BIG data archetypes for crops from EO

Background

> *As population increases, so does the demand for food and food security. Field crops supply a large proportion of this globally, but pressures on crop area and productivity continue to grow. Earth Observation (EO) already provides much information, and new datasets offer more quality and detail for crop development and yield prediction at the field-scale and beyond. We use algorithms to interpret the data, but they are a complex response to a large number of soil, crop, and atmospheric conditions that must be disentangled to get at the crop information. We will test a new technique to map crop condition from EO for winter wheat in South Africa. We use physics models of the EO to describe their response to biophysical parameters, then machine learning to calculate the parameters from the measurements. But many parameters are uncertain when estimated like this. We use big data analysis globally to build typical models of the parameters that we call archetypes. We use these to provide robust estimates of all parameters for the entire growth cycle. But this science needs testing and application to areas of need, which is what we will do here in a collaboration between UK and South African scientists.*

Field data collection

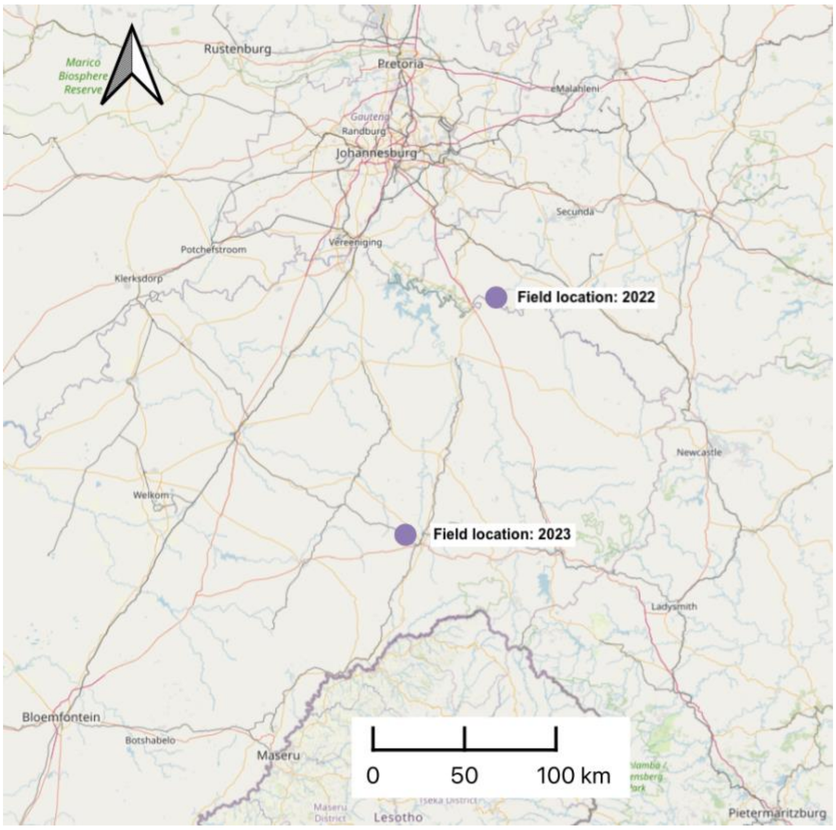

Figure 1 Locations for field data collected in 2022 and 2023.

In this study, researchers from South Africa conducted field measurements to collect biophysical data for the years 2022 and 2023. The 2022 data collection took place in the Dipaleseng Local Municipality, South Africa, at coordinates longitude: -26.963854 and latitude: 28.744026. For the 2023 data collection, the research was conducted at Syngenta's Marné Research Farm, located on Kaallaagte Road in Meets, Bethlehem, South Africa, at unspecified coordinates. The specific locations used for data collection are illustrated in Figure 1. In both instances, rainfed wheat was sown at the beginning of September. The frequency of measurements varied between the two years; in 2022, measurements were taken weekly, whereas in 2023, they were recorded every 15 days. This change in the measurement interval was due to a modification in funding arrangements with UCL. The harvesting period for the crops spanned from the end of November to early December. The Leaf Area Index (LAI) and leaf chlorophyll content were the primary biophysical parameters measured, utilizing xxx and xxx.

