DATA ANALYTICS AND MACHINE LEARNING ON WEATHER IN SINGAPORE AND IMPACT ON SENSORS PERFORMANCE

Liang Yu Fei[1], Poh Kai Li[2], Valerie Chia Boon Hwan[3], Koh Geng Hao Bernard[4]
[1] Nanyang Girls' High School, 2 Linden Drive, Singapore 288683
[2] National Junior College, 37 Hillcrest Rd, Singapore 288913
[3] River Valley High School, 6 Boon Lay Ave, Singapore 649961
[4] Defence Science and Technology Agency, 1 Depot Rd, Singapore 109679

ABSTRACT
Weather attenuates radiofrequency, hence a sensor's performance is significantly affected by the weather, which can vary from day to day and follow certain trends and patterns in a year. This study utilises historical weather data sets of Singapore available to analyse and infer patterns that can be observed in the rainfall and temperature and subsequently the signal attenuation a sensor experiences. We developed a Machine Learning model using Logistic Regression and XGBoost algorithms to study the relationship between the parameters of temperature, wind speed and dates to predict the daily rainfall amounts expected in order to gauge the sensor's attenuation on specific days. This can be applied in sensors used in island air defence systems and for the optimisation of sensor performance under various weather conditions.

1 INTRODUCTION

1.1 Background & Objective

Radiofrequency (RF) plays an important role in our military defence as many sensors used in electronic warfare (EW) rely on RF to receive and send signals. EW is responsive to threat signals in its environment that are detected by receivers and then processed for operators to recognise. However, the concentration of atmospheric aerosols affected by weather can attenuate RF, leading to the performance of sensors being affected as well as inaccurate data. Understanding the climatic conditions of Singapore is therefore essential to finding out weather's impact on sensors' performance. We aim to analyse Singapore's weather, assess its impact on the RF spectrum, and forecast rainfall. Additionally, we intend to use the Machine Learning (ML) model's role in predicting the impact of weather conditions on RF sensor performance.

1.2 Fundamentals of Radiofrequency

Radio frequency is a rate of oscillation in the range of around 3 kHz to 300 GHz, which corresponds to the frequency of radio waves. [1]

<u>Conversion to dB form</u>

$$N = 10^{N(dB)/10}$$

N (dB) is the dB form of the linear number, N

dB (Decibel) is the difference (or ratio) between two signal levels. It is used to describe the effect of system devices on signal strength. For example, a cable has 6 dB signal loss or an amplifier has 15 dB of gain. This is useful since signal strengths vary logarithmically, not linearly. Since the dB scale is a logarithmic measure, it produces in simple numbers for large-scale variations in signals. It is very useful because system gains and losses can be calculated by adding and subtracting whole numbers.[2]

Free Space Path Loss

The propagation of all radio signals is subject to Free Space Path Loss (FSPL), which is a mathematical definition of the geometric property that the further away you are located from the source of a radio transmission, the energy level in that signal drops as a function of the square of the distance.[3]

Common dB Definitions

dBm	= dB value of Power / 1 milliwatt	Used to describe signal strength
dBW	= dB value of Power / 1 watt	Used to describe signal strength
dBsm	= dB value of Area / 1 meter^2	Used to describe antenna area or radar cross-section
dBi	= dB value of antenna gain relative to the gain of an isotropic antenna	0 dBi is, by definition, the gain of an omnidirectional (isotropic) antenna

Table 1.2.1

Free-Space Path Loss Formula[3]

The free-space path loss (FSPL) formula derives from the Friis transmission formula. This states that in a radio system consisting of a transmitting antenna transmitting radio waves to a receiving antenna, the ratio of radio wave power received P_r to the power transmitted P_t is:

$$\frac{Pr}{Pt} = D_t D_r \left(\frac{\lambda}{4\pi d}\right)^2$$

- D_t is the directivity of the transmitting antenna
- D_r is the directivity of the receiving antenna
- λ is the signal wavelength
- d is the distance between the antennas

(ref. Appendix 1 for more info)

1.3 Weather Impact on Electromagnetic (EM) Waves

Singapore has a tropical climate, which leads to abundant rainfall and high humidity and temperature. There are 2 monsoon seasons, the Northeast Monsoon from December to early March, and the Southwest Monsoon from June to September. Afternoon thunderstorms are also common throughout the year, especially during the inter-monsoonal periods from late March to May and October to November, caused by strong surface heating and by the sea breeze circulation that develops in the afternoon. These high precipitation rates and temperature affect the atmospheric attenuation that a sensor experiences.

Atmospheric attenuation is a reduction in the intensity of EM radiation in the Earth’s atmosphere as a result of the absorption and scattering of the radiation, which occurs mostly due to the presence of hydrogen and oxygen molecules in the atmosphere. The amount of these molecules (humidity) is largely affected by precipitation rates and temperature. Temperature affects the capability of air to hold water molecules. An increase in temperature causes air to expand and its ability to hold more water molecules, leading to an increase in humidity.

High relative humidity (the amount of water vapour present in air expressed as a percentage of the amount needed for saturation at the same temperature) occurs when the air temperature approaches the dew point value, which is the temperature at which air reaches saturation by water molecules. Singapore’s temperature is usually higher than that of its dew point (24°C) and hence Singapore has high humidity and high moisture content.

Humidity is also affected by wind speed. The faster the wind, the faster water evaporates to become water vapour in the atmosphere hence the humidity increases, causing greater atmospheric attenuation. Additionally, wind speed affects the refraction capabilities of the medium which leads to aberration in radio propagation. Winds affect telecommunication signals as Ultra-High Frequency of radio waves in the KU band spectrum are of shorter wavelengths and attenuations in these radio paths are mostly by absorption and scattering of dust and sand particles caused by the movement of wind.[6] However, due to the fact that Singapore is not near to desert regions, wind speed is likely not a major factor in atmospheric attenuation.

