Lessons Learnt report on the use of advanced machine learning algorithms, in conjunction with multi-temporal satellite imagery, in detecting FAW outbreaks in Malawi

Project	**MH5: ARMYWORM - Armyworm Research in Malawi Yielding Workable and Original Remote-sensing Methods (Malawi)**
Location:	Balaka District, Malawi
Associated Project(s)	MH4 BLF
Budget:	USD 29,717.40
Budget source:	PO#7186834. World Bank Innovation Fund "Collaborative Data Innovations for Sustainable Development" Competition.
Start Date:	March 2018
Completion Date:	March 2020
SHA Lead(s):	Paul Wagstaff Peter Soko Thokozani Sentala Raphaële Ng Tock Mine
Lead Investigator(s):	Dr. Jerome O'Connell Professor Nicholas M. Holden
Collaborating Institutions:	Orbas CRT Ltd, UCD School of Biosystems and Food Engineering
Data types	Microdata & Geospatial
Project Objectives	a. develop a new machine-learning algorithm that integrates multi-platform, multi-temporal high-resolution satellite images and meteorological data to build a robust and transferrable model that can be easily updated with additional data when available, b. utilize a research approach that will complement concurrent efforts to tackle FAW and explore potential knowledge gaps.

This project, ARMYWORM - Armyworm Research in Malawi Yielding Workable and Original Remote-sensing Methods (Malawi), 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), is 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.

Executive Summary

This report summarises the key lessons learnt from the **ARMYWORM - Armyworm Research in Malawi Yielding Workable and Original Remote-sensing Methods (Malawi)** funded by the World Bank Innovation Fund "**Collaborative Data Innovations for Sustainable Development**" Competition. The project developed an AI algorithm to detect the damage caused by Fall Armyworm in maize fields in Balaka District Malawi. Full details of the project can be found in the final report.

The lessons learnt from this project have been applied in the extension of the Fall Armyworm (*Spodoptera frugiperda*) research project funded by the US Foundation for Food and Agriculture Research (FFAR)[1] and in recent proposals to adapt the software to detect damage caused by Desert Locusts *(Schistocerca gregaria*) and Banana Bunchy Top Virus.

Data collection

The iPads used for data collection were not ideal for field use in the tropics as they quickly overheat. The field data needed extensive cleaning and the time required for data cleaning was underestimated in the budget, requiring additional funds from SHA to employ a student to clean the data.

Collecting GPS coordinates

Given the small size and high levels of heterogeneity of fields in Malawi small errors in the coordinates invalidate the ground data. The project has tried various options to improve accuracy and is currently testing Emlid Reach RS+ RTK GNSS units with 30cm resolution.

Availability of Images and Image resolution.

For the research methodology to work good quality satellite images are required, taken as close as possible to the date of the field surveys. The area of Malawi chosen for the project had very limited satellite coverage and some rounds of field survey data could not be used as there were no corresponding satellite images. This will limit the usefulness of the algorithm in areas with poor satellite coverage.

Larger quadrats (15m x15m) improve the accuracy of the analysis as there are more pixels per quadrat but occupy too much of the farmer's small fields and the project had to settle for 10m x 10m quadrats. With 3.7m resolution (Planet) images the project had only 9 full pixels in a 10m x 10m quadrat.

Damage Scale

The project used a scale of 0-5, which proved to be unnecessarily complicated as the differences between points on the scale were not apparent in satellite imagery. The scale has been revised to a 3-level scale.

Data distribution.

The learning datasets for Random Forest Algorithm require data that is evenly distributed across all variables. As the FAW populations change as the maize develops the survey data is skewed, with implications for the various modelling scenarios which was not recognised when the surveys were designed. To overcome this problem the data had to be resampled to create balanced learning datasets, reducing the amount of data available for analysis. Skewed data is a fundamental factor in the population dynamics of many pest and disease infestations and may limit the usefulness of current AI tools in analysing fluctuations in pest and disease infestations.

High level of Variability across fields and quadrats

The project averaged the FAW damage scores and the reflectance across quadrants and fields. Averaging FAW damage scores and the colour of pixels (reflectance) per quadrat is questionable due to the highly uneven distribution of FAW damage in fields and quadrates.

Spectral Reflectance

The project found that the blue part of the spectrum had the highest mean importance score for detecting FAW.

