Bayesian Approach for Indoor Wave Propagation Modeling

Abdullah Al-Ahmadi*, Yazeed Qasaymeh, Praveen R. P., and Ali Alghamdi

Abstract—This paper presents a parsimonious Bayesian indoor wave propagation model for predicting signal power in multi-wall multi-floor complex indoor environments. The received power is modeled as a Bayesian multiple regression model. The parameters of the model are assessed and validated using a two-tier validation strategy in which Bayes factor and posterior probability are used in the first tier and second tier, respectively. The performance of the two-tier strategy is then assessed using Bayesian information criterion. The proposed indoor propagation model is tested in a two-storey building with access points operating at 2.4 GHz.

1. INTRODUCTION

With the expanding of wireless communication technologies in the recent years, it is becoming increasingly difficult to ignore the fact that these technologies have become more complex to design and more expensive to install and maintain. In the new global economy, cost has become a central issue for developing such technologies. A major area of interest within the field of minimizing the cost of wireless communication systems is by reducing the human intervention. A considerable amount of literature has been published on this matter. Some of the studies emphasize the propagation of Radio Frequency (RF) and multipath phenomenon, where RF signals arrive at the receiver from multiple directions due to reflection, diffraction, and scattering. By incorporating the knowledge of RF propagation, researchers were able to reduce the human interference factor by using theoretical propagation models instead of empirical methods. However, a major problem with this kind of propagation models is their accuracy compared to empirical methods.

More recently, literature has emerged, which offers extremely meritorious findings about the use of Bayesian models in various applications. In location determination systems, Bayesian models are used to significantly reduce the size of the data set required for model training [1–3], in predicting the popularity of a tweet in social media [4], in record linkage and duplication detection [5], in the field of blind image deconvolution [6], and many more. Three essential operations are needed in Bayesian data analysis [7]. The first step is defining a probability model composed of observed variables and hidden causes, then conditioning on these observed variables in the second step, and lastly evaluating the fit of the model.

The objective of this study is to design a Bayesian model that is capable of predicting the signal power level in a complex multi-floors and multi-walls indoor environments. This study aims to contribute to this growing area of research by integrating Bayesian models with the field of signal propagation.

This study is divided into four sections. Section 2 provides a literature review on the area of wave propagation modeling. Section 3 introduces the proposed model and describes the methods used to validate the model. Section 4 shows the results and performance analysis of the model, and Section 5 describes summary and conclusion.

2. CLASSIFICATION OF INDOOR PROPAGATION MODELS

Generally, indoor propagation models can be divided into two categories, Empirical models where the received power is reported with mean and variance reflecting their accuracy, and Deterministic models where the estimation of received power is based on simulating the physics of radio wave propagation [8].

Tarng & Liu [9] proposed a deterministic site-independent hybrid model where the path loss is determined using direct transmitted ray along with the support of ray tracing technique. At first, the system searches for the direct Line-of-Sight (LoS) to determine whether the transmitted ray was blocked by an obstacle then traced to a predetermined reference direction.

Ray launching model [10] is a deterministic type propagation model for small-cell areas. A number of angle-separated rays are broadcasted in various directions interacting with the obstacles surrounding the receiver. A ray is terminated either by reaching a predetermined number of obstacles or if it declins below a predetermined threshold. The received ray $r(t)$ is formulated as a sum of phase-shifted Dirac functions:

$$r(t) = \sum_{k=1}^{n} \gamma_k \cdot \delta(t - \tau_k) \cdot e^{j\Phi_k} \qquad (1)$$

where γ is the amplitude, τ the time-of-arrival, and Φ the phase-of-arrival. However, the downside of this approach is the increased probability of rays missing obstacles as the separation between transmitter and receiver increases.