Humidity affects atmospheric attenuation as it affects the air density which can cause radio signals to refract, or bend, which can affect the direction and distance in which they travel. High frequencies have shorter wavelengths and when the wave hits an obstacle like water molecules in the atmosphere, the water molecule absorbs energy from the wave and scatters the rest of it, hence weakening the signal. A higher relative humidity means that there will be more signal loss. The EM wave attenuation due to rain (rainfall attenuation) is one of the most noticeable components of excess losses, especially at frequencies of 10 GHz and above.

1.4 Impact of Sensors Performance

Sensors are classified into two categories, mainly active and passive sensors. The window bands in infrared (IR) and RF portions of the EM spectrum where most sensors operate are listed below. [8]

IR: 3 - 5 µm, 8 - 12 µm
RF: 94GHz, 35GHz, 10GHz

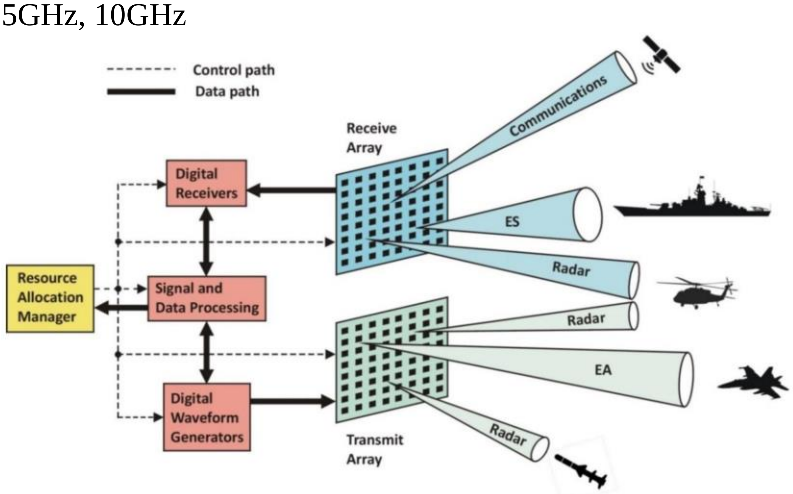


Fig 1.4.1 [7]

Passive sensors only require the use of a receiver. They detect EM energy emitted by other sources in their field of view. These include electronic support measures, passive remote sensing and passive electro-optical sensors. Active sensors transmit their own EM waves into the terrain, interacting with the terrain which produces a backscatter of energy detected by the sensor. These include radar, communications, active remote sensing and active electro-optical sensors.

As RF signals travel through rainwater, the EM waves in the atmosphere are absorbed by water molecules. This is known as rain fade which mostly affects frequencies higher than 11 GHz. [10] EM interference can cause systems to be unable to send or detect signals, as well as disrupting communication devices; in the case of EW, change in transmitted frequency may cause threat signals from received RF signals to go unidentified.

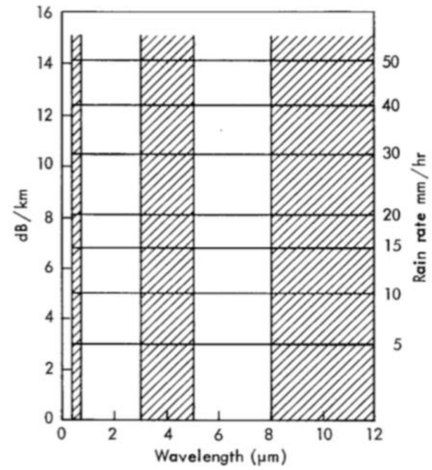


Fig 1.4.2 [8]

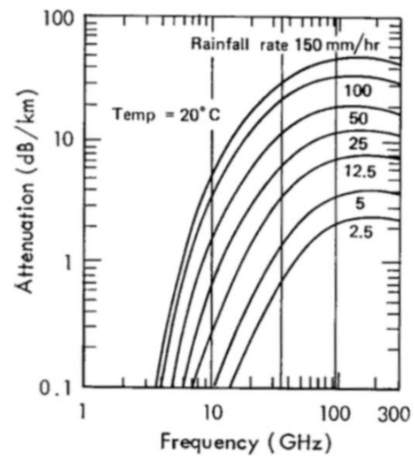


Fig 1.4.3 [8]

Figure 1.4.2 plots optical attenuation at different rain rates against EM wavelength. Optical attenuation due to the various rainfall rates remains constant throughout the IR wavelengths, which can be accounted for easier. Figure 1.4.3 plots RF attenuation at different rain rates against RF frequencies. At 10 GHz, RF attenuation is almost insignificant for rainfall rates of 2.5 to 50mm/hr. However, at 35GHz and 94GHz, there is significant RF attenuation which renders the reliability of RF sensors under heavy rainfall.

Radar systems use radio waves to determine the distance and velocity of targets they hit. A radar system usually consists of a transmitter that sends out radio signals and a receiver that catches reflected energy from the target.

In EW, sensors are used for RF threat identification, where threats are identified from the parameters of received RF signals. The parameters of a threat signal include, effective radiated power, antenna pattern, antenna scan type, antenna scan rate, transmitted frequency, types of modulation and modulation parameters.[9] A higher frequency is subject to greater atmospheric attenuation as shown in the graph above.

Under rainfall conditions, RF sensors would face a significant signal attenuation. To overcome it, a suite of systems (consisting of Electro Optics Sensors, radio communication and radar) will be optimal to provide situation awareness under rainfall conditions. However, electro optic sensors have a short detection range, therefore under heavy rainfall conditions, it will not be as effective due to the field of view being obstructed.

