

## Research Article

# Hybrid ARIMA and Neural Network Modelling Applied to Telecommunications in Urban Environments in the Amazon Region

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

*Post Graduation Program in Electrical 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 30 October 2019; Accepted 23 December 2019; Published 24 February 2020

Academic Editor: Lorenzo Luini

Copyright © 2020 Ramz L. Fraiha Lopes et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This study explores the use of a hybrid Autoregressive Integrated Moving Average (ARIMA) and Neural Network modelling for estimates of the electric field along vertical paths (buildings) close to Digital Television (DTV) transmitters. The work was carried out in Belém city, one of the most urbanized cities in the Brazilian Amazon and includes a case study of the application of this modelling within the subscenarios found in Belém. Its results were compared with the ITU recommendations P. 1546-5 and proved to be better in every subscenario analysed. In the worst case, the estimate of the model was approximately 65% better than that of the ITU. We also compared this modelling with a classic modelling technique: the Least Squares (LS) method. In most situations, the hybrid model achieved better results than the LS.

## 1. Introduction

The growth in the population density of big cities has made residential and commercial buildings a common feature of the urban landscape, as they optimize the spatial exploitation of a terrain. Sometimes, though, these buildings are located near transmitter towers. Depending on the operating frequency and received power intensity of these transmitters, there is a real concern about the excessive exposure of the general public to nonionizing radiation (NIR), which in particular can affect citizens, building constructors, and supervisory agencies.

This article sets out a hybrid modelling system to estimate the propagation of the electric field along vertical paths close to transmitter towers that operate around 600 MHz. We chose this frequency range because it is adopted by Brazilian Digital Television (DTV) and, at least in Belém and similar cities, one of the most significant sources of outdoor/indoor NIR exposure are VHF/UHF transmitters operating at a high-voltage power.

Finding a definitive solution to this problem, such as, for example, fitting frequency filters to the walls of the buildings or relocating the transmitter towers, requires great political will and a high financial cost. In the meantime, people are being continuously exposed to NIR in their own homes/offices. This subject is still under discussion and has led to various approaches, such as in [1–6]. Hamiti et al. [1] describe a study of exposure to NIR involving multiple frequencies in the city of Prishtina. Koprivica et al. [2] carry out a statistical analysis of electromagnetic radiation, but are restricted to GSM frequencies in Serbia. Bernhardt [3] addresses general subjects related to NIR. Belpomme et al. [4] examine some of the expected effects of long-term exposure to NIR. Kuzniar et al. [5] conduct a study of the biological effects of short-term exposure to NIR with three frequencies (50 Hz, 2.1 GHz, and 5.8 GHz) in mammalian cells. Bernabò et al. [6] carry out a survey of the scientific literature with regard to the effects of NIR exposure on fertility (the authors examined 104 papers published from 1989 to 2017).

There have been suspicions about the health effects of continuous NIR exposure on these people which have arisen because of the uncertainty of the results obtained so far. In light of this, the aim of this study is to provide a degree of comfort to the general public by helping to estimate the level of exposure to NIR (i.e., electric field exposure), in vertical path scenarios satisfactorily. With the aid of this information and on the basis of the exposure safety standards for NIR, ordinary citizens can check if their level of exposure to NIR (in a selected frequency) exceeds the permitted limit, which may mean their health is at risk. The modelling proposed in this article is designed to help involved parties to make an assessment of their risk of exposure to NIR.

The system used in this work is a hybrid ARIMA and neural network model. This type of model is well known for its ability to address problems that have both linear and nonlinear components in their mathematical formulation, since single ARIMA models are widely used to represent linear problems and neural network models can represent nonlinear problems satisfactorily. The purpose of this hybrid approach is to use the advantages of both modelling techniques to tackle a single problem.

Neural network modelling is a technique inspired by how the human brain operates, with regard to the synaptic connections of its neurons. It is widely used (either alone or combined with other techniques) in many areas, including for the solution of electromagnetic propagation problems. This wide range of applications is the most important advantage of a neural network.

Models such as ARIMA, that is, based on the time series theory, are widely used in areas such as economics and biology, as well as for any problem that can be modelled using a series of time-based data as its main variable. However, this type of modelling is rarely used to meet the requirements of electromagnetic propagation, which is the main problem in this work.

Some studies closely related to the model set out in this work are [7–12]. Wang et al. [7] analyse the phenomenon of tuberculosis incidence by means of a hybrid ARIMA model and nonlinear autoregressive neural network and compare its estimates to those of a single ARIMA model. In [8], a hybrid seasonal ARIMA (SARIMA) and neurofuzzy system network is used to predict the monthly inflow of water, as it is an extremely important variable in water resource planning. This model was compared with a combined SARIMA and neural network model and achieved better results for its purpose. Naveena et al. [9] employ a joint hybrid ARIMA and neural network model to predict the price of “Robusta Coffee” in India and compares this with single ARIMA and neural network models. On the basis of the chosen evaluation metric, the hybrid model predicted the prices more accurately. In [10], the main benchmark for the model is chosen. A large number of studies which use similar hybrid models have been influenced by [10]. Finally, Khandelwal et al. and Aladag et al. [11, 12] provide variations of the methodology devised by [10] and applied to literature-based data.

## 2. Materials and Methods and the Proposed Model

*2.1. NIR Problem and General Methodology.* This particular problem of human exposition to NIR was analysed in the central zone of Belém city (located in the Brazilian state of Pará, in the Brazilian Amazon region), where the transmitters of its main television/radio stations are located. Some of these towers are as close as 50 m to residential buildings. An aggravating factor is that, instead of having repeater stations throughout the city operating at a lower power, there is only one transmitter for each station that usually operates at an excessively high power. For instance, the highest transmission power level in Belém is around 20 kW and this tower is located in one of the most urbanized neighbourhoods of the city, where there is a residential building that is almost as high as the tower situated approximately 100 m from it.

Table 1 shows the limits of exposure to NIR for the general public adopted in Brazil. In this work, we take into account the values of electric field exposure limits, since the measurements were of electric field intensity. It should be noted that the Brazilian limits to NIR exposure follow the public exposure guidelines defined by the International Commission on Non-Ionizing Radiation Protection (ICNIRP) for the frequency range of 8.3 kHz to 300 GHz.

As the transmitter analysed in this work operates at 600 MHz, the electric field intensity exposure limit equals 33.6805 V/m. This is the limit of the general public exposure and should not be exceeded by the mean value of the electric field over a time lag of 6 minutes.

The assessment of human exposure to NIR at any given point can be measured by using the quadratic relationship following [13]

$$ER = \sum_i \left( \frac{E_{m,i}}{E_{L,i}} \right)^2, \quad (1)$$

where  $E_{m,i}$  is the measured electric field at the given point in frequency  $i$  and  $E_{L,i}$  is the exposure limit for frequency  $i$ . Equation (1) shows that the effect from exposure to NIR is cumulative, if more than one source is present. As only one frequency is considered in this work, the sum in equation (1) becomes a single quadratic ratio for each point, as variable  $i$  has a single value for every measured point (see equation (2)). The value of the ER ratio must not be higher than the unity, that is,  $ER \leq 1$ . If the ER value exceeds 1, it means there is an excessive exposure to NIR in the considered location:

$$ER = \left( \frac{E_m}{E_L} \right)^2. \quad (2)$$

The studied scenario and its measurement campaigns resulted in three datasets, namely, “data1,” “data2,” and “data3.” They represent a single scenario: a building close (less than 250 m) to a transmitter tower operating at approx. 600 MHz. These three datasets can be regarded as three subscenarios, that is, three similar situations. Figure 1 shows a representation of the studied scenario. All three datasets

TABLE 1: Limits of the general public exposure to electromagnetic fields in the range of frequencies between 8.3 kHz and 300 GHz [13].

