Comparative Study of RMSE and Functional Composition of Residual - Based Tuning of Hata Pathloss Model in the Suburban Area

Wali Samuel, Nwiido Friday, Udoka Uduak Etim

Department of Electrical/Electronic and Computer Engineering, University of Uyo, Uyo Nigeria

Email address:
nwiidofriday@yahoo.com (N. Friday)

To cite this article:

Received: October 16, 2016; Accepted: January 4, 2017; Published: January 29, 2017

Abstract: In this paper, RMSE and functional composition of residual are used as correction factors for tuning Hata model in the suburban area and 800-900MHz GSM frequency band. The study is based on empirical measurements conducted at Abak town, a suburban area in Akwa Ibom state, Nigeria. The tuned model is obtained by adding the correction factor to the original Hata pathloss model for the suburban area. The results showed that the functional composition of residual - based tuning approach has better prediction performance when compared with the RMSE-based tuning approach. Particularly, when the functional composition tuning approach is employed Hata model has the lowest RMSE value of 4.47, the highest prediction accuracy of 97.19% and the highest competitive success rate of 64.29%. On the other hand, the RMSE-tuned Hata model has a higher RMSE value of 7.03, lower prediction accuracy of 96.19% and the lower competitive success rate of 35.71%.

Keywords: Pathloss, Empirical Model, Functional Composition, Residual, Prediction Accuracy, Competitive Success Rate, Hata Model

1. Introduction

Accurate prediction of pathloss is very essential in GSM network planning and optimization. Pathloss is the reduction in power density of an electromagnetic wave signal at it propagates from the transmitter to the receiver [1]. Propagation pathloss models are used to calculate pathloss during transmission of a signal so as to predict the mean signal strength for an arbitrary transmitter-receiver separation distance [2-5]. In general, pathloss models are categorized as empirical, stochastic and deterministic [6-8]. Among the three categories, the empirical models are frequently used for outdoor pathloss predictions. However, in practice, empirical pathloss model tuning is usually required due to significant drop in prediction performance of empirical models when applied in the environments other than the ones they are designed.

Generally, the goal of model tuning is to minimize the difference between measured pathloss and corresponding model predicted pathloss [8]. The tuning can be done by adding correction factor to the original model or by modifying the coefficients of some of the model’s parameters. Among the different model tuning approaches, the Root Mean Square Error (RMSE) based approach has been the easiest and most popular. In the RMSE-based tuning approach, the RMSE between the predicted and the measured pathloss is used as the correction factor which is either added or subtracted from the model to minimize its prediction error. After tuning, the pathloss model prediction performance is evaluated using statistical parameters such as the RMSE, the coefficient of determination ($R^2$), among others. According to available literatures, the performance of a pathloss model is considered acceptable if it provides an overall RMSE of about 6-7dB for urban areas and 10 to13dB for suburban and rural areas [6-8]. Studies have also shown that most empirical pathloss models have high prediction errors with RMSE above the given acceptable range for the particular environment being studied. Therefore, model tuning is usually employed to reduce the model prediction error so that the RMSE falls within the
acceptable range.

Although adding or subtracting the RMSE to the original model may bring down the prediction error to the acceptable range, however, the model with its RMSE closest to zero is usually preferred. Functional composition of residual can be used to improve on the prediction accuracy of pathloss model as well as minimize the prediction error better than the RMSE. In this paper, functional composition of residual and RMSE based correction factors are used for tuning Hata model for suburban area in Abak community located in Akwa Ibom state. The prediction performance of the tuned models is compared in terms of RMSE, prediction accuracy and competitive success rate.

2. Review of Hata’s Propagation Model

Hata pathloss model is a closed-form empirical mathematical expressions use to represent the graphical pathloss data provided by Okumura. Essentially, Hata model simplifies calculation of pathloss based on measurements made by Okumura in urban and suburban areas at Japan in 1968 [9, 10]. Okumura’s presented his pathloss data graphically whereas, Hata simplifies calculation of Okumura’s pathloss by articulating the pathloss in a closed-form empirical mathematical expressions for the different kinds of environments provided in the original Okumura pathloss graph plots. Hata's equation are classified into three models based on the environment, namely, urban, suburban and rural areas [9-11]. Rural are includes open space and areas with no tall trees or building in path. The suburban area includes village highway scattered with trees and house with some obstacles near the mobile but not very congested. The urban area is built-up city or large town with large building and houses [9].

