RESEARCH ARTICLE

A semi-automated camera trap distance sampling approach for population density estimation

Maik Henrich[1,2] (iD), Mercedes Burgueño[2], Jacqueline Hoyer[3], Timm Haucke[4] (iD), Volker Steinhage[4,†], Hjalmar S. Kühl[3,5,6,†] & Marco Heurich[1,2,7,†]

[1]Chair of Wildlife Ecology and Wildlife Management, University of Freiburg, Freiburg, Germany
[2]Department of National Park Monitoring and Animal Management, Bavarian Forest National Park, Grafenau, Germany
[3]German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Leipzig, Germany
[4]Institute of Computer Science IV, University of Bonn, Bonn, Germany
[5]Senckenberg Museum für Naturkunde Görlitz, Görlitz, Germany
[6]International Institute Zittau, Technische Universität Dresden, Zittau, Germany
[7]Institute for Forest and Wildlife Management, Inland Norway University of Applied Sciences, Koppang NO-34, Norway

Abstract

Camera traps have become important tools for the monitoring of animal populations. However, the study-specific estimation of animal detection probabilities is key if unbiased abundance estimates of unmarked species are to be obtained. Since this process can be very time-consuming, we developed the first semi-automated workflow for animals of any size and shape to estimate detection probabilities and population densities. In order to obtain observation distances, a deep learning algorithm is used to create relative depth images that are calibrated with a small set of reference photos for each location, with distances then extracted for animals automatically detected by MegaDetector 4.0. Animal detection by MegaDetector was generally independent of the distance to the camera trap for 10 animal species at two different study sites. If an animal was detected both manually and automatically, the difference in the distance estimates was often minimal at a distance about 4 m from the camera trap. The difference increased approximately linearly for larger distances. Nonetheless, population density estimates based on manual and semi-automated camera trap distance sampling workflows did not differ significantly. Our results show that a readily available software for semi-automated distance estimation can reliably be used within a camera trap distance sampling workflow, reducing the time required for data processing, by > 13-fold. This greatly improves the accessibility of camera trap distance sampling for wildlife research and management.

Introduction

The mass extinction of species is progressing at a rate unprecedented in Earth's history (Diáz et al., 2019). It is mainly driven by the synergistic effects of anthropogenic activities, especially the destruction of habitats, the overexploitation of populations and climate change. While the conservation of threatened species is a priority because of their intrinsic value as well as their cultural and economic benefits for humans (Diáz et al., 2019), effective population control measures are demanded for animal populations that fuel human–wildlife conflicts, because of damage to forestry, agriculture and private property (Massei et al., 2011). These problems have never been as relevant as they are today and they have highlighted the need for readily available methods to monitor the population status and developments of these species, in order to evaluate the effectiveness of management measures and adjust them if necessary.

In principle, the easiest way to estimate animal abundance is to count individuals in a random sample of plots, with subsequent extrapolation to the whole survey area. While this approach is straightforward for sessile organisms, the high mobility of vertebrates means that an observer would need to count all animals in all plots at one instance in time and with a high degree of certainty, to avoid double-counting and missing animals. A number of methods for the counting of vertebrates have been developed, ranging from the recording of indirect signs of animal presence to the capture and marking of individuals (Schwarz & Seber, 1999). Distance sampling, which models the probability of observing an animal dependent on its distance from a transect point or line (detection probability), is among the most widely applied approaches to estimate animal abundance (Buckland et al., 2004) and its application goes beyond traditional human observer-based surveys of animals along transects. Distance sampling theory can be applied to data collected by autonomous recorders, such as passive acoustic sensors or camera traps (CTs), which observe their surroundings continuously, in defined time intervals or following a trigger event. When these recorders are randomly placed with respect to landscape features, they can be used for special forms of point transect distance sampling. While acoustic sensors are mainly used for songbirds and marine mammals (Marques et al., 2013), camera trap distance sampling (CTDS) is a promising approach for many terrestrial bird and mammal species (Howe et al., 2017). Distance sampling functions are also useful to estimate the size of the sampled area for alternative methods to estimate population densities from CT data, such as the Random Encounter Model and the Random Encounter

and Staying Time Model (Nakashima et al., 2018; Rowcliffe et al., 2008).

While camera trapping requires little fieldwork, the generated datasets often consist of several thousand photos or videos whose manual processing requires enormous amounts of time and effort. In the Snapshot Serengeti project, more than 30 000 h were needed to classify and count animals on 5.5 million photos (Norouzzadeh et al., 2018). Estimation of the distances to the observed animals, necessary to estimate detection probabilities, further increases the workload. Observation distances are most often either measured in the field, using a tape measure and compass, after the positions of observed animals have been reconstructed based on natural landmarks (Rowcliffe et al., 2011) or estimated by eye, based on reference images of an object photographed at a known, regular interval (Hofmeester et al., 2017). Comparison of the position of one animal with 4–15 different positions of the reference object takes, on average, 12 sec under optimal conditions (Haucke et al., 2022). If additional manual steps are needed to load and align photos or if the horizontal observation distance to the centre of the field of view (FOV) must be considered, then the process may take several minutes.

The automation of these tasks is therefore desirable. This has been hindered by the fact that the animals of interest are often only partly visible, lighting conditions are frequently challenging, and the background can be highly dynamic (Norouzzadeh et al., 2018). The first approaches to automatically detect animal presence or absence were based on motion detection in videos and background identification in time lapse photos (Price Tack et al., 2016; Swinnen et al., 2014). Thereupon successive video frames were used to localize and count animals (He et al., 2016). The development of MegaDetector, an object detection model trained on several hundred thousand bounding boxes from a variety of environments (Beery et al., 2019), allowed the same for single photos. The automated identification of animal species gained interest following the introduction of conventional image classification algorithms for use with cropped photos (Yu et al., 2013), leading to the application of deep convolutional neural networks with increasing accuracy (Norouzzadeh et al., 2018). Nonetheless, obtaining accurate estimates of observation distances for large datasets remains problematic, with the exception of recently described photogrammetric approaches to estimate the distances of objects of known size (Leorna et al., 2022; Zuleger et al., 2022). Such an approach was already adopted for population density estimation of 35 species on Borneo based on a pole as reference object, but it still required the manual annotation of pixel positions for observed animals (Wearn et al., 2022). Haucke

et al. (2022) developed a semi-automated approach for estimating the distances to animals of any size from a single photo, with only a minimum of two reference images of an object at a known distance required.

In this study, we incorporated the semi-automated estimation of observation distances in a workflow to estimate the population densities of ten animal species *via* CTDS: European badger *Meles meles*, Eurasian beaver *Castor fibre*, domestic cat *Felis catus*, European hare *Lepus europaeus*, North American raccoon *Procyon lotor*, red deer *Cervus elaphus*, red fox *Vulpes vulpes*, roe deer *Capreolus capreolus*, water rail *Rallus aquaticus* and wild boar *Sus scrofa*. Data were obtained from two camera trapping studies that employed different CT models and settings, included different subsets of the aforementioned species, and covered a broad range of habitats within a wetland nature reserve in Eastern Germany and a mountainous national park in Southern Germany. In the first step (1), we assessed the recall (% of photographed animals detected) and precision (% of true positives in animal detections) of MegaDetector, dependent on the distance to the CT. For animals that were correctly automatically detected (2), we then investigated the relationship between different ranges of semi-automated distance estimates and their divergence from the manual estimates to test the hypotheses that their agreement is the highest closest to the CT and decreases linearly with distance. Finally, (3) we compared the CTDS-based abundance estimates obtained from manually versus semi-automatically derived observation distances and assessed the overlap of their confidence intervals.

Materials and Methods

Study area descriptions

The conservation area "Hintenteiche bei Biesenbrow" (1.04 km²) is located in the north-eastern section of the Biosphere Reserve Schorfheide Chorin (S-C) (Brandenburg, Germany). Over half of the area consists of three ponds, surrounded, respectively, by alder-ash forests, alluvial softwood forests and grasslands within a farmland area. Our study site included 0.461 km² of the total area, excluding ponds and farmland. The Bavarian Forest National Park (BFNP, 242.3 km²) is located 448 km to the south (Fig. 1). Its forests, which are dominated by spruce (Cailleret et al., 2014), cover 90% of the area.

Study design

The 27 CTs placed in the S-C were separated from each other by a distance of at least 150 m. The 50 CTs in the BFNP had a minimum spacing of 1 km.

