The application of drones for mosquito larval habitat identification in rural environments: a practical approach for malaria control?

Authors:

Michelle C Stanton[1,2], Patrick Kalonde[3], Kennedy Zembere[3], Remy Hoek Spaans[1,2], Christopher M Jones[1,3]

[1] Vector Biology Department, Liverpool School of Tropical Medicine, Liverpool, UK

[2] Lancaster Medical School, Lancaster University, Lancaster, UK

[3] Malawi-Liverpool-Wellcome Trust Clinical Research Programme, Blantyre, Malawi

Corresponding author: Michelle C Stanton, Vector Biology Department, Liverpool School of Tropical Medicine, Liverpool UK L3 5QA

Email: michelle.stanton@lstmed.ac.uk

Running headline: *Drones for mosquito larval habitat identification*

Abstract

1. Spatial and temporal trends in mosquito-borne diseases are driven by the locations and seasonality of larval habitat. One method of disease control is to decrease the mosquito population by removing habitat and/or reduce the likelihood of larvae developing into adults, known as larval source management (LSM). In malaria control, LSM is currently considered impractical in rural areas due to perceived difficulties in identifying target areas. High resolution drone mapping is being considered as a practical solution to address this barrier. In this paper, we use our experiences of drone-led larval habitat identification in Malawi to assess the accuracy and practicalities of this approach.
2. Drone imagery and larval surveys were conducted in Kasungu district, Malawi between 2018-2020. Water bodies and aquatic vegetation were identified in the imagery using both manual methods and geographical object-based image analysis (GeoOBIA) and the performance of the classifications were compared. Larval sampling sites were characterised by biotic factors visible in drone imagery (e.g. vegetation coverage, type), and generalised linear mixed models were used to determine their association with larval presence.
3. Imagery covering an area of 8.9km^2 across eight sites was captured. Characteristics associated with rural larval habitat were successfully identified using GeoOBIA (e.g. median accuracy = 0.98, median kappa = 0.96 using a standard RGB camera), with a median of 18.3% being classed as surface water, compared to 20.1% using manual identification. The GeoOBIA approach, however, required greater processing time and technical skills. Larval samples were captured from 326 sites, and a relationship was identified between larval presence and vegetation (log-OR=1.44, p=0.01). Vegetation type was also a significant factor when considering late stage anopheline larvae only.
4. Our study demonstrates the potential for drone-acquired imagery as a tool to support the identification of mosquito larval habitat in rural areas where malaria is endemic. There are, however, technical challenges to overcome before it can be smoothly integrated into malaria control activities. Further consultations between experts and stakeholders in the fields of drones,

image analysis and vector control are needed to develop more detailed guidance on how this technology can be most effectively exploited.

Key words

Drones, machine-learning, object-based image classification, mosquito, Anopheles, malaria, larval habitat, mapping

Introduction

Malaria cases in Africa have reduced by over half in the last two decades making transmission more heterogeneous. This has led to a growth of studies applying spatial and temporal analyses to determine where and when remaining transmission foci exist (Stresman, Bousema, & Cook, 2019), and a focus on how new and existing control methods can be best utilised to reduce this residual transmission (Bousema et al., 2016; Hsiang et al., 2020; Sy et al., 2019).

The geographical spread and extent of malaria transmission is limited by the seasonally-driven mosaic of water bodies available for female mosquitoes in which to lay their eggs. The ecology of preferred breeding grounds for mosquito oviposition vary both within and between species. For example, two of the main sibling species of the *Anopheles gambiae* sensu lato complex, *An. gambiae* and *An. arabiensis*, are found in transient, sunlit, small pools whereas *Anopheles funestus* is associated with more permanent, larger vegetated water (Nambunga et al., 2020). At the micro-geographic scale, the presence of mosquito larvae may differ over the course of just a few metres (Eneh, Fillinger, Borg Karlson, Kuttuva Rajarao, & Lindh, 2019; Gowelo et al., 2020; Musiime et al., 2020). Biotic and abiotic factors such as the presence of specific types of vegetation, microbiota, predators, algal density, shade, and water depth influence larval development.

Mosquito larval populations are fixed in space for the duration of their development to adulthood. Typically, eggs hatch into larvae within 2-3 days of oviposition and take 5-10 days to metamorphosise into pupae, although the speed of this process is highly dependent on temperature (Beck-Johnson et al., 2013). One method of controlling diseases transmitted by mosquitoes is to reduce the population by reducing the availability of oviposition sites and/or reduce the likelihood that resulting larvae develop into the adult stage (World Health Organization, 2013). Larval source management (LSM) involves the environmental, biological or chemical manipulation of the environment in which mosquitoes are present for the purpose of targeting the immature, aquatic stages of the mosquito and hence reducing the adult mosquito population. In the early days of mosquito control, an

aggressive approach to searching and removing mosquito breeding sites was successful at reducing (and even eliminating) disease, with historical examples including its use during the construction of the Panama Canal in the early 20th Century, and its role in the elimination of *Anopheles gambiae* in Brazil by 1940 (Tusting et al., 2013). In sub-Saharan Africa, LSM was responsible for large reductions in malaria incidence in Zambia copper mines between 1929 and 1949 (Fillinger & Lindsay, 2011). LSM is, however, a labour-intensive exercise and following the introduction of IRS by DDT in the 1950s and subsequently the development of ITNs in the 1990s, it fell out of favour as a viable control option, particularly in Africa where the long rainy seasons produce countless sites for *Anopheles* development (Fillinger & Lindsay, 2011). As such, LSM is currently only recommended as a complementary vector control intervention to bed nets and IRS to target residual transmission and as a method of combating insecticide resistance (Killeen, 2014). While its value is acknowledged by WHO and national malaria control programmes (NMCPs) there are several barriers to its widespread implementation.

The primary barrier to implementing LSM is the issue of determining where and when the intervention should be implemented. In rural settings the WHO recommend the application of LSM in areas where there is high coverage of long-lasting insecticidal nets (LLINs), evidence of outdoor biting and/or insecticide resistance and where larval sources are '*few, fixed and findable*'. Despite the lack of a clear definition of what can be considered 'few' or 'findable', this has led to many considering LSM to be impractical in rural areas with diffuse seasonal larval habitats. The perception of these terms may evolve as technology and processes for implementing LSM advance. This paper focuses on challenging the '*findable*' component of this trio of conditions.

Geospatial technology is rapidly evolving and what now constitutes as '*findable'* may switch from less reliance on exhaustive ground-based searches to remotely sensed data. Drone mapping is being touted as at least equivalent (if not superior) to and more cost-effective than mapping larval habitat manually (Carrasco-Escobar et al., 2019; Hardy, Makame, Cross, Majambere, & Msellem, 2017) or using remotely sensed satellite imagery. While the latter can cover vast areas in a single day, images

are often obscured by clouds and although very high-resolution commercial satellite imagery exists, the resolution (at best 30 cm) is still inferior to that obtained by drones (2-10 cm) with the time of image captured out of the data user's control.