Radio frequency range	Electric field intensity (V/m)	Magnetic field intensity (A/m)	Power density, seq (W/m <sup>2</sup> )
9 kHz to 65 kHz	87	5	—
0.065 MHz to 1 MHz	87	$0.73/f$	—
1 MHz to 10 MHz	$87/f^{1/2}$	$0.73/f$	—
10 MHz to 400 MHz	28	0.073	2
400 MHz to 2000 MHz	$1.375 f^{1/2}$	$0.0037 f^{1/2}$	$f/200$
2 GHz to 300 GHz	61	0.16	10

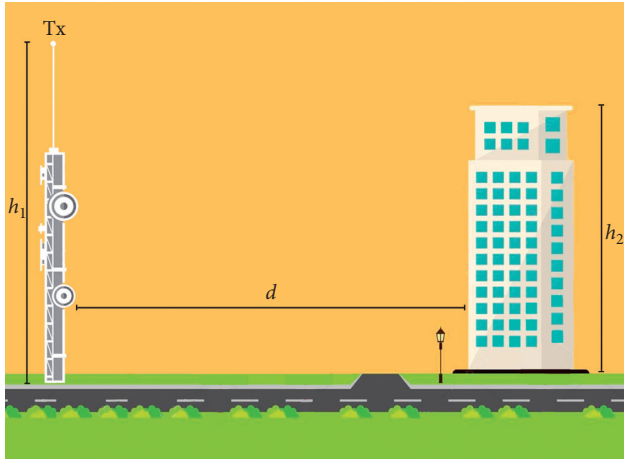


FIGURE 1: Representation of the studied scenario. Tx abbreviation stands for “transmitter tower.” Variable  $d$  is the horizontal distance between the measured building and the transmitter tower (not to be confused with the distance of the receiver equipment from the transmitter source). Variable  $h_1$  is the height of the transmitter tower and  $h_2$  represents the height of the measured building.

were obtained through measurement campaigns and consist of values of electric field intensity (V/m) in every floor of each measured building. The receiver equipment was always located in such a way that it was directly facing the transmitter tower and the data was acquired for 6 minutes, as recommended by [13], which resulted in a series of values of the electric field for each floor. The mean value of each series was taken to represent the received intensity of the electric field in each floor. Table 2 provides information of the variables in Figure 1.

The proposed model was developed from a hybrid methodology that combined an ARIMA model and a neural network, inspired by [10–12]. Let  $E$  be the value of the electric field intensity. This can be written as the sum of two components (equation (3)), one containing its linear part ( $L$ ) and the other containing its nonlinear part ( $N$ ), that is,

$$E = L + N. \quad (3)$$

In this work, the linear component  $L$  is first adjusted by an ARIMA model. Secondly, the difference between the ARIMA estimate  $\hat{L}$  and the measured data ( $E$ ), i.e., the residuals of the ARIMA estimates, is adjusted by the neural network.

Datasets representing other scenarios (different from those found in Belém) must be acquired to ensure that the model operates effectively in these new scenarios, as it is an

TABLE 2: Information on the tested subscenarios.

Subscenario	“data1”	“data2”	“data3”
$d$	48 m	100 m	200 m
$h_1$	212 m	125 m	125 m
$h_2$	81 m	11 m	81 m

empirically based model. New datasets can be acquired through measurement campaigns and/or simulations. These are some of the continuous upgrades designed for this work, as described in Section 4. Another advantage of this model is that it can be deployed to address a wide range of problems, as it can design both linear and nonlinear systems precisely. The necessary calculations and programs were carried out on MATLAB [14] software, by means of internal functions, both for ARIMA and the adjusted neural network.

**2.2. ARIMA Fitting Methodology.** The ARIMA adjustment was made by adopting the usual strategies from the time series theory, mainly autocorrelation function (ACF) and partial autocorrelation function (PACF) analysis for the original, or (somehow) transformed, time series, as in [15]. The neural network is a radial basis network with two layers.

It should be noted that, when using the ARIMA model, the (usual) variable “time” is replaced with the “height from ground” variable, which characterizes a vertical path. In other words, it is assumed that the electric field intensity on one floor of a building is determined by the values (the exact quantity depends on the order of the model) from the floors below. The standard Cartesian coordinate system is used for the  $y$ -axis with a variable height (zero at the base and increasing as it rises).

The analysis and the results of this work were divided into two groups: (1) Original Measured Data and (2) Interpolated Measured Data. As the names suggest, the results of Group 1 were obtained from the datasets with their original number of samples. With regard to the second group, the results were obtained after adjusting the model to the interpolated measured data. We did this to increase the number of samples of each measured dataset, thus allowing the ARIMA model to work with more samples, thus refining the linear fitting. We used a shape-preserving piecewise cubic interpolation (SPPCI) to increase the number of samples of each dataset from 25 (datasets: “data3” and “data1”) or 31 (dataset: “data2”) to 200. In addition, the interpolated group of datasets was able to simulate a “nonstop” measurement campaign, which is more desirable than a “stop-and-go” campaign. Instead of having to stop at

each floor of a building to carry out the measurement, we could use a device that allows continuous measurements to be made (a receiver attached to a drone moving at constant speed, for example) with no stops. However, our measurement campaigns were of the “stop-and-go” type.

There are some steps that have to be taken to adjust the ARIMA model. Firstly, it is necessary to determine if a nonlinear transformation (e.g., logarithm transformation) of the original series is necessary to stabilize its variance. Secondly, the tendency of the original data must be calculated and isolated, if necessary, so that the adjustment of the series can proceed without its tendency. Then, on the basis of the ACF and PACF analysis, it must be decided whether differences in the series should be made. Finally, the type/order of the ARIMA model will be obtained when the ACF and PACF graphs make correct evaluations after the necessary interventions in the original series. All these stages follow the standard [15] approach when modelling with time series, where it is necessary for the series that must be adjusted to be stationary (or “close” to it). The original series of graphs for the three datasets are shown in Figure 2, while the graphs for the interpolated datasets are shown in Figure 3.

In the process of achieving the best result for the proposed modelling, we carried out a wide range of tests that followed different stages. The diagram in Figure 4 shows all the stages of the testing procedure followed in this study. Optional stages are to take a nonlinear transformation and taking differences on the original/interpolated series. These stages are indicated in the diagram by dashed lines.

When analysing the diagram, it was noticed that the decision about the interpolation was made right at the beginning of the process required for this work. We did this to simulate the two scenarios under consideration (i.e., the “non-stop” and “stop-and-go” measurement campaigns) and allow a more refined ARIMA estimate, as stated previously. We isolated the tendencies of the series, which is a standard procedure for the ARIMA fitting and replicated in the LS fitting to make a fair comparison between the fitting methods. We, then, continued with the ARIMA fitting, by analysing if a nonlinear transformation is necessary (that is the reason for the dashed lines in this stage, since this stage is optional). Following this, we analysed the ACF and PACF functions of the series under study and decided whether to take difference(s) or proceed to define the ARIMA model. After the ARIMA fitting, we proceeded with the neural network fitting, which is carried out by finding the difference between the measured data and ARIMA estimates, i.e., the residuals from the linear estimate. Finally, we obtained the results of the modelling, which were ready for making comparisons and reaching conclusions.

The optional stages in the diagram were tested and the best results were obtained when dataset “data3” was used as a calibration set (for both the interpolated and original datasets) without nonlinear transformation, but isolating the tendency and fitting the ARIMA to the series without it. This means that the ARIMA coefficients were calculated by only using the “data3” group as an adjustment/training set.

Groups “data1” and “data2” were only used for making comparisons and estimates.

In Sections 3.1 and 3.2, we briefly explain how we carried out the fitting by employing the LS method. This was carried out to act as a counterpoint to the modelling, as it is a usual procedure for tackling problems like this. We will, later, compare this method with the LS and the ITU recommendation P. 1546-5 in Section 4 of this study. In addition, we explain in detail how we obtained the best linear results in Sections 3.3 and 3.4. As for the neural network fittings, these are described in Sections 3.5 and 3.6.

*2.3. LS Fitting Methodology.* To ensure a fair comparison is made between the combined ARIMA and Neural Network fitting, “data3” will be used as a calibration dataset and the fitting procedure will be carried out in the series without its tendency.

The general problem was represented by choosing a recursive second order polynomial given by

$$f_h = a_1 f_{h-1} + a_2 f_{h-2} + a_3. \quad (4)$$

This represents a situation where the current value of  $f$ , i.e.,  $f_h$ , depends on the two preceding values (based on the adopted metric in the problem, of course), which is similar to how the ARIMA model makes its estimates.

