A Gaussian Process Regression Model to Predict Path Loss for an Urban Environment

Seyi E. Olukanni*
Department of Physics, Confluence University of Science and Technology Osara, Nigeria
Email: olukannise@custech.edu.ng
ORCID iD: https://orcid.org/0000-0003-1410-8971
*Corresponding Author

Ikechi Risi
Department of Physics, Rivers State University, Port Harcourt, Nigeria
Email: Ikechi.risi@ust.edu.ng
ORCID iD: https://orcid.org/0000-0001-83316550

Salifu. F. U.
Department of Physics, Confluence University of Science and Technology Osara, Nigeria
Email: salifufu@custech.edu.ng
ORCID iD: https://orcid.org/0000-0001-9015-2347

Johnson Oladipupo S.
Department of Statistics, Kogi State Poly Technique, Lokoja, Nigeria
Email: oladipuposamueljohnson@gmail.com
ORCID iD: https://orcid.org/0009-0002-8312-2726

Received: 01 February, 2023; Revised: 10 March, 2023; Accepted: 18 April, 2023; Published: 08 May, 2023

Abstract: This research paper presents a Gaussian process regression (GPR) model for predicting path loss signal in an urban environment. The Gaussian process regression model was developed using a dataset of path loss signal measurements acquired in two urban environments in Nigeria. Three different kernel functions were selected and compared for their performance in the Gaussian process regression model, including the squared exponential kernel, the Matern kernel, and the rotational quadratic kernel. The GPR model was validated and evaluated using various performance metrics and compared with different regression models. The results show that the Gaussian process regression model with the Matern kernel outperforms the linear regression and the support vector regression, but the decision tree and the random forest regression did better than the GPR in both cities. In the city of Port Harcourt, the GPR has a RMSE value of 3.0776 dB, the DTR has 2.0005 dB, the SVR has 3.6047 dB, the RFR has 1.0459 dB, and the LR 3.5947 dB. The proposed GPR model provides more accurate and efficient approach to predict path loss compared to traditional methods. The extensive data collection and analysis conducted has resulted in a well-developed and accurate model.

Index Terms: Gaussian process regression, wireless communication, Path Loss, Machine Learning, regression.

1. Introduction

In the field of wireless communication, the ability to accurately quantify the signal strength that gets to the receiver from the transmitter is an essential task for accurate network planning. The propagation loss model describes the reduction of signal due to various factors like distance, obstacles, and interference[1]. The loss in received signal strength (RSS) is more prominent in the urban environment due to the presence of high-rise buildings and many other obstacles[2]; hence, accurate path loss modeling is crucial in an urban environment. Accurate path loss prediction is essential for network planning, optimization, and performance evaluation. Gaussian Process Regression (GPR) [3] is a powerful machine learning technique that has been widely used for path loss prediction due to its ability to model
complex non-linear relationships between variables. In this research, we propose a Gaussian Process Regression model to predict path loss for an urban environment.

The research objectives for this research are:

1. To develop a GPR model that accurately predicts path loss in an urban environment.
2. To evaluate the performance of the proposed GPR model and compare it with existing solutions.
3. To identify the limitations of the proposed GPR model and suggest possible improvements.

The main problem addressed in this research is the accurate prediction of path loss in an urban environment. Existing path loss models are often based on empirical measurements or simplistic assumptions that do not capture the complexity of urban environments. This can lead to inaccurate predictions, which can have a significant impact on network planning and optimization.

Several existing solutions have been proposed for path loss prediction in an urban environment. These include empirical models such as the Okumura-Hata [4] model and the COST-231[5] model, as well as more advanced techniques such as ray tracing[6] and neural networks [7]. The best solution depends on several factors such as the specific characteristics of the urban environment, available data, and computational resources. However, these existing solutions have limitations such as oversimplification of the environment or high computational requirements.

We hope to develop a GPR model that accurately predicts path loss in an urban environment while addressing the limitations of existing solutions. The proposed model aims to provide a more accurate and efficient alternative to existing methods for path loss prediction.