The One-Slope Model (1SM) [8] considers only the separation distance between the transmitter and the receiver, and assumes a linear dependence on the path loss (L_p)

$$L_p = L_0 + 10n \log(d) \qquad (2)$$

where L_0 is the first-meter path loss, n the path loss index, and d the distance in meters (m) between the transmitter and the receiver. The main drawback of the 1SM model is that it does not address the very complex characteristics of the indoor environments. Instead, it only relies on the distance to calculate the path loss. The Multi-Wall Model (MWM) [8] suggests that Floor Attenuation Factor (FAF) has a nonlinear relationship with the number of floors between transmitter and receiver.

$$L_p = L_{FS} + L_c + \sum_{i=1}^{I} n_{wi} L_{wi} + n_f^{\left[\frac{n_f+2}{n_f+1} - \beta \right]} L_f \qquad (3)$$

where L_{FS} is the free space loss between transmitter and receiver, L_c a constant loss, n_{wi} the number of penetrated walls of type i, n_f the number of floors penetrated, L_{wi} the loss of type i wall, L_f the loss between neighboring floors, β an empirical parameter, and I the number of wall types. Multiple linear regression is used to calculate L_c for wall losses which usually approaches zero. The model depends on the number and type of walls for the total wall attenuation between transmitter and receiver. Linear Attenuation Model (LAM) [11] assumes linear association between separation distance and path loss:

$$L_p = L_{FS} + \alpha \times d \qquad (4)$$

where α is the attenuation coefficient, and d is the distance from transmitter in meters. A major drawback for this model is the need for large data set and a complete propagation model for a small testbed area.

Malnar and Jevtic [12] proposed a site-dependent multi-room multi-obstacle indoor propagation model. The multi-parameter model integrates total attenuations caused by rooms, wall, windows, and doors.

$$L_p = L_{\text{LoS}} + L_{\text{rooms}} + L_{\text{obstacles}} \qquad (5)$$

where L_{LoS} is the path loss caused by first room encountered while signal travels from transmitter towards the receiver, L_{rooms} the total power loss caused by passing through the walls of all rooms, and $L_{\text{obstacles}}$ the total obstacles power loss. However, this system requires a detailed site knowledge by categorizing rooms into different types based on their usage and specifying walls materials and thickness.

3. BAYESIAN INDOOR PROPAGATION MODEL

The Bayes' Theorem describes the joint probability between random variables α and β as:

$$P(\alpha, \beta) = P(\alpha|\beta) P(\beta) = P(\beta|\alpha) P(\alpha) \tag{6}$$

The conditional probability of α given β is called the *posterior probability* which can be obtained by arranging Equation (6):

$$P(\alpha|\beta) = \frac{P(\beta|\alpha) P(\alpha)}{P(\beta)} \tag{7}$$

where $P(\alpha|\beta)$ is the marginal probability of the prior $P(\alpha)$ multiplied by the likelihood $P(\beta|\alpha)$:

$$posterior \propto prior \times likelihood$$

The Bayes Factor (BF) states that for any two models M_1 and M_0, the BF is [13]:

$$BF = \frac{P(\text{Data}|M_1)}{P(\text{Data}|M_0)} \tag{8}$$

where M_1 is the candidate model, and M_0 is the null model. Table 1 lists the interpretation of BF values [14].

Table 1. Interpreting Bayes factors.

BF_{10}	Evidence
< 1	Supports the null model M_0
1 to 3	Anecdotal evidence for M_1
3 to 12	Moderate evidence for M_1
12 to 150	Very strong evidence for M_1
> 150	Extreme evidence for M_1

Based to the parsimony principles, in model selection there is a tradeoff between bias and variance [14]. As the number of model's parameters increases so does the average error between measured and estimated values.

Table 2. Model comparison.

Models	$P(M)$	$P(M|\text{Data})$	BF_{10}	R^2
Null model	0.125	$3.921e-10$	1.000	0.000
Floors + Distance + Walls	0.125	0.894	$2.279e+9$	0.869
Floors + Walls	0.125	0.091	$2.333e+8$	0.817
Walls	0.125	0.012	$2.955e+7$	0.753
Distance + Walls	0.125	0.002	$5.456e+6$	0.754
Floors + Distance	0.125	0.001	$3.028e+6$	0.742
Distance	0.125	$5.624e-7$	1434.402	0.477
Floors	0.125	$7.159e-11$	0.183	0.001

