Models for RF & FSO Channel Attenuation Final Report

Tongtingna Qi
STUDENT NO. 1775622

November 21, 2024

Report submitted for **Data Science Research Project** at the School of Mathematical Sciences, University of Adelaide

Project Area: **Hybrid RF/FSO Communication Channel Model**
Project Supervisor: **Siu Wai Ho**

Abstract

This report is a model analysis of how weather conditions influence the attenuation in a hybrid Radio Frequency/Free Space Optical (RF/FSO) communication channel. There are 25 predictors in the dataset for this report such as absolute humidity, distance, particulate, rain intensity, temperature, visibility, and wind speed. After data pre-processing, a random forest is used as the main machine learning algorithm. Moreover, coarse tuning and fine-tuning are applied to hyperparameter optimization, and the final decision tree is calculated after the tuning process. Furthermore, two types of models are implemented using the Out-of-bag (OOB) information methodology. The specific models are separate models based on different SYNOPCodes whilst the generic model is deduced by incorporating the SYNOPCode as the categorical variables. All the models demonstrate that the features selected under various weather conditions and channel attenuation are different and of different importance levels. The comparison of generic and specific models is based on Root Mean Square Error (RMSE) and R^2. Regarding RF channel attenuation, the RMSE and R^2 improvement of the specific model over the generic model under different weather conditions are: for clear weather (RMSE: 0.094; R^2: -1.151%), for dust storm (RMSE: -0.011; R^2: 0.803%), for fog (RMSE: -0.121; R^2: 9.883%), for drizzle (RMSE: -0.283; R^2: 0.955%), for rain (RMSE: -0.068; R^2: 2.810%), for snow (RMSE: -0.189; R^2: -2.162%), and for showers (RMSE: 0.085; R^2: -7.517%). Regarding FSO channel attenuation, the RMSE and R^2 improvement of the specific model over the generic model under different weather conditions are: for clear weather (RMSE: -0.106; R^2: 1.708%), for dust storm (RMSE: -0.773; R^2: 2.819%), for fog (RMSE: -0.318; R^2: -1.173%), for drizzle (RMSE: 0.144; R^2: 18.731%), for rain (RMSE: -0.105; R^2: -38.420%), for snow (RMSE: 0.877; R^2: 5.069%), and for showers (RMSE: 0.009; R^2: 1.128%). Although there are several exceptions to model performance under foggy and snowy weather, the overall performance of generic models exceeds that of specific models. Therefore, the generic model is the final model selected in this report with the principle of the Markov chain and RMSE and R^2 of the final model will be discussed. Furthermore, two hybrid models will be built to analyse if the two channels can compensate each other on attenuation, and the metrics are Pearson Correlation Coefficient(PCC) and mutual information(MI). There is a strong correlation under dust storms with $\frac{I_{EO}}{H_{EO}} = 0.8$. Regarding RMSE and R^2, the generic model outperforms the hybrid for the FSO channel and is generally better for the RF channel, the detailed comparison is in Table 5.4 and 5.5.

1 Introduction

Hybrid Radio Frequency/Free Space Optical (RF/FSO) is a communication system with two parallel links: FSO and RF links as a backup to cooperate [1]. Specifically, FSO is suitable for long-distance transmission with an efficient data rate [3] It is applied to various applications such as 5th generation (5G) and satellites with high reliability and capacity [3]. However, weather conditions, such as temperature and humidity, could cause channel attenuation. Although other factors such as elevation angles are significant, this project primarily focuses on atmospheric conditions.

Three models used in this report are specific, generic, and hybrid models. Feature selection is generated using Out-of-bag (OOB) information. After comparing the Root Mean Square Error (RMSE) and R^2 between the first two models, generic models have better performance for both FSO and RF models.

Pearson Correlation Coefficient(PCC) and mutual information(MI) are two metrics to measure the linear and non-linear correlation between two models, and they are used in hybrid models. The output demonstrates that foggy weather witnessed the most significant compensation between FSO and RF channels.

2 Work Since Last Submission

All the further work during this trimester are:

- Generated an FSO channel attenuation hybrid model using all features and predicted RF channel attenuation
- Generated an RF channel attenuation hybrid model using all features and predicted FSO channel attenuation
- Feature importance plots and feature selection for two hybrid models
- Calculated PCC and generated a table to interpret the correlation
- Calculated MI and $\frac{I_{EO}}{H_{EO}}$ and plotted it to find the compensation between two channels
- A heat map to visualise the accuracy of models
- A comparison between hybrid and generic models using RMSE and R^2

3 Background

The project aims to deduce the models for Hybrid Optical/Radio Frequency (RF/FSO) communication channels: one is how weather conditions influence FSO communication channel attenuation, whereas the other is on RF channel attenuation. Various weather conditions will have significant impacts on the attenuation of both communication channels. After an overall understanding of weather predictors, a random forest is implemented for further model selection. Both coarse tuning and fine-tuning are applied to find the best hyperparameters, which is followed by a more accurate decision tree with a maximum depth of the tree, minimum number of samples of being a leaf node and to split the internal nodes.

Here is a brief description of three models built in this report:

- Specific model: contains six models with separate SYNOPCodes
- Generic model: a single model taking SYNOPCode as a categorical predictor
- Hybrid model: is built using all features and the predicted attenuation of the other channel

Random forest is the major method used in this report, which can reduce the variance after the reduction of trees correlation [6]. There are several advantages of using random forest for this report: it is suitable for both continuous and categorical variables, builds both linear and non-linear relationships, and is computationally inexpensive [5]. It is based on the decision trees and deduces the variable importance rank using the Out-of-bag (OOB) methodology [8]. Moreover, OOB is implemented by iteratively removing the features during the training process and selecting the significant features under certain metrics criteria [4]. Specifically, given the set of training data containing all the predictors, the Root Mean Square Error (RMSE) and R^2 are calculated for the random forest. The feature with the least importance is removed based on the OOB information, and the RMSE and R^2 are calculated again with one feature removed. All the following step is repeated until all the predictors are removed, and the last predictor removed is the most important feature.

When building the hybrid model with random forest, the metrics used to determine whether the correlation between channel attenuation is linear or non-linear are PCC and MI. Specifically, PCC can measure linear correlation whereas MI can measure both linear and non-linear correlation.

Therefore, the method calculates $\frac{I_{EO}}{H_{EO}}$ to find which metric has better observations on correlation. The formula of PCC and MI will be explained in Section 4.3.2.

Formula

In this report, for N observations in the dataset, let y_n be the attenuation at time n, $\bar{y_n}$ be the mean attenuation for all n, and $\hat{y_n}$ be the predicted attenuation using the model containing several predictors. Two metrics are used as standards of testing model performance, which are Root Mean Square Error (RMSE) and R squared (R^2), which are

$$RMSE = \sqrt{\frac{1}{N}\sum_{n=1}^{N}(y_n - \hat{y_n})^2}$$

$$R^2 = 1 - \frac{\sum_{n=1}^{N}(y_n - \hat{y_n})^2}{\sum_{n=1}^{N}(y_n - \bar{y_n})^2}$$

4 Methods

4.1 Data Properties

The dataset is synthesized based on real data from six different cities. There are 27 variables: attenuation of FSO (GHz) and RF (GHz), humidity, the distance between transmitter and receiver, signal frequency, particulate in the air, rainfall intensity, relative humidity, SYNOP code for weather conditions, temperature, time (0-23), visibility, wind direction, and speed. There are 91,377 observations in this dataset and no missing value, and the first row is removed since the first value of TemperatureDifference is the default value which will influence the data analysis. Specifically, the Frequency is either 83,500,000,000 or 73,500,000,000 for RF attenuation only. SYNOPCode is the Surface Synoptic Observations Code that represents the weather condition for transmitting surface [2]. Table 4.1 demonstrates the SYNOPCode and the corresponding weather condition.

Weather Condition	Clear	Dust Storm	Fog	Drizzle	Rain	Snow	Showers
SYNOPCode	0	3	4	5	6	7	8

Table 4.1: SYNOPCode for Different Weather Conditions

The distribution of FSO and RF attenuation is depicted in Figure 4.1. Specifically, the distribution of FSO attenuation is bimodal and right-skewed whilst the distribution of RF attenuation is unimodal and right-skewed.

Three variables (Frequency, SYNOPCode, and Time) are converted into categorical variables for better visualization. The distribution of Frequency is even shown in Figure 4.2 since they represent the frequency of two directions. The histogram of the distribution of SYNOPCode in Figure 4.3 demonstrates that the data of clear weather accounts for the lion's share with 56,963 observations precisely, which is around 60% of the whole dataset. Similarly, there are 25,018 observations under the rain condition, and the observations with the rest of the other 5 weather conditions only occupy about 10% in the dataset. The distribution of time is roughly evenly distributed as demonstrated in Figure 4.4.