Equation (5) expresses the system of equations representing each value of  $f_h$  from  $f_3$  to  $f_n$ . We consider values  $f_1$  and  $f_2$  to be the first two values of the training dataset:

$$\begin{cases} f_3 = a_1 f_2 + a_2 f_1 + a_3, \\ f_4 = a_1 f_3 + a_2 f_2 + a_3, \\ \vdots \\ f_n = a_1 f_{n-1} + a_2 f_{n-2} + a_3. \end{cases} \quad (5)$$

System in equation (5) is incompatible, but a least square solution can be obtained by minimizing the sum of squares of the errors between the estimates and the (theoretically) correct values, i.e., its residuals. The system with the errors is given by

$$\begin{cases} \varepsilon_1 = f_3 - (a_1 f_2 + a_2 f_1 + a_3), \\ \varepsilon_2 = f_4 - (a_1 f_3 + a_2 f_2 + a_3), \\ \vdots \\ \varepsilon_{n-3} = f_n - (a_1 f_{n-1} + a_2 f_{n-2} + a_3). \end{cases} \quad (6)$$

Our problem, hence, is to minimize  $\sum_{k=0}^{n-3} \varepsilon_k^2$ .

It should be stressed that the original problem is linear, but, since the matrices of the system that have to be minimized are dynamic, we have a nonlinear LS problem to solve.

*2.4. Neural Network Fitting.* A neural network can be used for fitting the difference between the ARIMA estimate and the original data. That is, let  $\hat{L}$  be the ARIMA estimate of one measured dataset  $Z$ . We can write it as in

$$Z = \hat{L} + N. \quad (7)$$

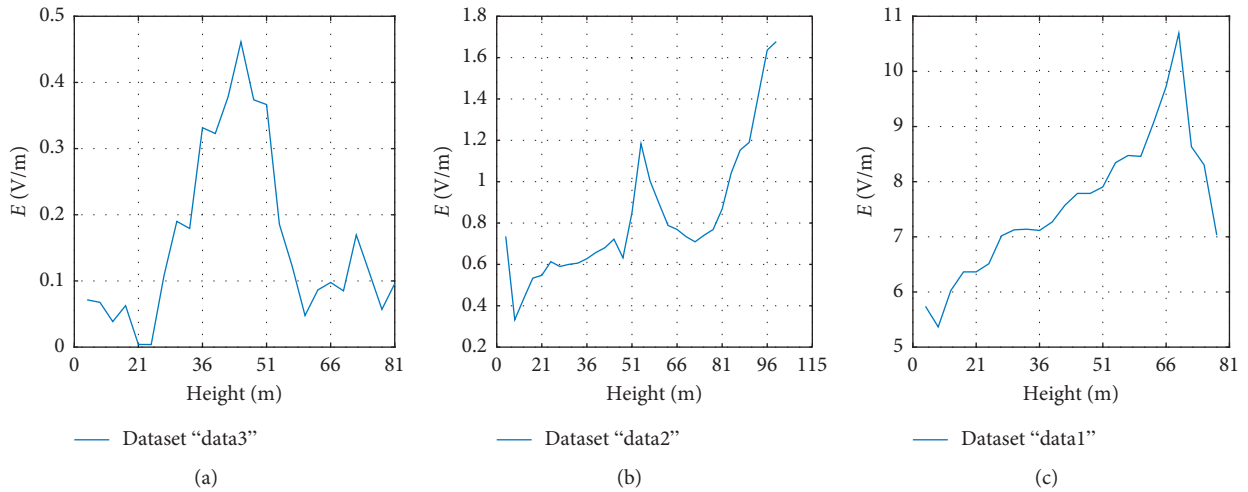


FIGURE 2: Graph for original (a) “data3” dataset; (b) “data2” dataset; (c) “data1” dataset.

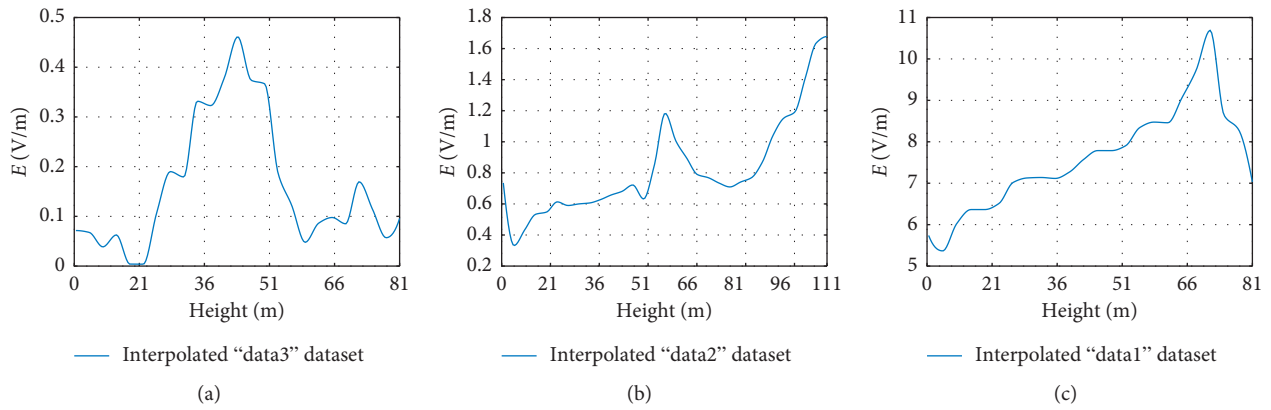


FIGURE 3: Graph for interpolated (a) “data3” dataset; (b) “data2” dataset; (c) “data1” dataset.

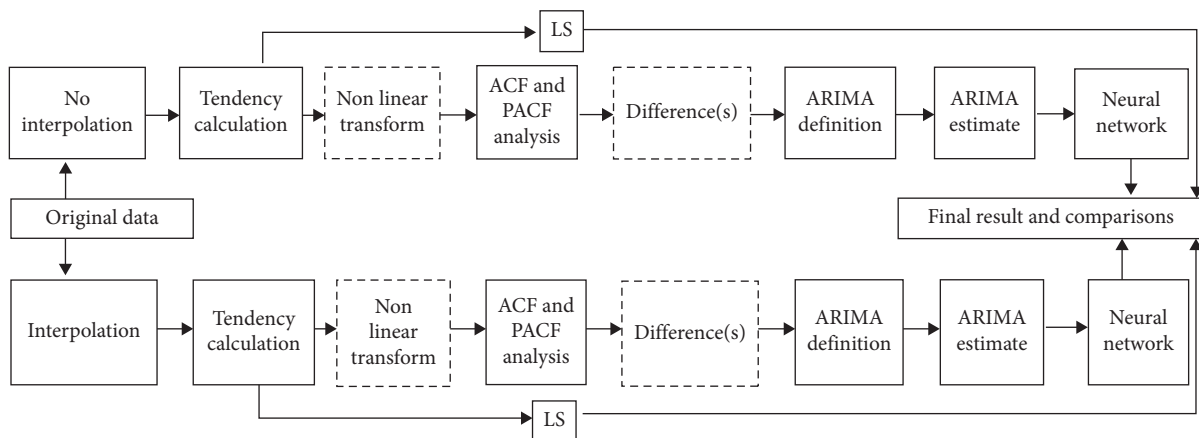


FIGURE 4: All steps in the fitting process.

In equation (7), the nonlinear term of  $Z$ , which will be fitted by the neural network, is represented by  $N$ .

In this study, a radial basis function neural network with two layers is employed. 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 normalizes the values previously found (see MATLAB documentation on *newgrnn* neural network [14]).



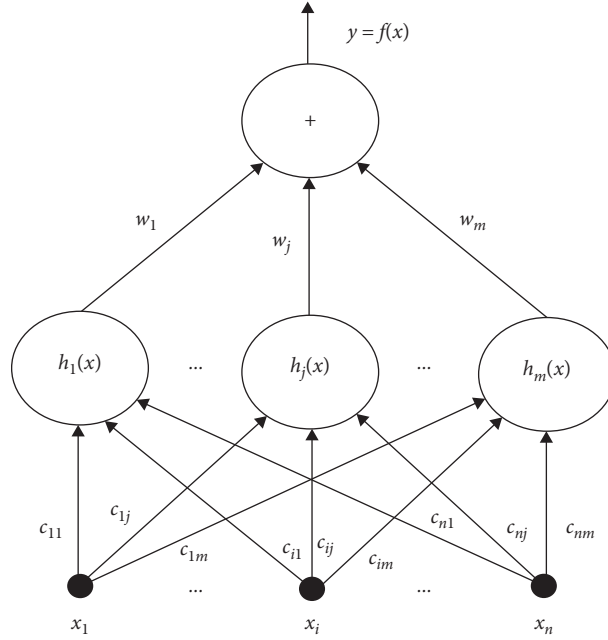


FIGURE 5: Radial basis neural network diagram.

