

RESEARCH ARTICLE

Camera trapping with photos and videos: implications for ecology and citizen science

Sian E. Green^{1,2} , Philip A. Stephens², Mark J. Whittingham³ & Russell A. Hill¹¹Department of Anthropology, Durham University, Durham DH1 3LE, UK²Conservation Ecology Group, Department of Bioscience, Durham University, Durham DH1 3LE, UK³School of Natural and Environmental Sciences, Newcastle University, Newcastle NE1 7RU, UK**Keywords**

activity level, camera traps, citizen science, occupancy, species diversity

Correspondence

Sian E. Green, Department of Anthropology, Durham University, Durham, DH1 3LE, UK.

E-mail: sian.e.green@durham.ac.uk; sian.green92@outlook.com

Editor: Marcus Rowcliffe

Associate Editor: Rahel Sollmann

Received: 25 May 2022; Revised: 5 September 2022; Accepted: 10 September 2022

doi: 10.1002/rse2.309

Remote Sensing in Ecology and Conservation 2023, **9** (2):268–283**Abstract**

Camera traps are increasingly used in wildlife monitoring and citizen science to address an array of ecological questions on a wide variety of species. However, despite the ability of modern camera traps to capture high-quality video, the majority of studies collect still images, in part because of concerns with video performance. We conducted a camera trap survey of a forested landscape in the UK, using a grid of paired camera traps, to quantify the impact of using video compared to photos on the outcomes of ecological research and for participation and engagement of citizen scientists. Ecological outputs showed no difference between photo and video datasets, but comparison between expert and citizen science classifications showed citizen scientists were able to classify videos more accurately (average accuracy of 95% for video, 86% for photo). Furthermore, citizen scientists were more likely to volunteer additional information on age (provided for 61% videos and 30% photos) and sex (provided for 63% videos and 45% photos) of animals in video footage. Concerns over slow trigger speeds for videos did not appear to affect our datasets or the inferences gained. When combined with citizen science, video datasets are likely to be of higher quality due to increased classification accuracy. Consequently, we encourage researchers to consider the use of video for future camera-trapping projects.

Introduction

The use of camera traps for research has seen exponential growth over the last two decades (Blount et al., 2021). While this growth may slow, development of technology, analytical methods and coordinated data sharing platforms will allow for continued diversification of the topics and questions that can be addressed using camera trap data (Delisle et al., 2021). Furthermore, the practical advantages of camera traps during times of uncertainty and restricted travel have recently become more evident (Blount et al., 2021). In light of this, there is a pressing need to consider how to optimize camera trap set-ups for specific purposes.

While the majority of camera trap research uses photographs, video footage may be more suitable for applications such as behavioural studies (Caravaggi et al., 2017; Janisch et al., 2021; Tagg et al., 2018), monitoring group

size (Balestrieri et al., 2016; Green-Barber & Old, 2018; Medeiros et al., 2019) or density estimation of unmarked species (Corlatti et al., 2020; Howe et al., 2017; Nakashima et al., 2018). Video footage may also increase public engagement and facilitate easier identification of species and individuals for citizen scientists (Reyes et al., 2017; Swinnen et al., 2014). Despite these advantages, a number of issues may deter researchers from using video. First, camera traps can have slower trigger speeds and longer recovery times when set to video and therefore risk missing some events (Apps & McNutt, 2018; Findlay et al., 2020). Second, videos have larger file sizes, leading to faster filling of memory cards, and increased power consumption when recording, leading to shorter battery life (Blount et al., 2021; Janisch et al., 2021). A final concern is the longer processing time needed to view videos, which may be exacerbated by a lack of support for video management in software

designed to streamline camera data management. Processing times are already an issue for many projects collecting photo data (Glover-Kapfer et al., 2019; Rovero & Zimmermann, 2016; Young et al., 2018) and can slow research and lead to potentially valuable data on non-target species being left unanalysed (Scotson et al., 2017).

Despite these concerns, little work has quantified the impacts of using video on the outcomes of ecological research. While some research has explored the impact of different camera trap settings, this has focused on controlled scenarios, such as using domestic animals to trigger cameras (Apps & McNutt, 2018; Yajima & Nakashima, 2021) and typically uses relatively small numbers of camera traps and sites. Fewer studies have compared photo and video settings; those that have were focused on a small number of species (Findlay et al., 2020; Glen et al., 2013; Palencia et al., 2019), or on the identification of individuals of a species (Reyes et al., 2017). Instead, studies have focused on the influence of camera model (Driessen et al., 2017; Yajima & Nakashima, 2021), camera position (Apps & McNutt, 2018; Jacobs & Ausband, 2018; Meek et al., 2016; Seidlitz et al., 2021), flash type (Herrera et al., 2021) or sensitivity settings (Palencia et al., 2021).

To support the time-consuming matter of data processing, camera trap researchers may turn to citizen science (Meek & Zimmermann, 2016). Camera-trapping citizen science projects have burgeoned recently and been shown to provide ecologically meaningful data (Hsing et al., 2018; Lasky et al., 2021; McShea et al., 2016; Swanson et al., 2016). Factors such as camera settings and location can impact classification accuracy; specifically, sequences with multiple photos have been found to have higher classification accuracy than those with single photos (Egna et al., 2020). Camera trap videos could allow for easier species identification because movement can make animals easier to locate within the footage, and because more information is available to an observer, such as different views of an animal, their gait or movement profile, and sound. However, probably owing to the concerns outlined above, most camera trap citizen science projects use photographs and there has been little assessment of citizen science classification accuracy of videos (but see McCarthy et al., 2021). Gaining adequate numbers of classifications is important for timely processing and for combining multiple classifications to achieve higher confidence in classification accuracy (Anton et al., 2018; Egna et al., 2020; Hsing et al., 2018; Swanson et al., 2016). Attracting participants and maintaining engagement are, therefore, important considerations for citizen science projects (Meek & Zimmermann, 2016). Including blank photos in a dataset can increase engagement, leading to longer classification sessions (Bowyer

et al., 2015). This is thought to be due to the increased feeling of reward when an image containing an animal is seen (Bowyer et al., 2015). Other than this, little work has addressed how citizen scientists engage with different types of camera trap content. The sound and movement provided by videos could create a more immersive and engaging experience, but we are not aware of studies comparing how citizen scientists engage with photo *versus* video content.

Here, we present data collected from a camera trap survey in the Forest of Dean, UK. Paired cameras were placed across the site with one set to take photos and the other set to take video. We ran common camera trap analyses, including species richness, occupancy, activity level and detection rate, to determine whether there were any ecologically meaningful differences between the photo and video datasets. Data were uploaded to the citizen science platform, MammalWeb (www.mammalweb.org) (Hsing et al., *in press*), into parallel photo and video projects for classification. We compared classification accuracy between different lengths of photo sequence and video footage as well as looking at participation and engagement with both types of media. Our aim was to inform the choice of photos *versus* videos in camera settings in future ecological surveys, particularly those looking to engage citizen scientists in the classification process.

Materials and Methods

Study site – Forest of dean

The Forest of Dean ($51^{\circ}46'59.99''\text{N}$, $-2^{\circ}32'59.99''\text{W}$), extends across Gloucestershire, Herefordshire and Monmouthshire, UK, and consists of a mixture of broadleaf and conifer woodland, with patches of young trees. Our survey covered two forest areas, the larger covered approximately 65 km² with a smaller patch of approximately 20 km². Land is managed by Forestry England and fieldwork was undertaken in partnership with the Gloucestershire Wildlife Trust. The site is ecologically interesting to citizen scientists, as it is home to a variety of UK mammal species, including reintroduced populations of wild boar (*Sus scrofa*) (Dutton et al., 2015) and pine marten (*Martes martes*) (Macpherson & Wright, 2021). The mammal assemblage (Table S1) enabled ecologically meaningful comparisons between photo and video footage for species ranging in abundance, body size and activity level.

Fieldwork

A grid of points spaced 1 km apart was overlaid on a map of the Forest of Dean using QGIS (QGIS

Development Team, 2018). The grid covered the main forest area, plus the additional patch around Symonds Yat approximately 2.5 km west of the main block (Fig. 1). The main forest area was divided into four sections, each containing 15 points; the forest at Symonds Yat constituted a fifth area. Fifteen pairs of camera traps were deployed between the 19 November 2019 and 24 March 2020, with the survey conducted during this period to avoid dispersal and breeding/birthing seasons. This period is characterized by reduced vegetation cover, increasing the field of view available to camera traps, and reducing the risk of false triggers. Camera traps were placed as close as possible to the specified grid points, while ensuring sites were accessible for servicing, avoided the river, and, to reduce risk of damage or theft, were out of sight of public footpaths. Mean displacement of camera stations from planned locations was 94 m. Cameras were deployed at sites in the main forest for between 25 and 30 nights before being rotated to new sites. Camera traps in the fifth and final location were in place for either 14 or 15 nights, as surveying was curtailed by the Covid-19 pandemic and uncertainty over site accessibility.