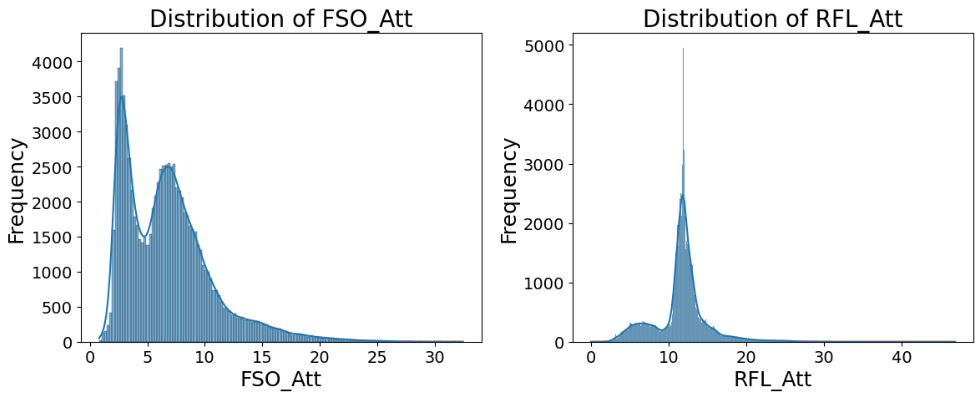

Figure 4.1: Distribution of FSO & RF Attenuation

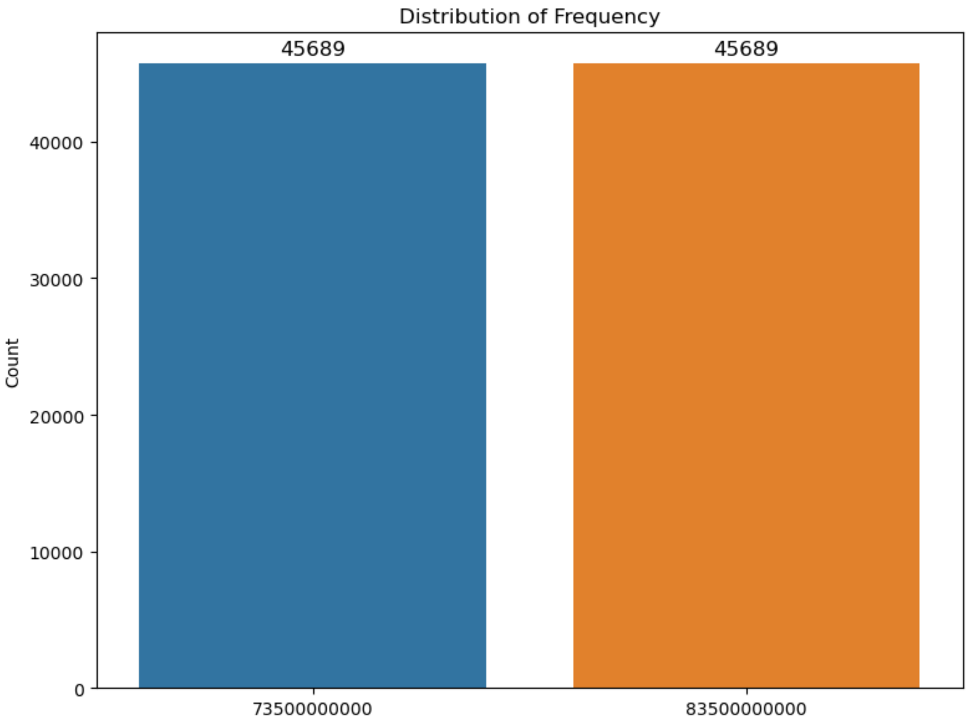

Figure 4.2: Distribution of Frequency

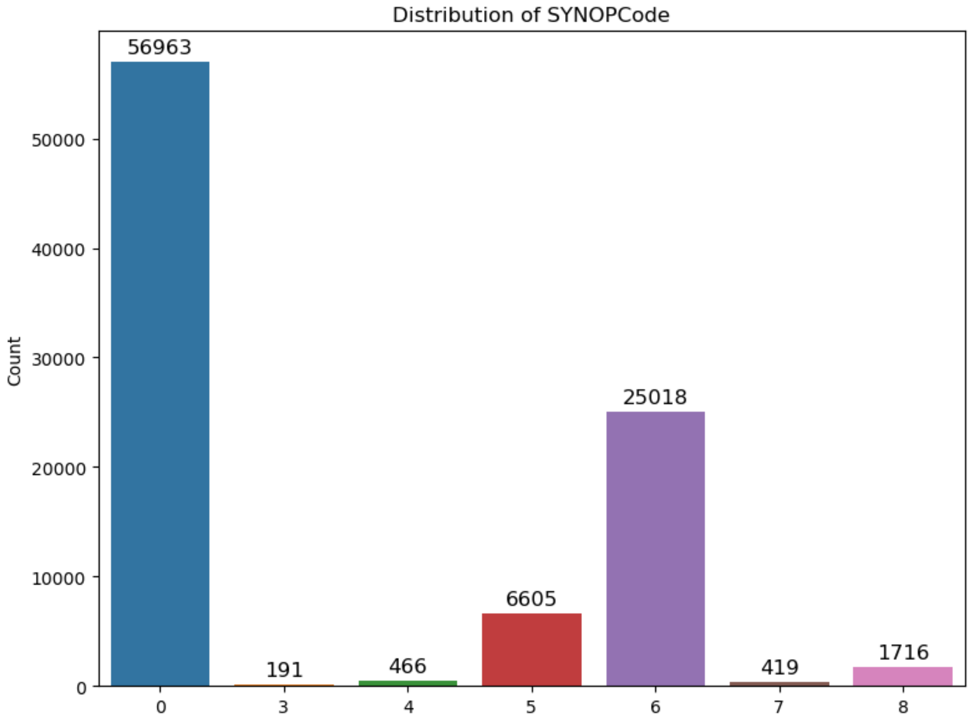

Figure 4.3: Distribution of SYNOPCode

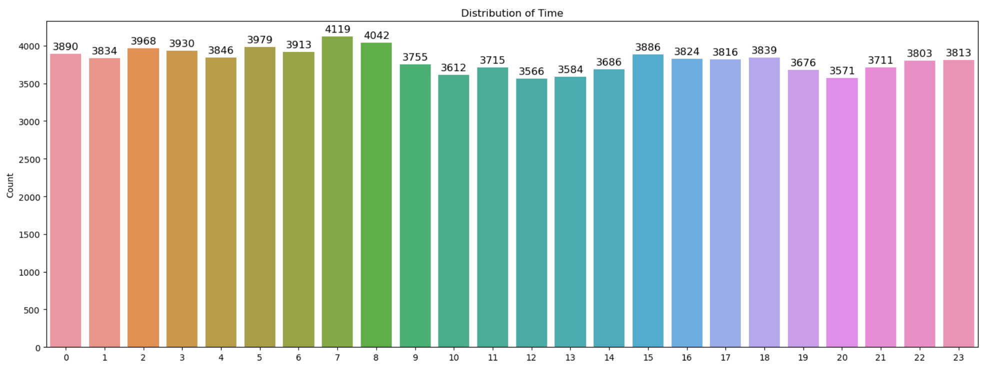

Figure 4.4: Distribution of Time

Table 4.2 demonstrates all the significant values of variables except attenuation and categorical variables. The table is more suitable for visualization than Boxplot since the data is scattered so a huge amount of data is displayed as outliers.

	mean	**std**	**min**	**25%**	**50%**	**75%**	**max**
AbsoluteHumidity	9.55	5.86	1.14	4.96	6.87	14.05	24.79
AbsoluteHumidityMax	10.03	6.16	1.24	5.21	7.21	14.78	26.41
AbsoluteHumidityMin	9.08	5.58	1.05	4.71	6.52	13.38	24.27
Distance	3297.94	1224.31	2012.00	2019.43	2959.86	4820.89	4828.00
Particulate	27.07	72.13	0.00	0.00	0.00	16.95	1621.00
ParticulateMax	28.42	75.76	0.00	0.00	0.00	17.78	1753.75
ParticulateMin	25.72	68.60	0.00	0.00	0.00	16.04	1500.67
RainIntensity	0.25	1.64	0.00	0.00	0.00	0.00	87.26
RainIntensityMax	0.26	1.71	0.00	0.00	0.00	0.00	92.02
RainIntensityMin	0.23	1.55	0.00	0.00	0.00	0.00	82.27
RelativeHumidity	76.60	17.74	8.94	67.62	80.94	90.13	100.00
Temperature	12.54	9.13	-6.89	4.68	10.28	21.47	37.26
TemperatureDifference	-0.07	0.80	-11.16	-0.40	-0.10	0.25	10.92
TemperatureMax	13.17	9.59	-6.69	4.91	10.80	22.47	40.77
TemperatureMin	11.90	8.69	-7.27	4.44	9.77	20.32	36.72
Visibility	32986.06	24713.77	10.15	11158.42	26376.26	53492.49	75005.00
VisibilityMax	34636.27	25978.69	11.03	11721.41	27686.04	56090.03	82503.13
VisibilityMin	31331.24	23497.73	9.48	10588.64	25047.01	50797.44	74999.34
WindDirection	89.04	26.42	0.00	82.98	90.87	98.48	360.00
WindSpeed	0.74	0.87	0.00	0.08	0.44	1.09	7.80
WindSpeedMax	1.85	1.76	0.00	0.51	1.32	2.68	16.42
WindSpeedMin	0.70	0.83	0.00	0.07	0.42	1.04	7.25

Table 4.2: Summary statistics of various parameters

The histogram for several variables is not clear since most observations are within a narrow range with several outliers, so the density plot is used for visualization. The density plot of particulate and rain intensity is plotted in Figure 4.5. All the distributions of six variables are unimodal and right-skewed and the overall distributions of rain intensity are narrower than those of particulate.

The overall distributions of absolute humidity are right-skewed and the peaks are around 5 in 4.6

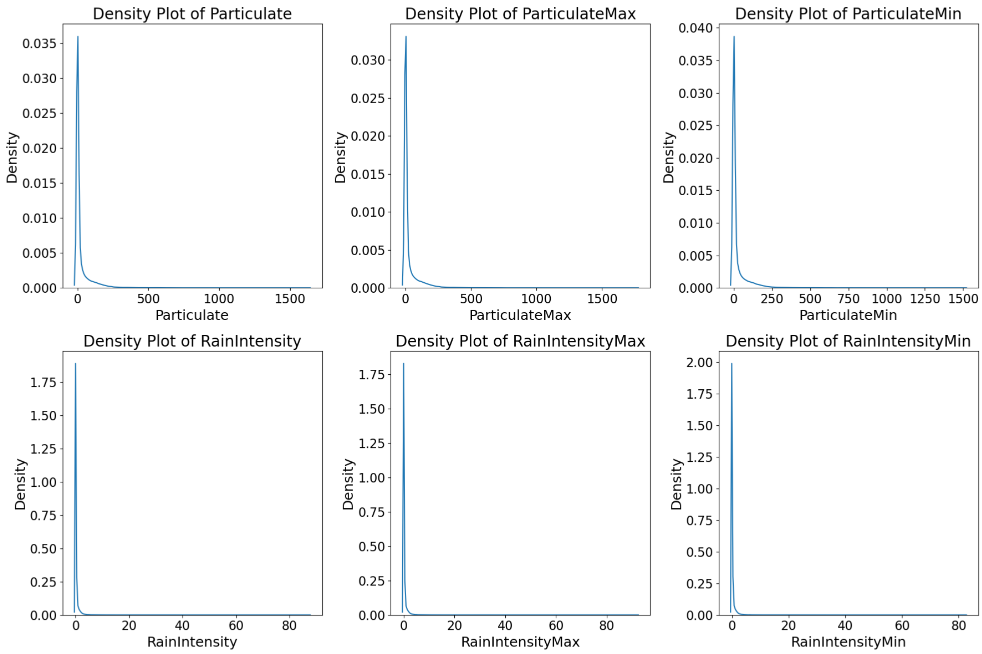

Figure 4.5: Density Plot of Particulate & Rain Intensity

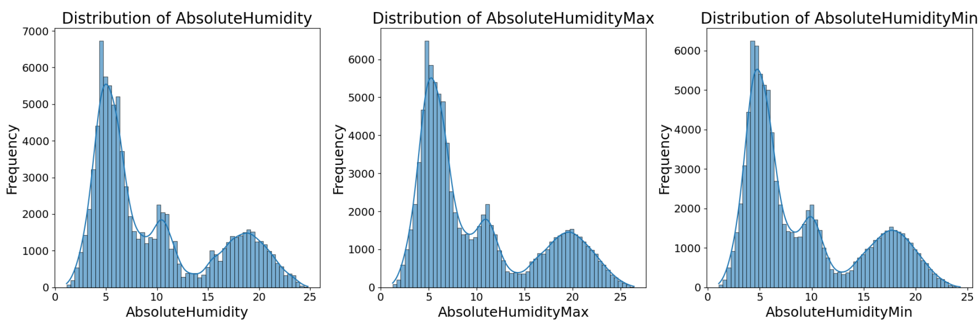

Figure 4.6: Distribution of Absolute Humidity

From Figure 4.7, the peaks of distance are approximately 2,000 and 4,800. The distribution of relative humidity is unimodal and left-skewed peaking between 80 and 100. Moreover, the overall distributions of visibility are similar, which are bimodal and peaking at the range of minimum and maximum values. Furthermore, the distributions of wind speed are unimodal and right-skewed.