Similar to Okumura model, the Hata model is presented in the urban area propagation loss as a standard formula and provide correction equations for suburban and rural areas [11-13]. For urban rural areas, the Hata model median pathloss equations are given by:

\[
LP_{\text{HATA/(urban)}} = 69.55 + 26.16 \times \log_{10}(f) - 13.82 \times \log_{10}(h_b) - a(h_m) + H \\
H = (44.9 - 6.55 \times \log_{10}(h_b)) \log_{10}(d) 
\]

(1)

Where

- \(a(h_m)\) = correction factor for effective mobile antenna height. For a small to medium size city \(a(h_m)\) is given as;

\[
a(h_m) = [1.1 \times \log_{10}(f) - 0.7] \times h_m - [156 \times \log_{10}(f) - 0.8] \quad (2)
\]

- \(f\) = frequency in MHz; 150 MHz\(\leq f \leq 1000\)MHz
- \(d\) = link distance in km; 1 km \(\leq d \leq 20\)km
- \(h_b\) = height (in metres) of the base station antenna; 30m \(\leq h_b \leq 200\)m
- \(h_m\) = height (in meters) of the mobile antenna; 1m \(\leq h_m \leq 10\)m

\[
LP_{\text{HATA/(suburban)}} = LP_{\text{HATA/(urban)}} - 5.4 - 2 \left[ \log_{10}\left(\frac{d}{10}\right) \right]^2 
\]

(4)

3. Methodology

Measurement of Received Signal Strength (RSS) for the GSM network was done using Handheld Samsung Galaxy S Duos S7562. The path measurements were taken is in Abak town, a suburban area in Akwa Ibom state, Nigeria. The Samsung Galaxy S Duos S7562 phone has Netmonitor Aroid application installed on it. With the Netmonitor application, the android phone can monitor GSM/CDMA/LTE network’s current and neighboring cell information, Received Signal Strength (RSS) in dB and the current cell’s CID, LAC. The Netmonitor application, uses the GPS/geolocatio to generate the longitude and latitude of the mobile phone and also shows the location of the mobile phone and the GSM base stations on a map. The distance between each measurement point and the base station are determined using haversine formula [13-15].

3.1. Calculation of the Measured Pathloss from the Measured RSS

The measured RSS values are converted to measured pathloss (\(PL_{m(dB)}\)) as follows [13-16]:

\[
PL_{m(dB)} = P_{\text{BTS}} + G_{\text{BTS}} + G_{\text{MS}} - L_{\text{FC}} - L_{\text{AB}} - L_{\text{CF}} - \text{RSS(dBm)} 
\]

(5)

where \(PL_{m(dB)}\) is the measured pathloss

RSS is the measured Received Signal Strength (RSS) in dBm

- \(P_{\text{BTS}}\) is the transmitted power (dBm),
- \(G_{\text{BTS}}\) is the transmitter antenna Gain (dBi),
- \(G_{\text{MS}}\) is the receiver antenna gain (dBi),
- \(L_{\text{FC}}\) is the feeder cable and connector loss (dB),
- \(L_{\text{AB}}\) is the antenna body loss (dB) and
- \(L_{\text{CF}}\) is the combiner and filter Loss (dB).

The values of these parameters are given as [16-18]:

- \(P_{\text{BTS}} = 46\) dBm, \(G_{\text{BTS}} = 18.15\) dBi, \(G_{\text{MS}} = 0\)dBi, \(L_{\text{FC}} = 3\) dB, \(L_{\text{AB}} = 3\) dB, \(L_{\text{CF}} = 4.7\) dB. Then,

\[
PL_{m(dB)} = 53.5\text{ dBm} - \text{RSS(dBm)} 
\]

(6)

3.2. Performance Analysis of the Model

i. Root Mean Square Error (RMSE): Root Mean Square Error (RMSE) is given as:

\[
\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \left( \frac{Y_{m(i)} - Y_{p(i)}}{Y_{m(i)}} \right)^2} 
\]

(7)

Where \(Y_{m(i)}\) is the ith actual or measured value and \(Y_{p(i)}\) is the ith predicted value.

ii. Prediction accuracy: Prediction accuracy is given as:

\[
\text{Prediction Accuracy} = \left( 1 - \frac{1}{n} \sum_{i=1}^{n} \left| \frac{Y_{m(i)} - Y_{p(i)}}{Y_{m(i)}} \right| \right) \times 100\% 
\]

(8)

iii. Competitive Success

The competitive success metric is the percentage of times
in a given data set that a given model has made the best prediction. For instance, for each of the n measurement points, the model that makes the best prediction with the smallest prediction error is noted and the total count of the number of best predictions for each model is noted and eventually divided by the total number of measurement points considered in the study.

