ARMYWORM
(Armyworm Research in Malawi Yielding Workable and Original Remote-Sensing Methods)

Research summary: Can the application of machine learning algorithms to remote sensing data successfully detect Fall Armyworm outbreaks in Malawi?

Figure 1: A mature Fall Armyworm caterpillar feeding on maize

Figure 2: Fall Armyworm damage to a maize cob

Figure 3: Maize showing leaf damage and leaf colour typical of FAW attack.

Summary

Supported by the World Bank's Trust Fund for Statistical Capacity Building (TFSCB) ProvEye Ltd (formerly Orbas Consulting), an Irish company specialising in the application of AI to remote sensing, and Gorta-Self Help Africa (SHA), an NGO working in Africa, have succeeded in developing a machine learning algorithm that can detect the damage caused by the Fall Armyworm (*Spodoptera frugiperda*) to smallholder maize fields in Southern Africa in satellite images. The Fall Armyworm, the caterpillar stage of a moth, is a major pest in North and South America which crossed the Atlantic to West Africa in 2016 and rapidly spread across the continent. The Fall Armyworm feeds on over 80 plant species but prefers maize, which is the staple food crop for much of East and Southern Africa, causing up to 30% crop losses. The software can detect the presence or absence of FAW damage with up to 87% accuracy and the level of FAW damage with 63-75% accuracy, depending on the data available. The success of this "proof of concept" research provides the foundation for the development of remote sensing software that can identify FAW damage hot spots at the landscape level, enabling governments to target scarce control resources on these hotspots. The success also indicates that this approach can be used to detect other crop pests and diseases that change the colour of the crop canopy.

Problem statement

In August 2016 the Fall Armyworm moth (*Spodoptera frugiperda*, FAW) crossed the Atlantic to West Africa and rapidly spread across Africa. The results were dramatic, with maize leaves torn to shreds by the caterpillars and the cobs destroyed. The damage to maize fields is very distinctive which suggested the research question: if infested fields can be readily identified by the naked eye can infested fields be detected from satellite images?

Researchers had successfully identified FAW damage in large homogenous maize fields in North America from satellite images, but the challenge is to build software that can analyse satellite images to spot FAW hotspots in small and highly heterogeneous fields in Southern Africa. Machine Learning algorithms are now widely used in remote sensing so SHA partnered with Dr Jerome O'Connell, AI remote sensing specialist at ProvEye Ltd, and with the Malawi Ministry of Agriculture and Planet Labs to test the use of machine learning algorithms to "learn" to spot FAW damaged maize in satellite photos. ProvEye Ltd had already built machine learning algorithms and image processing tools for remote sensing to detect land use changes and so were ideal partners for the project.

The ultimate aim is to build software that can rapidly detect FAW hotspots from remote sensing data to enable Governments, UN agencies, NGOs to focus limited resources on the hotspots. The software may even be accurate enough to guide precision agriculture equipment used by large scale farmers. This research, however, was a proof of concept project: could FAW be detected through remote sensing and which part of the spectrum gives the strongest correlations with FAW damage?

What was involved

The theory behind the approach is to look for correlations between the reflectance of various bands of the light spectrum of FAW infested fields in satellite images infestations (figures 4 and 5) and the level of FAW measured by field surveys. Part of the data was used to detect these correlations, which form the basis of the algorithm. The remaining data is used to test the algorithm: the algorithm is given satellite images of sites with known levels of FAW damage and asked to calculate the level of damage. The advantage of a machine learning algorithm is that it learns from its mistakes and the accuracy of the predictions improves with each cycle. Many factors can affect the reflectance of a maize field: the maize variety, age of the crop, crop canopy cover, colour of the soil, the height of the crop, nutrient deficiencies and water stress, so these variables also need to collected and tested.

