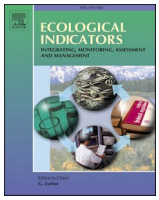
Ecological Indicators 182 (2026) 114593

Contents lists available at ScienceDirect

Ecological Indicators

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



Scalable computer vision-based assessment of bait lamina sticks to quantify soil fauna activity

Adrija Roy [a,*], Lukas Thielemann [a], Masahiro Ryo [a,b], Juan Camilo Rivera-Palacio [a,b], Konlavach Mengsuwan [a,b], Kathrin Grahmann [a]

[a] Leibniz Centre for Agricultural Landscape Research (ZALF), Müncheberg, Germany
[b] Brandenburg University of Technology Cottbus-Senftenberg, Cottbus, Germany

ARTICLE INFO

Keywords:
Feeding activity
Image analysis
Automated assessment
Biological indicators
Ecosystem functions

ABSTRACT

Soil fauna plays a critical role in ecosystem functions such as nutrient cycling, organic matter decomposition, and soil structure maintenance. Accurately assessing their activity is therefore essential for monitoring soil health. Traditional methods like the bait lamina test, while widely used, rely on manual visual scoring, which can be subjective, time-consuming, and difficult to scale. In this study, we present an automated computer vision approach to quantify soil fauna activity by assessing bait consumption on bait lamina sticks, using high-resolution imagery processed with a Python-based pipeline. We implemented this approach on 159 bait sticks gathered from field plots in Brandenburg, Germany, and compared the automated findings with assessments from five independent human operators. The automated method displayed a strong agreement with manual evaluations, yielding Pearson's r between 0.80 and 0.92, depending on the operator, and Cohen's kappa of 0.48 in categorical concordance. The Bland-Altman analysis revealed that over 90 % of the automated scores were within +/− 0.2 of the manual measurements. This automated technique reduced the time required for analysis in comparison to manual scoring, along with removing operator subjectivity and bias. Although there was an underestimation in identifying fully consumed bait holes, the average difference between the automated and manual scores was only 0.02 ($p = 0.0049$), suggesting a negligible effect size. The automated approach is straight-forward, reproducible, and flexible, which facilitates the efficient and impartial evaluation of soil fauna activity for large-scale soil health monitoring. Possible improvements could involve enhancing the image-analysis workflow, such as improving hole-detection robustness, reducing sensitivity to coating or lighting variation, and exploring more advanced classification models.

1. Introduction

Evaluating soil health is essential for sustainable agriculture and ecosystem management, as soil biological processes directly influence nutrient cycling, organic matter decomposition, and plant productivity (van der Heijden and Wagg, 2013). Among many indicators of soil health, the activity of soil mesofauna, such as collembolans, mites, and enchytraeids or soil macrofauna, especially earthworms, serves as a sensitive proxy, reflecting both the current state and resilience of the soil ecosystem (Ritz et al., 2009; Bardgett and van der Putten, 2014). Meso- and macrofauna are principle drivers of litter breakdown and microbial interactions, and their activity has been shown to respond rapidly to changes in land management, pollution, and climate conditions (Brussaard et al., 2007). Quantifying how biological activity affects soil structural parameters and biochemical processes is key to assessing soil health (Franciska T. de Vries et al., 2013).

Von Törne Von (1990) introduced the bait lamina test (BLT) as a rapid and straightforward method for the visual assessment of plant debris consumption by soil organisms. Detailed descriptions of the BLT can be found in other sources (ISO, 2016). In essence, the method involves filling holes in PVC strips with bait material made from cellulose and plant material (such as wheat bran or nettle leaf powder) and then inserting the strips in soil for a certain period, depending on the feeding activity. After exposure, an operator visually counts the number of pierced holes, providing a measure of feeding activity of soil fauna. Due to its simplicity, the BLT has been widely utilized for soil health assessment, particularly to evaluate the impacts of land use and management changes in agroecosystems (Larink and Sommer, 2002; Förster

* Corresponding author.
E-mail addresses: adrijar94@iitb.ac.in, adrija.roy@zalf.de (A. Roy).

https://doi.org/10.1016/j.ecolind.2025.114593
Received 19 August 2025; Received in revised form 27 December 2025; Accepted 28 December 2025

et al., 2015; Römbke et al., 2017), including pesticide or fertilizer application in agricultural systems (Birkhofer et al., 2022). Beyond agroecosystems, the BLT has also been applied in other fields, such as soil pollution (Filzek et al., 2004; André et al., 2009; Vorobeichik and Bergman, 2020), ecotoxicological testing (Jänsch et al., 2006; Bart et al., 2018) and assessment of ionizing radiation (Beresford et al., 2022), fire impacts (Musso et al., 2014; Podgaiski et al., 2014), forest fragmentation (Simpson et al., 2012), urbanization (Bergman et al., 2017), and plant invasion (Pehle and Schirmel, 2015).

Although the traditional approach of visually assessing the BLT has been widely applied and accepted as a standard practice to quantify soil fauna activity (ISO, 2016), it comes with several limitations. The most critical issue is operator variability: different assessors may interpret bait consumption inconsistently, which introduces subjectivity and affects comparability across studies (Eisenhauer et al., 2014). A second challenge is the coarse scoring scales commonly used in manual assessment, typically two classes (any or no consumption) or three classes (total, medium or no consumption), which limits resolution and reduces the accuracy of derived activity measures (Vorobeichik and Bergman, 2021). Finally, manual scoring is relatively more time-consuming and labor-intensive, as each stick must be inspected hole by hole. Different existing manual scoring systems thresholds are shown in Table 1.

To address the limitations of manual bait lamina evaluation, this study aims to develop and validate an automated, impartial, and scalable method to assess bait lamina consumption using computer vision. We present a reproducible Python-based image analysis pipeline that detects and quantifies feeding activity on bait lamina sticks from high-resolution images.

The specific objectives of this study are to:

(1) develop a computer vision workflow to automatically assess bait lamina feeding activity;
(2) evaluate the accuracy and reliability of the automated scoring method compared to visual assessments; and.
(3) assess the potential of this method for large-scale, high-throughput soil biological activity monitoring.

We hypothesize that the automated approach will yield results comparable to manual scoring while reducing subjectivity and improving scalability. This tool is intended to support standardized, reproducible soil fauna monitoring in both ecological and agricultural research.

Table 1
Comparison of bait lamina scoring systems with examples from the literature.

Scoring	Manual Classification			
	2-point		3-point	5-point
	(Törne Von, 1990*)*	*(ISO 18311, 2016)*	*((*Thakur et al., 2018*;* Siebert et al., 2019*) Manual scoring in this study*	*((*Bergman et al., 2017*;* Vorobeichik and Bergman, 2021, 2023*)*
0	Bait is not perforated	Consumption of the bait by less than half	Filled	Bait untouched
0.25				~ 25 % hole area is empty
0.5			Partly empty	~ 50 % hole area is empty
0.75				~ 75 % hole area is empty
1	Bait is perforated to any extent	Consumption of the bait by at least half	Empty	No bait left

2. Materials and methods

2.1. Site description and field nstallation

As the scope of this study focuses on the verification of a new evaluation method, the experimental site is only described briefly: The data collection took place in 2024 in the patchCROP landscape laboratory. The experimental site is located in Brandenburg (52.4426°N, 14.1607°E) and characterized by heterogeneous, sandy soils due to historic glaciation events, displaying cambisol, luvisol, and truncated luvisol soil types (Hernández-Ochoa et al., 2025). The long-term annual mean temperature from 1980 to 2010 was 9.2 °C, while the average annual rainfall was 568 mm, ranging from 373 to 774 mm. For context on field conditions during sampling, refer to Appendix A and Fig. A1 in Appendix, which shows the daily precipitation, temperature, and volumetric soil moisture in two selected plots.

The bait lamina sticks (Terra Protecta GmbH, Berlin, Germany) used in this study measure 120 mm × 6 mm × 1 mm and contain 16 circular perforations (diameter 1.5 mm) spaced at 5 mm intervals, starting at 5 mm from the insertion tip. The sticks had either gray or white color, varying with purchased batches. The perforations are filled with a standardized bait mixture, that was prepared using 70 % cellulose powder, 27 % finely ground and sieved wheat bran and 3 % charcoal and water to produce a paste-like consistency. To ensure a complete filling in each of the 16 perforations, 4–6 rounds of drying and refilling were conducted to close drying gaps and cracks.

Bait lamina sticks were placed in summer 2024 in plots of grain maize (Fig. A2) that captured a gradient in both soil texture and management practices (Table A1 in Appendix). Specifically, plots ranged in sand content (63–67 % to 80–83 %; sandy loams to loamy sands) and weed control strategies (chemical vs. mechanical weed control). This gradient allowed us to evaluate whether the automated classification and detection algorithms remained accurate and consistent despite differences in soil background or bait stick color, which can affect image analysis. Activity was checked regularly using test sticks and all sticks were removed after 21 days and individually wrapped in aluminum foil for transport. The exposure period lies in between the ones reported for other studies on similar soils in Brandenburg of 14 days on grassland (Birkhofer et al., 2022) and 28 days on arable soils (Joschko et al., 2008; Birkhofer et al., 2022). After field retrieval, the bait sticks were often covered with adhered soil particles. In the laboratory, the sticks were cleaned using moistened paper towels to remove soil particles that might obscure bait consumption. Any residual soil within the perforations was gently dislodged using needles and brushes to prepare them for the manual and automated evaluation. This cleaning step was essential to prevent false pixel detection in the automated algorithm, as soil specks can mimic bait consumption.

