Chapter 1

Introduction

1.1 Introduction

Satellite communication is an important part of today's telecommunication, broadcasting, and navigation systems. The performance and reliability of satellite links are largely affected by several atmospheric parameters that attenuate the signal. Understanding and being able to accurately predict path loss due to environmental effects is required in order to design and optimize satellite communication networks. For Bangladesh, where the weather is diverse, it is preferable to model satellite path loss for enhancing the quality of satellite-based communication services.

Satellite communication path loss is caused by various factors such as free-space attenuation, atmospheric loss, rain attenuation, and cloud interference. Ground stations utilized to quantify attenuation by transmitting a dummy signal to the satellite and estimating its loss. This, however, takes resources and infrastructure on top of what is already needed. If a reliable predictive model can be established, it would eliminate the necessity of transmitting these dummy signals, thereby lessening operating expenses and boosting efficiency.

In this study, we take Bangladesh Satellite-1 and its ground base stations at Gazipur and Betbunia into account. Collecting historical attenuation data from Bangladesh Communication Satellite Company Limited (BCSCL) and relative weather data from Bangladesh Meteorological Department (BMD), a comprehensive dataset has been formed. The dataset integrates cloud, atmospheric, rain, and free-space attenuation data with various meteorological parameters. In the estimation of total accumulated attenuation, two approaches are investigated: a predictive model based on machine learning and a mathematical model. Both models' accuracy and viability are compared to ascertain their efficacy in being used as substitutes for the conventional dummy signal method. This study not only helps in the enhancement of satellite communication in Bangladesh but also offers a model that can be applied in other parts of the globe that have similar climatic conditions. By having a robust path loss model, satellite communication service providers can render their networks more stable, simplify resource planning, and enhance overall service quality. The findings of this research will give valuable insights on how to mitigate atmospheric attenuation effects and offer uninterrupted satellite communication services in Bangladesh and beyond.

1.2 Motivation

The inspiration behind this research is the growing dependence on satellite communication for a number of important applications such as telecommunication, broadcasting, weather forecasting, and emergency services. It would be best to render these services more efficient, particularly in a developing nation such as Bangladesh, where satellite technology is emerging as a vital part of national infrastructure. One of the most essential satellite communications issues is attenuation of signals caused by atmospheric conditions. Conventional attenuation monitoring methods use dummy signals, which impose cost and complexity on satellite operations. By having a good predictive model, we can make these dummy signals obsolete, promoting operational efficiency and cost savings to satellite operators.

Besides, Bangladesh has diverse and extreme climatic conditions like high monsoons and humidity, which have a pervasive impact on signal transmission. A good path loss model will help mitigate the consequences of these challenges by providing real-time attenuation predictions, thus improved signal reliability and uninterrupted communication services.

The worldwide relevance of this study cannot be ignored since atmospheric attenuation is experienced by all satellite operators around the globe. The result of this study can be used as a reference for other regions of the world with the same meteorological climate, which will add to the general body of satellite communication and path loss modeling.

In general, this study is motivated by the demand for an inexpensive, dependable, and efficient means of enhancing satellite communication in Bangladesh and elsewhere. With the use of sophisticated machine learning techniques and mathematical modeling, we seek to establish a scientifically credible method of attenuation prediction for the general goal of increasing the level of performance of satellite-based communication systems.

1.3 Objective

The main goal of this research is to create an effective and precise model for predicting satellite path loss from atmospheric attenuation. Based on machine learning approaches and mathematical modeling, this research intends to establish the most appropriate method to foresee total attenuation. The main aims of this research include:

- Collecting and merging BCSCL's historical attenuation data with BMD's weather data to create a comprehensive dataset.
- Building a machine learning model to predict total accumulated attenuation based on meteorological data.
- Creating a mathematical model for predicting attenuation and comparing the precision with the machine learning model.
- Comparing the validity and accuracy of both models for determining their viability in replacement of conventional dummy signal-based attenuation monitoring.
- Providing data that can be utilized for optimizing satellite communication networks, reducing operational costs, and enhancing the performance of Bangladesh Satellite-1 and other similar systems worldwide.

Through these goals, this research seeks to play a role in the creation of a more efficient and resource-saving method of monitoring satellite communication signal attenuation.

1.4 Scope of Work

The study's importance transcends the prediction of satellite path loss—it presents a novel and globally applicable method for optimizing the performance of satellite communications. Unlike other models needing location-dependent parameters like latitude and longitude, the mathematical model proposed in this work is not geographically limited. That is, wherever a ground station is positioned or whichever country it is in, the model is still valid.
Instead of being dependent on coordinates, the mathematical model only needs to be provided with typical weather parameters like rain rate, surface temperature, surface pressure, relative humidity, cloud base height, liquid water density in clouds, liquid water temperature in clouds, uplink frequency, polarization type, and path length. All these as inputs make the model deployable anywhere on the earth without any changes, which is a huge advantage when it comes to scalability and deployment ease.
By eliminating reliance on location-specific data, this model has the potential to be a globally viable solution for satellite communication path loss prediction. Even though improvements in accuracy are still possible, the groundwork for global usability is already there. This study, thus, not only has value in the enablement of satellite communication in Bangladesh but also makes a wider contribution by providing a model that can readily be incorporated into satellite networks worldwide.