As a tropical country that has abundant rainfall, it is essential to implement measures to mitigate the influence of rainfall on sensor performance. Understanding the trends of Singapore's rainfall can allow us to use ML models to predict RF attenuation as well as implement power adaptive management strategies for sensors, allowing them to dynamically adjust transmission power based on predicted rainfall attenuation. This ensures optimal performance under varying weather conditions.

2 METHODOLOGY

2.1 Data Analytics of Past Weather

2.1.1 Data Analytics of Weather across different regions of Singapore

In order to understand the weather conditions of Singapore, we utilised historical weather data sets from Changi, Ang Mo Kio and Tuas South to observe the annual rainfall of these 3 regions. The locations were chosen to represent the East, Central and West regions of Singapore as they had the most complete set of datasets from 2014 to 2020. Together, they give an overview of the difference in rainfall across Singapore and an insight into which locations would be most suitable for sensors to be placed.

2.1.2 Data Analytics of Historical Weather data sets from Changi to observe trends for rainfall and temperature and obtaining rainfall attenuation values

Historical weather data sets of Changi from 1983 to 2022 were used as Changi is the only location in Singapore to have records of daily rainfall rates from 1983 onwards, providing us with a large enough sample size to observe trends across the years. Using Power BI, we plotted graphs for the annual rainfall and mean temperature of Changi from 1983-2022, as well as the average amount of rainfall per day in each month. Using those values, we used the average RF attenuation due to rain using the formula below as the rainfall attenuation model.

$$L = kR^{\alpha} \cdot d_{eff} \text{ [13]}$$

Where R is rain rate (mm/h), k and α are linear polarisation and horizontal paths, and d_{eff} is effective propagation distance. The path elevation angle and polarisation tilt angles are assumed to be zero and the daily rain rate is assumed to have lasted for 1 hour as convectional rain (short but heavy rainfall) is the most common type of rain in Singapore. The effective propagation distance is assumed to be 1km.

2.2 Machine Learning Model

To predict future weather conditions and the impact of these sensors, we used a predictive ML model. We tested different ML models, logistic regression and XGBoost and combined the 2 best ML models: logistic regression to predict if it will rain or not *(ref. Appendix 2A)* and XGBoost *(ref. Appendix 2B)* to predict daily rainfall rate. XGBoost was chosen as it operates by employing a sequence of decision trees, where each tree learns from the errors made by its predecessors. If the initial tree makes a prediction mistake, the subsequent trees are designed to take that mistake into account and rectify it. This iterative process continues until reaching the final prediction, where the combined wisdom of all the trees is leveraged to deliver an improved and more accurate outcome. Our dataset includes mean wind speed, month, year, day and mean temperature of Changi from 1983 to 2020 to predict rainfall. 80% of data is used to train the model and the remaining 20% is used to test the effectiveness of our model.

3 RESULTS & DISCUSSION

3.1 Data Analytics of Weather across different regions of Singapore

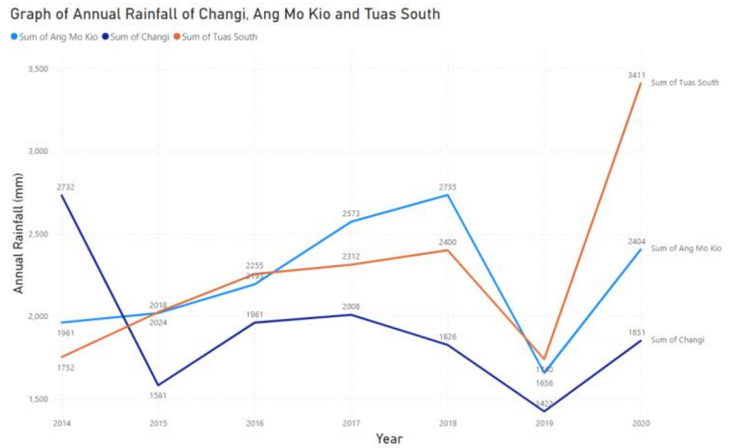


Fig 3.1.1

The annual rainfall rate differs in each location, especially in 2014 and 2020. Although there was a general increase in rainfall from 2014 to 2020 in all locations, the east region (Changi), generally has less rainfall than the western region (Tuas South). This is attributed to the rain shadow effect, which happens due to Bukit Timah Hill in the west. When air approaches one side of the hill, it rises and cools down and thus is unable to hold as much moisture, resulting in rainfall. When the air moves beyond the peak of the elevation point, the air warms as it sinks, hence there is less rainfall in the eastern region of Singapore. The eastern region becomes a rain shadow and is a drier region as a result of the rain shadow effect [11].

This means that the East side of Singapore is less affected by rainfall and experiences less RF attenuation, hence is more suitable for the use of RF sensors. Given the influence of rainfall on the Tuas South region, a strategic approach would be to install sensors in the North-west region of Singapore to address the impact of rainfall.