Data collection strategies

In each field, the researchers designated five plots for measurements, each spanning an area of 50x50 meters. Within these plots, five subplots, each covering an area of 10x10 meters, were identified to collect representative data for the larger plot. To ensure accurate representation of conditions within each subplot, five separate measurements were taken and subsequently averaged. This approach provided a detailed and representative dataset of the conditions within each plot. An illustration of this sampling design is provided in Figure 2 to visually depict the structured approach to data collection. This systematic method ensures a comprehensive and accurate representation of the biophysical parameters for the field.

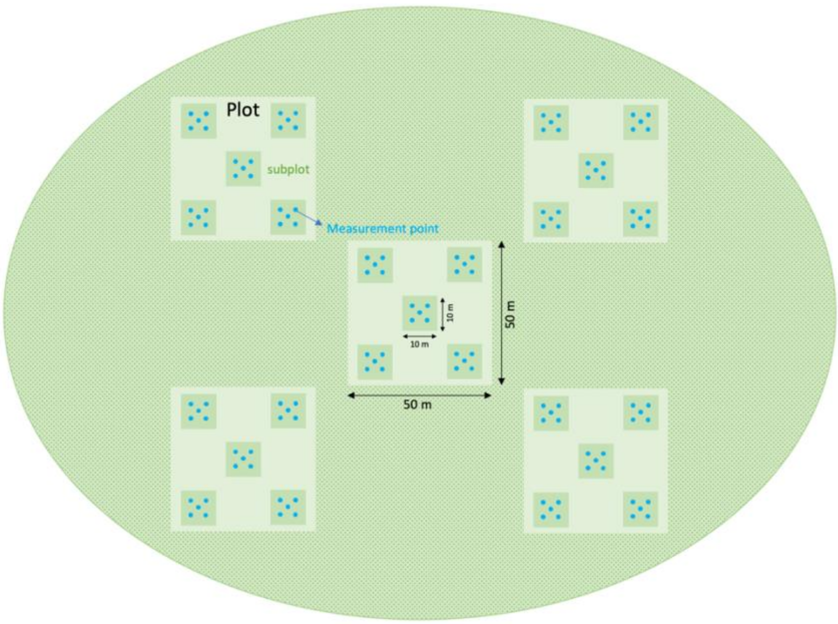

Figure 2 Field sample design illustration. 5 plots are placed within 1 field, with 5 subplots within each plot. 5 measurements are taken within each subplot to provide the averaged values for each subplot.

