Smart Agricultural Technology 12 (2025) 101667

Contents lists available at ScienceDirect

Smart Agricultural Technology

journal homepage: www.journals.elsevier.com/smart-agricultural-technology

Use of an AI-based annotation tool reduces inter-observer error linked to cover estimation of arable crops and weeds

Alessandra Virili [a], Alex Falcon [b], Beatrice Portelli [a,b,*], Alessandro Peressotti [a], Giuseppe Serra [b], Elisa Marraccini [a]

[a] University of Udine, Department of Agriculture, Animal, Environmental and Food Sciences, 33100 Udine, Italy
[b] University of Udine, Department of Mathematics, Computer Science and Physics, 33100 Udine, Italy

ARTICLE INFO

Keywords:
Vegetation
Surveys
Observation
Intercropping
SAM
VegAnn

ABSTRACT

Vegetation surveys are employed in agro-ecosystems to understand weed eco-biology and develop integrated weed management solutions. Vegetation cover through visual estimation is the most common, non-destructive measurement but is often associated to inter- and intra-observer variability. The present study investigates whether the inter-observer error linked to experience can be reduced after using an AI-based annotation tool that we developed. The tool combines general and domain-specific knowledge using Segment Anything Model and VegAnn to assist the annotator. Two vegetation surveys were performed at different growth stages of a lentil-buckwheat intercropping trial. Vegetation and soil cover were evaluated independently by two observers (expert and novice) on two sampling areas per plot. Each sampling area was photographed (Samsung Galaxy A42 smartphone) and the pictures were processed by the novice with the AI-based tool. The effect of the observer type (expert, novice, AI-tool) on the cover of crops, weeds and bare soil was tested using generalized least squares-mixed effect models. Overall, after using the AI-tool the gap between expert and novice decreased by 44 % for weeds but increased by 8 % for lentil, by 11 % for buckwheat, and by 4 % for soil. Nonetheless, when the spatial layout and the growth stage of the crops were considered, the gap between expert and novice was reduced in almost all cases. We also proved that the AI tool is useful for novice observer training prior to entering the field and which can be further developed to aid researchers and farmers in estimating vegetation cover.

1. Introduction

The extensive use of herbicides has led to few and competitive weed species dominating arable fields [1]. Integrated weed management encompasses a wide range of practices which aim to reduce weed populations in the field while maintaining diversified communities, contributing to the overall resilience of the agro-ecosystem [2,3]. The effect of agro-ecological management practices (e.g., cover crops, intercropping, false seed beds) on weed communities is assessed though the identification of weed species and the quantification of the weight of each species in the community [4,5]. This method allows to establish whether certain practices lead to a community with overall positive or negative traits based on the number of species, their competitiveness, nutrient requirements or habit [6]. Vegetation surveys are coupled with the measurement of biomass, density or cover [7]. The decision on which measurement to use depends on both the objective of the study and the resources available. Biomass can be considered as the most unbiased measurement, but being destructive it prevents repeated measures on the same sampling areas [8]. The weighing and separation processes are also time consuming, although the processing of the collected samples requires a certain rapidity due to the wilting of vegetative tissues which can make species identification difficult. Furthermore, indices such as cover or density are better suited to answer some agronomical questions, such as the effect of cropping practices on weed suppression [9,10], have been found to be linearly correlated to biomass in low open herbaceous vegetation [8,11] and in some cases have been found to be better predictors for yield loss compared to weed biomass [12].

Cover assessments through visual estimation allow to carefully observe the community and employ less time compared to other sampling methods, such as the random-point quadrat (RPQ) [13]. Similarly to biomass and density, vegetation cover assessments can be performed on the entire plant community or by dividing the population in categories (forbs and grasses, crops and weeds, etc..), up to species level. On

* Corresponding author at: University of Udine, Department of Mathematics, Computer Science and Physics, 33100 Udine, Italy.
E-mail address: beatrice.portelli@uniud.it (B. Portelli).

https://doi.org/10.1016/j.atech.2025.101667
Received 31 March 2025; Received in revised form 7 October 2025; Accepted 23 November 2025
Available online 25 November 2025

the contrary, visual cover assessments are an acquired skill and are prone to both inter-observer and intra-observer error [14].