3.2 Data Analytics of Past Weather

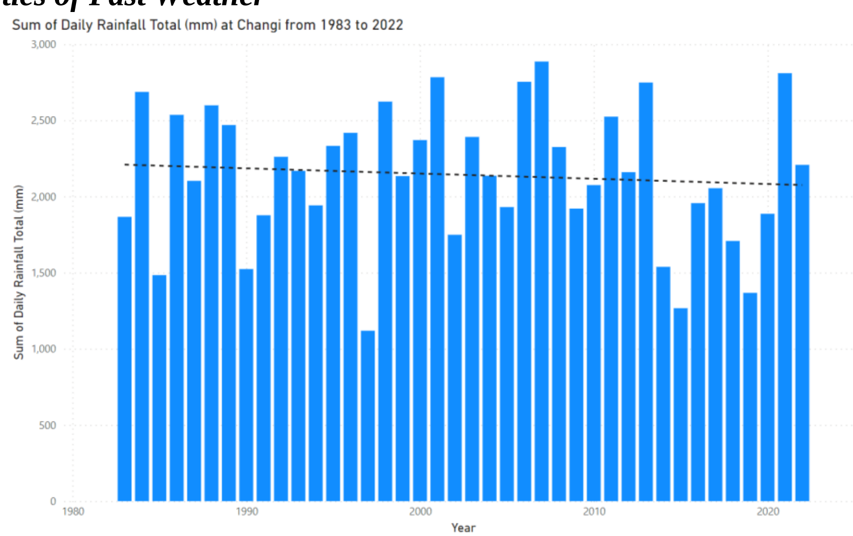


Fig 3.2.1

Fig 3.2.1 shows the annual rainfall of Changi. The overall annual rainfall rate at Changi shows a slight decrease from 1983 to 2022. This can be attributed to global warming, where a warmer climate causes water vapours in the atmosphere to take longer to condense into clouds and fall as rain, decreasing the amount of rain over land and increasing the amount of rain over oceans.[12]

The rise and fall of precipitation in certain years also correspond to El Niño (a climate pattern that occurs every 2-7 years, bringing drier weather to Singapore) and La Niña (a weather pattern that brings warmer-than-normal temperatures to Singapore, resulting in higher precipitation rates).

Month	**Jan**	**Feb**	**Mar**	**Apr**	**May**	**Jun**	**Jul**	**Aug**	**Sep**	**Oct**	**Nov**	**Dec**
Avg Rainfall per day (mm)	7.66	3.75	5.19	5.29	5.31	4.57	4.81	4.80	4.97	5.35	8.69	9.81
L **at 10 GHz (dB)**	0.3070	0.1381	0.1986	0.2029	0.2037	0.1723	0.1824	0.1820	0.1892	0.2054	0.3536	0.4052
L **at 35 GHz (dB)**	3.5830	2.0093	2.6138	2.6545	2.6626	2.3580	2.4577	2.4536	2.5237	2.6788	3.9690	4.3792
L **at 94 GHz (dB)**	7.9580	5.1133	6.2527	6.3271	6.3419	5.7791	5.9652	5.9575	6.0873	6.3714	8.6053	9.2770

Using the formula in section 2.1.2, the average RF attenuation *L* (dB) of each month is calculated.

Table 3.2.1

Average daily rainfall per month from 1983-2022 and RF attenuation at 10, 35 and 94 GHz in different months

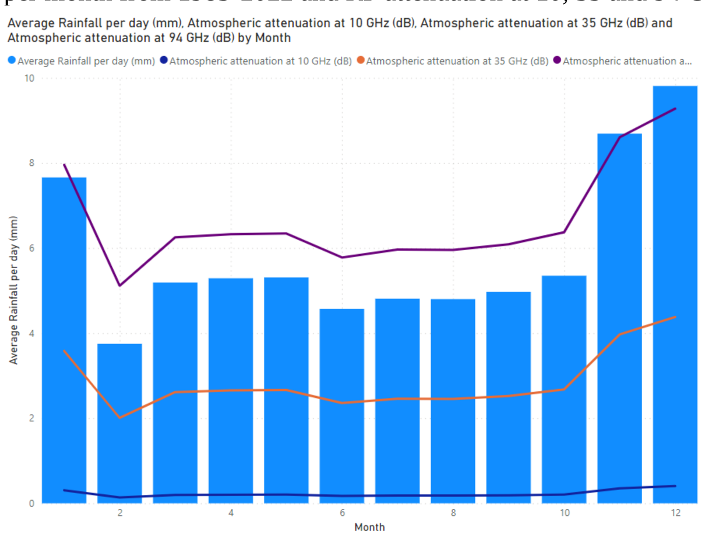


Fig 3.2.2

Figure 3.2.2 shows the average daily rainfall for each month from 1983 to 2022. December, November and January are the months with the highest amount of rainfall with February having the lowest amount. This is due to the Monsoon seasons, Northeast Monsoon from December to early March, and the Southwest Monsoon from June to September. The RF attenuation shows a similar pattern in variation for 10, 35 and 94 GHz and the trend are similar to that of the daily rainfall rate, with 94 GHz having the highest amount of attenuation, followed by 35 and 10 GHz. In our frequency selection for analysis, we included 10 GHz to mirror the frequency of commercial radar. Additionally, we incorporated 35 GHz, to position it as a midpoint between 10 and 94 GHz. This choice enhances our ability to better visualise attenuation trends. Lastly, the inclusion of 94 GHz allows us to investigate attenuation patterns at higher frequencies, providing a comprehensive perspective on the impact of rainfall across the spectrum. Hence with these values, we can conclude that rainfall directly affects the RF attenuation a sensor experience. The observed trend suggests that the higher the rainfall rate, the higher the attenuation the sensor experiences. Following the variations in the daily rainfall, sensors used in February would experience the least attenuation and sensors used in the months of December to January would experience the most attenuation.

3.3 Machine Learning Model

To check the accuracy of our logistic regression model, we used the metrics function to calculate our accuracy, recall and precision. *(ref. Appendix 2A)* From Table 3.3.1, we found that our logistic regression model has an accuracy of 70.28% and high precision and recall, where precision is the ratio of correctly predicted positive observations to the total predicted positives while recall is the ratio of correctly predicted positive observations to all positive observations in the actual class. Therefore, the high precision and recall of the model indicate that a high proportion of positive identifications and actual positives were identified correctly. The F1 score is a metric that combines both precision and recall into a single value, providing a balance between these two metrics. Our model has an f1 score of 0.699 which indicates a good balance between precision and recall. From the ROC Curve, we can see that the AUC is 0.778 which signifies our model has a good performance in predicting whether a day will rain or not. Hence our logistic regression model is suitable to predict future weather.