The rest of the paper is organized as follows: Section 2 provides a review of the literature on path loss modeling and Gaussian process regression. Section 3 describes the methodology used in this study, including data collection, model development, and experiment design. Section 4 presents the results and discussion of the study, including an evaluation of the Gaussian process regression model and a comparison with existing models. Section 5 summarizes the findings and contributions of the study, along with its limitations and future research directions.

2. Related Works

Gaussian Process Regression (GPR) is a powerful non-parametric Bayesian approach for modeling and predicting functions that are observed with noise [8]. GPR has been used in a variety of applications, including finance, robotics, and engineering, due to its ability to provide uncertainty estimates and handle non-linearities without specifying a particular parametric form for the underlying function. One of the key components of GPR is the kernel function, which determines the covariance structure between the input data points [9]. The choice of kernel function can have a significant impact on the performance of GPR, and different kernel functions may be more appropriate for different types of data.

One popular kernel function used in GPR is the squared exponential (SE) kernel, also known as the radial basis function (RBF) kernel or Gaussian kernel. The SE kernel is isotropic and produces a smooth function; it has been used in a variety of applications, including speech recognition and image classification [9, 10].

In [12], the authors used different GPR kernels for in-depth extrapolative analysis of Covid-19 Daily Cases Drift Rates. They argued that GPR outperformed other machine learning techniques such as neural networks, Neural-Fuzzy networks, Random forest, Regression tree, Support Vector machines, K-nearest neighbor and Discriminant linear regression. In the field of wireless communication, GPR has been used to model and predict path loss signals in urban environments. One study by [13] used GPR with the SE kernel to model path loss signals in a complex suburban area in Korea. They claimed their model accurately predicts the measured data with a coverage error of less than 1.6% and also performed better compared to the path loss linear log-normal shadowing model. Another study by [14] used GPR to model radial mean function and local shadow fading term for the prediction of path loss in a wireless networks. They affirmed that their model will predict better than the log-distance path loss model, and can be used to simulate path loss in any environment.

In conclusion, GPR with different kernel functions can be used to model and predict functions in a variety of applications, including wireless communication. The choice of kernel function should be made based on the characteristics of the data and the desired properties of the resulting function. The SE, Matern, and RQ kernels are all viable options for GPR, depending on the complexity and non-linearities of the input data.

3. Methodology

The step-by-step approach to achieve the research objectives includes:

i. Collecting path loss data from two urban cities through a drive test survey. This data is preprocessed and analyzed to identify the most relevant features affecting path loss in an urban environment.
ii. Developing a Gaussian Process Regression (GPR) model based on the identified features to predict path loss in an urban setting.
iii. Validating the GPR model using a hold-out dataset or cross-validation techniques, ensuring the model’s robustness and ability to generalize well to new data.

3.1. Data collection

Measurement Environment

The data collection took place in two Nigerian cities, Onitsha and Port-Harcourt, located in the south-east and south-south regions of the country, respectively. Onitsha is the most populated city in Anambra State, with a population of 2 million people, and the commercial hub of the southern geo-political zone in the country. It is situated at 6.1329°N, 6.7924°E. Port Harcourt is the fifth most populous city in Nigeria, with a population density of about 3.4 million people. It is the capital city of Rivers State, located at latitude 4.8472°N and longitude 6.9746°E. Both cities have all the qualities of urbanization, like high-rise buildings, factories, and so on. For a more detailed description of the cities investigated, an aerial view of the environments is given in Figs. 2 and 3, respectively.

Measurement tools and procedure

The signal of a radio network was measured with the aid of experimental drive-test tools. A predefined route was selected around two transmission antennae in each of the study locations. The tools used are TEMs investigation and drive-test software installed on a mobile phone and connected to a Dell laptop. The other tools are a GPS, an inverter, connecting cables, and a car for mobility. The car is driven at a speed of 30 km/h to mitigate the impact of Doppler effects. The carrier frequency is 2600 MHz. The measured data was analyzed using MS-Excel and Python.