The three main sources of error in visual cover estimation are related to: 1) the characteristics of the vegetation, 2) the environmental conditions, and 3) the observer [14]. As an example, vegetation cover estimation in a vineyard is likely to be less precise compared to an assessment in a corn field after an herbicide application [15]. Weather conditions, mosquitoes or harsh sunlight may disturb the observer [14, 15]. Finally, mental or physical fatigue and differences in experience may affect the precision of the assessment [14,16,17]. Although the assessment of each site may be quick, vegetation surveys can be demanding for one observer to complete when many sites are involved, thus often two or more observers carry out the observation in the field. This poses the additional task of calibrating observations between observers every so often, especially if they have different levels of experience [14].

Several papers have considered assessment of vegetation cover percentage through remote-sensing techniques, digital image processing or indices [18–21]. An interesting agronomical application of image-processing techniques was developed by Wiles [10], who created a user-friendly application for the estimation of weed cover in fallow fields using GIS files. Although promising, these methods seem best suited for monitoring of large areas such as forests or grasslands or seem to be targeted towards precision agriculture, while researchers often study weed communities at plot level. Furthermore, the time investment in using some of these techniques may overweigh the benefit particularly for users who are not accustomed to using programs like GIS.

As pointed out by Morrison [14], picture analysis is not always an adequate solution when layers of vegetation are present. Visual vegetation surveys in small areas such as 0.25 m^2 take around 2–5 min to complete, including the identification of species. Furthermore, visual estimations were found to be more repeatable compared to other in situ assessments [13].

This study compares the cover estimation from two observers: an expert (5+ years of experience) and a novice (a bachelor student with no prior experience). In addition, we developed and tested a novel AI-based tool based on two models for shape and vegetation recognition (SAM) [22] and VegAnn [23]. The tool provides cover estimates of species/categories through manual image annotation assisted by the AI models. The efficacy and efficiency of the AI-tool were also evaluated by tasking the novice of processing the same observed areas with the tool after the field surveys.

Morrison [14] reports that comparisons between observers have been carried out in ecological survey settings, such as forests, woodlands, grassland, or meadows. To the best of the authors' knowledge, the present study is one of the few to focus on an agricultural area (see also [9]). Specifically, observations were carried out in an intercropping trial involving lentil (*Lens culinaris* Medik) and buckwheat (*Fagopyrum esculentum* Moench). This trial was particularly well-suited for this study as it included cropping systems with different complexities, allowing to investigate an interaction between the characteristics of the vegetation (first source of error) and the experience of the observer (third type of error). This paper provides considerations on species-specific cover based on the precision of the different observers at estimating vegetation categories and soil cover. The main objectives of the study were to: 1) measure the reliability of different observers with varying levels of expertise; 2) measure the effect of the AI tool in reducing the observation gap between observers; 3) understand the trade-off between the duration of the observations and the variability of the obtained data. The estimations provided by the novice were hypothesized to have a greater coefficient of variation compared to the expert, making intra- and inter-observer errors dependent on experience. The AI-tool was hypothesized to reduce the gap between novice and expert, and even to provide more precise values compared to visual estimation, in the early growth stages when vegetation is sparse. In conditions of stratified vegetation, the AI-tool was hypothesized to under-estimate values compared to human observations.

2. Materials and methods

2.1. Experimental design

Lentil and buckwheat were sown on the 9th of April 2024 in an intercropping trial with two spatial arrangements: within row (MIX) and alternate row (ALT) intercropping. Two buckwheat plant densities were considered for the MIX plots, resulting in plots with 25 % of buckwheat seed rate (MIX25) and 50 % of buckwheat seed rate (MIX50). In the ALT plots, buckwheat seed rate was also halved compared to the sole plots. Plots had four rows with 34 cm row spacing in the pure lentil and MIX plots, while the ALT and pure buckwheat plots had eight rows spaced 17 cm apart (Fig. 1). Trials followed a completely randomized design with four replicates per treatment.

