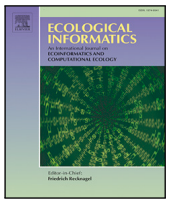
Ecological Informatics 95 (2026) 103682



Contents lists available at ScienceDirect

Ecological Informatics

journal homepage: www.elsevier.com/locate/ecolinf





Distance-based population density estimates from camera traps: Strong impact of technical choices in automated pipelines

Phil Garthen [a,e,*], Hari Surya Charan Mudragada [b], Stefanie Schwarz [c], Hendrik Edelhoff [d], Manuel J. Steinbauer [a,e], Wibke Peters [d], Kagan Kuecuekaytekin [b], Ludwig Bothmann [c,f]

[a] Bayreuth Center of Sport Science (BaySpo), University of Bayreuth, Universitätsstr. 30, 95447 Bayreuth, Germany
[b] Department of Computer Science, Technical University of Munich (TUM), Boltzmannstr. 3, 85748 Garching, Germany
[c] Department of Statistics, Ludwig-Maximilian-University Munich (LMU), Ludwigstr. 33, 80539 München, Germany
[d] Research Unit Wildlife Biology and Management, Bavarian State Institute for Forestry (LWF), Hans-Carl von Carlowitz Platz 1, 85354 Freising, Germany
[e] Bayreuth Center of Ecology and Environmental Research (BayCEER), University of Bayreuth, Universitätsstr. 30, 95447 Bayreuth, Germany
[f] Munich Center for Machine Learning (MCML), Ludwigstr. 33, 80539 München, Germany

ARTICLE INFO

Dataset link: https://github.com/slds-lmu/wildlife-experiments-server/releases/tag/v1.0.0, https://doi.org/10.5281/zenodo.18282969

Keywords:
Camera trap distance sampling
Deep learning
Image segmentation
Monocular depth estimation
Model sensitivity
Population density

ABSTRACT

Camera Trap Distance Sampling (CTDS) enables population density estimation of unmarked species from camera trap data by incorporating animal-to-camera distances. In this study, we systematically evaluate the sensitivity of CTDS-derived density estimates to variation in key pipeline parameters, using a high-resolution camera trap dataset from Central Europe. Over a 12-month period, 110 cameras were deployed across a forested area of 94 km^2. Animal-to-camera distances were estimated via monocular depth estimation. To reduce manual annotation effort, we introduce an automated masking approach for calibration images based on the Segformer model, saving an estimated minimum of two minutes per image. Using red deer (*Cervus elaphus*) as model species, a full-factorial sensitivity study (n = 135 parameter configurations) was conducted, varying MegaDetector confidence threshold, depth sampling method, mean trigger intervals and right-hand distance truncation schemes. Results show that population density estimates are highly sensitive to the choice of depth sampling method, MegaDetector confidence threshold and mean trigger interval. Parameter-induced variation led to differences of up to 200% in density estimates relative to the lowest estimate, twice as much as the variation observed between spatial subsets of the study area differing in vegetation characteristics. However, relative differences between the two sub-areas remained largely consistent across configurations. These results highlight the critical impact of technical choices within CTDS workflows and emphasize the need for transparent reporting and standardization of key pipeline parameters. All CTDS analyses using automated pipelines should therefore be accompanied by dataset-specific sensitivity analyses to account for parameter uncertainty and validation challenges. Overall, our study demonstrates both the methodological challenges and practical applicability of automated CTDS for large-scale wildlife monitoring.

1. Introduction

Monitoring of wildlife abundance and density is fundamental for quantifying population dynamics, understanding biotic interactions, and designing effective management and conservation strategies (Callaghan et al., 2024; Williams et al., 2001). Camera trapping plays a significant role in this process by providing non-invasive, continuous, and extensive data collection on species presence, behavior, and population trends (Caravaggi et al., 2017).

For species where individuals can be individually identified (e.g., some fields), capture-recapture models allow reliable population density estimates (Royle et al., 2008). However, most wildlife species lack distinctive features for individual recognition. Here, Camera Trap Distance Sampling (CTDS) is increasingly recognized as a valuable method to estimate population density of wildlife species. If cameras are distributed randomly with respect to both habitat characteristics and animal movement patterns (i.e., not preferentially along trails or roads), this method can provide reliable population density estimates across a range of ecological settings (Howe et al., 2017).

In CTDS, the detection probability of an animal is modeled as a function of its distance from the camera, accounting for decreasing detectability with increasing distance. Distance estimates have been

* Correspondence to: Bayreuth Center of Sport Science, Department of Sport Ecology, Universitätsstr. 30, Germany.
E-mail address: phil.garthen@uni-bayreuth.de (P. Garthen).

https://doi.org/10.1016/j.ecoinf.2026.103682
Received 22 August 2025; Received in revised form 18 January 2026; Accepted 26 February 2026
Available online 12 March 2026
1574-9541/© 2026 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

commonly derived from manually comparing animal locations to reference objects at known distances from the camera (Caravaggi et al., 2016). However, this method requires extensive laborious effort. To address this issue, Haucke et al. (2022) proposed an automated pipeline based on monocular depth estimation and depth image calibration, reducing the manual effort by a factor greater than 20.

While automatization in the CTDS pipeline significantly improves efficiency, it also introduces new challenges regarding the accuracy of population density estimates (Haucke et al., 2022). Previous studies showed that CTDS-based population density estimates are relatively robust to image loss arising from automatization (Zampetti et al., 2024; Henrich et al., 2024), but it has never been systematically quantified to which extend they are affected by decisions and errors in the automated pipeline. To address this gap, we assess how variation in the following pipeline parameters propagates to CTDS-based population density estimates. Specifically, we examine the influence of four key pipeline parameters:

(1) Confidence threshold for object detection: An important prerequisite for automatizing population density estimation via CTDS is the identification of animals on photos via artificial intelligence (AI; Zampetti et al., 2024): The MegaDetector as an AI-based object detection model is able to find and localize animals on camera trap images from diverse ecosystems worldwide (Beery et al., 2019). Each detected animal is assigned a bounding box and accompanied by a confidence score. As typical for general AI models, some previous studies applying MegaDetector rely on comparably high thresholds of 0.7 to 0.95 (Leorna and Brinkman, 2022; Mitterwallner et al., 2024). However, with new updates or additional training data, the optimal threshold may change. Bothmann et al. (2023) argue that the optimal threshold can be a performance-relevant hyperparameter worth to being tuned, as a higher probability to detect animals with low thresholds is accompanied by an increase in false detections (e.g., empty images). Consequently, the confidence threshold directly affects both the quantity and the reliability of animal detections.

(2) Depth sampling model: Within each bounding box, the depth relative to the camera location can vary significantly. For example, the animal itself will be closer to the camera than vegetation in the background. Several approaches have been suggested to automatically sample a depth that coincides with the animal's position. A straightforward option is to calculate the depth at the bottom center of the bounding box (Henrich et al., 2024), as the bottom line roughly represents the standing position of the animal. Haucke et al. (2022) suggest that the probability of the sampled depth to lie directly on the detected animal will be maximal, when sampling at the 20th percentile of all depths contained inside the bounding box. Alternatively, each animal can be assigned a mask in addition to the bounding box, using an object identification AI such as Detectron2 (Wu et al., 2019) or Segment Anything (Kirillov et al., 2023). In this way, the depth can be sampled from a representative part of the animal's mask (essentially a center-like point).