All potential sites located in water or human settlements were excluded. The CTs were attached to trees at a height of 50–80 cm and oriented north to avoid backlight. The CT settings are listed in Table 1. The CTs in the BFNP and S-C were active throughout the summer (June–August) of 2018 and 2019 respectively. Because of their non-invasive nature, these CT studies did not require ethical approval.
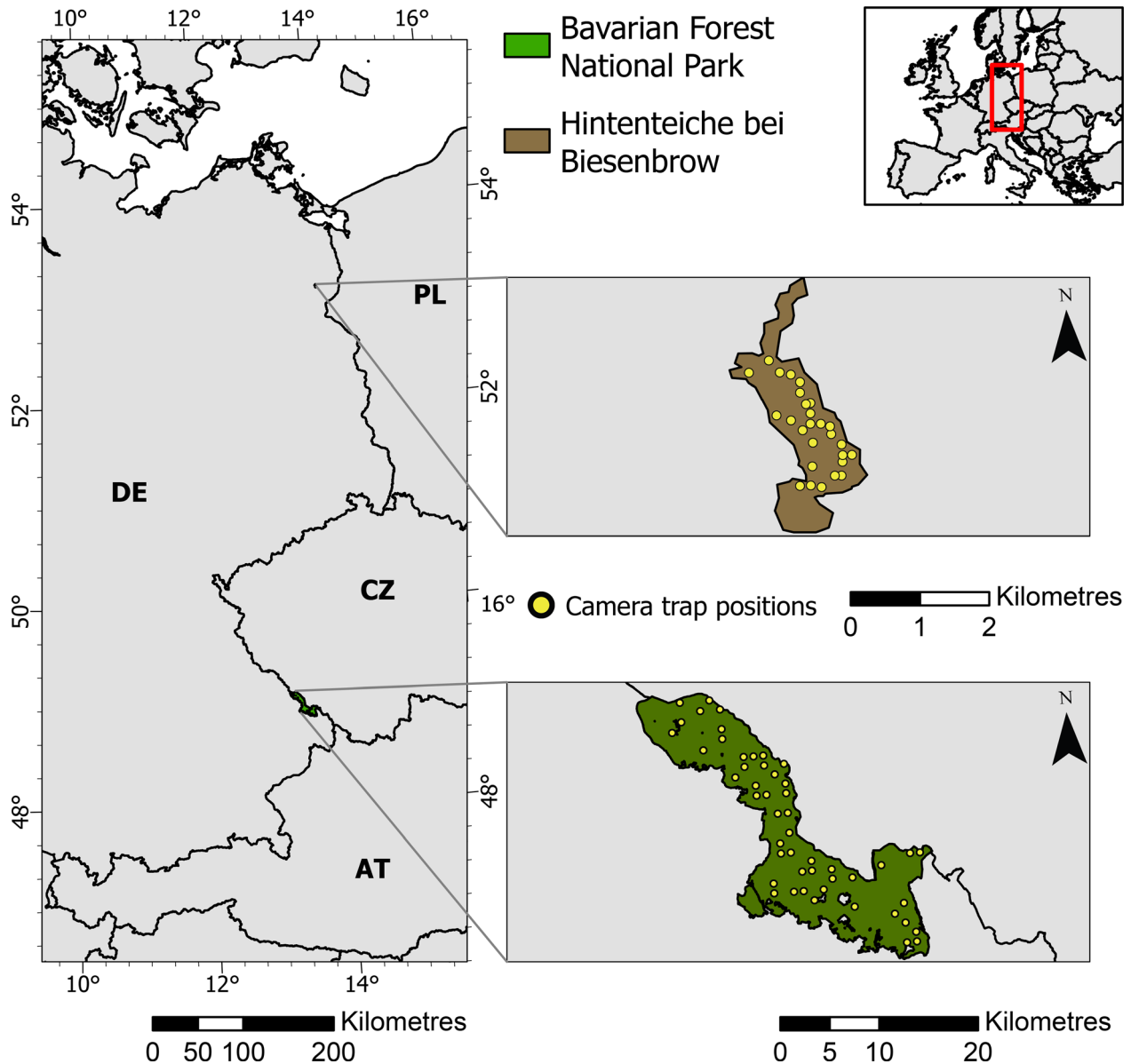
Manual distance estimation

At each CT location, a set of reference photos (BFNP) or videos (S-C) was obtained by recording a ranging pole (BFNP) or a person (S-C) positioned at distances of 1–15 m at 1-m intervals along the visual axes of each CT (larger intervals were used in the S-C for distances > 8 m; see Table S1). For the animal detections, still images were extracted from the videos recorded in the S-C at 2-sec intervals (Howe et al., 2017), with the animal species and number of individuals noted for all ground-living species observed in at least 10 videos from any number of CT locations (9 of the 10 species included in our study except for red deer). The distances to the observed animals were derived from the two closest positions of the person in the reference videos.

In the BFNP, the first photos of the photo series were defined as snapshot moments to account for variable delays between triggers (Kühl et al., submitted; Henrich et al., 2022). Red deer and roe deer were counted on these photos; images taken directly after behavioural reactions to the CTs were excluded. An event was defined as all images of an animal species at a CT location obtained within < 5 min from each other. Ten events per species, CT location, and month were randomly sampled to estimate observation distances to the animal that had moved the furthest into the FOV upon first detection (see Henrich et al., 2022). Transparent photos of the ranging pole positions were superimposed on the animal photos using Microsoft PowerPoint 2016 to find the best fit. This distance along the visual axis and the horizontal distance of the animal to the visual axis were used to trigonometrically calculate the observation distance (following Pfeffer et al., 2018).

Semi-automated distance estimation

For the S-C, the reference images for the manual and semi-automated distance estimation were the same. However, the ranging poles used in the BFNP were not well suited as reference objects for the semi-automated approach, since their outlines became very thin at larger distances and covered only a small number of pixels, increasing the risk that a large proportion of them is

Figure 1. Map of the camera trap locations within their respective study areas.

obscured by vegetation. Independent distance estimates were instead created using a person carrying an ultrasonic hypsometer (Vertex IV, Haglöf Sweden AB) to mimic the position of observed animals, with the distances between that person and the CT then measured. Two animal positions were used as references for each CT location, covering the range of observation distances.

All reference images and photos with animal observations were cropped to exclude the banners along the bottom. Masks were created for the outlines of the reference objects and text files containing the information on the corresponding distances were generated. The photos were sorted in the required folder structure and processed using the Distance Estimation Workbench of Haucke et al. (2022) (latest version available here: https://github.com/timmh/distance-estimation). Relative depth images of the reference photos are created using a deep learning algorithm (Dense Prediction Transformers, Ranftl et al., 2021), depicting pixel-wise inverse distances (D^{ref}) to objects in the FOV with the unknown scale and shift parameters m and c. These parameters are obtained by aligning at least two reference images before minimizing the error between the median value within the binary mask covering the reference object in each image and its

TABLE 1. Overview of the two test datasets used in the current study.

	Schorfheide-Chorin	Bavarian Forest National Park
Camera model	Bushnell Trophy CAM HD Aggressor 119876 (Bushnell Outdoor Products, Overland Park, Kansas, USA)	Cuddeback C2 (Cuddeback, Green Bay, Wisconsin, USA)
Angle of view	40°	55°
Flash	Infrared flash ("medium")	Infrared flash ("far")
Passive infrared sensor settings	High sensitivity	High sensitivity, wide angle
Recording settings	60-sec video, 2 sec between videos	5-pictures series, no delay between photos ("fast as possible")
Resolution	30 frames per second, resolution = 1920 × 1980 pixels	resolution = 2576 × 1984 pixels

known distance (RANSAC, Fischler & Bolles, 1981). Metric distances Z^{ref} of all pixels in the images are then obtained by Equation 1.

$$Z^{ref} = \frac{1}{m \times D^{ref} + c} \qquad (1)$$

The metric scale of these depth images is transferred to the relative depth images of the animal observations. Animals are detected in the photos using MegaDetector 4.0 (Beery et al., 2019), which creates bounding boxes around them, within which the 20th percentile of the calibrated depth values is extracted to obtain an estimate of the distance to the animal (Fig. 2). Since animals are often slightly occluded by objects in the foreground, the 20th percentile offers the highest probability that the value is sampled within the outline of the photographed animal and not any other pixel of the rectangular bounding box (Haucke et al., 2022). The confidence threshold of Mega-Detector was set to 0.75, since the suggested typical value ranges between 0.7 and 0.8 (Osner, 2022). Lowering the threshold means a higher probability to detect animals, but also an increase in false detections. With a threshold of 0.75, the percentage of motion-triggered images with false positive detections was <1% in Leorna and Brinkman (2022). The maximum distance was defined as 15 m in the S-C and 25 m in the BFNP, depending on the known maxima in the datasets (S-C: 13.5 m, BFNP: 21.7 m based on the hypsometer measurements). At distances exceeding 10 m, animals blend into the background and the corresponding depth image regions do not allow meaningful distance estimations. Depth sampling was therefore modified to allow a linear interpolation of the original 20th percentile (z_{20th}) and the depth value at the bottom centre of the bounding box (z_{bottom}) for z_{20th} values between 10 and 25 m in the BFNP data (Eq. 2, see Fig. S1). Thus, instead of extracting the distance estimate from within the outline of the animal, the point on the ground where the animal stood is used.

$$z_{sampled} = (1-a)z_{20th} + a\,z_{bottom} \qquad (2)$$

$$a = \frac{(z_{20th}-10)}{(25-10)}$$

Analyses of false negatives, false positives and differences in distance estimates

The datasets of the manual and semi-automated distance estimates (D_m/D_a) were merged based on the photo file names, such that D_a was assigned to an animal species based on manual species classification. In case of multiple distance estimates of animals in the same image in the S-C, both D_m and D_a were sorted in ascending order and matched accordingly. Animal detections in which D_m was > 14.5 m were excluded, since larger D_m could not be estimated manually due to a lack of reference photos and were frequently assigned to 15 m, the maximum distance of the reference object.