Five operators (M1–M5) independently evaluated all bait lamina strips from the same set of images. Operators M1 and M5 were authors of this paper, and M2, M3 and M4 were graduate students assisting in data evaluation. None of the operators, except M5, had prior experience with the bait lamina method or formal training in soil ecology. This composition was chosen to represent typical variability among non-expert users performing routine assessments. Operators visually inspected each perforation and assigned one of the 3 categories: full activity, partial activity, and no activity with activities assigned as 100 %, 50 % and 0 %.

2.2. Image acquisition and preprocessing

To digitize the bait lamina sticks for automated analysis, we developed a controlled imaging setup using a mounted Android smartphone camera positioned perpendicular above a grid-lined A4 background (Appendix Fig. A3). Images were captured using a smartphone (Android 13, 16 MP rear camera, f/1.8 aperture). All AI-based enhancements (HDR, scene optimizer, auto-beautification) were disabled to avoid pixel

interpolation or sharpening artifacts. The sticks were placed in a fixed 8-stick arrangement per image, with each stick spaced equally. This arrangement ensured consistent orientation across samples and minimized parallax distortions. The camera was positioned 21 cm above the surface, and images were captured under diffuse lighting provided by a custom light table, enhancing contrast across both gray and white bait stick variants (Appendix Fig. A4).

Every image was converted to grayscale via OpenCV's cvtColor, which implements the standard 0.299 R + 0.587 G + 0.114 B formula (Gonzalez and Woods, 2017) a process that reduces a color image to shades of gray based on luminance, thereby simplifying analysis. Gaussian blurring was applied to minimize background noise by smoothing the image, using a Gaussian kernel to suppress high-frequency variations (Flusser et al., 2016; Bergstrom et al., 2023). The Hough Circle Transform, a feature extraction technique commonly used in computer vision for detecting circular shapes (Kierkegaard, 1992; Kerbyson and Atherton, 1995; Li and Wu, 2020), was then employed to identify perforations. Finally, the pixel intensity histograms inside each detected hole were analyzed to quantify grayscale variation, which can indicate levels of material removal or activity inside the perforations. Feeding intensity was then quantified for each detected circle by computing the percentage of white (background) pixels within its area.

2.3. Pipeline steps

The automation workflow (Fig. 1) consisted of two sequential scripts developed in Python using the OpenCV, NumPy, and PIL libraries. The pipeline is explained in subsequent subsections and is available in GitHub (https://github.com/Adrija1/baitstick-analysis).

2.3.1. Strip extraction and background classification

High-resolution JPEG images were cropped to isolate the sticks and then split into eight vertical strips, each representing one bait lamina stick. For each strip, the mean grayscale pixel intensity was calculated to classify background color. Strips with mean intensity above 190 were assigned as white background, and those at or below 190 as gray. This classification, based on the distribution of grayscale values in our imaging setup, ensured consistent separation of stick types and informed the parameter settings used in subsequent image processing steps, such as circle detection. This threshold may require adjustment under

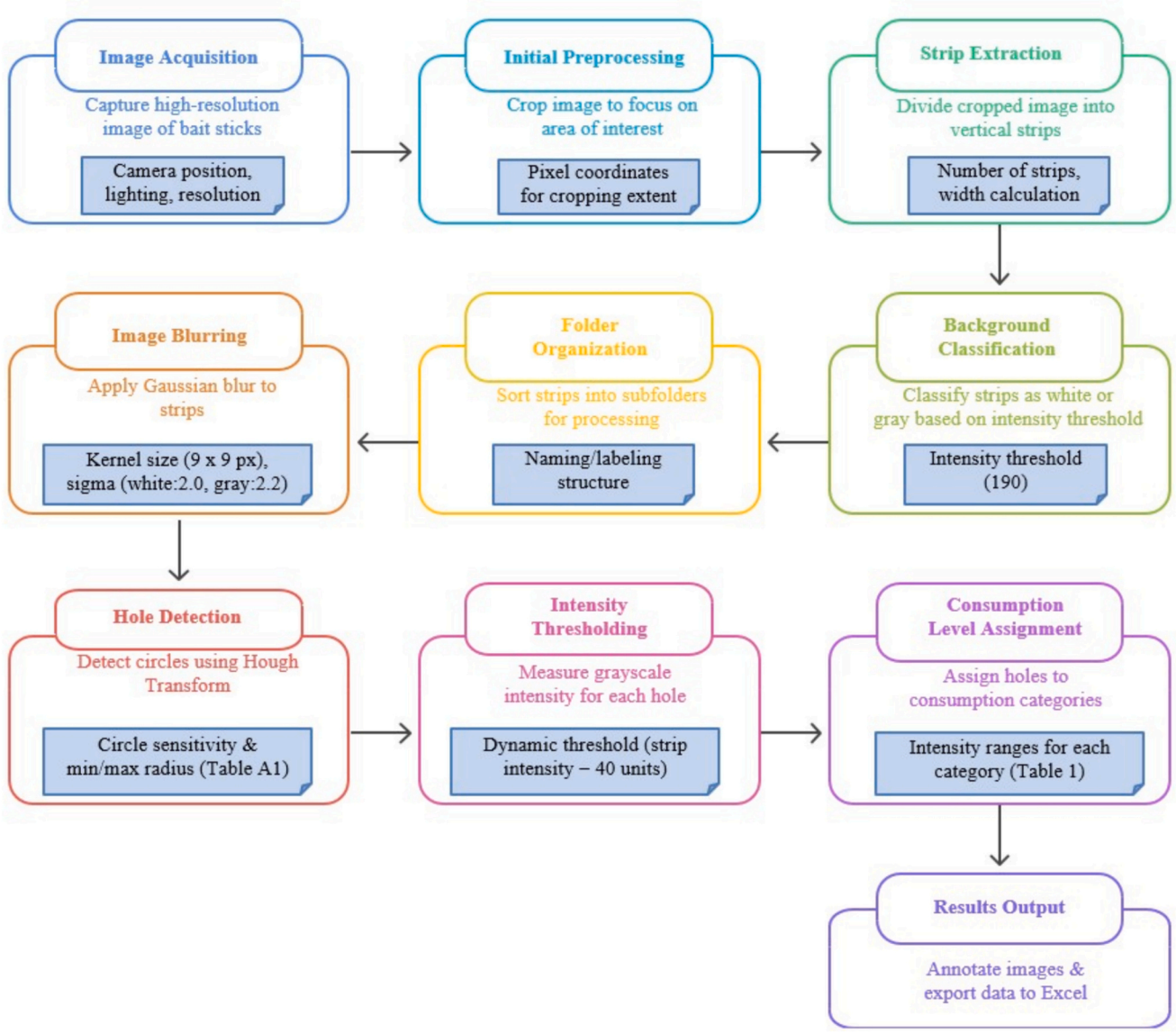


Fig. 1. Workflow diagram of the automated bait lamina sticks analysis pipeline. Each step indicates key parameters or decisions that can be fine-tuned, such as intensity thresholds, blur settings, and classification criteria.

different lighting conditions or camera settings. Background classification also determined the parameter settings for subsequent image processing: white and gray sticks required different levels of contrast enhancement and circle sensitivity during Hough Circle detection. The individual strip images were then automatically organized into structured subfolders, labeled according to the original image and segment number.

2.3.2. Hole detection via hough transform

Each strip image underwent analysis using a computer vision pipeline that initiated with adaptive Gaussian blurring, where the parameters for kernel size and blur intensity were tailored according to the strip's background tone (either white or gray) to improve contrast and reduce noise. We used a Gaussian filter (OpenCV's GaussianBlur) using a 9×9 px kernel. The standard deviation (σ) was set to 2.0. These values were chosen empirically to provide optimal noise reduction and edge enhancement in our dataset but may require fine-tuning for other image conditions or different bait stick colors. Next, the Hough Circle Transform was utilized to identify circular perforations that correspond to the bait-filled holes. Detection was performed with OpenCV's HoughCircles (HOUGH_GRADIENT), applying the parameters shown in Table A2 in the Appendix.

Each bait lamina hole is biconical, with an inner diameter of approximately 1.5 mm and an outer diameter of 2 mm. The segmentation algorithm detects the visible outer contour of each hole in the image, corresponding to the 2 mm outer diameter. Although the bait lamina design is nominally standardized, we observed deviations in both hole alignment and diameter across strips, likely due to manufacturing variability. Fixed-position or fixed-radius approaches were tested but led to frequent missed detections when holes were slightly displaced or noncircular. For this reason, the adaptive circle-detection step was used to locate the actual hole boundaries in each image, which improved robustness across heterogeneous strip batches and imaging angles.

2.3.3. Pixel-intensity thresholding & consumption-level assignment

For every identified circle, the pixel intensity in that area was examined to determine the proportion of white (unconsumed bait) pixels. To binarize holes against the background, we applied a threshold equal to the mean gray-level of each strip minus 40 intensity units. In practice, this dynamic threshold typically equates to approximately 220 for white-stick images and about 170 for gray-stick images. The value of 40 units was selected to maximize the separation between bait and background pixels based on visual inspection of image histograms. Nevertheless, these values should be regarded as starting points and may need to be adjusted for optimal performance in different imaging environments. Annotated images were saved with overlaid circles and labeled percentage of feeding activity.