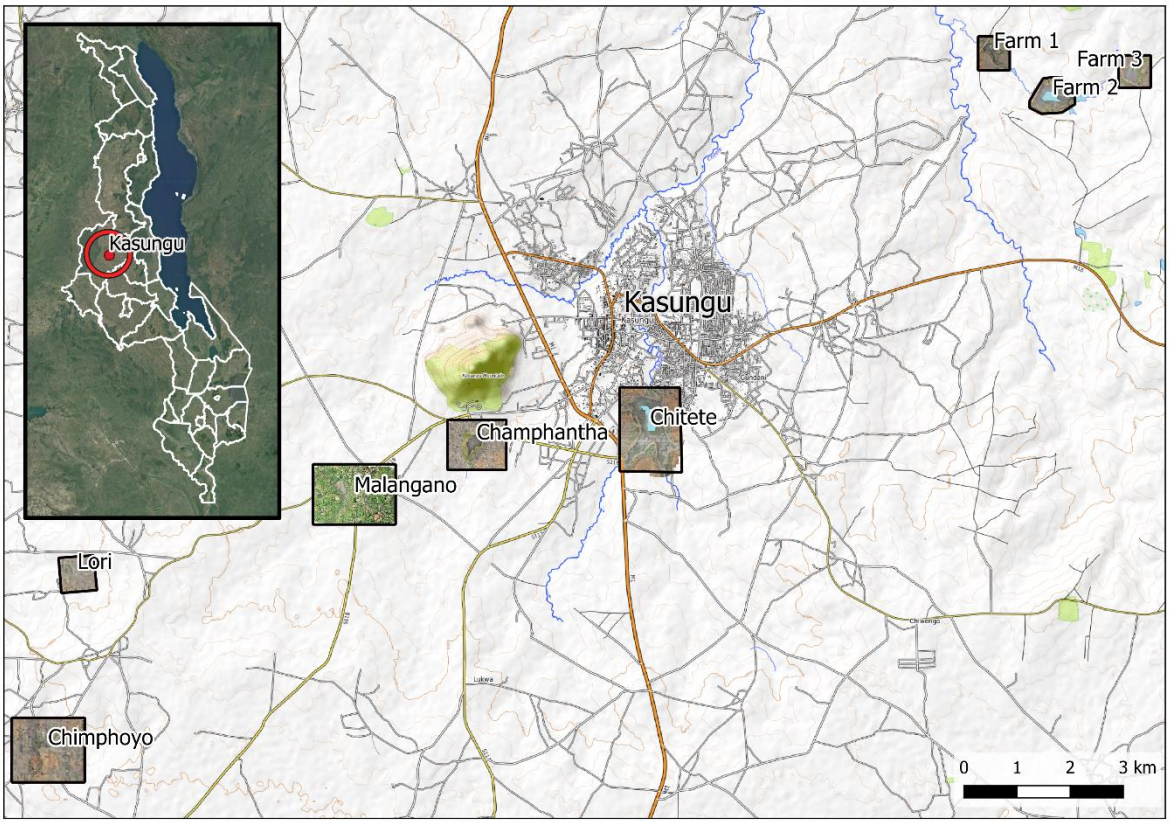
In this paper we explore the use of drones as a method for collecting very high resolution, contemporary imagery of an area for the purposes of identifying larval habitat. We tackle issues relating to the process of capturing drone imagery (by who, how much, how often), processing the images to extract the required information (what software, image classification methods, computer processing requirements), collecting 'ground-truth' data (entomological sampling), and subsequently summarising this information into recommendations that can be used by the control program implementers to guide their LSM activities.

Materials and Methods

Drone image capturing

A series of image data capture exercises were conducted within Kasungu district, central Malawi in an area that has been designated by the Government of Malawi, in collaboration with UNICEF, as a 'humanitarian drone testing corridor' (Fig. 1). Authorisation to conduct these flights was obtained from the Malawi Department of Civil Aviation. Malaria transmission occurs all year round in this area, with parasite prevalence in children between 2 and 10 years old estimated at 19% in 2017 (Chipeta et al., 2019). This transmission is potentially driven by a number of reservoirs which provide permanent sources of water within which female *Anopheles* can lay their eggs (Kibret, Lautze, McCartney, Nhamo, & Wilson, 2016). Images were captured over three visits in June 2018 (early dry season), October 2019 (late dry season) and February 2020 (wet season) using two drones, both of which were able to capture images using a standard RGB camera, plus a near-infrared (NIR) camera. Both drones were purchased off-the-shelf from commercial vendors. The first was a multirotor (quadcopter) type aimed towards the 'hobbyist' market (the DJI Phantom 4 Pro), supplemented by an additional NIR sensor by

120 Sentera (Sentera, 2020). The second was a fixed-wing drone marketed towards the agriculture
121 industry (eBee SQ) which incorporated a Parrot Sequoia multispectral camera.



122

123 Figure 1: Locations of sites surveyed within the 'humanitarian drone testing corridor', centred on
124 Kasungu town, Central Malawi (inset). Coordinates can be found in Table S1.

125 *Drone image processing*

126 Individual images captured during each mapping mission were stitched together into orthomosaics
127 using the commercial image processing software Agisoft Metashape Professional (version 1.4.2). A
128 subset of images captured during the wet season were classified using a geographical object-based
129 image analysis (GeoOBIA), using the LargeScaleMeanShift algorithm within the open source software
130 Orfeo Toolbox (version 7.1.0), applied within the QGIS environment (version 3.8.1). GeoOBIA involves
131 grouping contiguous pixels into 'objects' or 'segments' such that each segment is relatively
132 homogenous (within a prespecified threshold) with respect to pixel characteristics. In this instance,

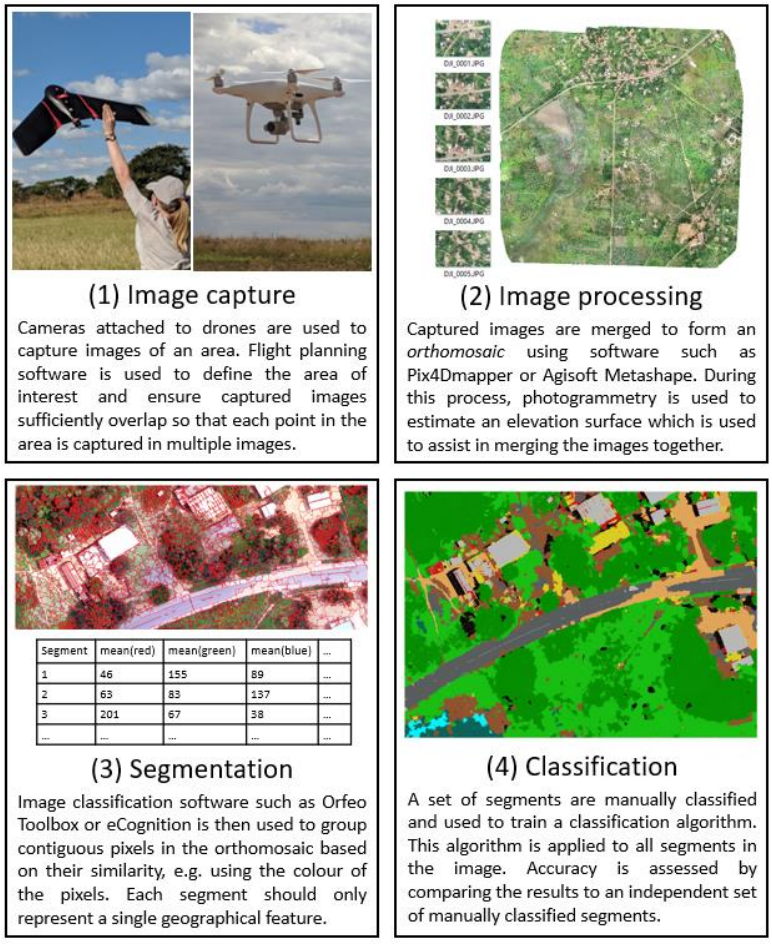
pixels were grouped into segments according the values of red, green, blue and elevation, with the latter being estimated using photogrammetric methods within Agisoft Metashape and then rescaled to lie between 0-255 to match the scale of the RGB values. We used trial and error to select the optimal segmentation parameters i.e. the spatial radius, range radius and minimum segment size. The smoothing radius determines how the amount by which the image is smoothed or filtered prior to the segmentation algorithm being implemented, whereas the range radius determines the similarity between pixels for grouping within the same segment. Similarity in this context refers to the Euclidean distance between two pixels. Supervised classification was then undertaken to assign each segment to one of 12 land cover classes (Table S2), including open water and aquatic vegetation (floating, emerging, submerged) based on the characteristics of that segment. Table S3 displays the segment-level characteristics used, which incorporated characteristics related to segment texture (Haralick textural features (Haralick, Dinstein, & Shanmugam, 1973)) in addition to a range of water and vegetation indices. The mean and variance of each of these were used in the classification.

