A Neural Network Propagation Model for LoRaWAN and Critical Analysis with Real-World Measurements

Salaheddin Hosseinzadeh 1, Mahmood Almoathen 2, Hadi Larijani 1,* and Krystyna Curtis 1

1 School of Engineering and Built Environment, Glasgow Caledonian University, Glasgow G4 0BA, UK; Salaheddin.Hosseinzadeh@gcu.ac.uk (S.H.); krystyna.curtis@gcu.ac.uk (K.C.)
2 Computer Technology Department, Qatif College of Technology, Al Qatif 32636, Saudi Arabia; malmoathen@tvtc.gov.sa
* Correspondence: H.Larijani@gcu.ac.uk

Received: 8 October 2017; Accepted: 9 December 2017; Published: 14 December 2017

Abstract: Among the many technologies competing for the Internet of Things (IoT), one of the most promising and fast-growing technologies in this landscape is the Low-Power Wide-Area Network (LPWAN). Coverage of LoRa, one of the main IoT LPWAN technologies, has previously been studied for outdoor environments. However, this article focuses on end-to-end propagation in an outdoor−indoor scenario. This article will investigate how the reported and documented outdoor metrics are interpreted for an indoor environment. Furthermore, to facilitate network planning and coverage prediction, a novel hybrid propagation estimation method has been developed and examined. This hybrid model is comprised of an artificial neural network (ANN) and an optimized Multi-Wall Model (MWM). Subsequently, real-world measurements were collected and compared against different propagation models. For benchmarking, log-distance and COST231 models were used due to their simplicity. It was observed and concluded that: (a) the propagation of the LoRa Wide-Area Network (LoRaWAN) is limited to a much shorter range in this investigated environment compared with outdoor reports; (b) log-distance and COST231 models do not yield an accurate estimate of propagation characteristics for outdoor−indoor scenarios; (c) this lack of accuracy can be addressed by adjusting the COST231 model, to account for the outdoor propagation; (d) a feedforward neural network combined with a COST231 model improves the accuracy of the predictions. This work demonstrates practical results and provides an insight into the LoRaWAN’s propagation in similar scenarios. This could facilitate network planning for outdoor−indoor environments.

Keywords: LoRaWAN; LPWAN; propagation analysis and modeling; feedforward neural networks; COST231 multi-wall model

1. Introduction

The low-power wide-area network technology has several advantages, such as low price and low power consumption. LoRa and SigFox are two of the main Internet of Things (IoT) Low-Power Wide-Area Network (LPWAN) technologies [1]. LoRa has more advantages, in addition to having a low power consumption and long range of coverage. One of its features is a proprietary chirp spread spectrum (CSS) modulation, which is resistant to interference and the Doppler effect [2]. The CSS modulation also improves the sensitivity of the device and increases the overall available link budget [3]. It is also internet protocol version 6 (IPv6) compatible; hence, it is capable of providing better security, scalability, and end-to-end connectivity [2,4]. Unlike SigFox, the LoRa alliance is not an IoT network provider; therefore, no subscription is required, and there are no uplink or downlink restrictions regarding the number of messages per day [5–7]. The aforementioned advantages made LoRa favorable for the purpose of this research. Currently, Semtech is the main manufacturer of LoRa modules. LoRa commonly refers to a physical layer using the CSS, and LoRaWAN is an open standard
Media Access Control (MAC) layer protocol developed by the LoRa Alliance, which allows end devices to gateway communication.

The outdoor propagation of LoRa and LoRaWAN has been studied in [8–14]. These studies were mainly focused on the: (a) coverage range of line of sight (LoS), (b) signal strength in central business districts, and (c) impact of modulation spreading factor on its propagation. The propagation ranges in these studies are therefore reported from two to 20 km, which is much lower than 50 km [15]. There have been a few studies where point-to-point LoRa communication has been investigated in indoor environments for the purpose of network sensor implementation [16,17]. In this study, an outdoor–indoor scenario is investigated with an end-to-end connection, to observe the structural penetration of LoraWAN and its gateway. Therefore, practical measurements were collected, analyzed, and compared against common propagation models. These models include a log-distance, COST231 Multi-Wall Model (MWM).