Chapter 2

Background Work

2.1 General Path Loss Models

To understand and predict signal attenuation in wireless communication, we studied several **path loss models** that describe how the strength of a transmitted signal weakens with distance. Since wireless signals are subject to varying degrees of loss due to obstructions, reflections, and environmental conditions, multiple models help estimate actual behavior.

To compare these models, we ran **MATLAB simulations** using a 1.5 GHz (randomly selected) carrier frequency. The simulation compares three commonly used path loss models:

1. **Free-space path loss model** – perfect line-of-sight (LOS) propagation.
2. **Log-distance path loss model** – encompasses distance-dependent attenuation in different environments.
3. **Log-normal shadowing model** – encompasses random variations in signal strength due to obstacles.

All the models were run with mathematical equations for the evaluation of **signal attenuation over distance** considering different **antenna gains and path loss exponents**.

2.1.1 Free-space Path Loss Model

The **free-space path loss model** assumes that a transmitted signal propagates without any obstructions. The power loss in decibels (dB) is given by:

$$PL_{free}(d) = 20 \log_{10} \left(\frac{4\pi d f_c}{c} \right) - G_t - G_r \tag{2.1}$$

where:

- d = Distance between transmitter and receiver (m)
- f_c = Carrier frequency (Hz)
- c = Speed of light (3×10^8 m/s)
- G_t = Transmitter antenna gain (dB)
- G_r = Receiver antenna gain (dB)

This model is useful for **satellite and open-area communication**, but it does not consider environmental factors like obstacles or reflections, making it **unrealistic for terrestrial communication**.

2.1.2 Log-distance Path Loss Model

In real-world environments, signal attenuation is influenced by the surroundings. The **log-distance path loss model** accounts for this using a path loss exponent (n):

$$PL_{logdist}(d) = PL(d_0) + 10n \log_{10} \left(\frac{d}{d_0} \right) \tag{2.2}$$

where:

- $PL(d_0)$ = Path loss at reference distance d_0
- d_0 = Reference distance (100 m in this study)
- n = Path loss exponent (e.g., 2 for free space, 3 for urban areas, 6 for highly obstructed environments)

The path loss exponent (n) varies based on terrain and building density. This model provides a **more practical estimation** than the free-space model but does not incorporate fading effects.

2.1.3 Log-normal Shadowing Model

In environments with obstacles, signal strength can fluctuate unpredictably. The **log-normal shadowing model** extends the log-distance model by adding a **Gaussian-distributed random variable** to represent signal variations due to shadowing:

$$PL_{lognorm}(d) = PL(d_0) + 10n \log_{10} \left(\frac{d}{d_0} \right) + X_\sigma \tag{2.3}$$

where:

- X_σ = A Gaussian random variable with standard deviation σ
- σ = 3 dB in our simulation

This model is more realistic for urban and obstructed environments as it introduces signal variation, but it requires **statistical modeling** and does not explicitly account for multipath fading.

2.1.4 Implementation in MATLAB

To analyze these models, we performed MATLAB simulations using:

- Carrier frequency: 1.5 GHz
- Reference distance: $d_0 = 100$ m
- Path loss exponents: $n = 2, 3, 6$
- Antenna gains: G_t and G_r with varying values

The distances were taken as squared values $(1, 4, 9, ..., 961)$ to better observe attenuation trends. Three separate plots were generated:

1. Free-space model with different antenna gains.
2. Log-distance model with varying path loss exponents.
3. Log-normal model showing shadow fading effects.

2.1.5 Advantages and Limitations

Free-space Path Loss Model:

- Advantage: Provides a simple reference for signal attenuation.
- Limitation: Does not account for obstacles or environmental interference.

Log-distance Path Loss Model:

- Advantage: More adaptable for different environments by adjusting n.
- Limitation: Assumes a single path loss exponent, which may not always be accurate.

Log-normal Shadowing Model:

- Advantage: Captures random signal variations due to obstacles.
- Limitation: Requires statistical modeling and does not include multipath fading.

2.1.6 Practical Considerations

While these models provide useful insights, real-world wireless communication is affected by:

- Multipath fading due to reflections and interference.
- Weather effects such as rain and atmospheric conditions.
- Dynamic obstacles like moving vehicles and buildings.
- Antenna characteristics that affect radiation patterns.

For real-world applications, more advanced models such as Hata, COST-231, ITU-R, and Rayleigh/Rician fading models are used for network design and signal prediction.

2.2 Okumura/Hata Model

The **Hata model** is an empirical path loss model that is used for the prediction of signal attenuation in macrocellular environments (150 MHz – 1500 MHz). It estimates path loss in urban, suburban, and open areas based on the heights of the transmitter and receiver from the ground. Our thesis is about satellite communication, where antennas are not on the ground, so this model is not applicable for our research.