(3) Mean trigger interval: In CTDS, population densities are estimated as a function of effective survey effort, which is defined as camera runtime divided by the number of potential detection events at which an animal could be photographed if continuously present in front of the camera, that is, the number of so-called snapshot moments (Buckland, 2001). The number of snapshot moments results from the mean time interval between consecutive triggers generated by the same individual remaining within the camera's field of view (FOV). Underestimating this interval leads to an overestimation of effective survey effort and, consequently, to an underestimation of density. Both technical recovery times and species-specific, behavior-driven retrigger delays affect trigger intervals, because animals present in front of a camera do not constantly move (e.g., due to sleeping, grazing, or freezing in reaction to the camera), and therefore do not immediately reactivate the motion sensor. While the first is specified by the manufacturer and can also be validated experimentally, the latter is highly species-specific and depends on population densities, habitats, and camera characteristics. Therefore, mean trigger intervals must be inferred directly from the respective dataset, but manual checks are time-consuming and often unreliable, because individuals are hardly recognizable and animals may temporarily leave the camera's FOV without the possibility of detecting this. To address this, Kühl et al. (2023) present a statistical approach for estimating dataset-specific mean trigger intervals. However, the results are highly sensitive to the truncation threshold used to decide which intervals to include in the calculation of the mean: Choosing too large thresholds mixes in many intervals from different individuals or between the same individual leaving and returning later, whereas too small thresholds cut off intervals that actually stem from the same individual. The selection of an appropriate truncation threshold is therefore a trade-off, for which objective selection criteria are not available.

(4) Maximum distance of observations: Right-hand truncation of the distance data helps to select more realistic conditions. Depth beyond a certain threshold may not be reliable or relevant (Buckland, 2001). The actual maximum visible distance between camera and animal depends on site-specific characteristics such as vegetation. Additionally, it is expected to differ between species and habitats. E.g., large animals in open habitats are more likely to be detected at far distances compared to small animals in dense forests (Zuleger et al., 2022).

Our systematic approach aims to address the following research questions: (a) Do variations in parameter configurations within an automated CTDS pipeline translate into differences in population density estimates? If so, (b) which parameters require the most attention? Finally, we address the risk that (c) variation induced by pipeline parameter settings may outweigh natural spatial variation. To explore this, we estimate red deer (*Cervus elaphus*) population densities within the Veldensteiner Forest in Northern Bavaria, Germany, focusing particularly on the contrast between the eastern and western sections, which are separated by a river and a parallel railway line, and differ in vegetation characteristics. These real-world differences serve as a reference framework to assess whether the density differences produced by model configurations are meaningful or negligible in the context of our study area.

2. Materials and methods

2.1. Study site

Camera trap data were collected in the Veldensteiner Forest and adjacent forest areas. The 94 km^2 study site is located in Northern Bavaria (Germany) and part of the nature park "Fränkische Schweiz and Frankenjura". It is partitioned into two subregions by the river Pegnitz and an adjacent railway line that follows its course (Fig. 1). The mountain range within this region spans elevations from 440 to 610 m above sea level. A wide range of wildlife species inhabits the Veldensteiner Forest, including ungulates (e.g., roe deer (*Capreolus capreolus*), red deer (*Cervus elaphus*), wild boar (*Sus scrofa*)) and carnivores (e.g., wolf (*Canis lupus*), red fox (*Vulpes vulpes*), wild cat (*Felis silvestris*)). The forest is predominantly shaped by human influence, including significant recreational use, hunting, and forestry operations. The major part of the area, especially in its western part, is covered by coniferous forests with the dominant tree species Scots pine (*Pinus silvestris*) and Norway spruce (*Picea abies*). The distribution of the main deciduous species, *Fagus sylvatica*, is largely concentrated in the south-eastern part.

2.2. Camera trap deployment and data collection

A total of 110 camera traps were placed in a grid of 950 m distance to each other, following a systematic design with a random start. All potential sites located in fenced areas or human settlements were excluded. Cameras were installed approximately 50 cm above the

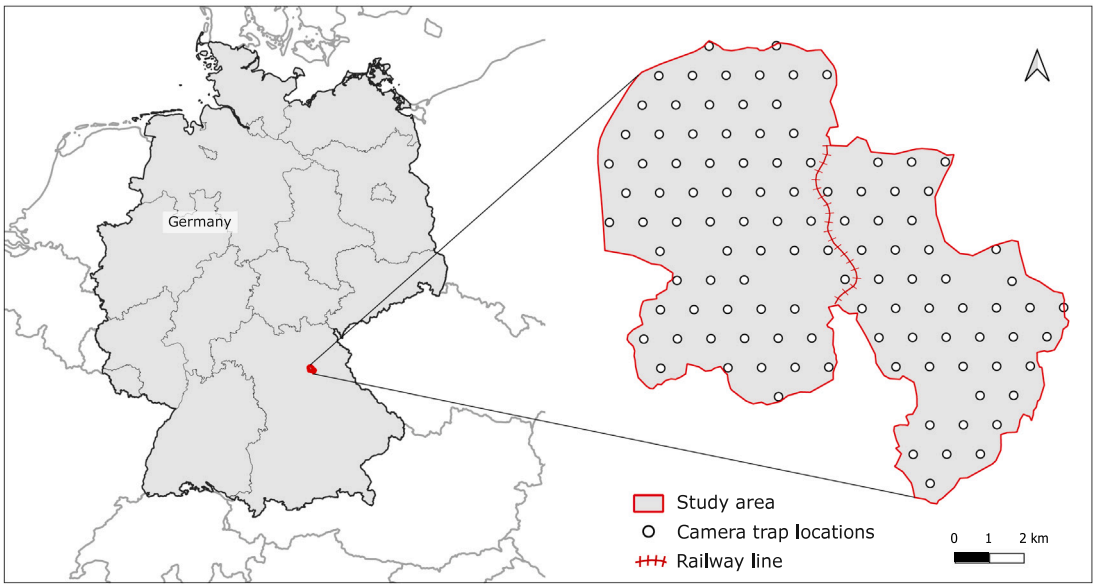


Fig. 1. Location of study area within Germany and grid of camera traps within the study area.

Table 1
Overview of camera settings.

Camera model	Cuddeback Power House Black G-5031
Recording mode	Multi-shot mode: 5 pictures series, minimum delay + Time lapse mode: one picture per day
Horizontal field of view	~40°
Vertical field of view	~31.5°
Flash type	Black flash (no visible glow at night)
Flash range	Far (~30 m)
Image format	Full (4 × 3)
Image size	5 MP
Zone control	Wide

ground, with an angle parallel to the ground. The deployment period of the cameras was from September 2020 to August 2021.

Camera traps that malfunctioned during deployment were assumed to have failed at the time of their last recorded photo. Each camera took at least one image per day (time lapse mode), so any uncertainty regarding the exact failure time is negligible. Further camera settings are specified in Table 1.

Camera trap images were classified as animals, humans, vehicles and empty images using the object detection model MegaDetector (MDv4.1; Beery et al., 2019), with low-confidence classifications being manually reviewed to prevent the loss of animal images. Subsequently, all images containing red deer (n = 13,909) were manually selected. Red deer was chosen as model species due to its high management relevance in the study area.

2.3. Depth estimation

A critical component of the CTDS workflow described in Section 2.4 is the estimation of the distance between an animal and the camera. Our approach for deriving these distances from camera trap images uses state-of-the-art deep learning methods, with the goal of minimizing manual effort. We build upon the pipeline proposed by Haucke et al. (2022) for generating absolute depth maps and extend it with a preprocessing step that automates the masking of calibration images. We used release version v.1.6.5 of the respective python package, which includes a revised computation of world coordinates based on an implementation issue identified during the course of this study.