LAI measurements protocol

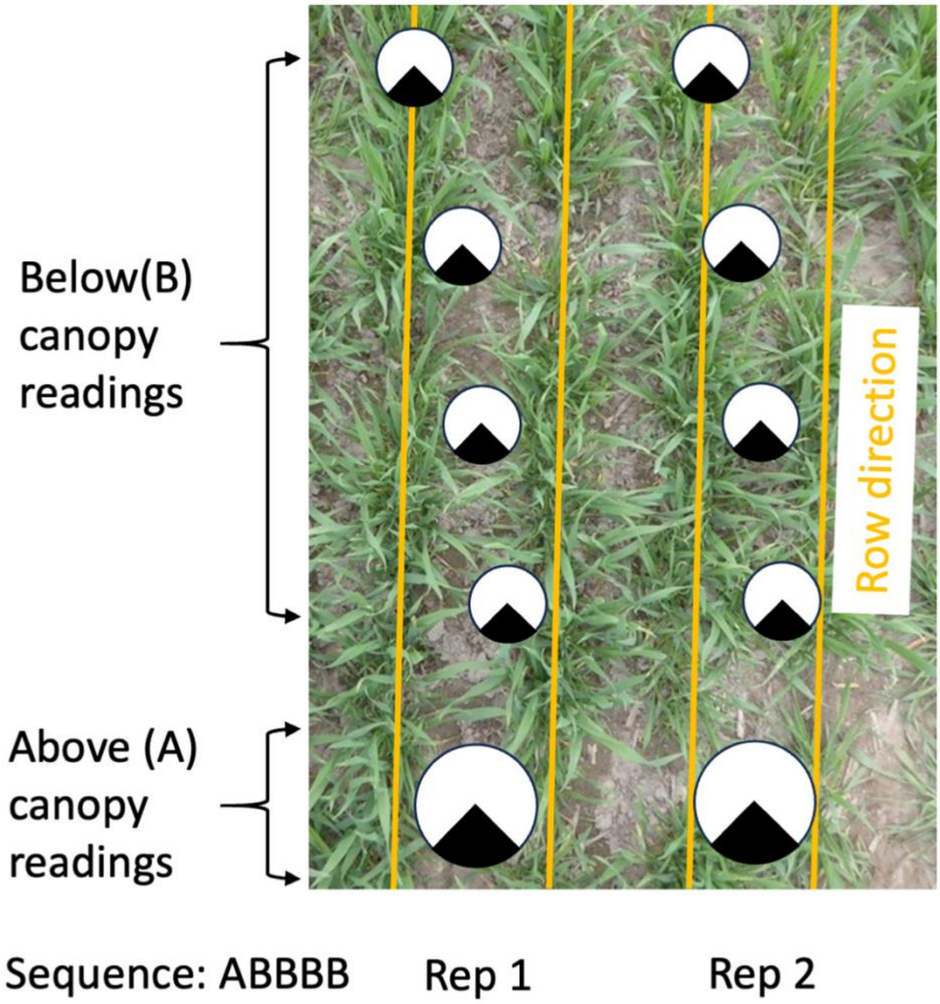

Figure 3. LAI measurement design, adapted from Fang et. al. 2012.

1. Select five distinct locations within each field for conducting LAI measurements. These locations should represent the variability in the density of the plant population within the field and be positioned at least 10 meters away from the edges of the field to avoid edge effects.
2. Determine and mark the specific rows for measurement within each subplot. Record the GPS coordinates for each marked area for precise relocation during subsequent measurements.
3. At every marked location, perform one measurement above the plant canopy and then four measurements beneath the canopy, as depicted in Figure 3. This approach ensures comprehensive coverage of the canopy's light interception capability.
4. Maintain a consistent orientation for the view cap during all measurements at each location. Utilize a 270° view cap to prevent the sensor from detecting the operator, ensuring accurate readings.
5. Calculate the mean of all measurements taken across the sampling points to derive the LAI value for each subplot.

Key Considerations:

- Increasing the number of readings taken below the canopy enhances the spatial representation of the measurements.
- For canopies that are 1 meter in height, position the optical sensor at least 3 meters away from the canopy edge to avoid interference from adjacent areas.
- Ensure all below-canopy measurements are conducted at a uniform height of 5 cm above the soil surface and aligned in the same direction as the above-canopy measurement to guarantee consistency.

Chlorophyll Measurement Protocol:

1. Within each marked area, identify five plants for ongoing chlorophyll measurements.
2. Label the fifth and sixth leaves on each selected plant for targeted measurements.
3. On each tagged leaf, conduct measurements at three points: one-quarter, one-half, and three-quarters of the distance from the base of the leaf to its tip.
4. Execute measurements on both tagged leaves, resulting in a total of six measurements per plant, to ensure a thorough assessment.
5. Use the “Average” function to compute the mean SPAD reading for each plant, providing a representative value of its chlorophyll content.

The SPAD readings are collected from plants located in the same five sampling points selected for the LAI measurements to offer a comprehensive analysis of the chlorophyll content across different parts of the field.

Satellite data

The study uses surface reflectance data from the Sentinel-2 satellite, specifically the Level 2A product, accessible via Google Earth Engine for the years 2022 and 2023. This data covers the fields where measurements were conducted and is used in estimating biophysical parameters. The Sentinel Hub's s2cloudless algorithm is applied to eliminate pixels affected by clouds.