D_ms without designated D_as were classified as false negatives, and D_as without designated D_ms as false positives. The number of false negatives was summarized for each species (within each study area), CT location, and rounding interval of D_m (Table S1). Subsequently, recall was calculated for each category (species/species and distance interval/ species, distance interval and CT location). We modelled the interacting effects of D_m and animal species on the recall for each distance class, species and CT location per study area, including the identification code of the CT location as a random intercept. For this purpose, a beta-binomial generalized linear mixed model was chosen to prevent a decline in model fit with increasing D_m that occurs in binomial models ('glmmTMB', Brooks et al., 2017). Model fit was checked using "DHARMa" (Hartig, 2020). The deviation of the species-specific effect of D_m on recall from the average over all species was compared using the emtrends function with false discovery rate correction to account for multiple testing (Lenth, 2020). False positives were counted in the S-C for each distance interval based on D_a. No false positives occurred in the BFNP because there was only one D_a per photo (for the centre-most animal only). When both D_m and
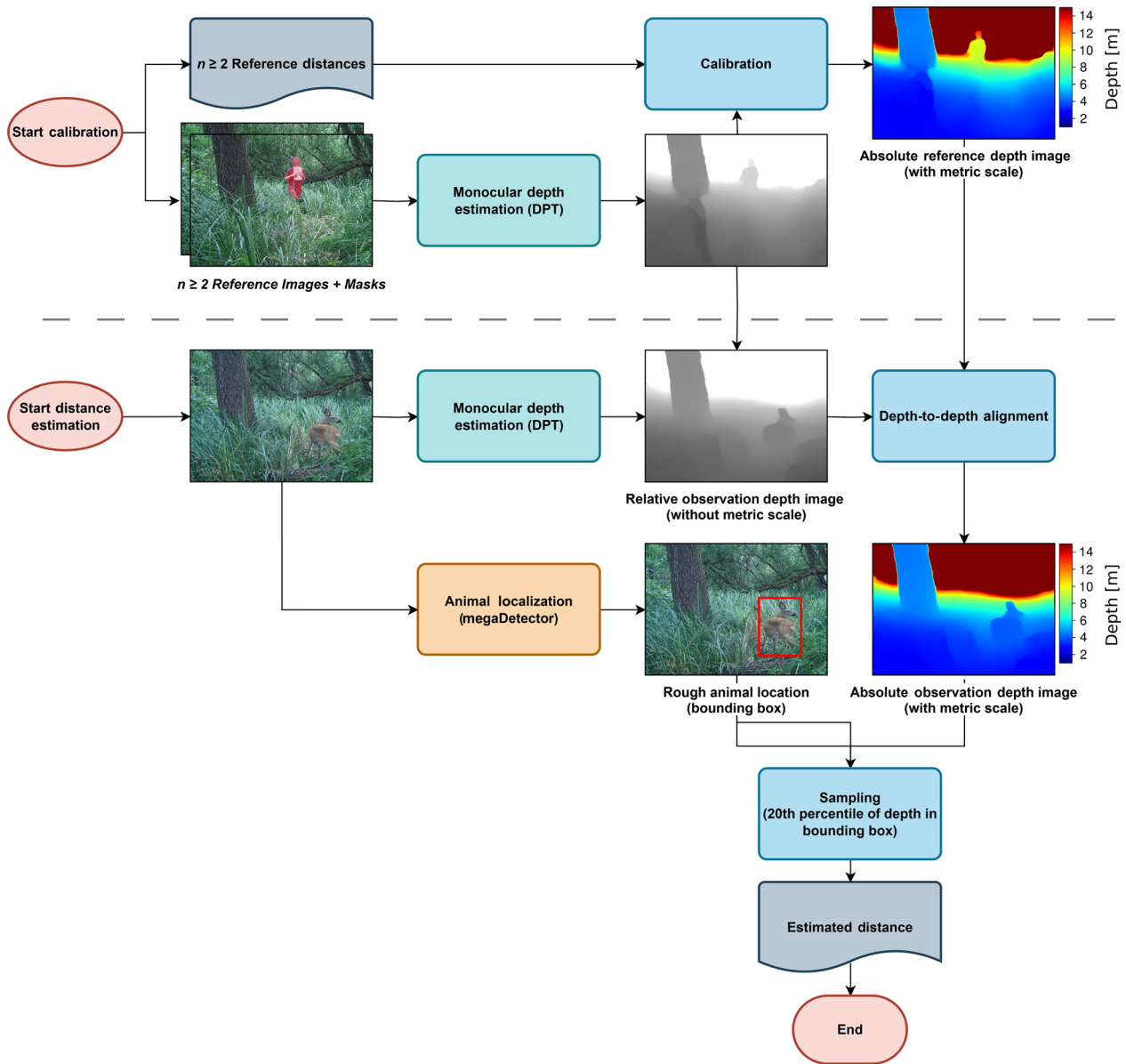
Figure 2. Graphical depiction of the semi-automated workflow to estimate distances to observed animals, featuring example photos of a camera trap location in the Schorfheide-Chorin study area.

D_a could be assigned to an observed animal (in total n observations), the values were used to compute the root mean square error (RMSE) per species and time of day (Eq. 3). All photos, for which the flash was used, were considered night-time photos.

$$\text{RMSE} = \frac{\sqrt{\sum_1^n (D_m - D_a)^2}}{n} \qquad (3)$$

The RMSEs of day-time and night-time detections were compared for each species by resampling the animal

detections with replacement and recalculating the difference of the RMSEs for each of 10 000 non-parametric bootstrap iterations. In addition, the difference D_a minus D_m was modelled depending on D_a, using Gaussian-family generalized linear and additive models, in which the identification code of the CT location was used as a random intercept ('mgcv', Wood, 2011). In a first model, D_a and the animal species were included as interacting parametric coefficients (GLMM). In a second model, a spline was fitted to D_a for each species (GAMM). The models were re-fitted based on maximum likelihood and compared with a log-likelihood ratio test, which could

otherwise not be applied since the likelihoods of models with different fixed effect structures would not be directly comparable (Faraway, 2016).

Distance sampling

Detection probabilities and population density estimates based on them were calculated for each animal species and distance estimation method based on the full dataset, including D_ms for detections with missing D_as. The R package "Distance" 1.02 (Miller et al., 2019) was used to fit detection functions to the distributions of the distances (Fig. S2). Bins were based on the rounding intervals (Table S1), but the two intervals of the closest distances were combined to account for increased uncertainty in the distance estimates when the position of the animal on the ground was not visible. Truncation distances were determined by the bin with the maximum estimate or the bin preceding a hump in the tail of the distribution, with the smaller value from the distributions of D_m and D_a being used for both datasets (except for water rail, where detection functions could otherwise not be fitted).

For the S-C, the model selection procedure of Howe et al. (2019), with nine candidate detection functions, was applied. Model fitting and selection for the BFNP data were the same as in Henrich et al. (2022). Population density estimates were computed following Equation 4.

$$\widehat{D} = \frac{2t \sum_{k=1}^{K} n_k \, p}{\theta_v \, \omega^2 \sum_{k=1}^{K} T_k \, \widehat{P}} \frac{1}{\widehat{A}} \qquad (4)$$

t = interval between snapshot moments (replaced by $\widehat{d}$ in the BFNP); k = camera trap location index; K = number of camera trap locations; n_k = sum of individuals counted during all snapshot moments for a given species at location k; p = proportion of all estimated distances that are within ω for a given species in the BFNP; takes the value 1 in the S-C; θ_v = horizontal angle of view (radians); ω = truncation distance (km); T_k = camera trap deployment time at location k (seconds); $\widehat{P}$ = detection probability of a given species; $\widehat{A}$ = activity level of a given species (daily proportion of time spent active).