Table 2 lists all 2^p possible models' combinations where p is the number of explanatory variables. A total of 8 models are considered in this study with BF_{10} compared to the null model in the fifth column where 6 of these models have extremely high BF_{10} values. BF_{10} of the null model is equal to 1 since $M_1 = M_0$, and it is considered a better model candidate than a single "Floors" explanatory variable model. Hence, the posterior probability is used to pick and choose from multiple candidates with high BF as:

$$P(M_i|\text{Data}) = \frac{P(\text{Data}|M_i) P(M_i)}{\sum_{j=1}^{k} P(\text{Data}|M_j) P(M_j)} \tag{9}$$

where k is the number of candidate models. The third column in Table 2 shows the posterior probability for candidate models, here we only consider the top 6 models with BF higher than 150.

Bayesian Information Criterion (BIC) is another strategy used to select and validate predictor variables to be used in the model. This technique leverages the model's likelihood and the size of the data set for parameters selection as follows:

$$\text{BIC} = \ln(n)(p+1) - 2\ln\left(\hat{L}\right) \tag{10}$$

where n is the number of observations in the data set, p the number of parameters, and $\hat{L}$ the maximum likelihood value. Table 3 shows the marginal inclusion probability for the Bayesian model predictors (Floors, Distance and Walls) in addition to the intercept. It is clear that all predictors have a high inclusion probability that exceeds 0.5 with $p(\text{Floors}) = 0.986$, $p(\text{Distance}) = 0.897$ and $p(\text{Walls}) = 0.999$ indicating the contribution of each predictor on the final model as shown in Figure 1 using the Bayesian Adaptive Sampling (BAS) package for R [15].

Table 3. Posterior summaries of coefficients.

Coefficient	Mean	SD	P(incl)	P(incl—data)	$\text{BF}_{inclusion}$
Intercept	65.488	1.187	1.000	1.000	1.000
Floors	15.773	3.338	0.500	0.986	71.856
Distance	1.252	0.396	0.500	0.897	8.703
Walls	6.700	1.352	0.500	0.999	840.953
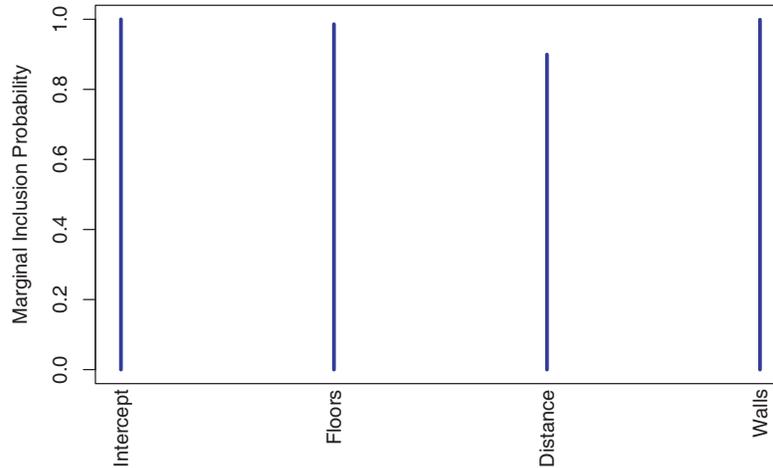
Figure 1. The marginal inclusion probability for three variables in addition to the intercept.

Figure 2 shows all possible models where each column corresponds to a unique model combination. These columns are ordered based on their log posterior odds divided by the null model from highest score in the left to the lowest in the right. The null model refers to the model that has only the intercept as a predictor. As shown in Figure 1, the first model which includes all three predictors has the best performance based on its posterior score of 23.29. The second best model with the posterior score of 20.983 eliminates only one predictor (Distance) which is marked in Figure 2 by a black rectangle. Interestingly, the third model with only one predictor (Walls) has a higher posterior probability than the forth and fifth models with two predictors (Distance and Walls) and (Floors and Distance), respectively.

