Hybrid Optical/Radio Frequency Communication Channel Model

Nithin Manyamvollu
A1879720

November 24, 2024

Report submitted for **Research Project A** at
The University of Adelaide

Project Title: **Hybrid Optical and RF Communication Model**
Project Supervisor: **Siu Wai Ho**

In submitting this work I am indicating that I have read the University's Academic Integrity Policy. I declare that all material in this assessment is my own work except where there is clear acknowledgement and reference to the work of others.

I give permission for this work to be reproduced and submitted to other academic staff for educational purposes.

OPTIONAL: I give permission this work to be reproduced and provided to future students as an exemplar report.

Abstract

This study focuses on the creations of models that predict signal loss in Free Space Optics (FSO) and Radio Frequency (RF) communication systems using Random Forest algorithms. FSO systems are vulnerable to atmospheric factors such as fog and rain. On the other hand, the RF systems are bandwidth-limited and dependent on rain intensity. By utilising the Random Forest algorithms, models are developed to estimate signal attenuation in real-time based on various environmental factors. These factors include the ones such as visibility, rainfall, and temperature. The models are evaluated using Root Mean Square Error (RMSE) and R-Squared metrics, which are critical part of this work. A major insight has revealed on the use of specific and generic models that can change the selection of models for this task. A comparison was done based on their performance with established models. In this case, the specific model has outperformed their generic model in most of the cases. In some cases having improvements as 18.96% in Dust Storm conditions, among others in FSO systems. The model also showed much lower RMSE compared to the generic model, with as low as -69.69% in RFL systems to snow conditions. Therefore, such better predictive models are properly suited for this task. Therefore, it was used in this case. Results show that the Random Forest models can provide reliable predictions for signal loss. Therefore, using this approach can improve the modelling.

Contents

List of Figures

1 Introduction and Background

1.1 Introduction and Rationale

FSO and RF communication systems are critical in the modern communication landscape. FSO systems use light to transmit data through the atmosphere. This means that the effects of the different atmospheric features are present. This is also true for various other effects at different situations. However, these systems offer significant bandwidth advantages. This allows for efficient communication using optics. Despite having such a strong capability, they are highly susceptible to atmospheric attenuation. These are caused by factors such as fog, rain, and turbulence (Khan et al., 2022). On the other hand, RF systems use radio waves to transmit data. These systems offer greater stability even when there are difficult conditions. However, they are limited by the bandwidth capacity.

The increasing demand for reliable and high-capacity communication systems requires the development of hybrid models. They are expected to predict signal attenuation in both FSO and RF systems (Wu et al., 2023). These systems should be functional under various weather conditions. Many weather conditions are difficult to handle for the FSO (Free Space Optics) model. On the other hand, some weathers the RF (Radio Frequency) model performs well. The aim therefore becomes to check whether such a stable model can be achieved through understanding the different RF models. In this case, this type of model is chosen to use Random Forest. These models are highly capable decision trees that use advanced algorithmic capabilities to fill the gap in the understanding of the data (Fernandes et al., 2021). These algorithms allow for finding complex and nuanced patterns in the data. Thus, this study aims to fill that gap by using Random Forest algorithms to predict attenuation levels. It also tries and assess the impact of different variables on communication system performance. By improving our understanding of how weather conditions affect signal transmission, these models can help optimise the planning.

The advancement of these technologies would allow for a more robust and enhanced system designed to be operated in an automated manner. This would enhance the industrial communication methods and would enhance the businesses in the commercial markets. Currently, it is a necessity to understand the signal processing to its core as the information is expanding at a rapid pace. Therefore, the prediction of signal attenuation holds an important place that will change the trajectory of the world in a much smarter way.

1.2 Aim, Objectives, and Research Questions

The aim of this study is to develop and evaluate predictive models for signal attenuation in FSO and RF systems. The objectives are:

1. To identify key environmental factors affecting signal attenuation in FSO and RF systems.
2. To create Random Forest models that can accurately predict signal attenuation.

The main research questions are:

1. How does treating environmental factors as categorical variables improve the predictive accuracy of signal attenuation models for FSO and RF systems?
2. To what extent can specialised Random Forest models improve predictive performance over generic models across diverse environmental conditions?

1.3 Review of current literature

In their research, Haluška et al. (2020) focused on predicting the RSSI (Received Signal Strength Indicator) parameter. This parameter controls the hard switching in hybrid FSO and RF systems. This parameter is critical in analysing the system to its fullest extent. The paper explored different machine learning approaches related to decision trees. This is because the decision trees are a powerful algorithm for these types of tasks providing with higher levels of R squared values and predictive accuracies. The decision trees and the AdaBoost regressor both was compared to predict signal strength based on environmental conditions. Their results indicated that the AdaBoost-enhanced decision tree method achieved higher predictive capability than normal decision trees. However, there is a longer required training time associated with it. Haluška et al. (2020) also highlighted the importance of training models on diverse weather conditions. The paper claims that it improves the prediction accuracy of the model. The paper recommended investigating random forests and neural networks as possible candidates to optimise hybrid FSO and RF systems performance.