2.2. Data collection

Crop emergence was registered on the 20th of April. Two vegetation surveys were carried out three weeks apart: 1) T1 - 15 days after crop emergence (third true leaf stage of lentil; second true leaf stage of buckwheat) and 2) T2 - 38 days after crop emergence (branching stage of lentil; full flowering stage of buckwheat). Vegetation (lentil, buckwheat, weeds) and soil cover were evaluated independently by two observers (expert and novice) on two sampling areas of 0.35 m^2 (0.5 m x 0.7 m) in each plot, resulting in eight observed sampling areas for each treatment at each sampling date ($n = 40$ at T1, $n = 40$ at T2). Cover was estimated by the projection of the vegetation category on the ground surface expressed as percent covered by each category. Prior to the first survey, the novice was trained in the field by the expert (explanation provided in Appendix A). The total duration of the observation was timed for each observer. Each sampling area was then photographed from a consistent height (1 m) facing the same direction and with the same orientation of the device (Samsung Galaxy A42 smartphone), the pictures were then cropped to the size of the sampling frame (Fig. 1).

2.3. Annotation tool

The tool is a desktop application developed in Python and leverages PyTorch, Pandas, and Numpy extensively. It offers a graphical user interface supported by the PyGame library (Fig. 2A). The tool allows the user to load images of any size, manage labels (e.g., adding labels for lentil or buckwheat), and perform pixel-wise annotations directly on the image. Annotations can be made manually using a brush, but the primary strength of the tool lies in its AI-assisted annotation feature, which suggests regions of pixels to annotate through a simple click on the image.

The AI annotation is powered by the Segment Anything Model (SAM) [22] and VegAnn [23]. SAM is a state-of-the-art foundation model for image segmentation designed to segment any object within an image based on minimal hints such as points, boxes, or freeform masks. It allows for general-purpose segmentation tasks across diverse domains. VegAnn, on the other hand, is a specialized tool tailored for vegetation analysis and agricultural imagery. It combines traditional computer vision techniques with deep learning to distinguish between vegetation and non-vegetation areas.

The tool supports three methods of annotation, depicted in Fig. 2B: "AI" tool with a single click, "AI" tool by clicking and drawing, and "Brush". In the two "AI" cases, the user identifies an object of interest by clicking or drawing and the SAM model is prompted to generate a mask based on the user's input. On the contrary, when using the "Brush" annotation method, the user manually colors all the areas they wish to annotate. In all cases, before finalizing the annotation, the mask generated via the "AI" or "Brush" tool is optionally combined via intersection with the binary VegAnn mask to filter out non-vegetation

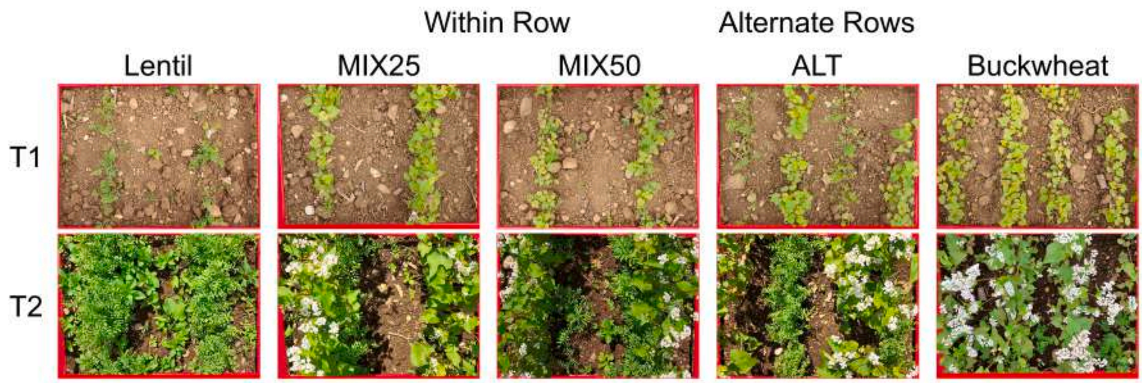

Fig. 1. Photos of the surveyed areas in the five cropping systems taken at 15 days after crop emergence (T1) and at 38 days after crop emergence (T2). MIX25: mixed row intercropping with buckwheat sown at 25 % of its pure sowing density; MIX50: mixed row intercropping with buckwheat sown at 50 % of its pure sowing density; ALT: alternate row intercropping with buckwheat sown at 50 % of its pure sowing density. The red border visible in the photos is the 0.5 × 0.7 m frame.

