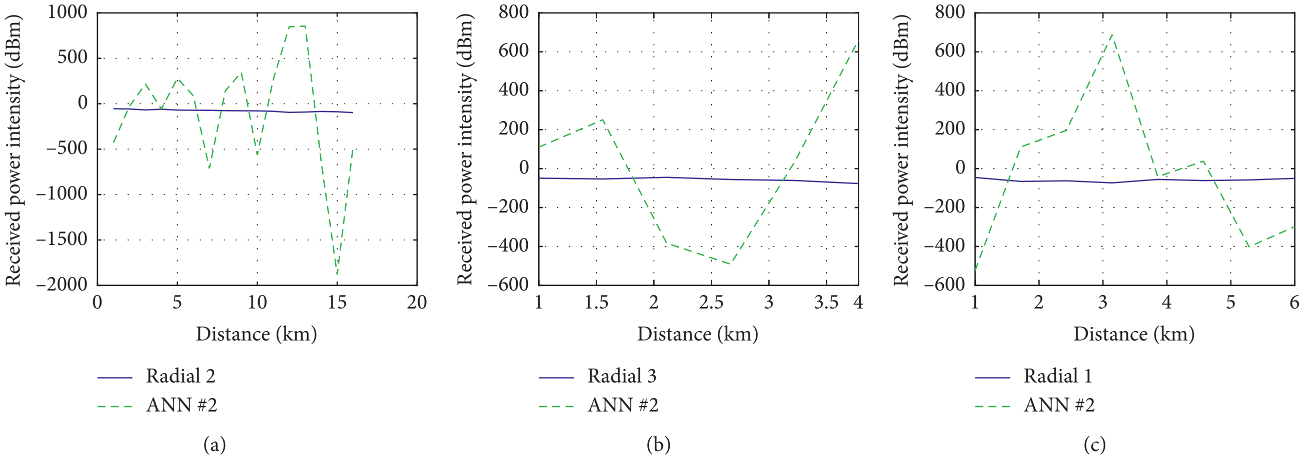
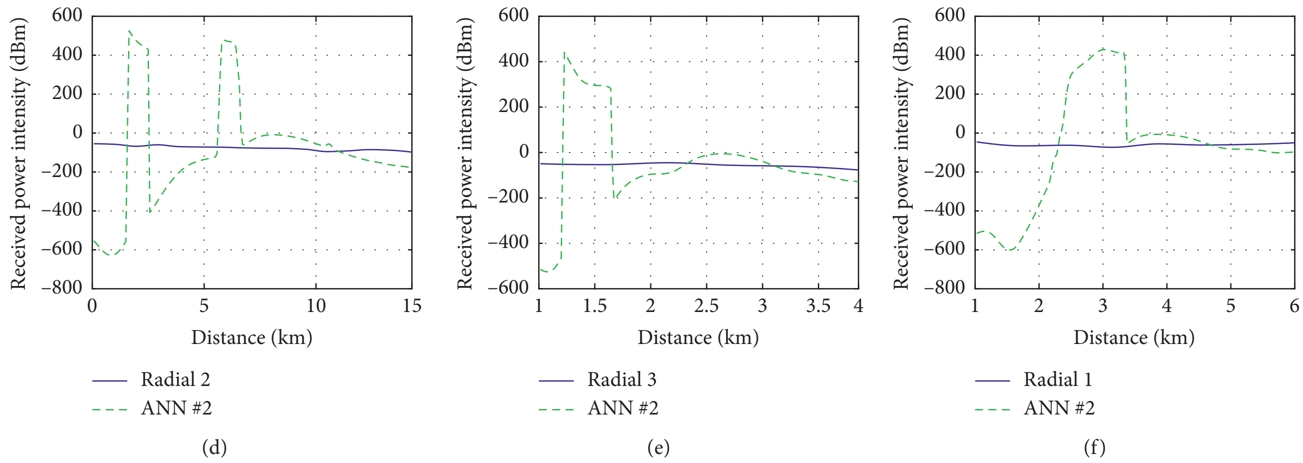


FIGURE 10: Continued.