The correlation between the residuals and fitted values is shown in Figure 3 showing a constant spread of values with only three outliers. It also proves that the residuals and fitted values are uncorrelated suggesting that linear regression is a credible approximation.
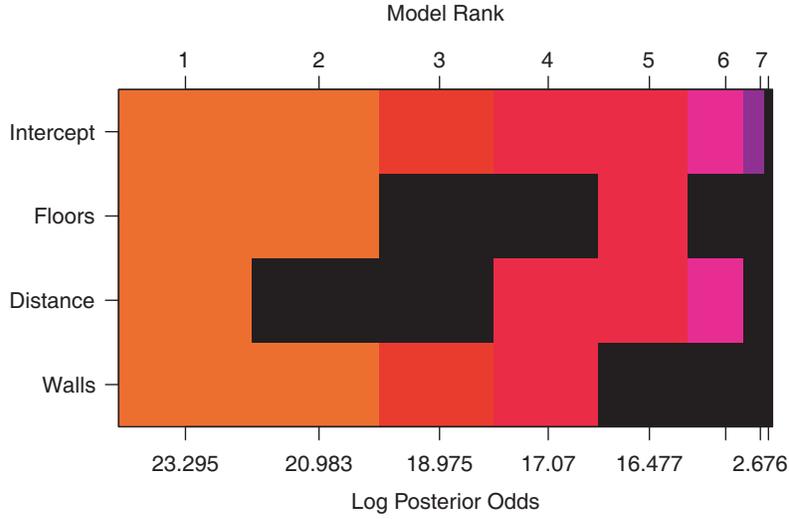
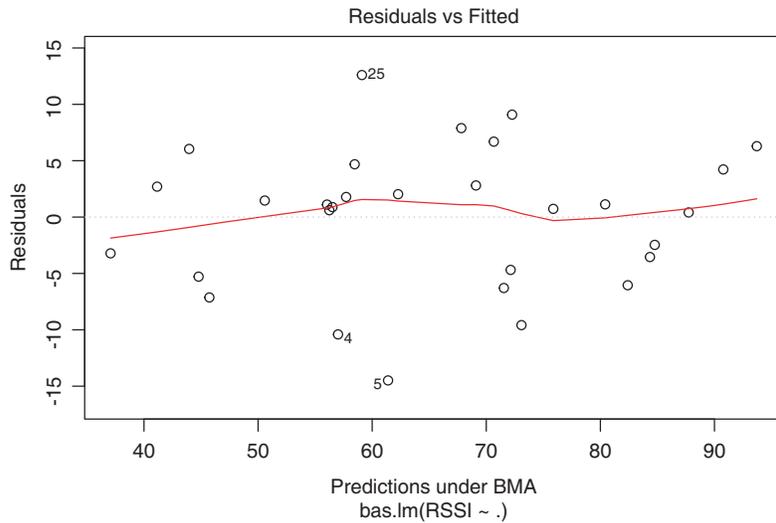**Figure 2.** Model rank for all possible combinations.

Figure 3. Residuals vs. fitted values.

4. DATA COLLECTION AND RESULTS

The data are collected using a custom made PC application for wireless interface [16] in a two-storey building as shown in Figure 4. A user carrying a PC collects data at random location points. The used data set contains 30 *tuples* collected at random location points in a two-storey building [17]. 21 of these tuples are recorded in the first floor of the testbed and 9 in the second floor.

Each tuple comprises six variables, $x_i, y_i, fl_i, D_i, Pr_i$ and w_i. The model is set as:

$$Pr_i = \alpha + \beta_1 x_{fl_i} + \beta_2 x_{D_i} + \beta_3 x_{w_i} + \epsilon_i \tag{11}$$

where Pr_i refers to the received power in dBm; fl_i represents the number of floors between the transmitter and receiver; D_i is the Euclidean distance in meters between the location points and the APs; w_i is the number of walls separation between transmitter and receiver; and ϵ_i is the error term of the ith location point from the Access Point (AP). From Equation (11), the distance D_i is:

$$D_i = \sqrt{(X - x_i)^2 + (Y - y_i)^2} \tag{12}$$

Figure 4. Floor maps for the wireless communication center at Universiti Teknologi Malaysia. (a) First floor. (b) Second floor.
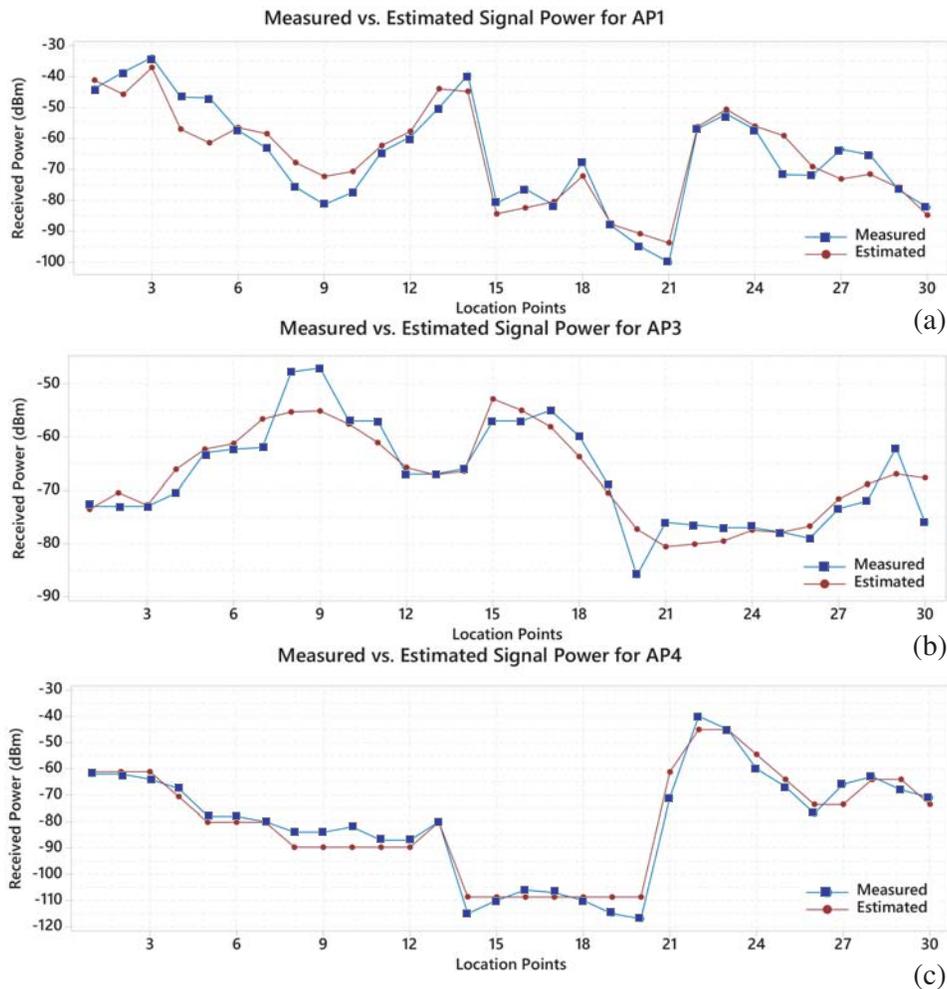
Figure 5. Measured versus estimated signal power for (a) AP1, (b) AP3 and (c) AP4.

where x_i and y_i are the coordinates of the ith location point, and X and Y are the coordinates of AP. The error term ϵ_i is assumed to be Independent and Identically Distributed (iid) with common variance σ^2 as:

$$\epsilon_i \stackrel{\text{iid}}{\sim} \text{Normal}\,(0, \tau) \tag{13}$$

From Equation (11), α, β_1, β_2 and β_3 are the priors for the multiple regression model drawn from a normal distribution with means centered at parameters a, b_1, b_2 and b_3, and τ is Gamma prior distribution for the precision:

$$\alpha, \beta_1, \beta_2, \beta_3 \sim \text{Normal}\,((a, b_1, b_2, b_3)\,, \tau) \tag{14}$$

$$\tau \sim \text{Gamma}\,(g_1, g_2) \tag{15}$$

The BAS package for R which provides a predict function using Bayesian Model Averaging (BMA) is used to test the performance of the proposed model. Three models are built for three individual APs replicating distinct indoor environments. Each aforesaid model is supplied with a Zellner-Siow Cauchy prior and a uniform prior. Figure 5 shows the measured and estimated received powers at all location points from AP1, AP3, and AP4, respectively.