Given the absence of official uncertainty metrics within the Sentinel-2 surface reflectance product, a nominal uncertainty of 10% is adopted for the analysis. This assumed uncertainty is incorporated into the process of retrieving biophysical parameters, serving as a conservative estimate to account for potential variations in satellite data accuracy and the official atmospheric correction method.

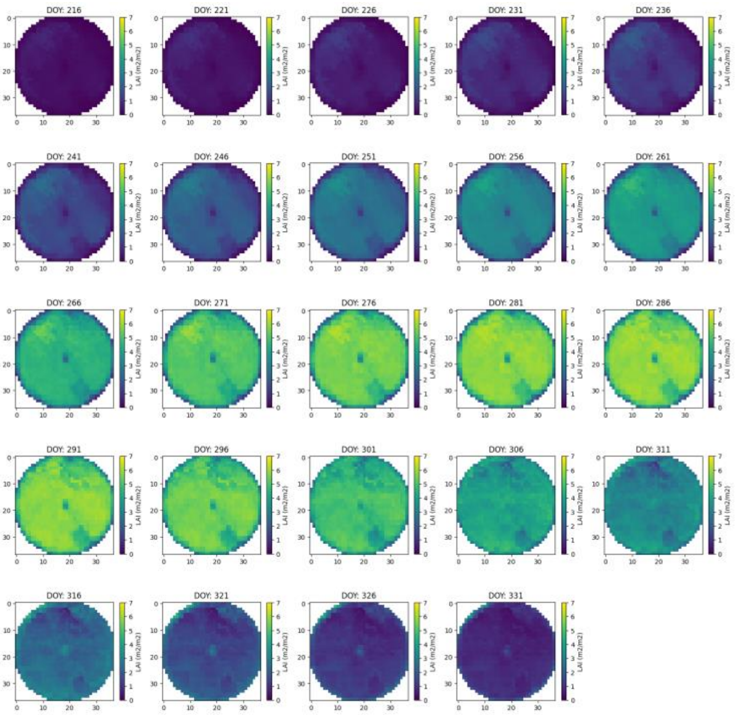

Figure 4 Time series of LAI maps retrieved with the UCL ARC algorithm over the 2022 field.

UCL ARC algorithm

UCL team has developed a dynamic model for simulating biophysical parameters, such as Leaf Area Index (LAI) and leaf chlorophyll content, as well as surface reflectance time series, building on their previous work (Yin et al., 2024). This model is used in the ARC algorithm, where a Monte Carlo solver is proposed to solve time series of PROSAIL biophysical parameters from time series of satellite observations. A corresponding Python package has been made available on GitHub: https://github.com/MarcYin/ARC. Using this package, we can retrieve time series data of LAI and chlorophyll content from Sentinel-2 surface reflectance imagery. The package is installed on a JupyterLab instance hosted by the EO AFRICA R&D Facility Innovation Lab, leveraging the facility's computational resources for the retrieval of biophysical parameters from Sentinel-2 data. An illustration of this application is the generation of LAI maps over time, exemplified in Figure 4.

Results

The study involves a comparative analysis of satellite-derived and ground-measured Leaf Area Index (LAI) and leaf chlorophyll content, focusing on plot-level assessments. For both LAI and chlorophyll content, the mean and standard deviations calculated from measurements across five subplots represent the values and uncertainties for each plot, respectively. The findings, illustrated in **Error! Reference source not found.** and Figure 6, reveal a close agreement between satellite estimations and ground observations.

For LAI, the comparison highlights a robust match between satellite retrieval and ground measurements, showcasing the effectiveness of satellite-derived estimations in capturing LAI variations over time. Similarly, the comparison of leaf chlorophyll content demonstrates a strong alignment between them, indicating the satellite's capability to accurately estimate chlorophyll levels. However, discrepancies were noted towards the end of the growth season, where satellite estimations tend to overestimate chlorophyll content. This deviation is likely due to the satellite sensors primarily capturing the canopy's top layer, where leaves typically exhibit higher chlorophyll concentrations, even in later growth stages, compared to those at the canopy's mid or lower levels.