In addition to these models, artificial neural network (ANN) models are used to estimate the propagation. Since hybrid models are based on an ANN, they are capable of learning the propagation. Learning is achieved through a training process, where the ANN is exposed to sets of input–output parameters. ANN models are more accurate and less computationally demanding compared with non-deterministic and deterministic models, respectively [18]. However, one disadvantage of the ANN models is that they require a considerable amount of data collection over a vast area for the purpose of training, validation, and testing. For instance, in [19], 600,000 indoor data samples were collected, and learning took several hours. Authors in [18–25] used ANNs to predict the propagation in indoor environments. In these studies, the main ANN inputs were the: distance between transceivers, number of walls, number of doors, number of windows, frequency of transmission, antenna gains, and even transmission power. However, two of the authors [20,22] used “free space path attenuation” as a single input. This is an intuitive preprocessing of two individual inputs of distance and frequency, by taking advantage of the free space path loss formula. This approach (a) reduces the number of inputs to the ANN, which should help the learning by reducing the number of parameters; (b) makes the network needless of learning the basic propagation principles, such as calculating space path loss. A similar approach can also be applied to unify all of the attenuating factors such as the number of walls, windows, and doors or transmission power and antenna gain, which can form the effective radiated power.

In these studies, the ANN had the responsibility of inferring the relations between the propagation parameters. This in turns requires comprehensive training with several measurements. However, not only collecting several measurements in a relatively small indoor environment can be tedious, but it also defeats the purpose of estimation. To facilitate data collection and yet also be able to improve predictions, COST231 is utilized as the first stage of modeling, where its results were used to train the ANN toward acquiring better estimations. We therefore did not directly use these physical parameters to train the ANN.

This article is arranged as follows: in Section 2, the measurement setup and data collection are explained; Section 3 explains the modeling and the optimization; Section 4 demonstrates the results; and Section 5 concludes the outcomes of this research.

2. Data Collection Setup

The mobile device used during the measurement was comprised of a Multitech mDot module [26], which is controlled by a Raspberry pi single board computer. The Kerlink gateway was equipped with LoRa SX1301, a mobile network, and microprocessors [27]. The gateway was located on the rooftop of the George Moore Building at Glasgow Caledonian University (GCU), which is about 27 m high. The mobile device was configured to regularly transmit and receive a sequentially increasing message. This sequential increment also helped to identify any lost messages during the data analysis. Data was collected on the eighth and seventh floor of the Hamish Wood Building at GCU, which has a height of 27 m, length of 60 m, and width of 22.5 m. To cover the entire building, the mobile device
was moved to several different locations inside the building. The equivalent isotropic radiation power and frequency of transmission were 14 dBm and 867.1 MHz, respectively. Other configurations were a spreading factor of nine, a bandwidth of 125 kHz, and antenna gains of 2 dBi. Nearly 10 received signal strength indicators (RSSI) were collected at each location. These RSSI values were logged both on the network server and locally on the mobile device. Figure 1 demonstrates the test environment.

![Image of Hamish Wood Building and George Moore Building](image)

**Figure 1.** Studied environment at Glasgow Caledonian University (GCU).

The internal structure of the eighth floor is presented in Figure 2. The mobile LoRaWAN transceiver was moved to sampling locations marked from 1 to 27 to cover the entire area, and the averaged RSSI was calculated at each location. The location numbers and averaged RSSIs are indicated in Figure 2, where they are separated by a comma. There are also a total of 43 walls on each floor; however, most of the walls along the width of the building were not blocking the LoS due to the positioning of the LoRaWAN gateway. The gateway was mounted on the rooftop of the George Moore Building, which is 55 m away from the first obstructing wall, that is indicated by a dashed-line in Figure 2.

![Image of Indoor structural layout of eighth floor](image)

**Figure 2.** Indoor structural layout of eighth floor, sampling locations, and their corresponding received signal strength indicators (RSSI).

3. Propagation Modeling

All of the models were optimized by using the collected practical measurements. This was done in order to find the propagation characteristics of the environment and benchmark the performance of the LoRaWAN. The models used in this research were log-distance, the indoor COST231 (multi-wall) model, a proposed variation of the COST231, and the hybrid model, which is comprised of the alternated COST231 and an ANN. These models are briefly discussed in this section.
3.1. Log Distance

Log distance is a commonly used model for both indoor and outdoor environments. It has a few parameters to characterize the propagation, and therefore, it is relatively easy to implement. Although it lacks the accuracy of sophisticated models, it is not computationally demanding [28]. The log-distance model is presented in Equation (1) [29].

\[ PL = PL_{d_0} + 10n \log \left( \frac{d}{d_0} \right) + X_g \]  

(1)

where \( PL \) is the path loss (dB), \( d \) is the distance (meter) between the transceivers, \( n \) is the path loss exponent, \( X_g \) is the shadow fading with mean (\( \mu \)) zero and standard deviation \( \sigma \) (dB), and \( PL_{d_0} \) is the loss at a reference distance (usually 1 m) from the transmitter.