2.2.1 Hata Path Loss Model Equation

For urban situations, the formula is as follows:

$$PL_{urban} = 69.55 + 26.16\log_{10}(f_c) - 13.82\log_{10}(h_{tx}) - C_{Rx} + (44.9 - 6.55\log_{10}(h_{tx}))\log_{10}(d/1000) \tag{2.4}$$

where:

- f_c = Carrier frequency (MHz)
- h_{tx} = Transmitter height (m)
- h_{rx} = Receiver height (m)
- d = Transmitter-to-receiver distance (m)
- C_{Rx} = Receiver correction factor (function of f_c)

Correction factors modify the path loss equation to account for suburban and open areas.

2.2.2 MATLAB Implementation

The MATLAB code:

- Computes path loss for urban, suburban, and open areas.
- Takes a carrier frequency of 1.5 GHz with htx = 30m, hrx = 2m.
- Plots results against logarithmic distance values.

2.2.3 Why Hata Model is Not Used in This Thesis

- Is derived from ground antenna heights, and hence cannot be used for satellite antennas.
- Is designed for use in terrestrial links, not space-based links.
- Does not take into consideration satellite path loss, where the free-space loss and atmospheric attenuation are more prominent.

Thus, alternative models like free-space path loss and satellite-specific models need to be used for our study.

2.3 IEEE 802.16d Model

The **IEEE 802.16d model** is an empirical path loss model used to predict signal attenuation for fixed broadband wireless systems. It classifies environments into three categories (A, B, C), based on propagation conditions, and incorporates shadowing corrections using factors from ATnT and Okumura models. Additionally, a modified version exists, which adjusts reference distance (d_0) based on correction factors.
Since this model requires ground-based antenna heights, it is not applicable for our thesis on satellite communication, where antennas are not ground-based.

2.3.1 IEEE 802.16d Path Loss Model Equation

The general form of the IEEE 802.16d model is:

$$PL(d) = A + 10\gamma \log_{10}\left(\frac{d}{d_0}\right) \tag{2.5}$$

where:

- $A = 20\log_{10}\left(\frac{4\pi d_0}{\lambda}\right) + PL_f + PL_h$ is the reference path loss at distance d_0.
- λ = Wavelength of the signal ($\lambda = \frac{c}{f_c}$).
- γ = Path loss exponent, given by:

$$\gamma = a - bh_{tx} + \frac{c}{h_{tx}} \tag{2.6}$$

with values depending on the environment type:

- Type A: $a = 4.6, b = 0.0075, c = 12.6$
- Type B: $a = 4, b = 0.0065, c = 17.1$
- Type C: $a = 3.6, b = 0.005, c = 20$

Correction Factors for Shadowing:

- ATnT Correction: $PL_f = 6\log_{10}(f_c/2000)$, $PL_h = -10.8\log_{10}(h_{rx}/2)$.
- Okumura Correction:

$$PL_h = \begin{cases} -10\log_{10}(h_{rx}/3), & h_{rx} \leq 3m \\ -20\log_{10}(h_{rx}/3), & h_{rx} > 3m \end{cases} \tag{2.7}$$

A modified version of IEEE 802.16d adjusts the reference distance:

$$d_0' = d_0 \times 10^{-\frac{(PL_f + PL_h)}{10\gamma}} \tag{2.8}$$

where d_0' is the adjusted reference distance.

2.3.2 MATLAB Implementation

The MATLAB code:

- Computes path loss for Type A environments using ATnT correction.
- Compares the original and modified IEEE 802.16d models.
- Uses carrier frequency = 2 GHz, transmitter heights = 30m, and receiver heights = {2m, 10m}.
- Plots results against logarithmic distance.

2.3.3 Why IEEE 802.16d Model is Not Used in This Thesis

- Requires ground-based antenna heights, which are not relevant for satellites.
- Designed for terrestrial fixed wireless systems, not space-to-ground communication.
- Does not consider satellite-specific losses, such as atmospheric attenuation or free-space propagation.

Since this model is unsuitable for satellite networks, alternative satellite-specific models are needed for our study.

2.4 Integration of ITU-R Propagation Models

2.4.1 Significance of ITU-R Propagation Models

Accurate prediction of satellite signal attenuation is crucial for developing effective and dependable telecommunication systems. Attenuation affects the strength, quality, and overall performance of signals and systems. The International Telecommunication Union Radiocommunication Sector (ITU-R) provides standardized models with accurate methodologies to estimate signal degradation caused by atmospheric effects. The three ITU-R models discussed in this section—**ITU-R P.838-3 (Rain Attenuation), ITU-R P.840-9 (Cloud and Fog Attenuation), and ITU-R P.676-13 (Atmospheric Gases Attenuation)**—are the core of satellite communication system planning. They ensure consistency and reliability in prediction models, resulting in the establishment of robust communication links.
Microwave and millimeter-wave satellite signals suffer severe degradation due to atmospheric absorption, scattering, and reflection. The most significant causes of signal attenuation are rain, cloud, fog, and atmospheric gases, and these must be accounted for in any satellite path loss calculation. The ITU-R models give empirical and analytical approaches to the computation of these losses and are therefore a necessity for engineers and researchers.