Let \( J \) be the number of models considered in the study.
Let \( N \) be the total number of data measurement points considered in the study.
Let \( n_j \) be the number of times the model \( j \) makes the best prediction with the smallest prediction error.

Let \( CS_j \) be the Competitive Success of the model \( j \) where:
\[
CS_j (\%) = \left( \frac{n_j}{N} \right) 100\% \quad (9)
\]

Meanwhile, \( n_j \) can be determined using algorithm that is based on the values of the prediction errors. Let \( \epsilon_{j,x} \) be the prediction error of the model \( j \) at the data measurement point \( x \), where \( j = 1, 2, \ldots, J \) and \( x = 1, 2, \ldots, N \). Also, let \( B_{j,x} \) be the an indicator that takes the value of 1 if the absolute value of the prediction error of the model \( j \) at the data measurement point \( x \) is the smallest, otherwise, the value of \( B_{j,x} \) is 0. The algorithm can be stated as follows;

The algorithm for computing \( n_j \)
1: Initialize the counter: \( n_j = 0 \) for all \( j = 1, 2, \ldots, J \)
2: For \( x = 1 \) to \( N \) Step 1
3: Minimum \( \epsilon_x = \text{Minimum} (|\epsilon_{1,x}|, |\epsilon_{2,x}|, \ldots, |\epsilon_{j,x}|) \)
4: Next \( x \)
5: For \( x = 1 \) to \( N \) Step 1
6: For \( j = 1 \) to \( J \) Step 1
7: If \( \{ |\epsilon_{j,x}| \leq \text{Minimum} \epsilon_x \} \) Then
8: \( B_{j,x} = 1 \)
9: Else
10: \( B_{j,x} = 0 \)
11: Endif
12: \( n_j = n_j + B_{j,x} \)
13: Next \( j \)
14: Next \( x \)

3.3. Model Tuning Process and Model Correction Factors

The measure pathloss, the model predicted pathloss and the prediction error or prediction residual is related as functions of distance as follows:

\[
P (d) = P(d) + e (d) + \bar{P} \quad (10)
\]
\[
e (d) = P(d) - \bar{P} \quad (11)
\]

where,
\( P (d) \) is the empirically measured pathloss at distance \( d \) from the transmitter,
\( \bar{P}(d) \) is the model predicted pathloss at distance \( d \) from the transmitter and
\( e (d) \) is prediction residual at distance \( d \) from the transmitter.

The prediction residual consists of both predictable and random error components. The predictable component of the residual at distance \( d \) from the transmitter is denoted as \( E (d) \) whereas the random component is denoted as \( \epsilon \). The random component \( \epsilon \) is not a function of \( d \) so it is modeled as a lump sum of all the random errors associated with the measurement. Hence,

\[
e (d) = E (d) + \epsilon \quad (12)
\]
\[
\bar{P} (d) = \bar{P}(d) + E (d) + \epsilon \quad (13)
\]

Model tuning or optimization process seeks to adjust the model so that the tuned model can as well predict the predictable components of the error thereby reducing the error to only the random component. Hence the tuned model denoted as \( \bar{P} \bar{T} (d) \) is given by;

\[
\bar{P} \bar{T}(d) = \bar{P}(d) + E(d) \quad (14)
\]

Essentially, \( E (d) \) is the model correction factor that can be used to minimize the prediction error.

In most literatures examined, pathloss model tuning is performed by adding or subtracting the RMSE to the original model [16-21]. In this case, \( E (d) \) which is the predictable component of the residual is approximated by a constant, namely, the RMSE between the measured and the predicted pathloss. Hence, the model correction factor, \( E (d) \) is equal to the RMSE,

\[
E (d) = \text{RMSE} \quad (15)
\]
\[
P (d) = \bar{P}(d) + \text{RMSE} + \epsilon \quad (16)
\]
\[
\bar{P} (d) = \bar{P}(d) + \text{RMSE} \quad (17)
\]

The predictable component of the residual can be predicted with respect to the \( \bar{Y} (d) \) is the model predicted pathloss at distance \( d \). In this case, \( E (d) \) is modeled as a function of the predicted pathloss, where;

\[
E (d) = F (\bar{P}(d)) \quad (18)
\]

\( F (\bar{P}(d)) \) is a composition function whereby the residual error \( E (d) \) is predicted as a function of the predicted pathloss, \( \bar{P}(d) \) which is a function of distance, \( d \). Hence, in this case, \( F (\bar{P}(d)) \) is the model correction factor that can be used to minimize the prediction error.