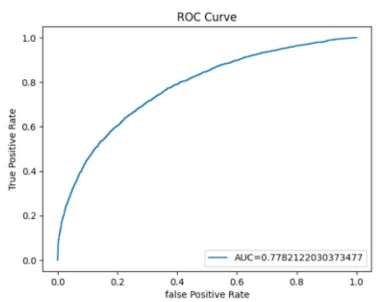
<table>
<tr><td>Accuracy of log model</td><td>0.703</td></tr>
<tr><td>Recall of log model</td><td>0.760</td></tr>
<tr><td>Precision of log model</td><td>0.757</td></tr>
<tr><td>f1_scoreof log model</td><td>0.699</td></tr>
<tr><td>ROC Curve</td><td>
</td></tr>
</table>

Table 3.3.1

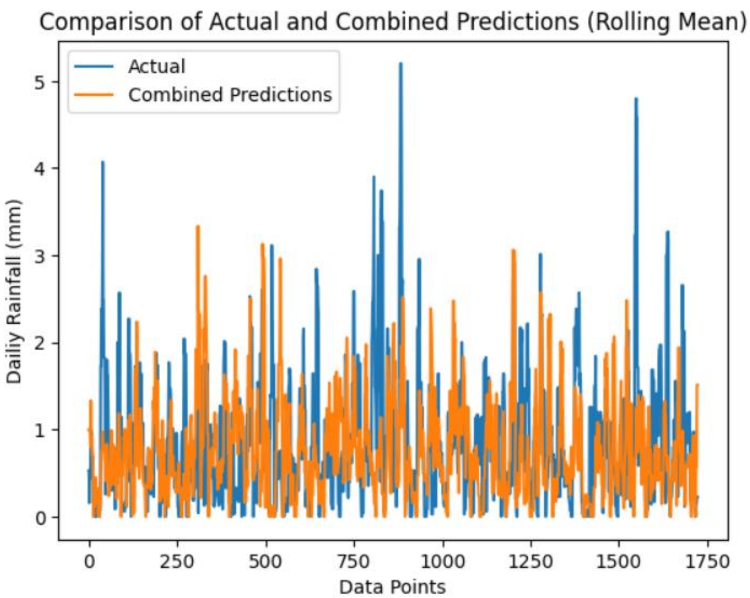
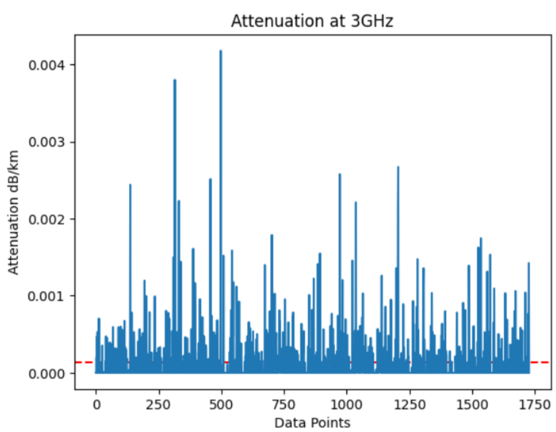
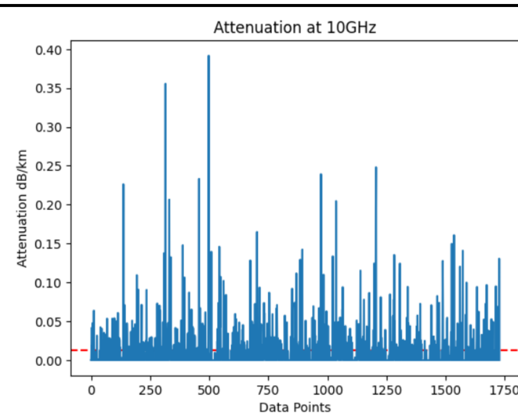


Fig 3.3.1

Figure 3.3.1 shows the rainfall our combined logistic regression and XGBoost model predicted against the actual rainfall amount. The mean squared error of our combined model is very low at 4.548, signifying that our rainfall prediction is close to the actual rainfall amount. However, our model is unable to predict outliers in rainfall due to the sensitivity of XGBoost to outliers.