3.2. COST231

One of the frequently used propagation models for indoor environments is the COST231, which is formulated in Equation (2) [30]. Implementation of this model is presented in [31,32]. Unlike the log-distance model, the COST231 model requires site-specific information, such as the layout of the indoor environment; as a result, it offers better accuracy [33].

\[ PL = L_{FS}(d) + L_C + \sum \alpha_j k_j \]  

(2)

where \( L_C \) is the constant loss (dB), \( L_{FS}(d) \) is the free space path loss at distance \( d \) between transceivers, and \( \alpha_j \) and \( k_j \) are the attenuation coefficient and the number of walls of type \( j \) between the LoS.

3.3. Adjusted COST231

The testing environment in this research consists of both indoor and outdoor propagation. Whereas the COST231 model is only intended for indoor modeling, as it has an optimistic space path loss prediction, it only accounts for the free space loss. Hence, we made an adjustment to the COST231 model to account for the outdoor propagation parameters. Since this adjustment is inspired by the log–distance model, we added a path loss exponent \( (n_{A231}) \) to the free space path loss \( (L_{FS}) \). This adjusted model is presented in Equation (3). All of the other parameters in Equation (3) are defined similar to Equation (2).

\[ PL = n_{A231} \times L_{FS}(d) + L_C + \sum \alpha_j k_j \]  

(3)

3.4. Optimization

Since there are multiple locations on each floor at which the RSSI is measured, a multi-objective min–max optimization is used to find the optimum model parameters. For the log-distance model, the overall fading \( (X_{gi}) \) at each location \( (i) \) is optimized as an independent variable. Constant loss \( (L_E) \) is also added to the optimization functions, which compensates for any possible additional losses in the system such as cable loss, antenna mismatch, or polarization mismatch. Therefore, \( L_E \) and \( n/n_{A231} \) are not changing in any of the objective functions. The overall form of the objective functions and their constraints are formulated for the log-distance model in Equation (4). Where \( M_i \) is the summation of transmission power, antenna gains, and the measured RSSI at location \( i \)th \( (M_i < 0) \). A similar optimization is also applied to COST231 and its adjusted variation, except \( F_i \) in the adjusted COST231 is a function \( n_{A231} \), \( L_C \) and \( \alpha_j \).

\[
\min_{n, X_{gi}} \max_{L_E, F_i} F_i(n, X_{gi}, L_E, F_i) \quad \left\{ \begin{array}{l}
1 \leq n \leq 8 \\
-20 \leq L_E \leq 20 \\
-20 \leq X_{gi} \leq 20
\end{array} \right.
\]

(4)

\[ F_i(n, X_{gi}, L_E) = -PL_0 - M_i - L_E - 10n \times \log (d) - X_{gi} \]
3.5. Artificial Neural Network

The feedforward neural network model consists of an input layer, two hidden layers, and an output layer. The five inputs of the network are:

- Total number of walls blocking the LoS ($\sum k_i$)
- Total attenuation caused by the blocking walls ($\sum a_i k_i$)
- Attenuation caused by distance and loss exponents ($n_{A231} \times L_{FS}(d)$)
- Set of $(x_i, y_i)$ coordinates to specify the measurement location inside the building

The first and second hidden layers consist of seven and three neurons, respectively. The number of layers and the size of each layer were heuristically found to avoid overfitting, while generalizing the problem. The only output of the network is the error between the optimized or best estimate of the adjusted COST231 model ($F_{A231}$) and the measured RSSI. The transfer functions are linear for the first hidden layer, and hyperbolic tangent sigmoid [34] for the second and the output layer. Also, the Levenberg–Marquardt backpropagation algorithm [34] is used for training. An overall schematic of this implemented network in MATLAB R2016b is depicted in Figure 3, where “w” and “b” are indicating the weights and biases of each layer of neurons, respectively.

![Figure 3. Schematic of the implemented neural network in MATLAB.](image)

4. Data Analysis Result

The model parameters that are derived from the optimization identify the characteristics of the LoRaWAN propagation.

4.1. Log-Distance

For the log-distance model parameters, the path loss exponent ($n$) derived as 3.9, and $L_E$ was found as 3.4 dB. The mean and standard deviation of shadow fading ($X_g$) were found as $\mu = 0.25$ and $\sigma = 6.8$ dB, respectively. These extracted parameters resemble the propagation in an “office with hard partitions” at 1.5 GHz [35]. It is necessary to emphasize that these tests were conducted at 867 MHz.