\[
P (d) = \bar{P}(d) + F (\bar{P}(d)) + \epsilon \quad (19)
\]
\[
\bar{P} (d) = \bar{P}(d) + F (\bar{P}(d)) \quad (20)
\]

4. Results and Discussion

Table 1 shows the location of the measurement point locations and their corresponding distance from the base station along with the RSS, measured pathloss and Hata suburban model predicted pathloss the suburban and rural area in Aba town.
From Table 1, Hata model for suburban area has RMSE of 47.28 dB with Prediction Accuracy of 67.78%. The model’s performance is unacceptable since the overall RMSE is greater than 10 dB for suburban area [22-24]. Therefore, two different model correction factor tuning approaches are used to minimize the prediction error. Specifically, RMSE-based correction factor and the functional composition of prediction residual, $F(\hat{P}(d))$-based correction factor are used.

In the first case, the RMSE of 47.28 dB is added to each prediction of the Hata model for the suburban area. In the second case, the functional composition of the residual is generated by fitting nonlinear equation, $E(d)$ to the graph of $e(d)$ versus $\hat{P}(d)$ as shown in Table 2 and figure 1 where $E(d)$ is given as:

$$E(d) = \frac{\hat{P}(d)}{0.0005689198299(P(d))^2+0.0954435263(P(d))-2.6091686093} \quad (21)$$

<table>
<thead>
<tr>
<th>Hata Suburban Model Predicted Pathloss, $\hat{P}(d)$ in dB</th>
<th>Prediction Residual, $e(d)$ in dB</th>
<th>Functional composition Of The Hata Suburban Model Prediction Residual $E(d)$ in dB</th>
</tr>
</thead>
<tbody>
<tr>
<td>72.5198</td>
<td>37.9802</td>
<td>38.7076939</td>
</tr>
<tr>
<td>80.8228</td>
<td>45.6772</td>
<td>39.0176944</td>
</tr>
<tr>
<td>87.4116</td>
<td>44.0884</td>
<td>40.3379845</td>
</tr>
<tr>
<td>91.944</td>
<td>43.556</td>
<td>41.9174979</td>
</tr>
<tr>
<td>95.4021</td>
<td>43.0979</td>
<td>43.5781862</td>
</tr>
<tr>
<td>98.1987</td>
<td>42.3013</td>
<td>45.1446166</td>
</tr>
<tr>
<td>100.406</td>
<td>43.094</td>
<td>46.8430817</td>
</tr>
<tr>
<td>102.4465</td>
<td>42.0535</td>
<td>48.5296317</td>
</tr>
<tr>
<td>104.2377</td>
<td>42.2623</td>
<td>50.2123616</td>
</tr>
<tr>
<td>105.834</td>
<td>45.666</td>
<td>51.7850858</td>
</tr>
<tr>
<td>107.1862</td>
<td>47.3138</td>
<td>53.4619292</td>
</tr>
<tr>
<td>108.504</td>
<td>53.996</td>
<td>55.1389304</td>
</tr>
<tr>
<td>109.7132</td>
<td>60.7868</td>
<td>56.8183094</td>
</tr>
<tr>
<td>110.8304</td>
<td>62.6696</td>
<td></td>
</tr>
</tbody>
</table>

Figure 1. Graph of prediction residual ($e(d)$), functional composition of the Hata prediction residual ($E(d)$) versus Hata predicted pathloss, $\hat{P}(d)$ for suburban area.
Afterwards, the functional composition, $E(d)$ is added to the Hata predicted pathloss for the suburban area. Table 3 and figure 2 show the results of the two tuned Hata model for suburban area. From Table 3, the functional composition of prediction residual tuned Hata model has the lowest RMSE value of 4.47, the highest prediction accuracy of 97.19% and the highest competitive success rate of 64.29%. On the other hand, the RMSE- tuned Hata model has a higher RMSE value of 7.03, lower prediction accuracy of 96.19% and the lower competitive success rate of 35.71%. The functional composition of prediction residual based tuning approach performed better than the RMSE based tuning approach.

### Table 3. The results of tuning of the Hata model by addition of the RMSE and by addition of the Functional composition of Prediction Residual of the Hata model for suburban area.