Yahia et al. (2021) proposed another hybrid RF and FSO communication strategy for satellite communications. The target of this paper was optimising performance of the existing systems. Their paper found that in a signal-based communication, soft-switching setup allows a LEO satellite to adapt its transmission power. This in turn enables the selection between RF and FSO links based on weather conditions. This switching uses a context-aware sensor system. This enables the system to be performant in the long term. The paper used outage probability and simulations to assess the model's performance. Therefore, the paper was able to demonstrate the improved power efficiency over conventional hybrid RF and FSO systems. On the contrary, Haluška et al. (2020) utilises context-aware switching for efficiency, which is different from the conventional approach. Their paper highlighted the importance of using smart switching. This method enables the system to tackle the weather conditions. The model also showed the drastic effect the different parameters of the weather on the very critical parameter known as the signal attenuation.

Lionis (2021) investigated the performance prediction of FSO systems in maritime environments. The focus was strongly on the communication systems. The measurement of success is based on the accuracy of various machine learning (ML) algorithms. The study utilised a large dataset of RSSI and seven atmospheric parameters. The list of parameters included the wind speed, temperature, and humidity. The values were collected over a year from a commercial FSO system and a weather station. The paper used various commonly known algorithm, which were k-nearest neighbours, decision trees, random forests, gradient boosting, and ANN. This in turn allowed the system to model RSSI performance. This contrasts with the approach of Haluška et al. (2020).

Lionis (2021) in their paper used the comparison system of RMSE and the R-Squared. These are two of the standard metrics for such cases. Their findings indicated that all ML models significantly outperformed traditional regression techniques. The ANN model achieved the best result surpassing the other models in terms of RMSE and R-Squared. The model achieved the highest R-Squared of 0.94867, which is 94.9%. This means that the parameters in the model can explain the 94.9% variation in the data. The random forests models yielded the best RMSE of 7.37. The study concluded that ML methods provide a robust framework for accurately predicting RSSI. Yahia et al. (2021) on the other hand shows that it is more useful when predicting using the hybrid approach for the signal attenuation on the FSO and RF systems. The paper also showed that in complex atmospheric conditions, such ML algorithms are extremely powerful. The paper suggested that further improvements can be made by integrating additional data for ongoing predictions.

Esmail (2023) proposed a hybrid optical fibre and FSO communication system in their paper. This is to meet the increasing demand for network capacity. This is driven by new digital applications and the increasing market share. This system utilises the different ML techniques. This is specifically Gaussian Process Regression (GPR), to predict key channel impairments. Different key features such as turbulence, optical signal-to-noise ratio (OSNR), and chromatic dispersion (CD) are used to build the model. Lionis (2021) shows that other models outperform regression techniques in this regard, which is supported by the approach in this paper. This shows the diverse amounts of features to be considered for such models. The model's performance was assessed using metrics such as RMSE and R-squared. These are standard metrics where the higher RMSE and R-Squared indicates higher prediction accuracy. This is particularly true under varying channel conditions. However, the GPR model faced challenges in accurately predicting light turbulence parameters.

Esmail (2023) showed that there are difficulties the model faces in predictive capabilities in terms of different situations and conditions that are present in the modelling. The authors compared their GPR approach with RF and SVM models. They found that these models are equal or more effective in most scenarios. This model can aid telecommunications operators in optimising FSO performance. This is done using the adaptive modulation and digital signal processing techniques. Future research directions they showed include predicting additional impairments and exploring advanced methods. Lionis (2021) supports this argument and states that in many situations these models can model complex phenomena. These methods include the ones like quantum and graph neural networks for better enhanced performance of the models.

Kaur and Sharma (2023) analysed the quality of the received signal under various weather conditions. These conditions are like the SYNOP Code used in the dataset. The actual conditions include clear air, low haze, heavy haze, and light fog. ML techniques, specifically ANN and SVM were employed to predict signal impairments at the receiving end. The ANN model achieved impressive performance with a RMSE of 0.148 and an R-squared value of 0.98. In contrast, the linear SVM model exhibited RMSE and R-Squared values of 0.937 and 0.76, respectively. Lionis (2021) supports the use of ANNs and other specific models for this purpose and highlight the benefits of using those algorithms. This indicated it as the best-fit model for estimating the quality factor of the received signal. This research showed the effectiveness of ML models in predicting and validating the performance in predicting the attenuation of the signals.

Literature Gap: There is a lack of predictive modelling using Random Forest algorithms on the different datasets of signal attenuation. Very few studies focus on using this algorithm and in many situations, they have chosen the ANNs to a more suited model. This has led to this research choosing this segment as there is a severe gap in using RF algorithms on the attenuation prediction for the RFL and FSO systems.

2 Methods

This study uses different Random Forest algorithms to predict signal attenuation in both FSO and RF systems. Random Forest is a decision tree-based machine learning technique. It is a well-established technique for these types of tasks where the different conditions may have indirect effects on the outcome of the quality. Therefore, it is well-suited for non-linear relationships and complex interactions among input features. In this case, two main models are considered. These are the FSO and RF models (Haluška et al., 2020). The key objective lies in understanding the different effects of the environmental factors on the two different types of models. Another key objective is to understand whether these effects influence the outcome of the model's accuracy. Therefore, for this case, the consideration of the two models is necessary.