2.3.1. Pipeline of Haucke et al. (2022)

The pipeline has two primary workflows: (1) calibration and (2) distance estimation:

The goal of (1) calibration is to construct a reliable depth scale for each camera location. This requires a set of calibration images for each camera location with masked landmarks at known distances. To achieve this, each calibration image is prepared with a person standing at a known distance from the camera while holding a sign indicating the distance (Fig. 3A). The standard approach is then to manually draw masks around the landmarks, which works well but requires a lot of manual effort. Here, our own contribution improves efficiency by automating this step (see Section 2.3.2). These annotated calibration images then undergo monocular depth estimation using the Dense Prediction Transformer (DPT) model (Ranftl et al., 2021). This process results in a series of uncalibrated depth disparity images that are then aligned using a RANSAC-based method (Fischler and Bolles, 1981), such that the depth readings of the landmarks correspond to their actual values. This alignment produces a unified depth scale across all reference images, which serves as a reliable baseline for subsequent distance measurements for one camera location. For each of the 110 camera trap locations in our study, around 3 to 27 calibration images were captured, depending on the visibility range within the camera's FOV.

In (2), the distance estimation workflow, disparity maps are also generated for animal observation images using monocular depth estimation with the DPT model. These disparity maps are then aligned with the calibrated reference. Fig. 2 illustrates the intermediate steps in this workflow: Starting with a camera trap image (left image), animals are localized automatically by the MegaDetector model (Beery et al., 2019), which generates bounding boxes for each detected animal based on a predefined confidence threshold (red rectangle). In the following, the distance estimation pipeline estimates a depth image, assigning a distance value to each pixel (center image), which is then aligned with the calibrated reference obtained from the calibration workflow (right image). Finally, a representative sample point within the bounding box (⋄) is selected. Three different depth sampling methods were implemented (see Section 2.5). The sample point's depth becomes the animal-to-camera distance (6 m in the example). We selected 30 m as maximum distance in the pipeline. A manual inspection of distance distributions revealed one location with failed depth estimation due to sunglare, which was removed from subsequent analyses.

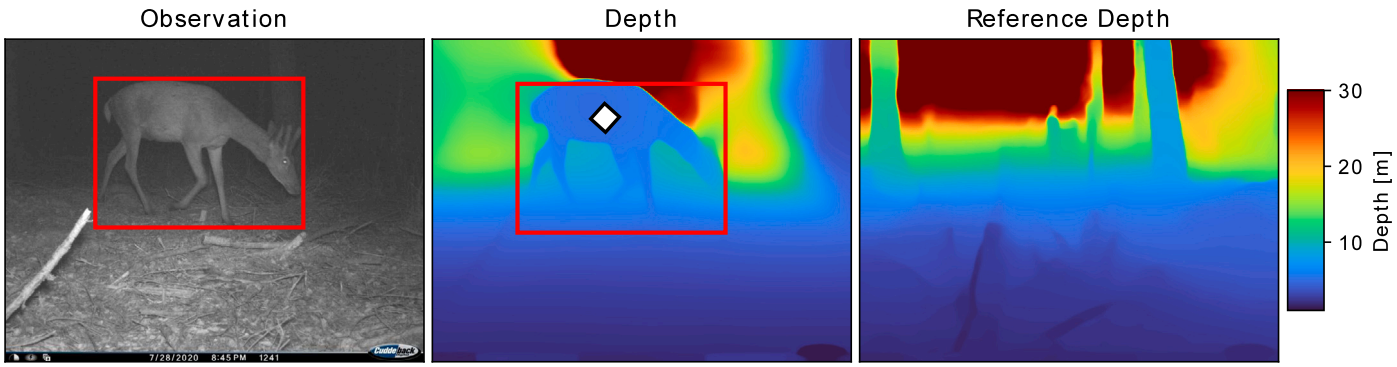


Fig. 2. Depth estimation for a camera trap scene (exemplary depth sample point marked as ⋄).



(A) (B)

Fig. 3. Automated masking of calibration images by Segformer. (A) Calibration image at distance 3 m and (B) mask for person in calibration image.

2.3.2. Our proposal: Automated calibration image masking

Given the large number of images required across multiple locations, manually drawing masks for each image, as proposed in the original workflow by Haucke et al. (2022), would be highly labor-intensive and impractical. Consequently, automating the masking step becomes essential to maintain both efficiency and consistency across datasets of this scale.

To automate the process, a segmentation model is employed to identify and label specific area in each image. In this pipeline, the Segformer model (Xie et al., 2021) is used, as it combines both segmentation and labeling tasks, with various pretrained weights offering different label options and performance scores. Two segmentation approaches were explored: segmenting the sign itself and segmenting the person holding it. The latter approach proved to be more reliable, as it covers a larger segmented area (Fig. 3). By automating this process, we reduce the manual effort by about two minutes for each calibration image.

2.4. Camera Trap Distance Sampling (CTDS)

For calculation of population density estimates $\hat{D}$, CTDS was applied (Buckland, 2001; Howe et al., 2017), following the Eq. (1):

$$\hat{D} = \frac{2t \sum_{k=1}^{K} n_k}{\theta_v \omega^2 \sum_{k=1}^{K} T_k \hat{P}} \frac{1}{\hat{A}} \quad (1)$$

t = Mean trigger interval [s]
K = Number of camera trap locations
n_k = Number of individuals on the first photo of a photo series at location k
θ_v = Horizontal angle of view of the camera [rad]
ω = Truncation distance beyond which distances are discarded [km]
T_k = Camera deployment time at location k [s]
$\hat{P}$ = Detection probability
$\hat{A}$ = Activity level (proportion of daily time where target species is active)

Analysis were conducted in R 4.2.3 (R Core Team, 2023) using the Distance package (version 2.0.0; Miller et al., 2019).

The first photos of each series were defined as snapshot moments. Mean trigger intervals t were estimated based on our own survey data: As proposed by Kühl et al. (2023), we calculated mean time intervals between pairs of consecutive pictures supposed to be triggered by the same individual, which ranged from 9.86 and 11.46 s depending on the truncation of the time interval distribution (see Section 2.5). In this way, periods where animals are not photographed though present in front of the camera, are accounted for.

The activity level $\hat{A}$, i.e. the estimated proportion of time animals were active, was derived from circular kernel distributions fitted to the daily distribution of independent events using the 'activity' package (version 1.3.4; Rowcliffe, 2023), allowing the complete 24-hour data to be included. We defined an event as independent based on a minimum time interval of five minutes between consecutive images, following the same principle as Rovero and Marshall (2009) and Gray (2018) (Supplementary Figure A.1). Activity levels were calculated as yearly average, with uncertainty assessed via 1000 bootstrap replicates implemented in the fitact() function (Supplementary Figure A.2).

The camera's FOV was provided by the camera manufacturer. Additionally, detection angles θ_v were computed from the results of the depth estimation pipeline. Our data show a uniform horizontal distribution of events within a FOV of 40° (Supplementary Figure A.3). Observations outside a FOV of 40° were discarded (Howe et al., 2017).

For animals that were detected directly in front of the camera, species identification often failed because only a part of the animal was visible. This resulted in fewer than expected detections at small distances, so the data were left truncated at 1.5 m. Likewise, images at large distances were discarded, as reliability of species identification

and distance estimation decreases also with increasing distances (see Section 2.5; Buckland, 2001).

Detection probabilities $\hat{P}$ were fitted using eight candidate detection functions, including uniform key functions with 1, 2 and 3 cosine adjustments, half normal key functions with 0, 1 and 2 cosine adjustments and hazard rate key functions with 0 and 1 simple polynomial adjustments. To account for potential overdispersion and avoid overly complex models, we applied the two-step model selection procedure suggested by Howe et al. (2019). First, the best model within each key function was identified using Akaike's information criterion adjusted by overdispersion (quasi-AIC), with the overdispersion parameter $\hat{c}$ estimated as the ratio of the chi-square (χ^2) statistic from the most highly parameterized model within each key function to its degrees of freedom. Second, among these candidates, the model with the lowest overdispersion was chosen.