At each camera station, a pair of Browning Recon Force Extreme (2017 model) camera traps was mounted side-by-side on a metal bracket. Cameras were placed at a mean height of 51 cm from the ground and secured to a suitable tree with a camera strap and a python lock. Signs were attached to each camera station, informing people of the purpose of the study, requesting that the cameras not be disturbed, and providing contact information. One camera trap from each pair was set to record bursts of eight photos, and the other to record 20-s videos. Half-way through deployment at each site, cameras were serviced, with the batteries checked and memory cards changed. Batteries used were either Varta alkaline or Ene-loop rechargeable, with the same type of battery used in both cameras in each pair. In order to account for any slight differences in camera position, the camera settings were switched half-way through deployment so that the camera taking photos would take videos, and *vice versa*. Cameras use an infrared flash and manufacturer specifications suggest a trigger speed of 0.4 s and 0.8 s recovery time for photos. Trigger interval was set to 5 s.

Data processing

To explore the impacts of video length and number of photos in a sequence, we created three versions of each camera trap sequence. Each video was clipped to create two additional versions, the first containing only the first 10 s and the second only the first 5 s of the clip. Images were first allocated into sequences with a greater than 10-s interval between images used to define a new sequence.

Images were then labelled according to sequence and image number within sequence for each camera deployment. This followed the standard image processing method for footage added to MammalWeb (Hsing et al., 2018) and allowed manipulation of the number of images from each sequence. Three versions of the photo sequence dataset were created: one containing the first eight images in each sequence, one containing the first three images in each sequence, and one containing only the first photo from each sequence. There were, thus, six versions of the dataset: short (5 s), medium (10 s) and long (20 s) videos and short (one photo), medium (three photos) and long (eight photos) photo sequences.

All videos and photo sequences were tagged in the open-source photo management tool 'digiKam' (<https://www.digikam.org>) to create expert classified datasets. Metadata and tags were extracted from photos and videos using the R package 'camtrapR' (Niedballa et al., 2016).

MammalWeb

Separate Forest of Dean photo and video projects were established on MammalWeb, with matching descriptions and display images, so that the projects differed only in containing either photos or videos for classification. To prevent potential bias from a user repeatedly classifying the same piece of footage, only one version (short, medium or long) of each video or photo sequence was uploaded to MammalWeb. Footage was uploaded to MammalWeb between the 20 February and 3 June 2020 and the two projects first became available for public classification on 9 March 2020. MammalWeb contributors can choose to participate in a specific project or select the 'classify all' button, which will then serve the user a selection of footage from all active projects on the site. Users were able to participate in the Forest of Dean projects as 'Spotters', which involves viewing the footage and adding tags to identify the species present. Participants classifying an animal in a photo sequence or video could supply additional information about the animal, including sex (male or female) and age (adult or juvenile). A default option of 'Unknown' was set for both age and sex. Spotters could also 'like' the sequence or video that they were viewing. Time, date and anonymous user ID number were recorded by the website for each classification submitted.

Classification data were downloaded on 11 June 2021, after all footage had been available for classification for at least 1 year and all footage had received at least one classification. Classifications submitted by citizen scientists via the MammalWeb platform were compared to expert classifications to determine whether each classification was correct. All classifications submitted were used in

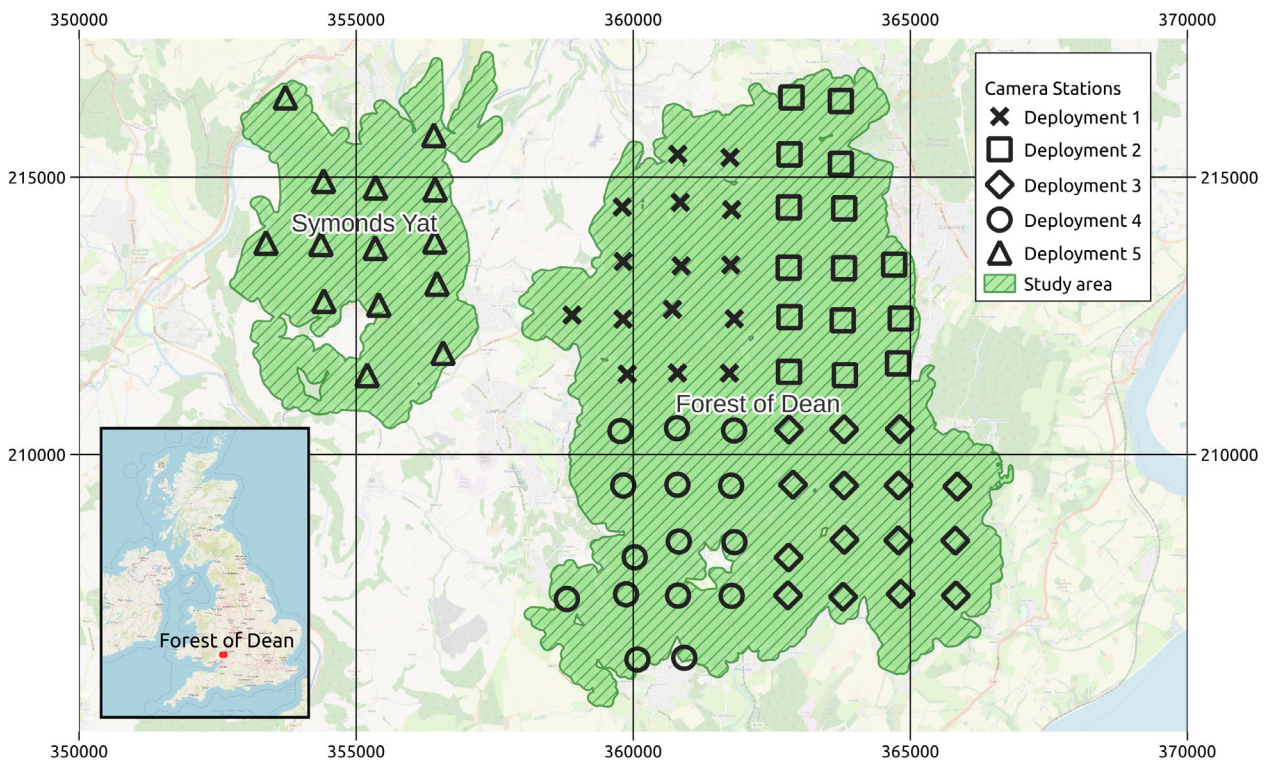


Figure 1. Forest area included in the study with locations of camera trap stations. Inset shows map of the UK with the location of the study site. Different deployments indicate the different time periods during which each area of the forest was surveyed. The time periods for each deployment are as follows: Deployment 1 = 19/11/2019–17/12/2019; Deployment 2 = 18/12/2019–13/01/2020; Deployment 3 = 13/01/2020–11/02/2020; Deployment 4 = 11/02/2020–09/03/2020; Deployment 5 = 09/03/2020–24/03/2020.

accuracy analysis, but data were split into discrete classification sessions to compare participation rates between the photo and video projects. The minimum requirement for a session was three or more consecutive classifications of footage within either the photo project only or video project only by one user within a 30-min period. This was designed to exclude classifications by participants who had selected 'classify all' and had randomly been served footage from the Forest of Dean projects. This ensured that sessions were analysed only where a user had specifically chosen to classify from that particular project. A session ended if there was a greater than 30-min interval between classifications submitted.

Data analysis

Ecological outputs

Analyses were conducted using R 3.6.2 (R Core Team, 2019). Diversity and richness estimates were generated for all six datasets using 'iNEXT' (Hsieh et al., 2020). More detailed ecological analysis focused on a subset of species. Selection criteria were mammals with

a body size greater than 250 g (a small rodent), which might reasonably be expected to be detected by our camera trap set up, and that yielded sufficient detections ($n = 40$ detections at a 30-min independence level). The species comprised Eurasian badger (*Meles meles*), red fox (*Vulpes vulpes*), fallow deer (*Dama dama*), Reeves muntjac (*Muntiacus reevesi*), roe deer (*Capreolus capreolus*), wild boar (*Sus scrofa*), European rabbit (*Oryctolagus cuniculus*) and grey squirrel (*Sciurus carolinensis*).

Detections were compared between the datasets by generating presence-absence data at each camera station for each of the focal species detected at that site, for every half-hour during the active period of each camera station. A half-hour period was chosen as this is a common interval for discerning independent detections (Burton et al., 2015; Rovero & Zimmermann, 2016; Sollmann, 2018). To assess differences in species detections between different media types and lengths, we used general linear mixed models (GLMMs) with a binomial distribution in 'lme4' (Bates et al., 2015). We defined separate models to check for the effect of length within each media type and then to check for differences between photo and video datasets. Analyses were

separated to avoid replication of the same datasets (across length variants) when comparing photos and videos. Species and camera station were specified as random factors. Half-hour time slot or 'survey period' was included as an additional random factor in the photo/video comparison model. Model comparison tables were generated using the MuMIn package to assess whether including media type or length-improved model performance (Barton, 2020).

For each focal species, activity levels (the proportion of time species spent active per day; see Rowcliffe et al., 2014) from each length of video or photo sequence were compared using the R-package 'activity' (Rowcliffe, 2021). Data were species detections with 5-min intervals between events. Even though some events may not be independent, this time period was chosen to trade-off the risk of non-independence with the aim of resolving activity to a reasonably fine scale. A Wald test was used to assess differences in activity level between the datasets produced by the different media and media lengths for each individual species and the combined dataset.