[1] FFAR and Gates Foundation Help Farms Combat Pests, Disease

Lessons Learnt

Data Collection

Ground truth data was collected using iPads, with Filemaker Go software integrated into Salesforce software. The iPads overheated in the midday sun and took time to cool down enough to continue the surveys.

The original plan was to take overhead photos of each quadrat using the iPads on a selfie-stick, but the i-Pads also proved to be too heavy to take overhead photos of the quadrats using the selfie-sticks procured for the project.

The data needed extensive cleaning. The time required for data cleaning was underestimated in the budget and additional funds from SHA were used to employ a student to clean the data.

The survey methodology was revised for the 2018-19 season. As SHA no longer had staff permanently based in Balaka District Government Agriculture Extension staff were selected and trained as FAW surveyors.

Collecting GPS coordinates

Given the small size and high levels of heterogeneity of fields in Malawi small errors in the coordinates invalidate the data. The field surveys were conducted using iPads and the first survey round also used the iPads for collecting the GPS coordinates. The internal GPS system on the iPads proved to be far too inaccurate for our requirements (>5m error).

The 2017-18 shapefiles showed considerable discrepancies with between the GPS coordinates of the field boundaries, quadrate and soil sample sites and the satellite imagery, with some soil sample sites falling outside field boundaries. Cleaning the data and removing these discrepancies reduced the number of sites to 24 fields, a very limited dataset in terms of number of observations for a machine learning algorithm like Random Forest.

The sentinel fields and quadrats were resurveyed using hand-held Garmin GNSS units. This reduced the error to <3m but this was still a significant error for the 10m x 10m quadrates when used to geolocate the 3m resolution Planet imagery. With the 3m resolution of the Planet images each quadrat was represented by only 9 pixels so a 2m error significantly affected the results.

To further improve accuracy the project conducted a market survey of high precision GNSS units. Three basic solutions are available for increasing the accuracy of GNSS equipment:

- Correction via a satellite correction service
- Correction via the internet (NTRIP)
- Differential GNSS/ RTK

None of the suppliers had tested their equipment in Malawi and were unable to provide a guarantee of access to satellite correction services in Malawi. No internet-based correction services are currently available in Malawi (a commercial company has the equipment to provide a service but has not yet received a government licence). The project therefore decided to test Emlid Reach RS+ GNSS receivers using RTK correction. Though the cheapest option the Emlids have proved easy to use and highly accurate. The Emlid Reach RS+ units were used in a Base station + Rover configuration, in combination with a smartphone running the ReachView app. These provide sub-meter accuracy, within 30cm in most cases.

Availability of Imagery and Image Resolution

For the research methodology to work good quality satellite images were required, taken as close as possible to the date of the field surveys.

A time series of Planet Labs PlanetScope 3m resolution satellite imagery[2] were downloaded from the Planet Labs database. The data was provided free through the **Planet Application Program Interface: In Space for Life on Earth**. The total number of images processed was 4, with software written to automatically process (geometric, radiometric and spectral) the data in a fully automated workflow so that the outputs could be directly inputted into an AI model.

The area of Malawi chosen for the project had very limited satellite coverage and some rounds of field survey data could not be used as there were no corresponding satellite images. Ironically, had the project been run in the drought and flood prone south of Malawi far more satellite images would have been available. This will limit the usefulness of the algorithm in areas with poor satellite coverage (typically those areas of limited economic or geopolitical interest).

The project selected 60 maize fields. Larger quadrats would have improved the accuracy and statistical rigour of the analysis by providing more pixels per quadrat but 15m x15m quadrats occupied too much of the farmer's small fields and the project had to settle for only two 10m x 10m quadrats per field. With 3.7m resolution (Planet) images the project had only 9 full pixels in a 10m x 10m quadrat. Any inaccuracies the coordinates, or corrupted pixels, greatly reduced the statistical power of the analysis.

Damage Scale

The project used a scale of 0-5, based on scales used by plant breeders testing for FAW resistance in maize varieties:

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).

With hindsight it was obvious that this scale was unsuitable for a remote sensing project. Level 1 damage (no visible damage but eggs present) was identical with Level 0 from a remote sensing perspective, as is cob damage. The descriptions for level 2 (" Shot holes and elongated lesions") and 3 ("Elongated lesions on >50pc of leaves") of FAW infestation are very similar, therefore confusion between these in the model based on satellite imagery at 3.5m using training data from a 10x10m quadrat was not surprising.