2.4. Activity assessment

The percentage of whitened pixels within the circular mask represents the proportion of bait removed and yields a continuous eaten fraction between 0 and 1. Strip-level activity was calculated as the average of these per-hole fractions. This value is used as the primary automated indicator of feeding activity.

For scoring the activity fraction from manual method, the simplified common three-point classification system was used, with medians assigned as follows: 0 % for No activity, 50 % for Partial, 100 % for Full activity.

The activity for a strip was then computed as:

$$Activity_{frac} = \frac{\sum_{i=1}^{3} n_i * m_i}{N}$$

where n_i is the number of holes in class i, m_i is the median activity fraction of that class, and N is 16, number of holes in the strip.

2.5. Validation metrics

All statistical analyses were conducted in Python 3.10 using NumPy 1.23 and SciPy 1.9, with auxiliary routines from scikit-learn 1.2 and statsmodels 0.14. A two-sided significance threshold of $\alpha = 0.05$ was applied throughout. To control the family-wise error rate across multiple Pearson correlations and paired *t*-tests, *p*-values were adjusted using the Bonferroni method. Sample sizes are reported alongside each test. A summary of applied tests and tools is provided in Table 2.

In addition to these metrics, we evaluated the agreement between the automated and manual assessments using a set of complementary validation tests. The automated continuous feeding-activity fraction was compared against each operator's assessment and against the manual consensus (mean of the operators). Agreement was quantified using Pearson correlation, mean absolute error (MAE), and root mean square error (RMSE), Deming regression, and Bland-Altman analysis. For completeness, we also derived categorical classes (no, partial, full feeding) from the continuous per-hole percentages using 5 % and 95 % as thresholds and computed the corresponding confusion matrix to compare automated and manual classifications. Practical equivalence was tested using a two-one-sided (TOST) procedure with a tolerance of $\pm 1/16$ of a hole (≈6 %). Intra-operators reproducibility was assessed separately using intraclass correlation coefficients ICC(2,1) and ICC (3,1).

2.6. Computational reproducibility assessment

To evaluate the numerical stability of the hole-based continuous activity metric, we analyzed a subset of 25 representative strip images. To mimic small but realistic changes in image acquisition, we generated a perturbed version of each strip by combining a 0.5° rotation, a 1-pixel translation, and a 5 % global increase in brightness. The same pipeline

Table 2
Summary of statistical tests, their purposes, software implementations, and sample sizes.

Test	Purpose	Software / Library	Sample size (n)
Mean Absolute Error (MAE)	Quantify average absolute deviation between methods	Python 3.10, scikit-learn	159 strips
Root Mean Square Error (RMSE)	Quantify error magnitude with higher penalty on larger deviations	Python 3.10, scikit-learn (mean_squared_error)	
ICC (2,1)	Assess reproducibility (absolute agreement)	Python 3.10 (ICC routine; two-way random effects)	159 strips
ICC (3,1)	Assess reliability (consistency)	Python 3.10 (ICC routine; two-way mixed effects)	159 strips
Pearson's r	Assess linear agreement in activity fractions	Python 3.10, SciPy 1.9 (pearsonr)	159 strips
Cohen's κ	Quantify categorical agreement (No/ Partial/Full feeding)	Python 3.10, scikit-learn 1.2 (cohen_kappa_score)	2528 hole observations
Bland-Altman analysis	Evaluate bias and 95 % limits of agreement	Python 3.10, statsmodels 0.14	159 strips
Deming regression	Model systematic bias between methods	Python 3.10, statsmodels 0.14	159 strips
Paired t-test	Test for mean differences in continuous and categorical scores	Python 3.10, SciPy 1.9 (ttest_rel)	159 paired strips

was rerun on these perturbed images, and agreement between original and perturbed activity scores was quantified using Pearson correlation, Bland-Altman analysis, and a two-way mixed-effects intraclass correlation ICC(3,1).

2.7. Processing workflow and time accounting

To quantify the total time requirement for both approaches, each stage from strip cleaning to final data export was recorded. For both workflows, cleaning soil residues from each bait stick required approximately 40 s. The manual method then involved around 30 s per stick for visual assessment under a lamp and 10 s for recording the scores in a spreadsheet.

For the automated workflow, the cleaned sticks were arranged on an A4 sheet (8 per frame; ≈ 40 s per photo), the frame was positioned under the camera (≈ 10 s), and the image was captured and manually checked for brightness and visibility (≈ 20 s per photo). Batch processing and result export through the Python pipeline required ≈ 10 min of unattended computation. Overall, the total operator time for evaluation of ~160 strips was reduced from about 3.5 h in the manual workflow to approximately 2 h in the automated one.

3. Results

We performed a stepwise validation process that included both human evaluators (operators) and the automated method to quantify precision and consistency of the automated bait lamina analysis pipeline.

3.1. Inter-operator agreement

Five operators (M1, M2, M3, M4 and M5) evaluated bait lamina strips independently using a three-point classification system: no activity, partial activity, and full activity. The operators exhibited strong agreement in their classification of bait lamina feeding activity but visual comparison of activity fractions across plots revealed observable inter-operator variability (Fig. 2). In some plots, such as Plot 3 and Plot 10, the interquartile ranges differ notably among operators, with M2 frequently showing larger variance. Subjectivity in interpreting bait removal likely arises from differences in lighting conditions, bait color contrast, or perceptual thresholds for classifying full vs. partial feeding. Discrepancies were particularly evident in strips with high feeding activity, where differentiating between near-complete and fully consumed bait is inherently ambiguous. Such inconsistencies attest the risk for bias in manual scoring and reinforce the need for standardized protocols or automated methods to improve consistency and reliability. Pairwise Pearson correlation analysis indicated strong agreement, except for M2 (r ranging between 0.84 and 0.86) (Fig. 3f). The inter-operator consistency across the operators was high, with ICC(2,1) = 0.88 and ICC(3,1) = 0.90.

3.2. Agreement between manual and automated methods

The automated pipeline was validated by comparing its output to operator assessments. First, we compared the strip-wise average activity fractions derived from both, the manual method per operator and the automated pipeline. Across 159 bait lamina strips, the automated activity fraction showed strong linear agreement with the operators (Fig. 3f). The correlations were highest for M3 ($r = 0.92$), followed by M4 and M5 ($r = 0.91$). The automated estimate also aligned well with the manual consensus (r = 0.92). These relationships were reflected in the corresponding error metrics. Overall, the automated activity fraction closely tracked manual scoring patterns across strips, with strongest agreement for operators whose scoring distributions most closely matched the consensus. One operator showed a systematic tendency toward higher activity, which contributed to increased deviation from the automated estimates. Full agreement statistics are provided in Table 3.

To allow direct comparison with the standard bait lamina scoring system, the continuous per-home percentages were converted into 3 categorical counts (no feeding, partial feeding, full feeding) by applying fixed thresholds. Holes with ≤5 % white pixels were assigned as *no activity*, holes with ≥95 % white pixels as *full activity*, and all intermediate values (>5 % and < 95 %) as *partial activity*. These categories are only used for agreement testing with the manual method and are not involved in computing the automated activity fraction. Fig. A5 in Appendix shows correlation matrices for counts of different classes between operators and automated methods. The No-activity class shows the highest consistency among all operators (r ≈ 0.94–0.99), reflecting its relatively unambiguous visual signature. Full-activity counts also show high agreement across most operators (r ≈ 0.82–0.94), with the automated method correlating moderately to strongly with operators. As holes in No and Full classes are visually unambiguous, their counts vary consistently between operators, resulting in higher correlations. In

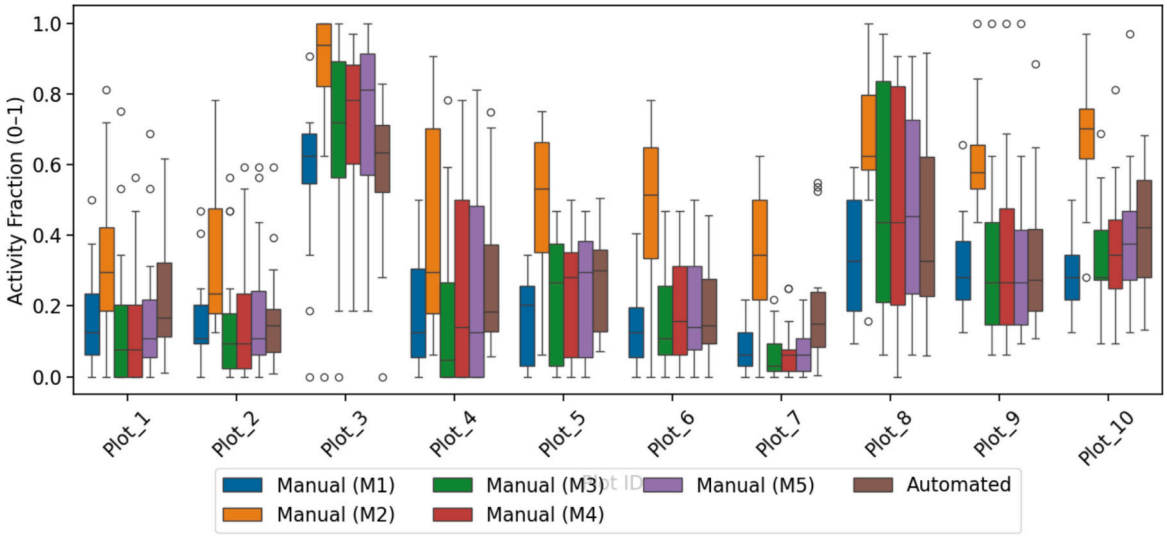


Fig. 2. Boxplots comparing distributions of bait stick activities ($n = 16$) per plot assessed by five different operators and the automated pipeline. Boxplots show the median (center line), interquartile range (box), whiskers extending to 1.5 × IQR, and outliers (points).