<table>
<thead>
<tr>
<th>$d$ (km)</th>
<th>Measured Pathloss (dB)</th>
<th>Un-tuned Hata Predicted Pathloss, $P_d$ in dB for the suburban area</th>
<th>RMSE Tuned Hata Predicted Pathloss, in dB for the suburban area</th>
<th>Functional composition of Prediction Residual Tuned Hata Predicted Pathloss, in dB for the suburban area</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.060129</td>
<td>110.5</td>
<td>72.5198</td>
<td>119.7958858</td>
<td>114.8109564</td>
</tr>
<tr>
<td>0.104353</td>
<td>126.5</td>
<td>80.8228</td>
<td>128.0988858</td>
<td>119.5304939</td>
</tr>
<tr>
<td>0.163601</td>
<td>131.5</td>
<td>87.4116</td>
<td>134.6876858</td>
<td>126.4292944</td>
</tr>
<tr>
<td>0.222888</td>
<td>135.5</td>
<td>91.944</td>
<td>139.2200858</td>
<td>132.2819645</td>
</tr>
<tr>
<td>0.28132</td>
<td>138.5</td>
<td>95.4021</td>
<td>142.6781858</td>
<td>137.3195979</td>
</tr>
<tr>
<td>0.340977</td>
<td>140.5</td>
<td>98.1987</td>
<td>145.4747858</td>
<td>141.7768862</td>
</tr>
<tr>
<td>0.396284</td>
<td>143.5</td>
<td>100.406</td>
<td>147.6820858</td>
<td>145.5604616</td>
</tr>
<tr>
<td>0.453649</td>
<td>144.5</td>
<td>102.4465</td>
<td>149.7225858</td>
<td>149.2895817</td>
</tr>
<tr>
<td>0.512036</td>
<td>146.5</td>
<td>104.2377</td>
<td>151.5137858</td>
<td>152.7673317</td>
</tr>
<tr>
<td>0.570857</td>
<td>151.5</td>
<td>105.834</td>
<td>153.1100858</td>
<td>156.0463616</td>
</tr>
<tr>
<td>0.625741</td>
<td>154.5</td>
<td>107.1862</td>
<td>154.4622858</td>
<td>158.9712858</td>
</tr>
<tr>
<td>0.684713</td>
<td>162.5</td>
<td>108.504</td>
<td>155.7800858</td>
<td>161.9659292</td>
</tr>
<tr>
<td>0.742933</td>
<td>170.5</td>
<td>109.7132</td>
<td>156.9892858</td>
<td>164.8521304</td>
</tr>
<tr>
<td>0.801693</td>
<td>173.5</td>
<td>110.8304</td>
<td>158.1064858</td>
<td>167.6487094</td>
</tr>
<tr>
<td>RMSE</td>
<td>47.2760858</td>
<td>7.032222031</td>
<td>4.470521962</td>
<td>35.71428571</td>
</tr>
<tr>
<td>Prediction Accuracy (%)</td>
<td>67.77617854</td>
<td>67.77617854</td>
<td>96.18760293</td>
<td>96.18760293</td>
</tr>
<tr>
<td>Competitive Success</td>
<td>0</td>
<td>0</td>
<td>35.71428571</td>
<td>64.28571429</td>
</tr>
</tbody>
</table>

The RMSE based tuning approach has the correction factor which is equal to the RMSE value of 7.03. Then, with respect to Eq. 1, the RMSE-based tuned Hata pathloss model ($TLP_{\text{HATA (suburban)RMSE}}$) for suburban area is given as:

$$TLP_{\text{HATA (suburban)RMSE}} = LP_{\text{HATA (suburban)}} + 7.03 (22)$$

Similarly, by equation (21), the functional composition of residual tuned Hata pathloss model ($TLP_{\text{HATA (suburban)FCR}}$) for suburban area is given as:

$$TLP_{\text{HATA (suburban)FCR}} = \frac{\hat{P}(d)}{0.0005689198299[P(d)]^2 + 0.0954435263[P(d) - 2.609168693]} (23)$$

### 5. Conclusion

Tuning and comparative prediction performance analysis of Hata pathloss model for suburban area is presented. RMSE-based and functional composition of residual-based correction factor pathloss tuning approaches are used to tune the Hata pathloss model for suburban area. The study is conducted for the GSM network in the 800-900MHz frequency band. The study is based on empirical measurements conducted at Abak town, a suburban area in Akwa Ibom state, Nigeria. The prediction performance of the two tuning approaches is compared in terms of RMSE, prediction accuracy and competitive success rate. The results show that the functional composition of residual-based approach performed better with lower RMSE, higher prediction accuracy and higher competitive success rate.

### References