Figure 5 Comparison of LAI retrieval from Sentinel-2 using ARC algorithm and ground measurements over the 5 plots for the field in 2022.

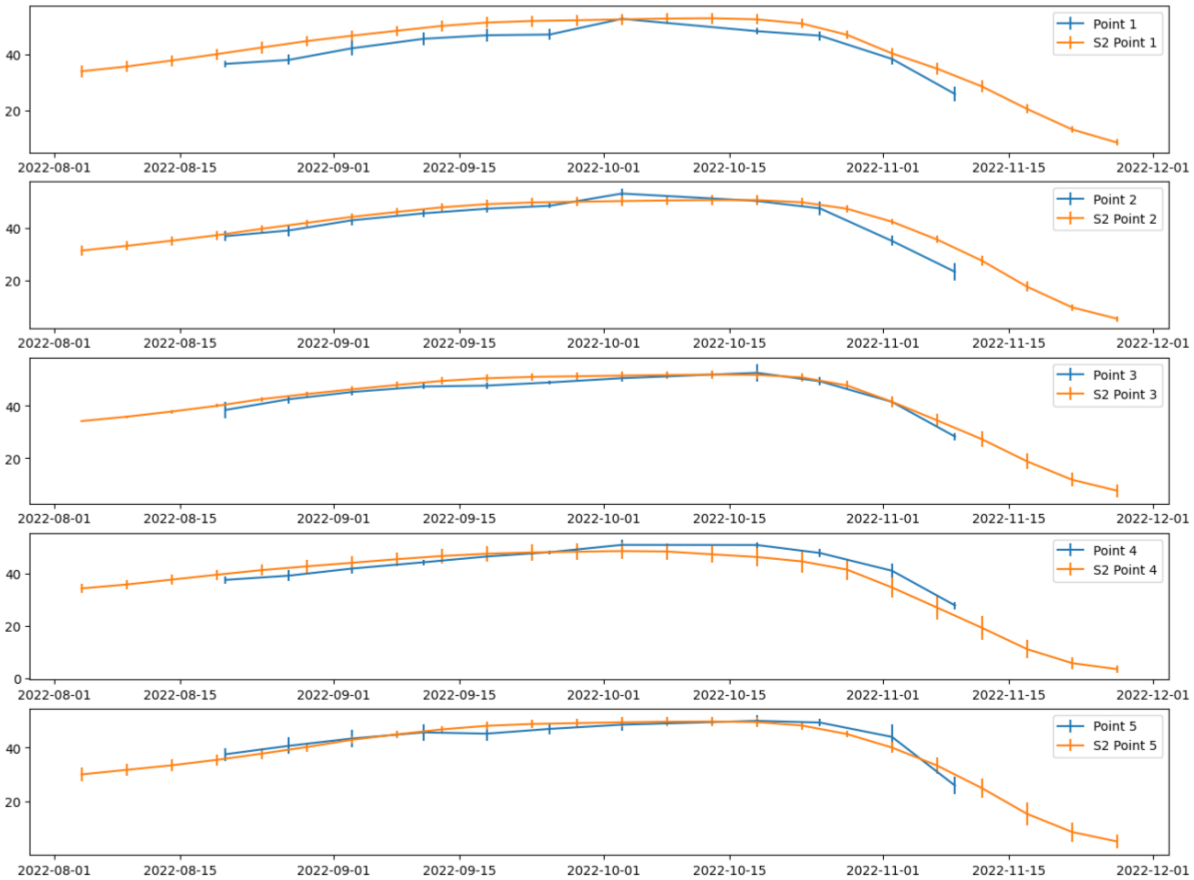

Figure 6 Comparison of leaf chlorophyll content retrieval from Sentinel-2 using ARC algorithm and ground measurements over the 5 plots for the field in 2022.