Figure 6 shows the Cumulative Distribution Function (CDF) of estimation error for all APs where each model is based on data collected from that particular AP. The average of estimation error is 4.87 dB for AP1, 3.06 dB for AP3, and 3.55 dB for AP4. The marginal posterior values for coefficients α, β_1, β_2, and β_3 are listed in Table 4 where the received power can be estimated by substituting the values for each coefficient listed in the table with their associated variables, x_{fl_i}, x_{D_i}, and x_{w_i} at the ith location point. Either β_1 or β_3 will be ignored if the transmitter and receiver are at the same floor (i.e., $x_{fl_i} = 0$), or there is a direct LoS between the AP and the location point (i.e., $x_{w_i} = 0$).
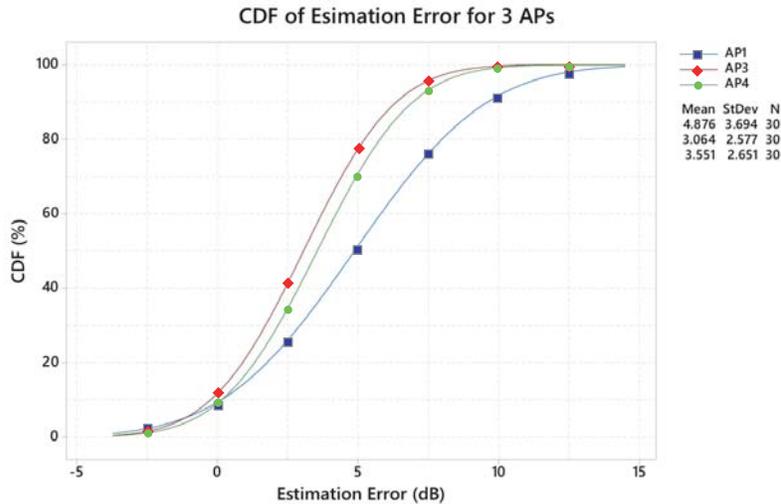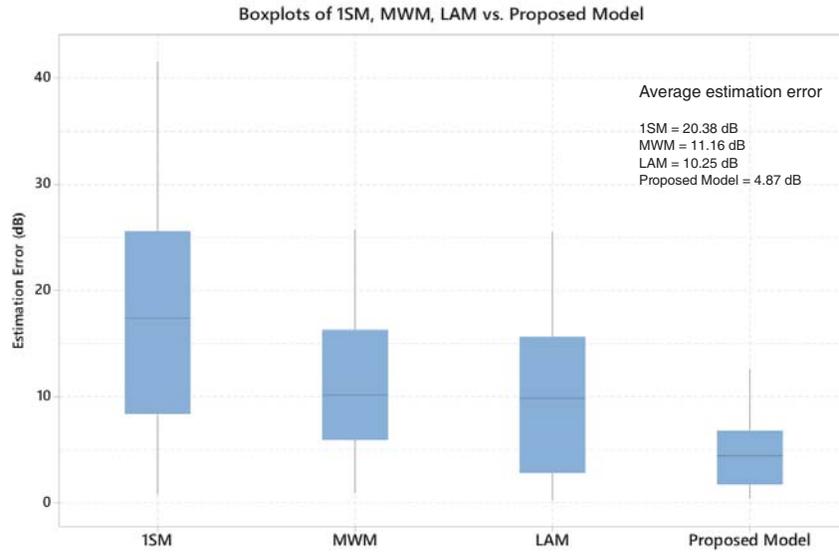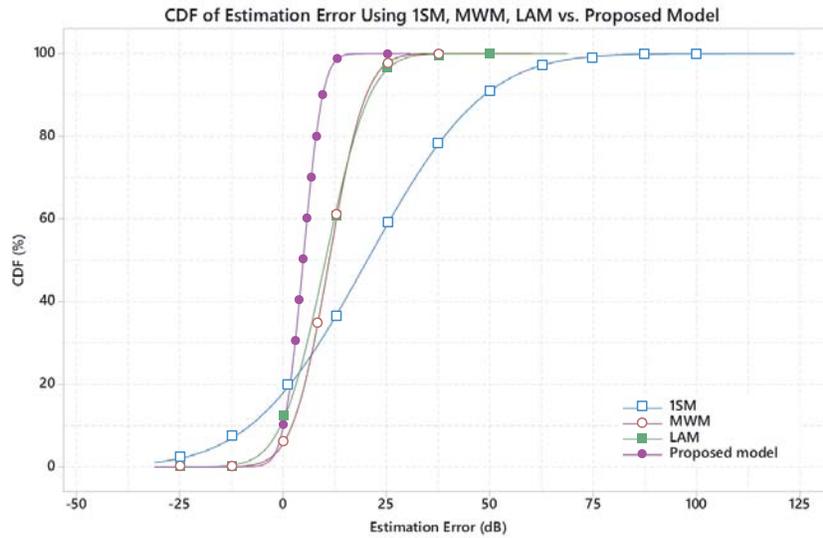