Density estimates were calculated as yearly averages for the hole study area as well as for the two subregions separately. Uncertainties were derived using a design-based approach (method "P2" of Fewster et al. (2009)), which estimates variance by post-stratifying adjacent camera locations and weighting within-stratum variability relative to its sampling effort. This approach thereby accounts for the systematic camera placement with a random start and unequal sampling effort among locations. Density estimates were statistically compared running pairwise Wald tests, with statistical significance determined from the χ^2 distribution with one degree of freedom (alpha = 0.05; Wald and Wolfowitz, 1940).

2.5. Sensitivity study

We conducted a sensitivity study, focusing on how parameter variation in automated CTDS pipelines influence the resulting population density estimates. The parameter grid for this sensitivity study is defined as follows (Fig. 4).

Key parameters to be defined within the depth estimation pipeline by Haucke et al. (2022):

- (1) MegaDetector confidence threshold: 10%, 30%, 50%, 70%, 90%
  MegaDetector (MDv5a.0.0) was used to localize animals in the images of the preselected dataset (Beery et al., 2019). The MegaDetector's confidence threshold is varied to assess its influence on the number of animals detected. A range of confidence levels, from very low (10%) to very high (90%), is tested to evaluate how the model's certainty in detecting animals affects the resulting population density estimation. The effect of the MegaDetector confidence threshold on the absolute number of detections is shown in Supplementary Table A.1.
- (2) Depth sampling method: BBOX_BOTTOM, BBOX_PERCENTILE, SAM
  To sample the depth within the bounding box, three different methods are explored. As initially mentioned by Haucke et al. (2022) and Henrich et al. (2024), each method offers a different approach to handling depth estimation within the bounding boxes. We sampled
  (a) the lower central pixel of the bounding box (BBOX_BOTTOM), based on the assumption that the detected animal is standing on the ground;
  (b) the 20th percentile of all depth values within each bounding box (BBOX_PERCENTILE), which may offer a more accurate depth estimate despite potential variability in field conditions; and
  (c) the centroid of the detected animal, applying the Segment Anything Model (SAM) (Kirillov et al., 2023) for segmentation within the bounding box. The SAM approach required approximately twice the computational time compared to the bounding-box-based methods.

Key parameters to be defined in the downstream steps of the CTDS pipeline:

- (3) Mean trigger interval: 9.86 s, 11.21 s, 13.46 s
  Mean trigger intervals are affected by the truncation threshold used to decide which intervals to include in the calculation of the mean. Based on the approach of Kühl et al. (2023), we applied truncation thresholds of 15.5 s, 40.5 s, and the third quartile of the 0–5 min time-interval distribution aiming to exclude trigger intervals from different individuals, from individuals temporarily leaving the camera's FOV or being hidden by vegetation. These thresholds yielded estimated mean trigger intervals of 9.86 s, 13.46 s and 11.21 s, respectively.
- (4) Maximum distance truncation: 15 m, 30 m, 85%
  The maximum distance is capped at 15 m or 30 m, with distances beyond that being truncated (right-hand truncation). The 30 m cut-off corresponds to the camera manufacturer's specified infrared flash range. Each truncation results in varying portions of the data, ranging between 66% (SAM, MegaDetector confidence threshold = 30%) and 95% (BBOX_BOTTOM, 90%) for truncation at 15 m and between 86% (SAM, 30%) and 99% (BBOX_BOTTOM, 90%) for truncation at 30 m. Therefore, a uniform right-hand truncation of 15% of the largest observed distances was additionally applied to each individual dataset, resulting in truncations distances between 10 m (BBOX_BOTTOM, 90%) and 28 m (SAM, 30%).

By combining these different parameter settings, the total number of experiments conducted in the sensitivity study is 135.

3. Results

3.1. Effects of pipeline parameter variation on population density estimates

We systematically tested the sensitivity of population density estimates to parameter variation in automated pipelines. Overall, substantial variation in density estimates was observed across the tested parameter configurations. Variation of the depth sampling method had the highest impact, followed by the mean trigger interval and the MegaDetector confidence threshold (Fig. 6D). The largest positive deviation from the overall mean was obtained using the depth sampling method BBOX_BOTTOM in combination with low MegaDetector confidence thresholds and large mean trigger intervals, irrespective of maximum distance truncation (+40%). In contrast, the largest negative deviation was produced using the depth sampling method SAM with high MegaDetector confidence thresholds and small mean trigger intervals, reaching up to –20% deviation from the overall mean (Fig. 6A). Pairwise Wald comparisons partially revealed significant differences among pairs of density estimates. Except for one configuration (SAM-10%-13.46 s-15 m), all parameter configurations differed significantly from at least one other. The largest density estimate (BBOX_BOTTOM-10%-13.46 s-30 m) exhibited the largest number of significant differences to other estimates (88). When only a single parameter was varied while holding the other three constant, significant differences between density estimates were observed only for variations in the depth sampling method. These differences involved nine pairwise comparisons between BBOX_BOTTOM and SAM.

3.1.1. Effects of parameter variation within depth estimation pipeline

The depth sampling method had a visible impact on the distribution of estimated distances (Fig. 5). While BBOX_BOTTOM resulted in the smallest average distance estimates, SAM yielded a comparatively high number of larger distances, whereas BBOX_PERCENTILE produced intermediate results. Variation of the MegaDetector confidence threshold had mostly no influence on the distance distribution, except for a high threshold of 0.9, which resulted in relatively more close detections (Supplementary Fig. A.4).

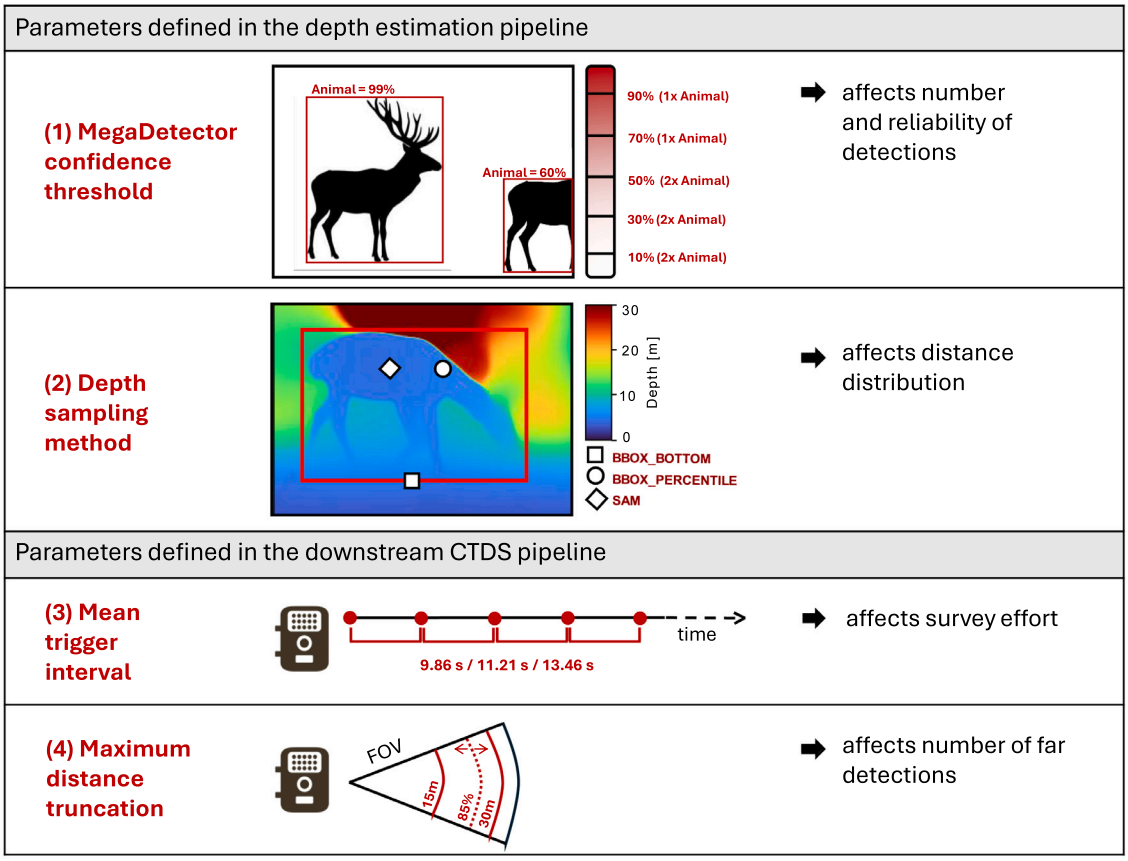


Fig. 4. Overview of pipeline parameters included in the sensitivity study.