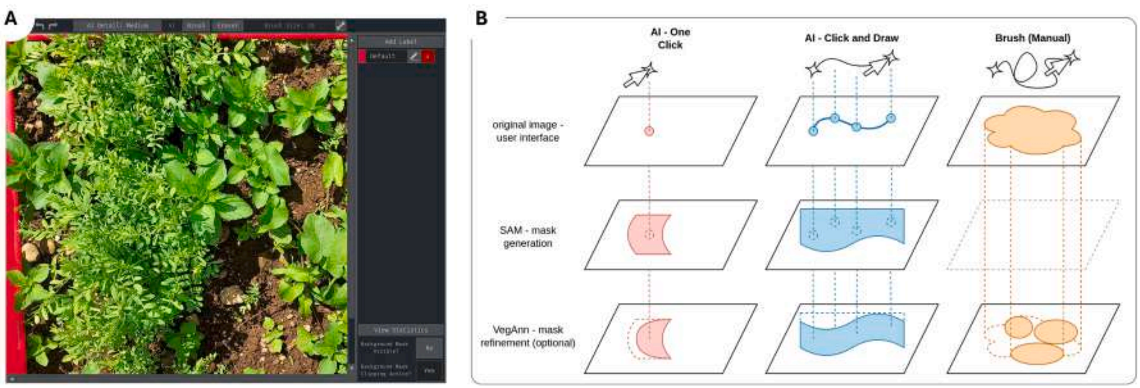

Fig. 2. A) Screenshot of the annotation tool interface. B) Graphical depiction of the three annotation methods supported by the developed tool. From left to right: "AI" tool with a single click, "AI" tool by clicking and drawing, and "Brush" tool. The VegAnn mask refinement (bottom row) is optional (active by default) and can be applied to both AI and manual annotations.

areas or background pixels (Fig. 2B, bottom row).

2.3.1. *Interface and details of the annotation tool*

Upon loading the image for the first time, a series of preparatory steps is performed to setup the AI models: the image is resized to 1024×1024 and normalized, SAM is used to compute and store the embeddings of the image to enable repeated segmentations, and the image is processed once by VegAnn to generate and store a binary vegetation / non-vegetation mask.

Following these preprocessing steps, the user is presented with an interface (Fig. 3) that shows the original image and a series of tools that can be used to annotate it (a), in particular "AI" and "Brush". When "AI"

Fig. 3. Detailed screenshot of the annotation tool interface.

is selected, the user can use a single click or draw lines to indicate regions for segmentation. This prompts SAM to generate a mask for the selected object (i.e., a plant). SAM can generate three different masks per prompt corresponding to different levels of granularity (low, medium, and high). For example, if a user clicks on an image of a car, SAM might propose a mask for the entire car, a mask for the car door, or a mask for the window. The application defaults to "medium" granularity and the user can change it at any moment using the "AI Detail" button (b). If the "Brush" tool is selected (a), the user can manually paint over the image to annotate it, changing the brush size as needed.

In all cases, before finalizing the annotation, the mask generated via the "AI" or "Brush" tool is combined via intersection with the binary VegAnn mask to filter out non-vegetation areas (background pixels). By clicking on the toggle "Background Mask Visible?" (c) the user can visualize the precomputed VegAnn vegetation mask as a colored overlay (d). In addition, they can temporarily disable the effect of the background mask clicking on the "Background Mask Clipping Active?" toggle (c), allowing them to annotate pixels that VegAnn might have misinterpreted as non-vegetation. Fig. 2B summarizes the functioning of the three available annotation methods.

Finally, the panel on the right allows the user to switch between different labels (e) or create new ones. At any stage, the user can generate a summary of plant coverage (f), which provides the percentage of annotated pixels for each label, along with general statistics (e.g., the number of annotated and unannotated pixels). This functionality offers insights into plant distribution and annotation progress.