For future research a simpler 3-level scale will be used: No leaf damage; ≤30% leaf damage; > 30% leaf damage.

Data distribution.

The learning datasets for Random Forest algorithm require data that is evenly distributed across all variables and it has been shown that highly unbalanced training data like the FAW dataset can bias the Random Forest model (Chen, Liaw, & Breiman, n.d.). For Random Forest to work best the number of quadrates with an average score in each category (0-5) should be more or less even in each survey round.

Distribution for the various FAW levels across the 3 surveys in the 2018-19 season varied significantly with only the 3rd survey (February 2019) giving observations across 4 of the 5 FAW levels (figure 2). For the December survey over 80% of the training data was in Level 1 (No visible damage, eggs present), with the remainder at Level 2 (Shot holes and elongated lesions). For the January survey 100% of the training data was in Level 2.

When the first round (December) surveys are conducted (V5, 5-leaf stage) there is very little visible FAW damage as the FAW population is low, those insects present are still very small and are confined to a few plants.

---

[2] Planet Team (2019). Planet Application Program Interface: In Space for Life on Earth. San Francisco, CA. https://api.planet.com

The damage scores are therefore skewed to the left. In the second round surveys (February, V10) the caterpillars are maturing and have started spreading to more plants in the quadrates. By the start of the reproductive phase (R0) almost all plants are infested, and a wide range of damage scores recorded.

To overcome this problem a threshold was set within the software which automatically assesses the distribution of the training data (Figure 1) and either up-samples the minority class or down-samples the majority class, depending on the distribution, to create a binary distribution (present/ absent).

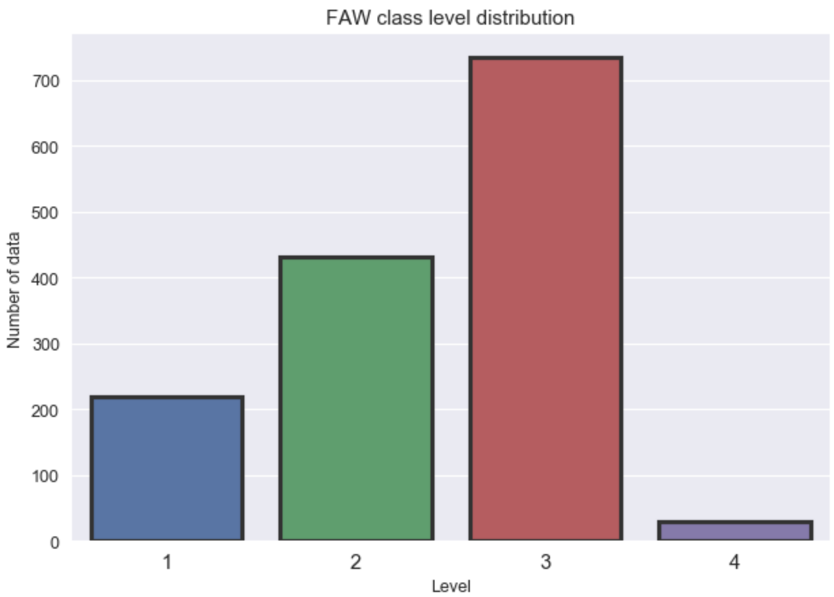

Figure 1: Distribution of FAW levels for February 2019 survey.

- A binary presence of FAW, yes or no, where Levels 0 and 1 were converted to a “No” response and Levels 3-5 were converted to a “Yes” response.
- FAW categorical model, where Levels 1-5 were classified (if present) individually, with Level 0 merged with Level 1 since both show no evidence of leaf damage according to the survey scale.

The distribution of FAW categories across the 3 training datasets means that there are limitations on the modelling scenarios that can be achieved. For example, for the January dataset has 100% of the training data in damage level 2, therefore it is not possible to use any of this data on the 3 modelling scenarios. The December training data had to be balanced by up-sampling the level 2 category from 40 observations to 108. Binary modelling of this dataset was not possible since all the observations were in the No FAW damage binary category (i.e. merger of levels 0, 1 and 2).