Classification was undertaken using a set of 1800 segments all of which were firstly manually classified by the research team. One-third of the segments (n = 600) were within a 400m by 400m area and were used to train the classification algorithm. An additional one-third were in the same 400m by 400m area and were used for evaluating the accuracy of the classification within the same geographical area used for training (spatial interpolation), whereas the remaining 600 segments were distributed outside of the area used for training (spatial extrapolation). Classification was undertaken in R (version 3.6.1) using the caret package and the Random Forests classification algorithm (Kuhn, 2008). Manually classified segments within the 400m by 400m area were randomly split into training and testing segments. A ten-fold cross validation approach was used to determine the optimal tuning parameters for the Random Forests algorithm, and the resulting model was then applied to the testing segments. An accuracy statistic (% of classifications that were correct), and the Cohen's kappa agreement statistic (Cohen, 1960) were then calculated for the interpolated and extrapolated testing segments initially considering all 12 land cover classes, followed by a reduced classification that only differentiated

159 between surface water (open water and aquatic vegetation) and any other class. The classification
160 was then applied to the entire 600m by 800m area, and the percentage of the area classed as being
161 covered in surface water was calculated. This process of randomly splitting the segments into training
162 and testing groups was repeated 100 times, and the median and inter-quartile range of the resulting
163 accuracy, kappa agreement and percentage surface water cover were reported.

164 A manual classification of surface water was also undertaken which involved systematically scanning
165 through the image and creating a polygon around each area of surface water. The surface area of the
166 image manually identified as being covered in water was then compared with that classified as surface
167 water using the automated approach. The time taken and the computer resources required for each
168 of these tasks were also recorded.

169 An overview of the image capture, processing and classification procedures is presented in Fig. 2.



170

Figure 2: Processes undertaken to identify larval habitat from drone imagery.

Entomological sampling

Larval surveys were conducted concurrently to the drone image capture during each field visit (Fig. 1). Permission to collect these data were obtained from the Kasungu District Council, Kasungu Water Board and private landowners. As the first two field visits were undertaken during the dry season, sampling was focused in and around permanent water bodies which in this case were local reservoirs that provide drinking water and irrigation to the area. Sampling was undertaken at regular intervals around the periphery of reservoirs. Reservoirs were selected purposively based on their proximity to Kasungu town (for accessibility) and their proximity to human settlements. The third visit was conducted during the wet season, and sampling was focused around one of the reservoirs sampled in the previous dry season. Both temporary and permanent water bodies were sampled, with sampling sites identified from drone imagery captured the previous day. A subset of sites was sampled on four consecutive days to determine their consistency with respect to larval presence.

At each site, the presence and number of larvae were recorded using 10 repeated dips of the surface water, categorised by stage (L1/L2 or L3/L4) and either anopheline or culicine. To characterise malaria vectors in Kasungu as part of our broader efforts to understand transmission in the area, we raised all anopheline larvae to adult stage and identified morphologically to species (Coetzee, 2020). The location, description and photographs of each site were recorded using an Android Smartphone and Open Data Kit (ODK). These photographs were later used to classify each site according to the amount and type of vegetation present plus turbidity. A generalised linear mixed model was then fitted to the resulting presence/absence data to predict the likelihood that larvae were present from biotic site information obtained via drone imagery (vegetation type, coverage, turbidity). Models were fitted to presence/absence data for any larvae irrespective of stage or genus, and for late stage larvae only as characteristics of habitat containing late stage larvae are considered by WHO to be of greater

196 importance than early stage (World Health Organization, 2013). The productivity of the sampled sites
197 with respect to the number of early or late stage larvae collected was also considered.

198 **Results**

199 *Image capturing*

200 During the three sampling periods we captured a total of 10 distinct areas in Kasungu, covering an
201 area of 8.9km^2. The two drones significantly differ in relation to operational costs, equipment and
202 software requirements and usage. Tables 1 and 2 describe the primary differences in relation to initial
203 costs and operational usage respectively. The fixed wing drone (eBeeSQ) had a greater initial cost than
204 the multirotor Phantom 4 Pro due to it being inclusive of a NIR sensor, costing approximately £7000
205 (inclusive of an educational discount), compared to £3,300 for a standard (RGB sensor) Phantom 4 Pro
206 drone on which a NIR sensor was retrofitted. The eBeeSQ also required a high-spec laptop (£1000+)
207 on which to run the software required to plan and conduct missions, whereas the Phantom 4 Pro was
208 operated using free apps installed on GPS-enabled Android or iOS smartphone or tablet devices. On
209 an operational level, the primary differences are between the flight times per battery, and the ease of
210 use (Table 2). While, overall the Phantom 4 Pro is easier to use due to the small amount of open space
211 required for take-off and landing, its limited battery life means that to cover a relatively modest area
212 of 1km^2, 2-3 individual flights are needed depending on whether a fixed launch site is used or whether
213 this is adapted to minimise flight time. This comes at a cost of both time and money, particularly given
214 each battery comes at a price of £150. The eBeeSQ fixed wing drone requires less energy to fly and
215 therefore batteries last approximately twice as long (up to 1 hour in comparison to 30 minutes) than
216 the Phantom 4 Pro. Therefore, while the time required to cover 1km^2 is longer (66 minutes compared
217 to 38 minutes when flown at 120m above sea level [asl] which an 80% overlap in captured images),
218 this area can be comfortably covered using two batteries and a single launch site, meaning that in
219 practice the process is more efficient. The fixed-wing drone is however more difficult to operate than

220 the Phantom 4 Pro, requires a larger space for take-off and landing, and therefore cannot be used in
221 more densely vegetated areas.

222 Table 1: Comparison of approximate costs required to capture and process imagery captured by the
223 Phantom 4 Pro (rotor) and eBeeSQ (fixed wing).

			Phantom 4 Pro	**GBP (£)**	**eBeeSQ**	**GBP (£)**
Costs	Initial costs	Standard drone	RGB only	£1500	RGB only	N/A
		Supplementary sensor	Sentera NDVI	£1800	Parrot Sequoia	£7000+
	Supplementary hardware		Tablet	£150	High spec laptop	£1000+
		Spare batteries	<=30 mins flight time (per battery)	£150	<=1hr flight time (per battery)	£90
	Supplementary software	Mission planning	Pix4D capture	£0	eMotion Ag	£0
		Image processing	Agisoft MetaShape Professional Edition (Educational Licence)	£425	Agisoft MetaShape Professional Edition (Educational Licence)	£425
		Image classification	Orfeo Toolbox	£0	Orfeo Toolbox	£0

224

225 Table 2: Practical and operational differences between the drones used in this study

		Phantom 4 Pro	**eBee SQ**
Type of drone		Multirotor/multicopter	Fixed wing
Battery life (mins)		30	60
Practical flight time, accounting for take-off and landing (mins)		22	45
Area (km^2) covered per battery	120m asl and 80% overlap	0.49	0.64
Time (mins) required to cover $1km^2$	120m asl and 80% overlap	38	66
Image resolution at 120m asl (cm/pixel)	RGB camera	3.3	3.7
	NIR sensor	11	11cm
Ease of use	Mission planning	Via app on tablet/smartphone	Via software installed on laptop computer
	Take-off and landing	Vertical take-off and landing	Manual launch, and gradual descent in clear area

226

227 *Image processing & classification*

228 We used Agisoft Metashape to process all images captured. Using the Phantom 4 Pro flying at 120m

229 above surface level, with an 70% overlap in images, a total of 782 individual images (6.2GB) covered

an area of 1.77km^2. Processing these images in Agisoft Metashape in order to produce an orthomosaic of the area and an accompanying digital surface model took a total of 250 minutes using a computer with an Intel Core i7-6700 processor, 32GB RAM, resulting in an orthomosaic with a spatial resolution of 3cm (file size = 4.2GB). A subset of the image covering an area of 0.48km^2 (800m by 600m, file size = 1.9GB) was then selected for classification (Fig. 3).