Figure 6. CDF of estimation error.

Table 4. Marginal posterior summaries of coefficients.

Model	Intercept (α)	Floors (β_1)	Distance (β_2)	Walls (β_3)
AP1	65.48	15.15	1.13	7.11
AP3	67.20	5.7	0.83	2.23
AP4	80.11	16.11	0.0023	9.49

Figure 7 shows the performance of AP1 based Bayesian model versus 1SM, MWM, and LAM. All models are tested in the same environments. The *Distance* variable in 1SM model is modified to

Figure 7. Boxplots of estimation error for the proposed model vs. 1SM, MWM and LAM.

Figure 8. CDF of estimation error for the proposed model versus 1SM, MWM and LAM.

substitute for *floors* effect. The average estimation error for 1SM is 20.38 dB and 11.16 dB for MWM and 10.25 dB for LAM in comparison with average estimation error of 4.87 dB for the proposed model. The attenuation coefficient α for LAM model is set to 2.8. Wall Attenuation Factor (WAF) and FAF in MWM model are set to 7 dB and 20 dB, respectively. Note that MWM and LAM perform similarly in terms of average estimation error, but this does not apply to the error variance for both models where variance for MWM is 49.11 dB and 70.93 for LAM. This is also visible by the CDF plot for all models in Figure 8.

Figure 9 shows the CDF of estimation error for Bayesian model based on AP1 coefficients listed in Table 4 with different APs data (AP3 and AP4 in particular) demonstrating system performance in different environments. With the new data, the system achieves a good performance with an average estimated error of 9 dB for AP3 and 6.28 dB for AP4 with low variance compared to MWM and LAM.
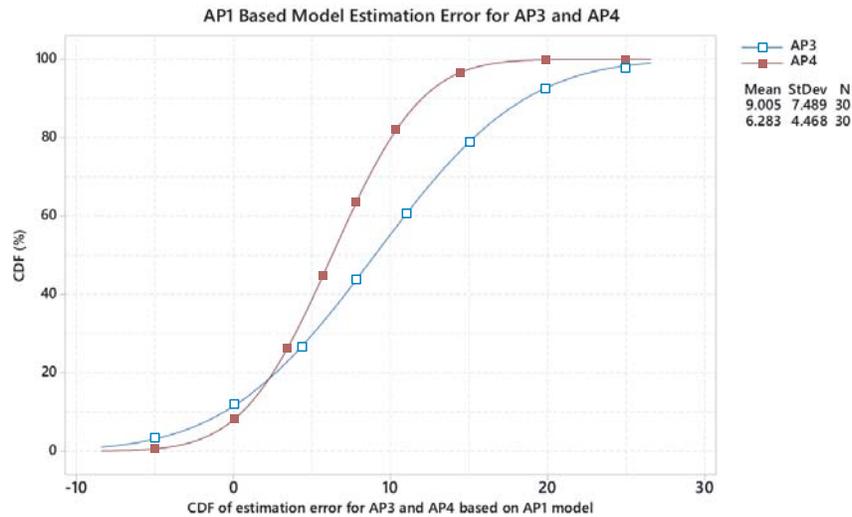
Figure 9. CDF of estimation error for AP3 and AP4 based on AP1 model.