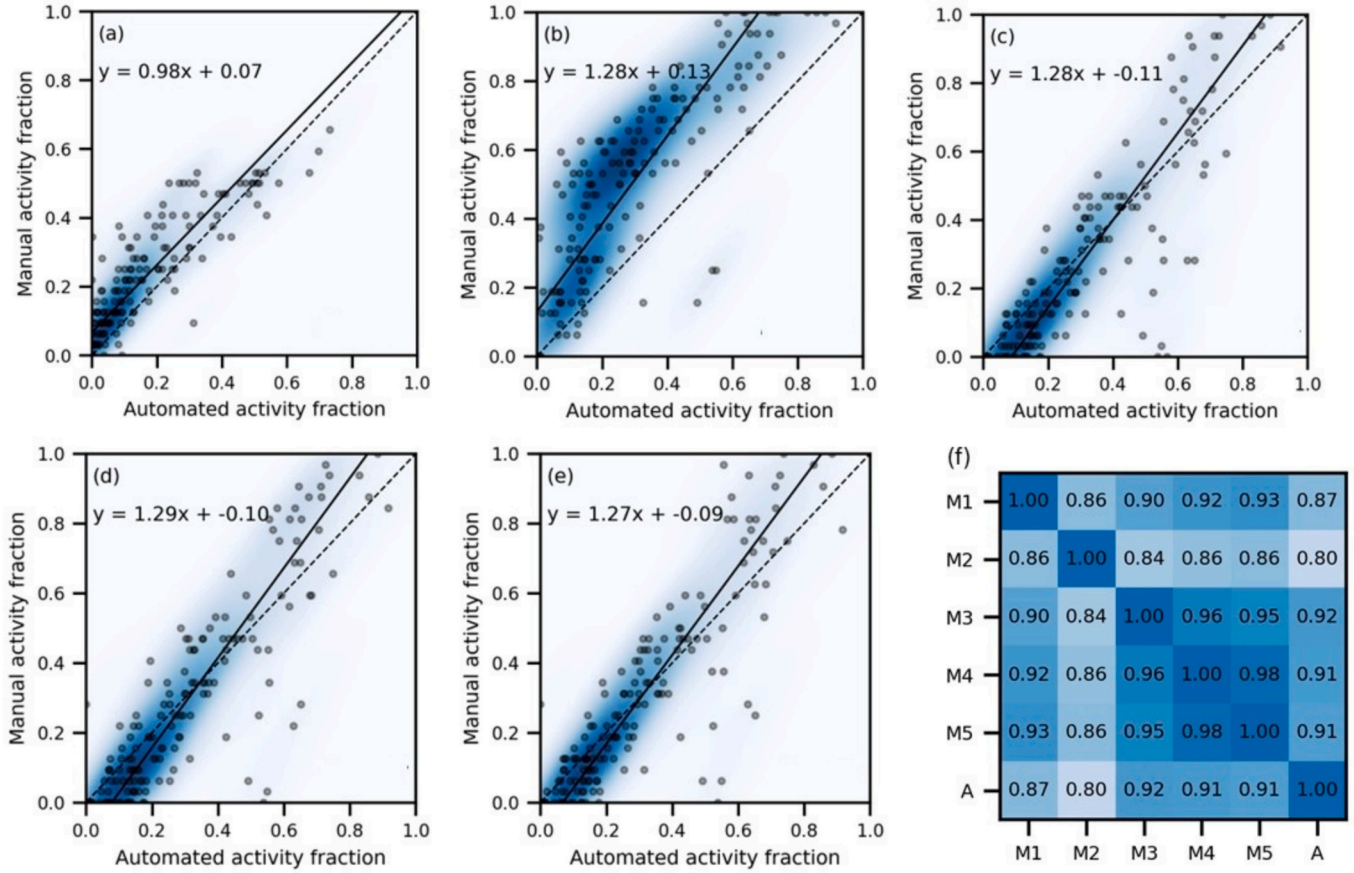


Fig. 3. (a–e) 2D kernel density scatter plots for comparing total activity fractions estimated by automated method with that of each operator: (a) M1, (b) M2, (c) M3, (d) M4, and (e) M5. The dotted line represents the Deming regression line and the equation for the regression line is shown. (f) Correlation matrix showing pairwise Pearson correlation coefficients between the automated method (A) and operators M1 – M5.

Table 3
Summary of agreement metrics between the automated feeding-activity estimates and the five operators (M1–M5), along with the manual consensus.

Operator	Pearson r	MAE	RMSE	Deming slope	Intercept	Mean bias	Accuracy
M1	**0.87**	**0.08**	**0.11**	**0.98**	+0.07	**+0.06**	**0.81**
M2	0.80	0.36	0.40	1.28	+0.13	+0.36	0.62
M3	**0.92**	0.13	0.18	1.28	−0.11	+0.12	0.84
M4	0.91	0.14	0.19	1.29	−0.10	+0.13	0.83
M5	0.91	0.15	0.20	1.27	−0.09	+0.14	0.79
Consensus	0.90	0.16	0.19	1.13	−0.02	+0.18	0.74

contrast, almost all borderline judgments fall into the Partial class, so even small scoring differences accumulate there, leading to lower correlations for Partial despite the additive constraint.

In this part of the analysis, we shift from hole-level classifications to the activity class assigned to the entire strip. Examining the calculated activity of entire strips, most of them fell into the partial activity category, and none was under full activity (Fig. A6 in Appendix). Using M1 as reference, manual scoring yielded 68 partial activity, 90 no activity, and 1 full activity strip. In contrast, the automated method assigned 107 partial and 51 no-activity classifications. Agreement was obtained for 66 partial-activity cases and 49 no-activity cases, with remaining mismatches concentrated at the partial/no-activity boundary. Overall accuracy was 0.81, and Cohen's κ was 0.48, indicating moderate agreement once the continuous values are collapsed into discrete categories. The remaining operators showed the same qualitative pattern but with lower agreement. M3 provided the closest categorical match among them (κ = 0.21, accuracy = 0.48), followed by M5 (κ = 0.19, accuracy = 0.45), and M4 (κ = 0.19, accuracy = 0.44). M2 showed the weakest categorical alignment (κ = 0.13, accuracy = 0.47). In all cases, the disagreements arose almost entirely from strips whose activity fraction lay close to the 5 % threshold: small shifts in the underlying continuous value were enough to move a strip from “partial” to “no activity,” creating categorical mismatches that do not reflect large differences. Representative cases of mismatches between manual and automated classifications, including missed holes and false positives, are shown in Fig. A7 (Appendix).

3.3. *Bias and systematic deviation*

To characterize operator-specific biases in the strip-wise activity fractions, we examined the signed difference distributions (manual - automated) for each operator (Fig. 4a–e). All five histograms are approximately unimodal and roughly symmetric, but they differ in location and spread. M1 (Fig. 4a) shows a relatively narrow distribution centered close to zero, in line with its near-unity Deming slope and minimal intercept. M2 (Fig. 4b) is clearly right-shifted, indicating that

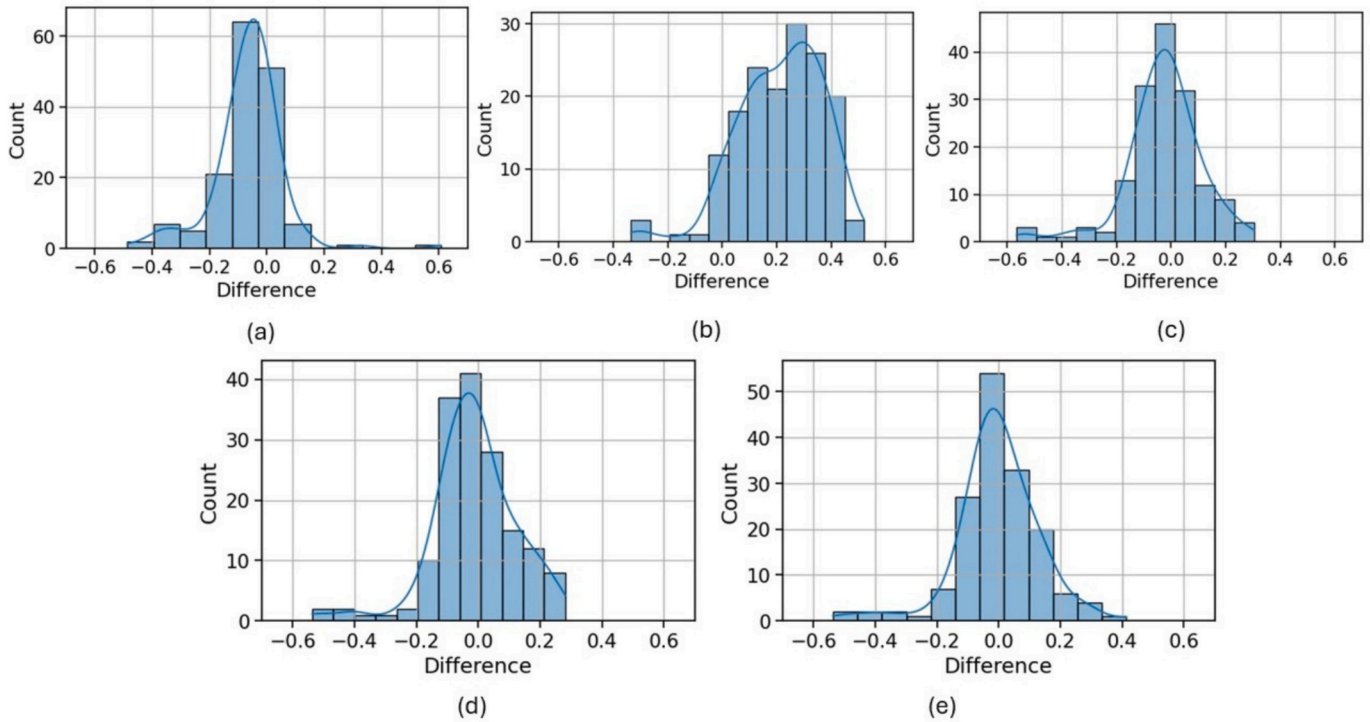


Fig. 4. (a – e): Histograms of the difference distributions, with kernel density overlays, illustrating the spread and skew of manual–automated deviations for operators M1 to M5, respectively.