2.1 Feature Selection

Feature selection plays a critical role in the performance and accuracy of both the FSO and RF models. This is because each model depends on distinct environmental factors that directly affect the key variable signal attenuation. This is the target variable in the dataset and is the topic of research. In the FSO model, the focus is on visibility conditions (Nebuloni and Verdugo, 2022). This is due to the optical signals transmitted in free space. The free space is always affected by factors such as fog and humidity. These factors significantly affect how light passes through the atmosphere. Hence, it leads to varying degrees of signal loss. Such a severe case would be heavy fog. This can cause scattering of the optical signal. This will result in higher attenuation. Temperature and wind speed are also factors for such models. They also have the capability to influence the stability of atmospheric conditions. This will further contribute to signal loss by making the attenuation higher. In contrast, the RF model emphasizes weather conditions such as the rainfall intensity, cloud cover, wind speeds, and temperature fluctuations. Rainfall is a particularly critical factor in RF communication. This can be ascertained as water droplets can absorb and scatter RF signal. This has led to signal degradation and attenuation. Wind speed and temperature can also affect the signal strength. This is done by altering atmospheric density and causing diffraction (Ghoname et al., 2020). Along with that, the amount of moisture in the air, humidity plays a role in RF attenuation. This is truer at higher frequencies, which makes it an essential feature to include in the model. Such a feature selection is critical. This is because selecting features that align with the unique physical properties of both systems, models are better equipped. These models will be able to handle the specific challenges posed by atmospheric conditions. Therefore, the model will allow for more accurate predictions of signal attenuation.

This shows that it is extremely crucial that appropriate feature sets are selected for both the specific and generic models. The feature sets comprise of different factors and since, the signal mechanisms are different, they would impact the systems differently. The selected features are understood from the context of the signal processing. In the context of the FSO, thus the selected features consist of visibility factors, which are understood to be affecting the FSO based attenuation in a signal. On the other hand, the RF model requires careful selection of cloud cover, temperature and wind speed factors. In this scenario, there are no such factors as cloud cover and therefore, it has not been considered here. These features are selected with the approach that since, the actual dataset consists of all the features, most of the features will not be impactful for the purpose of the model development and therefore, such the selected features were considered.

2.2 Dataset creation

The dataset creation step consisted of different features and ratios that were considered for forming the different dataset parts, which were used for the training of the models. It is essential for training the model that randomness must be controlled and there should be a standard test size. In this case, the test size has been consistent throughout with each model having the test size of 0.2. This means that for the testing purposes, each of the models will be tested on 20% of the data. This reduces the redundant sampling and inconsistent performance across different dimensions of the datasets (Liu et al., 2020). In this scenario, such an approach is a necessity. It enables for developing efficient models. After that, the randomness has been controlled in a deterministic manner. This is done by keeping the seed same for all the operations.

This reduces the possibility of having different results each time the comparison is done and establishes a stable modelling approach. This approach benchmarks its data with the other standard values after the model has completed its tasks. Then the dataset is created with separated feature sets for each of the models. This enables for creating models that use the separated feature sets that are thought to be the most important for the modelling part. By doing this, the models become much more robust and effective (Mishra et al., 2020). Therefore, this is vital that the model is carefully trained with appropriate data, which was done here. The selection of the data for the model training shows that there are factors such as the determinism. This can influence the model's outcome. On the other hand, it is also essential that the selected datasets carry the most amount of information.

2.3 SYNOP Code based model

The SYNOP (Surface Synoptic Observations) code is a standardised system used for reporting weather observations globally. In this system, a SYNOP code-based modelling approach is used first to determine the accuracy of the models in specific weather conditions (Yabra et al., 2024). There are 7 SYNOP code-based models in this consideration as the dataset contains different SYNOP codes. Therefore, first the data is split across 7 different sets while training and then each of these sets are trained on a Random Forest model. Therefore, this system utilises this standardised weather data format to predict signal attenuation in communication systems. Hence, both the FSO and RF models are considered. The SYNOP format includes key meteorological variables such as temperature, humidity, cloud cover, and visibility. This makes it work like a standard model. However, in terms of predictive capability it would be a specific model. That makes it a valuable source of data for assessing environmental impacts on signal transmission. In this model, the relevant SYNOP code data is ingested to train machine learning algorithms such as Random Forests. The objective is same as predicting the signal loss or attenuation of the two different types of models. For the FSO system, features such as visibility and humidity used (Verdugo et al., 2023). On the other hand, for the RF system, typical features are rainfall, and wind speed. This separation is necessary as modelling of these two algorithms rely on the existing knowledge. By integrating SYNOP codes, this model benefits from specific conditional changes on the data.

2.4 Generic FSO & RF Model

The FSO model predicts signal attenuation based on key environmental factors. These includes visibility, temperature, and wind speed. The signal attenuation is expressed in decibels per kilometre (dB/km). Data for these environmental variables are separated from the data source appropriately. Then the model is trained on historical data to predict how these factors affect signal strength. On the other hand, the RF model focuses on different environmental factors. These are the strength of rainfall, particulates, temperature, and wind speed (Khan et al., 2021). Like the FSO model, the RF model also predicts signal attenuation in dB/km. The historical data used to train this model includes real-time weather observations and signal strength measurements. This allows the model to capture the complex interactions between the weather conditions and signal attenuation.

2.5 Model Evaluation

The performance of both models is evaluated using two key metrics. These are RMSE and the R-squared.

Root Mean Square Error (RMSE): The RMSE measures the average magnitude of prediction error. It is defined mathematically as:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{1}$$

where y_i represents the observed values, $\hat{y}_i$ represents the predicted values, and n is the total number of observations. A lower RMSE indicates a better-performing model, as it reflects fewer errors in the predictions.

Coefficient of Determination ($\mathbf{R^2}$): The R^2 metric measures the proportion of variance in the dependent variable that is predictable from the independent variables. It is computed as:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{2}$$

where $\bar{y}$ is the mean of the observed values. R^2 values range from 0 to 1, with higher values indicating that the model explains a greater proportion of the variance in the data. A well-performing model is characterized by a higher R^2.