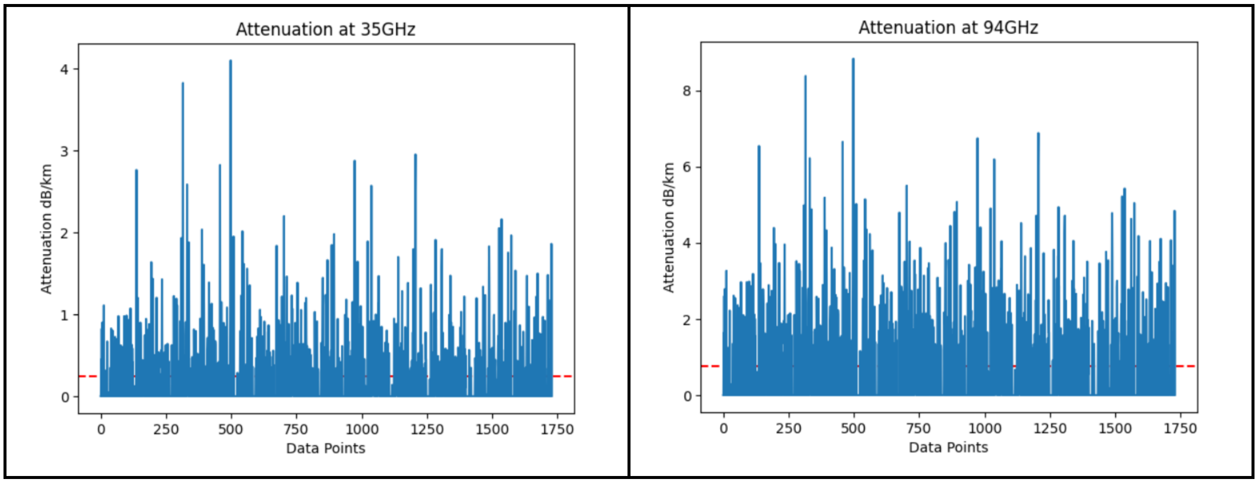


Table 3.3.2

Linking back to RF attenuation, the graphs in table 3.3.2 shows the calculated RF attenuation *L* for a 1km distance on different data points at different frequencies, with each data point being a certain date from 1983 - 2020. It can be seen that average RF attenuation for 3 and 10 GHz are insignificant while that at 35GHz and 94GHz are more likely to be affected, similar to our findings in section 1.4. Therefore, during heavy rainfall, sensors work best to operate at low frequencies of 10GHz and below due to the minimal attenuation and hence minimal disruption to sensor performance.

4 CONCLUSION & FUTURE WORK

Overall, Singapore being a tropical country is subjected to constant rainfall, therefore it is an underlying concern to the performance of RF sensors due to attenuation. This study has found that installing sensors of lower frequencies are recommended as they are less affected by rainfall attenuation, as well as positioning sensors in places with lesser rainfall. Our ML model can be used to predict the attenuation at different frequencies due to rainfall, helping relevant users to get an estimate of the RF attenuation. This helps to implement adaptive power management strategies for sensors, allowing them to dynamically adjust transmission power based on predicted rainfall attenuation, ensuring optimal performance under varying weather conditions. This information can be used to help in the planning for the sensors in island and defence where the different sensors can be operated as a suite of systems in order to strategically overcome the rainfall attenuation.

Future work of our model can include using weather data other than rainfall, such as haze, fog and clouds to form relationships and use these data points to improve our rainfall forecasting model, as well as using data from different parts of Singapore to predict the rainfall of those places.

5 ACKNOWLEDGEMENTS

We would like to acknowledge the efforts of our mentor, Bernard, for his invaluable support and guidance throughout our project. We would also like to thank Eric for sharing his expertise and knowledge and providing us with helpful suggestions.

REFERENCES

[1] RF Global Solutions Ltd (n.d). *What is radiofrequency (RF)*. Retrieved 2023, September 15 from https://www.rfglobalsolutions.co.uk/what-is-radio-frequency-rf/

[2] Young, M.F. (n.d). *Understanding Decibels and Their Use in Radio Systems.* YDI Wireless.

[3] Encyclopedia (n.d). *Free-space path loss*. Retrieved 2023, October 3 from https://encyclopedia.pub/entry/31362#:~:text=The%20free-space%20path%20loss%20(FSPL)%20formula%20derives%20from,%CE%BB%204%20%CF%80%20d%

[4] Netcontrol (n.d). *Radio Fade Margin.* Retrieved 2023, October 3 from https://www.netcontrol.com/services/radio-networking-tools/radio-fade-margin/

[5] Electricity - Magnetism (n.d). *Directivity of an antenna equation.* Retrieved 2023, October 20 from https://www.electricity-magnetism.org/directivity-of-an-antenna-equation/

[6] Chima, A.I., Onyia, A.I., Udegbe, S.U. (2018, March). *The Effects of Atmospheric Temperature and Wind Speed on Uhf Radio Signal.* Retrieved 2023, November 10 from https://www.iosrjournals.org/iosr-jap/papers/Vol10-issue2/Version-1/L1002018390.pdf

[7] Moo, P.W. and DiFilippo D.J. (2018, June). *Multifunction RF Systems for Naval Platforms.* Retrieved 2023, December 12 from https://www.mdpi.com/1424-8220/18/7/2076

[8] Chen, C.C. (1975, April). *Attenuation of Electromagnetic Radiation by Haze, Fog, Clouds, and Rain.* Retrieved 2023, November 17 from https://www.rand.org/content/dam/rand/pubs/reports/2006/R1694.pdf

[9] Adamy, D (2001). *A First Course in Electronic Warfare.* Artech House.

[10] everythingRF (2021, May). *What is Rain Fade?* Retrieved 2023, December 5 from https://www.everythingrf.com/community/what-is-rain-fade#:~:text=Rain%20Fade%20is%20a%20phenomenon%20that%20negatively%20affects,signals%20that%20have%20frequencies%20higher%20than%2011%20GHz.

[11] National Geographic (n.d). *Rain Shadow.* Retrieved 2023, December 5 from https://education.nationalgeographic.org/resource/rain-shadow/

[12] Staedter, T. (2005). *Warmer Climate Produce Less Rain.* Retrieved 2023, December 6 from ttps://www.scientificamerican.com/article/warmer-climate-produces-l/#:~:text=New%20climate%20simulations%20from%20NASA%20show%20that%20under,amount%20decreased%20over%20land%20but%20increased%20over%20oceans.

[13] MathWorks (n.d). RF signal attenuation due to rainfall. Retrieved 2023, December 6 from https://www.mathworks.com/help/phased/ref/rainpl.html

APPENDICES

Appendix 1: RF fundamentals

<u>Fade Margin</u>

Fade Margin = PRX – Rx Sensitivity

Fade Margin is an expression for how much margin – in dB – there is between the received signal strength level and the receiver sensitivity of the radio.[4] It's typically added to the received signal strength to ensure reliable communication under adverse conditions.