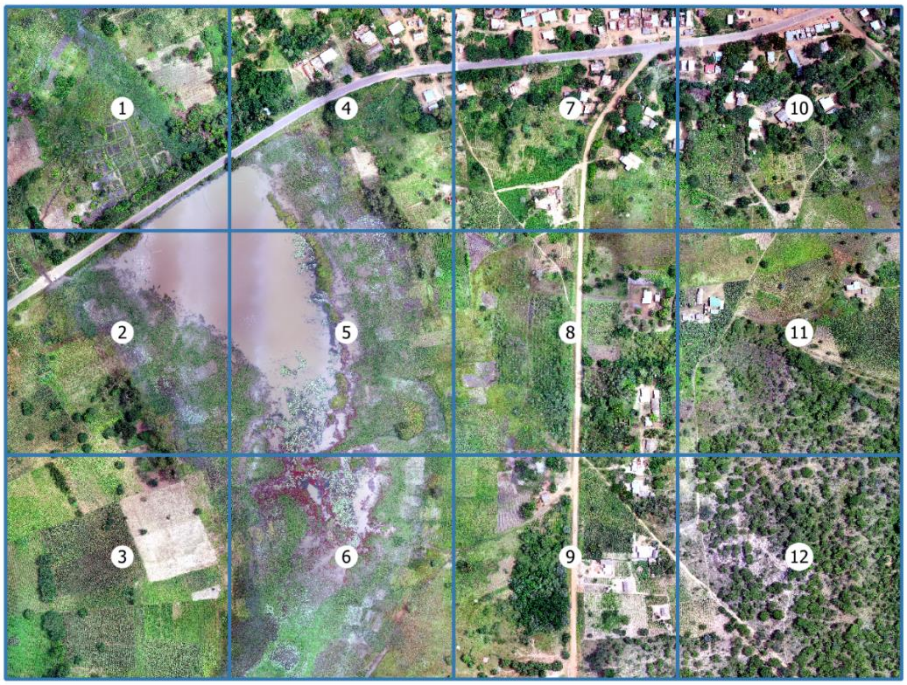


Figure 3: Image captured by the Phantom 4 Pro in Kasungu in February 2020 covering 800m by 600m, with each grid representing 200m by 200m. Grids 4, 5, 7 and 8 were used for training the classification algorithm, and an assessment of its accuracy was made using features both within this area (interpolation) and in the surrounding grids (extrapolation).

A set of 1800 training and testing segments were then generated for the 12 identified land classes (Table S2), plus an additional category representing areas that were in shadow. The study area was partitioned into cells of 200m by 200m, labelled as cells 1-8 (Figure 3), and the training and testing segments were proportionally distributed throughout the cells as follows: two thirds (1200) of segments were within cells 4, 5, 7 & 8 covering an area of 400m by 400m. We refer to these as the *internal* segments. One third of segments (600) were within the remaining cells (1-3, 6, 9, 10-12) which we refer to as *external* segments.

Segmentation was performed using Orfeo Toolbox (OTB) functions within the QGIS environment. Figure 4 demonstrates the impact of varying values of the spatial and range radius on the resulting segmentation and the time taken to perform this segmentation over a 100m by 100m area using the computer specifications previously specified (see Methods). While increasing the spatial radius provided a more adequate balance between over-segmentation (single discrete features of interest being split into many segments) and under-segmentation (multiple discrete features of interest being grouped into a single segment), this came at the price of substantially increasing the processing time. Additional processing time is required to calculate the segment-level summaries (mean, variance) of each of the variables being used to classify the imagery, with processing time increasing as the number of segments increases.

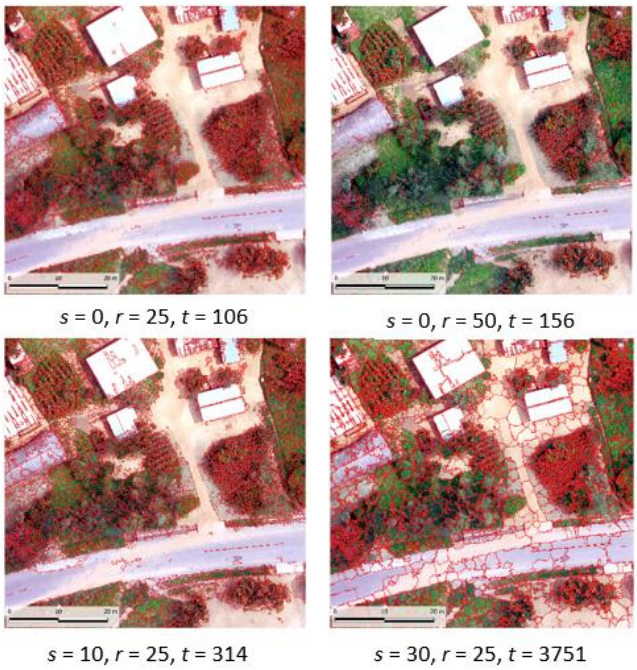


Figure 4: Examples of the segmentation process under different values for spatial radius *s* (0 ,10, 30), range radius *r* (25, 50) with a minimum segment size of 100. Time *t* corresponds to the time taken in seconds to segment a 100m by 100m image with a spatial resolution of 3cm using the LargeScaleMeanShift algorithm in Orfeo Toolbox. This process includes calculating the mean and variance of the RGB and elevation values for each segment.

261 The segmentation process was then applied to the entire 800m by 600m area using the parameters
262 10 (spatial radius), 25 (range radius) and 200 (minimum segment size), creating close to 800,000
263 segments. The total processing time, which includes calculating the segment-level mean and variance
264 of the RGB and elevation values, was 24.5 hours with an additional 15 hours taken to calculate the
265 mean and variance of each of the additional variables under consideration (Table S2). Two
266 classifications were then undertaken, one of which included the NIR-derived variables and one of
267 which did not.

268 The resulting accuracies of these classifications are presented in Table 3, and a representation of the
269 classified output from one area of Kasungu excluding NIR-derived variables is shown in Fig. 5, with the
270 resulting image which includes NIR-derived variable presented in Figure S1. The corresponding
271 variable importance plots are available in the SI (Figure S2). The results are very similar for both models
272 fitted with and without the NIR-derived variables. The variable making the greatest contribution to
273 the classification model in both cases is the mean elevation, with mean red, blue, green and brightness
274 also important. While the NIR-derived variables NDVI and SAVI make the greatest contribution to the
275 classification algorithm in the second model (Figure S2), Table 3 indicates that the inclusion of NIR-
276 derived variables does not make any significant impact on classification accuracy, with overall median
277 interpolated accuracy obtained using NIR-derived variables being marginally lower (0.904) than that
278 obtained without using NIR-derived variables (0.910). There is a clear drop in both accuracy and kappa
279 agreement when considering data from the extrapolation area, with accuracy reducing to 0.761 and
280 0.798 when considering overall accuracy without and with NIR-derived variables respectively. This
281 reduction in classification quality is less pronounced when considering surface water (open water or
282 aquatic vegetation) accuracy alone compared with trying to distinguish between all 12 land cover
283 classes. Trends in values of kappa are similar.

284 There is a close agreement in the percentage of the 400m by 600m area that is covered in surface
285 water obtained by fitting the model with and without NIR-derived variables (without NIR: median =

18.3%; with NIR: median = 16.6%), with more variability observed in the NIR-inclusive models (Table 3). A manual review of the image, independently undertaken by two researches, resulted in a larger percentage of the area being identified as surface water (21.2% and 20.1% by researchers 1 and 2 respectively), with the intersection of these two outputs (our '*manual surface water classification'*) covering 20.1% of the area. This process of manual classification took approximately two hours to complete.

Table 3: Summaries of classification accuracy (proportion of segments correctly classified) and kappa agreement for all 12 classes (*overall)* and for *surface water* (including open water and aquatic vegetation) versus all other classes for GeoOBIA obtained with and without NIR-derived variables.