These two are critical metrics that determine the quality of a model that does prediction on continuous variables. RMSE provides a measure of the average prediction error. The R-Squared indicates how closely the predictions match the actual data and the explained variability on the dataset. The lower the RMSE of the model, the better it has performed on the test set. On the other hand, the higher the R-Squared, the better the model explains variation in the original data (Chicco et al., 2021). A lower RMSE and higher R-Squared, signify a well-performing model. The two different model types are compared with one another and then the better one is finalised.

3 Results

	Clear	Dust Storm	Fog	Drizzle	Rain	Snow	Showers
FSO RMSE	1.549	1.819	0.824	1.381	1.523	1.754	1.108
FSO R^2	0.811	0.962	0.956	0.851	0.873	0.885	0.917
RFL RMSE	0.850	0.458	0.523	0.752	0.909	0.282	1.039
RFL R^2	0.890	0.981	0.904	0.934	0.948	0.957	0.874

Table 1: Performance of FSO and RF Models under Different Weather Conditions

From the Table 1, the comparison of the results between the seven distinct random forest models and the generic model shows important findings. The specific model reveals significant insights regarding their predictive capability. The models have outperformed the generic random forest model in some weather conditions. On the other hand, in some SYNOP based categories, the generic model has performed better. This can be attributed to specific nature of the feature values corresponding to different conditions. For instance, the Random Forest model for FSO in fog conditions achieved an RMSE of 0.824 and an R-Squared of 0.956. These results are much stronger results than the generic FSO model's RMSE of 1.687 and R-Squared of 0.810. Here, it is critical to observe the RMSE (Khan et al., 2022). This suggests that a model trained for specific conditions can accurately capture the nuances of signal attenuation in some of those conditions. However, in other conditions such as a clear weather, the generic model provides improved performance metrics.

	Generic Model
FSO RMSE	1.687
FSO R^2	0.810
RF RMSE	0.897
RF R^2	0.932

Table 2: Performance of Generic Model

In the table 2, it can be observed that the FSO models generally performed better than the RFL models in terms of R-Squared values across some weather conditions. This is a striking output, as that means the dependencies across the different feature sets are critical for some models. It also means that the model is sensitive to different changes, which are majorly influential. It can be observed that the FSO model in snow conditions achieved an R-Squared of 0.917, while the RFL model achieved an R-Squared of 0.957. This shows that both models had comparable but strong predictive capabilities, with the RFL model slightly outperforming the FSO model. In contrast, the FSO model achieved an R-Squared of 0.811 under clear sky conditions, whereas the RFL model performed much better with an R-Squared of 0.890. These results for the RFL model under clear conditions demonstrate its capability to predict effectively with a well-selected feature set (Babatunde et al., 2022). Hence, it can be stated that without proper modeling and sufficient data, a generic model with standard feature sets will fall short compared to an enhanced specialist model.

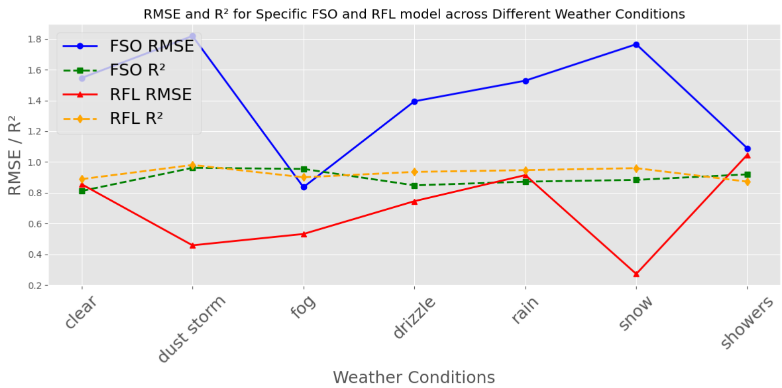

Figure 1: RMSE and R-Squared values plotted for the two specific models.

The random forest models for drizzle and snow showed remarkable results for both FSO and RFL predictions. After plotting the RMSE and R-Squared for both the FSO and RFL models, distinctive patterns emerge (Fig. 1). The plot shows that the FSO model demonstrates excellent R-Squared values across the board, with values as high as 0.917 in snow conditions and 0.962 in dust storm conditions. However, the FSO model also exhibits higher RMSE in certain scenarios, such as 1.754 in snow conditions and 1.523 in rain conditions. On the other hand, the RFL model demonstrates consistently lower RMSE values in challenging weather conditions, highlighting its effectiveness. For instance, the RFL model achieved an RMSE of 0.282 in snow conditions, significantly outperforming the FSO model. It also achieved an R-Squared of 0.957 in snow conditions, surpassing the performance of the generic random forest models. These results indicate that the RFL model's specialized approach allows for more precise predictions in difficult weather situations, enabling the potential use of a context-sensitive switching mechanism for RF models. The generic random forest models, by contrast, delivered an RMSE of 1.687 for the FSO model and 0.897 for the RF model, along with R-Squared values of 0.810 and 0.932, respectively. These results underscore that the generic models fall short of the predictive accuracy achieved by the specific models in various weather conditions. The models trained for specific SYNOP codes have consistently delivered higher R-Squared values and lower RMSE in targeted scenarios. This adaptability factor, as shown by the performance of the specific models, enables more efficient switching to appropriate conditions while maintaining high prediction accuracy (Mohamed et al., 2023). However, the use of specific models will require more computational capabilities due to it specifically built individual weather conditions.

