Utilizing digital elevation models for camera-trap distance estimation

Leopold Böss[1], Stephanie Wohlfahrt[2] and David C. Schedl[1]

[1]Digital Media Lab, University of Applied Sciences Upper Austria, Hagenberg, Austria
[2]Büro für Wildökologie und Forstwirtschaft, Klagenfurt am Wörthersee, Austria

Abstract

Camera traps have become a norm for estimating wildlife populations utilizing random encounter models to express metrics such as abundance and density. One essential parameter required to evaluate such models is the speed of the trapped animal. This speed can be estimated by labeling the contact point of instances with the ground across image sequences and projecting the resulting pixels onto a model describing said ground. Our approach proposes using digital elevation models recorded via airborne laser scanning as an alternative to manually calibrating simpler ground models. A study to investigate the impact of DEMs' level of complexity on the accuracy of projected pixel distance estimates was conducted using a realistic dataset of 2629 2D labels and DEMs of three different resolutions. Reducing the resolution of the used DEM from 3 m to 20 m and 50 m leads to an average skew in distances of 2.84 m and 3.94 m, respectively, with widely dispersed individual errors. Further work is needed to assess the impact of these errors on speed and wildlife density calculations. Despite these challenges, the method shows promise as an alternative to currently used methods.

Keywords

wildlife density estimation, elevation models, camera traps, biodiversity monitoring, geospatial techniques

1. Introduction

Understanding system metrics, such as density and abundance, is crucial for effective wildlife management [1, 2]. In reality, however, it is seldom feasible to perform exhaustive measurements of these variables, necessitating estimates based on samples and specialized models.

In recent years, camera traps—fixed cameras triggered by motion sensors—have become the preferred tool for sampling as they are non-invasive, robust, and widely applicable [3]. Camera traps can be used to generate abundance indices such as trapping rates for a quick insight into population size and trends, but these indices have limitations [4]. Although the trapping rate significantly correlates with separate density estimates [5], their precise relation may vary per deployment [6], necessitating a deployment-specific calibration. Random Encounter Models (REMs) address this concern by describing the contact rate between animals and camera traps via two-dimensional ideal gas models [1]. This method enables the expression of wildlife density in terms of contact rate, animal speed, deployment duration, and the radius and angle of the detection zone. Thus, REMs provide a more generic abundance estimation, even applicable for species with indistinguishable appearances. All parameters described can be extracted directly from the camera trap data using in-situ measurements [1], trigonometric measurements with marks at known distances [7], or with automated distance estimation methods based on monocular depth and robust camera calibration workflows [8].

Modern camera traps often allow capturing a sequence of images per contact along with rapid re-triggering. These enable the estimation of animal speed required for REMs by tracking animal features along the sequence. Typically, the contact point of the animal with the ground closest to the camera when entering the camera's field of view (FOV) is chosen as such a feature and labeled respectively. By projecting the pixel coordinates of a label onto a specially fitted model of the ground, 3D coordinates—for such use cases, oftentimes expressed as distance and angle on a plane—can be

4th International Workshop on Camera Traps, AI, and Ecology, September 5 – 6, 2024, Hagenberg, Austria
leopold.boess@fh-hagenberg.at (L. Böss); wohlfahrt@wildoekologie.at (S. Wohlfahrt); david.schedl@fh-hagenberg.at (D. C. Schedl)
0009-0001-5177-0342 (L. Böss); 0009-0008-0508-8781 (S. Wohlfahrt); 0000-0002-7621-3526 (D. C. Schedl)

derived, enabling the computation of the speed of animals across the sequence. Naturally, the quality of the resulting estimate depends on the accuracy of the employed ground model.

Fitting a ground model requires a manual calibration routine per deployment location plus one camera calibration for the operating camera type and model. One such calibration routine involves placing a standardized pole (typically 1 m tall and marked at known intervals of 10 cm to 20 cm) perpendicular to the ground within the FOV and recording images. Afterward, specified pole features, such as both ends, are annotated in each image. These annotations, along with their actual distances and the camera intrinsics (i.e., sensor resolution, FOV, and optical center), allow fitting a ground model. For instance, the R-package CTtracking[1] fits a flat or unidirectional (i.e., bending on one axis) planar ground model per deployed camera trap to predict animal positions and speed. The quality of the estimated model depends on the expressiveness of the pole positions, i.e., how well they represent the respective environment.