Detection histories were generated for each of the focal species from each length of video and photo sequence dataset. Detection histories were based on 24 h survey periods and were generated using 'camtrapR' (Niedballa et al., 2016) and then used to fit occupancy models using the package 'unmarked' (Fiske & Chandler, 2011). Outputs were back-transformed to give occupancy probability and the probability of detection. No covariates were included because our aim was the comparison between inferences from different media (and the paired design of the data collection ensured covariate differences did not introduce bias in these parameters), rather than to identify the factors driving occupancy of each species.

Citizen science classifications

Citizen science classifications were analysed using GLMMs to assess species classification accuracy, likelihood of submitting age and sex classification data, likelihood of footage receiving a 'like', and length of classification session.

Classification accuracy models were initially split into photo and video datasets, using sequence or video length as a fixed factor. Length did not have a detectable influence on classification accuracy for either photo or video, so it was excluded from models using the combined photo and video datasets. Classification accuracy models were then fitted to the full data set and to each focal species' data. The response variable for classification accuracy was a binary indicator of whether or not a citizen science classification matched the expert classification for that footage (1 if classifications matched, 0 if they did not). Age classification likelihood models were fitted to footage

containing any of the eight focal species as these are common mammals for which participants might reasonably be expected to identify adult or juvenile forms. Sex classification likelihood models were fitted only to footage containing one of the three deer species present at the study site, because only these species show clear sexual dimorphism. The response variables for age and sex classification models were also binary indicators of whether or not an age or sex classification had been provided alongside a species classification (1 if an age or sex classification was given, 0 if the corresponding classification was not given). Models determining the probability of liking footage used the full dataset. Again, the response variable was a binary indicator (1 if the footage was liked, 0 if it was not). Fixed factors in all classification models were media type (photo or video) and whether or not the flash was activated (i.e. whether footage was full colour or black and white/grey scale). Length of video or photo sequence was included as a fixed factor in age, sex and like classification models. Random factors were camera site ID and anonymous user ID, which were used in all classification models other than for accuracy of rabbit classifications where only anonymous user ID was used, due to the small number of different sites at which rabbits were detected. For each model, we first fitted the full model and then used the dredge function of the MuMIn package (Barton, 2020) to fit all possible additive variable permutations for the fixed factors described above. For the classification accuracy model a single interaction between media type and activation of flash was also included. Models were ranked according to AIC (Tables S5–S8). We then used effect size and p value of factors in the top model (using model averaging where there was more than one model within six AIC of the top) to assess strength and classical statistical significance of the effect of these predictors.

Length of session was measured in two ways: time difference between first and last classifications submitted in a session, and number of classifications submitted in a session. For both models, media type was used as fixed factor and anonymous user ID as a random factor.

Results

Camera traps taking photos were active and functional for a total of 1734 trap nights across 73 camera stations, while cameras taking videos were active for 1730 nights across 73 stations. The slight differences were due to camera malfunction. Of the original 75 planned sites, one was excluded due to bracken causing high levels of false triggers and filling memory cards, with no meaningful data collected from that site. A second site was excluded due to the theft of the cameras during the fourth rotation.

This meant only 14 sites were available for the fifth rotation; since this was the smaller forest patch, coverage was not greatly affected. Other than the one theft, cameras were left undamaged and were not tampered with at any other site, despite evidence they were detected by people on multiple occasions.

The displays showed photo cameras used, on average, 0.8% of battery per week with video cameras using 3.8%. Photo cameras recorded a mean of 415 individual photos per week and video cameras a mean of 29 videos. Each 20-s video had a file size of approximately 31 MB and each photo had a file size of approximately 0.8 MB. Based on the above rates of capture, an average of 332 MB and 899 MB of memory storage were needed per week for photos and videos, respectively.

Ecological outputs

Only data from 70 sites where video and photo cameras were active at the same time were used in analysis of ecological outputs. Diversity and species richness and species accumulation rates were similar for all datasets (Table 1; Fig. S1). The same 13 mammal species were detected in all lengths of photo sequence and video, and 18 and 15 bird species were detected in video and photo footage, respectively (Table S1).

Species detection

Trapping rates for all of the eight focal species were very similar across all lengths of photo sequence and video (Fig. 2). Neither length of video or photo sequence, nor

choice of video *versus* photo influenced the probability of detecting a species event; model selection showed no improvement in model performance when media type or length were included compared to null models ($\Delta AICs$ of null models = 0, $\Delta AICs$ of models including fixed effects <6; Table S2).

Activity

There was no difference between the activity levels derived from the different lengths of photo sequence and video for any of the focal species (P -values between 0.4 and 1; Fig. 3; Table S3).

Occupancy

Due to lack of difference in species detections between the different lengths of video and photo sequence, occupancy analyses used only the 20-s video and eight-photo sequence datasets to compare the two media. Occupancy and detection probability estimates were the same or very similar for photo and video. Where slight differences occurred, standard errors overlapped, indicating no meaningful difference in outputs (Table S4).

Citizen science classifications

5,326 photo sequences and 5,610 videos were uploaded to MammalWeb for classification. All photo sequences and videos received at least one classification and, overall, 17,474 photo and 12,429 video classifications were submitted.

Table 1. Species diversity estimates from a camera trap survey of the Forest of Dean, UK, where photo bursts and video data were collected simultaneously using a paired camera setup. Diversity estimates are given based on the first photo, first three photos and first eight photos in each burst, and for the first 5 s, first 10 s and full 20 s of each video clip. Observed differences were due to three species of bird detected in video footage but not in photos. These species were blue tit, marsh tit and tree creeper, all of which are small birds for which we would not consider our camera trap setup a suitable method for surveying.

	Five-second videos	Ten-second videos	Twenty-second videos	One-photo sequences	Three-photo sequences	Eight-photo sequences
Species richness observed	30	31	31	28	28	28
Species richness estimator (SE)	36.25 (7.55)	37.25 (7.55)	33.67 (3.48)	28.25 (0.73)	28.25 (0.73)	28.5 (1.32)
Shannon diversity observed	10.2	10.54	10.78	10.43	10.42	10.45
Shannon diversity estimator (SE)	10.3 (0.23)	10.64 (0.26)	10.87 (0.27)	10.51 (0.25)	10.49 (0.27)	10.52 (0.22)
Simpson diversity observed	7.46	7.6	7.74	7.57	7.55	7.56
Simpson diversity estimator (SE)	7.48 (0.18)	7.63 (0.18)	7.76 (0.18)	7.6 (0.18)	7.56 (0.18)	7.59 (0.16)

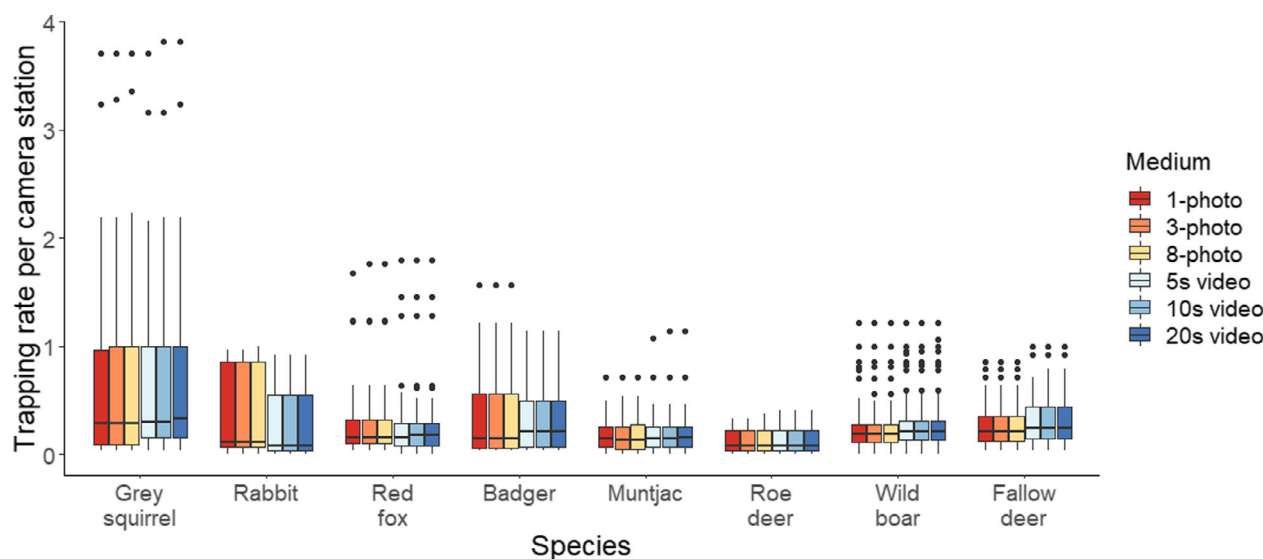


Figure 2. Boxplots of trapping rates for each of eight focal species detected during a camera trap survey of the Forest of Dean, UK, using 73 camera stations with paired cameras recording videos and photo bursts. Trapping rates were acquired for each camera trap station at which the species was detected, calculated as number of half-hour periods in which a species was detected, divided by number of operational camera days. Trapping rates were calculated using detections in datasets comprising: the first photo in every photo sequence; the first three photos in a sequence; the first eight photos in a sequence; the first 5 s of each video clip; the first 10 s; and 20-s video clips (the full length of each video). Bars and boxes indicate median and interquartile range (IQR), whiskers show the largest and smallest values within 1.5*IQR, with individual outliers plotted as solid fill circles.