The value of θ was based on the manufacturer's specifications and that of $\widehat{A}$ was derived from circular probability density functions fitted to the daily distributions of independent events for each species ('activity', Rowcliffe et al., 2014). n_k corresponded to the number of MegaDetector detections for the semi-automated workflow in the S-C, but manual counts had to be used to that end in the BFNP, since only a subset of the animal observations was used for distance estimation and analysed by MegaDetector. p had to be included in the formula since the number of observations within the truncation distance was known

only for this subset. Furthermore, t had to be changed to $\widehat{d}$ to account for recovery and retrigger times between photos (see Henrich et al., 2022).

The 95% confidence intervals for the density estimates from the manual and semi-automated workflows and their differences were derived using a non-parametric bootstrap of CT locations, with the numbers of observations n and deployment times T recalculated for each bootstrap iteration (n = 10 000).

All analyses were performed in R 4.0.2 (R Core Team, 2020).

Results

Automated animal detection

The percentage of animals not detected by MegaDetector varied widely between species, ranging from 4% for water rail to 39% for wild boar (see Table S2). In general, recall across species was not related to the distance from the CT (see Table S3). However, distance had a negative effect on the proportion of automatically detected roe deer ($P = 0.03$), beaver ($P = 0.006$) and wild boar ($P < 0.0001$) relative to the overall trend in the S-C. In this study area, 3.8% of the automated detections were false positives and their distribution closely followed that of true positives, with a peak at the 4- to 5-m interval of D_a (Fig. S4).

Differences in distance estimation methods

The RMSE for D_m and D_a of the same animal detection was smallest for water rail and largest for red deer (Table 2). It was larger at night than during the day for roe deer, wild boar, and beaver in the S-C and for red deer in the BFNP, but the relationship was reversed for hare. For all species, the difference between the estimates was positively related to D_a (Fig. 3, see Fig. S3 for all species). The minimum in the differences, as derived from the GAMMs, occurred around the 4-m mark of D_a for most species, but at a shorter D_a for domestic cat (2 m) and a higher D_a for beaver (5 m) and the two deer species in the BFNP (7 m). Overall, the non-linear splines did not improve the model significantly (see Table S4). Roe deer and wild boar in the S-C deviated the most from linearity (effective degrees of freedom of $\sim$4), with little change ($\leq$0.5 m) in the predicted difference between D_m and D_a at 4–7 m and 3–7 m respectively.

Differences in estimated animal abundance

The differences between the population density estimates based on manual versus semi-automated workflows was

TABLE 2. Root mean square error between manual and semi-automated distance estimates of the same animal detection: per species, per time of day (without or with infrared flash) and per study area.

Species	Area	RMSE$_{day}$ [m]	RMSE$_{night}$ [m]	95% CI difference (RMSE$_{night}$ − RMSE$_{day}$)
Badger	S-C	2.61	2.22	−0.87, 0.10
Beaver	S-C	2.33	3.89	1.23, 1.87
Domestic cat	S-C	2.59	2.52	−0.73, 0.56
Hare	S-C	3.36	1.64	−2.55, -0.67
Raccoon	S-C	2.99	3.13	−0.15, 0.42
Water rail	S-C	1.87	NA	NA
Red fox	S-C	2.60	2.54	−0.50, 0.35
Roe deer	S-C	2.44	2.78	0.25, 0.42
Wild boar	S-C	2.31	2.95	0.55 0.73
Red deer	BFNP	3.25	5.63	0.94, 3.75
Roe deer	BFNP	1.94	2.01	−0.77, 1.05

BFNP, study area Bavarian Forest National Park; RMSE, Root mean square error; S-C, study area Schorfheide-Chorin.

<5% of the estimated densities for red deer, roe deer (S-C), hare, and badger, whereas the largest relative differences of 71% and 129% were found for water rail and domestic cat (Fig. 4; Table S2). The population density estimates for the same species and study area were never statistically significantly different, as the confidence intervals for their difference always overlapped zero (see Table S2).

Discussion

The semi-automated estimation of distances between camera traps and observed animals was generally unbiased in comparison to the manual estimation around 4 m, but the agreement between the methods decreased in larger distances. These differences did, however, not significantly impact the population density estimates of ten animal species at two different study sites, although the point estimates diverged to varying degrees.
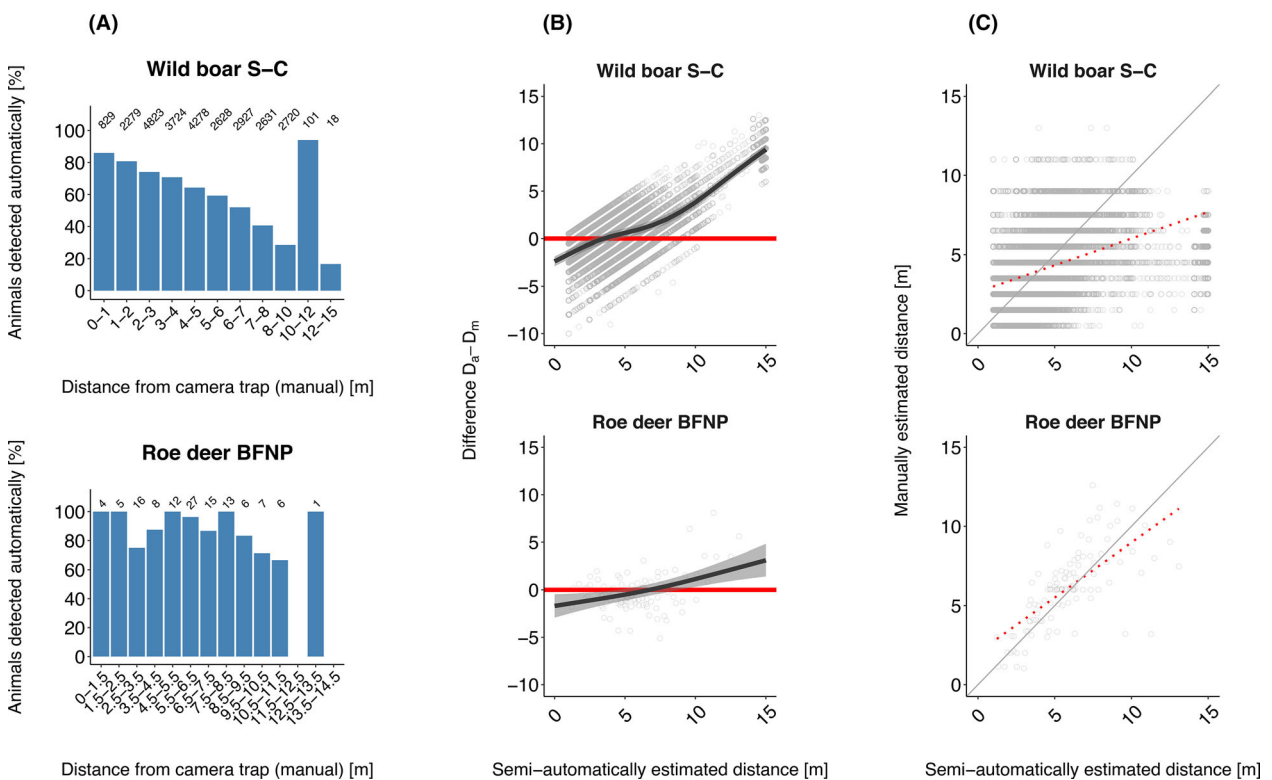
Figure 3. (A) Percentage of animal observations detected by MegaDetector, dependent on the distance from the camera trap, (B) the relationship between the semi-automated distance estimates and their differences with the manual estimates and (C) the semi-automated distance estimates plotted against the manual distance estimates, as exemplified by wild boar (Schorfheide-Chorin) and roe deer (Bavarian Forest National Park). The numbers above the bars in (A) indicate the total number of true positive detections in the respective distance class; a difference of zero is indicated by the red line in (B) and the solid grey line in (C). The trend line of the actual data is shown with a red dotted line in (C).
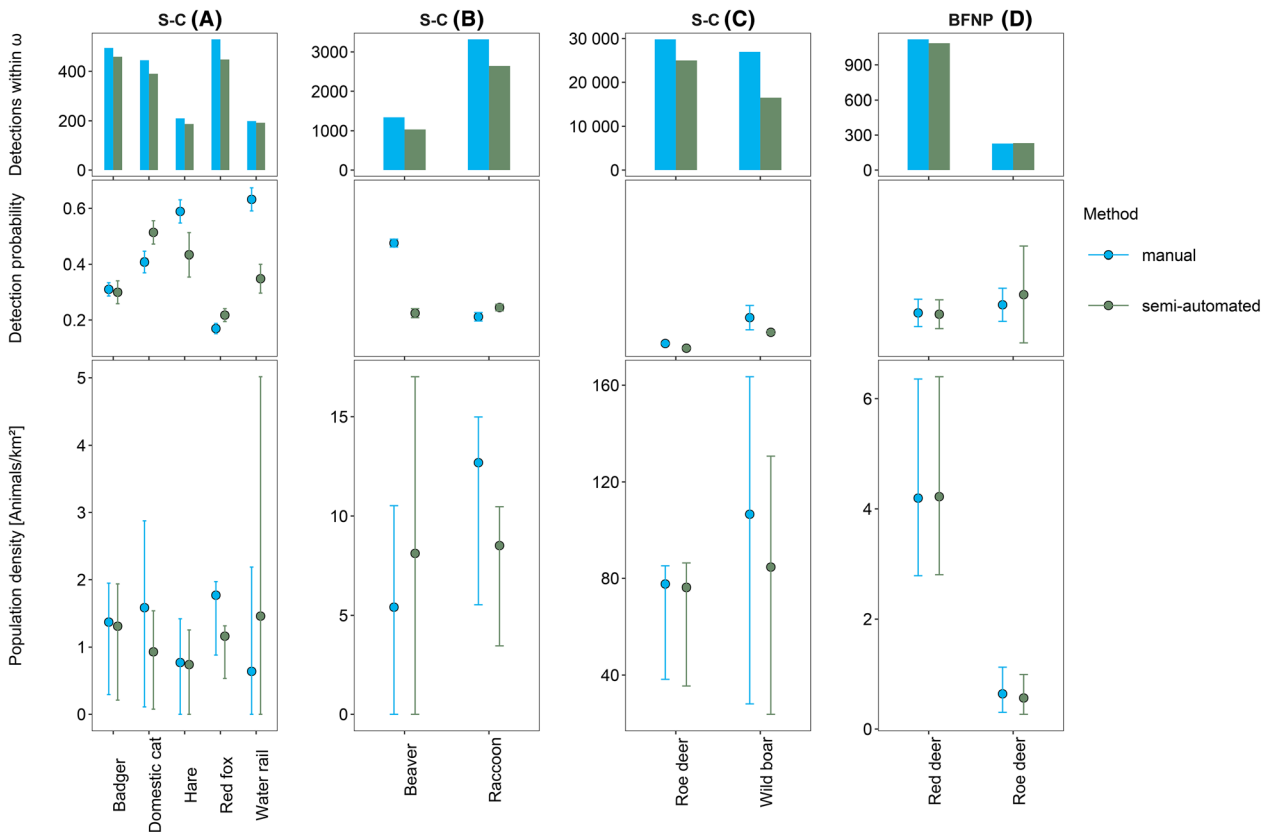
Figure 4. Detections, detection probabilities and population density estimates of nine animal species at 27 camera trap locations in the Biosphere Reserve Schorfheide-Chorin in the summer of 2019 (A–C) and of red and roe deer at 50 camera trap locations in the Bavarian Forest National Park in the summer of 2018 (D). Detection probabilities can only be compared within species due to different truncation distances ω. For the same number of animal detections, an increase in the detection probability leads to lower population density estimates. Apart from the number of detections and the detection probability, the other camera trap distance sampling parameters were constant between the manual and semi-automated workflows.