this operator systematically assigns higher activity than the automated pipeline, consistent with its strong proportional bias (slope ≈ 1.28). M3 (Fig. 4c) also shows a positive shift and a wider spread, reflecting scaling differences suggested by its Deming slope (≈ 1.28). M4 (Fig. 4d) exhibits a modest positive shift with moderate dispersion, indicating a lower but still consistent overestimation. M5 (Fig. 4e) again clusters close to zero with only a slight right shift, suggesting small but systematic deviations. These histograms, together with the regression results, confirm that the deviations between methods are not random noise but reflect operator-specific tendencies in the interpretation of feeding activity.

Fig. 5 presents a detailed method comparison between the automated pipeline and manual scoring by operator M1 across different levels of faunal feeding activity. Bland-Altman analysis was carried out using M1 as the reference method to quantify fixed bias and the 95 % limits of agreement (LOA) between manual scoring and the automated activity fraction. Panels (a–c) show Bland-Altman plots for partial, no, and full activity categories, respectively. The differences between methods are mostly centered around zero with relatively narrow limits of agreement for partial (a) and no activity (b), indicating strong concordance in those classes. In contrast, the full activity plot (c) displays a larger positive bias, suggesting that the automated method tends to underestimate the number of fully eaten holes, likely due to image contrast or classification thresholds. Fig. 5d shows the Bland-Altman analysis for total activity fraction; it illustrates a small positive mean bias of 0.02, with LOA spanning from −0.15 to 0.19. This narrow range indicates that the automated pipeline deviates only minimally from the operator's assessments and remains within the expected variability of a human operator. The BA plot for activity fractions computed from the automated method and the mean of 5 operators is shown in Fig. A8 in Appendix. The analyses showed minimal fixed bias (mean difference = ±0.02). The limits of agreement ranged from −0.24 to +0.20, indicating that most strip-level deviations fell within a narrow band of ±0.2 activity units.

A summary of agreement metrics between the automated feeding-activity estimates and the five operators (M1–M5), along with the manual consensus, is given in Table 3. Pearson's r quantifies linear association, MAE and RMSE represent strip-wise error relative to manual scores, and Deming regression provides proportional and additive bias terms while accounting for uncertainty in both measurements. Mean bias reflects the average manual-automated difference in activity fraction. Accuracy refers to categorical agreement based on the thresholds (5 % / 95 %). Together, these metrics show that M3 and M4 align most closely with the automated pipeline, M1 displays minimal proportional bias, and M2 consistently assigns higher activity levels than both the automated method and the other human operators. The consensus behaves as an intermediate reference, moderating individual operator tendencies.

In addition to the pairwise Deming regressions presented earlier, a consensus-level Deming regression was computed between the automated feeding-activity fraction and the mean of the operators (M1–M5) to assess proportional agreement with the collective human reference (Fig. 5e). The resulting slope was 1.13 with a small negative intercept (−0.02), indicating a mild proportional deviation. As activity increases, the consensus score rises more steeply than the automated estimate. This behavior reflects the influence of M2, M3, and M4, who showed strong positive proportional bias individually, on the consensus distribution.

To assess whether the automated scores were practically indistinguishable from manual scoring, we applied a two one-sided test (TOST) with equivalence bounds of ±1/16 of a hole (±0.0625 activity units). Using the continuous strip-wise activity fraction, the differences between the automated method and operators M3, M4, and M5 were statistically equivalent within this tolerance ($p_{TOST} < 0.001$ in all cases; mean differences 0.003–0.027, 90 % CIs fully within ±0.0625). The automated scores were also equivalent to the consensus of all five operators (mean difference − 0.021, 90 % CI [−0.036, −0.006], $p_{TOST} = 3.8 \times 10^{-6}$). In contrast, the deviations relative to M1 (mean difference 0.067, 90 % CI [0.050, 0.083], $p_{TOST} = 0.66$) and especially M2 (−0.215, 90 % CI [−0.235, −0.194], $p_{TOST} = 1.00$) exceeded the predefined equivalence bounds and were therefore not considered practically equivalent. To determine whether the automated pipeline

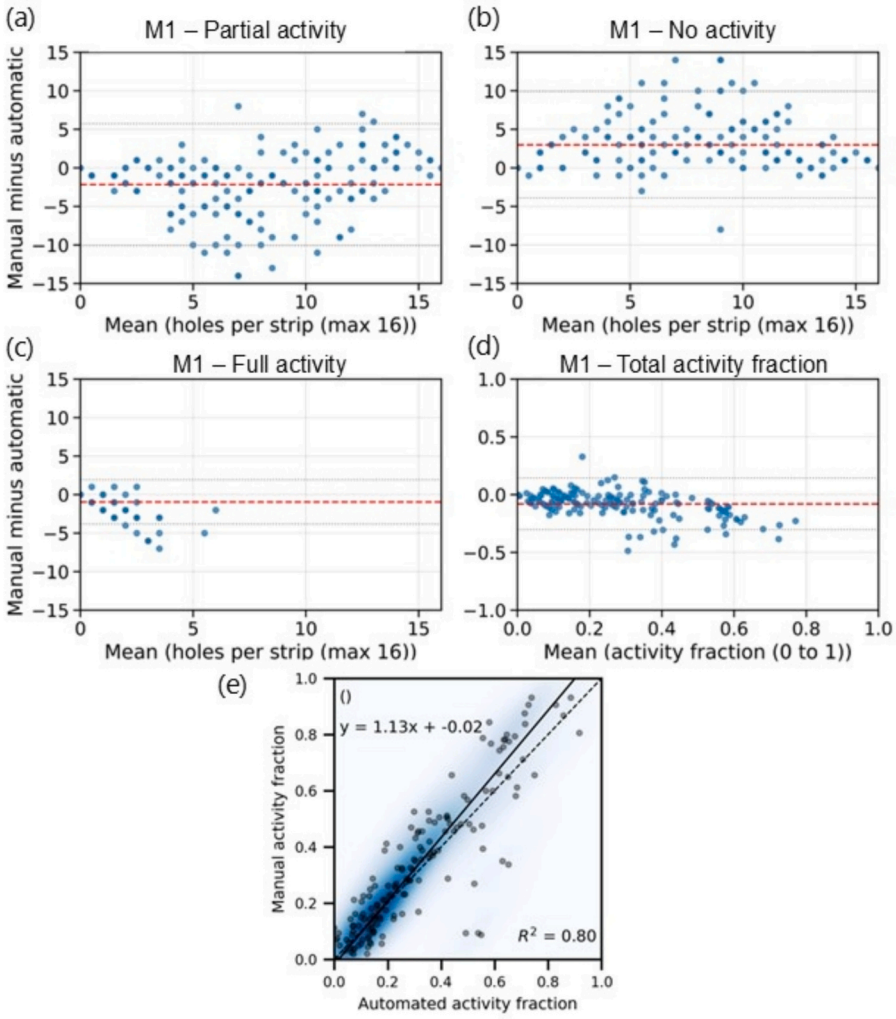


Fig. 5. Bland–Altman plot showing agreement range for: (a) Partial activity, (b) No activity, (c) Full activity, and (d) total activity fraction of operator 1(M1). (e) Deming regression comparing activity fractions between the manual (consensus) and automated method.

deviates systematically from the reference operator (M1), paired *t*-tests were conducted on two key metrics with 159 paired observations: the continuous total activity fraction and the discrete categorical score (0 = no feeding, 1 = partial, 2 = full). For the total activity fraction, M1's mean strip-wise activity was 0.226 (standard deviation [SD] = 0.095), where SD quantifies the typical deviation of individual strip measurements from the mean, compared with 0.206 (SD = 0.093) from the automated method. The mean difference of 0.020 (SD of the paired differences = 0.089) was statistically significant, $t(158) = 2.85$, $p = 0.0049$. Cohen's *d* (the mean difference divided by the standard deviation of the differences) was 0.23, indicating a small effect size. The comparison between automated and manual consensus scoring is summarized in Table A3 in the Appendix.