This laborious manual deployment calibration restricts the use of camera traps to locations that are accessible by humans and need to be repeated for each deployment session. Also, more complex terrains, such as those with ridges and slopes, require extensive calibration to map adequately. Simple ground models might miss such terrain features, leading to inaccuracies in the 3D coordinates. Therefore, we propose utilizing precise digital elevation models (DEMs)—typically recorded by airborne laser scanning—instead of simpler manually fitted models.

Although this approach appears straightforward, there is a deficit of studies investigating the utilization of DEMs for this specific use case. Notably unexplored is the impact of their level of complexity on the quality of the distances they assign to pixel coordinates. While the resource intensity increases with model complexity, a coarser model can lead to substantial inaccuracies in the distance estimates (see Figure 1). Findings in this area could provide valuable insights for improving camera-trap-based analysis. Hence, this work compares DEMs varying in precision to address the question: *"How critical is a DEM's resolution and detail to the accuracy of projected pixel distance estimates?"*

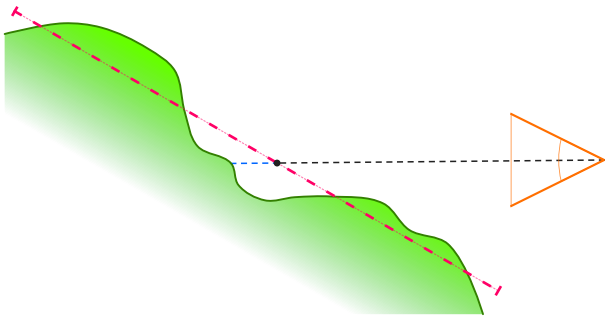


Figure 1: Illustration of potential errors when using low-resolution DEMs to approximate the actual elevation, such as a simple plane. The more non-planar the actual environment (green), the greater the possible error (blue) when projecting 2D coordinates from the camera (orange) onto a crude DEM (hot pink). The orange triangle indicates the camera's FOV.

2. Methods

Most of the methods required and applied in the presented approach represent problems in computer graphics and computer vision. Fortunately, as such, these problems are likely to have already been solved in the form of algorithms or even software libraries.

Fundamental to all camera-trap-related data processing is determining its intrinsic metrics. For this purpose, an OpenCV camera model is computed using substantial captures featuring a calibration checkerboard pattern and the OpenCV library [9]. The resulting model is used to correct camera-specific distortions or imperfections. This step is crucial for accurate 3D reconstructions, reducing the parameters needed for projection to a principal point and the FOV.

[1] https://github.com/MarcusRowcliffe/CTtracking

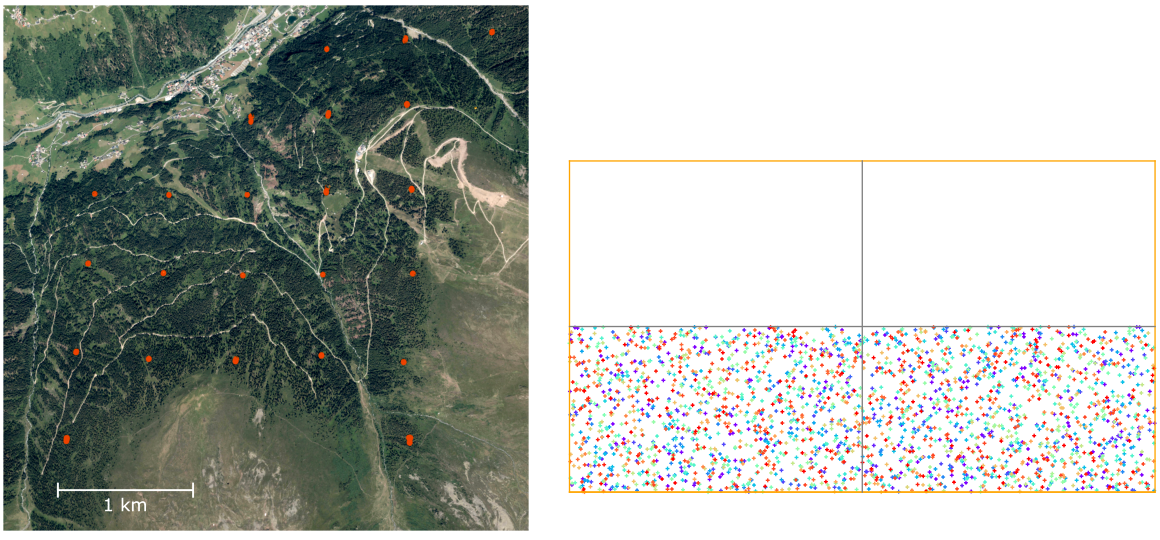
The more variable input required consists of a set of camera shots, each labeled to mark a feature of the triggering target. Further, the orientation data encompassing the camera's location in GPS coordinates and its rotation in the given deployment are necessary to fit a DEM. Although modern camera traps may include features that automatically determine orientation parameters, these can also be assessed via external tools or by following guidelines, such as consistently facing the camera north.