2.4.2 Explanation of Each ITU-R Model

ITU-R P.838-3: Model of Rain Specific Attenuation

Rain is the most powerful cause of attenuation of the signal in satellite communication, especially at frequencies above 10 GHz. **ITU-R P.838-3** provides a model to estimate rain attenuation using the power-law relation:

$$\gamma_R = kR^{\alpha} \tag{2.9}$$

Where:

- γ_R is the specific attenuation (dB/km),
- R is the rain rate (mm/h),
- k and α are frequency-dependent empirical constants.

The model encompasses polarization effects by distinguishing between horizontal and vertical polarizations. The model also provides coefficient values as a result of detailed scattering computations for precise prediction over a wide range of frequencies (1 GHz to 1000 GHz). The model plays a crucial role in **link budget analysis, fade margin design, and adaptive modulation schemes** in high-frequency satellite links.

ITU-R P.840-9: Attenuation in Cloud and Fog

Cloud and fog add additional attenuation, primarily at frequencies greater than 10 GHz. **ITU-R P.840-9** provides a formula for the estimation of signal loss due to cloud and fog as a function of integrated liquid water content (LWC). The model uses the following formula for attenuation:

$$\gamma_c(f, T) = K_l(f, T) \cdot \rho_l \tag{2.10}$$

Where:

- γ_c is the specific attenuation (dB/km),
- $K_l(f, T)$ is particular cloud liquid water attenuation coefficient,
- ρ_l is cloud or fog liquid water density (g/m$^{\mathbf{3}}$),
- f is frequency (GHz),
- T is temperature (Kelvin).

The model can be applied within a frequency band of **1 GHz to 200 GHz** and is therefore essential in the design of **Earth-space links, high-frequency terrestrial microwave systems, and next-generation satellite constellations**. Because cloud attenuation is meteorology dependent, the model further incorporates statistical distributions and global climatological information, enhancing its predictive power.

ITU-R P.676-13: Attenuation by Atmospheric Gases

Atmospheric gases like **oxygen and water vapor** absorb electromagnetic waves, causing additional path loss. **ITU-R P.676-13** provides procedures for calculating **slant path gaseous attenuation**, considering:

- **Oxygen and water vapor absorption** due to molecular resonance effects.
- **Frequency-dependent absorption coefficients** of different atmospheric layers.
- **Correction factors for varying atmospheric conditions (pressure, temperature, and humidity).**

The model provides two primary methods:

1. **Spectral line-by-line calculations** for highly accurate attenuation estimation from individual spectral lines.
2. **Empirical fits** for quick engineering calculations in engineering applications.

The model has a very wide frequency range (**1 GHz to 1000 GHz**), and therefore it is very useful for **radio astronomy, remote sensing, and high-frequency satellite communications**. It also accounts for path elevation angles, which is crucial for **geostationary and low Earth orbit (LEO) satellite systems**.

2.4.3 Reasoning Behind the Use of These Models

Use of ITU-R models in satellite path loss prediction ensures that the estimates of attenuation are:

1. **Scientifically Validated**: The models have been developed through extensive experimental and theoretical work, hence reliable.
2. **Frequency-Dependent**: Each model is specific to specific frequency bands, so they can be used across various bands utilized in satellite communications.
3. **Geographically Adaptive**: Global meteorological datasets are used in the models, which allows location-based predictions.
4. **Global Standardized Use**: ITU-R recommendations are commonly adopted, allowing interoperability among various communication systems and compliance with the regulations.

All the models are used for specific purposes:

- **ITU-R P.838-3** is critical for **rain attenuation predictions**, which have a great influence on satellite links in tropical and high-rainfall environments.

- **ITU-R P.840-9** deals with **cloud and fog attenuation**, which is important for systems that function in humid environments and high-frequency bands.

- **ITU-R P.676-13** also incorporates **gaseous attenuation**, important to apply to precise signal strength calculations, particularly in long-haul satellite transmissions.

Through their application, these models enable satellite communication systems to more effectively take advantage of **power budgets, antenna designs, frequency selection, and adaptive coding schemes**, leading eventually to a better link reliability and overall performance.

2.5 Selection of the Goff-Gratch Formula for Saturation Vapor Pressure Calculation

Accurate calculation of saturation vapor pressure (e_s) is crucial in climate modeling, hydrology, and atmospheric sciences with a high priority in high-latitude cold regions where temperature variation significantly influences vapor pressure calculations. In our thesis, we have employed the **Goff-Gratch formula** since it is more accurate than other available formulas. The selection is based on the comparison conducted by **Xu et al. (2012)** among various formulas, i.e., **Teten, Magnus, Buck, and Goff-Gratch**, and their performances according to the variations of temperature.