5. CONCLUSION

The purpose of the current study is to design a Bayesian model for predicting the signal power in indoor environments. The proposed model estimates the received power at the receiver in a complex indoor environment. The parameters of the Bayesian multiple regression model are validated using different strategies to ensure their suitability to the model.

ACKNOWLEDGMENT

REFERENCES

1. Madigan, D., E. Elnahrawy, R. P. Martin, W. H. Ju, P. Krishnan, and A. S. Krishnakumar, "Bayesian indoor positioning systems," *Proceedings — IEEE INFOCOM*, Vol. 2, 1217–1227, 2005.

2. Al-Ahmadi, A., A. I. A. Omer, M. R. Kamarudin, and T. A. Rahman, "Multi-floor indoor positioning system using Bayesian graphical models," *Progress In Electromagnetics Research B*, Vol. 25, 241–259, 2010.

3. Xu, H., Y. Ding, P. Li, R. Wang, and Y. Li, "An RFID indoor positioning algorithm based on Bayesian probability and K-Nearest Neighbor," *Sensors*, Vol. 17, No. 8, 1806, Switzerland, Aug. 2017.

4. Zaman, T., E. B. Fox, and E. T. Bradlow, "A Bayesian approach for predicting the popularity of tweets," *Annals of Applied Statistics*, Vol. 8, No. 3, 1583–1611, 2014.

5. Steorts, R. C., R. Hall, and S. E. Fienberg, "A Bayesian approach to graphical record linkage and deduplication," *Journal of the American Statistical Association*, Vol. 111, No. 516, 1660–1672, 2016.

6. Campisi, P. and K. Egiazarian, *Blind Image Deconvolution: Theory and Applications*, 1st Edition, CRC Press, 2007.

7. Kruschke, J. K. and T. M. Liddell, "Bayesian data analysis for newcomers," *Psychonomic Bulletin and Review*, Vol. 25, No. 1, 155–177, 2018.

8. Damosso, E., "Digital mobile radio towards future generation systems: COST action 231," European Commission, 1999.

9. Tarng, J. H. and T. R. Liu, "Effective models in evaluating radio coverage on single floors of multifloor buildings," *IEEE Transactions on Vehicular Technology*, Vol. 48, No. 3, 782–789, 1999.

10. Lawton, M. C. and J. P. McGeehan, "The application of a deterministic ray launching algorithm for the prediction of radio channel characteristics in small-cell environments," *IEEE Transactions on Vehicular Technology*, Vol. 43, No. 4, 955–969, Nov. 1994.

11. Herring, K. T., J. W. Holloway, D. H. Staelin, and D. W. Bliss, "Path-loss characteristics of urban wireless channels," *IEEE Transactions on Antennas and Propagation*, Vol. 58, No. 1, 171–177, Jan. 2010.

12. Malnar, M. and N. Jevtic, "Novel multi-room multi-obstacle indoor propagation model for wireless networks," *Wireless Personal Communications*, Vol. 102, No. 1, 583–597, Sep. 2018.

13. Rouder, J. N. and R. D. Morey, "Default Bayes factors for model selection in regression," *Multivariate Behavioral Research*, Vol. 47, No. 6, 877–903, Nov. 2012.

14. Posada, D. and T. R. Buckley, "Model selection and model averaging in phylogenetics: Advantages of akaike information criterion and Bayesian approaches over likelihood ratio tests," *Syst. Biol.*, Vol. 53, No. 5, 793–808, Oct. 2004.

15. Clyde, M., "BAS: Bayesian variable selection and model averaging using Bayesian adaptive sampling," R package version 1.5.3, 2018.

16. Al-Ahmadi, A. S., T. A. Rahman, M. R. Kamarudin, M. H. Jamaluddin, and A. I. Omer, "Single-phase wireless LAN based multi-floor indoor location determination system," *2011 IEEE 17th International Conference on Parallel and Distributed Systems*, 1057–1062, IEEE, Dec. 2011.

17. Al-Ahmadi, A. and T. A. Rahman, "One stage indoor location determination systems," *New Approach of Indoor and Outdoor Localization Systems*, F. B. Elbahhar and A. Rivenq, editors, Chapter 8, IntechOpen, Rijeka, 2012.