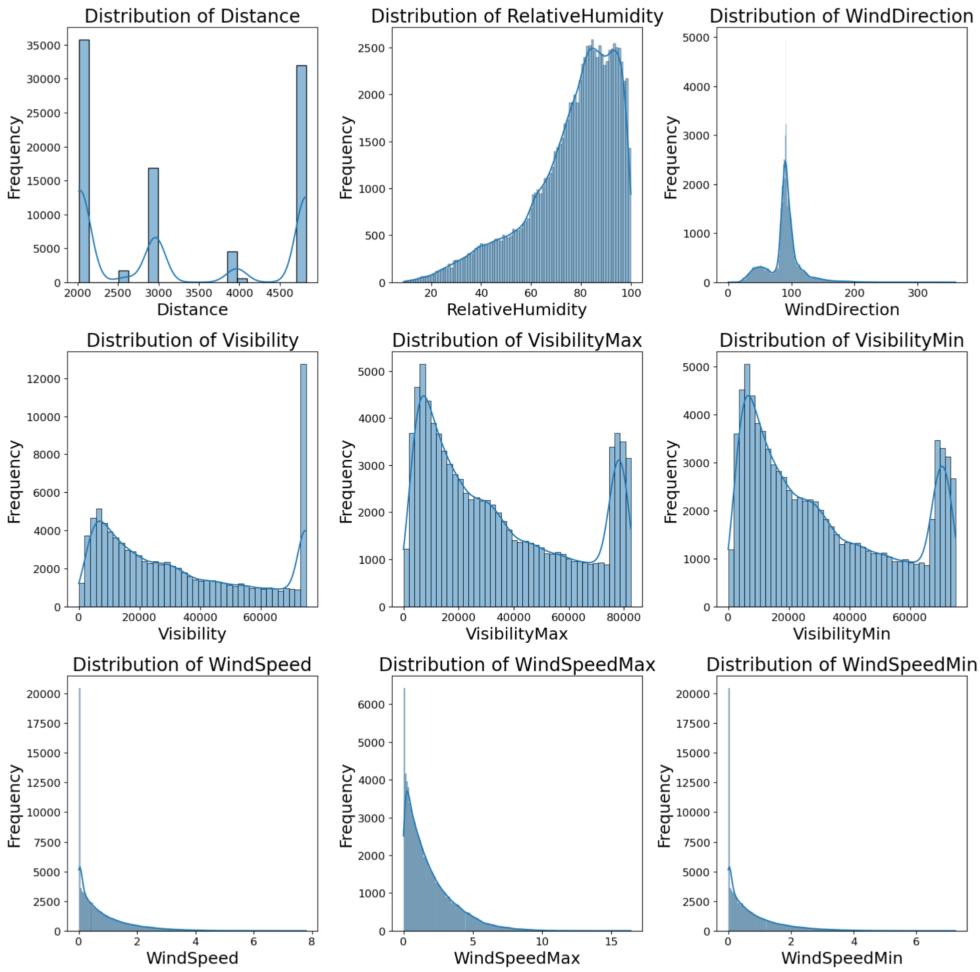

Figure 4.7: Distributions of Distance, Relative Humidity, Wind Direction, Visibility and Wind Speed

The distributions of temperature, maximum temperature, and minimum temperature are bimodal whilst that of temperature difference is unimodal, and it is shown in Figure 4.8

According to the correlation heatmap from Figure 4.9 for all variables in the dataset, other than several variables are highly correlated since they

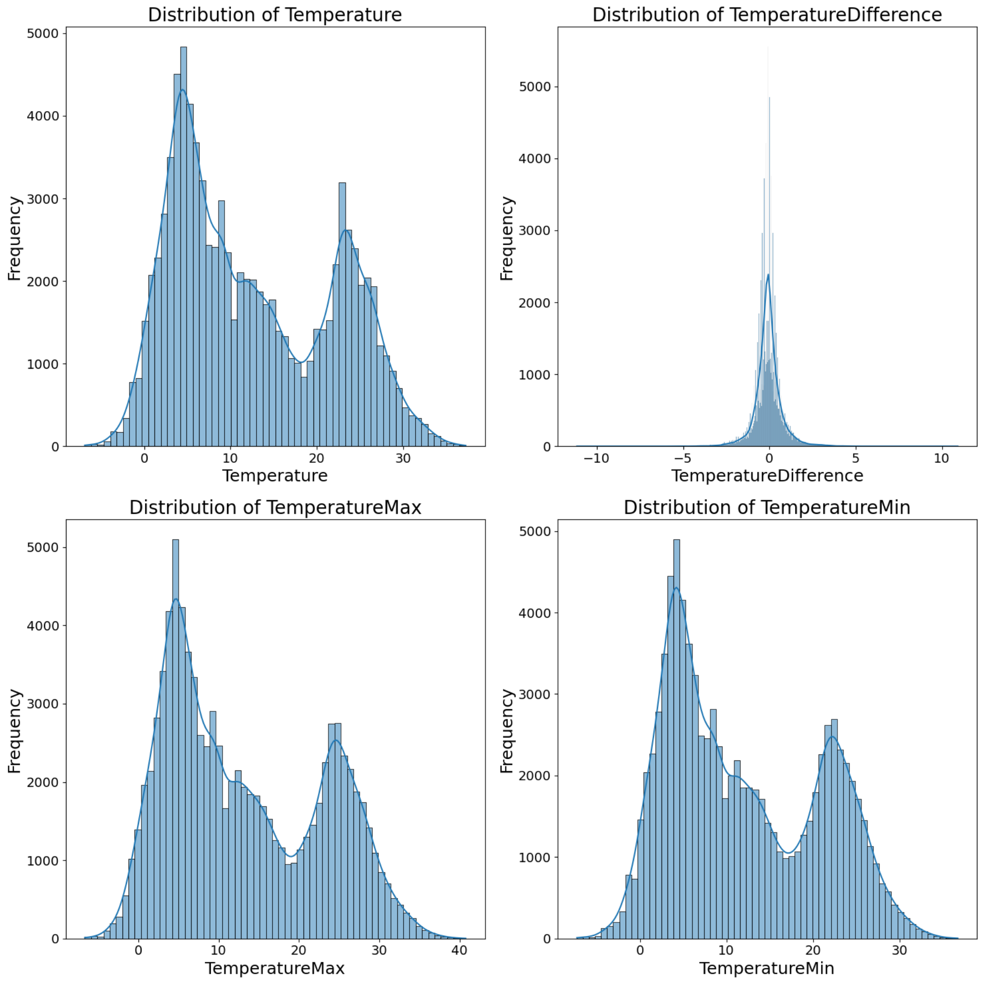

Figure 4.8: Distribution of Temperature

are within the same category such as maximum and minimum temperature, some variables are more related to attenuation. Specifically, attenuation of FSO is related to particulate and relative humidity. On the other hand, other than particulate and relative humidity, RF attenuation is also influenced by absolute humidity, rain intensity, and temperature, It is, however, a brief overview of the correlation between attenuation and other variables, a clear model selection will be deduced in the following section with corresponding principles.

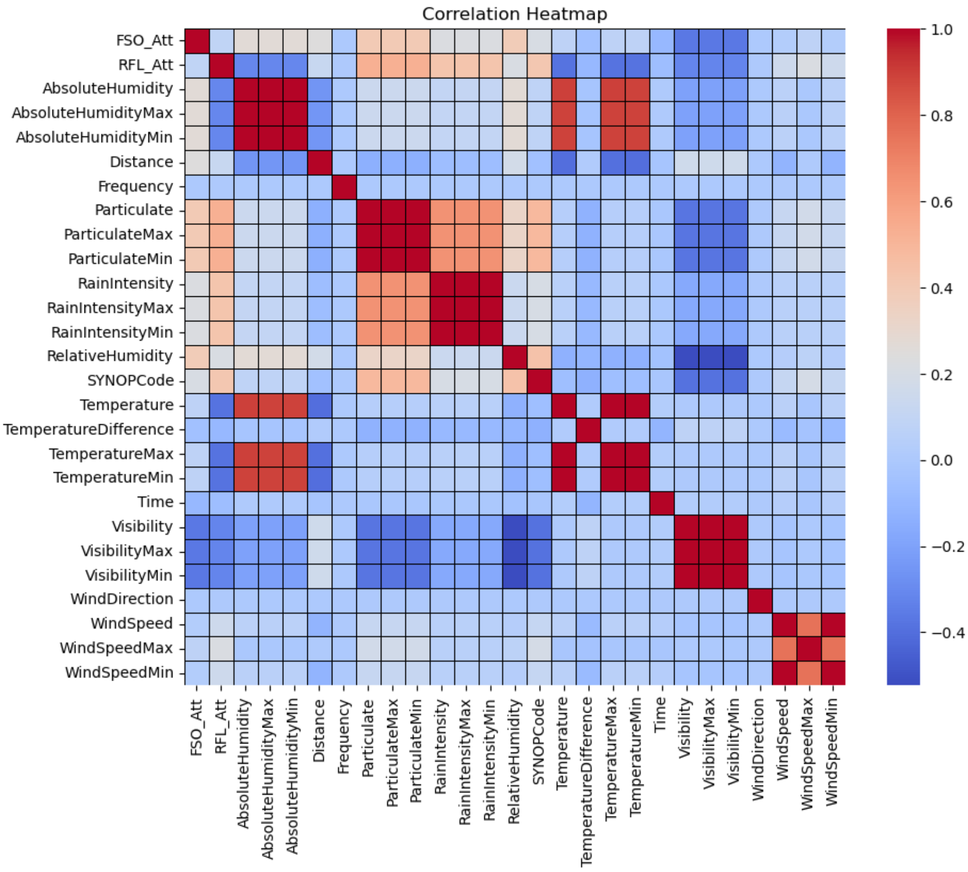

Figure 4.9: Heatmap of All variables

4.2 Model Selection

4.2.1 Decision Tree

Among the 91,378 observations in the dataset, 70% of observations will be split into training data, which is 63,964 precisely. The training data is used to build the random forest. Otherwise, the remaining 30% of the data is divided equally into validation and testing data, which is used for tuning hyperparameters and evaluating the final performance of the model. Decision tree is implemented as the first step. After setting the regression model and cross-validation, the model is tuned with the metric of MSE and R^2 respectively. Regarding the FSO model, the best scores of MSE and R^2 are approximately 1.528 and 0.899 respectively. The maximum depth of the tree is 24, the minimum number of samples of being a leaf node is 10, and the minimum number of samples to split the internal node is 2. On the other hand, the best scores of the RF model with the metric of MSE and R^2 are around 0.617 and 0.948 respectively. The maximum depth of the tree, the minimum number of samples of being a leaf node, and the minimum number of samples to split the internal node are 36, 1, and 11. As the depth grows, the MSE and R^2 will not climb dramatically, so the maximum depth of the tree will be chosen when the tendency is stable.

4.2.2 Tuning

Furthermore, coarse tuning is implemented to find the area where the hyperparameters lie. For the FSO attenuation model, the maximum depth of the tree is 30, the minimum number of samples of being a leaf node is 1, and the minimum number of samples to split the internal node is 2 as shown in Table 4.3. In terms of the RF attenuation model, the results are slightly different. The maximum depth of the tree, the minimum number of samples of being a leaf node, and the minimum number of samples to split the internal node are 30, 1, and 2 respectively from Table 4.3.