Before projection, the 2D coordinates undergo undistortion employing the camera calibration and OpenCV. Subsequently, the resulting undistorted coordinates are used to define rays from the camera origin through the corresponding pixel. These rays can then be cast onto the geometry expressed by the DEM, with their intersection point representing the 3D location depicted by the respective pixel.

The projected 3D coordinates allow for further analysis, e.g. speed computations as required for REM computation. The following section goes into experimentally verifying the described method.

3. Evaluation

The camera-trap dataset, which forms the basis for our experiments, was collected in the field using model BTC-PATRIOT-FHD (Browning International S.A., 84050 Morgan, Utah) camera traps, with a resolution of 2688×1520 (4 MP). The study area, located in See, Austria, is positioned geospatially between the latitudinal values of 47.053 920 41° N to 47.081 984 77° N and longitudinally between 10.450 032 78° E to 10.488 935 68° E as illustrated in Figure 2a. The dataset comprises 2629 labeled instances of the three resident ungulate species red deer (*Cervus elaphus*), roe deer (*Capreolus capreolus*) and chamois (*Rupicapra rupicapra*) and was gathered throughout 23 deployments, each lasting approximately 150 to 160 days. Only two periods, lasting 31 days in September–October 2023 and February 2024 respectively, were annotated, which form the baseline for subsequent experiments. Although 2D annotations exist, they could not be directly employed in our experiments due to missing information about the internal data handling of the applied labeling tool (i.e., Agouti[2]). Therefore, labels are simulated by randomly selecting 2D coordinates concentrated in the lower half of the image sensor (cf. Figure 2b). Due to camera trap alignment, the upper sensor half would lead to ray casts that do not intersect with the ground. Although the 2D image coordinates are randomly sampled, the number of labels per camera deployment corresponds to the in-field data, as shown in Table 1.



(a) Orthophoto of the study area.

(b) Labeled instances.

Figure 2: The deployment locations in See, Austria, are illustrated as orange dots on an orthophoto (a) with the respective label distribution on the camera-trap sensor for 23 deployments (b). Labels of the same color belong to the same deployment.

[2]https://agouti.eu/

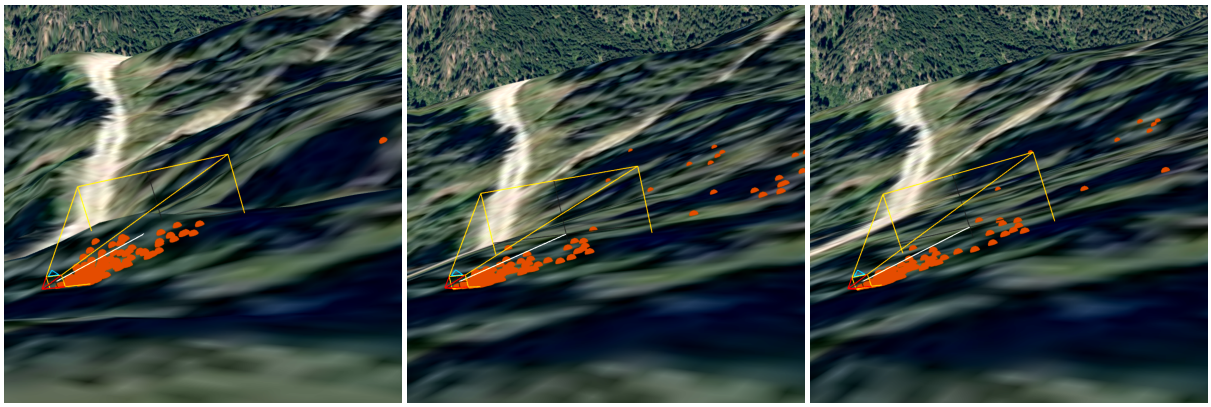
Table 1
Coordinates, altitude in meters (measured using the 3 m DEM), and number of labeled instances for the deployed cameras in the field.