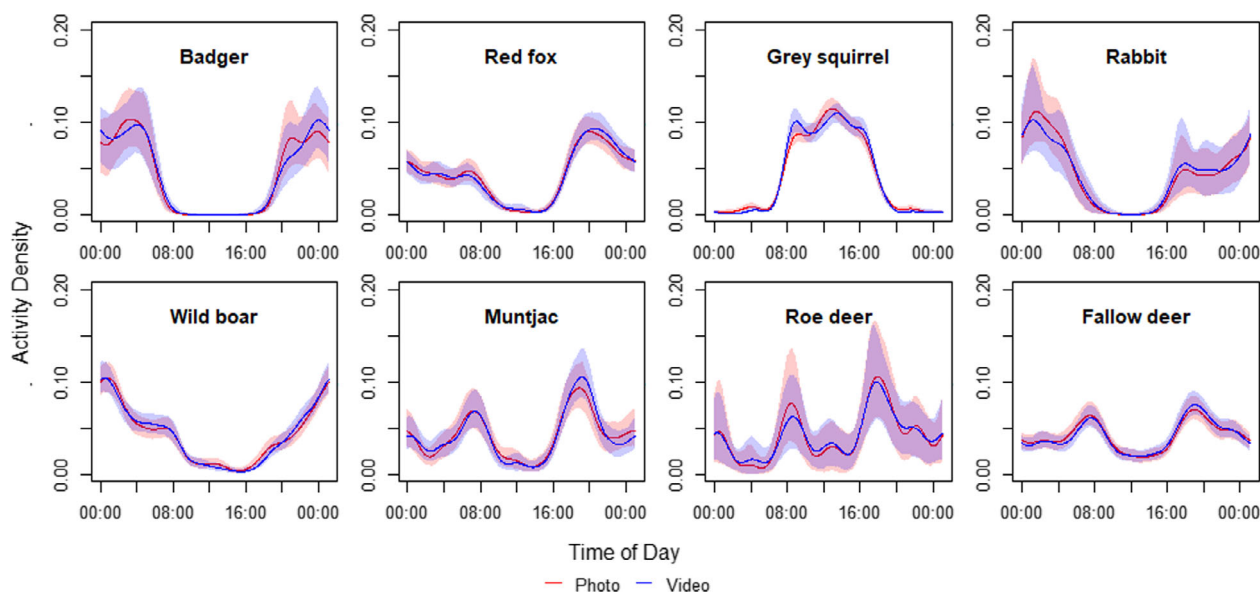


Figure 3. Activity distribution over a 24 hr. period with 95% confidence limits for each of eight focal species recorded during a camera trap survey of the Forest of Dean, UK, based on 8-photo and 20-s video datasets collected by paired cameras. Data were species detections with, at a minimum, 5-min intervals between events.

Species classification accuracy

Media type and flash activation both affected the probability of a citizen science classification being correct

(Table 2). Use of video had a positive effect with a higher probability of correct classifications than for photo sequences (Fig. 4; Table 2). Activation of flash had a negative effect on classification accuracy of both video and

Table 2. Coefficient estimate, standard error and p value for GLMMs examining classifications of camera trap data provided by citizen scientists. Fixed effects include media type (photo vs. video), whether the flash was activated or not, and the length of a photo sequence or video (short = the first photo taken in each sequence or the first 5 s of each video clip, medium = the first three photos in a sequence or first 10 s of a video and long = the first eight photos in a sequence and 20-s video clips. Response variable for accuracy models was a binary indicator of whether or not a citizen science classification matched the expert classification. Response variable for age, sex and likes was a binary indicator whether or not an age classification, a sex classification or a 'like' was provided alongside a species classification. Accuracy models were run on all citizen science classifications, then for footage containing each of eight focal species separately. Age models were run on data containing footage of one of eight focal species (badger, red fox, fallow deer, roe deer, muntjac deer, wild boar, rabbit, grey squirrel). Sex classification models were run on footage containing only deer (fallow, roe or muntjac). 'Like' models were run using all citizen science classifications. Where there are multiple models within six AIC units of the top model, model averaging is used to calculate coefficient estimates.

Model parameter	Estimate	SE	P value
All species classification accuracy			
Intercept	1.530	0.105	<0.001
Media (video)	0.906	0.109	<0.001
Flash (activated)	-0.412	0.053	<0.001
Media (video) X Flash(activated)	-0.009	0.063	0.887
Badger classification accuracy			
Intercept	1.668	0.429	<0.001
Media (video)	2.502	0.955	0.009
Red fox classification accuracy			
Intercept	1.564	0.788	0.0472
Media (video)	2.564	1.514	0.091
Flash (activated)	1.276	0.846	0.132
Media (video) X Flash(activated)	-1.750	1.559	0.262
Fallow deer classification accuracy			
Intercept	-0.385	0.327	0.240
Media (video)	1.237	0.253	<0.001
Flash (activated)	0.038	0.134	0.779
Media (video) X Flash(activated)	0.021	0.145	0.884
Reeves muntjac classification accuracy			
Intercept	1.467	0.477	0.002
Media (video)	1.001	0.469	0.033
Flash (activated)	-0.140	0.306	0.646
Media (video) X Flash(activated)	-0.256	0.268	0.906
Roe deer classification accuracy			
Intercept	2.752	0.637	<0.001
Media (video)	0.634	0.824	0.443
Flash (activated)	-1.206	0.481	0.012
Media (video) X Flash(activated)	-0.248	0.645	0.701
Wild boar classification accuracy			
Intercept	1.865	0.345	<0.001
Media (video)	1.133	0.367	0.002
Flash (activated)	0.200	0.287	0.487
Media (video) X Flash(activated)	0.011	0.276	0.968
Grey squirrel classification accuracy			
Intercept	2.060	0.157	<0.001

(Continued)

Table 2. Continued.

Model parameter	Estimate	SE	P value
Media (video)	0.795	0.206	<0.001
Flash (activated)	-1.630	0.126	<0.001
Media (video) X Flash(activated)	1.064	0.264	<0.001
Rabbit classification accuracy			
Intercept	2.318	0.446	<0.001
Media (video)	2.243	0.795	0.005
Age classification provision			
Intercept	-2.518	0.270	<0.001
Media (video)	1.104	0.087	<0.001
Flash (activated)	0.415	0.044	<0.001
Length (medium)	-0.240	0.122	0.046
Length (short)	-0.243	0.122	0.047
Sex classification provision			
Intercept	-0.797	0.225	<0.001
Media (video)	0.296	0.137	0.031
Flash (activated)	-0.005	0.047	0.918
Length (medium)	-0.268	0.143	0.062
Length (short)	-0.397	0.137	0.013
Likes given			
Intercept	-7.485	0.480	<0.001
Media (video)	2.132	0.258	<0.001
Flash (activated)	0.852	0.175	<0.001
Length (medium)	-0.246	0.281	0.382
Length (short)	-0.163	0.227	0.474

P values <0.05 are given in bold.

photo footage, with participants more likely to submit a correct classification when shown full colour than footage taken using infrared flash.

When analysed individually, all focal species were classified more accurately in video footage; this effect was statistically significant ($p < 0.05$) in models for all species apart from fox and roe deer (Table 2). Activation of flash was the only significant effect in the analysis of roe deer classification and had a negative influence on classification accuracy. For grey squirrel classifications, there was a significant interaction between media type and flash activation (Table 2) with the negative effect of flash activation being greater for photo footage than for video. Activation of flash was included in top models for fallow deer, red fox and wild boar with a positive effect, and in the top model for muntjac with a negative effect, but none of these were statistically significant (Table 2). Effect of flash could not be included in the models for badger or rabbit because the flash was activated for all detections.

Age and sex classification

Media type and length of video or photo sequence influenced whether an age or sex classification was provided

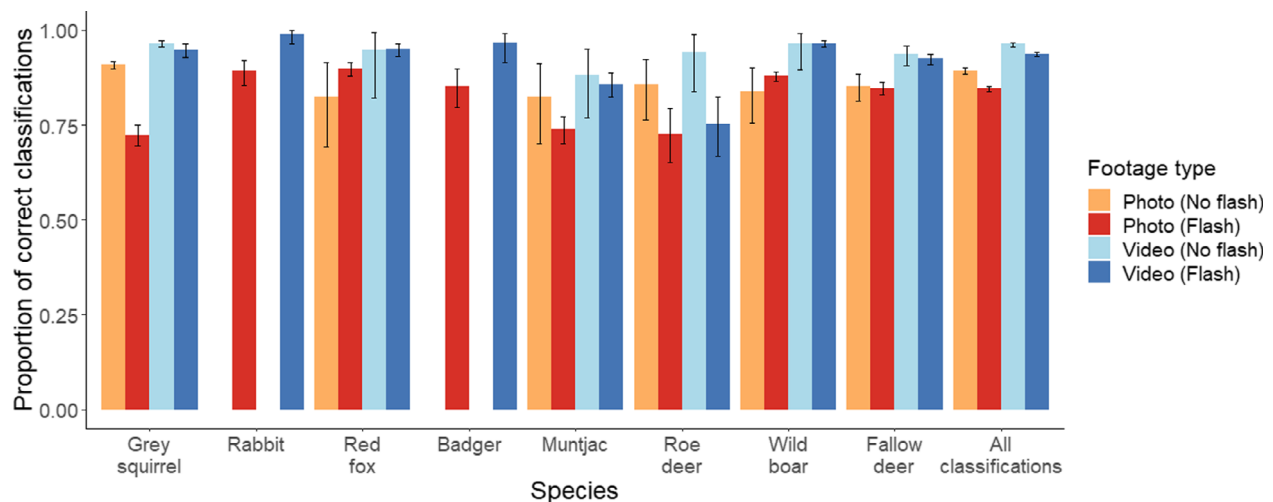


Figure 4. Proportion of citizen science classifications that matched expert classifications for each of eight focal species and for the full data set of classifications of footage collected in a camera trap survey of the Forest of Dean, UK. Proportions are given separately for photo and video footage, both with and without the flash activated. Error bars show 95% confidence intervals.