	Area	Without NIR-derived variables		With NIR-derived variables	
		Median	IQR	Median	IQR
Overall accuracy	Interpolated	0.910	0.898-0.917	0.904	0.899-0.908
	Extrapolated	0.761	0.754-0.772	0.798	0.758-0.811
Surface water accuracy	Interpolated	0.983	0.980- 0.986	0.982	0.979-0.986
	Extrapolated	0.942	0.941-0.947	0.936	0.931-0.939
Overall kappa	Interpolated	0.902	0.888-0.909	0.895	0.890-0.900
	Extrapolated	0.738	0.731-0.750	0.779	0.735-0.793
Surface water kappa	Interpolated	0.960	0.954-0.967	0.958	0.950-0.967
	Extrapolated	0.873	0.868-0.881	0.855	0.844-0.864

% surface water	All	18.3	17.3-20.9	16.6	15.7-22.0

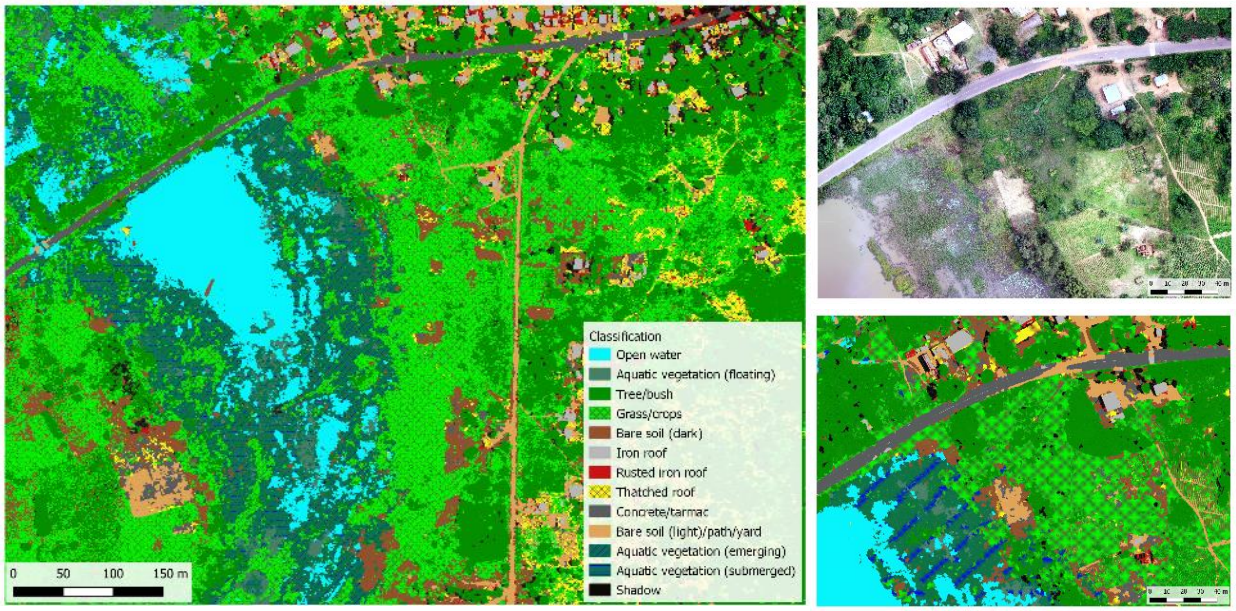


Figure 5: Example of a classification obtained for the entire study area using the random forests algorithm without including NIR-derived variables (left), with a more detailed view of a smaller area comparing the original image (top right) with the classified image (bottom right).

Entomological surveys

During three separate field visits (June 2018, October 2019, and February 2020) a total of 326 larval sites were sampled (available through the Figshare repository (Stanton, 2020)). During the dry periods these samples were focused along the shorelines of larger permanent water bodies (296 sites), with a mixture of 30 temporary and permanent sites surveyed during the wet season. Both anopheline and culicine larvae were found throughout the area during each sampling period (56% of sites sampled), with the lowest proportion of positive sites found in the late dry season (76% in June 2018, 16% in October 2019, 70% in February 2020). No clear sympatry was observed between anopheline and

culicine larvae in this study. For example, of the 321 sites where late stage larvae data were recorded (excluding 5 sites with missing data), larvae were observed in 31% (101) of samples with only 7% (24) sites containing both anophelines and culicines. In the area surrounding the Malangano site (Fig. 3) in February 2020, we morphologically identified 177 out of 297 anopheline specimens to species level, finding a predominance of *An. gambiae s.l.* (87.6%) followed by *An. coustani* (8.5%) and very few *An. pretoriensis* (2.3%) and *An. funestus* (1.7%).

At each site, GPS coordinates were recorded using the ODK app, photographs were taken using a smartphone, and aerial imagery was captured (Fig. 6). Samples were taken within approximately one metre of where the researcher stood to record the coordinate, however, as GPS coordinates have an accuracy of approximately three metres it was not possible to pinpoint precisely where the samples were taken within the aerial imagery.

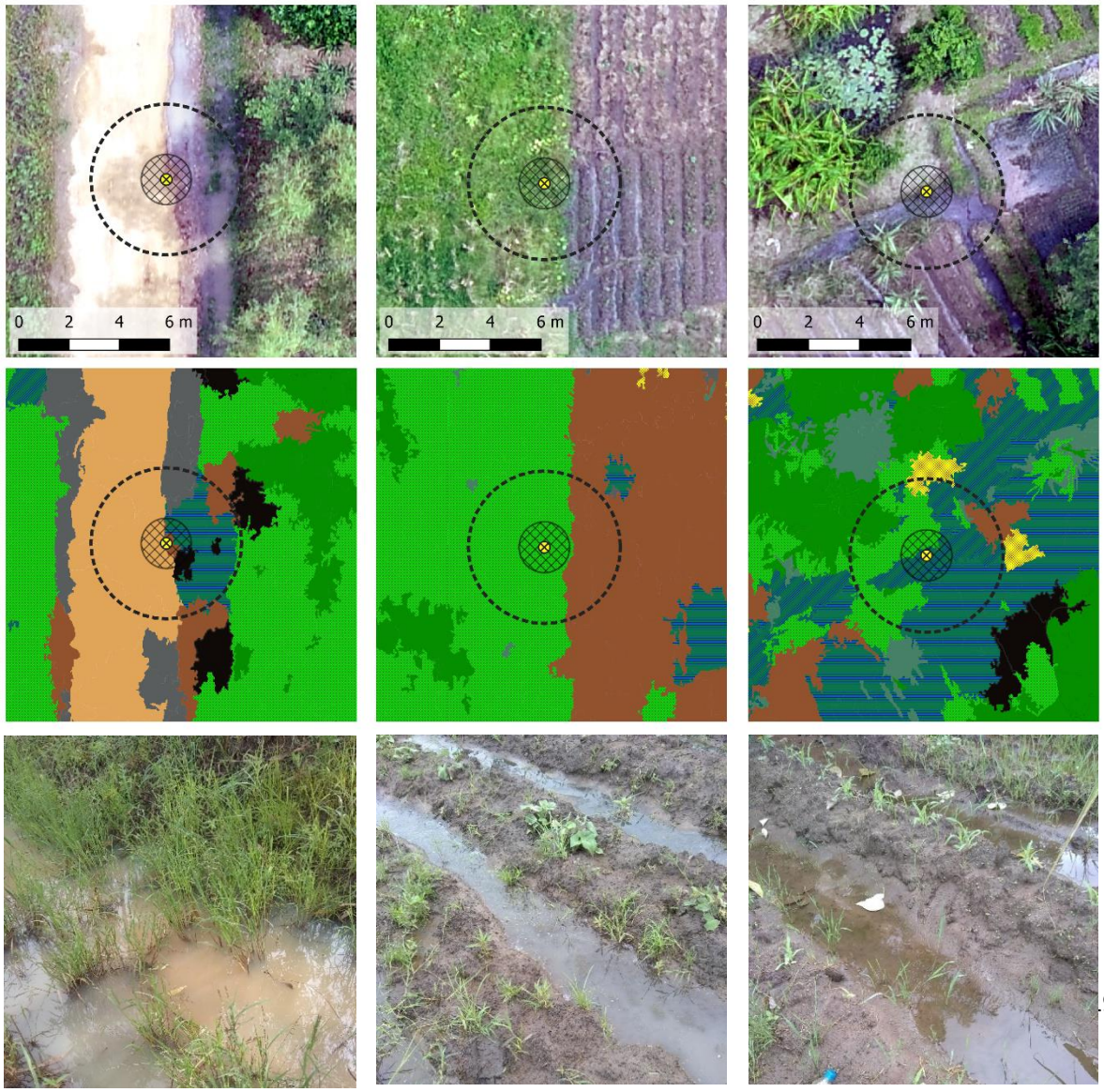


Figure 6: Examples of sampling sites where anopheline larvae were found. The top row indicates the precise GPS location captured using ODK (yellow circle), the expected sampling area based on these coordinates (1m radius), and the expected accuracy of the coordinates (3m radius), overlaid on top of the drone imagery. The middle row presents the classified imagery for these sites and the bottom row contains photographs of each site taken at the time of sampling.