3.4. Plot- and strip-level patterns

Fig. 6 illustrates the per-strip feeding category assigned by each operator alongside the automated pipeline for a representative subset of ten plots. Fig. 6 reveals that, at the level of individual plots, the automated pipeline consistently falls within the range of human scoring rather than producing anomalous classification. In more variable plots, the blue trace follows the same within-plot fluctuations captured by at least one operator. This pattern holds across plots exhibiting minimal feeding, where the pipeline's blue circles cluster tightly with human "No feeding" observations, as well as in plots with more variable or intermediate activity, where blue trace the same subtle within-plot fluctuations captured by the operators.

Moreover, this agreement is remarkably uniform across a diverse set of field conditions and plots. Whether examining highly homogeneous plots (e.g., Plot_7) or those with pronounced heterogeneity (e.g., Plot_3, Plot_4), the pipeline's classifications remain anchored within the human-judged range. Such strip-level agreement underscores the robustness of the image-analysis workflow. It reliably reproduces the spectrum of manual scorings while eliminating instances of complete operator consensus divergence.

3.5. Computational reproducibility assessment

The strip-wise activity scores of the perturbed images showed a clear 1:1 relationship between the original and perturbed outputs (Fig. 7a), with a strong Pearson correlation ($r = 0.92$) and a two-way mixed-effects intraclass correlation ICC(3,1) of 0.92. Differences between paired measurements were generally small: the mean shift was 0.006 activity units on the 0–1 scale, and the Bland-Altman analysis indicated limits of agreement from −0.19 to +0.21 (Fig. 7b). The spread reflects expected sensitivity of circle detection to slight changes in alignment and brightness, but the overall pattern of strip-level activity is preserved. These results confirm that the automated extraction of continuous feeding activity is numerically stable under realistic perturbations in image acquisition.

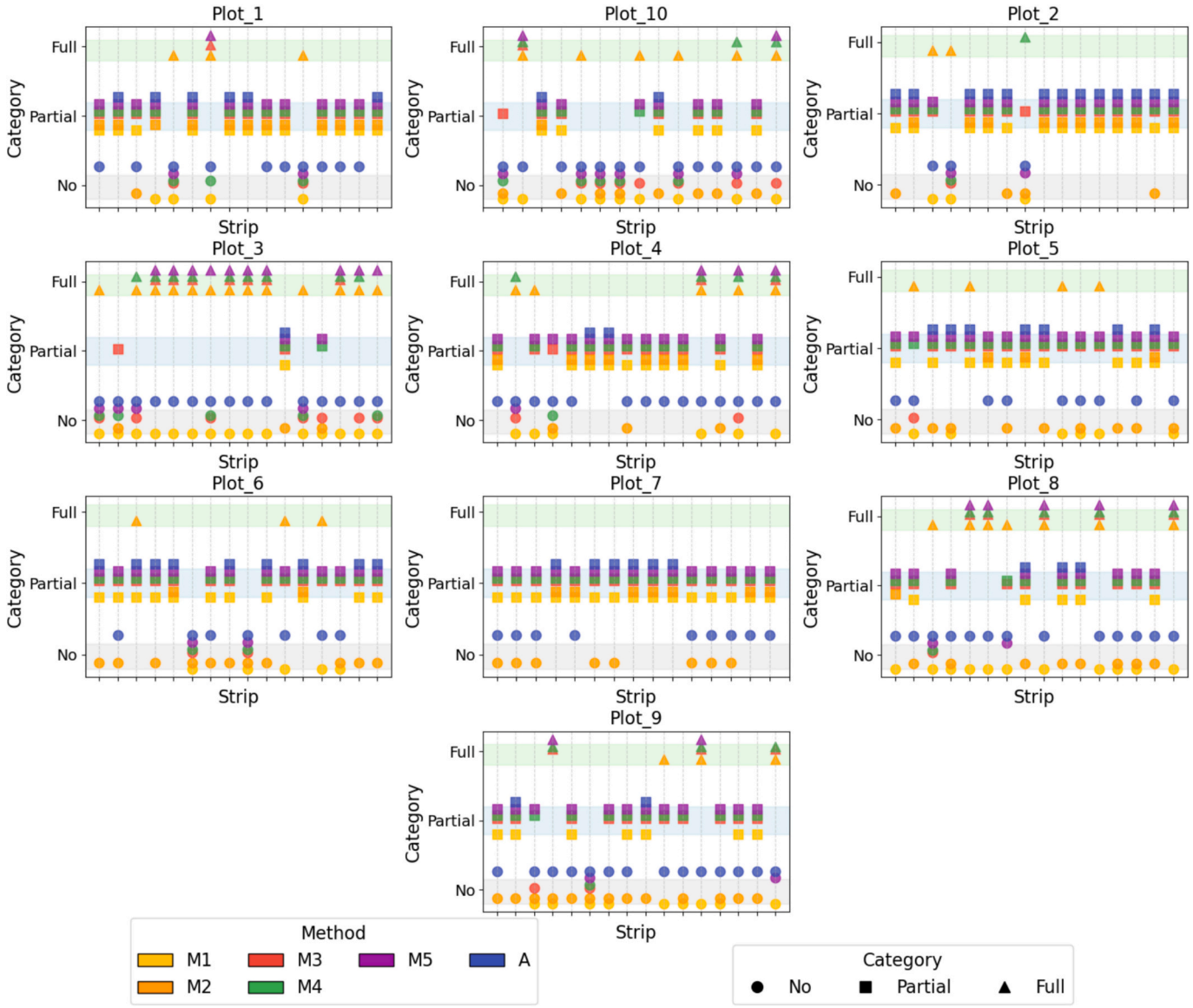


Fig. 6. Strip-level classification results for each bait lamina strip (1–16) across six scoring methods: operators (M1-M5), and the automated method.

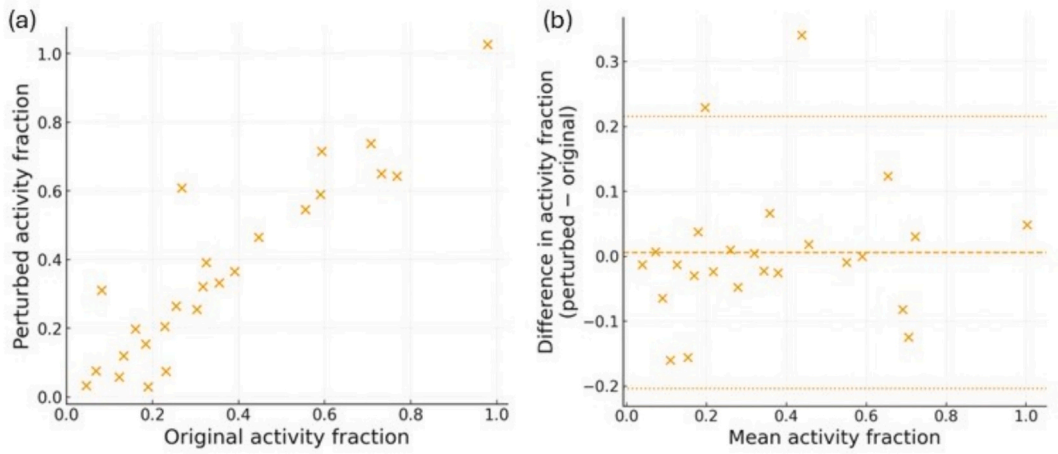


Fig. 7. (a) Relationship between activity fractions derived from the original strip images and from perturbed images combining a 0.5° rotation, a 1-pixel translation, and a 5 % increase in brightness. (b) Bland-Altman plot showing the difference in activity fraction (perturbed − original) against their mean, with the mean difference (dashed line) and 95 % limits of agreement (dotted lines).

4. Discussion

4.1. Advantages of automated workflow over manual scoring

This study revealed inconsistencies across operators for bait stick evaluation, indicating the importance of developing a clear, easy-to-follow observation protocol. Operator inconsistency in evaluating bait sticks was also reported by Eisenhauer et al. (2014) and is a common issue in environmental and ecological monitoring (Schmidt et al., 2023; Rivera-Palacio et al., 2025). Rivera-Palacio et al. (2025) identified that whether an operator follows the monitoring protocol or not can affect the agreement between human operators and computer vision-based observations by the degree of R^2 = 30 %. Identifying the highest agreement between a certain operator and a computer vision-based model necessarily guarantees neither the operator nor model. Manual classification of bait sticks is inherently subjective and may misclassify perforations as fully consumed even when minimal bait remains. Several annotated image snapshots illustrate the results produced by the automated system (Fig. 8). Unlike previous approaches that acknowledge variability in manual scoring as a methodological limitation, this study seeks to minimize that subjectivity through algorithmic standardization.

This research demonstrated that a computer vision-based system can successfully automate the evaluation of bait lamina stick consumption, providing a practical and scalable indicator of soil fauna activity. The strong correlation between automated findings and manual assessments confirms the method's reliability in capturing ecologically significant feeding behaviors. When all preparatory and handling steps are included, manual processing of 159 bait lamina strips requires roughly 3.5 h of operator time. In contrast, automated workflow, including arrangement, camera setup, and image quality check, requires about 2 h of active work and 10 min of unattended computation. This corresponds to about 40 % reduction in human effort. However, it should be noted that this time saving only accounts for the evaluation of the strips after the exposure. As the manual filling of the sticks is by far more time consuming than the evaluation and this step remains the same, the time saving needs to be seen in this context.