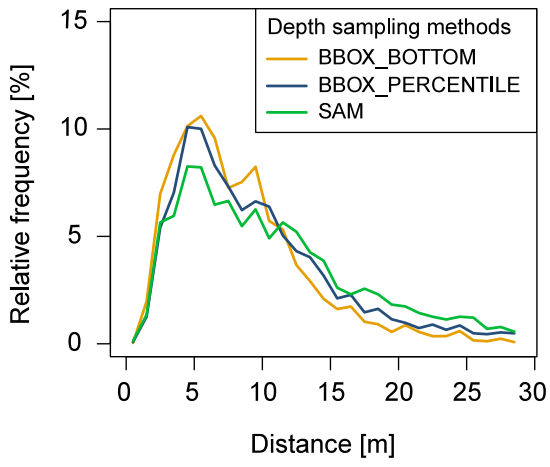


Fig. 5. Relative frequency of detections per meter by depth sampling method. Data for MegaDetector confidence threshold = 50% (pattern is representative of all tested thresholds).

Population density estimates resulting from BBOX_BOTTOM were on average (across all other parameter configurations) 29% higher, those resulting from SAM 11% smaller than the estimates resulting from BBOX_PERCENTILE. Decreasing MegaDetector confidence thresholds (90% to 10%) increased population density estimates by 23%, on average. An exemplary set of parameter configurations is depicted in Fig. 6B.

3.1.2. Effects of parameter variation within downstream CTDS pipeline

Population density estimates based on a mean trigger interval of 13.46 s were 20% higher, and those based on 9.86 s 12% lower than estimates obtained using 11.21 s. Instead, variation introduced by varying maximum distance truncation schemes was negligible (~5%). An exemplary set of parameter configurations is depicted in Fig. 6C.

3.2. Spatial variation in population density estimates

In the western part of the study area, population density estimates were on average twice as high as in the eastern part (Fig. 7). Estimates from 105 of 135 parameter configurations differed significantly between sub-areas, with non-significant results being dominated by configurations using a MegaDetector confidence threshold of 90%. Depending on the selected parameter configuration, density estimates for the eastern part ranged from approximately 0.3 to 0.8 individuals per km^2, while estimates for the western part ranged from 0.5 to 1.5 individuals per km^2. Thus, depending on parameter choices, density estimates varied by up to a factor of three within the same area.

4. Discussion

Our study systematically evaluated how wildlife population density estimates are influenced by different parameter configurations within an automated CTDS pipeline. We found that population density estimates were significantly affected by both the depth sampling method and the MegaDetector confidence threshold. Density estimates were further influenced by mean trigger intervals, while the selected maximum detection distance had a minimal effect. Notably, parameter-induced variation (up to 200% difference relative to the lowest density estimate) exceeded the considerable spatial variation in animal population densities between the western and the eastern part of the study area (~100% difference).

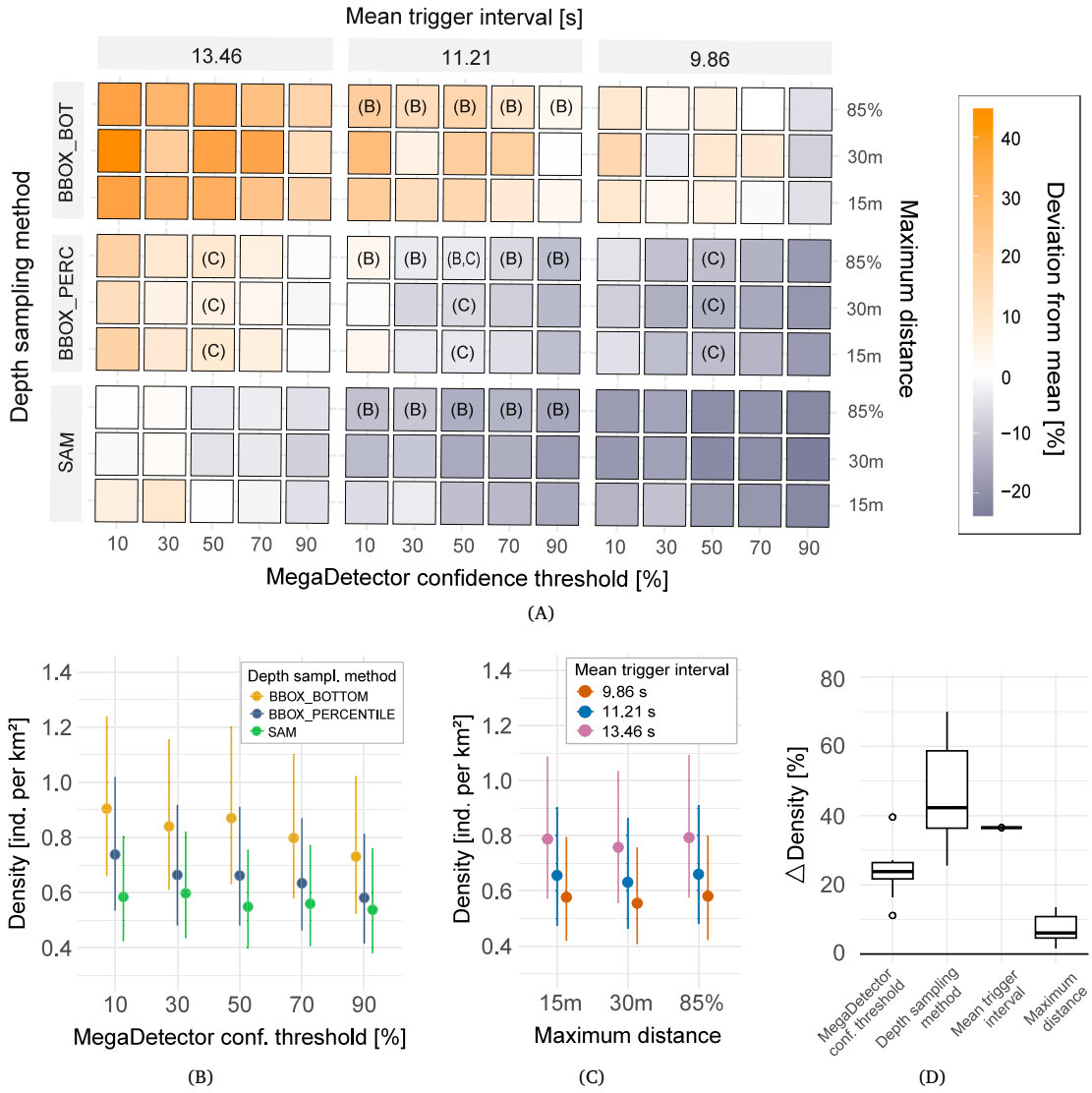


Fig. 6. Population density estimates and their variation across different parameter configurations. Panel (A) shows the percentage deviation from the overall mean of population density estimates across all tested parameter combinations. Panels (B) and (C) display exemplary population density estimates (individuals per km^2) for specific parameter configurations marked in (A). Points represent means, error bars the 95% confidence intervals. In (B), parameters vary within the depth estimation pipeline (Depth sampling method and MegaDetector confidence threshold; estimates are based on a mean trigger interval of 11.21 s and a maximum distance truncation at 85%). In (C), parameters vary within the downstream CTDS pipeline (Maximum distance and mean trigger interval; includes only data for depth sampling method = BBOX_PERCENTILE and MegaDetector confidence threshold = 50%). Panel (D) quantifies the sensitivity of density estimates to each specific parameter. For each parameter, differences are calculated as the percent increase from the lowest to the highest estimates, while all other parameters are held constant.