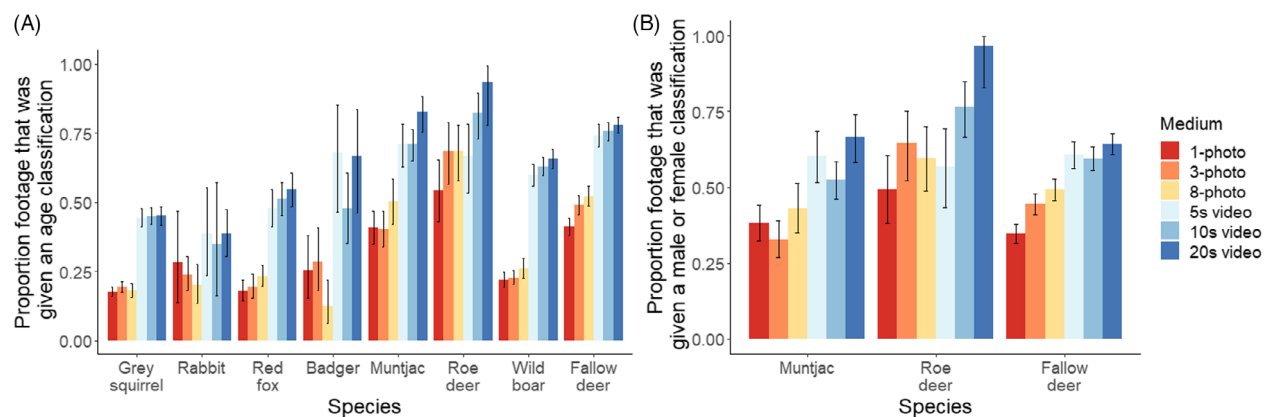


Figure 5. A) Proportion of camera trap footage for each of eight focal species for which an age category (adult or juvenile) was supplied by citizen scientists alongside a species classification and B) proportion of camera trap footage for three deer species for which a sex category (male or female) was supplied by citizen scientists alongside a species classification for different lengths of photo sequence and video. Error bars show 95% confidence intervals.

(Table 2; Fig. 5). Use of video had a positive effect on likelihood of an age or sex classification being submitted, as did the use of longer videos or photo sequences. Activation of flash was included in the top age classification model and had a slight positive effect on likelihood of an age classification being provided (Table 2; Fig. 5).

Citizen science engagement

A total of 183 contributors participated in at least one of the Forest of Dean projects; 126 users classified at least one sequence from the photo project, 117 users classified at least

one video from the video project; 60 users participated in both projects. Eighty-six users took part in a combined total of 411 classification sessions for the photo project and 59 users took part in 365 sessions for the video project. Video classification sessions had a mean duration of 27 min 5 s (range: 46 s–173 min 51 s) with a mean of 33 (range: 3–281) videos classified in a session. Photo classification sessions had a mean length of time of 26 min 34 s (range: 23 s–184 min 49 s) with 42 (range: 3–305) photo sequences classified per session. Mean time taken to classify was 42 s (range: 7.7 s – 259 s) for a photo sequence and 60 s (range: 15.3 s – 450 s) to classify a video.

Classification session length

More classifications were submitted per session for the photo project than for the video project. Model fit was improved by including media type as a predictor (Table S8) and the effect of media type was statistically significant ($p < 0.001$) in this model. However, there was still large variation and considerable overlap between the two projects (Fig. 6). For session duration, there was no difference in model performance when media type was included and model-averaged results show no significant difference ($p = 0.41$) between photo and video projects (Table S8; Fig. 6).

'Liked' footage

Media type and length of photo sequence or video had an effect on probability of footage being 'liked' by a citizen scientist (Table 2). The probability that a video was liked was 14 times greater than for photo sequences, with longer videos being most popular (Fig. 7). Single photos were more likely to be liked than both 3-photo and 8-photo sequences. Footage was more likely to be liked if the flash had been activated (Table 2).

Discussion

We tested whether camera traps set to video could collect ecological data of the same quality as that obtained by

cameras set to record photos. Moreover, we tested for differences in citizen science classification accuracy and engagement between photo and video datasets. We found that photo and video settings did not affect the ecological inferences, but that citizen scientists were more accurate and provided more detail when classifying video footage. Overall, there was a 9% difference in footage classification accuracy between photo (average accuracy 85%) and video (average accuracy 94%) footage when the flash was activated and a 7% difference when flash was not activated (average photo accuracy was 89%, video was 96%). The percentage of video footage containing one of the focal species that was given an age classification was twice that of age classifications given to photo footage (age classification provision for video was 61% and 30% for photo). The percentage of footage containing deer that was given a sex classification was also higher for video, with an 18% difference (sex classification provision for video was 63% and 45% for photo).

Ecological analyses

Based on datasets from expert classification, we found no differences in species diversity, occupancy, activity or detection rates between the photo and video datasets. This is in contrast to concerns that slow trigger and recovery speeds limit the value of video datasets, and is reassuring for those researchers wishing to use video based on its advantages in behavioural research

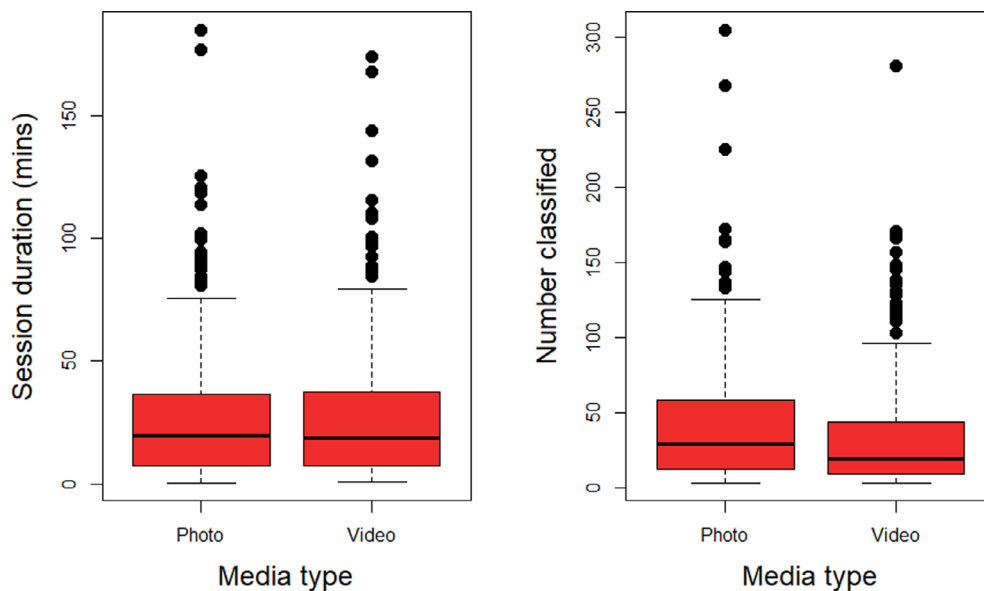


Figure 6. Boxplots of the length of time citizen scientists spent classifying camera trap footage from parallel photo and video projects in a continuous classification session, and the number of classifications that were submitted per classification session by MammalWeb participants. Bars and boxes indicate median and interquartile range (IQR), whiskers show the largest and smallest values within 1.5*IQR, with individual outliers plotted as solid fill circles.

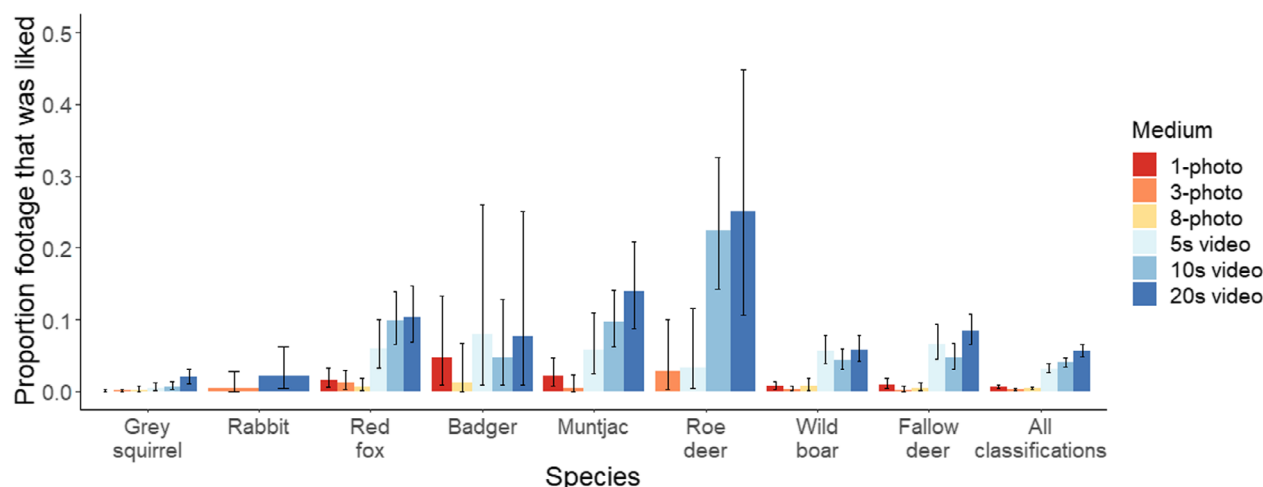


Figure 7. Proportion of camera trap footage of each species which was 'liked' by a citizen scientist while they were classifying species in that footage. Proportions are given for different lengths of photo sequence and video for footage containing one of eight focal species, and for all footage combined from a survey of the Forest of Dean, UK. Error bars show 95% confidence intervals.