4.2. Limitations of automated workflow

The automated pipeline showed limitations in accurately identifying perforations that were completely eaten and thus, an overall underestimation in activity fraction is observed. This difference may arise from both the limitations of the reference method itself and imperfections in the algorithm. Visual scoring, although the accepted ISO 18311 standard, introduces observer-dependent variation that cannot be eliminated entirely (Eisenhauer et al., 2014). This variability is also evident in our dataset. A further limitation is that, despite expanding to five scorers, the manual reference remains constrained by the limited pool of available trained operators, which prevents constructing the larger frequency distributions that would be required to fully characterize the variability of the standard method.

Shadow cast on the border between outer and inner diameter of the perforations, especially on slightly bent sticks where the angle to the light source changes, could be one reason for the underestimation of fully eaten holes. Improvements here could increase the accuracy for sticks with high activities. However, heterogeneity in observation has to be considered as a non-negligible error source that must be addressed independently from model performance (Rivera-Palacio et al., 2025). The analysis of cumulative agreements demonstrates that more than 90 % of automated scores are within ±0.2 of the manual reference. Agreement remained high across all validation metrics, with continuous-score correlations up to $r = 0.92$ and categorical agreement accuracies up to 84 %, indicating that the method has promising potential toward establishing a generally applicable workflow. However, it

Fig. 8. Example outputs of the automated pipeline overlaid on bait-lamina strips. Each column represents a single strip, showing detected hole locations (outlined circles) and assigned feeding-intensity labels.

is important to recognize the trade-off between automation and interpretation details. In studies where accuracy in the classification of completely consumed bait is essential, it may still be necessary to implement additional correction measures or conduct human assessment. A further limitation is that the current workflow reports only total consumption per hole and does not capture the vertical position of feeding along the strip, even though depth-specific patterns can carry important ecological information.

4.3. Future directions and methodological improvements

It is noteworthy that the tasks we performed can also be done using deep learning-based predictive modeling (e.g. quantifying the number of holes and weighted activity score). While deep learning models could potentially outperform our approach, the proposed method offers key advantages: transparency, tunability, and computational efficiency. Each processing step from image acquisition to parameter tuning can be explicitly observed, controlled, and corrected, allowing domain experts to guarantee the outcome's quality. Without advanced programming skills, one can manually adjust settings such as color thresholds or reflectance parameters in response to context-specific variations. This procedural transparency is particularly important for trustworthy monitoring outcomes. For instance, Hough Transform has been implemented in real-time applications for decades (Mukhopadhyay and Chaudhuri, 2015). Moreover, the proposed method is algorithmically lightweight and fast, making it suitable for low-power on-site device deployment compared to deep learning.

Recognizing these strengths, we also see the potential of deep learning for future applications. For instance, zero-shot learning, few-shot learning and foundation model do not require a large number of training data: e.g. SAM for image segmentation (Kirillov et al., 2023); T-Rex for object counting by visual prompting (Jiang et al., 2023), and; Grounding DINO for object detection with text-based prompting (Ren et al., 2024). They may ultimately outperform the proposed method in terms of accuracy and widespread applicability as these approaches may not require programming skills (Mengsuwan et al., 2024). Still, their black-box nature makes it difficult to trace errors or manually tune the model when systematic error occurs. Direct method comparisons should be a focus for future studies, though they are beyond the scope of this paper.

Although weighing the remaining bait could, in principle, provide a physical measure of consumption, this is currently not feasible for hole-level validation. Even whole-strip mass measurements would lose the resolution that makes the BLT valuable. Consequently, visual scoring remains the established reference standard (ISO, 2016), against which new analytical approaches, including computer vision, should be evaluated. Future work could additionally explore integrating textile-dyed bait substrates (Eisenhauer et al., 2014), which offer improved visual contrast, to assess whether such enhanced materials further increase the robustness and accuracy of automated image-based scoring.

While our study shows promising results, several critical aspects need to be addressed for improving robustness and generalizability. The present workflow is optimized for well-cleaned bait sticks; heavy soil adhesion could be a limiting factor, as its removal still requires more manual effort before imaging. Future developments could focus on adapting the segmentation to tolerate surface contamination. Full automation will still require periodic human quality checks, and implementing routine inspection of a small subset of strips can help ensure that algorithmic performance remains stable across batches and imaging sessions. The parameters applied in this study may not generalize across all image conditions. Standardized image acquisition protocols, including lighting, camera distance, and background, are crucial for ensuring reproducibility. Our pipeline is computationally reproducible, but full scientific reproducibility through repeated independent measurements will need to be addressed in future studies. Moreover, optimization algorithms are often required to identify parameter sets, searching the center of circular perforations (Cauchie et al., 2008), though such procedures were not examined schematically in this study. External validation using another independent dataset is necessary to evaluate method transferability. Another logical next step is to extend the workflow toward categorical, depth-resolved interpretation of feeding location along the strip, which will require a dedicated modeling approach beyond pixel-fraction thresholds.

5. Conclusion

This research introduces a reliable and scalable computer vision method designed to automate the assessment of bait lamina stick consumption, serving as a widely used indicator of soil fauna activity. The automated technique showed significant concordance with manual evaluations, especially concerning partial feeding levels, and effectively minimized subjectivity and manual labor compared to conventional visual assessments. This study validates the practical accuracy of the method. The adoption of intensity-based classification thresholds allowed for a more detailed analysis of feeding behavior, representing a significant improvement over binary or categorical manual approaches. Beyond accuracy considerations, an important strength of automation is the creation of a standardized and permanently archived image record, which ensures transparent and verifiable interpretation across studies and over time. While the time savings from automated scoring may be modest in small experiments, they become operationally meaningful in larger monitoring programs where hundreds to thousands of strips must be processed, directly reducing labor demands and overall project costs. In this sense, improved processing speed acts as an enabling factor for scaling bait-lamina testing beyond small research plots.

To address current limitations, future improvements could include integrating machine learning techniques for more adaptive classification, enhanced preprocessing algorithms to reduce background noise, and the use of multispectral imaging for better discrimination of consumption states. Field-deployable solutions or smartphone-based applications could also extend the tool's usability beyond laboratory conditions, supporting high-throughput monitoring in ecological field studies and sustainable soil management programs. In summary, the proposed system represents a significant step forward in soil biological monitoring, offering a transparent, reproducible, and efficient alternative to manual scoring methods.

CRediT authorship contribution statement

Adrija Roy: Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Formal analysis, Data curation, Conceptualization. **Lukas Thielemann:** Writing – review & editing, Validation, Formal analysis, Data curation. **Masahiro Ryo:** Writing – review & editing. **Juan Camilo Rivera-Palacio:** Writing – review & editing. **Konlavach Mengsuwan:** Writing – review & editing. **Kathrin Grahmann:** Writing – review & editing, Writing – original draft, Supervision, Project administration, Funding acquisition, Conceptualization.

Funding sources

Authors acknowledge funding of BMFTR research group SoilRob (Funding ID: 031B1391). MR, KM, and JCRP acknowledge WIR! - Land – Innovation - Lausitz (LIL) project “Landscape Innovations in Lausitz for Climate-adapted Bioeconomy and nature-based Bioeconomy-Tourism” (03WIR3017A), and the Brandenburg University of Technology Cottbus-Senftenberg (BTU) with the Graduate Research School cluster project “Integrated analysis of Multifunctional Fruit production landscape to promote ecosystem services and sustainable land-use under climate change” (BTUGRS2018_19).

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Kathrin Grahmann reports financial support was provided by BMFTR. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

We thank our student assistant Israt Shupon for taking photos and scoring the images, Inna Matiash and Emily Payne for preparing and handling the bait sticks, and Björn Wang for his support. We also thank Israt, Emily and Kira Wöber for scoring the feeding activity manually as operators. The maintenance of the patchCROP infrastructure is supported by the Leibniz Centre for Agricultural Landscape Research. We acknowledge the valuable contributions of two anonymous reviewers.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ecolind.2025.114593.

Data availability

The code used in this study is openly available on GitHub at https://github.com/Adrija1/baitstick-analysis.

References

André, A., Antunes, S.C., Gonçalves, F., Pereira, R., 2009. Bait-lamina assay as a tool to assess the effects of metal contamination in the feeding activity of soil invertebrates within a uranium mine area. Environ Pollut 157 (8), 2368–2377. https://www.sciencedirect.com/science/article/pii/S0269749109001481.

Bardgett, R.D., van der Putten, W.H., 2014. Belowground biodiversity and ecosystem functioning. Nature 515 (7528), 505–511. https://www.nature.com/articles/nature13855.

Bart, S., Roudine, S., Amossé, J., Mougin, C., Péry, A.R.R., Pelosi, C., 2018. How to assess the feeding activity in ecotoxicological laboratory tests using enchytraeids? Environ Sci Pollut Res 25 (34), 33844–33848.

Beresford, N.A., Wood, M.D., Gashchak, S., Barnett, C.L., 2022. Current ionising radiation doses in the Chernobyl exclusion zone do not directly impact on soil biological activity. PLoS One 17 (2), e0263600.

Bergman, I.E., Vorobeichik, E.L., Ermakov, A.I., 2017. The effect of megalopolis environment on the feeding activity of soil saprophages in urban forests. Eurasian Soil Sci 50 (1), 106–117.

Bergstrom, A.C., Conran, D., Messinger, D.W., 2023. Gaussian Blur and Relative Edge Response, Unpublished.

Birkhofer, K., Baulechner, D., Diekötter, T., Zaitsev, A., Wolters, V., 2022. Fertilization rapidly alters the feeding activity of grassland soil Mesofauna independent of management history. Front Ecol Evol 10–2022. https://www.frontiersin.org/journals/ecology-and-evolution/articles/10.3389/fevo.2022.864470.