3.2. Model development

Path Loss Model
The general large-scale path loss model (in dB) for any T-R separation (x) is given as:

$$P_l(x) = A \log(x) + B$$  \hfill (1)

Where $A = 10 \eta$ is the slope of the path loss curve, $\eta$ is the path loss exponent, and $B$ is the path loss at reference point.

Equation (1) can be rewritten as

$$y_i(x) = f(x_i) + \phi \quad i = 1, 2, \ldots, t$$  \hfill (2)

where $\phi$ designate the Gaussian zero mean noise.

Meanwhile, to obtain path loss from the measured signal we use equation (3).

$$P_{path} = P_t + G_t + G_r - f_l - C_l - RSP$$ \hfill (3)

where $P_t$ is the base station transmit power in (dBm), $G_t$ is the base station antenna gain in (dBi), $G_r$ is the mobile station antenna gain in (dBi) and, $f_l$ and $C_l$ are the feeder loss and cable loss respectively. The value of these parameters and other relevant parameters are listed in Table 1.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Definition</th>
<th>Numerical Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P_t$</td>
<td>Base station transmit power</td>
<td>43dB</td>
</tr>
<tr>
<td>$G_t$</td>
<td>Base station antenna gain</td>
<td>17.5dB</td>
</tr>
<tr>
<td>$G_r$</td>
<td>Mobile station antenna gain</td>
<td>0dB</td>
</tr>
<tr>
<td>$f_l$</td>
<td>Feeder loss</td>
<td>3dB</td>
</tr>
<tr>
<td>$C_l$</td>
<td>Cable loss</td>
<td>2dB</td>
</tr>
<tr>
<td>$H_t$</td>
<td>Base station antenna height</td>
<td>30m</td>
</tr>
<tr>
<td>$H_r$</td>
<td>Mobile antenna height</td>
<td>1.5m</td>
</tr>
<tr>
<td>$F_t$</td>
<td>Transmit frequency</td>
<td>2600 MHz</td>
</tr>
</tbody>
</table>

Gaussian Process Regression (GPR) model

In this work, we employ Gaussian process regression to predict signal attenuation in an urban environment.

Consider a nonlinear regression model given in equation (2), where the functional response $y(x)$ is the path loss in dB and the covariates function $x = (x_1, x_2, \ldots, x_Q)^T$ is the distance between the transmitter and the receiver at different instances. The path-loss $y$ from a GPR model can be modeled as:

$$P(y_i | f(x_i), x_i) \sim N(y_i | h(x_i)^T \beta + f(x_i), \sigma^2)$$  \hfill (4a)

where

$$x = \begin{pmatrix} x_1^T \\ x_2^T \\ \vdots \\ x_Q^T \end{pmatrix}, \quad y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_t \end{pmatrix}, \quad h = \begin{pmatrix} h(x_1) \\ h(x_2) \\ \vdots \\ h(x_t) \end{pmatrix}, \quad f = \begin{pmatrix} f(x_1) \\ f(x_2) \\ \vdots \\ f(x_t) \end{pmatrix}$$  \hfill (4b)

The observed target in the GPR model can equally be defined as follows:

$$P(f|X) \sim N(f|0, K(X, X))$$ \hfill (5)

$$K(X, X) = \begin{pmatrix} k(x_1, x_1) & k(x_1, x_2) & \cdots & k(x_1, x_t) \\ k(x_2, x_1) & k(x_2, x_2) & \cdots & k(x_2, x_t) \\ \vdots & \vdots & \ddots & \vdots \\ k(x_t, x_1) & k(x_t, x_2) & \cdots & k(x_t, x_t) \end{pmatrix}$$ \hfill (6)

where $K(X, X)$ is the covariance function.

Kernels
The Kernel, otherwise known as the covariance function is a vital part of GPR model; it determines the covariance between data points [15]. There are different kernel functions that impact on the GPR modeling. In this study, we experiment with three commonly used kernel function for path loss signal prediction in an urban environment. Table 2 presents our kernel of interest and their respective expressions.