The activation function of the neural network is a Gaussian function, given by equation (8). A general diagram of this kind of network is shown in Figure 5.

$$h_j(x) = \exp\left(-\frac{(x - c_j)^2}{2\sigma_j^2}\right), \quad j = 1, \dots, n, \quad (8)$$

with  $n$  being the number of inputs in the network. In this work, there is one neuron in the network for each training point. The number of training points vary from the original datasets and interpolated datasets.

From the diagram in Figure 5, we conclude that the function for the nonlinear estimate  $\hat{N}$  is given by

$$\hat{N} = \sum_{j=1}^n w_j \exp\left(-\frac{(x - c_j)^2}{2\sigma_j^2}\right), \quad j = 1, \dots, n, \quad (9)$$

with  $w_j$  being the weights of the neural network. The diagram of the representation of the network used in the original datasets fitting is shown in Figure 6.

In the original datasets, eight of the twenty-five original samples were used to train the network (as in Figure 6). With regard to the interpolated datasets, we used 24 of the 200 available samples. We proceeded in this way to avoid overtraining in the neural network. The boundary and the central samples are always used as training points. The other points are chosen at random. We used 1 as the spread value of the neural network (the standard value for MATLAB). 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. Finally, the estimated values from the neural network  $\hat{N}$  are then added to the estimated ARIMA values which gives the final model estimate for  $\hat{Z}$ , which is in

$$\hat{Z} = \hat{L} + \hat{N}. \quad (10)$$

### 3. Results and Discussion

**3.1. LS Fitting: Results of Original Samples.** We used the classical Levenberg–Marquardt [16] method to solve System 6. The values of the coefficients  $a_j$ ,  $j = 1, 2, 3$ , where  $a_1 = -0.1904$ ,  $a_2 = 1.0187$ , and  $a_3 = 0.0002$  for the original datasets. The Euclidean norm of residuals was 0.3302. The graphs of the LS curves for the three original datasets (calibration/training set and comparison sets) are shown in Figure 7. The relative and RMS errors values for this LS fitting are displayed on Table 3.

**3.2. LS Fitting: Results of Interpolated Samples.** By analogy, we used a similar polynomial as in equation (4) (recursive of the second order) for the interpolated datasets. This polynomial is represented by  $\tilde{f}_h$  (the symbol  $\sim$  in other variables through this article indicate they are originated from interpolated data). The analogous system to equation (6), after the matrices had been minimized, resulted in the following respective values of coefficients for  $\tilde{f}_h$ :  $\tilde{a}_1 = 0.9641$ ,  $\tilde{a}_2 = 1.9590$ , and  $\tilde{a}_3 = 0.0000$ .

The curves of each estimate for the interpolated datasets are displayed in Figure 8. The relative and RMS errors values for this LS fitting are given in Table 4.

We also tested the LS fitting using higher order polynomials. In light of the behaviour of the curves, this possibility is almost never considered. In the case of both the original and interpolated dataset curves, the second order LS fitting obtained good results. However, when the order of  $f$  was increased for the interpolated datasets, the LS could not find the optimal solution, no matter where the initial point was. Furthermore, comments on this will be made in Section 4.

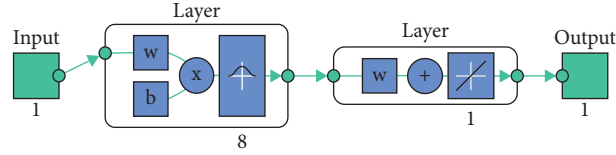


FIGURE 6: Neural network diagram for original training dataset [14].

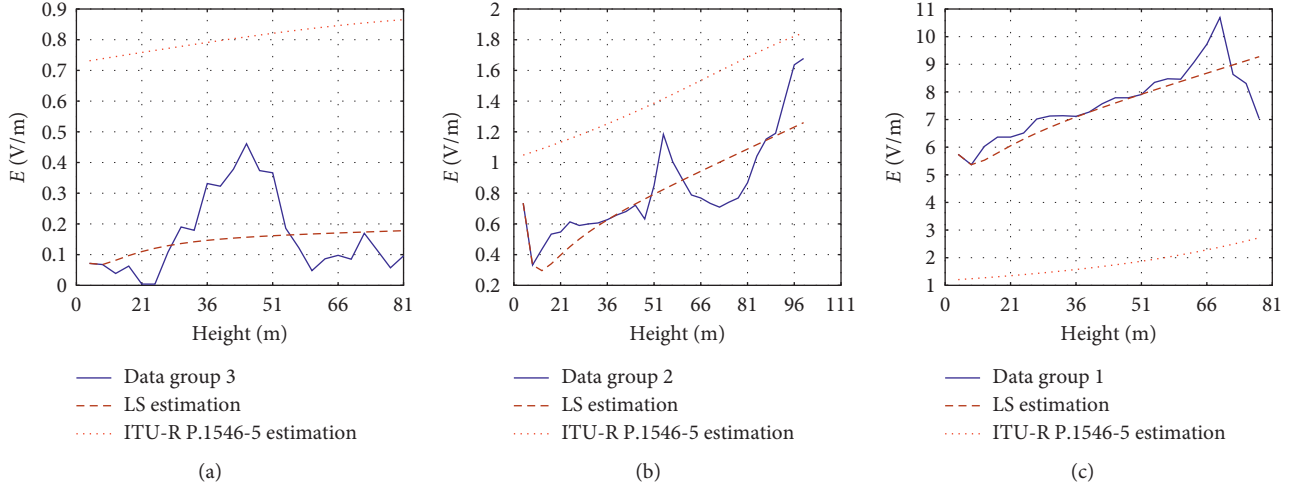


FIGURE 7: LS estimate for (a) “data3,” (b) “data2,” and (c) “data1.”

TABLE 3: Relative and RMS errors values for LS fitting of original datasets.

Dataset	RMS	Relative
“data1”	0.7096	0.0559
“data2”	0.1869	0.1662
“data3”	0.1247	2.8570

3.3. *ARIMA Fitting: Results of Original Samples.* Let  $Z_3$  be the mathematical notation for the original measured series of “data3” datasets. When analysing its “Amplitude vs. Mean” graph (see Figure 9), we decided not to take a logarithm transformation on  $Z_3$  in order to proceed with the ARIMA adjustment, even though the angular coefficient of the best linear fitting is not zero (as shown in the graph).

In view of the behaviour of all the measurement datasets acquired, we decided to calculate and isolate the tendency of all three series. In other words, fitting is carried out in the  $Z_3$  series without its tendency. The tendencies of  $Z_1$  and  $Z_2$  series are also isolated, since the model estimates for both datasets are made for these series without their tendencies (i.e., they are reintegrated after the estimates to allow comparisons to be made for the measured data). We examined the measured data without seasonal components and concluded that  $Z_3 = L_3 + N_3$  and  $L_3$  is

$$L_3 = T_3 + \alpha_3, \quad (11)$$

with  $T_3$  being the tendency for  $Z_3$  and  $\alpha_3$  the white noise.

We also took account of the polynomial trend line and estimated it by means of the linear least squares method,

which resulted in  $\hat{T}_3$ . Therefore, the series that must be estimated by the ARIMA model is represented by  $Y_3$ :

$$Y_3 = L_3 - \hat{T}_3. \quad (12)$$

In this case, the tendency is represented by a first degree polynomial. Coefficients of the polynomials representing the tendencies of  $Z_1$ ,  $Z_2$ , and  $Z_3$  are shown in Table 5. The studied series without tendency ( $Y_3$ ,  $Y_2$ , and  $Y_1$ ) are shown in Figure 10. The estimated series will be called  $\hat{Y}_j$ ,  $j = 1, 2, 3$ .