It was concluded in the research that while the Teten formula, being popular in application to evapotranspiration modeling (e.g., the FAO-56 Penman-Monteith equation), is good for the moderate temperature range (0°C to 40°C), it gives grave errors at cold temperatures. Teten's error in computed e_s increases proportionally to a decrease in temperature up to greater than 40% at -40°C. This leads to enormous errors in computing the vapor pressure deficit (VPD) and subsequently in estimating the reference evapotranspiration (ET_0).

In contrast, the Goff-Gratch formula, officially recommended by the World Meteorological Organization (WMO), provides a better estimation of saturation vapor pressure across a broad range of temperatures and is therefore the preferred option for studies dealing with extreme weather events.

2.5.1 Mathematical Representation of the Goff-Gratch Formula

The **Goff-Gratch equation** for saturation vapor pressure is expressed as follows:

For $T > 273.16K$:

$$\log e_s = a_1 \left(\frac{T_{01}}{T} - 1 \right) + b_1 \log \frac{T_{01}}{T} + c_1 \left(10^{d_1 (1 - T/T_{01})} - 1 \right) + e_1 \left(10^{f_1 (T_{01}/T - 1)} - 1 \right) + \log g_1 \tag{2.11}$$

For $T < 273.16K$:

$$\log e_s = a_2 \left(\frac{T_{02}}{T} - 1 \right) + b_2 \log \frac{T_{02}}{T} + c_2 \left(1 - \frac{T}{T_{02}} \right) + \log d_2 \tag{2.12}$$

where the constants are:

$$\begin{array}{llll} a_1 = -7.90298, & b_1 = 5.02808, & c_1 = 1.3816 \times 10^{-7}, & \\ d_1 = 11.344, & e_1 = 8.1328 \times 10^{-3}, & f_1 = -3.49149, & g_1 = 1013.246, \\ a_2 = -9.09718, & b_2 = -3.56654, & c_2 = 0.876793, & d_2 = 6.1071, \\ T_{01} = 373.16, & T_{02} = 273.16. & & \end{array}$$

These equations take into account the phase transitions of water vapor and provide highly accurate results across a broad temperature range.

2.5.2 Comparison with Alternative Formulas

The study by Xu et al. (2012) also evaluated other formulas:

- **Teten Formula**: Commonly used in FAO-56 Penman-Monteith equation:

$$e_s = 0.611 \times \exp \left(\frac{17.27t}{t + 237.3} \right) \tag{2.13}$$

  While computationally simple, it shows high errors at low temperatures.

- **Magnus Formula**: Offers improved accuracy by differentiating between $t > 0°\text{C}$ and $t < 0°\text{C}$:

$$e_s = 6.11 \times 10^{\frac{7.45t}{t+237.3}}, \quad t > 0°C \tag{2.14}$$

$$e_s = 6.11 \times 10^{\frac{9.5t}{t+265.5}}, \quad t < 0°C \tag{2.15}$$

- **Buck Formula**: Similar to Magnus but modified for improved performance:

$$e_s = 6.1121 \times \exp \left(\frac{18.678 - \frac{t}{234.5}}{257.14 + t} \times t \right), \quad t > 0°C \tag{2.16}$$

$$e_s = 6.1115 \times \exp \left(\frac{23.306 - \frac{t}{333.7}}{279.82 + t} \times t \right), \quad t < 0°C \tag{2.17}$$

Although the Buck and Magnus formulas provide reasonable approximations, the **Goff-Gratch formula remains the most authoritative** due to its basis in fundamental thermodynamic principles and experimental validation.

2.5.3 Conclusion

Based on the findings of Xu et al. (2012), the Goff-Gratch formula has been selected in this study due to its superior accuracy across a wide temperature range and its global acceptance as a standard method. Its ability to maintain precision under extreme cold conditions makes it particularly suitable for high-latitude climate analysis, ensuring reliable estimation of e_s and its associated impact on evapotranspiration calculations.

2.6 Machine Learning for Predicting Path Loss

2.6.1 Introduction and Objectives of the Paper

To see which model should be used to better predict the path loss, we read a paper on this named "Path Loss Prediction Based on Machine Learning Techniques: Principle Component Analysis, Artificial Neural Network, and Gaussian Process". This work explores the application of machine learning techniques for path loss prediction in wireless sensor networks. Classical models, such as linear log-distance models, can be inadequate for capturing the subtleties of environments. With the aim of improving the precision of predictions, the authors develop a machine learning-based framework that employs:

- **Principal Component Analysis (PCA)** for dimensionality reduction and feature selection.
- **Artificial Neural Networks (ANN)** for multi-dimensional regression and learning the structure of path loss.
- **Gaussian Process (GP)** for shadowing effect modeling and variance analysis.

The primary motivation behind this research is to develop a more flexible and accurate path loss model compared to conventional empirical approaches. By leveraging machine learning, the authors aim to create a model that better adapts to real-world variations in signal propagation, particularly in suburban environments.

2.6.2 Machine Learning Framework for Path Loss Prediction

The paper outlines a three-step process for predicting path loss using machine learning techniques:

1. **Feature Selection using PCA:**
    - The path loss dataset typically consists of multiple features, such as **distance, antenna height, and frequency**.