4.1. Strong effects of key pipeline parameters on density estimates

Increasing the MegaDetector confidence threshold consistently reduced population density estimates. This trend reflects the trade-off between false positives and false negatives: higher thresholds reduce false positives but also exclude true detections, thus reducing the total number of detections used in density estimation. Our results confirm findings by others (e.g., Beery et al., 2019; Bothmann et al., 2023), who observed that confidence threshold settings significantly affect detection rates and downstream ecological metrics. Notably, CTDS largely corrected for the number of detections, underscoring findings of previous studies (e.g., Zampetti et al., 2024). In our analyses, 2.3 times more detections were included for a threshold of 90% compared to 10%, yet the resulting density estimates increased by less than 50%, on average. Close detections are most important for selecting an appropriate detection function (see below), and with high thresholds (e.g., 90%), more detections near the camera will be missed, making the method not entirely immune to detection number variation in response to changes in the MegaDetector confidence threshold. In our study, a confidence threshold of 10% resulted in the number of detections (n = 3039) most similar to the number of manual detections (n = 3569). This suggests that still not all individuals were detected by the MegaDetector with the lowest tested confidence threshold. However, most undetected animals are located in the background of the image and will probably not significantly affect the density results due to the right-hand truncation applied in the density estimation pipeline. The low optimal MegaDetector confidence threshold stands in contrast to other studies (Mitterwallner et al., 2024; Zampetti et al., 2024), which recommend thresholds ranging from 70 to 95%. However, these studies rely on earlier versions of MegaDetector. Our results indicate that the updated version used here (MDv5a.0.0) performs best with considerably lower confidence thresholds. This finding is consistent

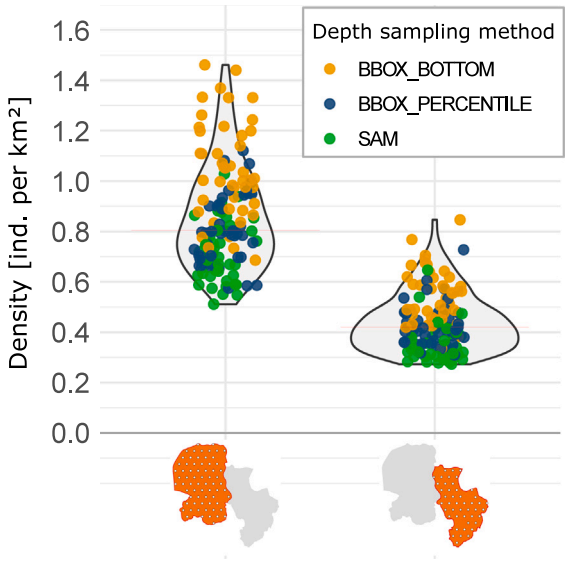


Fig. 7. Comparison of population density estimates between the western (left) and eastern (right) parts of the study area. Red lines represent medians. All tested parameter configurations included.

with reports from the model developers, who note that optimal confidence thresholds can vary substantially between model versions, with more recent releases allowing for markedly lower thresholds (down to 15%; Beery et al., 2019). In our study, threshold selection was further influenced by prior manual filtering of images to include only red deer, reducing the proportion of empty images and permitting lower thresholds (even below 15%) without substantially increasing false positives. This highlights that optimal confidence thresholds are highly dependent on both model version and dataset composition.

Depth sampling methods that resulted in shorter camera–animal distances produced higher population density estimates. These findings align with earlier research showing that if detections are clustered near the observer, the detection function declines more rapidly, which increases the estimated density (Buckland, 2001). Thus, the BBOX_BOTTOM method, which produces a steeper detection function, leads to higher density estimates because the model infers that more animals passing at larger distances from the camera remain undetected. On the other hand, SAM generates a higher proportion of larger distances, thus yielding lower density estimates. BBOX_PERCENTILE provides intermediate results between these two extremes. Part of the differences in distance distributions between depth sampling methods arise from the fact that animals detected far from the camera tend to blend into the background, either partially or entirely, leading SAM and BBOX_PERCENTILE to overestimate distances (Fig. 8A + B). Also, partial occlusion of the animal, e.g. by tree trunks or dense understory vegetation, can introduce bias because nearby pixels will then be overrepresented within the bounding boxes (Fig. 8C). In such cases, both the selected quartile and the bottom center of the bounding box may correspond to foreground vegetation rather than the animal's true position. In more open habitats, the observed differences in density estimates between the three depth sampling methods may be partially reduced, as the effect of occlusion is less pronounced. A systematic evaluation of the three methods across habitats with varying vegetation density and occlusion levels is needed to fully understand the effect of vegetation density on the depth estimates. Another source of bias is caused by animals being detected very close to the camera and dominating the FOV, distorting the overall depth profile used for estimation (Fig. 8D). In such cases, the depth estimation pipeline does not always recognize the animal as being directly in front of the camera, but instead interprets it as part of the background. In previous research, using the 20th percentile of depths in the bounding box and applying a linear interpolation with the bottom center pixel for distances beyond 10 m showed no statistically significant differences compared to manually measured ground-truth distances (Henrich et al., 2024). This approach may represent a compromise that, on the one hand, reduces underestimation at short distances, while also avoiding excessively large distances in the background of the image.

We also observed substantial effects of varying mean trigger intervals on population density estimates. This is expected, as density estimates in the CTDS model scale directly with the mean trigger interval (Eq. (1)). Selecting the third quartile as the truncation threshold has been proposed to result in reliable estimates: In the study by Kühl et al. (2023), CTDS density estimates based on third-quartile truncation closely matched independent estimates obtained via spatially explicit capture–recapture (Efford and Fewster, 2013), as well as random encounter model estimates (Rowcliffe et al., 2008). Another way to address issues related to the mean trigger interval would be to adapt camera settings and use dense time-lapse images (e.g., one per minute) rather than infrared-triggered images. This would create predefined snapshot events, making camera effort clearly defined. Similarly, long motion-triggered videos would allow images to be extracted from the video at a predefined interval (Howe et al., 2017). On the other hand, these approaches introduce additional challenges, including increased battery consumption, storage requirements and image processing time. In the case of time-lapse mode only, the possibility of missing rarely detected species entirely would also need to be considered, as animals could pass the camera between snapshots without being detected.

The right-hand truncation distance, instead, showed no clear influence on population density estimates, which is in line with previous studies (Zuleger et al., 2022). As reported in the literature, CTDS primarily relies on closer observations and detections far from the camera are less influential in determining population density estimates (e.g., Delisle et al., 2023).

Overall, our findings highlight the importance of the parameters defined in the depth estimation pipeline and confirm the need for accurate estimation of the mean trigger interval, all of which can cause density variations of a similar magnitude depending on parameter choices.