Now, we can analyse the ACF and PACF graphs for  $Y_3$  (this is the training set) to determine if their behaviour satisfies one ARIMA model. Figure 11 shows these graphs.

The main result at this stage is that the ACF plot moves to zero rapidly. This means that further transformation to the series is not necessary. The order of the ARIMA model can be defined.

The ACF function also moves rapidly towards zero and its behaviour is similar to a damped sinusoidal function, as well as the PACF graph. The ACF behaviour indicates that there is an autoregressive (AR) component in the model, and as it is infinite (i.e., exists in all lags), this is a sign that no moving average (MA) term is present. As PACF moves near zero after lag 1, we have a first-order AR model. It can thus be concluded that the best linear adjustment possible is an ARIMA (1,0,0) model. Its ACF and PACF are shown in Figure 12.

The adjusted/estimated ARIMA model is represented by

$$\hat{Y}_h = \phi_1 \hat{Y}_{h-1} + c, \quad (13)$$

with  $\phi = 0.8587$  and  $c = -0.0011$ .

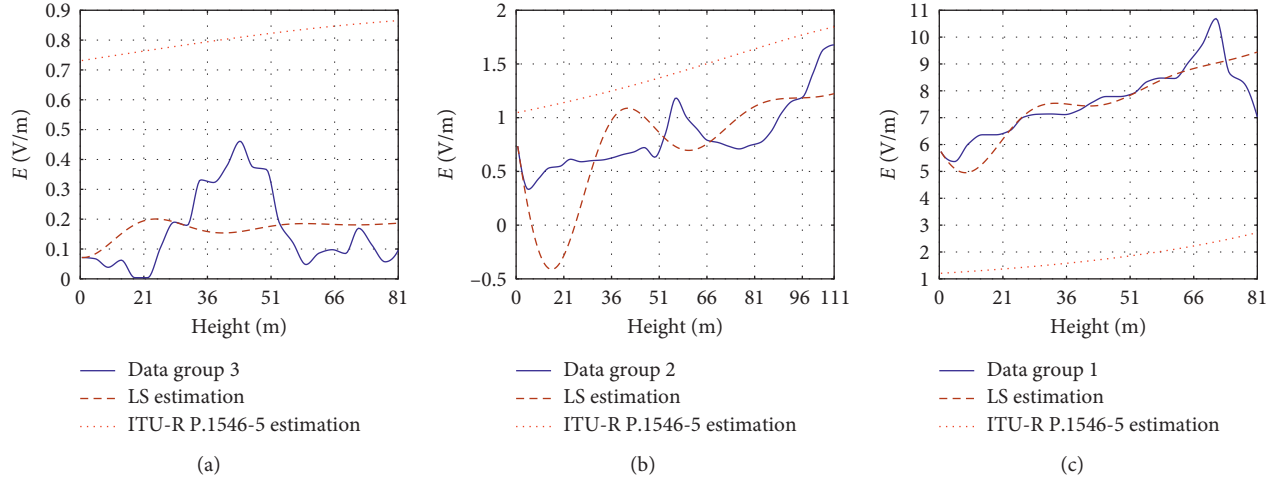
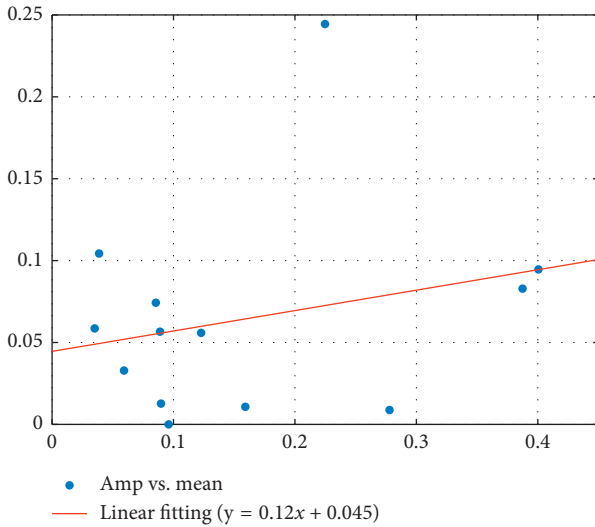


FIGURE 8: LS estimate for (a) “data3,” (b) “data2,” and (c) “data1.”

TABLE 4: Relative and RMS errors values for LS fitting of interpolated datasets.

Dataset	RMSE	Relative error
“data1”	0.6532	0.0591
“data2”	0.3954	0.4718
“data3”	0.1341	3.4521

FIGURE 9: Amplitude vs. mean graph for “data3” ( $Z_3$ ).TABLE 5: Linear tendency information  $y = ax + b$  coefficients  $a$  and  $b$  values.

Dataset	$a$	$b$
“data1”	5.7268	0.1479
“data2”	0.3731	0.0286
“data3”	0.1431	0.0014

The graph with the best adjustment for the “data3” dataset is shown in Figure 13. This is the graph that

originated from the estimation of equation (14) model when applied to its own adjusted dataset, i.e.,  $\hat{Z}_3$ :

$$\hat{L}_3 = \hat{Y}_3 + \hat{T}_3. \quad (14)$$

By analogy with  $\hat{L}_3$ , we can write  $\hat{L}_2$  and  $\hat{L}_1$ .

Figure 13 shows, as well, the graphs of the estimations of the ARIMA model to “data2” ( $\hat{L}_2$ ) and “data1” ( $\hat{L}_1$ ) datasets, respectively, that is, the comparative subscenarios. All these graphs also show the estimations of ITU-R P.1546-5 and LS for each subscenario. Tables 6–8 show the relative and RMS errors of both ARIMA, LS, and ITU estimations for every subscenario.

**3.4. ARIMA Fitting: Results of Interpolated Samples.** We have the variables  $\tilde{Z}_j$  and  $\tilde{Y}_j$ ,  $j = 1, 2, 3$ , representing the interpolated measured series with and without their tendencies, respectively.

The ARIMA adjustment process for the interpolated series gave, as its best result ( $\tilde{Z}_3$  is the training series) an ARIMA (4,0,0) model given by

$$\hat{Y}_h = \tilde{\phi}_1 \hat{Y}_{h-1} + \tilde{\phi}_2 \hat{Y}_{h-2} + \tilde{\phi}_3 \hat{Y}_{h-3} + \tilde{\phi}_4 \hat{Y}_{h-4} + \tilde{c}, \quad (15)$$

with  $\tilde{\phi}_1 = 3.1394$ ,  $\tilde{\phi}_2 = -3.7423$ ,  $\tilde{\phi}_3 = 2.0210$ ,  $\tilde{\phi}_4 = -0.4191$ , and  $\tilde{c} = 8.7149 \times 10^{-6}$ .

The graphs of the best ARIMA adjustment to  $\tilde{Z}_3$  and the comparisons to  $\tilde{Z}_2$  and  $\tilde{Z}_1$  are shown in Figure 14. The relative and RMS error for the interpolated datasets are displayed on Tables 9–11.

**3.5. Neural Network Fitting and Final Model Estimates: Original Samples.** Figure 15 shows the results for the adjusted final model versus the adjustment dataset (“data3”) and its estimates for the “data2” and “data1” datasets, respectively. These figures also show the ITU-R P.1546-5 and LS estimates for all the subscenarios. Tables 12–14 show the relative and RMS errors values of the combined model, the LS, and ITU estimates for all the datasets.