2.4. Evaluation of model size and granularity

The SAM model is publicly available in different sizes (e.g., Base, Large, Huge), hardware requirements, and segmentation performance. In addition, each of them allows to annotate images at various levels of granularity (coarse, medium, fine). It is therefore important to determine which variant to incorporate in the annotation tool. For this reason, we systematically tested all the SAM variants on the collected images while varying the granularity of the generated masks. Specifically, we tested all nine combinations (three SAM variants × three mask granularities) on the annotated images. For each image, we simulated 400 user clicks, uniformly distributed across a 20×20 grid. The label (plant category) assigned to each annotated pixel corresponds to the ground truth label at the click position, simulating the label selection process performed by an annotator.

The performance of the SAM combinations was analyzed considering the aggregated metrics over all the images (ALL) and category-level metrics (e.g., weed, buckwheat, lentil). We used three metrics commonly employed to assess segmentation tools: Precision (P), Recall (R), and F1-Score (F1), focusing particularly on the F1-Score as a summary metric. Given a class C (e.g., buckwheat): Precision measures how accurate the annotations are, by taking the percentage of pixels annotated as C that are from class C (high precision indicates fewer false positives, i.e. pixels incorrectly labeled as C). Formally:

$$P = \mathrm{TP} / (\mathrm{TP} + \mathrm{FP})$$

where TP and FP are the pixels correctly and wrongly labeled as C, respectively. Recall reflects how well the tool detects all instances of C, computing the percentage of actual C pixels that were correctly annotated (high recall means fewer false negatives, i.e. C pixels missed by the annotation). Formally:

$$R = \mathrm{TP} / (\mathrm{TP} + \mathrm{FN})$$

where FN are the pixels from class C that were not labeled as such by the model. Finally, the F1-Score is the harmonic mean of Precision and Recall. A high F1 indicates both good accuracy and comprehensiveness of the annotation process.

$$\mathrm{F1} = 2 * \mathrm{TP} / (2 * \mathrm{TP} + \mathrm{FP} + \mathrm{FN})$$

2.5. Data analysis

All statistical analyses were performed using R (v. 4.1.0) [24]. The effect of the observer type (expert, novice, AI-tool) on the cover of lentil, buckwheat, weeds and bare soil was tested using generalized least squares (GLS) mixed effect models (MEM). For each vegetation type model selection was performed using a linear mixed effect model, with (1|plot) as the random structure to account for the repeated measures of the vegetation. The main interest of this study was to assess whether the observer increased the strength of the model. Model selection was performed starting from an "a priori" model containing cropping system and sampling date, as well as their interaction. This null model was tested against other five models containing the observer, alone or with the different interactions. In this way we tested six different hypotheses on the effect of the observer on each response variable [25,26]. The six hypotheses for each response variable were compared based on the Bayesian Information Criterion (BIC) value. The final models for each response variable were validated by checking distribution and homogeneity of residuals. After confirming the non-homogeneous variance of the data [27] generalized least squares (GLS) were used. The weights of the models were selected for each response variable. The GLS model was then tested against a model containing a random structure ("random = ~1|plot") (MEM). The model with the random structure had the lowest BIC and thus was selected [28]. The models were validated by checking distribution and homogeneity of residuals. No post-hoc test was performed in this case since mixed effect models already perform a partial pooling of the estimates towards a common mean [29,30].

3. Results

3.1. Model size and granularities

As the detail level varies from low granularity (coarse) to high granularity (fine-grained), less pixels are annotated, yet those pixels are more accurate, meaning that they do not fall outside of the desired region. For example, on the left side of Fig. 4 (Base-Low, coarse masks), almost all the pixels belonging to lentil are detected (light blue), yet many incorrect pixels were also captured, e.g. near the buckwheat leaf (orange). Conversely, as the granularity becomes finer (Base-High), fewer pixels are captured with more precision. Therefore, it becomes important to evaluate how the different granularities, as well as the SAM variants, influence the annotation performance and the support offered to the user.

Quantitative results (F1 score) for all the combinations are reported in Table 1. Further details about P and R are discussed in Appendix B.

With respect to the F1 index (Table 1), SAM Base shows a more robust performance compared to the other model sizes. Base-Low has the best results for class-agnostic pixel coverage (93.0 and 96.5 % in T1 and T2 respectively), Base-Mid is the best performing combination for Lentil, Base-Low is the best model for Buckwheat at T2 and the second-best at T1. As regards the Weeds class, Base-Mid is the third-best combination in terms of F1 score due to the low Recall.