Table 2. Three types of kernels considered in this paper

<table>
<thead>
<tr>
<th>Kernel</th>
<th>Expression</th>
<th>Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Radial Basis Function (RBF)</td>
<td>( k(x,x'</td>
<td>\theta) = \sigma^2 \exp\left(-\frac{|x-x'|^2}{2\sigma^2}\right) )</td>
</tr>
<tr>
<td>Matern</td>
<td>( k(x,x'</td>
<td>\theta) = \sigma^2 \left[1 + \frac{\sqrt{3}r}{\rho} + \frac{\sqrt{3}r^3}{3\rho^3}\right] \exp\left(-\frac{\sqrt{3}r}{\rho}\right) )</td>
</tr>
<tr>
<td>Rotational Quadratic</td>
<td>( k(x,x'</td>
<td>\theta) = \sigma^2 \left[1 + \frac{r}{2\alpha^2}\right]^{-\alpha} )</td>
</tr>
</tbody>
</table>

RBF is also known as a squared exponential kernel. It works on the assumption that the correlation between data points decreases exponentially with the distance between them. It is both positive definite and isotropic. Meanwhile, the hyperparameters of RBF were learned from the data through cross-validation.

Matern kernel is a generalization of the RBF. It is more flexible and allows for different degrees of smoothness. However, it reduces to RBF if the smoothness parameters, \( \nu = \frac{1}{2} \).

Rotational Quadratic is a non-stationary kernel, meaning it can model functions with varying degrees of smoothness and non-linearities in different input directions. Its hyperparameters are the variance (\( \sigma \)) parameter, the center point (\( c \)), the positive definite matrix (\( A \)), and the smoothness parameter (\( d \)). They can all be learned from the data through cross-validation.

Model validation and Evaluation

In this paper, the model performance is evaluated using the mean absolute error (MAE), mean square error (MSE), and coefficient of determination (R-squared). MAE defines the absolute difference between the predicted and measured signals, while MSE represents the squared residual between the measured signal and the predicted signals. R-squared is a measure of the proportion of the variance in the dependent variable that is explained by the independent variables in the model.

To evaluate the model, we compare the performance of the Gaussian process regression model with other popular regression models, such as linear regression (LR), decision tree regression (DTR), random forest regression (RFR), and support vector regression (SVR).

3.3. Alternative Regression Models for Comparison

Apart from GPR, there are many other commonly used regression techniques that have been applied by researchers. However, for this paper, we will be applying the following:

a. Decision Tree Regression (DTR) is a non-linear machine learning model used for prediction. The input-output function is given as:

\[
y(x) = \frac{1}{\xi} \sum_{i=1}^{\xi} y_i
\]

Where \( \xi \) is the available observation number in a specific cell

b. Random Forest Regression (RFR) is a modified DTR, it uses an ensemble learning to improve the predictive performance and reduce over-fitting.

\[
y(x) = f_1(x) + f_2(x) + ... + f_n(x) + \epsilon
\]

where \( f_1(x), f_2(x), ..., f_n(x) \) are the tree-based function

c. Support Vector Regression (SVR) is a non-linear machine learning regression model; it maps the input data into a high-dimensional space and finds a hyperplane that best separates the output data. The input-output function is given as:
A Gaussian Process Regression Model to Predict Path Loss for an Urban Environment

\[ g(x) = \sum_{n=1}^{N} (a_n + \alpha_n^*) G(x_n, x) + b \]  

where \( a_n \) and \( \alpha_n^* \) are Lagrange multipliers, \( N \) is the bias term and \( k(\cdot) \) is the kernel function.

d. Linear Regression (LR) is arguably the most used regression model. It assumed a linear relationship between the independent and dependent variables.

\[ y(x) = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \ldots + \varepsilon \]  

where \( \alpha_0, \alpha_1, \ldots, \alpha_n \) are coefficients, and \( \varepsilon \) is the error term.

4. Results and Discussion

First, we examine the impact of different kernel functions (covariance functions), which determine the covariance between data points in a GPR for our datasets. Table 3 and Fig. 3 display the attained performance impact of the three kernels considered for the path loss prediction in an urban environment. The four evaluation metrics used are MAE, MSE, RMSE, and R2. The results show that the performances of the three kernels are almost equal. However, matern32 has the lowest RMSE of 3.04468 compared to the others. Therefore, Matern32 with GPR are used in the remainder of this paper.