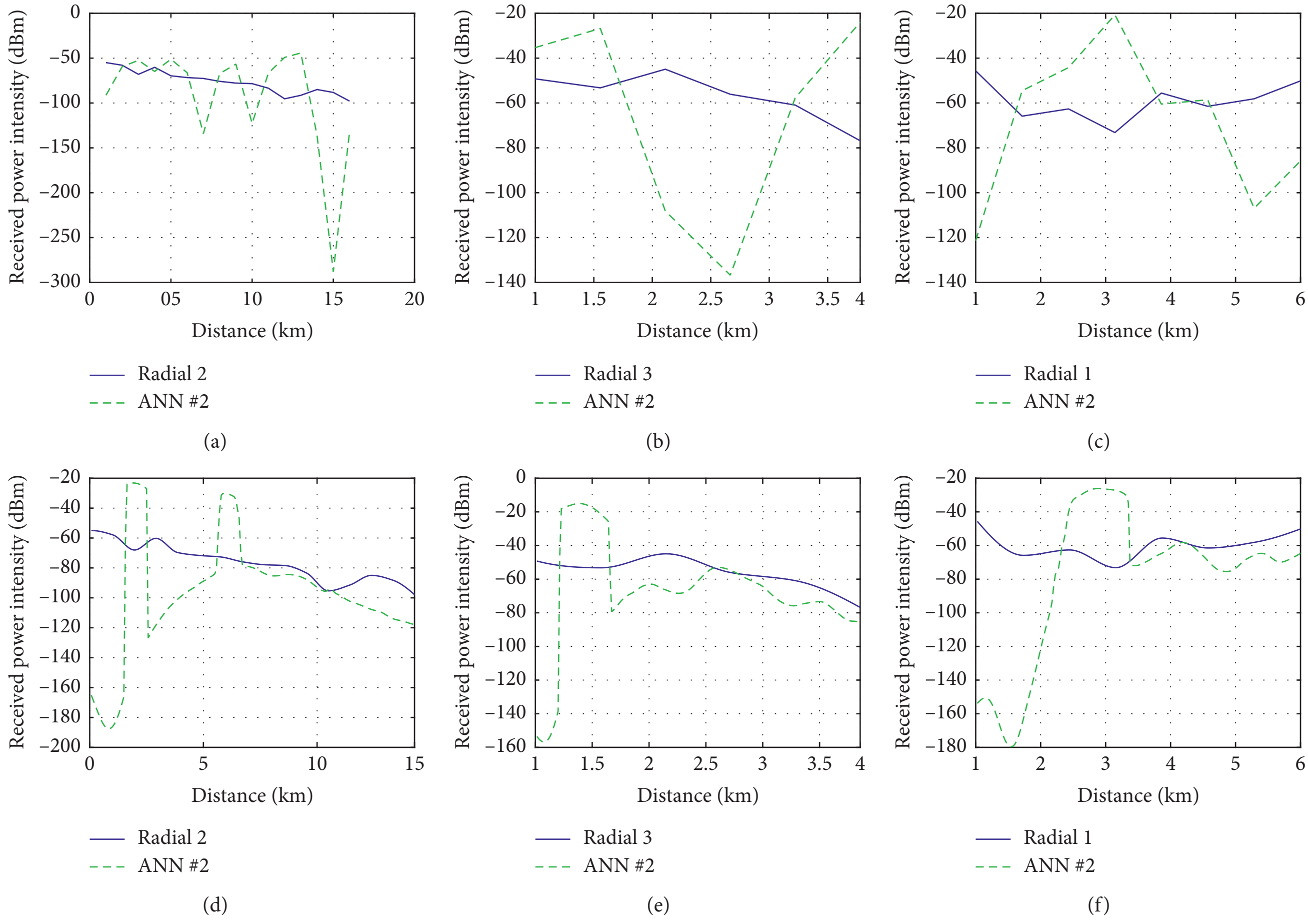


(d)

(e)

(f)

Figure 10: Product combination for original radial 2 (a); original radial 3 (b); original radial 1 (c); interpolated radial 2 (d); interpolated radial 3 (e);and interpolated radial 1 (f).



(a)

(b)

(c)

(d)

(e)

(f)

Figure 11: Exponential combination for original radial 2 (a); original radial 3 (b); original radial 1 (c); interpolated radial 2 (d); interpolated radial 3 (e);and interpolated radial 1 (f).

Table 4: RMSE values for different combinations for $\widehat{L}$ and $\widehat{N}$.

Combination	Radial 2 RMSE (dB)	Radial 3 RMSE (dB)	Radial 1 RMSE (dB)
Original sum combination	0.5673	1.2853	2.1471
Original product combination	652.2583	402.9861	370.2783
Original exponential combination	59.4781	48.6382	39.7458
Interpolated sum combination	3.7624	5.5556	4.4677
Interpolated product combination	273.0819	196.4770	291.8673
Interpolated exponential combination	46.5264	32.4439	49.0257

factor to the improvement of the results when using the RBF ANN.

Regarding the combination tests for $\widehat{L}$ and $\widehat{N}$, their graphs and RMSE values are shown in Figures 9–11 and Table 4, respectively.

From the graphs in Figures 9–11 and results in Table 4, we conclude that for the given datasets, the best combination for the linear and nonlinear terms is, indeed, a sum. Different from [15], we judged not necessary to let the ANN decide what combination was better. We chose, previously, some combinations deemed more probable of giving good results and tested, since we have empirical data and the proposed model does not have terms for environmental or physical influences (at least not yet).

9. Conclusions

When the original datasets are considered, the single ARIMA adjustment is, at least, equivalent to the usual LS fitting. In addition, it seems to be unnecessary to tackle this specific problem by complementing it with the ANN to fit the non-linear terms of the studied series, since when the network is applied, there is a slight increase in errors. Regarding the interpolated datasets, the LS fitting was not able to adjust to the training set properly. This was possibly due to the increased number of samples, which increases the size of the search area. The ARIMA fitting was, at least, 11% better than the LS fitting (it should be emphasized that this 11% improvement was found in the training set). In the sets used for comparison, the benefits of the ARIMA fitting were much greater than this. In a similar way to the original datasets, it is apparently not necessary to complement it with the ANN fitting, since the errors also increase slightly for radials 2 and 3, and there is a bigger rise for radial 1.