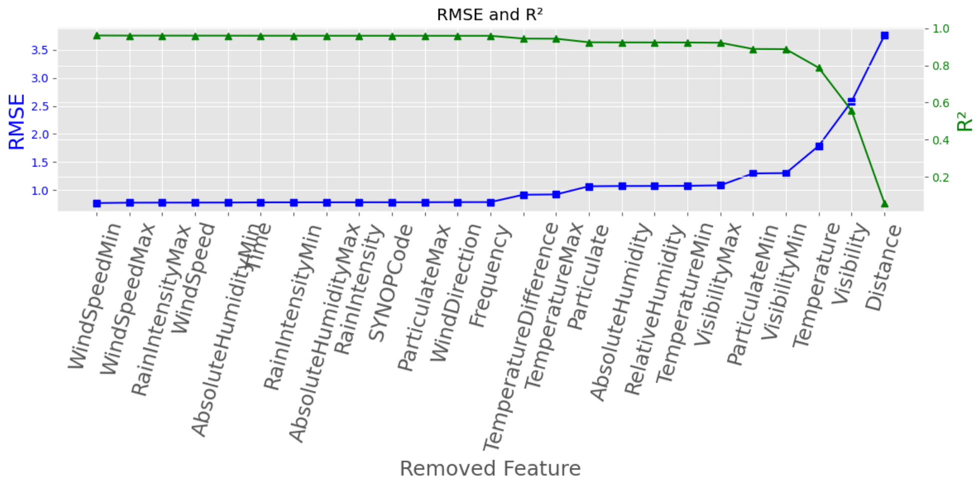

Figure 2: Declining R-Squared and inclining RMSE using OOB for the FSO model.

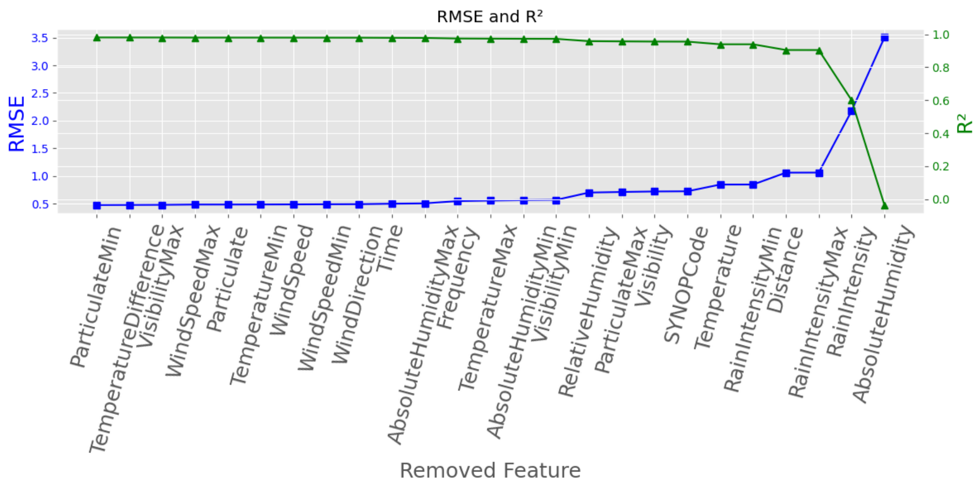

Figure 3: Declining R-Squared and inclining RMSE using OOB for the RFL model.

The OOB (Out of Bag) approach applied here showcases how important each features are. At each step in the process, first the least important feature is removed and then a new model is trained while keeping the R-Squared and RMSE score stored. Then the process is continued until each feature is tested. Then for the two systems, FSO and RFL, the different scores are plotted after sorting (Fig. 2 and 3). This helps visualise the impact of the feature removal for both conditions.

An important takeaway from these findings is the role of significant features in improving prediction accuracy. It has been shown that feature sets are a vital piece of information for modelling the data. The analysis of feature importance for both FSO and RFL models shows that parameters such as wind speed, absolute humidity, and temperature are critical to the model's success (Samy et al., 2022). In this scenario, the removal of wind speed and its minimum and maximum variations led to critical outcomes. It showed a significant increase in RMSE values and a decline in the R-Squared scores. This is true for both FSO and RFL models. This information highlights the dependency of these models on accurate data to predict signal attenuation. In the case of FSO models, wind speed-related variables were most important. This shows that for the FSO model, critical factors are the wind speed and wind related factors. Similarly, visibility and particulate-related features were significant contributors for RF models.

The performance of the generic model, was less beneficial. The RMSE of 1.694 was seen in FSO and an RFL RMSE of 1.701. On the other hand, the corresponding R-Squared values of 0.808 and 0.755 was seen. This indicates that the generic model can make predictions across all weather conditions. However, the capability of its predictive accuracy was lower on average than that of the specific models. This is an important information that should be considered when considering the different conditions. For a generic baseline performance, such a model can be used for its simplicity (Kiangala and Wang, 2021). However, the generic model failed to adapt to different diverse conditions represented by different SYNOP codes. This resulted in lower performance metrics of R-Squared and higher RMSE for the same predictions. This may be due to the model's inability to generalise across highly variable atmospheric factors.

