**GreenSight Project – Initial implementation plan (19/09/2024)**

During the meeting we identified two areas for contribution:

1. **Carbon Task:** Finding associations between carbon estimated from JULES and existing indices from EO data.
2. **Fires Task:** Mapping fires occurred at Green Belts over the last 20 years – comparing fine tuning of foundation model with traditional burned related indices (e.g., Burnt Area Index).

The aim is to include some of produce information in MetOffice dashboard. The design of the dashboard will take some time but both tasks need to be completed before the end of March 2025.

**Carbon Task:**

JULES provides carbon estimations from 1980-2024. We will focus on the last 10 years that Sentinel-1 (C band SAR) and Sentinel-2 (multi-spectral) data are available at 10m resolution. Sentinel-1 and Sentinel-2 compliments each other.

JULES provides estimates at 5,000m squares. Initially we will start with time-series extracted for each green belt polygon showing the averaged vegetation and soli carbon, as well as the average indices values across the green belt. We need to consider that each greenbelt polygon contains different types of Land (e.g., forested, grass, agriculture).

We need the following data from MetOffice:

* Vegetation carbon and Soil carbon time-series estimated from JULES for each greenbelt polygon
* Shapefiles of Greenbelts (<https://www.data.gov.uk/dataset/ccb505e0-67a8-4ace-b294-19a3cbff4861/english-local-authority-green-belt-dataset>)
* Any shapefiles available related to the landcover classification (<https://catalogue.ceh.ac.uk/documents/a413d1c9-5c2a-4864-a3ff-2b9f764ec32d>)

Note: We need to agree to a specific "subset" of the dataset (greenbelt(s) and time-period) to validate our approaches.

Associations that maybe identified:

* Local maxima in the time series, occurring between specific indices, may happen at the same time of the year as the annual local maxima in vegetation or soil carbon time-series.

We aim to create flexible implementations so that, in the future, we can scale the association identification and extraction of time-series to cover the entire UK. For that reason, we will add on Milto’s open-source tool PlotToSat (<https://github.com/Art-n-MathS/PlotToSat> - related paper under review but can be shared if needed to understand pre-processing steps). PlotToSat currently takes as input a CSV file with multiple locations associated with a given radius and a co-ordinate system. It creates a list of circular polygons and for each polygon, it extracts a time-series for a year of Sentinel-1 and/or Sentinel-2 data.

Tasks to be completed:

1. Adjust PlotToSat so that it takes as input a shapefile with multiple polygons and extract time-series of any given polygon. Then we can create shapefiles for e.g., a polygon for each JULES pixel and/or use the shapefiles of the Greenbelts. These shapefiles will be used for extracting time-series of Sentinel-1 and Sentinel-2 over the last ten years in .csv files.
2. Extract various indices from the exported .csv files. List of Vegetation, Soil and Water indices available here: <https://step.esa.int/main/wp-content/help/versions/10.0.0/snap-toolboxes/eu.esa.opt.opttbx.radiometric.indices.ui/OperatorsIndexList.html>
3. Correlate and compare those indices for finding associations between the soil and vegetation carbon time-series of JULES

Allocated tasks:

* Milto 1, 2
* Finlay 2, 3

Note: Finlay is on holidays. We aim to have the shapefile input to PlotToSat implemented before Finlay starts working. As a contingency plan, he can start working with approximated time-series created using locations of interest and a drawn circle around them (what PlotToSat currently supports).

Additional work: Investigate the possibility of adding more datasets to PloToSat e.g., Landsat for longer time-series investigations and/or thermal data (<https://github.com/sofiaermida/Landsat_SMW_LST>).

**Fires Task:**

We need an efficient approach for identifying burned areas in UK green belts, and presenting their statistics, e.g. location, size distribution, number per year for each green belt. => the result may be a datacube or yearly raster images showing the annual burnt areas within greenbelts each year and/or a shapefile that will contain the burnt areas with information in their attribute table about the year the fire occurred and burnt area size. This will help derive statistics about number of fires per year, size of areas burnt each year and create descriptive graphs from the information derived.

Remy has fine-tuned the Foundation Model of IBM to detect burnt marks. We want to investigate the correlation between Remy’s approach and traditional Remote Sensing approaches using Burnt related indices (e.g., Normalized Burn Ratio – NBR and Burn Area Index). Since we do not have ground data, if the two methods produce similar results, there will be a higher confidence in the maps of burnt areas created.

Google Earth Engine (GEE) is free for non-commercial purposes (Specific rules for free usage available here: <https://earthengine.google.com/noncommercial/>). A GEE account though needs to be connected to Google Cloud, so it does ask for credit card but as long as the projects created are listed as non-commercial/educational purposes, there is no charge. These tutorials are a good starting point for understanding how GEE works (<https://www.youtube.com/@noelgorelick3818>). PlotToSat has Sentinel-1 and Sentinel-2 classes that could be a good starting point for using the Python API of GEE.

Tasks:

1. Find a representative set of the Indices related to Burn Areas. It also worth writing a paragraph as a literature review of what each index does. This paragraph can be used in the paper later.
2. Implement to indices related approach (e.g., thresholding and some filtering)
3. Compare the results of the two approaches

Allocated tasks:

* Remy 1, 2, 3
* Milto 1, 2 (possibly creating a Landsat class)

Additional Potential Contributions:

* Hensel forest database (<http://earthenginepartners.appspot.com/science-2013-global-forest>) contains maps of global forest fires over the last 20 years. It can be used as additional training data for the fine tuning of the foundation model. Also <https://forest-fire.emergency.copernicus.eu/>. We need to be considerable of uncertainties as Green Belts contain multiple types of lands.
* Looking into associations with proximity to National Parks, population, droughts (daily precipitation estimates available at BOKU Climate data - <https://boku.ac.at/en/wabo/waldbau/wir-ueber-uns/daten>) and heatwaves (thermal data available on Landsat 30m and Sentinel-3). MetOffice may have in-situ temperature and precipitation measurements.

**Time Allocation:**

* Hywel (PI) 5%
* Milto (Co-PI) 30%
* Remy 50%
* Finlay 40%