For the February dataset there was (with up-sampling) a good distribution of observations across the FAW levels, which all 3 modelling scenarios achievable. Overall accuracy for the single RF model was 97%, although the k-fold model was significantly lower (73%) indicating that spatial autocorrelation maybe at play for the single model result.

The binary model was very accurate, with 96% accuracy in distinguishing between the presence or absence of FAW based solely on data from satellite observations.

Skewed data is a fundamental factor in the population dynamics of many pest and disease infestations which may limit the usefulness of current AI tools in analysing fluctuations in pest and disease infestations.

High level of variability across fields and quadrats

The project averaged both FAW damage scores and the reflectance across quadrants and fields. Fig. 2 shows the high level of variation between the pixels in each quadrat. Using an average score for the reflectance risks missing important information and weakening the model.

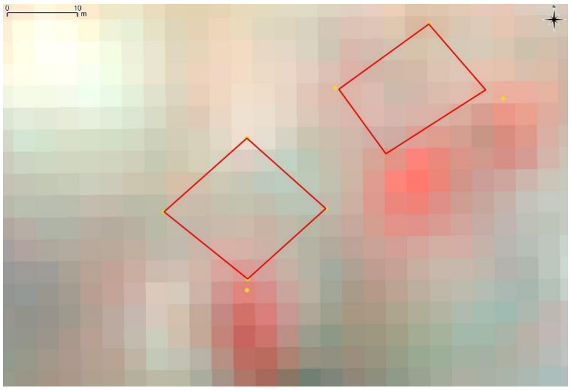

Figure 2: close up of quadrats to show pixel coverage

One solution the project tested was to adjust the sampling strategy for the satellite imagery from random points within the 10x10m quadrate to a single observation with a weighted distance mean, giving pixels located at the centre of the quadrat a greater importance than pixels on the edge. This is to minimise the adjacency effect which is common in remote sensing

To maximise the use of the data in the images, and the 30cm accuracy of the Emlid Reach GNSS future surveys will use a point sampling approach, sampling plants within 30cm of each coordinate:

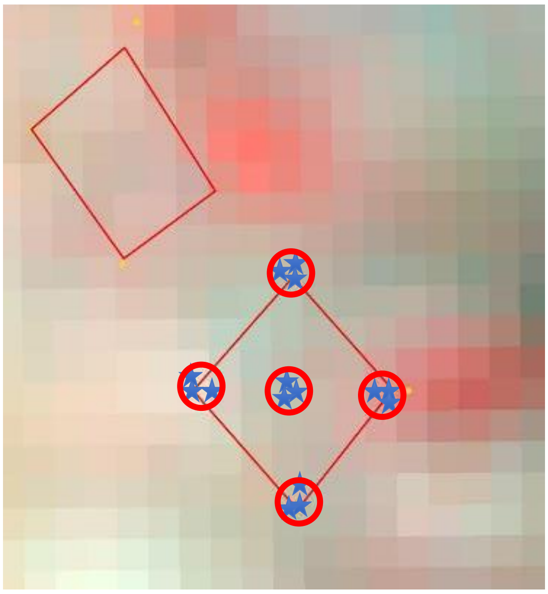

Figure 3: point sampling strategy

Spectral reflectance

One of the objectives of the research was to determine the variables that gave the strongest indication of FAW damage. One of the key advantages of the Random Forest algorithm is its ability to assess a wide range of variables. The following spectral bands and indices were tested, with the blue spectral band giving the strongest signal of FAW damage.

Spectral Band	Mean Importance Scores
Blue	0.275862
Red	0.173081

Green	0.169393
NIR (Near Infra-Red)	0.077198
Red, Standard Deviation	0.050376
NIR, Standard Deviation	0.046781
EVI (Enhanced Vegetation Index)	0.037190
SR (Surface Reflectance)	0.031657
Blue, Standard Deviation	0.028727
EVI (Enhanced Vegetation Index)	0.026975
MSAVI (Modified Soil Adjusted Veg Index)	0.024779
NDVI (Normalized Difference Vegetation Index)	0.024313
Green, Standard Deviation	0.018134
GNDVI (Green Normalized Difference Vegetation Index)	0.015534

Collaborations

This project is an example of NGO/commercial sector collaboration. The field work was possible through the good relationship SHA has developed with the Balaka District Administration; the Malawian Ministry of Agriculture, Irrigation & Water Development; the Malawian FAW Task Force and FAO Malawi.