## Automated animal detection

With MegaDetector, 77%–96% of the observations of most animal species were automatically detected on the CT images, except wild boar, which were often present in groups, decreasing the probability to detect all individuals on a photo (Fig. S5). The significant decrease in recall with distance indicated that the further away an animal was from the CT, the higher the likelihood that it would be partly concealed by other group members (see, e.g., Fig. 5c). Additionally, the chance of an animal being largely hidden by vegetation, such as a fallen tree at one CT location with beaver observations, also increases with the distance from the CT. Thus, if only automated animal detections are counted for population density estimations, as was done in the S-C, a steady increase in false negatives at larger distances would be fully accounted for by a lower detection probability. After all, distance sampling was specifically developed to consider false negatives as a

function of distance. However, if the distance ranges with low recall are more irregularly distributed such as for beavers, the detection probability cannot account effectively for the missed detections.

Beside plant parts being classified as animals, false positive detections by MegaDetector could often be ascribed to the separate detection of body parts of the same animal (see, e.g., Fig. 5D), which explains why the distribution of false positives over different distance classes follows that of the true positives. Consequently, the estimation of the detection probability would not be biased if false positives were included. However, false positives were not assigned to an animal species and therefore not included for population density estimation in our study.

Differences in distance estimation methods

If MegaDetector correctly detected an animal, the species-specific RMSE of the D_m and D_a was 2–3 m, except for
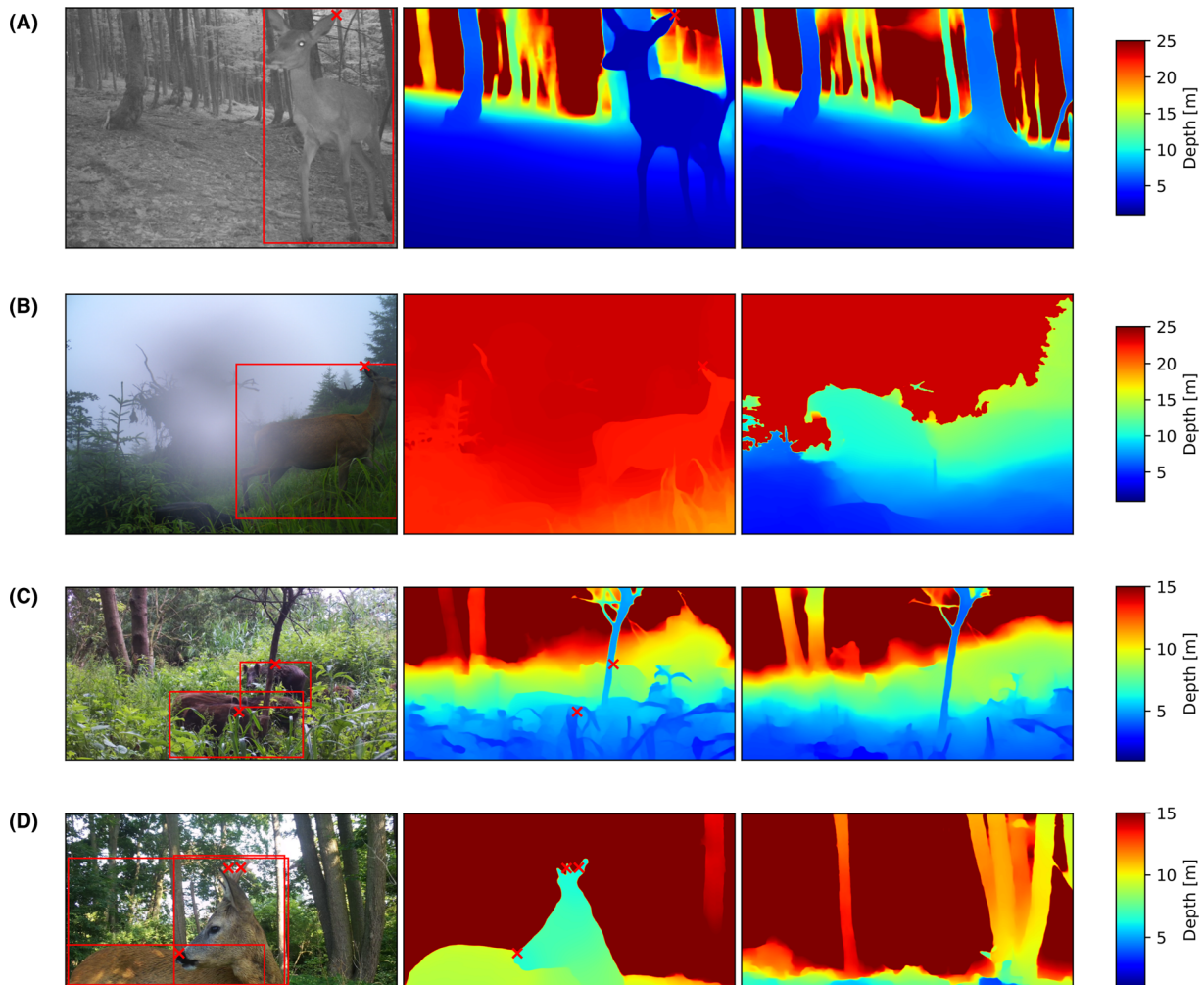
Figure 5. Examples of photos with animal observations, the corresponding calibrated depth images and the reference depth images with the reference object at the farthest distance (not always visible against the background). MegaDetector created the bounding boxes in the panels on the left; the cross indicates the pixel of the sampled distance estimate. In the depicted scenarios: (A) the animal is detected both manually and automatically and the difference in the estimated distance to the camera trap is minimal (0.48 cm in this case); (B) the animal is detected manually and automatically, but there is a large difference in the resulting distance estimates (15.9 m in this case); (C) some animals are not detected automatically (such as two of the four wild boar); (D) the automated detection produces false positives (as multiple crosses on different body parts of the same roe deer show).