Table 3. Statistics analysis for 3 different kernels investigated

<table>
<thead>
<tr>
<th>Kernel</th>
<th>MAE</th>
<th>MSE</th>
<th>RMSE</th>
<th>R-Squared</th>
</tr>
</thead>
<tbody>
<tr>
<td>Matern32</td>
<td>2.36704</td>
<td>9.27007</td>
<td>3.04468</td>
<td>0.895355</td>
</tr>
<tr>
<td>Rational Quadratic</td>
<td>2.39502</td>
<td>9.47136</td>
<td>3.07756</td>
<td>0.892899</td>
</tr>
<tr>
<td>RBF (Squared exponential)</td>
<td>2.39378</td>
<td>9.43541</td>
<td>3.07171</td>
<td>0.893326</td>
</tr>
</tbody>
</table>

Fig. 3. Statistics analysis plot for 3 different kernels

Figs. 4 to 8 display the predictive analysis of propagation path loss estimation in Port Harcourt environment using the GPR model with the matern32 kernel and four other regression models for comparison. From the figures and Table 4, the GPR; the focus of this paper only outperformed the linear regression and the support vector regression with approximately 2.395 MAE value when compared to the decision tree regression, support vector regression, random forest regression, and linear regression. However, it has the smoothest curve. Meanwhile, for our dataset, the random forest regression performs best among the models with 0.3156 MAE value and the LR performance is the least with 2.7992 MAE value.
Predictive Analysis with Gaussian Process Regression (GPR) Model

Fig. 4. Predictive analysis with GPR for the city of Port Harcourt

Predictive Analysis with Decision Tree Regression (DTR) Model

Fig. 5. Predictive analysis with DTR for the city of Port Harcourt

Predictive Analysis with Support Vector Regression (SVR) Model

Fig. 6. Predictive analysis with SVR for the city of Port Harcourt
Predictive Analysis with Radom Forest Regression (RFR) Model

Fig. 7. Predictive analysis with RFR for the city of Port Harcourt

Predictive Analysis with Linear Regression (LR) Model

Fig. 8. Predictive analysis with LR for the city of Port Harcourt

Table 4. Prediction Errors and R2 Values for 5 Regression models for Port Harcourt environment

<table>
<thead>
<tr>
<th>Model</th>
<th>MAE</th>
<th>MSE</th>
<th>RMSE</th>
<th>R-squared</th>
</tr>
</thead>
<tbody>
<tr>
<td>GPR</td>
<td>2.3950</td>
<td>9.4714</td>
<td>3.0776</td>
<td>0.8929</td>
</tr>
<tr>
<td>Tree</td>
<td>1.4750</td>
<td>4.0019</td>
<td>2.0005</td>
<td>0.9562</td>
</tr>
<tr>
<td>SVR</td>
<td>2.8069</td>
<td>12.994</td>
<td>3.6047</td>
<td>0.8504</td>
</tr>
<tr>
<td>Random</td>
<td>0.3156</td>
<td>1.0940</td>
<td>1.0459</td>
<td>0.9882</td>
</tr>
<tr>
<td>LR</td>
<td>2.7992</td>
<td>12.9215</td>
<td>3.5947</td>
<td>0.8504</td>
</tr>
</tbody>
</table>

Figs. 9 to 13 display the predictive analysis of propagation path loss estimation in Onicha environment using the GPR model with the matern32 kernel and four other regression models for comparison. From the figures and Table 5, The GPR; the focus of this paper only outperformed the linear regression and the support vector regression with approximately 1.94514 MAE value when compared to the decision tree regression, support vector regression, random forest regression, and linear regression. However, it has the smoothest curve. Meanwhile, for our dataset, the random forest regression performs better than the others with 0.187666 MAE value and the LR performance is the least with 3.92886 MAE value.
Predictive Analysis with Gaussian Process Regression (GPR) Model