Replicability

The next stage of the project is to test the replicability of the results across a range of soil types and agro-ecological zones, initially in Malawi and then across Southern Africa. SHA leads the consortium implementing the Farmer Field School component (BETTER) of the EU's KULIMA agriculture program in Malawi (https://kulimamalawi.org/). The BETTER project covers 10 districts across Malawi and the EU is keen that this research should be extended across these districts. Together with the EU SHA has selected 4 BETTER districts in distinct agro-ecological zones for testing the replicability of the results, funding permitting. The potential for replicability is important: the same methodology could be applied to different pests (locust, Banana Bunchy Top Virus, etc) in any geographical areas with good satellite imagery available.

Scalability

The small budget of the project limited software development to detecting FAW damage in sentinel maize fields demarcated at the start of each season. To be usable at a large scale the software will need a **Crop Mask**, a separate algorithm that detects the crop type in each field. The crop mask would be used to automatically identify maize fields in the landscape so the ARMYWORM software can be applied to the maize fields. Developing a Crop Mask in-house will require a large training data set, far larger than the data set developed for FAW damage detection.

Many research groups and commercial companies are developing crop masks and it is probably more cost, and time, efficient to request access to existing crop masks than to develop a project specific crop mask. Crop masks have been developed by OneSoil (https://map.onesoil.ai/2018#2/44.35/-43.66); the European Space Agency (Sentinel-2 Agriculture dynamic cropland mask, open source); NASA & Google (GFSAD1000), and others. These crop masks are well developed for European and North American crops but are still limited for African crops for the same reason that ARMYWORM did not develop a crop mask: the challenge of collecting sufficient ground truthing data. The project team are currently looking at the available crop masks and their suitability for African crops. The Ugandan tech startup Geogecko are keen to join the project to test their crop mask developed for Uganda (https://www.geogecko.com/).

The Malawian Ministry of Agriculture as requested that the approach used in the FAW project should be used to develop similar software for Banana Bunchy Top Virus. The World Bank's agriculture team in Malawi and the Irish Aid Mission to Malawi have proposed scaling-up the software into a tool for assessing pre-harvest crop losses across the country.

The team are also looking at feasibility of adapting the software to track Desert Locusts (*Schistocerca gregaria*) damage to crops and rangeland in East Africa. As locust swarms and bands are highly mobile the sentinel site approach used for gathering the FAW learning data will not work. An alterative approach is to use of the Penn State University / FAO Plant Village eLocust3M app[3] to collect real time data on locust movements. The locations will be available on the Plant Village website and will be used to procure multispectral satellite imagery to build the algorithm. Testing will be done on the severity of affected areas based on change detection analysis of biomass before and after an attack.