<u>Link Budget</u>

Received power (dBm) = transmitted power (dBm) + gains (dB) − losses (dB)

In wireless communication, there is a quantity used to describe how much power arrives at a wireless receiver: the link budget. In a wireless system connecting two devices, the link budget is intended to account for all sources of gain and loss that will impact power delivery to a receiving antenna.

<u>Antenna directivity</u>

Antenna directivity refers to the ability of an antenna to focus its radiation pattern in a particular direction, rather than radiating energy uniformly in all directions. A higher directivity corresponds to a more focused radiation pattern, which is advantageous for long-distance communication, as it maximises the signal strength in the desired direction while minimising it in others.[5]

The directivity of an antenna can be mathematically expressed using the following equation:

$$\mathrm{D}(\theta,\varphi) = U(\theta,\varphi) \div Pavg$$

Where:

- $\mathrm{D}(\theta,\varphi)$ is the directivity of the antenna at angles θ (elevation) and φ (azimuth).
- $\mathrm{U}(\theta,\varphi)$ is the radiation intensity of the antenna in the direction defined by angles θ and φ.
- P_{avg} is the average power radiated by the antenna, which can be calculated by integrating the radiation intensity over all possible directions.

<u>Link budget equations</u>

One way range equation

$$P_r = \frac{P_t \cdot Gt \cdot G_r \cdot \lambda^2 \cdot \sigma}{(4\cdot\pi)^3 \cdot {R_{Tx}}^2 \cdot {R_{Rx}}^2} \text{ (wavelength)}$$

$$P_r = \frac{P_t \cdot Gt \cdot G_r \cdot c^2 \cdot \sigma}{(4\cdot\pi)^3 \cdot f^2 \cdot {R_{Tx}}^2 \cdot {R_{Rx}}^2} \text{ (frequency)}$$

$$P_r = 10 \cdot log_{10}\,[p_r] \text{ (convert to dBW)}$$

Bistatic radar where the transmitter and receiver are in different locations

Two way range equation

$$P_r = \frac{P_t \cdot Gt \cdot G_r \cdot \lambda^2 \cdot \sigma}{(4\cdot\pi)^3 \cdot R^4} \text{ (wavelength)}$$

$$P_r = \frac{P_t \cdot Gt \cdot G_r \cdot c^2 \cdot \sigma}{(4\cdot\pi)^3 \cdot f^2 \cdot R^4} \text{ (frequency)}$$

$$P_r = 10 \cdot log_{10}\,[p_r] \text{ (convert to dBW)}$$

Monostatic radar where the transmitter and receiver are in the same location

Where

$$\boldsymbol{p_r = received\ peak\ power\ (W)}$$
$$\boldsymbol{p_t = transmitted\ peak\ power\ (W)}$$
$$\boldsymbol{G_r = received\ antenna\ gain\ (ratio)}$$

$$G_t = \boldsymbol{transmitter\ antenna\ gain}\ (\boldsymbol{ratio})$$
$$\lambda = \boldsymbol{transmitted\ wavelength}\ (\boldsymbol{m})$$
$$f = \boldsymbol{frequency}\ (\boldsymbol{Hz})$$
$$\sigma = \boldsymbol{radar\ cross\ section}\ (\boldsymbol{m^2})$$
$$R = \boldsymbol{monostatic\ range\ to\ target}\ (\boldsymbol{m})$$
$$R_r = \boldsymbol{bistatic\ transmitter\ range\ to\ target}\ (\boldsymbol{m})$$
$$c = \boldsymbol{speed\ of\ light}\ (\boldsymbol{m/s})$$

Appendix 2

Appendix 2A: Logistic Regression Model

```
import pandas as pd
import numpy as np
from sklearn.linear_model import LogisticRegression
path = "/content/drive/MyDrive/Downloads/weather datasets/CHANGI EVERYTHING 1983-2020 - changi 1983-2020.csv"
df = pd.read_csv(path, encoding='utf-8-sig')
def logregmodel(X_train, X_test, y_train, y_test, newdp):
  R = []
  for i in range(len(df)):
    if df['Daily Rainfall Total (mm)'][i] == 0:
      R.append(0)
    else:
      R.append(1)
  def is_float(string):
      try:
          # float() is a built-in function
          float(string)
          return True
      except ValueError:
          return False
  df['Rain'] = R

  y_test = pd.Series(y_test)
  logmodel= LogisticRegression(solver='sag', max_iter = 100000)
  logmodel.fit(X_train, y_train)
  y_pred = logmodel.predict(X_test)
  yprednewdy=  logmodel.predict(newdp)
  return (y_pred.tolist(), yprednewdy )
```