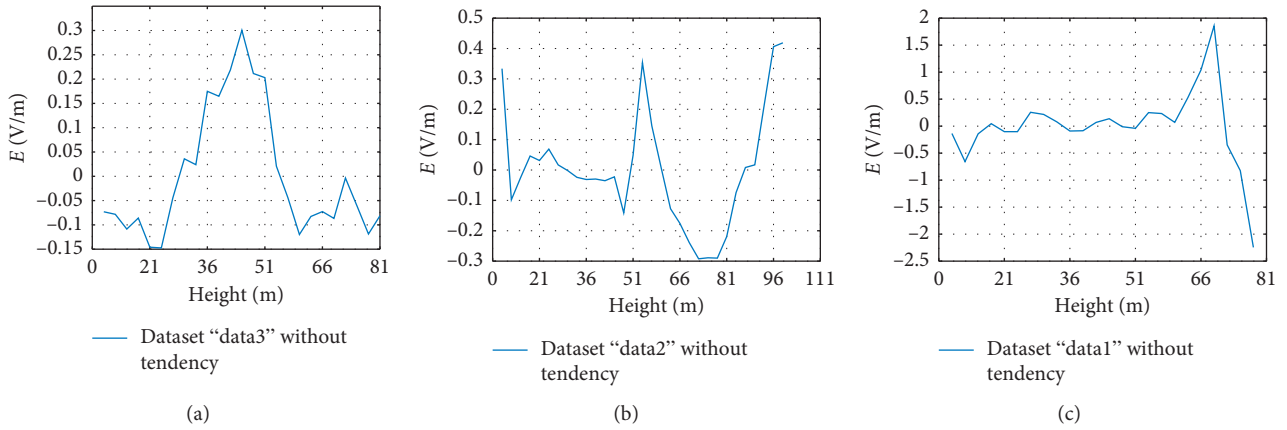


FIGURE 10: Graph of (a) “data3” without tendency ( $Y_3$ ), (b) “data2” without tendency ( $Y_2$ ), and (c) “data1” without tendency ( $Y_1$ ).

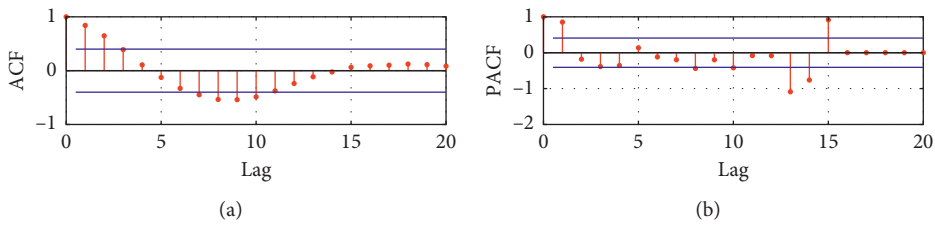


FIGURE 11:  $Y_3$  ACF and PACF graphs.

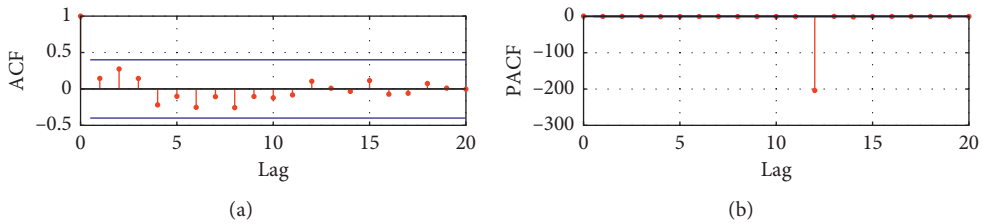


FIGURE 12: ARIMA (1,0,0) ACF and PACF.

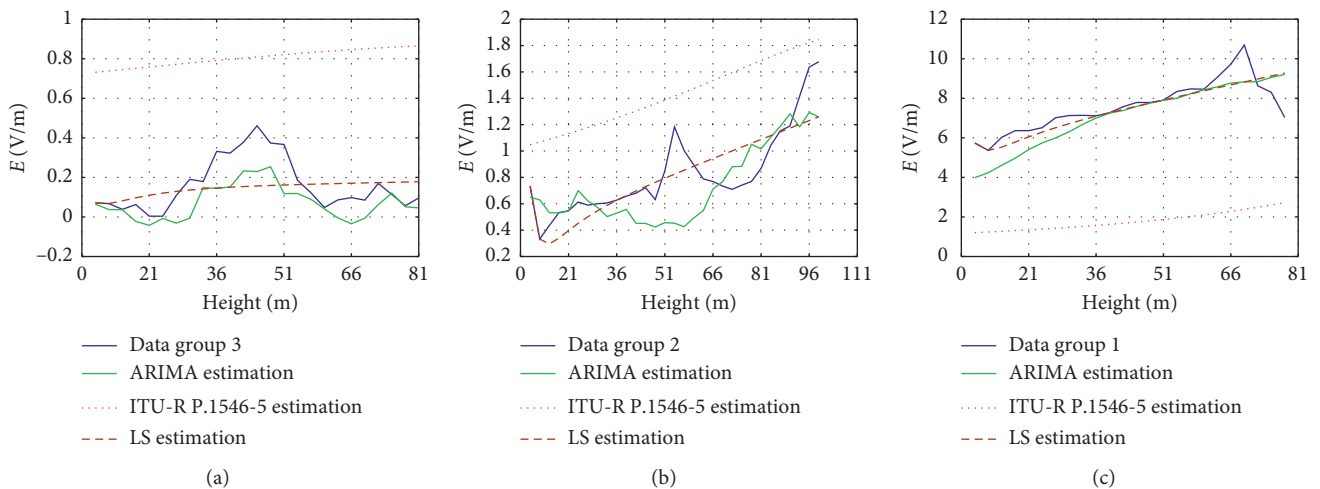


FIGURE 13: ARIMA estimation for (a) “data3” with tendency reintegrated ( $\hat{L}_3$ ), (b) “data2” with tendency reintegrated ( $\hat{L}_2$ ), and (c) “data1” with tendency reintegrated ( $\hat{L}_1$ ).

TABLE 6: ARIMA, ITU, and LS Relative and RMS errors for  $\hat{L}_3$ .

Dataset	RMSE	Relative error
ARIMA	0.0989	4.1562
LS	0.1247	2.8570
ITU	0.6577	21.4290

TABLE 7: ARIMA, ITU, and LS Relative and RMS errors for  $\hat{L}_2$ .

Dataset	RMSE	Relative error
ARIMA	0.1376	0.1496
LS	0.1869	0.1662
ITU	0.6279	0.8601

TABLE 8: ARIMA, ITU, and LS Relative and RMS errors for  $\hat{L}_1$ .

Dataset	RMSE	Relative error
ARIMA	0.9657	0.1090
LS	0.7096	0.0559
ITU	5.8504	0.7636

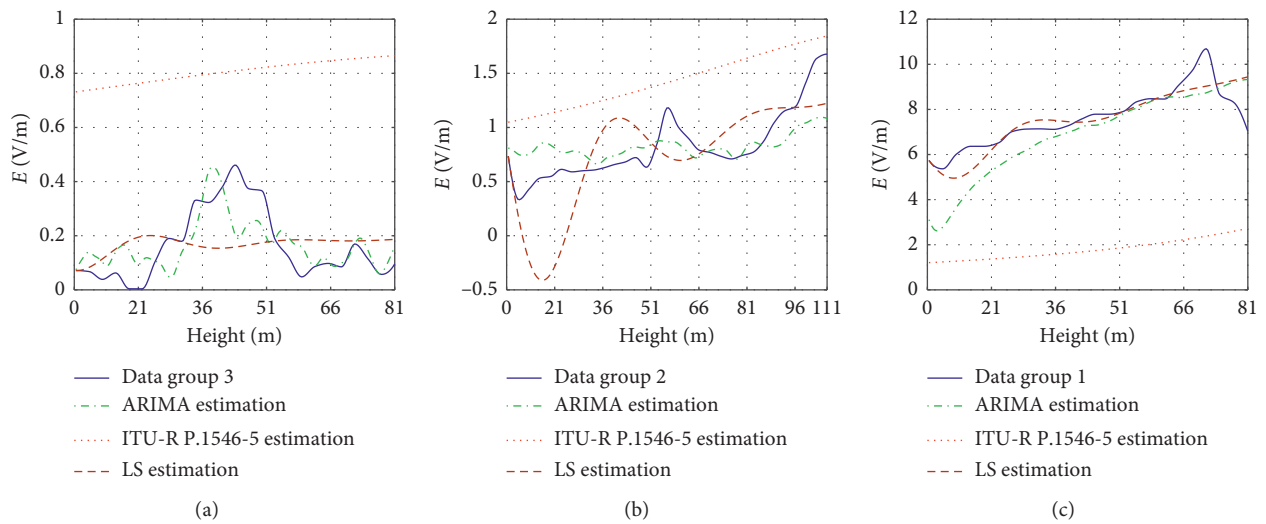


FIGURE 14: ARIMA estimation for (a) “data3” with tendency reintegrated ( $\hat{L}_3$ ), (b) “data2” with tendency reintegrated ( $\hat{L}_2$ ), and (c) “data1” with tendency reintegrated ( $\hat{L}_1$ ).