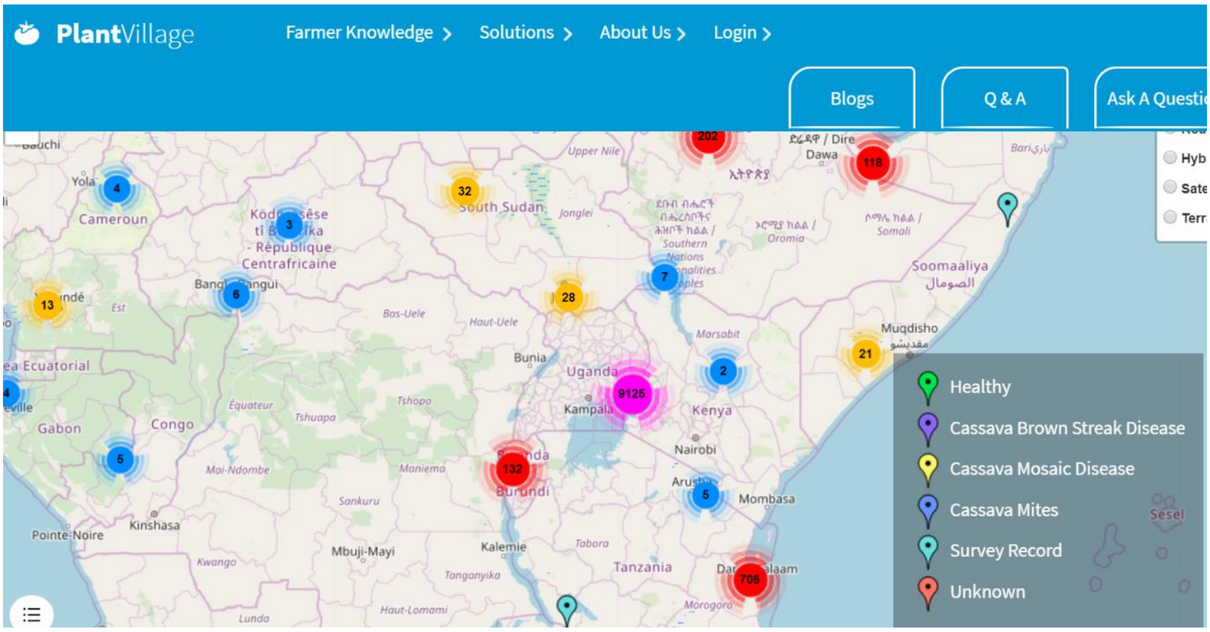

Figure 4: screen shot of the Plant Village Nuru web interface showing cassava disease incidence.

The work will also utilise other AI algorithms, such as Convolution Neural Networks (CNN's), which may be sensitive to structural changes in canopy vegetation before and after a locust swarm. Various satellite platforms will be utilised (e.g. Sentinel 2, Planet Labs, WorldView), so that the optimum temporal stack of images can be acquired in the shortest space of time over affected areas. Spatial and temporal resolution will be modelled as a function of accuracy so that recommendations can be made on the best image types and the number of images needed to achieve a sufficient level of accuracy.

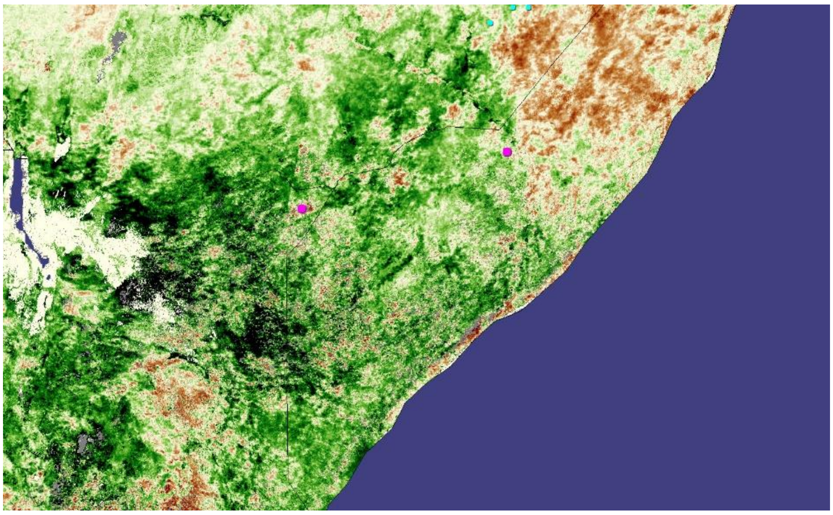
Figure 5: Satellite mapping of locust swarms (pink dots) in East Africa.

[3] https://play.google.com/store/apps/details?id=plantvillage.locustsurvey&gl=IE

Annex: Comparison of GNSS Equipment

Manufacture	Product	EUR	Correction services*	Claimed accuracy with correction	links
Trimble	R1	€2,843.20	RTX services,	60cm	https://geospatial.trimble.com/products-and-solutions/trimble-r1
Eos Positioning Systems	Arrow Lite	€1,668.63	Satellite correction services @ GBP 230 per month	50 cm	https://eos-gnss.com/product/arrow-series/arrow-lite
Eos Positioning Systems	Arrow 100	€2,784.77	Satellite correction services @ GBP 230 per month	40cm	https://eos-gnss.com/product/arrow-series/arrow-100
Eos Positioning Systems	Arrow Gold	€6,573.62	Satellite Correction service, Atlas H10, included in the price	4cm	https://eos-gnss.com/product/arrow-series/arrow-gold/
Juniper Systems	Geode Multi-GNSS Sub-meter Receiver, 1Hz	€1,919.76	NTRIP	30cm	https://www.junipersys.com/Products/Geode
EMLID	Emlid Reach RS+ survey kit (2 Reach RS+ units, base station and rover)	€1,500 set	Not required, differential GPS	7mm (unverified)	https://emlid.com/reachrs2/