Fig. 9. Predictive analysis with GPR for the city of Onitcha

Predictive Analysis with Decision Tree Regression (DTR) Model

Fig. 10. Predictive analysis with DTR for the city of Onitcha

Predictive Analysis with Support Vector Regression (SVR) Model

Fig. 11. Predictive analysis with SVR for the city of Onitcha
Predictive Analysis with Random Forest Regression (RFR) Model

Fig. 12. Predictive analysis with RFR for the city of Onitsha

Predictive Analysis with Linear Regression (LR) Model

Fig. 13. Predictive analysis with LR for the city of Onitsha

Table 5. Prediction Errors and R2 Values for 5 Regression models for Onitsha environment

<table>
<thead>
<tr>
<th></th>
<th>MAE</th>
<th>MSE</th>
<th>RMSE</th>
<th>R-squared</th>
</tr>
</thead>
<tbody>
<tr>
<td>GPR</td>
<td>1.94514</td>
<td>6.00224</td>
<td>2.44995</td>
<td>0.950795</td>
</tr>
<tr>
<td>Tree</td>
<td>1.80744</td>
<td>5.61919</td>
<td>2.37048</td>
<td>0.953594</td>
</tr>
<tr>
<td>SVR</td>
<td>3.94037</td>
<td>22.7011</td>
<td>4.76456</td>
<td>0.808276</td>
</tr>
<tr>
<td>Random</td>
<td>0.187666</td>
<td>0.31088</td>
<td>0.557566</td>
<td>0.997489</td>
</tr>
<tr>
<td>LR</td>
<td>3.92886</td>
<td>21.4887</td>
<td>4.63559</td>
<td>0.808276</td>
</tr>
</tbody>
</table>

5. Conclusion

This paper presents a Gaussian process regression (GPR) model for the prediction of propagation path loss in wireless communication in an urban environment. The GPR model was developed with a path loss dataset obtained in two Nigerian megacities, and three different kernel functions were compared for their performance in the model. The GPR model was validated with four performance metrics and compared with four other regression models. The results show that the GPR model with the Matern32 kernel only outperformed the linear model and the support vector model; however, the decision tree and the random forest regression model outperformed the GPR model. Nonetheless, the GPR model is a good model for improving the accuracy of path loss signal prediction in wireless communication systems. To improve the performance of the model, additional study can examine the usage of different kernel functions and different features.

Overall, the proposed GPR model will advance the field by offering a precise and efficient method for predicting path loss in urban settings. Utilizing machine learning techniques like GPR addresses limitations of traditional models that depend on oversimplified assumptions. This flexible, data driven approach effectively captures complex interactions between buildings, terrain, and environmental factors impacting path loss.
References


Authors’ Profiles

Seyi E. Olukanni is a Ph.D student in the Department of Physics, Federal University Lokoja, Kogi State, Nigeria. He received his M.Sc degree in 2017 and B.Sc degree in the year 2008. Both degrees are in physics from University of Ilorin and University of Abuja respectively. His area of interest includes Electromagnetism, Signal Processing, Internet of Things and Radio communications. He can be contacted through olukannise@custech.edu.ng

Ikechi Risi is lecturer at Rivers State University, Nigeria in Physics Department where he lectures physics. He obtained his B.Sc and M.Sc in Solid State Physics at the university in 2013 and 2019 respectively and currently pursuing his Ph.D at Ignatius Ajuru University of Education, Port-Harcourt. His interest areas are electronic circuit design/construction, embedded system, and radio signal propagation processing and modelling. He can be contacted through ikechirisi@ust.edu.ng.

Salifu Francis is a PhD student in the Department of Physics, University of Ilorin, Ilorin, Nigeria. He received his B.Sc. degree in 2014 from Kaduna State University and M.Sc. degree in 2019 from University of Ilorin. His areas of interest include space weather and equatorial ionosphere.

Johnson Oladipupo Samuel is a Mathematics PhD research student of Federal University Lokoja, Kogi State, Nigeria. He obtained B.Sc and M.Sc Mathematics in the year 2006 and 2017 respectively from the Federal University of Abuja, Nigeria. His area of specialization is Mathematical modeling and optimization.
A Gaussian Process Regression Model to Predict Path Loss for an Urban Environment