Using the photographs, we characterised each site according to presence/absence of larvae and sample site characteristics including dominant vegetation type (none, floating, submerged, emerging), vegetation cover (none, <1/3, 1/3 - 2/3, >2/3) and turbidity. Vegetation was present in most sites sampled (284/326, 87%). Of these, 64% (183) contained emergent vegetation, 20% (57) contained submerged and 15% (44) contained floating vegetation. Vegetation cover varied evenly across sites, with 32% having low (0-1/3) coverage, 35% having moderate (1/3-2/3 coverage) and 33% having high (>2/3) coverage. There was an interaction between vegetation type and coverage, such that sites with floating vegetation rarely had high vegetation coverage. With regards to turbidity, 40% (129) of sites were classed as turbid whereas the remaining 60% (197) were clear.

Table 4: Summaries of the larval sampling sites by presence/absence of mosquito larvae found. A similar table for late stage (L3-L4) larvae can be found in the SI (Table S4).

		Any Larvae				
		Absent		**Present**		**Total**
		N	**(%)**	**N**	**(%)**	**N**
Sampling period	**2018 (early dry season)**	31	24	97	76	128
	2019 (late dry season)	141	84	27	16	168

	2020 (wet season)	9	30	21	70	30
Vegetation	**Yes**	148	52	136	48	284
	No	33	79	9	21	42
Dominant vegetation type	**None**	33	79	9	21	42
	Floating	27	61	17	39	44
	Submerged	30	53	27	47	57
	Emerging	91	50	92	50	183
Vegetation cover	**0**	33	79	9	21	42
	<1/3	56	62	35	38	91
	1/3 - 2/3	52	53	46	47	98
	>2/3	40	42	55	58	95
Turbidity	**Turbid**	74	57	55	43	129
	Clear	107	54	90	46	197
Total		181	56	145	44	326

341

342 A strong interaction was observed between vegetation type and coverage, therefore when fitting the
343 GLMMs to the presence/absence data we did not consider these variables in the model
344 simultaneously, but rather explored which of the two resulted in the best fitting model with regards
345 to AIC. After counting for the effect of sampling period and site, there was a strong association
346 between the presence of vegetation and the likelihood of any larvae (log-OR=1.44, p=0.01), however

accounting for vegetation coverage or vegetation type did not improve the model further. When considering *Anopheles* L3 and L4 larvae only, the model was improved when vegetation type was considered such that larvae were more likely to be present when emerging (log-OR=1.14, p=0.07) or submerged (log-OR=1.90, p=0.01) vegetation were available, compared to sites with no vegetation. With regards to productivity, while there was variability in the abundance of larvae sampled per site (145 sites, min = 1, median = 4, max = 56), there were insufficient high productivity sites to formally explore any trends in their characteristics.

During the wet season, 10 sites were repeatedly sampled over four days, with larvae consistently observed in four sites and no larvae being found on at least one day in the remaining six sites. We observed that due to changes in the environment it was difficult to resample the same locations across larger time scales. Temporary surface water observed in the wet season dried up even after just a few days without rain and shorelines of permanent water bodies varied substantially both between seasons and between the same season over consecutive years (Fig. 7). For example, we observed that images captured later in the dry season (October) in 2019 were wetter than those captured in the early dry season (June) in 2018 at the Chitete reservoir.

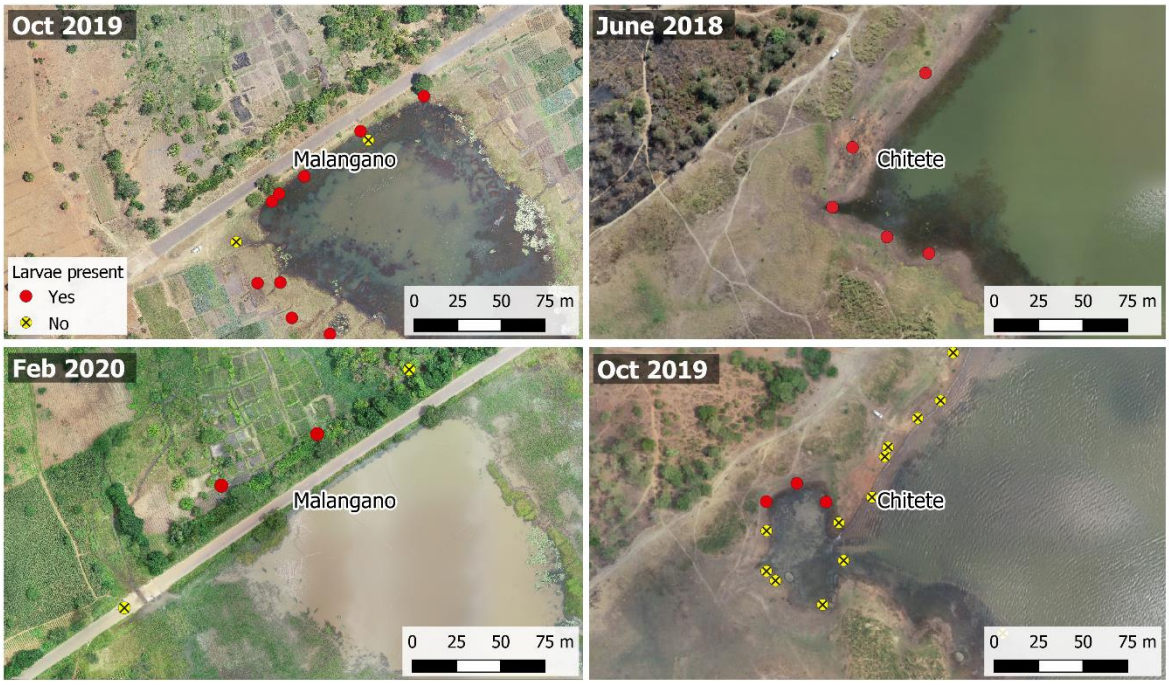


Figure 7: Comparisons of aerial images captured at different seasonal time points. Left images display a comparison between consecutive dry (Oct 2019) and wet (Feb 2020) seasons around the Malangano dam. Right images display comparisons between dry seasons over two consecutive years (June 2018, Oct 2019) around the Chitete dam.

Discussion

Image capturing

Image capture using drones inevitably leads to technical and skills-based challenges and we highlight a few of these here in the context of searching for water bodies in a rural setting. Aside from hardware and software issues we noted that flight experience was a key requirement to determine optimal flight times as neither the rotor or fixed wing drone could be flown in wet or windy conditions, and we experience the impact of extreme weather on the hardware with multiple occasions of over-heating on warm days. This highlighted the need for extensive drone piloting training by the operator. The country's drone regulations also need to be taken into careful consideration. In Malawi, data capture was facilitated by the relationship between UNICEF and the Department of Civil Aviation and a toolkit is currently being developed to outline the procedures that need to be followed by those wishing to fly drones for non-commercial purposes (https://www.updwg.org/wp-content/uploads/2019/12/Malawi-RPA-Toolkit-2019_Dec.-Final.pdf). While regulations vary by country, national civil aviation authorities are also requiring drone pilots to obtain accredited qualifications and seek appropriate permissions before using drones for research or humanitarian purposes. Training courses which cover both the operational and regulatory aspects of drone flying are currently quite sparse in sub-Saharan Africa and this may require future pilots to travel outside of their own country to gain the necessary experience. Should a malaria control implementer wish to use drone imagery within their programmes, they may therefore incur significant expense both in purchasing the equipment and training their staff. A solution to this would be to outsource the image capturing to qualified drone pilots operating in the area.