- PCA is used to reduce the dataset's dimensionality by identifying the most influential features, minimizing redundancy, and improving computational efficiency.
- The authors find that **log distance and log frequency** contribute to more than 70% of path loss variation.

2. **ANN-Based Regression Model:**
   - After dimensionality reduction, an **Artificial Neural Network (ANN)** learns the path loss pattern.
   - A multi-layer perceptron (MLP) architecture is used, with different activation functions (ReLU, sigmoid, and hyperbolic tangent) tested for performance evaluation.
   - The ANN-MLP model aims to predict **mean path loss values** from the dataset's reduced features.

3. **Gaussian Process for Shadowing Effects:**
   - The shadowing effect, caused by obstacles that interfere with the signal, is modeled using a **Gaussian Process (GP)**.
   - GP is used to estimate the **variance** in path loss predictions, allowing the model to quantify uncertainty in real-world scenarios.
   - This step helps improve the reliability of the path loss model by incorporating probabilistic reasoning.

2.6.3 Measurement System and Experimental Setup

The dataset used for model training and evaluation was collected in suburban areas of Korea with various environmental factors. Measurements were taken at three different frequencies (450 MHz, 1450 MHz, and 2300 MHz), with the following setup:

- **A fixed transmitter** placed on a four-story building with an antenna height of 15 meters.
- **A mobile receiver** mounted on a vehicle with an antenna height of 2 meters.
- **Path loss data collected** over distances ranging from 1 km to 3.5 km.
- **Preprocessing techniques** such as normalization and cross-validation applied to improve model robustness.

2.6.4 Key Findings and Performance Evaluation

The paper evaluates the proposed model using several performance metrics, including:

- Root Mean Squared Error (RMSE)
- Mean Absolute Error (MAE)
- Mean Absolute Percentage Error (MAPE)
- Coefficient of Determination (R^2)

Key observations include:

- The ANN-MLP model performs better than the classic log-linear models and two-ray models with results providing the smallest RMSE and highest R^2.
- PCA-based feature selection reduces the computation time without compromising on the prediction accuracy.
- Gaussian Process modeling results in better confidence intervals for shadowing effects, thereby allowing more dependable path loss estimation.
- With ANN, the sigmoid activation function resulted in the best performance.

2.6.5 Our Findings

Inspired by this study, we tried to **c**ouple PCA with ANN in our own investigation for path loss prediction. But we could not achieve satisfactory results. Some reasons for this may be:

- Trade-offs for Feature Reduction: PCA may have eliminated important features which have caused the most variations in the path loss.
- Nonlinearity of Path Loss Data: While PCA captures linear correlations, it may not well handle nonlinear dependencies in the data.
- Overfitting in ANN: The high variance nature of ANN models requires a careful balance of training data size, network architecture, and regularization techniques.
- Environmental Differences: The dataset on which the paper was conducted was gathered in a particular suburban region of Korea, while my experiment was carried out on some other terrain, environmental factors, or frequency bands.

- The Work not being done on Satellite: The dataset on which the paper was conducted was not for StG (Satellite to Ground) communication path loss. Since the Bangladesh Satellite-1 is in Line-of-Sight for both the ground stations, we excluded the Gaussian Process for shadowing effects. After all this adjustments, however, this process did not perform well.

2.6.6 Conclusion: Accepting the Study but Moving Ahead with Other Methods

Although the methodology is well-structured and innovative as proposed in the paper, my practical application of PCA with ANN did not give good results. So it may appear to be indicating that:

- Feature selection strategies should be carefully assessed to ensure essential information is retained.
- ANN models may not always benefit from PCA, especially if the dataset's nonlinear characteristics require direct learning from raw features.
- Alternative feature engineering techniques, such as autoencoders or deep learning-based dimensionality reduction, may be more effective for complex path loss modeling.

In conclusion, this paper provides valuable insights into machine learning-based path loss prediction. However, based on my findings, PCA and ANN together may not always be the best approach for every dataset and environment. Further exploration of nonlinear feature selection methods and adaptive learning techniques may yield better results in future work.

Chapter 3

Methodology

The proposed methodology of the study contains two approaches: a machine learning-based path loss prediction model and a mathematical modeling approach based on the ITU-R recommendations. The objective is to analyze and compare the performance of these two techniques to find out which one provides the most accurate and reliable path loss estimates. Both approaches were previously tested and refined to realistically mimic field conditions for an objective assessment of their effectiveness in predicting satellite communication path loss.

3.1 Machine Learning Approach

3.1.1 Dataset Preparation

One of the toughest but crucial aspects of this work was preparing the dataset. Initially, obtaining real satellite attenuation data was extremely challenging as it was not readily available. To circumvent this limitation, we first generated a synthetic dataset, attempting to mimic realistic attenuation patterns from theoretical models and environmental parameters that we understand. This synthetic dataset gave us a starting point to evaluate different machine learning models.