red deer and beaver. The error was largest for red deer in the BFNP (4.73 m), but it was substantially reduced (2.74 m) when detections with $D_a > 15$ m (14% of the estimates) were removed from the dataset. In contrast, the high RMSE of the beaver detections could be ascribed to three CT locations, which had an inherent bias in their reference depth images, as very small or very large distances were ascribed to nearly the entire picture and sometimes even the range was inverted. It is not clear why the deep learning algorithm sometimes over- or underestimates distances, as its underlying mechanism largely remains a black box (Hu et al., 2019). However, a few issues are known: Dense vegetation close to the

camera trap can lead to an underestimation of distances to objects behind it, because the depth estimation algorithm tends to smooth over small gaps in the vegetation (Haucke et al., 2022). Furthermore, the accuracy of the depth estimation is reduced if the ground is only visible in a small part of a photo, since context information about the relative positions of objects is lost. While the semi-automated density estimation is robust to small changes in the FOV (e.g. Fig. S6), large changes can be detrimental. It is, therefore, recommended that at least one third of the FOV should be covered by ground, there should be no tall vegetation within 3 m and the camera should be fixed securely (Haucke et al., 2022). Moreover,

it would be advantageous for long-term camera trap studies to take reference photos seasonally to prevent difficulties due to major changes of the scene such as fallen trees or growing vegetation over time.

The agreement between the methods was not consistently negatively impacted by the lack of colour images at night. However, the RMSE was larger at night for ungulates and beaver in the S-C and for red deer in the BFNP. Except for beaver, these species had an above average number of observations at larger distances (90% quantile of D_m between 7.5 and 12 m). If an animal is observed outside the range of the infrared flash at night, its outline is often barely visible. A low visibility of animal contours is highly detrimental to the quality of the semi-automated distance estimates (Fig. S7), but this problem is not restricted to night-time images, as fog, sun flares or water on the camera lens can also result in blurry photos (see, e.g., Fig. 5B).

The hypothesis that the agreement between D_m and D_a depends on the distance from the CT was confirmed. Distances were best estimated when an animal was close to the CT and its body was fully visible (Fig. S3). In distances below 3 m, it was not possible to directly locate an animal's position on the terrain and D_m tended to be larger than D_a. Further away, the number of pixels representing a given interval decreased with distance, but D_a often overestimated this decrease. The large differences between the methods at the highest D_a values can be explained by a depth estimation failure, such that these values were produced by resorting to the maximum D_a. For roe deer and wild boar in the S-C (89% of all observations in the area), the difference between D_m and D_a was relatively constant within a range of 3 m around the optimum. This might be explained by robust predictions of the mean difference due to large sample sizes, since these distance ranges covered 44% (roe deer) to 50% (wild boar) of the species-specific observations. In general, the relationship between D_m and D_a did not deviate from linearity. This finding could be exploited to derive a correction factor that can be used to adjust D_a to D_m irrespective of the distance to the CT (Fig. S8 and S9). For a consistent and robust correction, measurement errors should, however, be explicitly incorporated in the fitting of detection functions in future (analogous to Borchers et al., 2010). Importantly, D_m does not equal ground truth, since it can also be expected to have a lower precision very close and very far from the CT. For the BFNP, the manual estimates could, however, be validated by comparing them with the hypsometer estimates for all animal observations, which revealed negligible bias until approximately 10 m (Fig. S10) and small RMSEs (red deer: 1.65 m, roe deer: 1.24 m).

Differences in estimated animal abundance

A high RMSE translated in some cases to large differences in the detection probabilities derived from D_m and D_a, for example, for beaver. Water rail, in contrast, had the lowest RMSE, but its manual and semi-automated detection probability estimates diverged strongly. This problem arose because of low-biased reference depth images at the CT location where all except one water rail detection occurred, leading to a systematic underestimation of the distances to all observed animals. While the impact of such issues at individual CT locations decreases with the number of CT locations at which a species is observed, efforts to obtain accurate reference images should never be neglected. For red deer, D_m and D_a were nearly identical despite a large RMSE. Since RMSE is a measure of precision and not accuracy, it may possess limited explanatory power for differences in the distribution of D_a and D_m. A small number of large estimates can also strongly influence the RMSE, but will have little impact on the distribution overall. These values will also be removed by truncation before the fitting of the detection function, to increase precision (Buckland, 2001). Furthermore, aggregation of the data into bins centred on favoured rounding values (Thomas et al., 2010) eliminates the influence of the smaller scale differences between D_m and D_a in the form of the detection function. The smallest differences between D_a and D_m were found in the bins with the largest number of D_as, except for beaver and water rail (Fig. S11). Together, these factors led to a lower divergence of detection probabilities than expected based on the absolute mean differences between D_m and D_a.

For most species, a small difference between detection probabilities implied a small difference in population density estimates relative to the absolute values and *vice versa*. However, in the case of raccoon, a small difference in the detection probability but a high proportion of not automatically detected individuals led to an underestimation of the population density with the semi-automated approach. While for hare the false negative detections were largely balanced out by a lower semi-automated detection probability, this was not the case for wild boar. Although the concrete outcome depended on the specific number and distribution of observations for each species, there were no significant differences between the population density estimates obtained with manual and semi-automated CTDS for any of the studied species. The differences between the estimates were obscured by the spread of the confidence intervals, which was governed by the variability in the number of observed animals across CT locations. Typically, the coefficients of variation (CVs) for camera trap distance sampling range between 0.2 and 0.5 (76% of the 29 estimates per species or species-group

in Bessone et al., 2020; Cappelle et al., 2019, 2021; Corlatti et al., 2020; Howe et al., 2017; Mason et al., 2022). The CVs in our study fall within this range in the BFNP (0.21, 0.31), but are higher in the S-C (0.7–1.7). However, even if we assume a CV of 0.2, only the estimates for the water rail might become significantly different for the two observation distance estimation techniques (Table S5).

Semi-automated distance estimation greatly reduces the time required for data processing. Based on the mean processing period per photo reported by Haucke et al. (2022) and on two reference photos per CT location, the manual observation distance estimation for the dataset used in the present study would require > 23 work days, whereas semi-automated processing would be completed in <2 days (Table S6). For the manual steps alone, this is a 55-fold reduction in the hours of work. Considering that manual distance estimation may need > 12 sec per photo, in many cases the amount of time saved will be even greater.

Our semi-automated distance estimation workflow can be easily applied by practitioners, since the whole process is based on freely available software, which can be controlled *via* a simple graphical user interface. After sorting the images in the appropriate folder structure and checking the settings, the whole workflow runs autonomously as soon as the start button is pressed. The method can be applied to a variety of species and settings, since it does not depend on the size and outline of the animals or the form of the terrain and just requires two reference images per camera trap location. Future advances might omit the need for reference images and fully automate the process (Johanns et al., 2022).

Conclusions

Accurate estimation of the distances from observed animals to CTs is an essential, but time-consuming, step for many methods to estimate the population density of unmarked species based on camera trapping data (Palencia et al., 2021). Our study demonstrates that distance estimates obtained with a deep-learning approach based on standard monocular photos can provide reliable population density estimates, assuming both a small number of missed animals not accounted for by the detection probability and unbiased reference depth images. Potential errors can be minimized by compliance with CT placement guidelines as well as analytical approaches. The method is an important step in the establishment of a largely automated pipeline for the analysis of CT data aimed at the estimation of population densities, which includes the detection of animals in the images, species classification and distance estimation. Given the reduced effort required to obtain this information, it offers a practical option for wildlife managers who lack the workforce for extensive analyses and for situations in which fast results are essential for decision making (Schulz et al., 2019).

Author Contributions

MHen (BFNP), MHeu (BFNP), JH (S-C) & HSK (S-C) designed the camera trap studies. MB (BFNP), MHen (BFNP), JH (S-C), and HSK (S-C) collected and processed the data. TH and VS developed and updated the semi-automated distance estimation software. MHen analysed the data, with the support of MB and led the writing of the manuscript. All authors contributed critically to the drafts and gave final approval for publication.

Conflict of Interest

The authors have no conflict of interest to declare.

Data Availability Statement

Data and R scripts related to the project can be found at Henrich, M. (2023, April 5). A semi-automated camera trap distance sampling approach for population density estimation. https://osf.io/abf3y/?view_only=d8cb52a5b6cb4a638198914511730973.