3.2. Effect of observer type and vegetation complexity on cover estimation

A summary of the best model for each cover type is included in Table 2. For lentil, buckwheat and soil, the full model including the three-way interaction between sampling time, cropping system and observer had the lowest BIC. In the case of weed cover, the explanatory capacity of the model was best represented when the three-way interaction was excluded, but all two-way interactions were included.

The range of the coefficient of variation (CV) was 2 %−107 % at the first survey date and 8 %−84 % at the second survey date. The expert

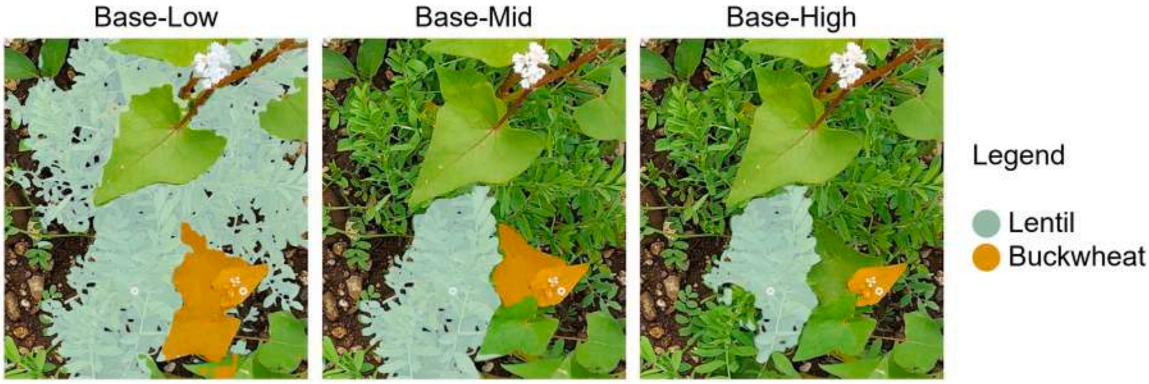

Fig. 4. Examples of masks suggested by SAM Base with different detail levels on one of the collected images. The masks are generated by prompting SAM with two clicks, highlighted with a white circle. From left to right: low granularity masks – Base-Low, medium granularity – Base-Mid, and high granularity masks – Base-High.

Table 1
Comparison of nine combinations for the SAM component of our annotation tool. Three SAM variants (Base, Large, Huge) with three granularities (Low, Mid, High) are tested on both annotation times (T1 and T2). Performance is reported using F1 both in aggregated form (ALL) and per class (Weeds, Buckwheat, Lentil). The best results for each time and class are bolded and underlined, the second-best results are bolded.

F1		ALL			Weeds			Buckwheat			Lentil		
		Low	Mid	High	Low	Mid	High	Low	Mid	High	Low	Mid	High
T1	Base	**93.0**	79.1	58.7	32.4	40.4	27.9	**83.3**	76.2	58.9	63.6	**70.0**	52.8
	Large	**92.9**	86.9	64.4	36.3	**41.9**	31.1	82.1	82.3	62.3	60.4	68.1	60.1
	Huge	92.7	86.2	66.8	37.3	**40.8**	33.0	**84.9**	81.0	63.3	63.9	**69.1**	63.5
T2	Base	**96.5**	88.4	75.2	30.6	57.6	44.6	**84.8**	**82.8**	71.7	74.0	**86.6**	75.1
	Large	96.1	89.6	78.1	32.7	**61.2**	47.9	78.8	80.2	74.2	68.0	**80.5**	77.9
	Huge	**96.2**	89.9	79.0	37.0	**59.5**	48.8	79.6	78.7	74.3	70.2	78.7	80.1

Table 2
Comparisons of the five regression models for lentil, buckwheat, weed and soil cover (%). The Bayesian Information Criterion (BIC) was used as a tool for model selection. The variables tested for each response variables were the sampling date (“date”), the cropping system (“CS”), and the observer (“O”). The models with the lowest BIC are underlined for each response variable.

