# Data Science Project Proposal

Tongtingna Qi
a1775622

## 1 Introduction

The project aims to deduce the models for Hybrid Optical/Radio Frequency communication channels: one is how weather conditions influence FSO communication channel attenuation, whereas the other is on RF channel attenuation. Furthermore, the relationship between the two models will be demonstrated. Finally, the model will be tested to ensure the predicted attenuation is accurate for diverse channels and weather conditions. The models are essential since they can demonstrate the clear relationship between channels and weather conditions, and the relationship between the two models compares the extent to which weather conditions affect different channels. All models could enhance the communication system performance by complementing the strengths and weaknesses of various channels. The final results will be summarised in this section.

## 2 Background

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]. It is applied to various applications such as 5th generation (5G) and satellites with high reliability and capacity [2]. 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. A further explanation of theories or formulas will be added in this section.

## 3 Methods

### 3.1 Dataset Properties

The dataset is synthesized based on real data from six different cities. There are 27 variables: attenuation of FSO and RF, humidity, the dis-

tance between transmitter and receiver, signal frequency, particulate in the air, rainfall intensity, relative humidity, SYNOP code for weather conditions, temperature, time, visibility, wind direction, and speed. All the measurements will be explained for further research. Data cleaning is the foremost action. Since most parameters are highly related, some will not be included in the final model based on evaluation and statistical results (e.g. P-value) with reasonable explanations. Furthermore, the initial value of the parameter 'TemperatureDiff' will be removed to ensure accuracy. The final dataset will be tidy for reliability and reproducibility.

### 3.2 Machine Learning Techniques

The main programming languages for this project are R, Python, and MATLAB. The relationship between attenuation and different weather variables will be demonstrated as plots. Random forest is the primary method for model construction, and PCA will be applied to get PVE, which will be completed by Week 7. Furthermore, a recipe will be created based on training data for model fitting. Specifically, the 'vip()' function will visualize the importance of parameters for attenuation, and 'roc()' will depict the best model. A clear explanation will be provided for the chosen model by Week 12. Furthermore, the relationship between the two models will be discussed since the effects of different parameters are various on each channel's attenuation. One feasible technique for model comparison is statistical comparison such as MSE, ANOVA, and cross-validation. The detailed research will be done in the next trimester.

## 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] 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.

---

**Notes:** More references will be included in further research. Due to the proposal length restrictions and requirements, the ethics statement, literature review, and detailed timeline will not be included.