Moreover, fine tuning is followed by coarse tuning, which is an optimization of hyperparameters. This optimization is implemented by narrowing the range of finding the hyperparameters. Specifically, regarding the FSO attenuation model, the maximum depth of the tree is narrowed to 29 whilst the minimum number of samples of being a leaf node and splitting the internal node remain the same from Table 4.3. For the RF attenuation model, the maximum depth of the tree is reduced by 3, which is 27, and the minimum number of samples of being a leaf node and splitting the internal node remains constant as demonstrated in Table 4.3

	FSO Coarse	FSO Fine	RF Coarse	FSO Fine
max_depth	30	29	30	27
min_samples_leaf	1	1	1	1
min_samples_split	2	2	2	2

Table 4.3: Table of Coarse & Fine Tuning for FSO & RF

The following table 4.4 represents the RMSE and R^2 for training and validation data. Out-of-bag (OOB) information is used for ranking the importance of predictors in this report, which is repeating the step of finding the most relevant predictor based on RMSE and R^2 and removing it from the dataset until the last predictor is removed. Therefore, the OOB R^2 is also included in the Table 4.4. It can be found from Table 4.4 that the RMSEs of both training and validation data for the RF model are smaller than those for the FSO model whereas the R^2 of training, validation data, and OOB for the RF model are bigger than those for the FSO model. Therefore, the overall performance of the RF model is better than the FSO model based on RMSE and R^2, but the statistics of the FSO are within the feasible intervals.

	Train RMSE	Validation RMSE	Train R^2	Validation R^2	OOB R^2
FSO	0.314	0.814	0.994	0.957	0.954
RF	0.202	0.524	0.997	0.977	0.975

Table 4.4: Table of RMSE & R^2 for Training, Validation and OOB

4.2.3 Variable Importance Plots

Furthermore, the RF and FSo attenuation variable importance plots are displayed in Figure 4.10 and 4.11. This method measures the reduction of all trees in the forest using impurity criteria such as the Gini index, demonstrating a better visualization of the correlation between weather predictors and attenuation. Distance is a significant predictor for both FSO and RF attenuation, which is the most and second most important variable for FSO and RF attenuation respectively. Specifically, visibility and temperature play an essential role in FSO attenuation whereas the maximum rain intensity only has the least impact on it. On the other hand, rain intensity accounts for the lion's share of RF attenuation importance, and absolute humidity has a similar vital importance with distance for RF attenuation whilst the wind direction only plays a little influence on it. It is, however, not the method for model selection in this report. The main method is still OOB due to its advantages, which are high efficiency and less overfitting.

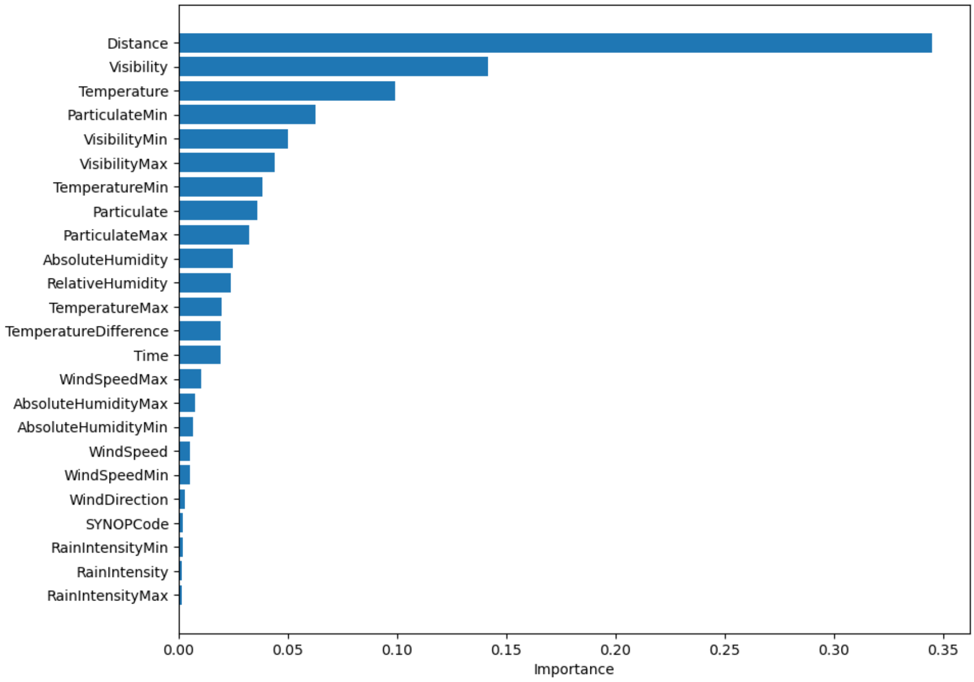

Figure 4.10: Variable Importance of FSO Attenuation

4.3 Hybrid Model

4.3.1 Two Methods for Hybrid Model

When generating generic and specific models, the principle for the two models is the same: predicting attenuation from all features using random forest and selecting the important features for the corresponding model. There are, however, two new methods that will be applied to hybrid models, as shown in Figure 4.12. Regarding the first method, after generating the predicted attenuation of FSO from all features using a random forest, the predicted RF attenuation is generated from all features and predicted FSO attenuation using a new random forest. Similarly, FSO attenuation is predicted from all features and predicted RF attenuation using random forest for method 2.

4.3.2 PCC & MI

PCC

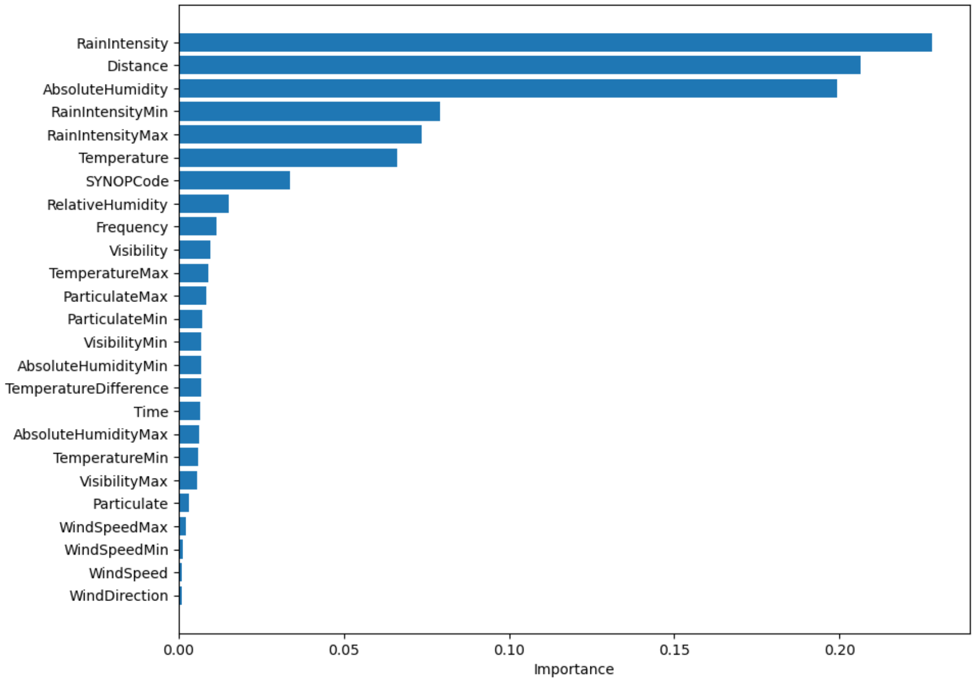

Figure 4.11: Variable Importance of RF Attenuation

PCC is used for measuring the linear correlation between FSO and RF attenuation in this project. Let r be the correlation coefficient, x_i and y_i be the attenuation of RF and FSO respectively, and $\hat{x}$ and $\hat{y}$ be the mean of RF and FSO attenuation, and here is the formula:

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}}$$

where $r \in [-1, 1]$. Specifically, the interpretation of PPC is shown in Table 4.5 [7]. This demonstrates that the $|r|$ is close to 1 when two sets of data have a strong positive/negative correlation whilst $|r|$ is close to 0 representing a negligible correlation.

MI

MI, as a measurement of defining the dependency of variables, is used to find both linear and non-linear correlation between FSO and RF entropy of attenuation in this project, so it can be compared with PCC to estimate whether the correlation between two attenuation is linear or non-linear [9]. The computational speed for MI is fast since the integral operation reduces to summation [9].

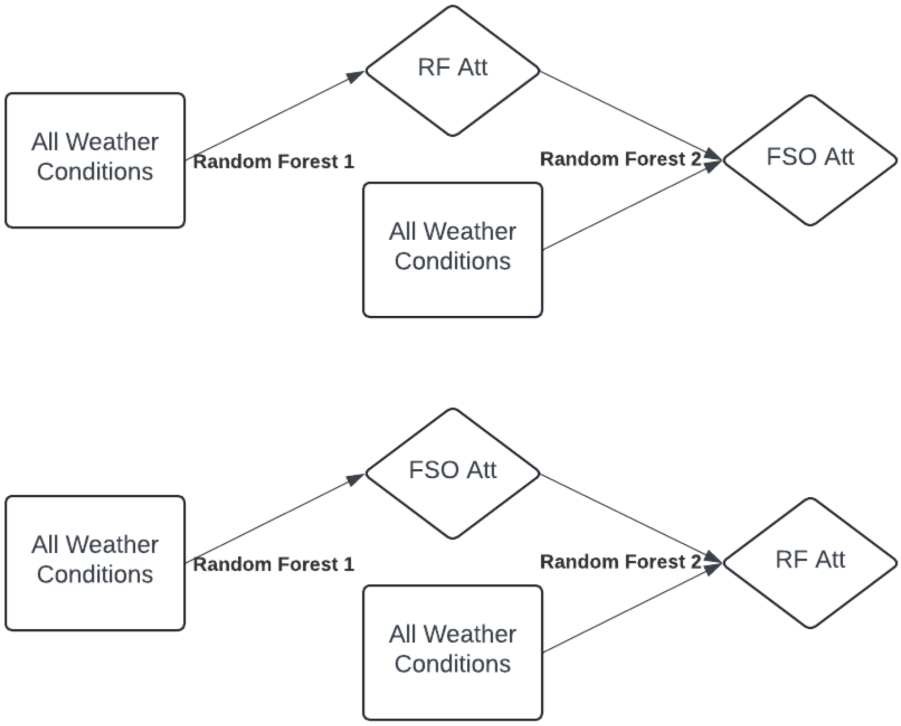

Figure 4.12: Methods of Hybrid Model

Entropy is a measurement of uncertainty, randomness, or amount of information. If the distribution of X is close to a uniform distribution, then entropy is large. If X is nearly a constant, the entropy is small. Specifically, if X is a constant, then the entropy is 0. Let X and Y be the RF and FSO attenuation respectively, p(x) and p(y) be the probability distribution of X and Y, then

$$H(X) = -\sum_{x} p(x) log p(x)$$
$$H(Y) = -\sum_{y} p(y) log p(y)$$
$$H(X,Y) = -\sum_{x,y} p(x,y) log p(x,y)$$
$$I(X;Y) = H(X) + H(Y) - H(X,Y)$$

where H(X,Y) is the entropy of X and Y, and I(X;Y) is the MI between X and Y. Regarding MI, I(X;Y) is non-negative with the property of

Size of Correlation	Interpretation
$\lvert r \rvert \in [0.9, 1]$	Very high correlation
$\lvert r \rvert \in [0.7, 0.9]$	High correlation
$\lvert r \rvert \in [0.5, 0.7]$	Moderate correlation
$\lvert r \rvert \in [0.3, 0.5]$	Low correlation
$\lvert r \rvert \in [0, 0.3]$	Negligible correlation

Table 4.5: Interpretation of PCC

which is smaller than the entropies (i.e. $I(X;Y) \leq H(X)$, $I(X;Y) \leq H(Y)$, $I(X;Y) \leq H(X,Y)$. X and Y are independent if $I(X;Y) = 0$. Conversely, if $I(X;Y) \approx H(X,Y)$, X and Y are highly correlated.