We began by attempting to train an Artificial Neural Network (ANN) model on this synthetic data set after applying Principal Component Analysis (PCA) to decrease the dimensionality. The outcomes using this approach were not satisfactory. The ANN model failed to generalize, which meant the synthetic data set likely did not capture all the complexity of real-world satellite signal attenuation. To handle this, we experimented with several other machine learning models, each designed to address different aspects of the prediction issue. The idea was that if a model performed well on the synthetic data, it might perform even better when we had actual real-world data.

We were eventually able to procure an actual dataset of satellite attenuation measurements spanning several years from the Bangladesh Communication Satellite Company Limited (BCSCL). These logs had the attenuation values measured on specific dates, providing us with useful historic information for our study. Attenuation alone was not sufficient for the purpose of making accurate predictions, however, as it is significantly meteorology-dependent. Therefore, we also gathered weather data for the respective dates from the Bangladesh Meteorological Department (BMD). The weather database had crucial atmospheric parameters such as temperature, humidity, precipitation, and cloud cover—parameters known to influence signal attenuation.

Having obtained both datasets, the initial step was to merge them into a single complete dataset. This involved concatenating the meteorological data and attenuation measurements based on timestamps to align them. Preprocessing was carried out carefully to handle missing values, eliminate inconsistencies, and normalize feature scales for improved model performance.

With this real dataset, we were able to re-train and re-test our models to get more accurate and realistic predictions of satellite path loss. The incorporation of real attenuation and meteorological data added a lot of credibility to our machine learning-based predictions, demonstrating the benefit of utilizing real-world measurements over synthetic approximations.

3.1.2 Model Performance Analysis

In our study, we attempted a number of machine learning models for path loss prediction. The following sections report attempts conducted by us, including the methods taken and, most significantly, all improvements accomplished and results achieved.

Attempt 01: ANN with PCA

We employed a machine learning pipeline that utilizes PCA for dimensionality reduction and an ANN for regression to predict Path Loss (dB) from various geographic and environmental parameters.

PCA is a technique that transforms correlated features into a lower dimensional space of uncorrelated principal components with maximal variance. It assists in noise reduction and improving computational efficiency. ANN, which is inspired by the human brain, is used in classification and regression tasks. It consists of layers of neurons that function on inputs through weighted connections, activation functions, and backpropagation to minimize errors.

Despite our efforts, this approach did not yield desirable outcomes. The final Mean Squared Error (MSE) was 78.35.

Attempt 02: Improved ANN with Regularization

To improve our model further, we incorporated regularization techniques, dropout layers, and early stopping to prevent overfitting and improve generalization.

Key Improvements:

- Regularization (L2), dropout layers, and early stopping to prevent overfitting.
- More controlled training process with early stopping, limiting excess epochs.
- Added scatter plots for actual vs. predicted comparison.

These modifications led to a reduction in MSE to 64.12.

Attempt 03: Hybrid Approach with Voting Regressor

Trying to improve our model further, we hybridized ANN, Random Forest, and a Voting Regressor with PCA, a hybrid machine learning model.

Key Enhancements:

- Combined ANN with Random Forest for better generalization.
- Random Forest performed well in non-linear relationship capturing.
- ANN was hyperparameter-tuned using ReduceLROnPlateau to dynamically change learning rates.
- Voting Regressor combined both model strengths, reducing bias and variance.
- Enhanced visualization with scatter plots and learning curves.

The findings show that Random Forest was way better than ANN. In fact, the optimal MSE in the model for this issue was 8.63 for Random Forest, while the optimal MSE for ANN was 93.45.

Attempt 04: Random Forest Model

In this attempt, we abandoned ANN and ensemble learning, and we only used the Random Forest Regressor (RF) as our main model.

Key Takeaways:

- Simpler approach with quicker training and assessment.
- Simpler hyperparameter tuning (i.e., number of estimators, max depth, etc.).
- Feature scaling is not required, which simplifies preprocessing.
- Random Forest showed better performance on structured tabular data.

This last method achieved somewhat less error than previously, which cemented Random Forest as the highest-performing model in our research.

Attempt 05: Enhanced Random Forest Model

Our new Random Forest model has significant changes and improvements over the previous version. We have significantly expanded our feature set, added categorical encoding, and adjusted the PCA components to match the increased input dimensions. We worked with a much larger synthetic dataset this time, which is a great step toward dealing with real data. Because the real dataset will have the same features.

Aspect	Previous Model	Updated Model
Number of Features	6	26+ (including environmental and satellite parameters)
Categorical Encoding	Not included	One-hot encoding for categorical features
PCA Components	6	27 (to match the expanded feature space)

Table 3.1: Comparison between the previous and updated Random Forest models

Comparative Analysis:

Key Takeaways:

- **More features** → Potentially higher accuracy
- **Categorical encoding** → Better handling of non-numeric features
- **More PCA components** → Retains more variance

This time, the efficiency stood around **73%**.

Attempt 06: Introducing XGBoost alongside Random Forest

Our latest model introduces XGBoost (`XGBRegressor`) alongside Random Forest (RF), enabling a comparative analysis between the two models. From this attempt and on, we used the real dataset that we prepared from BCSCL and BMD.