For choosing a final model, there are various aspects that must be considered. A final model is a type of model that is most suited for this task. IN this scenario, it is evident that two models are performing well in their respective conditions. For predicting signal attenuation in hybrid FSO and RF systems, both the generic and specific models have their strengths and weaknesses. The generic model provides the advantage of being simpler to implement. However, it only produces output from one trained model for all conditions. This is usually a problematic situation as shown here. It also suffers from lower predictive accuracy, which is particularly problematic when working with such different weather conditions (Hewage et al., 2021). On the other hand, the specific models trained for individual SYNOP codes deliver more accurate predictions. This is demonstrated by their higher R-Squared values and lower RMSE scores across different weather conditions.

The final goal in such a situation is optimising predictive performance of the models. In this regard, the generic model can be considered simpler for this task as these two generic RF and FSO models can produce results that are well suited. However, these results lack the considerations of the nuances across different weathers and therefore, are incapable of handling nuanced and complex data points. This ensures the most reliable communication performance from the specific models being the more suitable choice. Their ability to tailor predictions to the exact weather conditions enables them to achieve higher accuracy (Sanober et al., 2021). This makes them more valuable in applications where precision is critical. The generic model may still serve as a baseline. However, due to its nature of being less-focused means it cannot be used in a highly critical situation. This is because the output from the well suited specific Random Forest model outperforms the model. Therefore, the model cannot compare well with the tailored insights provided by the specific ones. Thus, the use of specific models, is recommended for both FSO and RFL in a real-world critical situation.

4 Discussion & Conclusion

4.1 Summarising the Main Results

Metric	Clear	Dust Storm	Fog	Drizzle	Rain	Snow	Showers
FSO RMSE	-8.17%	7.84%	-51.12%	-18.11%	-9.68%	4.01%	-34.28%
FSO R^2	0.08%	18.71%	17.97%	5.00%	7.71%	9.20%	13.11%
RFL RMSE	-5.20%	-48.95%	-41.66%	-16.18%	1.34%	-68.57%	15.79%
RFL R^2	-4.53%	5.20%	-3.03%	0.24%	1.67%	2.68%	-6.22%

Table 3: Percentage Improvements of the Specific Model Compared to Generic Model

The primary findings of the study reveal that the specific Random Forest model trained with the SYNOP code outperforms the generic model. From Table 3, it can be observed that for specific conditions, such as "Drizzle," the specific model achieves a higher R^2, which makes it more reliable for predictions. Here, the SYNOP code is treated as a categorical variable, allowing the specific model to capture the unique characteristics of each weather condition effectively. The Random Forest models trained for specific weather conditions demonstrate superior performance across key metrics such as RMSE and R^2. These metrics are crucial for assessing the quality of models predicting continuous variables. In nearly all weather scenarios, the specific model outperformed the generic model, as evidenced by higher R^2 values and lower RMSE for most conditions. For example, the FSO-specific model achieved a significant RMSE improvement of -51.12% in fog conditions and an R^2 improvement of 17.97% compared to the generic model. Similarly, the RFL-specific model showed an RMSE improvement of -68.57% in snow conditions and an R^2 improvement of 2.68%.

The higher R^2 values indicate that the specific models explain a greater proportion of the variance in the dataset compared to the generic models. This suggests that categorizing weather conditions as separate entities rather than using a generic feature set is a more effective approach. Specific models are better suited for capturing real-world effects that may be difficult to assess in a single, generalized framework. The adaptability of the specific models is another notable advantage. Each model is tailored to a distinct weather condition, enabling precise predictions across varying environmental scenarios. This flexibility allows for the implementation of a context-sensitive switching mechanism, where the system selects the most appropriate model based on prevailing weather conditions. This adaptability makes specific models especially valuable for real-time and automated systems in hybrid FSO and

RF communication networks. While the generic models performed adequately, they exhibited limitations in extreme weather conditions. For instance, under heavy fog or snow, the generic models struggled with accuracy. This resulted in higher RMSE values. This shows the potential for overfitting or underfitting of key features in a single model. The findings suggest that leveraging specific models for different weather conditions enhances the prediction process. However, in some cases for achieving superior results, a generic model can be chosen across key performance metrics. Although generic models may offer a broad approach, they lack the precision and adaptability required for challenging real-world scenarios.

Hyperparameter tuning: Hyperparameter tuning in the experiment was performed using RandomizedSearchCV, a method that efficiently searches the hyperparameter space to identify the optimal configuration for each Random Forest model. The tuning involved 10 iterations over a grid of hyperparameters, including the number of estimators (n_estimators), maximum depth (max_depth), minimum samples per split (min_samples_split), minimum samples per leaf (min_samples_leaf), and bootstrap sampling (bootstrap). The search was evaluated with 5-fold cross-validation, optimizing the negative mean squared error metric. The tuning demonstrated the model's ability to fit the appropriate data effectively and efficiently. This enabled the optimisation of its performance for each specific weather condition. Key hyperparameters for the best models were largely consistent across weather conditions, including a max_depth of 20 and n_estimators of 200. This process ensured that each model was tailored to the specific weather condition, enhancing predictive accuracy while preventing overfitting.

4.2 Contextualising the Results Relative to the Literature

The findings of this study align with the insights presented in the existing literature in the RFL and FSO systems. The results indicate that specific models tailored to distinct weather conditions outperform a generic random forest model. This is shown in the findings that the R-Squared is much higher on average and RMSE is much lower in comparison. This observation resonates with the conclusions drawn by Haluška et al. (2020). Their paper emphasized the critical role of environmental features in improving predictive accuracy. The paper recommended to explore more nuanced modelling approaches. This included the use of ML model like Random Forests, which is used here. This is further validated by our findings. This demonstrates that models trained on specific SYNOP codes yield significantly higher R-Squared values and are much more capable of explaining the data.