TABLE 9: ARIMA, ITU, and LS Relative and RMS errors for  $\hat{L}_3$ .

Dataset	RMSE	Relative error
ARIMA	0.0906	0.9199
LS	0.1341	3.4521
ITU	0.6555	17.0699

TABLE 10: ARIMA, ITU, and LS Relative and RMS errors for  $\hat{L}_2$ .

Dataset	RMSE	Relative error
ARIMA	0.2570	0.2975
LS	0.3954	0.4718
ITU	0.6356	0.8764

TABLE 11: ARIMA, ITU, and LS Relative and RMS errors for  $\hat{L}_1$ .

Dataset	RMSE	Relative error
ARIMA	1.1983	0.1358
LS	0.6532	0.0591
ITU	5.8982	0.7663

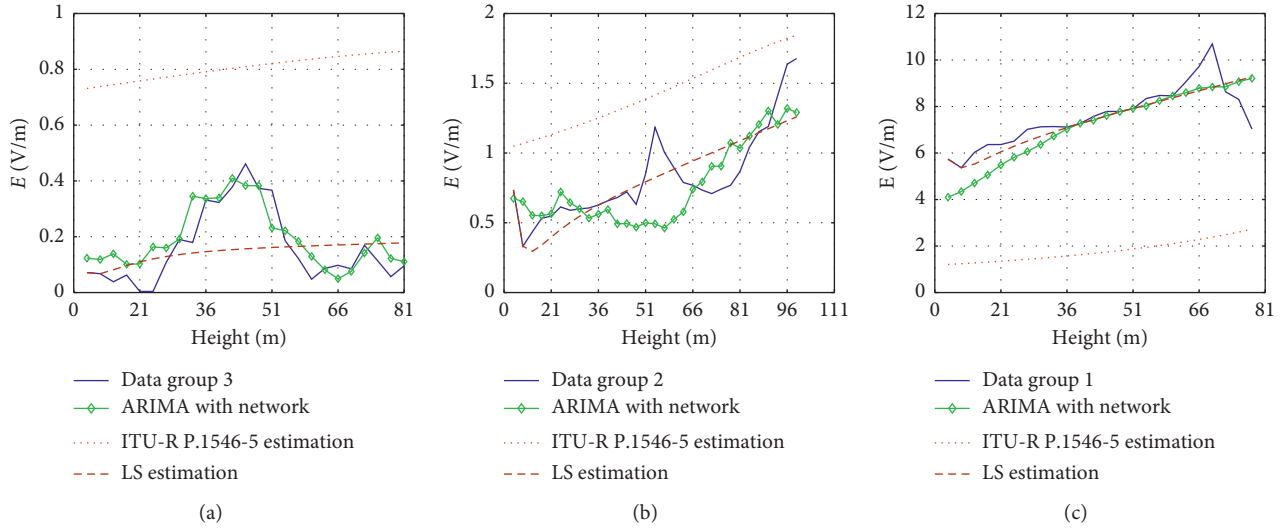


FIGURE 15: Comparison between measured data, ITU estimation, LS estimation, and proposed modelling curves for (a) “data3” subscenario, (b) “data2” subscenario, and (c) “data1” subscenario.

TABLE 12: ARIMA, ARIMA with neural network, LS, and ITU errors for “data3.”

Dataset	RMSE	Relative error
ARIMA	0.0989	4.1562
ARIMA w/Net	0.0752	2.7348
LS	0.1247	2.8570
ITU	0.6577	21.4290

TABLE 13: ARIMA, ARIMA with neural network, LS, and ITU errors for “data2.”

Dataset	RMSE	Relative error
ARIMA	0.1376	0.1496
ARIMA w/Net	0.1523	0.1582
LS	0.1869	0.1662
ITU	0.6279	0.8601

TABLE 14: ARIMA, ARIMA with neural network, LS, and ITU errors for “data1.”

Dataset	RMSE	Relative error
ARIMA	0.9657	0.1090
ARIMA w/Net	1.0097	0.1164
LS	0.7096	0.0559
ITU	5.8504	0.7636

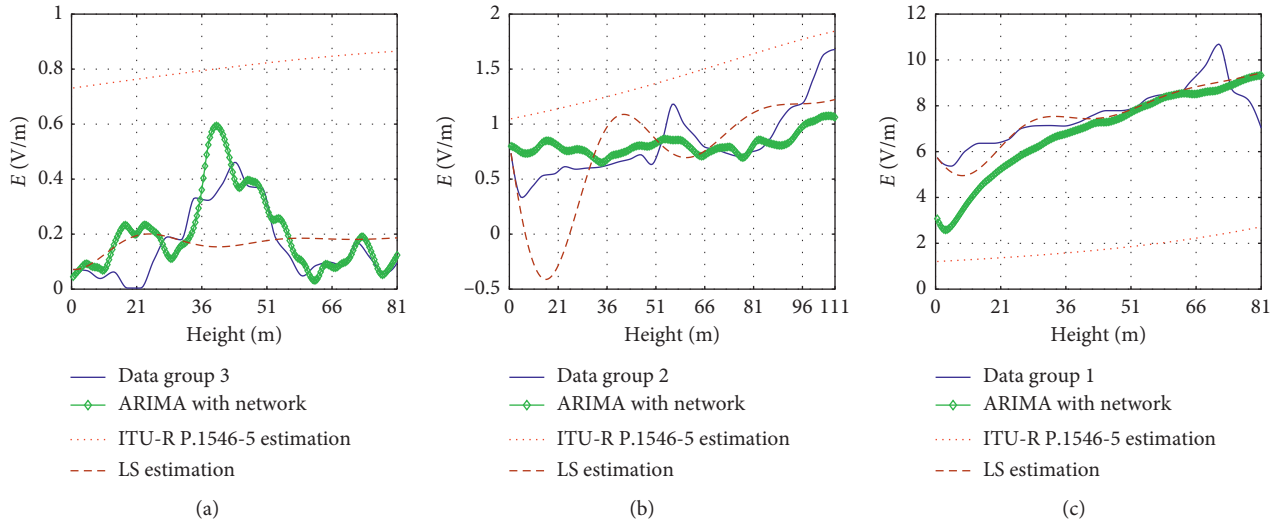


FIGURE 16: Comparison between measured data, ITU estimation, LS estimation and proposed modelling curves for (a) “data3” subscenario, (b) “data2” subscenario, and (c) “data1” subscenario.

TABLE 15: ARIMA, ARIMA with neural network, LS, and ITU errors for “data3.”

Dataset	RMSE	Relative error
ARIMA	0.0906	0.9199
ARIMA w/Net	0.0605	1.1375
LS	0.1341	3.4521
ITU	0.6555	17.0699

TABLE 16: ARIMA, ARIMA with neural network, LS, and ITU errors for “data2.”

Dataset	RMSE	Relative error
ARIMA	0.2570	0.2975
ARIMA w/Net	0.2564	0.2928
LS	0.3954	0.4718
ITU	0.6356	0.8764

TABLE 17: ARIMA, ARIMA with neural network, LS, and ITU errors for “data1.”

Dataset	RMSE	Relative error
ARIMA	1.1983	0.1358
ARIMA w/Net	1.2206	0.1392
LS	0.6532	0.0591
ITU	5.8982	0.7663

3.6. *Neural Network Fitting and Final Model Estimates: Interpolated Samples.* By analogy, Figure 16 shows the results for the fitting and estimates in the interpolated datasets and their respective relative and RMS errors are shown in Tables 15–17.

## 4. Conclusions

4.1. *General Public Exposure to NIR.* With regard to the general public exposure to NIR, the values found for the electric field do not exceed the limits currently in operation in

Brazil. On the contrary, they were significantly below the limits referred to in the frequency that was analysed (33.6805 V/m for electric field at 600 MHz). This means that the ER ratio value is less than 1, which can be regarded as a safe value by Brazilian standards. The greatest value of the ER ratio in all the scenarios (taking account of both the original and interpolated sets) was 0.3174 and was calculated through the ratio between the greatest value measured for the “data 3” dataset (10.6899 V/m) and the limit of 33.6805 V/m.