Comparative Analysis: Random Forest vs. XGBoost Model

Key Takeaways:

- **XGBoost Introduced**: More advanced boosting model to improve predictive accuracy.

Feature	Previous Model (RF only)	Current Model (RF + XGBoost)	Advancements
Model Types	Random Forest (RF)	RF + XGBoost (`XGBRegressor`)	Added Gradient Boosting Model
RF Hyperparameters	`n_estimators=100`	Same	No change
XGBoost Added	Not present	`n_estimators=100`, `random_state=42`	New model for boosting performance
Performance Metrics	Only RF results	RF vs. XGB comparison	Enables direct performance evaluation
Visualizations	RF predictions only	Separate plots for RF & XGB	Provides clear model comparison
Computational Cost	Moderate	Higher (XGB takes longer than RF)	XGB requires tuning

Table 3.2: Comparison between the Random Forest and XGBoost models

- **Direct Model Comparison**: Helps determine if XGBoost outperforms RF.
- **Same Preprocessing Pipeline**: Ensures a fair comparison.

Through this experiment, we observed that XGBoost outperformed the Random Forest model in terms of accuracy. XGBoost achieved 82% accuracy, whereas Random Forest provided 73% accuracy.

Attempt 07: Optimized XGB with Stacking

Our updated Current Model introduces Optimized XGB with Stacking.
Comparative Analysis-

Aspect	Previous Code	Updated Code
Models Used	Random Forest, XG-Boost	Random Forest, Optimized XGBoost, Stacking Regressor
Hyperparameter Tuning	None	Added Randomized-SearchCV for XGBoost
Feature Importance	Not included	Visualized feature importance (with/without PCA)
Dimensionality Reduction	Fixed PCA component count	Dynamic PCA component count based on feature availability
Ensemble Learning	Independent models	Stacked ensemble with Linear Regression as meta-learner
Result Visualization	Separate plots for each model	Visual comparisons for optimized and stacked models

Table 3.3: Comparative Analysis of Previous vs. Updated Model

What This Model Does Better

- **Optimized XGBoost**: RandomizedSearchCV improves learning rate, depth, and estimators.
- **Feature Importance Analysis**: Helps understand which variables drive predictions.
- **Stacking Regression**: Combines Random Forest + XGBoost, leading to higher accuracy.

Performance Comparison-

Model	Mean Squared Error (MSE)	R^2 Score
Initial XGBoost	mse_xgb	r2_xgb
Optimized XGBoost	mse_best_xgb	r2_best_xgb
Stacked Model (RF + XGB + Linear Regression)	mse_stacked	r2_stacked

Table 3.4: Performance Comparison of Models

The initial XGBoost Model used to offer 89.62% accuracy, whereas this model exceeds it and stands at 90.62% accuracy.

Attempt 08: Optimized Random Forest Model

This time, we have gone back to the Random Forest Model with some noticeable changes. The model has been significantly enhanced through several key optimizations. Initially, a standard Random Forest model using raw features and default settings was employed. The current optimized model now incorporates one-hot encoding for categorical data, tuned hyperparameters via RandomizedSearchCV, and feature importance visualization, resulting in improved accuracy and a better understanding of key drivers, while maintaining standard performance tracking with MSE and R^2.

Comparative Analysis-

Aspect	**Previous Code (XGBoost + Stacking)**	**Current Code (Optimized Random Forest)**
Model Types	XGBoost, Random Forest, Stacking	Random Forest only
Optimization	RandomizedSearchCV for XGBoost	RandomizedSearchCV for Random Forest
Dimensionality Reduction	PCA applied	No PCA (uses all features with feature importance)
Feature Importance	Based on PCA components	Based on original features (more interpretable)
Model Complexity	Higher (ensemble stacking)	Lower (single optimized model)
Training Speed	Slower (due to stacking and XGBoost)	Faster (Random Forest is quicker to train)
Result Visualization	Plots for both models (XGBoost + Stacking)	Single plot for Random Forest predictions

Table 3.5: Comparative Analysis of Previous vs. Current Model

What This Model Does Well

- **RandomizedSearchCV Optimization**: Tunes key hyperparameters for better model performance.
- **Feature Importance Analysis**: Identifies the most significant predictors in the model.
- **Visual Performance Evaluation**: Displays scatter plot for actual vs. predicted values, aiding in performance visualization.

Performance Comparison-

Model	**Mean Squared Error (MSE)**	$\mathbf{R^2}$ **Score**
Initial RF Model	mse_rf	r2_rf
Optimized RF Model	mse_best_rf	r2_best_rf

Table 3.6: Performance Comparison of Initial and Optimized RF Models

The optimized RF model should outperform the initial RF model in terms of lower MSE and higher R^2.
This optimized RF model has shown an impressive 91.36% accuracy. This much high accuracy is for training the model with the "Free Space Path Loss", not for other attenuations. Our thought was- if a model does not perform well for the Free Space Path Loss prediction, then it will have no chance for the Total Attenuation. After training the model with Total Attenuation, the accuracy rate has dropped down to 80.432%, which is actually promising.