The analysis finds that using separate models and switching the models is a more advantageous approach for this type of task. This is supported by Yahia et al. (2021). Their paper introduced a soft-switching mechanism in hybrid RF and FSO systems based on weather conditions. Their findings on the adaptability of satellite communication systems reinforce the importance of contextual awareness. This is because for a more accurate prediction of the signals, the model will detect better attenuation when the appropriate model is set up. This is fully supported by the results. The tailored random forest models developed show improved performance due to their categorical approach. Their paper strongly emphasizes the notion of separate modelling for separate type of weather condition as they are vastly superior for modelling.

The findings state that machine learning methods provide superior predictive capability compared to traditional regression. This aligns with the findings of Lionis (2021). In their paper they focused on maritime environments, employing various machine learning algorithms to predict RSSI. Their paper also highlights the effectiveness of Random Forests in predicting signal quality under varying conditions. The authors observation that the ANN model achieved the best performance among ML techniques show the importance of understanding the complex patterns underlying the data.

The findings conclude that it is critical to use random forests for understanding the non-linear patterns in the data. In line with this, according to Trichili et al. (2021), the RF models in ML are some of the most influential models in understanding the non-linear patterns inside the data. This supports the findings of the analysis. There is a constant change in the development of different ML models and therefore, it resonates with the findings that the adaptive nature of the modelling is crucial for success. However, in terms of the generic modelling there is a consideration for more simplified approach is to be thought through.

The research on the findings shows the usability of the specific random forest models as found by the predictive models developed in this research. In this context, Esmail (2023) explored a hybrid optical fibre and FSO communication system. Their paper showed the utility of GPR in predicting channel impairments. This shows a contrasting approach of using GPR, which is a different algorithm that is chosen here. However, the challenges faced by the GPR model suggest a limitation in its applicability under diverse conditions. These models outperformed the generic model in most scenarios. The emphasis on context-specific modelling shows the need for continuous improvement in predictive capabilities. The findings indicate that there is a superior performance benefit with tailored models. Kaur and Sharma (2023) conclude with the same results. They analysed signal quality under various weather conditions using ANN and SVM techniques. Their results show that the model achieved high predictive accuracy with RMSE and R-Squared values. The strong performance of the ANN model reflects the growing trend of domination of the Deep Learning models in the field of machine learning, which is contrasted using Random Forest, which is a conventional model.

The current analysis was done using a SYNOP code-based random forest models. In this context, according to Shao et al. (2024), the comparative analysis shows that in hybrid FSO and RF systems specific modelling approaches are greatly beneficial. This directly supports the analysis and findings of this research. It is evident that it can yield superior performance metrics compared to generic alternatives. This show that the importance of ML applications to FSO and RF communication systems. This paves the way for further advancements in the field.

4.3 Limitations of the Analysis

One limitation of the study is the reliance on historical weather data. This is since in many situations these data are not enough relevant for modelling. These data points may fail to capture the full variability of real-time conditions. Additionally, while the models performed well under most conditions. The second major limitation of the analysis is that the model takes parameters that are known to be affecting the signal quality (Song et al., 2021). However, there can be more such factors that may affect that and has not yet been recorded. Therefore, without using those parameters, the signal attenuation cannot be appropriately predicted. Lastly, there are many ML models out there and this research focuses on the Random Forest model, which is a decision tree. However, models such as ANNs can be a quite improvement over such models in predicting signal attenuation.

4.4 Ideas to Overcome Limitations and Extend the Analysis

Future work could focus on integrating real-time weather data from different regions to improve the correctness of the models, making it much more viable for many different regions. Along with that, using other ML techniques, such as deep learning can greatly enhance the quality of the model (Henna et al., 2023). The model can be enhanced to work with other model to make it much more capable in predicting signal attenuation. Finally, further testing with more advanced datasets could refine the models and make it suitable for various predictive tasks.

5 References

Babatunde, K.S., Ibikunle, F.A., Arowolo, M.O., Alabi, A.J., Jiya, E.A. and Kehinde, O.P. Machine Learning Model for Classifying Free Space Optics Channel Impairments. In 2022 5th Information Technology for Education and Development (ITED), 1-8, 2022.

Chicco, D., Warrens, M.J. and Jurman, G. The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. Peerj computer science, 7:e623, 2021.

Clerckx, B., Huang, K., Varshney, L.R., Ulukus, S. and Alouini, M.S. Wireless power transfer for future networks: Signal processing, machine learning, computing, and sensing. IEEE Journal of Selected Topics in Signal Processing. 15(5):1060-1094, 2021.

Esmail, M.A. Performance monitoring of hybrid all-optical fiber/FSO communication systems. Applied Sciences, 13(14):8477, 2023.

Fernandes, M.A., Brandão, B.T., Georgieva, P., Monteiro, P.P. and Guiomar, F.P. Adaptive optical beam alignment and link protection switching for 5G-over-FSO. Optics Express, 29(13):20136-20149, 2021.

Ghoname, S., Fayed, H.A., El Aziz, A.A. and Aly, M.H. Performance evaluation of an adaptive hybrid FSO/RF communication system: impact of weather attenuation. Iranian Journal of Science and Technology, Transactions of Electrical Engineering, 44:119-128, 2020.