4.3.3 Joint Probability Distribution

When calculating the joint probability distribution of RF and FSO attenuation, $P_{XY}(x,y) = \frac{\text{the count in the cell (x, y)}}{\text{the total number of data}}$ is used, and the heatmap is shown in Figure 4.13. The overall distributions of true and predicted attenuation under different weather conditions are similar.

Specifically, the attenuation distributions under dust, fog, and snow weather are wide, demonstrating the distributions close to a uniform distribution. Notably, there are slight differences between true and predicted attenuation distributions under dust and snow weather since the total number of data for those two weather conditions are 191 and 491 respectively, indicating the relatively low accuracy.

4.3.4 Feature Importance

The extent to which one attenuation affects another can be found by feature importance, as shown in Figure 4.14 and 4.15. The predicted RF attenuation is ranked fifth in the hybrid model for predicting FSO attenuation. In contrast, the predicted FSO attenuation is ranked seventh in the hybrid model for predicting RF attenuation, demonstrating that both attenuation impact each other.

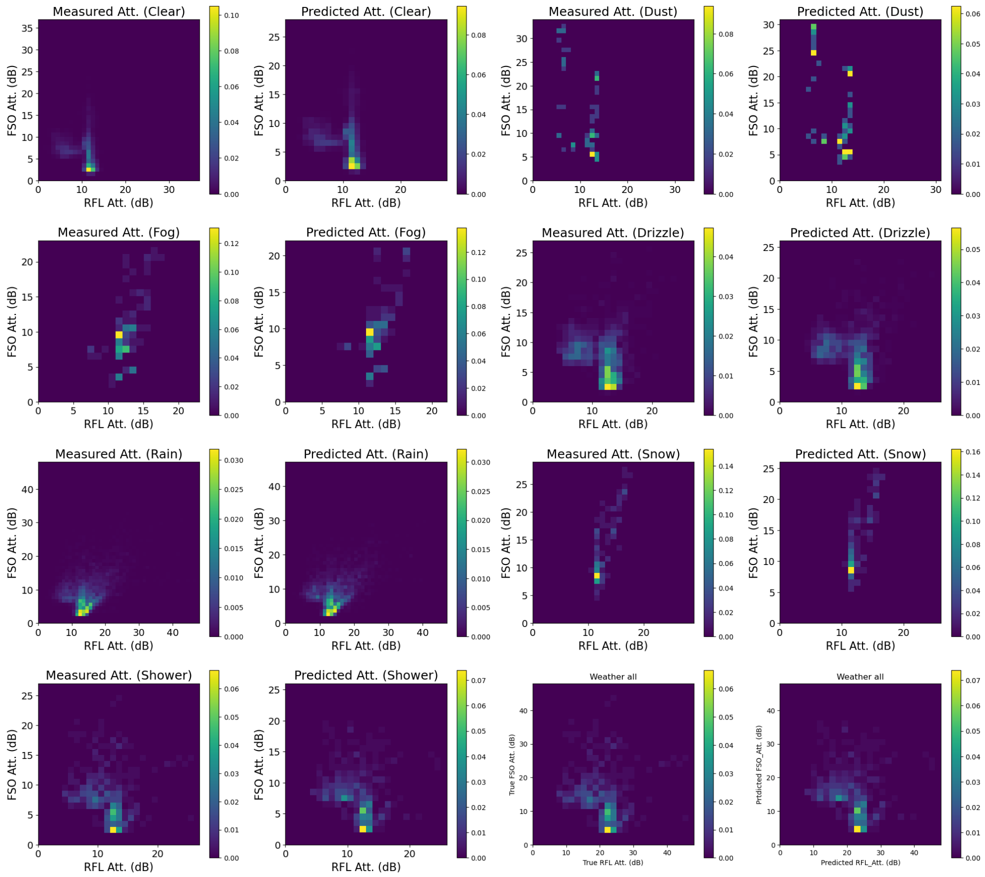

Figure 4.13: Comparison Between the Measured and Predicted Attenuation

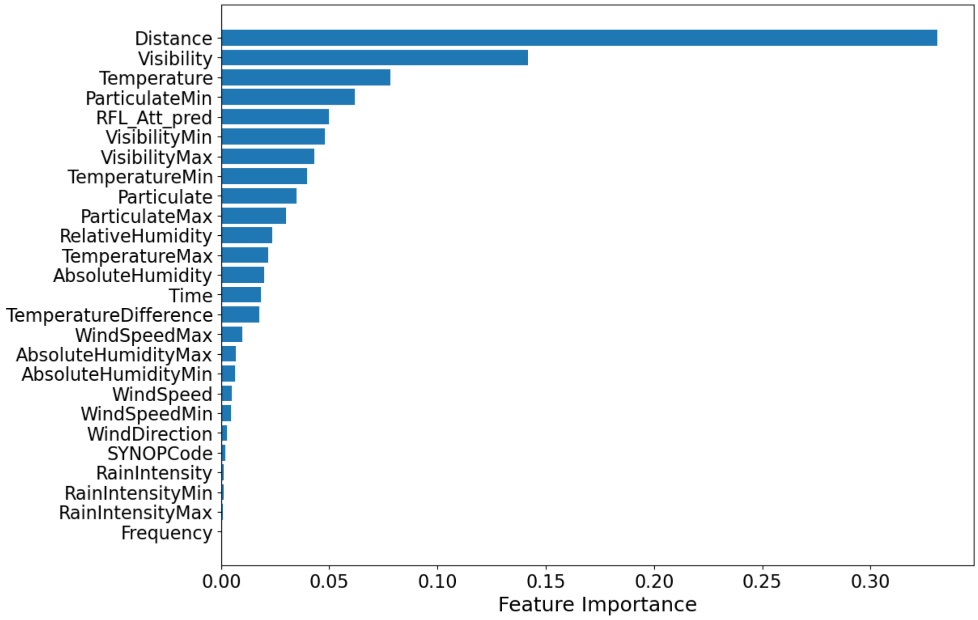

Figure 4.14: Feature Importance of Hybrid Model (FSO attenuation)

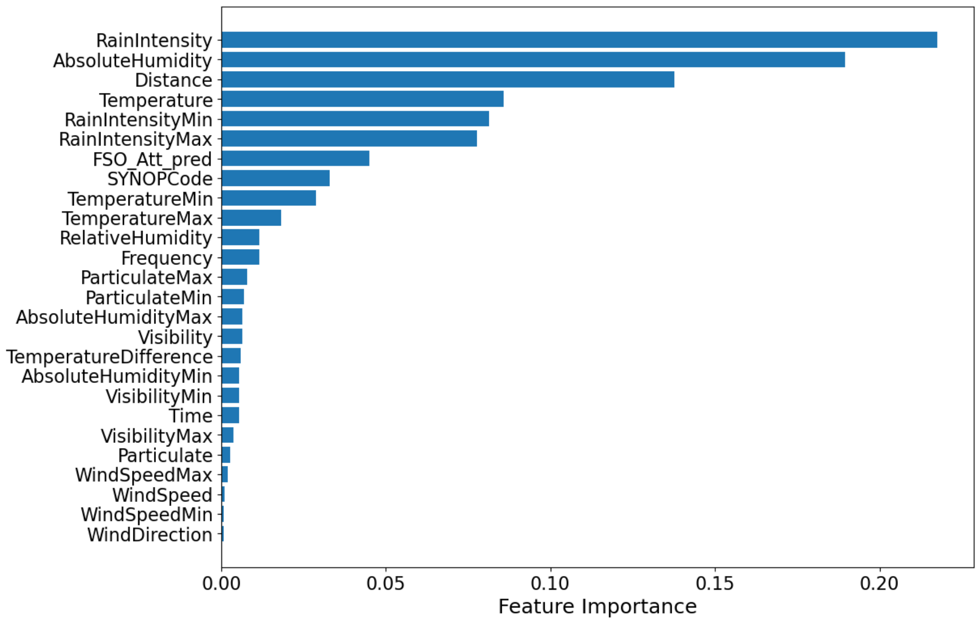

Figure 4.15: Feature Importance of Hybrid Model (RF attenuation)

5 Results

5.1 Specific & Generic Model

As discussed in Section 4.1, there are seven types of SYNOPCode representing different weather conditions, the specific model aims to split the data based on SYNOPCode and create seven models for both FSO and RF attenuation. All the plots based on the OOB method are demonstrated below.

Regarding the clear weather, RMSE and R^2 are stable until the Relative Humidity is removed from the RF model, and the rest of the 10 predictors are included to evaluate RF attenuation from Figure 5.1. For the FSO model from Figure 5.1, although there is a significant change for both RMSE and R^2 after deleting the minimum of absolute humidity, they remain constant for the following steps until the temperature difference is removed, so the rest of the 10 variables are the predictor when calculating the FSO attenuation under the clear weather. Notably, distance is the most significant predictor under clear weather for both RF and FSO weather.

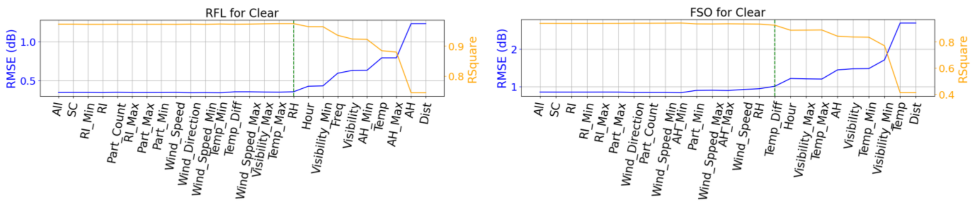

Figure 5.1: Metrics For Clear Weather

The visibility will be influenced by the dust storm, and it is confirmed in Figure 5.2. There is a huge decline after removing distance from the RF model for dust storms, and 12 predictors including visibility are the essential variables for the RF model. There is, however, a fluctuation after removing the minimum absolute humidity since there are only 191 observations with the SYNOPCode being 3, which will influence the accuracy of prediction. On the other hand, After several fluctuations, the accuracy of the FSO model under dust storms drops dramatically after omitting visibility, so the remaining 6 predictors are left for the FSO model.

For foggy weather, visibility is the first predictor kept for the OOB method, but there is an increase in accuracy after removing absolute humidity from Figure 5.3. This is because if the predictors are only selected from the maximum rain intensity, the number of predictors is small to

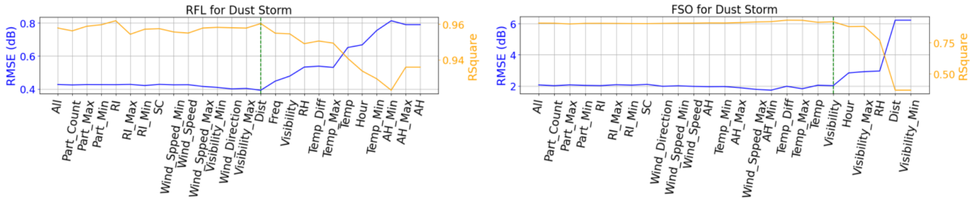

Figure 5.2: Metrics For Dust Storm

support the RF model and there are decrease in R^2 after removing the minimum temperature and maximum rain intensity. The number of predictors for predicting RF attenuation after adjustment is 9. Regarding the FSO model, the overall tendency is clearer in Figure 5.3. 9 predictors including maximum temperature are used for FSO attenuation model prediction.