An additional bottleneck is the availability of hardware within the country of operation. While it may be possible to purchase off-the-shelf drones in-country, should any technical issues arise, obtaining part replacements or repairs becomes problematic and expensive. Investments are therefore being made in 'home-grown' drones, to support local economies, decrease the cost of equipment, and make repairs much more easily accessible. In Malawi for example, MicroMek (https://www.micromek.net/) manufacture the low-cost fixed wing drone known as EcoSoar (Standridge, 2018), for both transporting goods and capturing imagery.

Image processing

Processing drone imagery to create the orthomosaics is time-consuming, requires a high-spec computer and a large capacity for data storage. Therefore, to use this imagery in the field, an NMCP would require people skilled in both image capture and processing, plus access to the relevant software. These skills are not usually taught as part of standard drone pilot training, however this may change as the potential for using drone technology for humanitarian purposes is increasingly realised. For example, the African Drone and Data Academy was launched in January 2020 in Malawi to build capacity in both drone piloting and drone image processing and analysis (UNICEF, 2020).

Image classification is appealing because once the algorithm has been trained, it can simply be applied to any additional imagery captured without any or only a little additional data being required. In our analysis we showed that there was a decline in classification accuracy in areas within very close proximity to that used to train the algorithm and noted that even in this small area there were important land cover classes in the extrapolation area that did not appear in the training area e.g. red algae in the water. This challenge is likely to be exacerbated when considering areas further apart, or data collected at different time points. As more data are collected these limitations may be overcome, however in the short term, the effort required by the end-user e.g. an individual NMCP to train a classification algorithm may outweigh its benefits. In this demonstration we implemented a

geographical object-based classification approach which generated the segments and computing segment-level characteristics prior to training and applying the classification algorithm. The segment-generating process can be very time-consuming depending on the values of the segmentation parameter referred to as the spatial radius, and the size of the area being classified. While other classification techniques such as pixel-based classification may be quicker to perform, an object-based approach is the most appropriate for very high-resolution images such as that generated by drones (Pande-Chhetri, Abd-Elrahman, Liu, Morton, & Wilhelm, 2017). The cost and benefits of accuracy against processing time therefore need to be considered should an NMCP wish to perform image classification in-house. The role of additional sensors in the image classification process is also unclear. In this analysis we compared the classification accuracy using imagery captured from a standard camera only (RGB), plus additional imagery captured by a much more expensive NIR sensor. While our performance metrics indicated very little difference in the accuracy obtained using the two approaches, a more extensive investigation would need to be undertaken over a more environmentally diverse area before we can conclude whether or not there is a benefit to incorporating this additional technology.

A more practical solution to 'automated' image classification may be to persevere with the less efficient, but lower skilled task of manual classification. This task is, however, not without its drawbacks, as human error can easily miss small areas, or misclassify water containing a lot of aquatic vegetation as land and vice versa. These latter '*missed*' areas are of significance, as *Anopheles* mosquitoes are generally found in water containing vegetation. The fact that there was a 10% discrepancy in the manual classification undertaken by two independent researchers, both of whom were familiar with the study area, demonstrates the fallibility in this method.

As with the drone image capture, an alternative is to outsource these activities to an organisation which specialises in image processing and classification. Additionally, cloud-based computing services such as DroneDeploy's Map Engine (DroneDeploy, 2018) and Google Earth Engine (Mutanga & Kumar,

2019) could be used as these allow individuals/groups to harness the power of remote servers to manage and manipulate the data. This approach could facilitate the development of more automated habitat classification approaches i.e. using data from other organisations, previous field or professional expertise in remote sensing to develop classification algorithms that don't require the use of bespoke training data. The TropWet tool developed by Hardy, Oakes, & Ettritch (2020) is a demonstration of this in which satellite imagery (Landsat, 30m resolution) is automatically classified for a user-specified area and time period using a Google Earth Engine interface.

There are still practical challenges with these approaches, particularly relating to the upload of large image files to enable these processes to be undertaken remotely, however these may be preferable to the more technical challenge of managing the data in-house.

Entomological survey

Larval surveys are an important part of the process of LSM both to confirm the species of mosquitoes found in the area, to characterise the types of surface water where larvae are likely to be found, and to monitor the progress of any subsequent intervention. Larval surveys are however a time-consuming process, particularly when undertaken during the wet season during which areas become inaccessible following heavy rains. The role of drones in LSM is not to completely remove the need for larval surveys, but to help differentiate between water bodies with respect to their potential as larval habitat and/or to differentiate sites according to their potential larval productivity.

We note that in our study that vegetation coverage appears to be important when considering presence/absence of late stage *Anopheles* larvae, with coverage correlated with the type of vegetation found i.e. coverage of floating vegetation was likely to be less than that of emergent or submerged vegetation. A full understanding the larval ecology of the local individual malaria vectors would greatly assist a targeted LSM approach aided by drone-imagery support. In south-eastern Tanzania, a basic

462 characterisation of *An. funestus* larval habitats provides support that this species occupies small
463 spring-fed pools, permanent natural ponds and slow-moving waters each of which fall under the 'few,
464 fixed and findable' paradigm (Nambunga et al., 2020). In a recent study in Southern Malawi (Gowelo
465 et al., 2020), *An. arabiensis* was the dominant species, with high densities being found in aquatic
466 habitats surrounded by bare soil. A species-specific approach to identifying larval habitat using drone
467 imagery may therefore be required, with imagery captured throughout the year to better understand
468 the temporal dynamics of larval habitat and thereby optimise the impact of any potential intervention.
469 These images could further be used to monitor the progress of LSM campaigns with, for example, a
470 more accurate estimates of LSM coverage and demonstratable changes in the landscape because of
471 habitat removal/modification. Further entomological surveillance remains pivotal to establish where
472 and when LSM should be deployed and measure the impact of the intervention on malaria
473 transmission potential.

474 Conclusions

475 Our study demonstrates the potential for drone imagery to be used as a tool to support the
476 identification of mosquito larval habitat in rural areas where malaria is endemic. While this technology
477 has the capacity to complement the more labour-intensive approach of identifying larval habitat from
478 the ground, there are technical challenges to overcome before it can be smoothly integrated into
479 malaria control activities. We believe that outsourcing the capturing and processing of drone imagery
480 to private companies with the equipment and skills necessary to extract the required information is a
481 more practical approach to developing equivalent skills in house. These services are becoming
482 increasingly available in other sectors such as agriculture, forestry and environmental monitoring and
483 there are promising developments in the African drone sector to support this local capacity. We do
484 however continue to emphasise that drone imagery should not be used to completely replace larval
485 surveys. Instead we envisage that this technology could provide supplementary information which
486 may help to reduce the time spent finding locations to be sampled, monitor environmental changes

over time and help to guide the frequency and scale of any LSM intervention, ultimately increasing its potential for success. Further consultations between experts and stakeholders in the fields of drones, image analysis and vector control are needed to develop more detailed guidance on how this technology can be most effectively exploited.

Author contributions

MCS and CMJ conceived the ideas and designed methodology; all authors collected the data; MCS, PK and KZ analysed the data; MCS and CMJ led the writing of the manuscript. All authors contributed critically to the drafts and gave final approval for publication.

Data availability

Entomological data can be accessed at: https://figshare.com/s/25102b084f56f41a1ca8

All drone imagery referenced in this paper has been uploaded to OpenAerialMap from which it can be freely downloaded (https://map.openaerialmap.org/#/33.489675521850586,-13.050633215031182,13/user/5f1fe6f357ddda00054a0647?_k=mmgy4y).