4.2. COST231

For COST231 models, 25 of the walls that are aligned parallel to the LoS (west to east orientation) are not involved in the optimization, as they are not blocking the LoS. The constant loss is found as $L_c = 17.32$ dB. The effective attenuations of the walls ($a_i$) are plotted in Figure 4. Attenuations ranged from 12.79 to 1 dB. These attenuations were mainly in agreement with the values reported in [36]. However, $k_1$, which is the first penetrating external wall and has several large windows, has a relatively much higher attenuation (12.79 dB) compared with the other walls. This is further investigated below.
4.3. Adjusted COST231

Since the measurement was conducted in an outdoor–indoor environment, a considerable part of the propagation was outdoors. Therefore, a similar path loss exponent \( n_{A231} \) was added to the COST231 indoor model to account for the outdoor propagation losses. For the adjusted model, the constant loss was found as \( L_e = 4.1 \text{ dB} \), and path loss exponent \( n_{A231} = 2.4 \). Attenuation of the walls for both the COST231 and the adjusted model are depicted and compared in Figure 4. For the adjusted model, these attenuations are range from 1 dB to 8 dB.

4.4. Hybrid Model

Since the adjusted COST231 model produced a more realistic and accurate estimation, its parameters are used as the inputs of the neural network. The total number of the LoS-blocking walls is extracted automatically during the execution of the COST231 model. However, the total attenuation of these walls are identified after the optimization of the COST231 model, which serves as the second input to the ANN. The third input is also calculated after the optimization of the path loss exponent \( n_{A231} = 2.4 \), which is a fixed parameter, while \( L_{FS}(d) \) is changing from one location to another. To consider the spatial influence of the building on propagation, such as fixtures and other clutter that is not included in the model, the fourth and fifth inputs are the position of the measurement locations inside the building. This \((x_i, y_i)\) pair is also extracted from the blueprint image of the COST231 model automatically. The schematics of the hybrid model during and after the training are depicted in Figure 5.

![Figure 5. Flowchart of the hybrid model, (a) depicting the model input/output during the training; (b) demonstrating the input/output after the artificial neural network (ANN) is trained.](image)

The network is then trained using all of the collected data on the eighth floor. Implementing the ANN with the adjusted COST231 assistance drastically facilitated the training process and yet increased the accuracy of the prediction. The desired output of the network was the difference between the optimized adjusted COST231 prediction \( F_{A231} \) and the RSSI measurement on the eighth floor; in other words, \( E_{A231} = \text{RSSI} - F_{A231} \). Therefore, the only responsibility of the network was to discern the deficiencies of the adjusted COST231 model, rather than learn the whole propagation mechanism. For the seventh floor, the ANN made an estimation \( T_{E_{NN}} \) of the adjusted COST231 error \( E_{A231} \).
which is depicted in Figure 6a. Figure 6b compares both of the optimized estimations of the adjusted COST231 ($F_{A231}$) and assisted ANN models ($F_{A231} + T_{ANN}$) with the practical measurements (RSSI). This is further explained in the discussion.

![Graphs showing comparison of COST231 and ANN estimations with practical measurements.](image)

**Figure 6.** (a) Comparison between the COST error ($E_{A231}$) and the ANN estimation of the COST error ($T_{ANN}$); (b) Comparison between measurement of the received signal strength indicators (RSSI), COST’s optimized estimation ($F_{A231}$), and compensated COST by ANN.

5. Discussion

After optimization, the resulted mean square errors (MSE) for the log-distance, COST231, and the adjusted COST231 models were derived as 45, 20.47, and 21.83, respectively. As expected, the log-distance model did not have the accuracy of the site-specific models. For the COST231 model and its adjusted variants, the difference in the MSE was almost negligible, and they performed the same in terms of propagation estimation.

It is observed that the major difference between the results of COST231 and its adjusted variation was the attenuation coefficient of the first penetrating wall ($a_1$). The COST231 model did not sufficiently account for the outdoor losses. However, this deficiency was compensated within the optimization by increasing the attenuation of the first penetrating wall ($a_1 = 12.79$ dB). This is because the $a_1$ was present in all of the optimization objective functions. In the adjusted model, however, outdoor loss compensation was handled by the $n_{A231} = 2.4$; therefore, resulting in a lower value of $a_1 = 5.10$ dB compared with the COST231 model. The lower $a_1$, which was derived from the adjusted model, is a better and more reasonable estimate, because the wall $k_1$ had plenty of large windows that should have facilitated the penetration [37–41]. This effect is clear at measurement location 13, which had the highest RSSI recorded (see Figure 2). Despite $n_{A231}$ increasing the outdoor loss by a factor of 2.4, none of the other attenuations were altered to a great extent to be noticeable, which further emphasized that it is only adjusting the outdoor propagation. In addition, according to the documented empirical coefficients [35], the path loss of $n_{A231} = 2.4$ resembled the propagation at 900 MHz. This adjustment not only corrected the attenuation of the first penetrating wall and the LoS, but also has identified the propagation characteristics. This simple adjustment in the COST231 model made it more applicable to an outdoor–indoor scenario.