WGS84 Ellipsoid Coordinates (EPSG:4326)	Altitude	Labels
47.060 185 16° N, 10.474 280 73° E	1834.10	205
47.076 552 61° N, 10.480 927 83° E	1833.99	89
47.070 996 68° N, 10.481 618 13° E	1930.61	117
47.064 879 84° N, 10.458 557 96° E	1693.33	81
47.065 223 27° N, 10.450 877 91° E	1627.59	91
47.081 588 14° N, 10.473 139 57° E	1340.92	155
47.070 199 19° N, 10.450 032 78° E	1406.78	124
47.070 627 03° N, 10.465 944 09° E	1503.02	47
47.076 237 39° N, 10.465 512 22° E	1264.99	23
47.059 371 90° N, 10.450 792 05° E	1738.15	226
47.076 098 57° N, 10.473 631 95° E	1624.18	43
47.060 179 20° N, 10.481 831 66° E	1919.38	192
47.065 750 80° N, 10.482 103 71° E	1957.65	109
47.059 459 84° N, 10.458 459 23° E	1926.54	65
47.081 984 77° N, 10.488 935 68° E	1715.79	99
47.070 332 51° N, 10.457 939 26° E	1472.01	58
47.053 920 41° N, 10.451 154 04° E	1889.37	146
47.065 403 13° N, 10.474 012 68° E	1678.85	100
47.055 142 51° N, 10.482 799 78° E	1917.85	51
47.065 029 98° N, 10.466 381 50° E	1765.97	119
47.059 691 51° N, 10.466 487 28° E	1972.68	274
47.081 822 98° N, 10.480 965 80° E	1524.77	93
47.070 845 55° N, 10.473 881 25° E	1686.51	122
Average	1710.32	109.5

The evaluation compares DEMs with resolutions of 3 m, 20 m, and 50 m (c.f., Figure 3), regarding the most precise DEM as ground truth. The federal government of Tirol offers the required elevation data[3], originally recorded with a step size of 1 m via airborne laser scanning in 2018. This provided data consists of height texture tiles in the MGI Austria GK West coordinate system (`EPSG:31254`) and is processed (i.e., merged and transformed) with the cartography software QGIS[4]. The coordinate reference system (CRS) `WGS84/UTM33N` (`EPSG:32633`) was chosen as the foundation for measuring distances and errors in meters. Given ThreeJS requires a 3D mesh for raycasting, the elevation data is lastly converted into a polygonal mesh using the Python libraries pyproj[5] and Rasterio[6].

Each camera deployment is given geospatial coordinates, which can be associated with 2D positions on the DEM. However, the precise height of the mounted cameras in situ is not measured. Therefore, these parameters are estimated using the DEMs and a relative height offset of 0.5 m. Similarly, the camera rotation had to be inferred based on the assumption that the cameras were generally installed parallel to the ground and facing north to avoid direct sunlight. Therefore, first, the forward orientation of the camera was defined by moving 5 m north on the DEM while retaining the relative height offset. Secondly, the camera roll (defining the upward direction) is estimated by sampling two elevations 3 m east and west of the camera's deployment position. Placement and orientation computations have been carried out individually for each DEM resolution, resulting in slightly changed parameters per DEM and deployment, as illustrated in Figure 3. Compared to the 3 m DEM, the variation in simulated mounting altitude was between 0.0 m to 3.27 m and on average 0.54 m and 1.14 m (cf. Table 2) for the lower resolution DEMs.