4.2. Parameter-induced variation exceeds spatial variation in our data set

In our study area, population density estimates were on average twice as high in the western part compared to the eastern part. The underlying reasons for this pattern remain unknown, but it is likely that the differences in densities reflect distinct habitat preferences. The two sub-areas are separated by a river and a railway line, creating a physical barrier between the western part, dominated by coniferous forests, and the eastern part, which is more rocky and has a higher proportion of deciduous trees. We expect differences in detectability to be minor, as no significant variation in understory vegetation density was observed, and the pattern appears to be specific to red deer and not influenced by other species. Despite large absolute differences in density estimates between parameter configurations, the relative differences between the two sub-areas remain largely consistent across configurations. This consistency underscores the general suitability of the method for practical applications, particularly when focusing on relative differences across space or time rather than absolute numbers. Despite this substantial spatial difference, the overall variation induced by model configurations exceeded the observed spatial variation within our data set significantly. Thus, consistent pipeline parameters are essential in practical application of automated CTDS. In particular, for assessing rare species under conservation criteria, even small changes in population densities can be of critical interest (e.g., Lankau and Strauss, 2011).

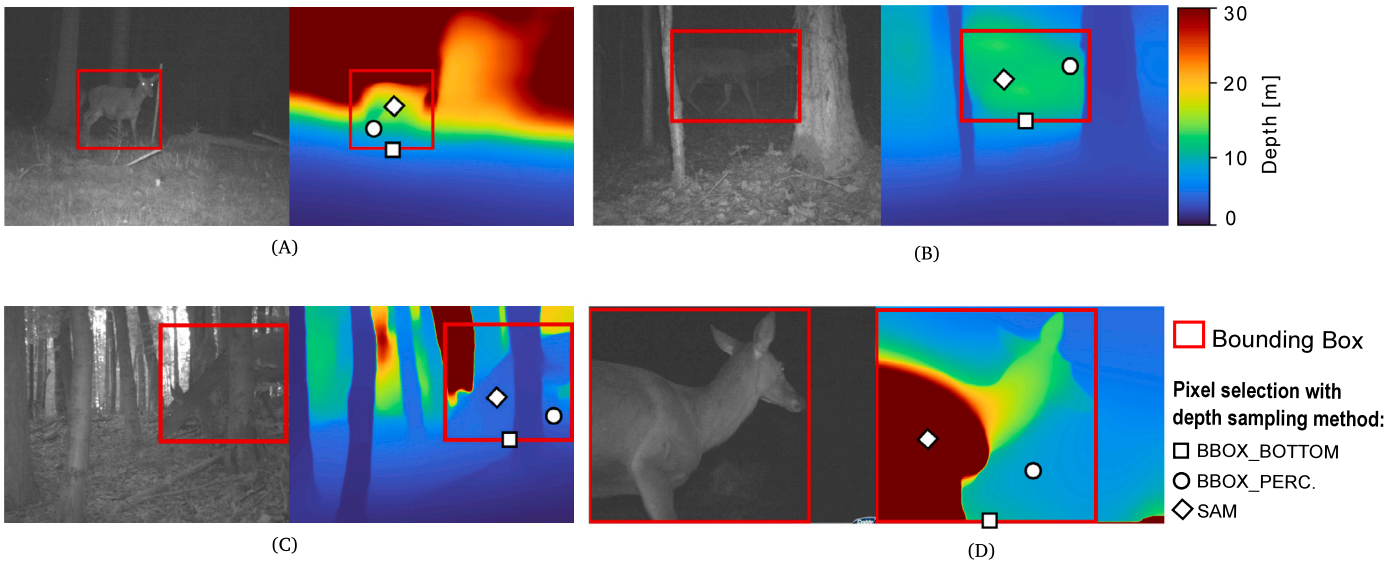


Fig. 8. Exemplary visualization of depth sampling methods and their effect on detection distances. Differences in depth estimates can arise from the animal blending (A) partly or (B) entirely into the background, (C) partial occlusion of the animal or (D) irritation caused by very close animals.

4.3. General implications for CTDS methodology

Although CTDS has been shown to be robust to random or moderate automatization errors (Henrich et al., 2024; Zampetti et al., 2024), our results demonstrate that systematic variation in pipeline parameters will still lead to substantial shifts in population density estimates. Parameter choices can act as significant sources of bias if left unexamined. This finding raises questions about the reliability and comparability of CTDS-derived density estimates, especially across studies using different parameter configurations.

Validation remains challenging, as true population densities are frequently unknown, as it was the case in our study. Also, most studies are conducted in small finite areas not subject to closure, allowing animals to freely move in and out, such that population densities are not constant over time. However, even with known population densities, multiple parameter configurations can produce similar results, as different parameters may compensate for each other. This highlights the importance of explicitly exploring and quantifying parameter sensitivity, rather than assuming that apparently accurate estimates necessarily reflect optimal or correct parameter settings. As a result, the objective selection of a concrete set of parameters is not straightforward: Parameter selection often involves trade-offs between competing objectives and optimal parameter settings are dataset-specific and depend on habitat characteristics, species-specific behavior, and camera properties. Manual verification of a subset of a dataset can partly address these uncertainties, for example, by inspecting animal detections obtained with different MegaDetector confidence thresholds, or by manually estimating animal distances. However, selecting an appropriate subset for manual validation is difficult, as numerous factors influence optimal parameter settings, including camera location, camera model, species, time, and detection distance. Future studies should therefore focus on quantifying how optimal parameter settings vary under different environmental and technical conditions, for instance by manually validating depth sampling methods in relation to detection distance and vegetation density. For the time being, when working with CTDS-derived population density estimates, we recommend performing a comprehensive sensitivity analysis for each dataset, testing multiple parameter configurations, to better understand potential sources of variance and uncertainty resulting from automated pipeline application.

5. Conclusion

Overall, our results underscore the sensitivity of CTDS-derived population density estimates to key pipeline parameters, highlighting the potential for substantial bias if parameter configurations are not carefully selected. Nevertheless, by conducting dataset-specific sensitivity analyses and thoughtfully choosing parameter settings, robust and reproducible density estimates are achievable. In this way, the automated CTDS pipeline offers substantial advantages in scalability and consistency, while emphasizing the need for careful consideration of parameter selection in future studies.

CRediT authorship contribution statement

Phil Garthen: Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Hari Surya Charan Mudragada:** Writing – review & editing, Software, Methodology. **Stefanie Schwarz:** Writing – review & editing, Software, Methodology, Conceptualization. **Hendrik Edelhoff:** Writing – review & editing, Resources, Funding acquisition, Conceptualization. **Manuel J. Steinbauer:** Writing – review & editing, Resources, Funding acquisition, Conceptualization. **Wibke Peters:** Writing – review & editing, Resources, Funding acquisition, Conceptualization. **Kagan Kuecuekaytekin:** Writing – review & editing, Software, Methodology. **Ludwig Bothmann:** Writing – review & editing, Supervision, Software, Funding acquisition, Conceptualization.

Funding

The study was funded by the Bavarian State Ministry of Food, Agriculture and Forestry (StMELF) [Grant Number JA018].

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We thank the regional forest administration (Bavarian State Forests) for granting access to the forest sites and permitting the installation of camera traps.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.ecoinf.2026.103682.

Data availability

The python code related to the distance estimation can be accessed through https://github.com/slds-lmu/wildlife-experiments-server/releases/tag/v1.0.0. Data and R code related to the density estimation is available at https://doi.org/10.5281/zenodo.18282969.

References

Beery, S., Morris, D., Yang, S., 2019. Efficient pipeline for camera trap image review. http://dx.doi.org/10.48550/arXiv.1907.06772.