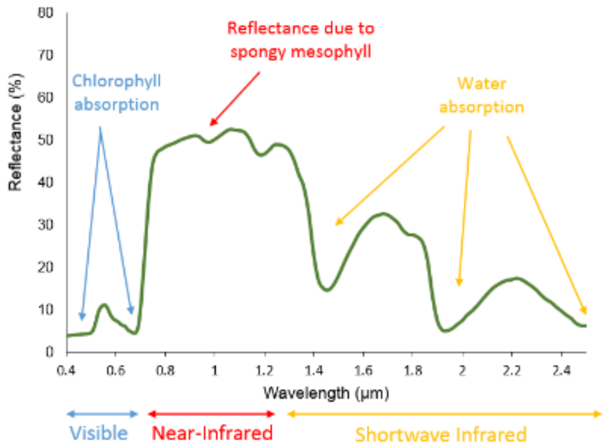

Figure 4: How plants absorb light

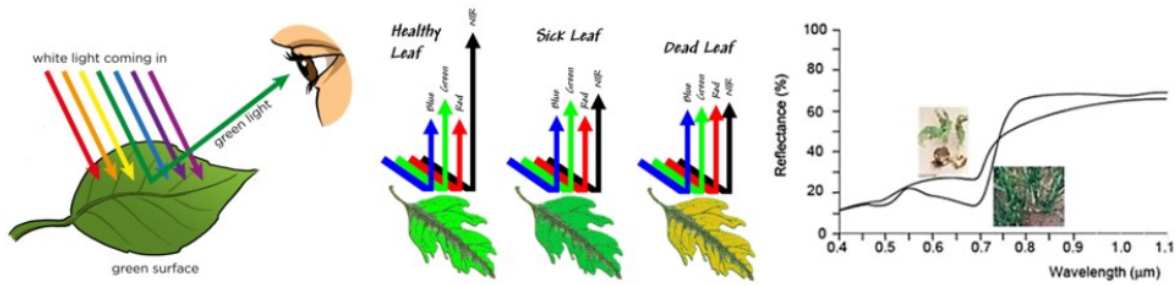

Figure 5: How the colour of a leaf changes when a plant is damaged.

To develop the algorithm the team needed training and testing data from the field. SHA had worked for several years in Balaka District, Malawi, an area heavily infested with FAW, so maize fields in Balaka District were selected as sentinel sites (figure 6). Each field was surveyed, and two 15x15 m quadrats were set up in each field. Teams of surveyors, initially from SHA's own staff but later from the District Agriculture Office, conducted three rounds of field surveys per growing season (December to April). The fields and quadrates were geolocated, first with the iPads used to collect the survey data. This proved to be too inaccurate, so the sites were resurveyed with hand-held Garmin GNSS, which improved the accuracy to 2-3m, still a significant error on a 15m x 15m quadrat, and finally with Emlid Reach RS+ GNSS RTK units with 30cm accuracy. The field team recorded the details of the fields (soil type, aspect, previous crops, planting date, variety) and then sampled the quadrates. 15 plants were sampled per quadrat and the level of FAW damage on each plant was assessed on a 0-5 scale.

0: No visible damage and no visible presence of either eggs or caterpillars.
1: No visible damage, eggs present.
2: Shot holes and elongated lesions (>2 cm) on <50% of leaves.
3: Elongated lesions on >50% of leaves.

4: Elongated lesions or tattering on most leaves (>75%).
5: Plant dead, dying (dead heart) or economically worthless (cob damage).

A large amount of data was collected from each site to capture all the factors that could affect reflectance (maize variety, growth stage, soil type, plant population, canopy height, canopy cover, other pests and diseases, water stress, nutrient deficiencies and nutrient applications) and FAW damage levels (crop rotations, previous FAW infestations, neighbouring crops).

3.7m resolution, 4 band, satellite images, were sourced from Planet Labs. There was very limited satellite coverage of the area, which made it challenging to obtain images taken at the time of each survey round and the small size of the fields (<0.5ha) meant that that there were only 6-9 pixels per field.