Appendix 2B: Combined model of xgboost and logistic regression

```
def xgboost_model(newdy):
  import pandas as pd
  from sklearn.model_selection import train_test_split
  from sklearn.metrics import mean_squared_error
  from sklearn.preprocessing import MinMaxScaler
  import xgboost as xgb
  path = "/content/drive/MyDrive/Downloads/weather datasets/CHANGI EVERYTHING 1983-2020 - changi 1983-2020.csv"
  df = pd.read_csv(path, encoding='utf-8-sig')

  R = []
  for i in range(len(df)):
    if df['Daily Rainfall Total (mm)'][i] == 0:
      R.append(0)
    else:
      R.append(1)
  df['Rain'] = R

  df['Mean Wind Speed (km/h)'] = df['Mean Wind Speed (km/h)'].astype("string")
  df['Max Wind Speed (km/h)'] = df['Max Wind Speed (km/h)'].astype("string")
  wrong_index_mean = []
  wrong_index_max = []
  def is_float(string):
      try:
          # float() is a built-in function
          float(string)
          return True
      except ValueError:
          return False
  for i in df['Mean Wind Speed (km/h)']:
    if i.isnumeric() == False and is_float(i) == False:
      wrong_index_mean.append(df[df['Mean Wind Speed (km/h)'] == i].index)
  for i in df['Max Wind Speed (km/h)']:
    if i.isnumeric() == False and is_float(i) == False:
      wrong_index_max.append(df[df['Max Wind Speed (km/h)'] == i].index)
  wrong_index_max = wrong_index_max[0].append(wrong_index_max[-1])
  wrong_index_mean = wrong_index_mean[0].append(wrong_index_mean[-1])
  df.loc[wrong_index_mean,'Mean Wind Speed (km/h)'] = '0'
  df.loc[wrong_index_max,'Max Wind Speed (km/h)'] = '0'
  df['Mean Wind Speed (km/h)'] = df['Mean Wind Speed (km/h)'].astype("float")
  df['Max Wind Speed (km/h)'] = df['Max Wind Speed (km/h)'].astype("float")
  mean = (df['Mean Wind Speed (km/h)']).mean()
  max =  (df['Max Wind Speed (km/h)']).mean()
  df.loc[wrong_index_mean,'Mean Wind Speed (km/h)'] = 1000
  df.loc[wrong_index_max,'Max Wind Speed (km/h)'] = 1000
  df['Mean Wind Speed (km/h)'] = df['Mean Wind Speed (km/h)'].replace(1000,mean)
  df['Max Wind Speed (km/h)'] = df['Max Wind Speed (km/h)'].replace(1000,max)

  df_sorted = df.sort_values('Mean Temperature (ýC)')
  rainfall_diff_sorted = df_sorted['Daily Rainfall Total (mm)'].diff().abs()
  df_filtered_sorted = df_sorted[rainfall_diff_sorted <= 24]
  print("Length of data after filtering outliers in temperature:", len(df_filtered_sorted))
  df_sorted = df_filtered_sorted.sort_values('Mean Wind Speed (km/h)')
  rainfall_diff_sorted = df_sorted['Daily Rainfall Total (mm)'].diff().abs()
  df_filtered1_sorted = df_sorted[rainfall_diff_sorted <= 5]
  X= df_filtered1_sorted[["Month", "Year", "Day", 'Mean Temperature (ýC)', "Mean Wind Speed (km/h)"]]
  y = df_filtered1_sorted['Daily Rainfall Total (mm)']

  X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
  logX = df_filtered1_sorted.drop(['Station', 'Daily Rainfall Total (mm)','Highest 30 Min Rainfall (mm)','Highest 60 Min Rainfall (mm)',
'Highest 120 Min Rainfall (mm)', 'Daily Rainfall Total (mm)','Rain', 'Maximum Temperature (ýC)','Minimum Temperature (ýC)','Max Wind Speed (km/h)' ], axis=1)
  logy = df_filtered1_sorted["Rain"]
  X_trainlog, X_testlog, y_trainlog, y_testlog = train_test_split(logX, logy, test_size=0.2, random_state=42)
  newdy = pd.Series(newdy)
  import numpy as np
  newdy = np.array(newdy)
  newdy = newdy.reshape(1, -1)
  logpredy = logregmodel(X_trainlog, X_testlog, y_trainlog,y_testlog, newdy)[0]
  logprednewdy = logregmodel(X_trainlog, X_testlog, y_trainlog,y_testlog, newdy)[1]

  model = xgb.XGBRegressor(n_estimators=100, random_state=42)
  model.fit(X_train, y_train)
  y_pred = model.predict(X_test)
  counting = 0
  for i in y_pred:
    if i == 0:
      counting += 1
  combined_predictions = []
  set_to_zero = False
  counter = 0
  for i in logpredy:
    if i == 0:
      counter += 1
  for log, xgboost in zip(logpredy, y_pred):
      if log == 0 and xgboost<= 1:
          set_to_zero = True
          combined_predictions.append(0)
      elif xgboost <0:
          combined_predictions.append(0)
      else:
        combined_predictions.append(xgboost)
  y_test_list = y_test.tolist()
  combined_predictions_list = combined_predictions
  df_comparison = pd.DataFrame(list(zip(y_test_list, combined_predictions_list)), columns=['y_test', 'combined_predictions'])
  xgboostpredict = model.predict(newdy)
  print("predicted log value for new data point:", logprednewdy)
  print("predicted xgboost value for new data point:", xgboostpredict)

xgboost_model([12, 2023, 21, 28.0, 20.88])
```

Appendix 3: Checking the accuracy and reliability of model

Appendix 3A

```
from sklearn import metrics
from sklearn.metrics import classification_report, confusion_matrix
print('Accuracy:', metrics.accuracy_score(y_test, y_pred))
print('Recall: ',metrics.recall_score(y_test, y_pred, zero_division=1))
print('Precision:', metrics.precision_score(y_test, y_pred, zero_division=1))
from sklearn.metrics import f1_score
print('f1_score:',f1_score(y_test, y_pred, average='weighted'))
```

Appendix 3B

```
y_pred_proba= logmodel.predict_proba(X_test) [:,1]
print(y_pred_proba)
false_positive_rate, true_positive_rate, _ = metrics.roc_curve(y_test, y_pred_proba)
auc= metrics.roc_auc_score(y_test, y_pred_proba)
plt.plot(false_positive_rate, true_positive_rate,label="AUC="+str(auc))
plt.title('ROC Curve')
plt.ylabel('True Positive Rate')
plt.xlabel('false Positive Rate')
plt.legend(loc=4)
```