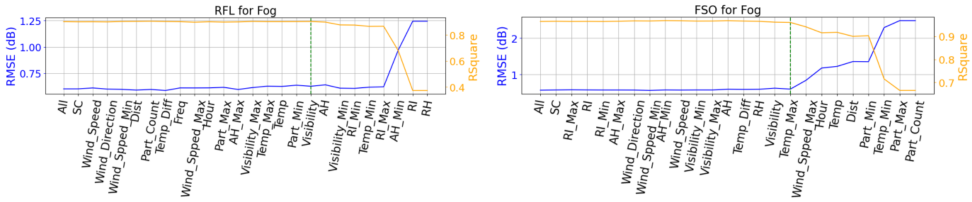

Figure 5.3: Metrics For Foggy Weather

There are 6,605 observations for drizzle weather, which ensures the relatively high accuracy of implementing the OOB method. There are 11 predictors including maximum particulate for the RF attenuation model for drizzle weather in Figure 5.4. On the other hand, the number of predictors for the FSO model is smaller (5 predictors). FSO and RF models keep the maximum particulate as the first variable during the removing iteration in Figure 5.4.

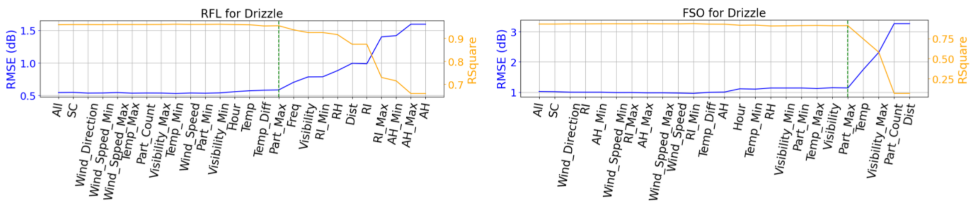

Figure 5.4: Metrics For Drizzle Weather

In terms of rainy weather, 8 variables including the minimum particulate are taken as the RF model prediction shown in Figure 5.5. On the other hand, there are more predictors used for the prediction of the FSO model under the raining conditions, which are 12 predictors including the maximum temperature since there is a constant climb for RMSE when removing the maximum temperature as shown in Figure 5.5, and all the rest predictors will be included.

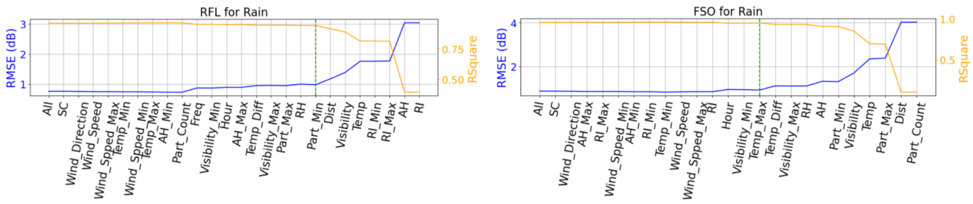

Figure 5.5: Metrics For Rainy Weather

Regarding the snow weather, the maximum wind speed and the temperature difference are considered significant features for both the FSO and RF attenuation models from Figure 5.6. Specifically, 6 predictors including the maximum rain intensity are included as the final model for RF model prediction whilst 8 predictors including temperature are chosen for the FSO model given in Figure 5.6. As shown in Figure 5.6RMSE of the FSO model fluctuates after removing the maximum particulate whereas the R^2 is constantly declined. This is might since only 419 observations of snow weather and the accuracy is not high enough.

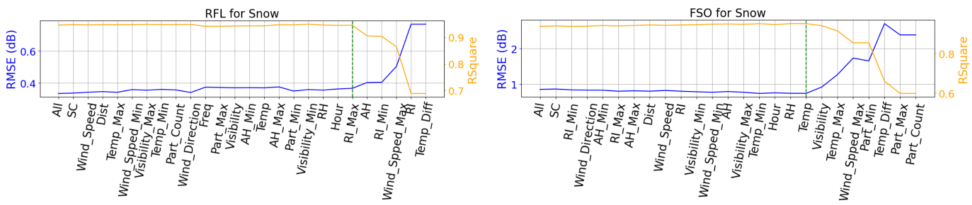

Figure 5.6: Metrics For Snow Weather

The last weather condition is showers. Absolute humidity and minimum visibility are the common predictors for RF and FSO model predictions from Figure 5.7. There is a significant change in accuracy after removing the maximum particulate for the RF model so the rest 6 predictors are included for RF model prediction shown in Figure 5.7. On the other hand, there are 10 predictors are chosen as the predictors for the FSO model under the showers condition as demonstrated in Figure 5.7.

Other than the Specific model, a generic model is also generated for comparison. The generic model simply converts the SYNOPCode as the categorical variable and only one model will be deduced for each

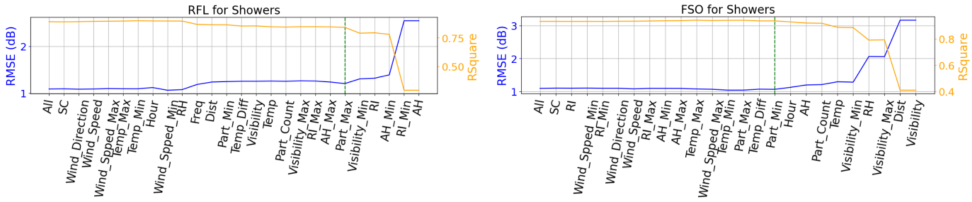

Figure 5.7: Metrics For Showers Weather

communication channel attenuation. As Figure 5.8 shows, the whole dataset is analyzed under all types of weather conditions. Specifically, there is a notable change for RMSE and R^2 for the RF model after removing minimum visibility, so the rest of the 13 predictors are used to predict the RF attenuation under all weather conditions. On the other hand, there are 14 variables are kept to build the FSO attenuation model including the maximum temperature.

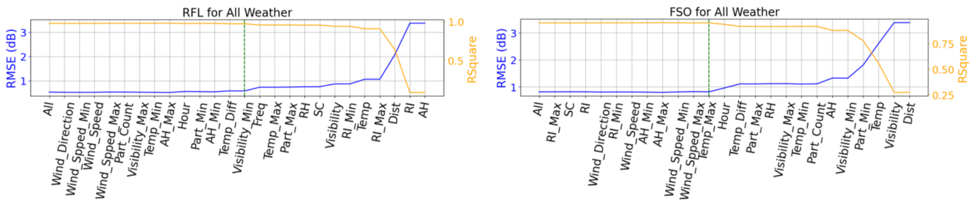

Figure 5.8: Metrics For All Weather

For better visualization of predictor selection, four tables are created in the A.

5.2 Comparison Between Specific & Generic Model

Both specific and generic models are feasible to implement the relationship between FSO & RF attenuation. It is, however, obvious that the specific model is more complex than the generic model, so the model will be evaluated based on RMSE and R^2 to compare the model performance. Bigger R^2 and smaller RMSE demonstrate better model performance, and the metrics are displayed in Figure 5.9.

Regarding the RMSE, FSO from the generic model has an overall better performance than FSO from the specific model except for the foggy and snowy weather. The total observations of foggy and snowy weather are 466 and 419, but there are 9 and 8 predictors are used for the specific model respectively. This might lead to a better performance of specific models due to a low number of observations and a high number of predictors. Similarly, the RF from the specific model for snowy weather also has a smaller RMSE whilst that for clear weather performs better than the generic model.

Furthermore, the R^2 performance of two models regarding FSO and RF has a similar comparison with the RMSE performance. Although there are several exceptions in which the specific model is better than the generic model, the overall performance of the generic is outstanding among the majority of datasets and weather conditions. Therefore, the generic model will be applied for both FSO and RF attenuation.

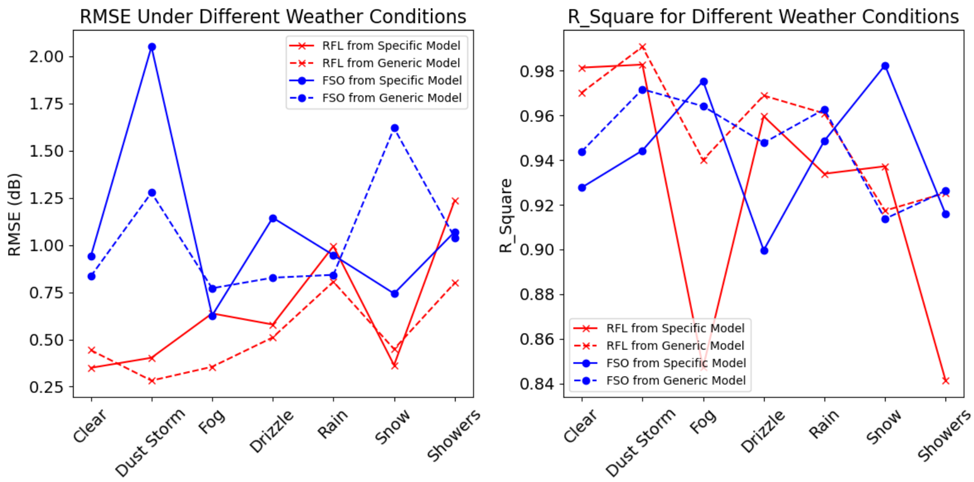

Figure 5.9: Metrics For Generic & Specific Models

Precisely, the changes in RMSE and R^2 can be demonstrated in Table 5.1, Regarding the RF attenuation, specific models are better in RMSE than the generic model only under clear and snowy weather, which increase by 21.1% and 18.9% respectively. The foggy weather witnesses the most significant decline in RMSE of the specific model over the generic model, which decreases by 0.283 (79.6%) precisely. The rest of the three weather conditions (dust storm, drizzle, and rain) have different levels of decrease with 42.9%, 13.4%, and 23.4%. Moreover, the overall changes for R^2 are the same with opposite tendency, with the most obvious increase and decrease under foggy and snowy weather, which are 9.9% and 2.2% respectively.

In terms of the FSO attenuation, specific models in RMSE exceed under foggy and snowy weather with 0.144 and 0.877 whilst the most significant for the generic model is under dust storm with a 60.5% drop. Regarding R^2, it declines by 0.059 under snowy weather whereas drizzle weather witnesses a dramatic model performance with the generic model with a 5.1% growth.

The reason why attenuation under snowy, or foggy weather can be predicted more precisely for specific models is the small sample size with a huge number of predictors. There is an exception of RF attenuation, which specific model performs better in RMSE with reasonable sample size.