(Caravaggi et al., 2017), individual ID (Reyes et al., 2017) and density estimation of unmarked individuals (Corlatti et al., 2020; Howe et al., 2017). While videos did take up more battery and memory, we found this had no real impact on our study, so need not be a deterrent to using video. Of course, in areas of higher animal activity, or where site access is necessarily infrequent, these factors may still need careful consideration. In such conditions, shorter videos could provide a good compromise, as we found that 5-s videos recorded species as reliably as 20-s videos.

We tested only one model of camera trap in one ecological system and other studies have found differences in performance between photo and video settings (Findlay et al., 2020; Palencia et al., 2019). As different camera traps perform differently (Apps & McNutt, 2018; Driessen et al., 2017; Palencia et al., 2021; Randler & Kalb, 2018; Rovero et al., 2013; Yajima & Nakashima, 2021), more research is needed to determine whether our findings will generalize across cameras and study species. Nevertheless, as camera traps continue to be upgraded, we would expect video performance to improve further over time.

Citizen science classifications

Overall, video footage was classified more accurately than photo sequences (Fig. 4). There was no difference in classification accuracy between videos of 5-s or 20-s length, highlighting that projects could benefit from improved accuracy even when short video clips are used. Citizen scientists were not only more accurate when classifying species in video, but they were also more likely to add

age and sex category classifications, suggesting this was easier to identify in videos than photos. Alternatively, the increase in age and sex classifications may reflect a deeper level of engagement with video footage. Preliminary examination of the data suggests that age and sex classifications, when provided, were accurate for both videos and photo sequences (approximately 95% correct) indicating that video footage could provide valuable additional demographic data. However, since the study was conducted when few juvenile animals were present and most male deer still had their antlers making them easier to recognize, further analyses across seasons are needed to determine more precisely the ability of citizen scientist to identify age and sex of animals accurately from photo and video footage. It is likely that more citizen science projects will start requiring human observers to move beyond species classifications; provision of additional detail is already evident in projects asking participants to identify age and sex (Thel et al., 2021) and individual ID (McCarthy et al., 2021; Tagg et al., 2018). Further research is needed to establish optimum camera settings for accurate identification of these traits, although video appears to offer clear benefits over traditional photos.

Confidence in accuracy and verification of citizen science ecological data is important for trust and acceptance of the value of the data (Baker et al., 2021; Freitag et al., 2016). The higher classification accuracy of video footage could, thus, increase the value of video datasets. Organizations harnessing citizen science invest considerable effort in data verification, with common methods including expert verification and community consensus; however, expert verification can be time consuming,

particularly for large camera trap datasets (Baker et al., 2021). Community consensus can be used to increase final classification accuracy but requires a large number of participants to gain enough classifications (Hsing et al., 2018; Swanson et al., 2016). Video footage could be advantageous, therefore, particularly for smaller projects, since community consensus could be achieved more easily.

Similar numbers of participants engaged with the MammalWeb photo and video projects although many spotters only participated in one project, suggesting they have different preferences for classifying photo or video footage. Participants spent similar amounts of time classifying per session in the photo and video projects but more photo sequences were classified per session than videos because people spent longer classifying each video. Participants spent around a minute, on average, classifying each video, which implies the video clip was watched multiple times. This could either be due to a determination to identify the video correctly or simply enjoyment of the clip. MammalWeb participants 'liked' more video footage than photos, suggesting that videos were more engaging and enjoyable to watch. Engaging enough participants is a challenge faced by many citizen science projects, particularly in light of the growing number of projects available to choose from (Follett & Strezov, 2015; Meek & Zimmermann, 2016; Pelacho et al., 2021; Willi et al., 2019). Using video could help with engagement while also gathering more accurate and detailed species classifications, thus generating higher quality datasets with fewer classifications needed per video. This would be advantageous to small projects as a high confidence in classification accuracy could be obtained with only a small number of participants viewing each video.

Video for camera trapping and citizen science

Detection probability in camera trap studies consists of several components, including the probability that an animal is identifiable in a photo or video (Findlay et al., 2020; Hofmeester et al., 2019). We found that citizen scientists classified videos more accurately, and it is likely that expert accuracy would also be improved through the use of video. Concerns over slow trigger speed reducing detection probability in videos were shown to be unfounded in this study. Therefore, due to increased animal identification accuracy, the use of video could increase detection probability, particularly for citizen science classified datasets. Consequently, we advocate for increased use of video in camera trap studies based on improved detection probability and citizen science engagement benefits.

More coordinated monitoring efforts are needed to identify global trends in biodiversity (Scotson et al., 2017; Steenweg et al., 2017) and there are now several initiatives combining camera trap footage from a range of participants in order to monitor wildlife across a wide area, such as MammalWeb (Hsing et al., *in press*), the Tropical Ecology Assessment and Monitoring (TEAM) Network (Rovero & Ahumada, 2017), and both Snapshot USA (Cove et al., 2021) and Snapshot Europe (<https://www.ab.mpg.de/358074/snapshot-europe>). These projects rely on large numbers of participants to collect enough data for meaningful ecological analysis. Our results show that it would be possible to combine datasets containing photo and video footage without risking loss of data quality. To support this, more camera trap data management software needs to be able to handle video footage. Artificial Intelligence (AI) is increasingly used to classify camera trap images and can be combined with citizen science for efficient data processing (Green et al., 2020). It is encouraging that AI for video classification is fundamentally no different to that for photo classification in that each video frame can be treated like a single photo (e.g. Chalmers et al., 2019), and that AI video analysis offers other potential advantages over photo classification (e.g. Johanns et al., 2022).

Integrating video could encourage a greater number of participants in collaborative projects, both from citizen science volunteers who prefer to use video and from researchers or practitioners undertaking surveys that fit other project criteria, but are currently unable to submit video footage. Allowing the use of video in such projects is of great importance for making efficient use of all camera trap data being collected, especially in the field of conservation, where resources are limited and data on even common species are still lacking (Croft et al., 2017).

Acknowledgements

We thank Forestry England for permission to deploy camera traps across the Forest of Dean. We also thank the Gloucestershire Wildlife Trust, in particular C. McNicol and J. Bridges for their assistance with fieldwork. This work is supported by a NERC IAPETUS DTP PhD scholarship for S. Green; grant number NE/L002590/1. We are grateful to the Associate Editor and two anonymous reviewers for comments that helped improve the final version of the manuscript.

References

- Anton, V., Hartley, S., Geldenhuis, A. & Wittmer, H.U. (2018) Monitoring the mammalian fauna of urban areas using