The introduced coefficient ($n_{A231}$) in the adjusted COST231 model correctly represented the propagation characteristics of LoRaWAN in an outdoor LoS condition; however, the log-distance path loss exponent ($n = 3.6$) and $\sigma = 6.8$ demonstrated a notable loss, as the LoRa encountered obstacles on the propagation path into the building. This susceptibility of LoRa resulted in the parameters that indicated the propagation in indoor environments, however, at the frequency of 1.5 GHz instead of 860 to 900 MHz. The $L_c = 4.1$ dB is nearly in an agreement with the $L_E = 3.4$ dB; these parameters were added to the COST231 and log-distance models respectively to account for any potential losses in the system.
Generally, the training of an ANN requires a large set of inputs–outputs. For propagation estimations, this translates to numerous measurements at different locations. For a relatively small indoor environment, this defeats the purpose of propagation estimation. Furthermore, due to the limited number of measurements, training the network so that it learns and generalizes the propagation mechanisms proved to be challenging, especially if the objective of the design is to achieve a better accuracy/precision. Therefore, to facilitate the training of the ANN with limited measurement samples, input data sets are preprocessed before being passed to the network. For instance, instead of providing the network with the LoS length ($d$); initially, $L_{FS}(d)$ is calculated, then scaled by the path loss exponent ($n_{A231}$), and passed to the network. Similarly, in addition to providing the ANN with the number of walls blocking the LoS, the total optimized attenuation of these walls are extracted from the COST231 model and passed as an input to further assist the network’s learning process. Above all, rather than entirely training the ANN to predict the propagation, it was trained to correct the inaccuracies of the COST231 model and improve upon its performance. With this approach, training is carried out using 270 data samples from the eighth floor, and then tested for the collected measurements on the seventh floor. This approach made the training process easy, quick, and needless of an intensive data collection.

Packet loss was observed at some of the locations during the data collection. This was compared against the fading standard deviation ($\sigma$) and RSSI; however, no particular correlation was found between them. This message loss might have been due to frequency interference, a destructive multi-path, or even the mobile data connection of LoRaWAN to the data server using a User Datagram Protocol (UDP).

6. Conclusions

The propagation of LoRaWAN was analyzed in an outdoor–indoor scenario and compared with commonly used propagation models. The performances of these models were briefly compared, and their advantages and shortfalls were discussed. An adjustment was made to the COST231 model, which made it more applicable to outdoor–indoor scenarios.

A hybrid model was proposed, comprising an ANN and an optimized Multi-Wall Model—in this case, the adjusted COST231 model. This combination made the training process faster and easier rather than relying on an ANN only. It also diminished the number of data samples required for training the ANN. By using the ANN, the propagation estimation accuracy was improved. This improvement was achieved by first optimizing the COST231 propagation ($F_{A231}$). Second, the trained ANN was used to generate $T_{ENN}$, which is an estimate of the error in the optimized COST231, Figure 6a. Finally, the estimated error was added to the optimized propagation results ($\hat{RSSI} = T_{ENN} + F_{A231}$), as shown in Figure 5, resulting in a more accurate prediction ($\hat{RSSI}$) of the practical measurements, Figure 6b. This hybrid model reduced the initial MSE of the COST213 from 21 to 11.23.

Acknowledgments: Authors would like to thank, A. Wixted, for his technical help. Glasgow Caledonian University, Stream Technologies, KTP (Innovate UK), and Qatif College of Technology for funding and facilitating this research.

Author Contributions: Salaheddin Hosseinzadeh and Mahmood Almoathen conceived and designed the experiments; Mahmood Almoathen and Salaheddin Hosseinzadeh performed the experiments; Salaheddin Hosseinzadeh analyzed the data; Mahmood Almoathen, Hadi Larijani and Krystyna Curtis contributed materials/analysis tools; Salaheddin Hosseinzadeh, Krystyna Curtis and Hadi Larijani wrote the paper.

Conflicts of Interest: The authors declare no conflict of interest.

References


© 2017 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).