Bothmann, L., Wimmer, L., Charrakh, O., Weber, T., Edelhoff, H., Peters, W., Nguyen, H., Benjamin, C., Menzel, A., 2023. Automated wildlife image classification: An active learning tool for ecological applications. Ecol. Inf. 77, 102231. http://dx.doi.org/10.1016/j.ecoinf.2023.102231.

Buckland, S., 2001. Introduction to Distance Sampling, first ed. Oxford University Press, ISBN: 978-0-19-850649-2.

Callaghan, C.T., Santini, L., Spake, R., Bowler, D.E., 2024. Population abundance estimates in conservation and biodiversity research. Trends Ecol. Evol. 39 (6), 515–523. http://dx.doi.org/10.1016/j.tree.2024.01.012.

Caravaggi, A., Banks, P.B., Burton, A.C., Finlay, C.M.V., Haswell, P.M., Hayward, M.W., Rowcliffe, J.M., Wood, M.D., 2017. A review of camera trapping for conservation behaviour research. Remote. Sens. Ecol. Conserv. 3 (3), 109–122. http://dx.doi.org/10.1002/rse2.48.

Caravaggi, A., Zaccaroni, M., Riga, F., Schai-Braun, S.C., Dick, J.T., Montgomery, W.I., Reid, N., 2016. An invasive-native mammalian species replacement process captured by camera trap survey random encounter models. Remote. Sens. Ecol. Conserv. 2 (1), 45–58. http://dx.doi.org/10.1002/rse2.11.

Delisle, Z.J., Henrich, M., Palencia, P., Swihart, R.K., 2023. Reducing bias in density estimates for unmarked populations that exhibit reactive behaviour towards camera traps. Methods Ecol. Evol. 14 (12), 3100–3111. http://dx.doi.org/10.1111/2041-210X.14247.

Efford, M.G., Fewster, R.M., 2013. Estimating population size by spatially explicit capture–recapture. Oikos 122 (6), 918–928. http://dx.doi.org/10.1111/j.1600-0706.2012.20440.x.

Fewster, R.M., Buckland, S.T., Burnham, K.P., Borchers, D.L., Jupp, P.E., Laake, J.L., Thomas, L., 2009. Estimating the encounter rate variance in distance sampling. Biometrics 65 (1), 225–236. http://dx.doi.org/10.1111/j.1541-0420.2008.01018.x.

Fischler, M.A., Bolles, R.C., 1981. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Commun. ACM 24 (6), 381–395. http://dx.doi.org/10.1145/358669.358692.

Gray, T.N.E., 2018. Monitoring tropical forest ungulates using camera-trap data. J. Zool. 305 (3), 173–179. http://dx.doi.org/10.1111/jzo.12547.

Haucke, T., Kühl, H.S., Hoyer, J., Steinhage, V., 2022. Overcoming the distance estimation bottleneck in estimating animal abundance with camera traps. Ecol. Inf. 68, 101536. http://dx.doi.org/10.1016/j.ecoinf.2021.101536.

Henrich, M., Burgueño, M., Hoyer, J., Haucke, T., Steinhage, V., Kühl, H.S., Heurich, M., 2024. A semi-automated camera trap distance sampling approach for population density estimation. Remote. Sens. Ecol. Conserv. 10 (2), 156–171. http://dx.doi.org/10.1002/rse2.362.

Howe, E.J., Buckland, S.T., Després-Einspenner, M.-L., Kühl, H.S., 2017. Distance sampling with camera traps. Methods Ecol. Evol. 8 (11), 1558–1565. http://dx.doi.org/10.1111/2041-210X.12790.

Howe, E.J., Buckland, S.T., Després-Einspenner, M.-L., Kühl, H.S., 2019. Model selection with overdispersed distance sampling data. Methods Ecol. Evol. 10 (1), 38–47. http://dx.doi.org/10.1111/2041-210X.13082.

Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., et al., 2023. Segment Anything. http://dx.doi.org/10.48550/arXiv.2304.02643.

Kühl, H.S., Buckland, S.T., Henrich, M., Howe, E., Heurich, M., 2023. Estimating effective survey duration in camera trap distance sampling surveys. Ecol. Evol. 13 (10), e10599. http://dx.doi.org/10.1002/ece3.10599.

Lankau, R.A., Strauss, S.Y., 2011. Newly rare or newly common: evolutionary feedbacks through changes in population density and relative species abundance, and their management implications. Evol. Appl. 4 (2), 338–353. http://dx.doi.org/10.1111/j.1752-4571.2010.00173.x.

Leorna, S., Brinkman, T., 2022. Human vs. machine: Detecting wildlife in camera trap images. Ecol. Inf. 72, 101876. http://dx.doi.org/10.1016/j.ecoinf.2022.101876.

Miller, D.L., Rexstad, E., Thomas, L., Marshall, L., Laake, J.L., 2019. Distance sampling in R. J. Stat. Softw. 89 (1), 1–28. http://dx.doi.org/10.18637/jss.v089.i01.

Mitterwallner, V., Peters, A., Edelhoff, H., Mathes, G., Nguyen, H., Peters, W., Heurich, M., Steinbauer, M.J., 2024. Automated visitor and wildlife monitoring with camera traps and machine learning. Remote. Sens. Ecol. Conserv. 10 (2), 236–247. http://dx.doi.org/10.1002/rse2.367.

R Core Team, 2023. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, URL https://www.R-project.org/.

Ranftl, R., Bochkovskiy, A., Koltun, V., 2021. Vision transformers for dense prediction. http://dx.doi.org/10.48550/arXiv.2103.13413.

Rovero, F., Marshall, A.R., 2009. Camera trapping photographic rate as an index of density in forest ungulates. J. Appl. Ecol. 46 (5), 1011–1017. http://dx.doi.org/10.1111/j.1365-2664.2009.01705.x.

Rowcliffe, M., 2023. Activity: Animal activity statistics. URL https://CRAN.R-project.org/package=activity, R package version 1.3.4.

Rowcliffe, J.M., Field, J., Turvey, S.T., Carbone, C., 2008. Estimating animal density using camera traps without the need for individual recognition. J. Appl. Ecol. 45 (4), 1228–1236. http://dx.doi.org/10.1111/j.1365-2664.2008.01473.x.

Royle, J.A., Nichols, J., Karanth, K., Gopalaswamy, A., 2008. Hierarchical model for estimating density in camera-trap studies. J. Appl. Ecol. 46, 118–127. http://dx.doi.org/10.1111/j.1365-2664.2008.01578.x.

Wald, A., Wolfowitz, J., 1940. On a test whether two samples are from the same population. Ann. Math. Stat. 11 (2), 147–162. http://dx.doi.org/10.1214/aoms/1177731909.

Williams, B.K., Nichols, J.D., Conroy, M.J., 2001. Analysis and Management of Animal Populations, first ed. Academic Press, New York, ISBN: 978-0-12-754406-9.

Wu, Y., Kirillov, A., Massa, F., Lo, W.-Y., Girshick, R., 2019. Detectron2: A PyTorch-based modular object detection library. https://github.com/facebookresearch/detectron2.

Xie, E., Wang, W., Yu, Z., Anandkumar, A., Alvarez, J.M., Luo, P., 2021. SegFormer: simple and efficient design for semantic segmentation with transformers. In: Proceedings of the 35th International Conference on Neural Information Processing Systems.

Zampetti, A., Mirante, D., Palencia, P., Santini, L., 2024. Towards an automated protocol for wildlife density estimation using camera-traps. Methods Ecol. Evol. 00, 1–13. http://dx.doi.org/10.1111/2041-210X.14450.

Zuleger, A.M., Holland, R., Kühl, H.S., 2022. Deriving observation distances for camera trap distance sampling. Afr. J. Ecol. 60 (3), 423–432. http://dx.doi.org/10.1111/aje.12959.