References

Beery, S., Morris, D. & Yang, S. (2019) Efficient pipeline for camera trap image review. Available from: https://doi.org/10.48550/ARXIV.1907.06772

Bessone, M., Kühl, H.S., Hohmann, G., Herbinger, I., N'Goran, K.P., Asanzi, P. et al. (2020) Drawn out of the

shadows: surveying secretive forest species with camera trap distance sampling. *Journal of Applied Ecology*, **57**(5), 963–974. Available from: https://doi.org/10.1111/1365-2664.13602

Borchers, D., Marques, T., Gunnlaugsson, T. & Jupp, P. (2010) Estimating distance sampling detection functions when distances are measured with errors. *Journal of Agricultural, Biological, and Environmental Statistics*, **15**(3), 346–361. Available from: https://doi.org/10.1007/s13253-010-0021-y

Brooks, M.E., Kristensen, K., van Benthem, K.J., Magnusson, A., Berg, C.W., Nielsen, A. et al. (2017) glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. *The R Journal*, **9**(2), 378–400. Available from: https://doi.org/10.32614/RJ-2017-066

Buckland, S.T. (2001) *Introduction to distance sampling: estimating abundance of biological populations*. Oxford; New York: Oxford University Press.

Buckland, S.T., Anderson, D.R., Burnham, K.P., Laake, J.L., Borchers, D.L. & Thomas, L. (2004) *Advanced distance sampling: estimating abundance of biological populations*. Oxford, UK: OUP Oxford.

Cailleret, M., Heurich, M. & Bugmann, H. (2014) Reduction in browsing intensity may not compensate climate change effects on tree species composition in the Bavarian Forest National Park. *Forest Ecology and Management*, **328**, 179–192. Available from: https://doi.org/10.1016/j.foreco.2014.05.030

Cappelle, N., Després-Einspenner, M., Howe, E.J., Boesch, C. & Kühl, H.S. (2019) Validating camera trap distance sampling for chimpanzees. *American Journal of Primatology*, **81**(3), e22962. Available from: https://doi.org/10.1002/ajp.22962

Cappelle, N., Howe, E.J., Boesch, C. & Kühl, H.S. (2021) Estimating animal abundance and effort–precision relationship with camera trap distance sampling. *Ecosphere*, **12**(1), e03299. Available from: https://doi.org/10.1002/ecs2.3299

Corlatti, L., Sivieri, S., Sudolska, B., Giacomelli, S. & Pedrotti, L. (2020) A field test of unconventional camera trap distance sampling to estimate abundance of marmot populations. *Wildlife Biology*, **2020**(4), wlb.00652. Available from: https://doi.org/10.2981/wlb.00652

Diáz, S., Settele, E.S., Brondízio, E.S., Ngo, H.T., Guèze, M., Agard, J. et al. (2019) *Summary for policymakers of the global assessment report on biodiversity and ecosystem services of the intergovernmental science-policy platform on biodiversity and ecosystem services*. Bonn, Germany: IPBES Secretariat.

Faraway, J.J. (2016) *Extending the linear model with R: generalized linear, mixed effects and nonparametric regression models*, 2nd edition. Boca Raton, FL: Chapman and Hall/CRC. Available from: https://doi.org/10.1201/9781315382722

Fischler, M.A. & Bolles, R.C. (1981) Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, **24**(6), 381–395. Available from: https://doi.org/10.1145/358669.358692

Hartig, F. (2020) *DHARMa: Residual diagnostics for hierarchical (multi-level/mixed) regression models*. Available from: https://CRAN.R-project.org/package=DHARMa [Accessed: 4th December 2020]

Haucke, T., Kühl, H.S., Hoyer, J. & Steinhage, V. (2022) Overcoming the distance estimation bottleneck in estimating animal abundance with camera traps. *Ecological Informatics*, **68**, 101536. Available from: https://doi.org/10.1016/j.ecoinf.2021.101536

He, Z., Kays, R., Zhang, Z., Ning, G., Huang, C., Han, T.X. et al. (2016) Visual informatics tools for supporting large-scale collaborative wildlife monitoring with citizen scientists. *IEEE Circuits and Systems Magazine*, **16**(1), 73–86. Available from: https://doi.org/10.1109/MCAS.2015.2510200

Henrich, M., Hartig, F., Dormann, C.F., Kühl, H.S., Peters, W., Franke, F. et al. (2022) Deer behavior affects density estimates with camera traps, but is outweighed by spatial variability. *Frontiers in Ecology and Evolution*, **10**, 881502. Available from: https://doi.org/10.3389/fevo.2022.881502

Hofmeester, T.R., Rowcliffe, J.M. & Jansen, P.A. (2017) A simple method for estimating the effective detection distance of camera traps. *Remote Sensing in Ecology and Conservation*, **3**(2), 81–89. Available from: https://doi.org/10.1002/rse2.25

Howe, E.J., Buckland, S.T., Després-Einspenner, M. & Kühl, H.S. (2017) Distance sampling with camera traps. *Methods in Ecology and Evolution*, **8**(11), 1558–1565. Available from: https://doi.org/10.1111/2041-210X.12790

Howe, E.J., Buckland, S.T., Després-Einspenner, M. & Kühl, H.S. (2019) Model selection with overdispersed distance sampling data. *Methods in Ecology and Evolution*, **10**(1), 38–47. Available from: https://doi.org/10.1111/2041-210X.13082

Hu, J., Zhang, Y. & Okatani, T. (2019) Visualization of convolutional neural networks for monocular depth estimation. In: *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*. Seoul, Korea (South): IEEE, pp. 3868–3877. Available from: https://doi.org/10.1109/ICCV.2019.00397

Johanns, P., Haucke, T. & Steinhage, V. (2022) Distance estimation and animal tracking for wildlife camera trapping. *arXiv: 2202.04613* [Preprint]. Available from: https://doi.org/10.48550/ARXIV.2202.04613

Kühl, H.S., Buckland, S.T., Henrich, M., Howe, E.J. & Heurich, M. (submitted) *Estimating effective survey duration in camera trap distance sampling surveys*.

Lenth, R. (2020) *emmeans: Estimated Marginal Means, aka Least-Squares Means*. Available from: https://CRAN.R-project.org/package=emmeans

Leorna, S. & Brinkman, T. (2022) Human vs. machine: detecting wildlife in camera trap images. *Ecological Informatics*, **72**, 101876. Available from: https://doi.org/10.1016/j.ecoinf.2022.101876

Leorna, S., Brinkman, T. & Fullman, T. (2022) Estimating animal size or distance in camera trap images:

photogrammetry using the pinhole camera model. *Methods in Ecology and Evolution*, **13**(8), 1707–1718. Available from: https://doi.org/10.1111/2041-210X.13880

Marques, T.A., Thomas, L., Martin, S.W., Mellinger, D.K., Ward, J.A., Moretti, D.J. et al. (2013) Estimating animal population density using passive acoustics. *Biological Reviews of the Cambridge Philosophical Society*, **88**(2), 287–309. Available from: https://doi.org/10.1111/brv.12001

Mason, S.S., Hill, R.A., Whittingham, M.J., Cokill, J., Smith, G.C. & Stephens, P.A. (2022) Camera trap distance sampling for terrestrial mammal population monitoring: lessons learnt from a UK case study. *Remote Sensing in Ecology and Conservation*, **8**(5), 717–730. Available from: https://doi.org/10.1002/rse2.272

Massei, G., Roy, S. & Bunting, R. (2011) Too many hogs? *Human-Wildlife Interactions*, **5**(1), 79–99.

Miller, D.L., Rexstad, E., Thomas, L., Marshall, L. & Laake, J.L. (2019) Distance sampling in R. *Journal of Statistical Software*, **89**(1). Available from: https://doi.org/10.18637/jss.v089.i01

Nakashima, Y., Fukasawa, K. & Samejima, H. (2018) Estimating animal density without individual recognition using information derivable exclusively from camera traps. *Journal of Applied Ecology*, **55**(2), 735–744. Available from: https://doi.org/10.1111/1365-2664.13059

Norouzzadeh, M.S., Nguyen, A., Kosmala, M., Swanson, A., Palmer, M.S., Packer, C. et al. (2018) Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning. *Proceedings of the National Academy of Sciences*, **115**(25), E5716–E5725. Available from: https://doi.org/10.1073/pnas.1719367115

Osner, N. (2022) *MegaDetector 5 evaluation*. Olifantsfontein, South Africa: WildEye.

