Methodology

Satellite data were firstly acquired from the Sentinel-2 MSI: Multispectral Instrument, Level-2A, accessed through Google Earth Engine. Data were loaded and prepossessed using Google’s Earth Engine Code Editor, scripted in JavaScript.

Then, to gather the representative spectral data for each remote sense site, the locations were loaded and a 15-metre radius created around them. The data were selected from a time period of 2022-06-01 to 2022-06-30. The site radii were then averaged and exported as a CSV.

Using iPython and Jupyter notebooks, further analysis was conducted: data preparation, cleaning, imputation, and prediction.

# Data Cleaning

Remote sense data were first converted to numeric datatypes, with invalid datapoints converted to NaN values. The satellite data were then joined by plot id #. The satellite data contain no invalid values.

# Data Imputation

The dataset sourced from the remote sites were in some cases extremely sparce, with attributes containing as many as 90% non-numeric values. For these cases, a K-nearest neighbours data augmentation technique was employed.

# Interpretation

The evaluation for the chosen prediction approach – a generalised additive model – was in the form of standard error. Noting that lower values indicate better performance,