- remote cameras and citizen science. *Journal of Urban Ecology*, **4**, 1–9. <https://doi.org/10.1093/jue/juy002>
- Apps, P. & McNutt, J.W. (2018) Are camera traps fit for purpose? A rigorous, reproducible and realistic test of camera trap performance. *African Journal of Ecology*, **56**(4), 710–720. <https://doi.org/10.1111/aje.12573>
- Baker, E., Drury, J.P., Judge, J., Roy, D.B., Smith, G.C. & Stephens, P.A. (2021) The verification of ecological citizen science data: current approaches and future possibilities. *Citizen Science: Theory and Practice*, **6**(1), 1–14. <https://doi.org/10.5334/CSTP.351>
- Balestrieri, A., Cardarelli, E., Pandini, M., Remonti, L., Saino, N. & Prigioni, C. (2016) Spatial organisation of european badger (*Meles meles*) in northern Italy as assessed by camera-trapping. *European Journal of Wildlife Research*, **62** (2), 219–226. <https://doi.org/10.1007/S10344-016-0996-Y/TABLES/3>
- Barton, K. (2020). MuMIn: Multi-Model Inference. R package version 1.43.17. <https://CRAN.R-project.org/package=MuMIn>.
- Bates, D., Mächler, M., Bolker, B.M. & Walker, S.C. (2015) Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, **67**(1), 1–48. <https://doi.org/10.18637/JSS.V067.I01>
- Blount, J.D., Chynoweth, M.W., Green, A.M. & Şekercioğlu, Ç.H. (2021) Review: COVID-19 highlights the importance of camera traps for wildlife conservation research and management. *Biological Conservation*, **256**, 108984. <https://doi.org/10.1016/J.BIOCON.2021.108984>
- Bowyer, A., Mädel, V., Lintott, C., Swanson, A. & Miller, G. (2015) This image intentionally left blank: mundane images increase citizen science participation. In: *Proceedings of the 2015 conference on human computation*. San Diego, California, United States.
- Burton, A.C., Neilson, E., Moreira, D., Ladle, A., Steenweg, R., Fisher, J.T. et al. (2015) Wildlife camera trapping: A review and recommendations for linking surveys to ecological processes. *Journal of Applied Ecology*, **52**(3), 675–685. <https://doi.org/10.1111/1365-2664.12432>
- Caravaggi, A., Banks, P.B., Burton, A.C., Finlay, C.M.V., Haswell, P.M., Hayward, M.W. et al. (2017) A review of camera trapping for conservation behaviour research. *Remote Sensing in Ecology and Conservation*, **3**(3), 109–122. <https://doi.org/10.1002/rse2.48>
- Chalmers, C., Fergus, P., Wich, S. & Curbelo Montanez, A. (2019). *arXiv*[preprint].
- Corlatti, L., Sivieri, S., Sudolska, B., Giacomelli, S. and Pedrotti, L. (2020) A field test of unconventional camera trap distance sampling to estimate abundance of marmot populations. **2020**(4), wlb.00652. <https://doi.org/10.2981/WLB.00652>.
- Cove, M.V., Kays, R., Bontrager, H., Bresnan, C., Lasky, M., Frerichs, T. et al. (2021) SNAPSHOT USA 2019: a coordinated national camera trap survey of the United States. *Ecology*, **102**(6), e03353. <https://doi.org/10.1002/ECY.3353>
- Croft, S., Chauvenet, A.L.M. & Smith, G.C. (2017) 'A systematic approach to estimate the distribution and total abundance of British mammals', *PLoS ONE*. Edited by G. Umapathy, **12**(6), e0176339. <https://doi.org/10.1371/journal.pone.0176339>
- Delisle, Z.J., Flaherty, E.A., Nobbe, M.R., Wzientek, C.M. & Swihart, R.K. (2021) Next-generation camera trapping: systematic review of historic trends suggests keys to expanded research applications in ecology and conservation. *Frontiers in Ecology and Evolution*, **9**, 97. <https://doi.org/10.3389/FEVO.2021.617996/BIBTEX>
- Driessen, M.M., Jarman, P.J., Troy, S. & Callander, S. (2017) Animal detections vary among commonly used camera trap models. *Wildlife Research*, **44**(4), 291. <https://doi.org/10.1071/WR16228>
- Dutton, J.S., Clayton, H.T. & Evans, S.M. (2015) The social aspects of wild boar in the Forest of dean. In: *Unpublished report for the forestry commission by the University of Worcester*[preprint].
- Egna, N., O'Connor, D., Stacy-Dawes, J., Tobler, M.W., Pilfold, N., Neilson, K. et al. (2020) Camera settings and biome influence the accuracy of citizen science approaches to camera trap image classification. *Ecology and Evolution*, **10**(21), 11954–11965. <https://doi.org/10.1002/ECE3.6722>
- Findlay, M.A., Briers, R.A. & White, P.J.C. (2020) Component processes of detection probability in camera-trap studies: understanding the occurrence of false-negatives. *Mammal Research*, **65**(2), 167–180. <https://doi.org/10.1007/s13364-020-00478-y>
- Fiske, I.J. & Chandler, R.B. (2011) unmarked: an R package for fitting hierarchical models of wildlife occurrence and abundance'. *Journal of Statistical Software*, **43**, 1–23. <https://doi.org/10.1002/wics.10>
- Follett, R. & Strezov, V. (2015) An analysis of citizen science based research: usage and publication patterns. *PLoS ONE*, **10**(11), 143687. <https://doi.org/10.1371/journal.pone.0143687>
- Freitag, A., Meyer, R. & Whiteman, L. (2016) Strategies employed by citizen science programs to increase the credibility of their data. *Citizen Science: Theory and Practice*, **1**(1), 2. <https://doi.org/10.5334/cstp.6>
- Glen, A.S., Cockburn, S., Nichols, M., Ekanayake, J. & Warburton, B. (2013) Optimising camera traps for monitoring small mammals. *PLoS ONE*, **8**(6), e67940. <https://doi.org/10.1371/JOURNAL.PONE.0067940>
- Glover-Kapfer, P., Soto-Navarro, C.A. & Wearn, O.R. (2019) Camera-trapping version 3.0: current constraints and future priorities for development. *Remote Sensing in Ecology and Conservation*, **5**, 1–15. <https://doi.org/10.1002/rse2.106>
- Green-Barber, J.M. & Old, J.M. (2018) Is camera trap videography suitable for assessing activity patterns in eastern grey kangaroos? *Pacific Conservation Biology*, **24**(2), 134–141. <https://doi.org/10.1071/PC17051>

- Green, S.E., Rees, J.P., Stephens, P.A., Hill, R.A. & Giordano, A.J. (2020) Innovations in camera trapping technology and approaches: the integration of citizen science and artificial intelligence. *Animals*, **10**(1), 132. <https://doi.org/10.3390/ani10010132>
- Herrera, D.J., Moore, S.M., Herrmann, V., McShea, W.J. & Cove, M.V. (2021) A shot in the dark: White and infrared LED flash camera traps yield similar detection probabilities for common urban mammal species. *Hystrix, Italian Journal of Mammalogy*, **32**(1), 72–75. <https://doi.org/10.4404/hystrix>
- Hofmeester, T.R., Crooms, J.P.G.M., Odden, J., Andrén, H., Kindberg, J. & Linnell, J.D.C. (2019) Framing pictures: A conceptual framework to identify and correct for biases in detection probability of camera traps enabling multi-species comparison. *Ecology and Evolution*, **9**(4), 2320–2336. <https://doi.org/10.1002/ece3.4878>
- Howe, E.J., Buckland, S.T., Després-Einspenner, M.-L. and Kühl, H.S. (2017) 'Distance sampling with camera traps. *Methods in Ecology and evolution*, edited by J. Matthiopoulos, **8**(11), 1558–1565. <https://doi.org/10.1111/2041-210X.12790>
- Hsieh, T.C., Ma, K.H. and Chao, A. (2020) iNEXT: iNterpolation and EXTrapolation for species diversity. R package version 2.0.20. <http://chao.stat.nthu.edu.tw/wordpress/software-download/>
- Hsing, P.-Y., Bradley, S., Kent, V.T., Hill, R.A., Smith, G.C., Whittingham, M.J., Cokill, J., Crawley, D., Stephens, P.A. and Stephens, P.A. (2018) Economical crowdsourcing for camera trap image classification. *Remote Sensing in Ecology and Conservation*, Edited by M. Rowcliffe and O. Wearn, **4**, 361–374. <https://doi.org/10.1002/rse2.84>
- Hsing, P.-Y., Hill, R.A., Smith, G.C., Bradley, S., Green, S.E., Kent, V.T. et al. *Large-scale mammal monitoring: the potential of a citizen science camera-trapping project in the UK*. Ecological Solutions and Evidence (in press).
- Jacobs, C.E. & Ausband, D.E. (2018) An evaluation of camera trap performance-what are we missing and does deployment height matter? *Remote Sensing in Ecology and Conservation*, **4**(4), 352–360. <https://doi.org/10.1002/rse2.81>
- Janisch, J., Mitoyen, C., Perinot, E., Spezie, G., Fusani, L. & Quigley, C. (2021) Video recording and analysis of avian movements and behavior: insights from courtship case studies. *Integrative and Comparative Biology*, **61**(4), 1378–1393. <https://doi.org/10.1093/ICB/ICAB095>
- Johanns, P., Hauke, T. & Steinhage, V. (2022) Automated distance estimation for wildlife camera trapping. *Ecological Informatics*, **70**, 101734. <https://doi.org/10.1016/J.ECOINF.2022.101734>
- Lasky, M., Parsons, A., Schuttler, S., Mash, A., Larson, L., Norton, B. et al. (2021) Candid critters: challenges and solutions in a large-scale citizen science camera trap project. *Citizen Science: Theory and Practice*, **6**(1), 1–17. <https://doi.org/10.5334/CSTP.343>
- Macpherson, J. & Wright, P. (2021) *Long-term strategic recovery plan for pine martens in Britain*. Vincent Wildlife Trust, Ledbury, UK.
- McCarthy, M.S., Stephens, C., Dieguez, P., Samuni, L., Després-Einspenner, M.-L., Harder, B. et al. (2021) Chimpanzee identification and social network construction through an online citizen science platform. *Ecology and Evolution*, **11**(4), 1598–1608. <https://doi.org/10.1002/ECE3.7128>
- McShea, W.J., Forrester, T., Costello, R., He, Z., Kays, R., McShea Tavis Forrester Robert Costello Zhihai He Roland Kays et al. (2016) Volunteer-run cameras as distributed sensors for macrosystem mammal research. *Landscape Ecology*, **31**(1), 55–66. <https://doi.org/10.1007/s10980-015-0262-9>
- Medeiros, K., Bastos, M., Jones, G. & Bezerra, B. (2019) Behavior, diet, and habitat use by blonde capuchin monkeys (*Sapajus flavius*) in a coastal area prone to flooding: direct observations and camera trapping. *International Journal of Primatology*, **40**(4–5), 511–531. <https://doi.org/10.1007/S10764-019-00103-Z/FIGURES/6>
- Meek, P. & Zimmermann, F. (2016) Camera traps and public engagement. In: Rovero, F. & Zimmermann, F. (Eds.) *Camera trapping for wildlife research*. Exeter, UK: Pelagic Publishing, pp. 219–236.
- Meek, P.D., Ballard, G.A. and Falzon, G. (2016) The higher you go the less you will know: placing camera traps high to avoid theft will affect detection, *Remote Sensing in Ecology and Conservation*. Edited by M. Rowcliffe and C. De Angelo, **2**(4), 204–211. <https://doi.org/10.1002/rse2.28>
- Nakashima, Y., Fukasawa, K. and Samejima, H. (2018) 'Estimating animal density without individual recognition using information derivable exclusively from camera traps. *Journal of Applied Ecology*. Edited by P. Stephens, **55**(2), 735–744. <https://doi.org/10.1111/1365-2664.13059>
- Niedballa, J., Sollmann, R., Courtiol, A. and Wilting, A. (2016) camtrapR: an R package for efficient camera trap data management, *Methods in Ecology and Evolution*. Edited by P. Jansen, **7**(12), 1457–1462. <https://doi.org/10.1111/2041-210X.12600>
- Palencia, P., Vicente, J., Barroso, P., Barasona, J., Soriguer, R.C. & Acevedo, P. (2019) Estimating day range from camera-trap data: the animals' behaviour as a key parameter. *Journal of Zoology*, **309**(3), 182–190. <https://doi.org/10.1111/jzo.12710>
- Palencia, P., Vicente, J., Soriguer, R.C. & Acevedo, P. (2021) Towards a best-practices guide for camera trapping: assessing differences among camera trap models and settings under field conditions. *Journal of Zoology*, **316**, 197–208. <https://doi.org/10.1111/JZO.12945>
- Pelacho, M., Ruiz, G., Sanz, F., Tarancón, A. & Clemente-Gallardo, J. (2021) Analysis of the evolution and