	Clear	**Dust Storm**	**Fog**	**Drizzle**	**Rain**	**Snow**	**Showers**
RF Att. RMSE (dB)	0.094 21.134%	-0.121 -42.933%	-0.283 -79.622%	-0.068 -13.387%	-0.189 -23.407%	0.085 18.937%	-0.438 -54.742%
RF Att. R^2	-0.011 -1.151%	0.008 0.803%	0.093 9.883%	0.009 0.955%	0.027 2.810%	-0.020 -2.162%	0.084 9.075%
FSO Att. RMSE (dB)	-0.106 -12.685%	-0.773 -60.484%	0.144 18.731%	-0.318 -38.420%	-0.105 -12.411%	0.877 54.124%	-0.030 -2.933%
FSO Att. R^2	0.016 1.708%	0.027 2.819%	-0.011 -1.173%	0.048 5.069%	0.014 1.469%	-0.069 -7.517%	0.010 1.128%

Table 5.1: Amount of Improvement Provided by the Specific Model over the Generic Model

The Markov Chain is the reason why the generic model has better performance. In this report, assuming X is the SYNOPCode, Y is all other predictors used in the model, and Z is the FSO and RF attenuation. When all other predictors are known, the SYNOPCode and FSO and RF attenuation are independent, which shows that the SYNOPCode-All other predictors-Attenuation is a Markov chain as shown in Figure 5.10.

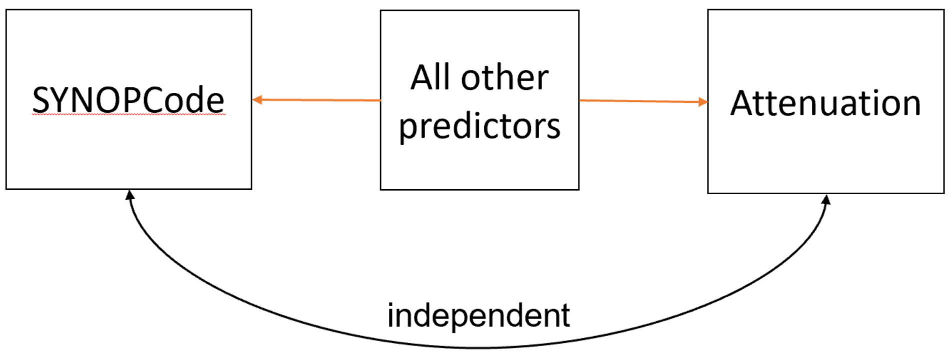

Figure 5.10: Markov Chain

After applying the new model to train data, the RMSE of RF and FSO models drop to 0.579 and 0.312 whereas R^2 for RF and FSO attenuation models increase to 0.970 and 0.994 respectively. These results imply model improvement after selecting the significant features.

5.3 Hybrid Model Comparison

According to the method mentioned in Figure 4.12, the feature selection plot is given in 5.11. Regarding M1, predicting FSO channel attention from all weather features and predicted RF attenuation, has same ranks as the generic model, except for predicted RF attenuation ranking at sixth. On the other hand, M2, predicting RF channel attenuation from all weather features and predicted FSO attenuation, selected the predicted FSO attenuation as seventh important features. Overall, there are 14 and 15 features selected during the feature selection of FSO and RF attenuation hybrid model respectively.

The two hybrid models performances are evaluated based on RMSE and R^2, and the precise output is demonstrated in Table 5.2. The RF hybrid model has smaller RMSE and bigger R^2, with 0.520 and 0.977 respectively, illustrating that the RF hybrid model has better performance than the FSO hybrid model.

	RMSE (Test)	**R^2 (Test)**
FSO	0.855	0.952
RF	0.520	0.977

Table 5.2: RMSE and R^2 for FSO and RF Hybrid Models

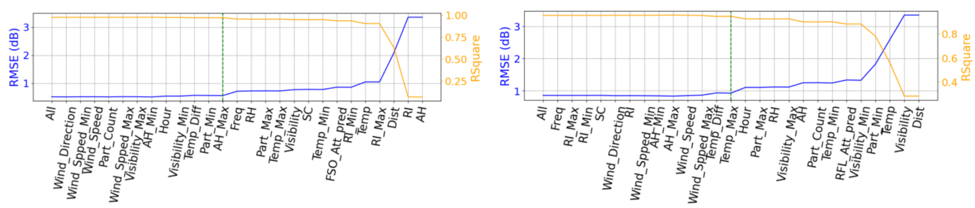

Figure 5.11: Hybrid Model Feature Selection: M1: FSO Model(right) & M2: RF Model (left)

5.4 PCC & MI for Hybrid Model

5.4.1 PCC

PCC, as one metric for comparing the hybrid model, the comparison is given in 5.3. It illustrates that the correlations between FSO and RF attenuation are high in snow, moderate in dust and fog weather, and low in rain. The attenuation for FSO and RF are similar under other weather conditions. Regarding $R_1 - R_2$, all differences are negligible under all weather conditions, which means the two methods are similar. There are, however, several differences between positive and negative correlations. Other than all negligible correlations, the correlation is negative only under dust weather, demonstrating that two different channels can compensate for each other. On the other hand, the correlations under fog, rain, and snow weather are positive, illustrating that when the attenuation of one model increases or decreases, the other channel attenuation will increase and decrease simultaneously.

Weather	r_1	r_2	$r_1 - r_2$	**Interpretation**
All	-0.1788	-0.1760	-0.0028	Negligible
Clear	-0.2966	-0.2958	-0.0008	Negligible
Dust	-0.5436	-0.5273	-0.0163	Moderate
Fog	0.6042	0.6064	-0.0022	Moderate
Drizzle	-0.1928	-0.1916	-0.0012	Negligible
Rain	0.3631	0.3627	0.0004	Low
Snow	0.8032	0.8023	0.0009	High
Shower	-0.2527	-0.2531	0.0004	Negligible

Table 5.3: PCC for Hybrid Models

5.4.2 MI

Regarding MI, it measures both linear and non-linear correlation between attenuation, and the output of both PCC and MI is shown in Figure 5.12. $\frac{I_{EO}}{H_{EO}}$ is higher than the correlation coefficient under dust, fog, and drizzle weather, demonstrating that the correlation between attenuation is closer to non-linear since MI captures more information than PCC. Specifically, the correlation is negligible or low under clear, fog, drizzle, and snow weather. Regarding rain and shower weather, r_{EO} is positive and higher than $\frac{I_{EO}}{H_{EO}}$, demonstrating that the attenuation pairs are more likely to have a positive linear correlation since r_{EO} is more accurate. Conversely, $\frac{I_{EO}}{H_{EO}}$ is the highest under dust storm whilst r_{EO} is negative, illustrating that the attenuation pairs are more likely to have a negative non-linear correlation since $\frac{I_{EO}}{H_{EO}}$ is more accurate. Therefore, attenuation might compensate each other under a dust storm.

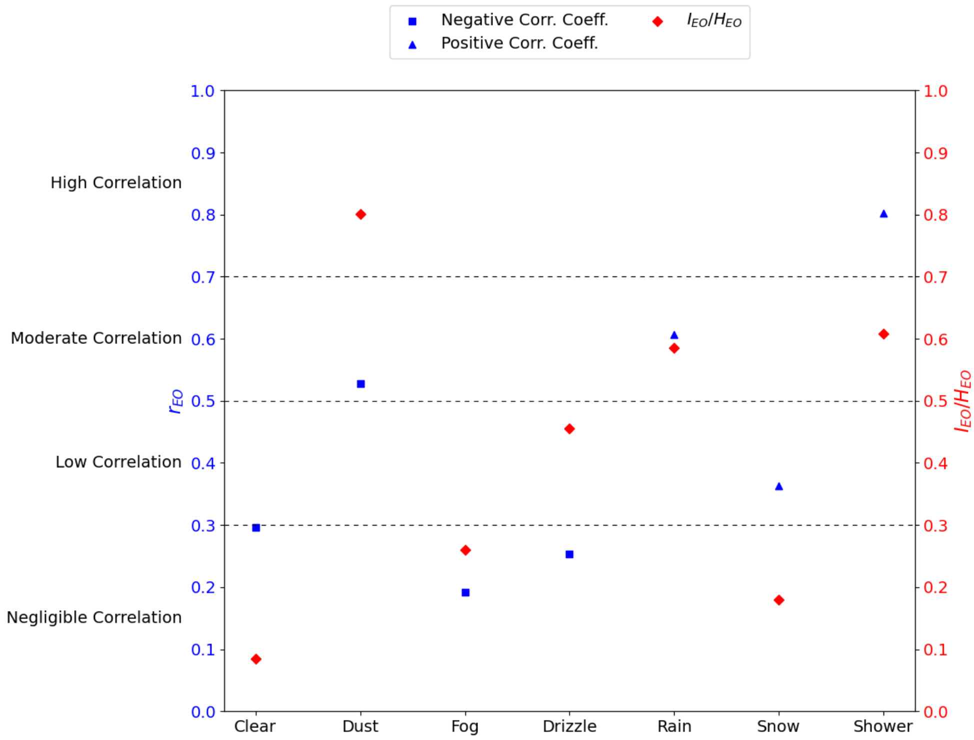

Figure 5.12: Correlation Coefficient and $\frac{I_{EO}}{H_{EO}}$. Interpretation is shown on the left-hand side of the plot.

5.5 Comparison Among Three Models

5.5.1 Hybrid & Generic Models

Based on the conclusion from Section 5.2, the overall performance of the generic model is superior to the specific model, so the hybrid model is compared with the generic model, as shown in Figure 5.13. It demonstrates that the generic model performs better than the hybrid model for FSO attenuation based on both R^2 and RMSE. On the other hand, regarding RF attenuation, the overall performances of generic and hybrid models are similar in terms of R^2 and RMSE. Therefore, there is no significant difference between generic and hybrid models, but the compensation can be displayed using PCC and MI. A clearer comparison among all models is discussed in the following section.

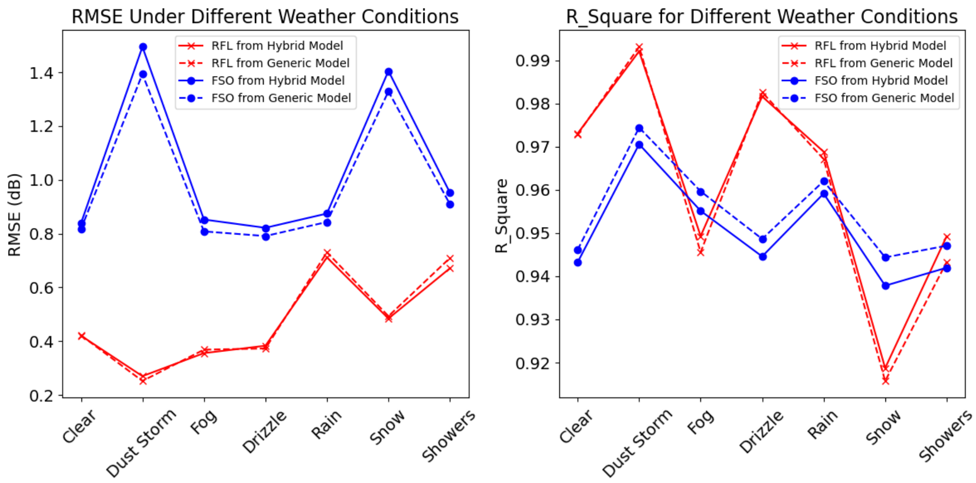

Figure 5.13: Metrics For Hybrid & Generic Models

5.5.2 Hybrid, Generic & Specific Models

The detailed metrics of RMSE are given in Table 5.4. In terms of FSO attenuation, the generic model has the lowest RMSE among the three models. Regarding RF attenuation, the hybrid model performs better under clear, fog, rain, snow, and shower weather conditions.