Palencia, P., Rowcliffe, J.M., Vicente, J. & Acevedo, P. (2021) Assessing the camera trap methodologies used to estimate density of unmarked populations. *Journal of Applied Ecology*, **58**, 1583–1592. Available from: https://doi.org/10.1111/1365-2664.13913

Pfeffer, S.E., Spitzer, R., Allen, A.M., Hofmeester, T.R., Ericsson, G., Widemo, F. et al. (2018) Pictures or pellets? Comparing camera trapping and dung counts as methods for estimating population densities of ungulates. *Remote Sensing in Ecology and Conservation*, **4**(2), 173–183. Available from: https://doi.org/10.1002/rse2.67

Price Tack, J.L., West, B.S., McGowan, C.P., Ditchkoff, S.S., Reeves, S.J., Keever, A.C. et al. (2016) AnimalFinder: a semi-automated system for animal detection in time-lapse camera trap images. *Ecological Informatics*, **36**, 145–151. Available from: https://doi.org/10.1016/j.ecoinf.2016.11.003

R Core Team. (2020) *R: a language and environment for statistical computing*. Version 4.0.2. Vienna, Austria: R Foundation for Statistical Computing. Available from: https://www.r-project.org/

Ranftl, R., Bochkovskiy, A. and Koltun, V. (2021) Vision transformers for dense prediction, *arXiv preprint arXiv:2103.13413* [preprint].

Rowcliffe, J.M., Carbone, C., Jansen, P.A., Kays, R. & Kranstauber, B. (2011) Quantifying the sensitivity of camera traps: an adapted distance sampling approach. *Methods in Ecology and Evolution*, **2**(5), 464–476. Available from: https://doi.org/10.1111/j.2041-210X.2011.00094.x

Rowcliffe, J.M., Field, J., Turvey, S.T. & Carbone, C. (2008) Estimating animal density using camera traps without the need for individual recognition. *Journal of Applied Ecology*, **45**(4), 1228–1236. Available from: https://doi.org/10.1111/j.1365-2664.2008.01473.x

Rowcliffe, J.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**(11), 1170–1179. Available from: https://doi.org/10.1111/2041-210X.12278

Schulz, K., Staubach, C., Blome, S., Viltrop, A., Nurmoja, I., Conraths, F.J. et al. (2019) Analysis of Estonian surveillance in wild boar suggests a decline in the incidence of African swine fever. *Scientific Reports*, **9**(1), 8490. Available from: https://doi.org/10.1038/s41598-019-44890-0

Schwarz, C.J. & Seber, G.A.F. (1999) Estimating animal abundance: review III. *Statistical Science*, **14**(4). Available from: https://doi.org/10.1214/ss/1009212521

Swinnen, K.R.R., Reijniers, J., Breno, M. & Leirs, H. (2014) A novel method to reduce time investment when processing videos from camera trap studies. *PLoS One*, **9**(6), e98881. Available from: https://doi.org/10.1371/journal.pone.0098881

Thomas, L., Buckland, S.T., Rexstad, E.A., Laake, J.L., Strindberg, S., Hedley, S.L. et al. (2010) Distance software: design and analysis of distance sampling surveys for estimating population size. *Journal of Applied Ecology*, **47**(1), 5–14. Available from: https://doi.org/10.1111/j.1365-2664.2009.01737.x

Wearn, O.R., Bell, T.E.M., Bolitho, A., Durrant, J., Haysom, J.K., Nijhawan, S. et al. (2022) Estimating animal density for a community of species using information obtained only from camera-traps. *Methods in Ecology and Evolution*, **13**(10), 2248–2261. Available from: https://doi.org/10.1111/2041-210X.13930

Wood, S.N. (2011) Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models: estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **73**(1), 3–36. Available from: https://doi.org/10.1111/j.1467-9868.2010.00749.x

Yu, X., Wang, J., Kays, R., Jansen, P.A., Wang, T. & Huang, T. (2013) Automated identification of animal species in camera trap images. *EURASIP Journal on Image and Video Processing*, **2013**(1), 52. Available from: https://doi.org/10.1186/1687-5281-2013-52

Zuleger, A.M., Holland, R. & Kühl, H.S. (2022) Deriving observation distances for camera trap distance sampling. *African Journal of Ecology*, **60**(3), 423–432. Available from: https://doi.org/10.1111/aje.12959

Supporting Information

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

Figure S1. An animal vanishing into the background of the depth image at a large distance. From left to right: infrared intensity image with bounding box, corresponding aligned depth image with bounding box, and reference depth image obtained by the calibration step.

Figure S2. Detection functions fitted to the manually and semi-automatically estimated observation distances of nine vertebrate species in Schorfheide-Chorin and two deer species in the Bavarian Forest National Park.

Figure S3. (A) Percentage of animal observations detected by MegaDetector, dependent on the distance from the camera trap, (B) the relationship between the semi-automated distance estimates and their differences with the manual estimates and (C) the semi-automated distance estimates plotted against the manual distance estimates, for all animal species in the study areas Schorfheide-Chorin (S-C) and Bavarian Forest National Park (BFNP). The numbers above the bars in (a) indicate the total number of true positive detections in the respective distance class; a difference of zero is indicated by the red line in (B) and the solid grey line in (C). The trend line of the actual data is shown with a red dotted line in (C).

Figure S4. Distribution of false positives in the Schorfheide-Chorin study area (S-C) over the range of the distance estimates. The numbers above the bars indicate the number of true positives in the respective distance class.

Figure S5. Percentage of wild boar detected by MegaDetector, dependent on the distance from the camera, for different group sizes.

Figure S6. Examples showing slight shifts of the camera trap field of view at two CT locations (upper two panels: T09 in Schorfheide-Chorin, lower two panels: G454 in the Bavarian Forest National Park).

Figure S7. Boxplots of the absolute differences between manual and semi-automated distance estimates of red deer and roe deer in the Bavarian Forest National Park dependent (A) on the proportion of the body of an animal inside the field of view and (B) the visibility of its outline.

Figure S8. The relationship between the population density estimates based on the semi-automated (A) or corrected semi-automated (B) distance estimates and the manual distance estimates. Corrections of the semi-automated distance estimates were based on the predictions of a linear regression with the manual estimates as the response variable: gam(Distance manual ∼ scale[Distance automatic] × Species + s[CT location ID, bs = "re"], data, family = Gaussian).

Figure S9. Detections, detection probabilities and population density estimates with the manual, semi-automated and corrected semi-automated workflows for nine animal species at 27 camera trap locations in the Biosphere Reserve Schorfheide-Chorin in the summer of 2019 (A–C) and of red and roe deer at 50 camera trap locations in the Bavarian Forest National Park in the summer of 2018 (D). Corrections of the semi-automated distance estimates were based on the predictions of a linear regression with the manual estimates as the response variable: gam(Distance manual ∼ scale[Distance automatic] × Species + s [CT location ID, bs = "re"], data, family = Gaussian). Detection probabilities can only be compared within species due to different truncation distances ω. Truncation distances were the same as for the comparison of manual and semi-automated density estimates in the main manuscript (Table S2).

Figure S10. (A) The relationship between the distance estimates to observed animals based on the ultrasonic hypsometer measurements (D_{vertex}) in the Bavarian Forest National Park (BFNP) and their differences with the manual estimates and (b) the hypsometer-based estimates plotted against the manual distance estimates. A difference of zero is indicated by the red line in (A) and the solid grey line in (B). The trend line of the actual data is shown with a red dotted line in (B).

Figure S11. Histogram of the semi-automatically estimated observation distances for each species in the two study areas Schorfheide-Chorin (S-C) and Bavarian Forest National Park (BFNP). The bins were chosen according to the rounding intervals in the respective study area. The colours of the bars represent the predicted absolute difference between the semi-automatically (D_a) and manually (D_m) estimated distances in the respective bin with the GAM that was also used for Figure 3B and Figure S3B, averaged for 0.1 m intervals. The vertical grey dashed lines indicate the truncation distances.

Table S1. Manual distance estimation process for animal observations on camera trap photos based on reference images depicting a person (Schorfheide-Chorin) or a ranging pole (Bavarian Forest National Park).

Table S2. Number of camera trap locations, sample sizes of manual and automatic detections, truncation distances, selected keys for detection functions, detection probabilities and population density estimates based on the manual and semi-automatic workflows and their differences.

Table S3. Regression coefficients and emtrends of the number of animal observations detected by MegaDetector in different distances from the camera traps, per animal species and study area.

Table S4. Model summaries and ANOVA of the two models describing the relationship of the semi-automatic

distance estimate and its difference to the manual distance estimate of the same animal observation (S-C = study area Schorfheide-Chorin, BFNP: study area Bavarian Forest National Park).

Table S5. Coefficients of variation (CV) of the population density estimates and 95% confidence intervals at an assumed CV of 0.2 were computed using the Wald method.

Table S6. Estimated time and effort for the processing of the dataset used in the present study (77 camera trap locations and 63 791 animal observations) with the manual and semi-automated distance estimation workflows.