- collaboration networks of citizen science scientific publications. *Scientometrics*, **126**(1), 225–257. <https://doi.org/10.1007/S11192-020-03724-X/FIGURES/20>
- QGIS Development Team. (2018) QGIS geographic information system. Open source geospatial foundation project. <http://qgis.osgeo.org>
- R Core Team. (2019) *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>
- Randler, C. & Kalb, N. (2018) Distance and size matters: A comparison of six wildlife camera traps and their usefulness for wild birds. *Ecology and Evolution*, **8**(14), 7151–7163. <https://doi.org/10.1002/ece3.4240>
- Reyes, A., Rodríguez, D., Reyes-Amaya, N., Rodríguez-Castro, D., Restrepo, H. & Urquijo, M. (2017) Comparative efficiency of photographs and videos for individual identification of the Andean bear (*Tremarctos ornatus*) in camera trapping. *Therya*, **8**(1), 83–87. <https://doi.org/10.12933/therya-17-453>
- Rovero, F. & Ahumada, J. (2017) The tropical ecology, assessment and monitoring (TEAM) network an early warning system for tropical rain forests. *Science of the Total Environment*, **574**, 914–923. <https://doi.org/10.1016/j.scitotenv.2016.09.146>
- Rovero, F. & Zimmermann, F. (2016) In: Rovero, F. & Zimmermann, F. (Eds.) *Camera trapping for wildlife research*. Exeter, UK: Pelagic Publishing.
- Rovero, F., Zimmermann, F., Berzi, D. & Meek, P. (2013) Which camera trap type and how many do I need? A review of camera features and study designs for a range of wildlife research applications. *Hystrix*, **24**(2), 148–156. <https://doi.org/10.4404/hystrix-24.2-6316>
- Rowcliffe, M., Kays, R., Kranstauber, B., Carbone, C. & Jansen, P.A. (2014) Quantifying levels of animal activity using camera trap data. *Methods in Ecology and Evolution*, **5**, 1170–1179.
- Rowcliffe, M. (2021). Activity: animal activity statistics. R package version 1.3.1. <https://CRAN.R-project.org/package=activity>.
- Scotson, L., Johnston, L.R., Iannarilli, F., Wearn, O.R., Mohd-Azlan, J., Ming Wong, W. et al. (2017) Best practices and software for the management and sharing of camera trap data for small and large scales studies. *Remote Sensing in Ecology and Conservation*, **3**(3), 158–172. <https://doi.org/10.1002/rse2.54>
- Seidlitz, A., Bryant, K.A., Armstrong, N.J., Calver, M. & Wayne, A.F. (2021) Optimising camera trap height and model increases detection and individual identification rates for a small mammal, the numbat (*Myrmecobius fasciatus*). *Australian Mammalogy*, **43**(2), 226–234. <https://doi.org/10.1071/AM20020>
- Sollmann, R. (2018) A gentle introduction to camera-trap data analysis. *African Journal of Ecology*, **56**(4), 740–749. <https://doi.org/10.1111/aje.12557>
- Steenweg, R., Hebblewhite, M., Kays, R., Ahumada, J., Fisher, J.T., Burton, C. et al. (2017) Scaling-up camera traps: monitoring the planet's biodiversity with networks of remote sensors. *Frontiers in Ecology and the Environment*, **15**(1), 26–34. <https://doi.org/10.1002/fee.1448>
- Swanson, A., Kosmala, M., Lintott, C. & Packer, C. (2016) A generalized approach for producing, quantifying, and validating citizen science data from wildlife images. *Conservation Biology*, **30**(3), 520–531. <https://doi.org/10.1111/cobi.12695>
- Swinnen, K.R.R., Reijnders, J., Breno, M. and Leirs, H. (2014) A novel method to reduce time investment when processing videos from camera trap studies, *PLoS ONE*. Edited by J. Goodrich, **9**(6), e98881. <https://doi.org/10.1371/journal.pone.0098881>.
- Tagg, N., McCarthy, M., Dieguez, P., Bocksberger, G., Willie, J., Mundry, R. et al. (2018) Nocturnal activity in wild chimpanzees (pan troglodytes): evidence for flexible sleeping patterns and insights into human evolution. *American Journal of Physical Anthropology*, **166**(3), 510–529. <https://doi.org/10.1002/ajpa.23478>
- Thel, L., Chamaillé-Jammes, S., Keurinck, L., Catala, M., Packer, C., Huebner, S.E. et al. (2021) Can citizen science analysis of camera trap data be used to study reproduction? *Lessons from Snapshot Serengeti program*, **2021**(2), wlb.00833. <https://doi.org/10.2981/WLB.00833>
- Willi, M., Pitman, R.T., Cardoso, A.W., Locke, C., Swanson, A., Boyer, A. et al. (2019) Identifying animal species in camera trap images using deep learning and citizen science. *Methods in Ecology and Evolution*, **10**(1), 80–91. <https://doi.org/10.1111/2041-210X.13099>
- Yajima, G. & Nakashima, Y. (2021) Can video traps reliably detect animals? Implications for the Density Estimation of Animals without Individual Recognition, **46**(3), 189–195. <https://doi.org/10.3106/MS2020-0055>
- Young, S., Rode-Margono, J. & Amin, R. (2018) Software to facilitate and streamline camera trap data management: a review. *Ecology and Evolution*, **8**, 9947–9957. <https://doi.org/10.1002/ece3.4464>

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Figure S1. Sample-size-based rarefaction and extrapolation curves for the 8-photo and 20-s video datasets. Data input was incidence rate and the sampling unit was one camera day at one camera station. Species accumulation rates were almost identical for video and photo surveys and reached a complete species list with the same number of sampling units required.

Table S1. Full list of species detected. *Species only detected in video footage. All other species detected in both photo and video

Table S2. Model comparison of GLMMS used to test for influence of video length, photo sequence length or media type on detections of the eight focal species. Comparison is given between null models and those including either media type or length as a fixed factor with the coefficient estimate, standard error and p value outputs for fixed factors from each full model (described in bold).

Table S3. Results of a Wald test comparing activity level derived from the short (5 s) and long (20 s) video data sets and short (one photo) and long (eight photo) photo sequence data sets for eight focal species. Long (20s videos and 8-photo sequence) data sets were used in video *vs.* photo comparisons. Difference = differences between activity level estimates produced by the difference data sets, SE = Standard error of the differences, W = Wald statistics, P = p values

Table S4. Occupancy and detection probabilities with standard errors (SE) for focal species for each media type and length data set based on expert classifications. *SE not available for these estimates due to lack or variability in presence across sites, that is this species was detected at all or almost all study sites.

Table S5. Species classification accuracy model comparisons showing all models within six AIC units of the top model for the global species set and individual focal species. Full model is given in bold with results of model selection below.

Table S6. Age classification model comparison showing all models within six AIC units of the top model. Full model is given in bold with results of model selection below.

Table S7. Sex classification model comparison showing all models within six AIC units of the top model. Full model is given in bold with results of model selection below.

Table S8. Results of GLMM model comparison for classification sessions measured in number of classifications and in length of time spent submitting classifications. Full model is given in bold with results of model selection below.

Table S9. 'Liked' footage model comparison showing all models within six AIC units of the top model. Full model is given in bold with results of model selection below.

Table S10. Breakdown of the number of photo sequences/videos in species categories for the different data sets.

Table S11. Number of detections of focal species used in activity analysis.