[3]https://www.tirol.gv.at/als
[4]https://www.qgis.org/
[5]https://github.com/pyproj4/pyproj
[6]https://github.com/rasterio/rasterio



(a) 3 m DEM (ground truth). (b) 20 m DEM. (c) 50 m DEM.

Figure 3: The impact of using DEMs of different precision—3 m to 50 m—as deployment surface model. The camera's view frustum is drawn as a 10 m high cone, indicating the placement and orientation of the camera. The 3D locations of labeled instances (i.e., animal sightings) are shown as orange dots and vary depending on DEM resolution. Note that fine details, such as the road and ridges in the background, are lost with low DEM resolution (c).

Based on the simulated camera mounting, the 2629 labeled instances, consisting of 2D coordinates in images, are assigned to the corresponding deployments and projected onto the DEM to compute 3D coordinates in the CRS of the DEM. In our experiments, ray-casting is performed using the ThreeJS JavaScript framework[7] utilizing a bounding-volume-hierarchy implementation[8] ensuring accuracy and computational efficiency. After projection, the labels are, on average, ~7 m away from the mounted cameras. The high-precision DEM has the highest label-to-deployment distance (cf. Table 2). Furthermore, the change in a label's 3D coordinates with a change in DEM accuracy is measured. Table 2 illustrates the impact on distances (between deployment and labels) and the 3D coordinate alone. All values indicate absolute mean values and are computed considering the altitude (3D) and disregarding altitude (2D). Decreasing the DEMs resolution to 20 m and 50 m skews the average distances by an average of 2.84 m and 3.94 m respectively, with the individual errors dispersing widely, as indicated by the standard deviation.

Note that due to the recalculation of the camera's mounting position for each DEM and the differences in the terrain data, not all 2629 labels could be projected to 3D locations in all situations (cf. Table 2). Furthermore, projected 3D locations are clipped if they exceed a distance of 100 m from the deployment camera. Missing 3D coordinates are omitted from the calculations.

The error values for our models indicate a logarithmic behavior as model resolution changes, i.e., changing from a 3 m to 20 m DEM introduces considerably more errors than the resolution decrease from 20 m to 50 m. Thus, as the detail level shrinks and resolution drops, the influence of the level of detail on the divergence decreases, resulting in less pronounced errors between models. However, this behavior is likely situational and might break once a model has reached a certain coarseness, at which critical landscape features are lost due to aliasing.

4. Conclusion

This paper takes the first steps toward estimating animal speeds required for REMs based on labeled instances and digital elevation models captured via airborne laser scanning. Converting these DEMs to 3D geometry allows instance distance computation by projecting their 2D coordinates onto this geometry. The described method presents a promising alternative to current methods by being less restrictive and reducing deployment calibrations to measure the location and rotation of the camera

[7] https://threejs.org/
[8] https://github.com/gkjohnson/three-mesh-bvh

Table 2
Results of varying DEM resolutions on deployment and labeled instance accuracy. All error calculations consider the 3 m elevation model as the ground truth. All distance and error values represent absolute averages in meters and include the standard deviation in parentheses.

DEM	3 m	20 m	50 m
Projected Labels	2422	2461	2445
Error in Label Positions			
3D (Std)	—	3.10 (7.54)	4.48 (9.87)
2D (Std)	—	2.73 (7.30)	3.77 (9.80)
Error Deployment Altitude (Std)	—	0.54 (0.61)	1.14 (0.93)
Labels to Deployment Distances			
3D (Std)	7.11 (11.66)	6.93 (10.04)	7.06 (11.20)
2D (Std)	6.68 (11.14)	6.52 (9.50)	6.67 (10.76)
Error in Distances			
3D (Std)	—	2.84 (8.11)	3.94 (10.77)
2D (Std)	—	2.72 (7.79)	3.77 (10.49)

trap. However, said method heavily relies on this calibration and the accuracy of the employed DEMs. This work mainly assesses the impact of a DEM's level of precision utilizing realistic data from in-situ deployed camera traps. The carried-out evaluation demonstrates that lowering model resolution not only distorts results but also results in strongly dispersed deviations from a more densely sampled DEM.