References

Beck-Johnson, L. M., Nelson, W. A., Paaijmans, K. P., Read, A. F., Thomas, M. B., & Bjørnstad, O. N. (2013). The effect of temperature on Anopheles mosquito population dynamics and the potential for malaria transmission. *PLoS ONE*, *8*(11), 79276. doi: 10.1371/journal.pone.0079276

Bousema, T., Stresman, G., Baidjoe, A. Y., Bradley, J., Knight, P., Stone, W., … Smith, D. (2016). The Impact of Hotspot-Targeted Interventions on Malaria Transmission in Rachuonyo South District in the Western Kenyan Highlands: A Cluster-Randomized Controlled Trial. *PLOS Medicine*, *13*(4),

510 e1001993. doi: 10.1371/journal.pmed.1001993

511 Carrasco-Escobar, G., Manrique, E., Ruiz-Cabrejos, J., Saavedra, M., Alava, F., Bickersmith, S., …
512 Gamboa, D. (2019). High-accuracy detection of malaria vector larval habitats using drone-based
513 multispectral imagery. *PLOS Neglected Tropical Diseases*, *13*(1), e0007105. doi:
514 10.1371/journal.pntd.0007105

515 Chipeta, M. G., Giorgi, E., Mategula, D., Macharia, P. M., Ligomba, C., Munyenyembe, A., … Kayange,
516 M. (2019). Geostatistical analysis of Malawi's changing malaria transmission from 2010 to
517 2017. *Wellcome Open Research*, *4*, 57. doi: 10.12688/wellcomeopenres.15193.2

518 Coetzee, M. (2020). Key to the females of Afrotropical Anopheles mosquitoes (Diptera: Culicidae).
519 *Malaria Journal*, *19*(1), 70. doi: 10.1186/s12936-020-3144-9

520 Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological*
521 *Measurement*, *20*(1), 37–46.

522 DroneDeploy. (2018). Introducing Map Engine. Retrieved June 29, 2020, from
523 https://medium.com/aerial-acuity/introducing-map-engine-cd3ef93bc730

524 Eneh, L. K., Fillinger, U., Borg Karlson, A. K., Kuttuva Rajarao, G., & Lindh, J. (2019). Anopheles
525 arabiensis oviposition site selection in response to habitat persistence and associated
526 physicochemical parameters, bacteria and volatile profiles. *Medical and Veterinary*
527 *Entomology*, *33*(1), 56–67. doi: 10.1111/mve.12336

528 Fillinger, U., & Lindsay, S. W. (2011). Larval source management for malaria control in Africa: myths
529 and reality. *Malaria Journal*, *10*(353).

530 Gowelo, S. A., Chirombo, J., Koenraadt, C. J. M., Mzilahowa, T., Berg, H., Takken, W., & McCann, R. S.
531 (2020). Characterisation of anopheline larval habitats in southern Malawi. *Acta Tropica*,
532 105558. doi: 10.1016/j.actatropica.2020.105558

533 Haralick, R. M., Dinstein, I., & Shanmugam, K. (1973). Textural Features for Image Classification. *IEEE*
534 *Transactions on Systems, Man and Cybernetics*, *SMC-3*(6), 610–621. doi:
535 10.1109/TSMC.1973.4309314

536 Hardy, A., Makame, M., Cross, D., Majambere, S., & Msellem, M. (2017). Using low-cost drones to
537 map malaria vector habitats. *Parasites and Vectors*, *10*(29).

538 Hardy, A., Oakes, G., & Ettritch, G. (2020). Tropical Wetland (TropWet) Mapping Tool: The Automatic
539 Detection of Open and Vegetated Waterbodies in Google Earth Engine for Tropical Wetlands.
540 *Remote Sensing*, *12*(7), 1182. doi: 10.3390/rs12071182

541 Hsiang, M. S., Ntuku, H., Roberts, K. W., Dufour, M. S. K., Whittemore, B., Tambo, M., … Gosling, R.
542 (2020). Effectiveness of reactive focal mass drug administration and reactive focal vector
543 control to reduce malaria transmission in the low malaria-endemic setting of Namibia: a
544 cluster-randomised controlled, open-label, two-by-two factorial design trial. *The Lancet*,
545 *395*(10233), 1361–1373. doi: 10.1016/S0140-6736(20)30470-0

546 Kibret, S., Lautze, J., McCartney, M., Nhamo, L., & Wilson, G. G. (2016). Malaria and large dams in
547 sub-Saharan Africa: future impacts in a changing climate. *Malaria Journal*, *15*(448).

548 Killeen, G. F. (2014, August 23). Characterizing, controlling and eliminating residual malaria
549 transmission. *Malaria Journal*, Vol. 13, pp. 1–22. doi: 10.1186/1475-2875-13-330

550 Kuhn, M. (2008). Building Predictive Models in R Using the caret Package. *Journal of Statistical*
551 *Software*, *28*(5), 1–26. doi: 10.18637/jss.v028.i05

552 Musiime, A. K., Smith, D. L., Kilama, M., Geoffrey, O., Kyagamba, P., Rek, J., … Lindsay, S. W. (2020).
553 Identification and characterization of immature Anopheles and culicines (Diptera: Culicidae) at
554 three sites of varying malaria transmission intensities in Uganda. *Malaria Journal*, *19*(1), 221.
555 doi: 10.1186/s12936-020-03304-7

556 Mutanga, O., & Kumar, L. (2019). Google Earth Engine Applications. *Remote Sensing*, *11*(5), 591. doi:

10.3390/rs11050591

Nambunga, I. H., Ngowo, H. S., Mapua, S. A., Hape, E. E., Msugupakulya, B. J., Msaky, D. S., … Okumu, F. O. (2020). Aquatic habitats of the malaria vector Anopheles funestus in rural south-eastern Tanzania. *Malaria Journal*, *19*(1), 219. doi: 10.1186/s12936-020-03295-5

Pande-Chhetri, R., Abd-Elrahman, A., Liu, T., Morton, J., & Wilhelm, V. L. (2017). Object-based classification of wetland vegetation using very high-resolution unmanned air system imagery. *European Journal of Remote Sensing*, *50*(1), 564–576. doi: 10.1080/22797254.2017.1373602

Sentera. (2020). Sentera. Retrieved June 18, 2020, from https://sentera.com/sensors/

Standridge, Z. (2018). *Design and Development of Low-cost Multi-function UAV Suitable for Production and Operation in Low Resource Environments* (Virginia Polytechnic Institute and State University). Retrieved from https://www.researchgate.net/publication/326582792_Design_and_Development_of_Low-cost_Multi-function_UAV_Suitable_for_Production_and_Operation_in_Low_Resource_Environments

Stanton, M. C. (2020). Figshare dataset: Larval sampling data, Kasungu. Retrieved from Figshare website: https://figshare.com/s/25102b084f56f41a1ca8

Stresman, G., Bousema, T., & Cook, J. (2019, October 1). Malaria Hotspots: Is There Epidemiological Evidence for Fine-Scale Spatial Targeting of Interventions? *Trends in Parasitology*, Vol. 35, pp. 822–834. doi: 10.1016/j.pt.2019.07.013

Sy, O., Niang, E. H. A., Diallo, A., Ndiaye, A., Konaté, L., Ba, E. H. C. C., … Faye, O. (2019). Evaluation of the effectiveness of a targeted community-based IRS approach for malaria elimination in an area of low malaria transmission of the central-western Senegal. *Parasite Epidemiology and Control*, *6*, e00109. doi: 10.1016/j.parepi.2019.e00109

Tusting, L. S., Thwing, J., Sinclair, D., Fillinger, U., Gimnig, J., Bonner, K. E., … Lindsay, S. W. (2013).

581 Mosquito larval source management for controlling malaria. In J. Thwing (Ed.), *Cochrane*
582 *Database of Systematic Reviews* (p. CD008923). doi: 10.1002/14651858.CD008923.pub2

583 UNICEF. (2020). African Drone and Data Academy. Retrieved from
584 https://www.unicef.org/malawi/african-drone-and-data-academy-malawi

585 World Health Organization. (2013). *Larval source management – a supplementary measure for*
586 *malaria vector control. An operational manual*. Retrieved from
587 http://www.who.int/malaria/publications/atoz/9789241505604/en/

588

589