Haluška, R., Šuľaj, P., Ovseník, Ľ., Marchevský, S., Papaj, J. and Doboš, Ľ. Prediction of received optical power for switching hybrid FSO/RF system. Electronics, 9(8):1261, 2020.

Henna, S., Minhas, A.A., Khan, M.S. and Iqbal, M.S. Ensemble consensus representation deep reinforcement learning for hybrid FSO/RF communication systems. Optics Communications, 530:129186, 2023.

Hewage, P., Trovati, M., Pereira, E. and Behera, A. Deep learning-based effective fine-grained weather forecasting model. Pattern Analysis and Applications, 24(1):343-366, 2021.

Kaur, S. and Sharma, A. Impairment analysis of WDM Ro-FSO system under different weather conditions employing machine learning. Journal of Optoelectronics and Advanced Materials, 25(9-10):444-452, 2023.

Khan, A.N., Saeed, S., Naeem, Y., Zubair, M., Massoud, Y. and Younis, U. Atmospheric turbulence and fog attenuation effects in controlled environment FSO communication links. IEEE Photonics Technology Letters, 34(24):1341-1344, 2022.

Khan, M.N., Kashif, H. and Rafay, A. Performance and optimization of hybrid FSO/RF communication system in varying weather. Photonic

Network Communications, 41:47-56, 2021.

Kiangala, S.K. and Wang, Z. An effective adaptive customization framework for small manufacturing plants using extreme gradient boosting-XGBoost and random forest ensemble learning algorithms in an Industry 4.0 environment. Machine Learning with Applications, 4:100024, 2021.

Lapčák, M., Ovseník, Ľ., Oravec, J. and Zdravecký, N. Investigation of machine learning methods for prediction of measured values of atmospheric channel for hybrid FSO/RF system. In Photonics, MDPI, 9(8):524, 2022.

Lionis, A., Peppas, K., Nistazakis, H.E., Tsigopoulos, A., Cohn, K. and Zagouras, A. Using machine learning algorithms for accurate received optical power prediction of an FSO link over a maritime environment. In Photonics, MDPI. 8(6):212, 2021.

Liu, S., Chen, P.Y., Kailkhura, B., Zhang, G., Hero III, A.O. and Varshney, P.K. A primer on zeroth-order optimization in signal processing and machine learning: Principals, recent advances, and applications. IEEE Signal Processing Magazine, 37(5):43-54, 2020.

Mishra, P., Biancolillo, A., Roger, J.M., Marini, F. and Rutledge, D.N. New data preprocessing trends based on ensemble of multiple preprocessing techniques. TrAC Trends in Analytical Chemistry, 132:116045, 2020.

Mohamed, P.H., El-Shimy, M.A., Shalaby, H.M. and Kheirallah, H.N. Hybrid FSO/RF system over proposed random dust attenuation model based on real-time data combined with G–G atmospheric turbulence. Optics Communications, 549:129891, 2023.

Nebuloni, R. and Verdugo, E. FSO path loss model based on the visibility. IEEE Photonics Journal, 14(2):1-9, 2022.

Samy, R., Yang, H.C., Rakia, T. and Alouini, M.S. Ergodic capacity analysis of satellite communication systems with SAG-FSO/SH-FSO/RF transmission. IEEE Photonics Journal, 14(5):1-9, 2022.

Sanober, S., Alam, I., Pande, S., Arslan, F., Rane, K.P., Singh, B.K., Khamparia, A. and Shabaz, M. An enhanced secure deep learning algorithm for fraud detection in wireless communication. Wireless Communications and Mobile Computing, 2021(1):6079582, 2021.

Shao, J., Liu, Y., Du, X. and Xie, T. Adaptive Modulation Scheme for Soft-Switching Hybrid FSO/RF Links Based on Machine Learning. In Photonics, MDPI, 11(5):404, 2024.

Song, S., Liu, Y., Xu, T., Liao, S. and Guo, L. Channel prediction for intelligent FSO transmission system. Optics Express, 29(17):27882-27899, 2021.

Trichili, A., Ragheb, A., Briantcev, D., Esmail, M.A., Altamimi, M., Ashry, I., Ooi, B.S., Alshebeili, S. and Alouini, M.S. Retrofitting FSO systems in existing RF infrastructure: A non-zero-sum game technology. IEEE Open Journal of the Communications Society, 2:2597-2615, 2021.

Verdugo, E., Mello, L.D.S., Colvero, C.P. and Nebuloni, R. Estimation of Rain Attenuation in FSO Links based on Visibility Measurements. In 2023 17th European Conference on Antennas and Propagation (EuCAP), 1-5, 2023. IEEE.

Wu, Y., Kong, D., Wang, Q. and Li, G. Performance analysis of UAV-assisted hybrid FSO/RF communication systems under various weather conditions. Sensors, 23(17):7638, 2023.

Yabra, M.S., De Elia, R., Vidal, L. and Nicolini, M. Intercomparison between METAR-and SYNOP-based fog climatologies. Pure and Applied Geophysics, 181(4):1337-1361, 2024.

Yahia, O.B., Erdogan, E., Kurt, G.K., Altunbas, I. and Yanikomeroglu, H. A weather-dependent hybrid RF/FSO satellite communication for improved power efficiency. IEEE Wireless Communications Letters, 11(3):573-577, 2021.