RMSE	Clear	Dust	Fog	Drizzle	Rain	Snow	Showers
FSO Specific	0.940	2.050	0.627	1.145	0.947	0.743	1.070
FSO Generic	0.816	1.393	0.808	0.791	0.843	1.329	0.909
FSO Hybrid	0.838	1.494	0.852	0.822	0.875	1.405	0.952
RF Specific	0.349	0.403	0.638	0.579	0.994	0.364	1.237
RF Generic	0.422	0.252	0.369	0.373	0.731	0.493	0.710
RF Hybrid	0.421	0.270	0.356	0.383	0.711	0.484	0.671

Table 5.4: RMSE for Different Weather Conditions and Models

Furthermore, the detailed metrics of R^2 are given in Table 5.5. In terms of FSO attenuation, the generic model has the highest R^2 among the three models. Regarding RF attenuation, the hybrid model performs better under fog, rain, snow, and shower weather conditions.

Therefore, the hybrid model for RF channel attenuation performs better under fog, rain, snow, and shower weather conditions for both RMSE and R^2 whilst the generic model is the best for predicting FSO

R^2	**Clear**	**Dust**	**Fog**	**Drizzle**	**Rain**	**Snow**	**Showers**
FSO Specific	0.928	0.944	0.975	0.900	0.949	0.982	0.916
FSO Generic	0.946	0.974	0.960	0.949	0.962	0.944	0.947
FSO Hybrid	0.943	0.971	0.955	0.945	0.959	0.938	0.942
RF Specific	0.981	0.983	0.847	0.960	0.934	0.937	0.841
RF Generic	0.973	0.993	0.946	0.983	0.967	0.916	0.943
RF Hybrid	0.973	0.992	0.949	0.982	0.969	0.919	0.949

Table 5.5: R^2 for Different Weather Conditions and Models

channel attenuation. This output demonstrates that when predicting FSO channel attenuation, it is more accurate using a generic model, and two channels can compensate for each other under fog, rain, snow, and shower weather conditions, but not to a significant extent.

6 Conclusion

This report discussed the weather influences on RF and FSO communication channel attenuation and two final models are built for attenuation. Data pre-processing and visualization are implemented before building the random forest, and there are 25 predictors included in the dataset. 70 % of the data is split as training data, and the rest of 30% is divided equally into validation and testing data for measuring the model performance. In terms of the decision tree, both coarse tuning and fine-tuning are applied for hyperparameter optimization.

For specific and generic models, the features are ordered based on OOB information with metrics of RMSE and R^2, and they are selected when there is a significant increase in R^2 and a decrease in RMSE. The number of predictors for RF from specific models ranges from 6 to 12 whilst that for FSO attenuation ranges between 5 and 11. On the other hand, there are 14 and 13 predictors selected for FSO and RF attenuation from generic models. Small sample sizes with a huge number of predictors in specific models predicting the attenuation under foggy and snowy weather result in a better performance than generic models, but the generic model is chosen as the final model because of high efficiency and better performance on RMSE and R^2, and the principle of Markov chain can also explain this.

Furthermore, the predictors that have an essential impact on RF communication channels are absolute humidity, rain intensity, distance, maximum rain intensity, temperature, minimum rain intensity, visibility, SYNOPCode, relative humidity, maximum particulate, maximum temperature, frequency, and minimum visibility. Moreover, the predictors in the FSO models of attenuation prediction are distance, visibility, temperature, minimum particulate, minimum visibility, absolute humidity, particulate, minimum temperature, maximum visibility, relative humidity, maximum particulate, temperature difference, time, and maximum temperature. Finally, the finalized models for both RF and FSO models perform better in RMSE and R^2 compared to the models using all predictors. RMSE for RF and FSO decline to 0.579 and 0.312 whilst R^2 increase to 0.970 and 0.994 respectively.

Moreover, of the two hybrid models, predicting FSO attenuation based on predicted RF (14 important features) and predicting RF attenuation based on predicted FSO attenuation (15 important features), RF hybrid model has better performances with RMSE and R^2 of 0.520 and 0.977

respectively. When using PCC to measure the correlation between FSO and RF attenuation, it demonstrates that the correlation between FSO and RF attenuation is high in snow, moderate in dust storms, and low in rainy weather, and the rest of the weather conditions witness a negligible correlation. Regarding MI, I_{EO}/H_{EO} has the highest value and negative correlation under dust storms, illustrating that the correlation is negative and non-linear. In other words, the two channels can compensate each other to the highest extent in dust storms because when one channel's attenuation increases, the other decreases.

Concerning model performance comparison among all three models, the specific model has the worst performance overall. For the FSO channel, the generic model performs better than the other two models in both RMSE and R^2. In terms of the RF channel, the hybrid model performs better under clear, fog, rain, snow, and shower weather for RMSE, and better under fog, rain, snow, and shower for R^2.

Therefore, the generic model has the best performance in general, but there are several notable compensations between the two channels. Specifically, the dust storm witnesses a negative non-linear correlation whilst the rainy weather has a positive linear correlation.

Acknowledgements

I extend my sincere gratitude to my project supervisor, Siu Wai Ho, for his invaluable guidance and help throughout the two trimesters. His insightful guidance and warm encouragement have significantly helped to improve and complete this report. He has also taught me many valuable knowledge and skills that will be essential in my future career. I am deeply grateful and honoured to have had the opportunity to learn and grow under his guidance.

I also extend my gratitude to the University of Adelaide for providing a platform that allowed me to receive invaluable guidance from all professors and tutors throughout my five years of undergraduate and postgraduate journey.

A Appendices

A.1 Abbreviation

Code link for this project: https://github.com/TinaQT/UoA-Report.git

Abbreviation	**Full Name**
RF	Radio Frequency
FSO	Free Space Optical
Att	Attenuation
OOB	Out-of-bag
PCC	Pearson Correlation Coefficient
MI	Mutual Information
SYNOPCode	Surface Synoptic Observations Code
RMSE	Root Mean Square Error
R^2	R Squared
Part	Particulate
Temp	Temperature
AH	Absolute Humidity
Freq	Frequency
RH	Relative Humidity
SC	SYNOPCode
RI	Rain Intensity
Dist	Distance

A.2 Feature Selection

	RF(Clear)	FSO(Clear)	RF(Dust)	FSO(Dust)
AH	✓	✓	✓	
AHMax	✓		✓	
AHMin	✓		✓	
Distance	✓	✓	✓	✓
Frequency	✓		✓	
Particulate				
ParticulateMax				
ParticulateMin				
RainIntensity				
RainIntensityMax				
RainIntensityMin				
RelativeHumidity	✓		✓	✓
SYNOPCode				
Temperature	✓	✓	✓	
TemperatureDiff		✓	✓	
TemperatureMax		✓	✓	
TemperatureMin		✓	✓	
Time	✓	✓	✓	✓
Visibility	✓	✓	✓	✓
VisibilityMax		✓		✓
VisibilityMin	✓	✓		✓
WindDirection				
WindSpeed				
WindSpeedMax				
WindSpeedMin				

Table A.2: Selected Features for RF & FSO Models Under Clear Weather & Dust Storm

	RF(Fog)	FSO(Fog)	RF(Drizzle)	FSO(Drizzle)
AH	✓		✓	
AHMax			✓	
AHMin	✓		✓	
Distance		✓	✓	✓
Frequency			✓	
Particulate		✓		✓
ParticulateMax		✓	✓	✓
ParticulateMin		✓		
RainIntensity	✓		✓	
RainIntensityMax	✓		✓	
RainIntensityMin	✓		✓	
RelativeHumidity	✓		✓	
SYNOPCode				
Temperature		✓		✓
TemperatureDiff				
TemperatureMax		✓		
TemperatureMin	✓	✓		
Time		✓		
Visibility	✓		✓	
VisibilityMax				✓
VisibilityMin	✓			
WindDirection				
WindSpeed				
WindSpeedMax		✓		
WindSpeedMin				

Table A.3: Selected Features for RF & FSO Models Under Foggy & Drizzle Weather

	RF(Rain)	FSO(Rain)	RF(Snow)	FSO(Snow)
AH	✓	✓	✓	
AHMax				
AHMin				
Distance	✓	✓		
Frequency				
Particulate		✓		✓
ParticulateMax		✓		✓
ParticulateMin	✓	✓		✓
RainIntensity	✓		✓	
RainIntensityMax	✓		✓	
RainIntensityMin	✓		✓	
RelativeHumidity		✓		
SYNOPCode				
Temperature	✓	✓		✓
TemperatureDiff		✓	✓	✓
TemperatureMax		✓		✓
TemperatureMin				
Time				
Visibility	✓	✓		✓
VisibilityMax		✓		
VisibilityMin				
WindDirection				
WindSpeed				
WindSpeedMax			✓	✓
WindSpeedMin				

Table A.4: Selected Features for RF & FSO Models Under Rain & Snow

	RF(Showers)	FSO(Showers)	RF(All)	FSO(All)
AH	✓	✓	✓	✓
AHMax				
AHMin	✓			
Distance		✓	✓	✓
Frequency			✓	
Particulate		✓		✓
ParticulateMax	✓		✓	✓
ParticulateMin		✓		✓
RainIntensity	✓		✓	
RainIntensityMax			✓	
RainIntensityMin	✓		✓	
RelativeHumidity		✓	✓	✓
SYNOPCode			✓	
Temperature		✓	✓	✓
TemperatureDiff				✓
TemperatureMax			✓	✓
TemperatureMin				✓
Time		✓		✓
Visibility	✓	✓	✓	✓
VisibilityMax		✓		✓
VisibilityMin		✓	✓	✓
WindDirection				
WindSpeed				
WindSpeedMax				
WindSpeedMin				

Table A.5: Selected Features for RF & FSO Models Under Showers & All Weather Conditions

References

[1] Anees, S & Bhatnagar, MR 2015, 'Performance evaluation of decode-and-forward dual-hop asymmetric radio frequency-free space optical communication system', *IET Optoelectronics*, vol. 9, no. 5, pp. 232–240.

[2] Australian Weather News n.d., *Meteorological Data Overview*, Australian Weather News, <http://www.australianweathernews.com/learn_about_meteorological_data_overvi

[3] Chowdhury, MZ, Hasan, MohK, Shahjalal, M, Hossan, MdT & Jang, YM 2020, 'Optical Wireless Hybrid Networks: Trends, Opportunities, Challenges, and Research Directions', *IEEE Communications Surveys and Tutorials*, vol. 22, no. 2, pp. 930–966.

[4] Guyon, I & Elisseeff, A 2003, 'An introduction to variable and feature selection', *Journal of Machine Learning Research*, pp. 1157-1182, March 2003.

[5] Ho, SW, Mitchell, L & Wang, V 2024, 'Modelling channel attenuation in hybrid optical/E-band system', *IEEE Transactions on Wireless Communications.*

[6] Hastie, T, Tibshirani, R & Friedman, J 2009, *Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Second Edition, Springer, New York.

[7] Mukaka, MM 2012, 'A guide to appropriate use of correlation coefficient in medical research', *Malawi Medical Journal*, vol. 24, no. 3, pp. 69-71.

[8] Naghibi, SA, Ahmadi, K & Daneshi, A 2017, 'Application of support vector machine, random forest, and genetic algorithm optimized random forest models in groundwater potential mapping', *Water Resources Management*, vol. 31, no. 9, pp. 2761–2775.

[9] Peng, H, Long, F & Ding, C 2005, 'Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 8, pp. 1226–1238.