From the first set of results, we chose to make other tests, inspired in the literature, with another architecture of ANN and by varying the spread value on the generalized regressor initially applied to this problem. These changes provided better results, improving significantly its reliability (for the studied datasets).

As future improvements to this work, the authors intend to acquire data from different areas, in order to improve the generality of the proposed model. It may, although, imply on changing its mathematical formulation, since the ARIMA fitting depends, basically, on the training series. Another way of improving this model is to isolate the weather influence. It would be done by measuring each point again on the rainy season, as in [2], and analysing the series acquired on this second measurement campaign. A third suggestion on how to improve this study is testing another hybrid technique in the same problem and assessing if there is one more suited to the electromagnetic propagation modelling.

Data Availability

The .xlsx ("radial_1_PSNR_POT_DIST," "radial_2_PSNR_POT_DIST," and "radial_3_PSNR_POT_DIST") datasets archives and.txt ("Intructions for.csv archives") instructions archives used to support the findings of this study are included within the supplementary information file(s).

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work was supported by the Federal University of Pará, by means of its infrastructure and measurement equipment.

Supplementary Materials

The datasets including the measurement data and the instructions on how to use this data in order to obtain the results in this work are supplementary materials. The datasets are archives in.xlsx. The instructions on how to use them are described in a .txt archive. (*Supplementary Materials*)

References

[1] G. P. Zhang, "Time series forecasting using a hybrid ARIMA and neural network model," *Neurocomputing*, vol. 50, pp. 159–175, 2003.

[2] I. R. Gomes, C. R. Gomes, H. S. Gomes, and G. P. dos Cavalcante, "Empirical radio propagation model for DTV applied to non-homogeneous paths and different climates using machine learning techniques," *PLoS One*, vol. 13, no. 3, 2018.

[3] M. Liangping and G. Sternberg, "Packet-based PSNR time series prediction for video teleconferencing," in *Proceedings of the International Conference on Multimedia and Expo*, Turin, Italy, June 2015.

[4] E. Rozal, E. Pelaes, J. Queiroz, and C. Salame, “Modeling of wireless networks using multivariate time models,” *EURASIP Journal on Advances in Signal Processing*, vol. 2012, p. 248, 2012.

[5] W. Wei, J. Jiang, H. Liang et al., “Application of a combined model with autoregressive integrated moving average (ARIMA) and generalized regression neural network (GRNN) in forecasting hepatitis incidence in Heng county, China,” *PLoS One*, vol. 11, no. 6, 2016.

[6] L. Yu, L. Zhou, L. Tan et al., “Application of a new hybrid model with seasonal auto-regressive integrated moving average (ARIMA) and nonlinear auto-regressive neural network (NARNN) in forecasting incidence cases of HFMD in shenzhen, China,” *PLoS One*, vol. 9, no. 6, 2014.

[7] P.-F. Pai and C.-S. Lin, “A hybrid ARIMA and support vector machines model in stock price forecasting,” *Omega*, vol. 33, no. 6, pp. 497–505, 2005.

[8] S. Panigrahi and H. S. Behera, “A hybrid ETS-ANN model for time series forecasting,” *Engineering Applications of Artificial Intelligence*, vol. 66, pp. 49–59, 2017.

[9] C. N. Babu and B. E. Reddy, “A moving-average filter based hybrid ARIMA-ANN model for forecasting time series data,” *Applied Soft Computing*, vol. 23, pp. 27–38, 2014.

[10] J. F. L. de Oliveira and T. B. Ludermir, “A hybrid evolutionary decomposition system for time series forecasting,” *Neurocomputing*, vol. 180, pp. 27–34, 2016.

[11] H. Zahra and K. Mehdi, “Hybrid structures in time series modelling and forecasting: a review,” *Engineering Applications of Artificial Intelligence*, vol. 86, pp. 83–106, 2019.

[12] G. E. P. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, *Time Series Analysis: Forecasting and Control*, John Wiley & Sons, Hoboken, NJ, USA, 4th edition, 2008.

[13] J. J. Moré, “The Levenberg-Marquardt algorithm: implementation and theory,” in *Numerical Analysis. Lecture Notes in Mathematics*, G. A. Watson, Ed., Vol. 630, Springer, Berlin, Germany, 1978.

[14] “MATLAB version 8.0.0.783,” in *Technical Documentation*, 2012.

[15] M. Khashei and M. Bijari, “A novel hybridization of artificial neural networks and ARIMA models for time series forecasting,” *Applied Soft Computing*, vol. 11, no. 2, pp. 2664–2675, 2011.

[16] S. O. Domingos, J. F. L. de Oliveira, and P. S. G. de Mattos Neto, “An intelligent hybridization of ARIMA with machine learning models for time series forecasting,” *Knowledge-Based Systems*, vol. 175, pp. 72–86, 2019.