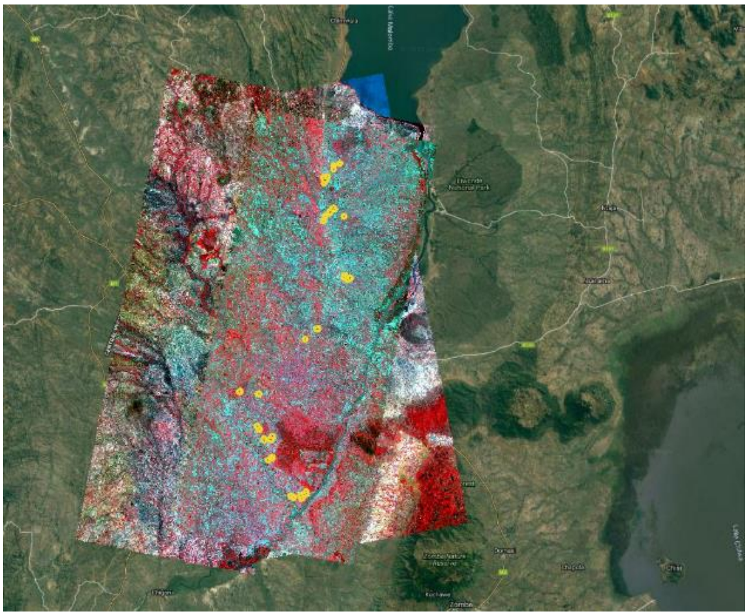

Figure 6: Planet Labs images of Balaka District, Malawi, used for analysis of 2018-2019 season. Yellow dots represent the location of the sentinel fields and quadrats.

Results

Of the variables tested by far the most important variable for detecting FAW damage is the **standard deviation of Near Infra-Red** (NIR) reflectance, distantly followed by the **Green Normalised Difference Vegetation Index** (GNDVI). This was as expected as discolouration of the green leaves of the maize plant, due to the reduction in chlorophyll levels, is one of the first indicators of FAW infestation and the NIR band is sensitive to the biochemical composition of vegetation, particularly chlorophyll, however it was surprising that other commonly used indices for plant health, like NDVI, did not rank higher.

It became clear that the 5 point maize damage scale, which was based on the scale used by maize breeders, was not very practical for remote sensing as it included factors, like the presence of eggs, that are invisible to satellites. To overcome this problem the damage levels were merged into 2 combinations: presence and absence (binary), and a simplified damage scale (0 and 1 merged). A second problem was that machine learning algorithms learn best when they have a more or less even number of records for each damage level, for example a 50:50 split of presence: absence records. For an insect like the Fall Armyworm the data will naturally be skewed. At the beginning of the season most of the plants sampled will have little or no FAW damage, by mid-season there will be a more even spread of damage levels, while by the end of season most of the plants will show high levels of damage. To overcome this problem the data had to be resampled to create even data sets.

Machine Learning algorithms learn from their mistakes and after over 1,000 cycles the ProvEye Random Forest-based algorithm can now detect the presence or absence of FAW damage with up to 87% accuracy. Using the simplified damage scale the level of damage can be predicted with 63-69% accuracy based on remote sensing data only, however accuracy of 75% can be achieved when field data is included. The critical field variable is **crop growth stage**. At the v2-v3 stage there is considerable reflectance directly from the soil, and from the soil through the leaf, which creates "noise". If the algorithm knows the crop growth stage it can compensate for this noise.

The first stage of this research used maize fields identified by the survey teams. To be usable by government agronomists the software needs to be able to detect maize fields in satellite images so the next stage of the research is to use a "Crop Mask", a piece of software that can detect the crop automatically.

The success of the this "proof of concept" research lays the foundations for developing remote sensing software that can identify FAW damage hot spots at the landscape level, enabling governments to target scarce control resources on these hotspots and estimate crop losses. The success also indicates that this approach can be used to detect other crop pests and diseases that change the colour of the crop canopy, like banana bunchy top virus. It is important to note that this approach only detects crop damage caused by the FAW and cannot make any predictions about where FAW outbreaks may occur.

National and International networking

The Ministry of Agriculture, Irrigation and Water Development in Malawi fully supported the project, allocating staff to collect field data, and the project reported to the Malawi National Fall Armyworm Task Force. The project team in Malawi have close links with the EU, FAO, World Bank and Irish Aid delegations in Malawi, who have followed the progress of the research with interest and suggested new applications for the software.

This project, ARMYWORM (Armyworm Research in Malawi Yielding Workable and Original Remote-Sensing Methods), submitted in response to the 2017 call for proposals by the World Bank's Development Data Group (DECDG) and the Global Partnership for Sustainable Development Data (GPSDD), was supported by the World Bank's Trust Fund for Statistical Capacity Building (TFSCB) with financing from the United Kingdom's Department for International Development (DFID), the Government of Korea, and the Department of Foreign Affairs and Trade of Ireland.