Further, the evaluation processes revealed significant inaccuracies caused by estimating altitude via a height offset and a DEM. For this dataset, using the same altitude values across all DEMs causes many projection rays to miss the DEM entirely or cameras to fall below the surface for both lower-resolution models. This discrepancy mainly occurs as the deployment locations do not align with the pixel corners within the height texture data. Therefore, camera deployment parameters have been recalculated for every DEM. This outcome emphasizes the need for models that reduce the regional domain as much as possible while staying faithful and precise. Despite some obstacles, this approach shows potential and, with further development, could offer a more effective solution than alternative methods.

Since an animal's speed is calculated based on its projected distances over an image sequence, the actual impact of this error on subsequent calculations such as REM is unclear. For example, this error might attenuate if all distances along a sequence share a similar error. Thus, further work should extend the evaluation to include speed estimations and compute the consequent wildlife density, ideally comparing them to independent estimates.

Although the 3 m DEM has clear advantages over its low-res counterparts, an even higher-resolution model with even further focus, possibly on specific deployment scenarios, might be beneficial. Such a context restriction allows a finer resolution and, thus, a more detailed depiction of the local environment. The grid-based airborne laser-sampling method may inherently not be ideal as it is susceptible to aliasing-related issues, potentially resulting in the loss of crucial features. It could be beneficial to explore using dynamic resolution approaches.

Additionally, focus should be shifted towards applying the approach in-field to ensure it represents an improvement over current methods. This focus involves researching techniques to accurately measure camera trap locations and rotations to minimize the effort connected to manual deployment calibration. Furthermore, the concept of 3D reconstruction may be incorporated to generate detailed 3D geometry per deployment, replacing laser-sampled height data. Such reconstructions may offer an adequate model accuracy-simplicity balance and can be created based on a short video capture of the respective area.

Acknowledgments

This project is funded by the Austrian Research Promotion Agency FFG (project *BAMBI* and *Equalize2*; program numbers: 892231 and 53988223) for which the budget is provided by the Federal Republic of Austria. Image data collection was part of a project funded by 'Wildbach- und Lawinenverbauung GBL Oberes Inntal.'

References

[1] J. M. Rowcliffe, J. Field, S. T. Turvey, C. Carbone, Estimating animal density using camera traps without the need for individual recognition, Journal of Applied Ecology 45 (2008) 1228–1236.

[2] T. G. O'Brien, Abundance, Density and Relative Abundance: A Conceptual Framework, Springer Japan, Tokyo, 2011, pp. 71–96. doi:10.1007/978-4-431-99495-4_6.

[3] A. F. O'Connell, J. D. Nichols, U. K. Karanth (Eds.), Camera Traps in Animal Ecology, Springer Japan, Tokyo, 2011. doi:10.1007/978-4-431-99495-4.

[4] A. Kalandarishvili, M. Heltai, Camera traps as a research method for carnivore population estimation: Strength, weaknesses, opportunities and threats, analysis and improvements, Columella 10 (2023) 13–24. doi:10.18380/szie.colum.2023.10.2.13.

[5] T. G. O'Brien, M. F. Kinnaird, H. T. Wibisono, Crouching tigers, hidden prey: Sumatran tiger and prey populations in a tropical forest landscape, Animal Conservation 6 (2003) 131–139. doi:10.1017/S1367943003003172.

[6] D. I. MacKenzie, J. D. Nichols, G. B. Lachman, S. Droege, J. Andrew Royle, C. A. Langtimm, Estimating site occupancy rates when detection probabilities are less than one, Ecology 83 (2002) 2248–2255. doi:10.1890/0012-9658(2002)083[2248:ESORWD]2.0.CO;2.

[7] S. E. Pfeffer, R. Spitzer, A. M. Allen, T. R. Hofmeester, G. Ericsson, F. Widemo, N. J. Singh, J. P. G. M. Cromsigt, Pictures or pellets? Comparing camera trapping and dung counts as methods for estimating population densities of ungulates, Remote Sensing in Ecology and Conservation 4 (2018) 173–183. doi:10.1002/rse2.67.

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

[9] G. Bradski, The OpenCV Library, Dr. Dobb's Journal of Software Tools (2000).