Brussaard, L., Ruiter, P.C. De, Brown, G.G., 2007. Soil biodiversity for agricultural sustainability. Agric Ecosyst Environ 121 (3), 233–244. https://www.sciencedirect.com/science/article/pii/S0167880906004476.

Cauchie, J., Fiolet, V., Villers, D., 2008. Optimization of an Hough transform algorithm for the search of a center. Pattern Recogn 41 (2), 567–574. https://www.sciencedirect.com/science/article/pii/S0031320307003160.

de Vries, Franciska T., Thébault, Elisa, Liiri, Mira, Birkhofer, Klaus, Tsiafouli, Maria A., Bjørnlund, Lisa, Jørgensen, Helene Bracht, Brady, Mark Vincent, Christensen, Søren, de Ruiter, Peter C., d'Hertefeldt, Tina, Frouz, Jan, Hedlund, Katarina, Lia Hemerik, W.H., Hol, Gera, Hotes, Stefan, Mortimer, Simon R., Setälä, Heikki, Sgardelis, Stefanos P., Uteseny, Karoline, van der Putten, Wim H., Wolters, Volkmar, Bardgett, Richard D., 2013. Soil food web properties explain ecosystem services across European land use systems. Proc Natl Acad Sci 110 (35), 14296–14301.

Eisenhauer, N., Wirsch, D., Cesarz, S., Craven, D., Dietrich, P., Friese, J., Helm, J., Hines, J., Schellenberg, M., Scherreiks, P., Schwarz, B., Uhe, C., Wagner, K., Steinauer, K., 2014. Organic textile dye improves the visual assessment of the bait-lamina test. Appl Soil Ecol 82, 78–81. https://www.sciencedirect.com/science/article/pii/S0929139314001656.

Filzek, P.D.B., Spurgeon, D.J., Broll, G., Svendsen, C., Hankard, P.K., Parekh, N., Stubberud, H.E., Weeks, J.M., 2004. Metal effects on soil invertebrate feeding: measurements using the bait Lamina method. Ecotoxicology 13 (8), 807–816.

Flusser, J., Farokhi, S., Höschl, C., Suk, T., Zitová, B., Pedone, M., 2016. Recognition of images degraded by Gaussian blur. IEEE Trans Image Process 25 (2), 790–806.

Förster, J., Barkmann, J., Fricke, R., Hotes, S., Kleyer, M., Kobbe, S., Kbler, D., Rumbaur, C., Siegmund-Schultze, M., Seppelt, R., Settele, J., Spangenberg, J.H., Tekken, V., Vclavk, T., Wittmer, H., 2015. Assessing ecosystem services for informing land-use decisions. A Problem-Oriented Approach Ecol Society 20 (3). http://www.jstor.org/stable/26270259.

Gonzalez, R., Woods, R., 2017. Digital image processing. Pearson International.

Hernández-Ochoa, I.M., Gaiser, T., Grahmann, K., Engels, A.M., Ewert, F., 2025. Within-field temporal and spatial variability in crop productivity for diverse crops—a 30-year model-based assessment. Agronomy 15 (3), 661.

Jänsch, S., Frampton, G.K., Römbke, J., van den Brink, P.J., Scott-Fordsmand, J.J., 2006. Effects of pesticides on soil invertebrates in model ecosystem and field studies: a review and comparison with laboratory toxicity data. Environ Toxicol Chem 25 (9), 2490–2501.

Jiang, Q., Li, F., Ren, T., Liu, S., Zeng, Z., Yu, K., Zhang, L., 2023. T-rex: Counting by Visual Prompting, Unpublished, p. 23.

Joschko, M., Oehley, J., Gebbers, R., Wiemer, M., Timmer, J., Fox, C.A., 2008. A spatial approach to soil-ecological experimentation at landscape scale. Journal of Plant Nutrition and Soil Science 171 (3), 338–343.

Kerbyson, D.J., Atherton, T.J., 1995. Circle detection using Hough transform filters. In: Fifth International Conference on Image Processing and its Applications, 1995, pp. 370–374.

Kierkegaard, P., 1992. A method for detection of circular arcs based on the hough transform. Mach Vis Appl 5 (4), 249–263. https://link.springer.com/article/10.1007/BF01212714.

Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A.C., Wan-Yen, Lo, Dollár, P., Girshick, R., 2023. Segment Anything, unpublished (30 p).

Larink, O., Sommer, R., 2002. Influence of coated seeds on soil organisms tested with bait lamina. Eur J Soil Biol 38 (3), 287–290. https://www.sciencedirect.com/science/article/pii/S1164556302011615.

Li, Q., Wu, M., 2020. An improved Hough transform for circle detection using circular inscribed direct triangle. In: 2020 13th international congress on image and signal processing. BioMed Eng Inform (CISP-BMEI) 203–207.

Mengsuwan, K., Rivera-Palacio, J.C., Ryo, M., 2024. ChatGPT and general-purpose AI count fruits in pictures surprisingly well without programming or training. Smart Agri Technol 9, 100688. https://www.sciencedirect.com/science/article/pii/S2772375524002934.

Mukhopadhyay, P., Chaudhuri, B.B., 2015. A survey of hough transform. Pattern Recogn 48 (3), 993–1010. https://www.sciencedirect.com/science/article/pii/S0031320314003446.

Musso, C., Miranda, H.S., Soares, Amadeu, M.V.M., Loureiro, S., 2014. Biological activity in Cerrado soils: evaluation of vegetation, fire and seasonality effects using the "bait-lamina test". Plant Soil 383 (1), 49–58.

Pehle, A., Schirmel, J., 2015. Moss invasion in a dune ecosystem influences ground-dwelling arthropod community structure and reduces soil biological activity. Biol Invasions 17 (12), 3467–3477.

Podgaiski, L.R., Da Silva, Goldas C., Ferrando, C.P.R., Silveira, F.S., Joner, F., Overbeck, G.E., Souza Mendonça, J.R.M., Pillar, V.D., 2014. Burning effects on detritivory and litter decay in Campos grasslands. Austral Ecol 39 (6), 686–695.

Ren, T., Jiang, Q., Liu, S., Zeng, Z., Liu, W., Gao, H., Huang, H., Ma, Z., Jiang, X., Chen, Y., Xiong, Y., Zhang, H., Li, F., Tang, P., Yu, K., Zhang, L., 2024. Grounding DINO 1.5: advance the "edge" of open-set object detection, unpublished, p. 25.

Ritz, K., Black, H.I., Campbell, C.D., Harris, J.A., Wood, C., 2009. Selecting biological indicators for monitoring soils: a framework for balancing scientific and technical opinion to assist policy development. Ecol Indic 9 (6), 1212–1221. https://www.sciencedirect.com/science/article/pii/S1470160X09000508.

Rivera-Palacio, J.C., Bunn, C., Ryo, M., 2025. Factors affecting deep learning model performance in citizen science–based image data collection for agriculture: a case study on coffee crops. Comput Electron Agric 232, 110096. https://www.sciencedirect.com/science/article/pii/S0168169925002029.

Römbke, J., Schmelz, R.M., Pélosi, C., 2017. Effects of organic pesticides on Enchytraeids (Oligochaeta) in agroecosystems: laboratory and higher-tier tests. Front Env Sci 5.

Schmidt, B.R., Cruickshank, S.S., Bühler, C., Bergamini, A., 2023. Observers are a key source of detection heterogeneity and biased occupancy estimates in species monitoring. Biol Conserv 283, 110102. https://www.sciencedirect.com/science/article/pii/S0006320723002033.

Siebert, J., Sünnemann, M., Auge, H., Berger, S., Cesarz, S., Ciobanu, M., Guerrero-Ramírez, N.R., Eisenhauer, N., 2019. The effects of drought and nutrient addition on soil organisms vary across taxonomic groups, but are constant across seasons. Sci Rep 9 (1), 639.

Simpson, J.E., Slade, E., Riutta, T., Taylor, M.E., 2012. Factors affecting soil Fauna feeding activity in a fragmented lowland temperate deciduous woodland. PLoS One 7 (1), e29616.

Thakur, M.P., Reich, P.B., Hobbie, S.E., Stefanski, A., Rich, R., Rice, K.E., Eddy, W.C., Eisenhauer, N., 2018. Reduced feeding activity of soil detritivores under warmer and drier conditions. Nat Clim Chang 8 (1), 75–78.

van der Heijden, M.G.A., Wagg, C., 2013. Soil microbial diversity and agro-ecosystem functioning. Plant Soil 363 (1–2), 1–5. https://link.springer.com/article/10.1007/s11104-012-1545-4.

Von, Törne E., 1990. Assessing feeding activities of soil-living animals. I. Bait-lamina-tests. Pedobiologia 34, 89–101.

Vorobeichik, E.L., Bergman, I.E., 2020. Bait-Lamina test in the assessment of polluted soils: choice of exposure duration. Russ J Ecol 51 (5), 430–439.
Vorobeichik, E.L., Bergman, I.E., 2021. Bait-lamina test for assessment of polluted soils: rough vs. Precise scales. Ecol Indic 122, 107277.
Vorobeichik, E.L., Bergman, I.E., 2023. Modification of the bait-lamina test to estimate soil macrofauna and mesofauna feeding activity. Soil Biol Biochem 183, 109047.