By using the planned modelling, any similar requirement that might arise in the future does not necessarily involve the

need for measurements since the model represents, to a satisfactory extent, the electric field propagation in the scenario and in the frequency in question. However, in a situation where the estimates of the model are as close as (or even closer to) the exposure limits than the error of the model (whether RMS or relative) with regard to the training set, there is a need for measurements and a cautious approach to the situation. Without measurements being required in every situation that raises doubts among the public, the process of evaluating the exposure to NIR would be less expensive, faster, and even more efficient. This helps in providing a precise and rapid response and in improving the service rendered by the regulatory agency.

*4.2. Original Datasets.* On the basis of the results of the 25 sample datasets (not interpolated), we observed that our modelling system shows a significantly better result than the commonly used ITU-R estimates. In the case of the “data1” group, where the estimates of the model were worst, it was still approximately 65% better than its ITU estimate (taking into account the relative errors). It can, thus, be concluded that the hybrid modelling achieves its goals with a high degree of accuracy.

However, when compared with the “single” ARIMA estimates, the combined ARIMA and neural network model achieves slightly worse results. This suggests that the additional feature (combining the ARIMA model with a neural network), in this particular problem, may not be necessary to represent it in a precise way.

Although there is this slight difference between the “single” and “combined” ARIMA models (as confirmed by the relative/RMS error values), we think that the combined model should be used in all cases, in view of its applicability to other scenarios. In the case of this problem, for example, in our view, both approaches (“single” and “combined” ARIMA) are equivalent, as the difference in the errors is slight. The largest RMSE difference in all the tested scenarios is approximately 0.1 for the “data3” dataset without interpolation.

Since this work comprises a case study, we think it should be stressed that the “single” ARIMA model gave the best results in this particular scenario. We expect, however, that in other scenarios, the combined model may become necessary, as it can represent nonlinear features, whereas the “single” ARIMA modelling cannot. When compared with the LS estimates (except for the “data1” dataset), the combined modelling system achieved better results.

*4.3. Interpolated Datasets.* On the basis of an analysis of the results obtained from 200 samples (interpolated datasets), the same conclusions can be reached as with results from the 25 samples, but with a caveat; although the LS was able to find an optimal solution to this problem, owing to the number of samples there is a risk that this may not be feasible in other scenarios. There may be too many alternatives for the LS algorithm to search and its result may “explode.” This also suggests that, in a desired scenario of “nonstop” measurements, when a higher number of samples is

naturally acquired, applying the classic LS fitting may not be the best alternative, since there is risk that an optimal solution will not be found. Nonetheless, the combined ARIMA modelling system (or even the “single” one) does not have this limitation and the neural network is able to model nonlinear features that the “single” ARIMA model cannot.

## 5. Future Improvements on the Model

As future improvements to this study, we recommend the following:

- (1) Including other scenarios and frequencies through measurement campaigns: this will enable the proposed modelling to tackle a wider range of problems, i.e., help to generalize the model
- (2) Adding datasets from near-field measurement campaigns: as it is, the proposed model cannot predict near-field propagation situations, since the training data consists of far-field measurement information
- (3) Implementing this modelling system through the deployment of an applicative service on a mobile device, such as a smartphone: we believe this applicative could be used by the general public to estimate electric field intensity in similar scenarios to the one(s) studied here, when doubts about the exposure to NIR arise

## Data Availability

The .xlsx (“data1,” “data2,” and “data3” datasets) and .txt (instructions on how to use the .xlsx archives) data used to support the findings of this study are included within the supplementary information file.

## 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 Brazilian National Agency of Telecommunications, by means of measurement equipment and personnel. The authors are grateful to the Brazilian Coordination of Superior Level Staff Improvement (CAPES) for its financial support. The authors would also like to thank the Federal University of Pará (UFPA) for the technical assistance given to the research undertaken in this paper.

## 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 provided. The datasets are archives in .xlsx. The instructions on how to use them is described in a .txt archive. (*Supplementary Materials*)



## References

- [1] E. Hamiti, M. Ibrani, L. Ahma, S. Dragusha, and R. Halili, "Comparative analysis of personal exposure levels induced by long-term evolution 1800 re-farming and other RF sources in an urban environment," *IET Microwaves, Antennas and Propagation*, vol. 12, no. 7, pp. 1185–1190, 2018.
- [2] M. Koprivica, V. Slavkovic, N. Neskovic, and A. Neskovic, "Statistical analysis of electromagnetic radiation measurements in the vicinity of GSM/UMTS base station installed on buildings in Serbia," *Radiation Protection Dosimetry*, vol. 168, no. 4, pp. 489–502, 2016.
- [3] J. H. Bernhardt, "Non-ionizing radiation safety: radio-frequency radiation, electric and magnetic fields," *Physics in Medicine and Biology*, vol. 37, no. 4, pp. 807–844, 1992.
- [4] D. Belpomme, L. Hardell, I. Belyaev, E. Burgio, and D. O. Carpenter, "Thermal and non-thermal health effects of low intensity non-ionizing radiation: an international perspective," *Environmental Pollution*, vol. 242, pp. 643–658, 2018.
- [5] A. Kuzniar, C. Laffeber, B. Eppink et al., "Semi-quantitative proteomics of mammalian cells upon short-term exposure to non-ionizing electromagnetic fields," *PLoS One*, vol. 12, no. 2, Article ID e0170762, 2017.
- [6] N. Bernabò, R. Ciccarelli, L. Greco, A. Ordinelli, M. Mattioli, and B. Barboni, "Scientometric study of the effects of exposure to non-ionizing electromagnetic fields on fertility: a contribution to understanding the reasons of partial failure," *PLoS One*, vol. 12, no. 12, Article ID e0187890, 2017.
- [7] K. W. Wang, C. Deng, J. P. Li, Y. Y. Zhang, X. Y. Li, and M. C. Wu, "Hybrid methodology for tuberculosis incidence time-series forecasting based on ARIMA and a NAR neural network," *Epidemiology & Infection*, vol. 145, no. 6, pp. 1118–1129, 2017.
- [8] H. Moeeni, H. Bonakdari, and I. Ebtehaj, "Integrated SAR-IMA with neuro-fuzzy systems and neural networks for monthly inflow prediction," *Water Resources Management*, vol. 31, no. 7, pp. 2141–2156, 2017.
- [9] K. Naveena, S. Singh, S. Rathod, and A. Singh, "Hybrid ARIMA-ANN modelling for forecasting the price of robusta coffee in India," *International Journal of Current Microbiology and Applied Sciences*, vol. 6, no. 7, pp. 1721–1726, 2017.
- [10] G. P. Zhang, "Time series forecasting using a hybrid ARIMA and neural network model," *Neurocomputing*, vol. 50, pp. 159–175, 2003.
- [11] I. Khandelwal, R. Adhikari, and G. Verma, "Time series forecasting using hybrid ARIMA and ANN models based on DWT decomposition," *Procedia Computer Science*, vol. 48, pp. 173–179, 2015.
- [12] C. H. Aladag, E. Egrioglu, and C. Kadilar, "Forecasting nonlinear time series with a hybrid methodology," *Applied Mathematics Letters*, vol. 22, no. 9, pp. 1467–1470, 2009.
- [13] ANATEL, "Resolution 700/2018—regulation on the assessment of human exposure to electrical, magnetic and electromagnetic fields associated with operation of radiocommunication transmitting stations," 2018, <https://www.anatel.gov.br/legislacao/resolucoes/2018/1161-resolucao-700>.
- [14] MATLAB Version 8.0.0.783. In Technical Documentation. 2012.
- [15] G. E. P. Box-Jenkins, *Time Series Analysis: Forecasting and Control*, John Wiley and Sons, Hoboken, NJ, USA, 4th edition, 2008.
- [16] J. J. Moré, "The Levenberg-Marquardt algorithm: implementation and theory," in *Lecture Notes in Mathematics*, G. A. Watson, Ed., vol. 630, Springer, Berlin, Germany, 1978.