model		Lentil		Buckwheat		Weeds		Soil	
		BIC	DF	BIC	DF	BIC	DF	BIC	DF
m0	(date : CS)	1215.29	7	1321.61	7	1477.43	8	1741.86	8
m1	(date : CS) + *O*	1214.78	9	1297.36	9	1427.46	10	1731.99	10
m2	(date : CS) + *O* + date : O	1215.47	11	1282.23	11	1358.29	12	1715.19	12
m3	(date : CS) + *O* + date : *O* + CS: O	1204.24	20	1277.61	20	**1357.28**	**24**	1712.66	24
m4	(date : CS) + *O* + date : *O* + CS : *O* + date : CS : O	**1183.43**	**26**	**1265.91**	**26**	1363.81	32	**1693.04**	**32**

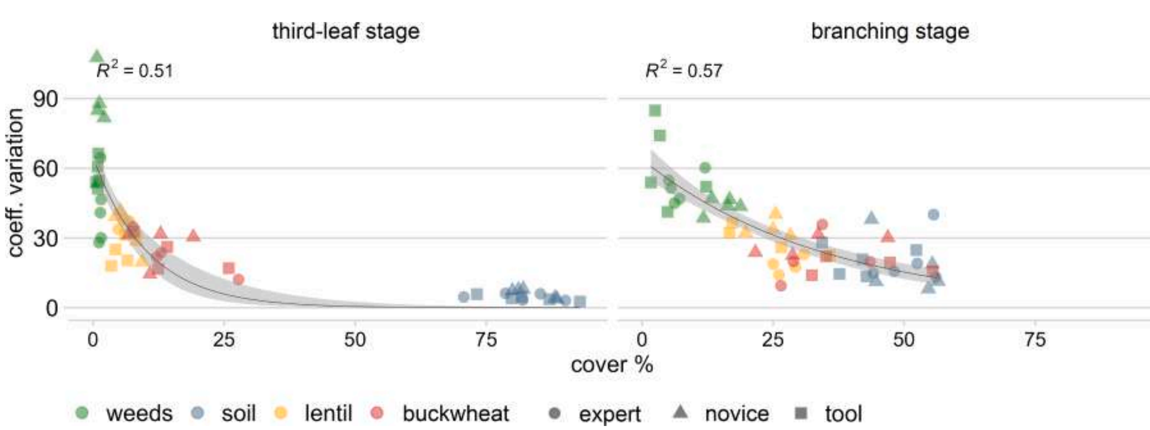

Fig. 5. Relationship between cover % and the coefficient of variation among observers for each cover category, in the two sampling dates (T1: lentil third-leaf stage,15 days after emergence; T2: lentil branching stage, 38 days after emergence).

observer recorded the lowest overall CV (27 %), followed by the AI-tool (29 %), and then the novice observer (33 %). The highest CV was observed for the weed category in the first survey date (60 %, averaged across observer type) (Fig. 5). At weed over values between 0–5 % the novice CV was of 83 %, compared to the expert (42 %) and the AI tool (57 %). At higher weed cover the inter-personal error decreased, with CV values of 44 % for the novice, 51 % for the expert and 61 % for the AI tool.

The CV range was lowest for the soil category at the first sampling date (2 %–8 %), when values were >70 %, but increased to 8 %–40 % when cover values decreased to around 30 %–50 % (Fig. 5). Inter-personal error was low in both cases, thus the higher range at T2 is attributed to intra-personal error. Coefficient of variation ranges remained constant for both the lentil and buckwheat categories between the two survey dates: the lowest inter-personal error was recorded at T1, while the ranges between observers increased for both crops at time 2 (Fig. 5).

In general, observer type could be discerned for all cover categories (Appendix C). Buckwheat and soil cover values estimated from the two human observers were generally closer compared to the cover values obtained using the AI-tool. Compared to the program, the human eye seemed to over-estimate soil cover (66 % and 67 % vs 62 % of the AI-tool) and under-estimate buckwheat cover (22 % and 24 % vs 28 % of the AI-tool). In the case of weed cover, the expert observer and the program were more closely aligned compared to the novice. The greatest similarities were found for lentil cover, averaged across cropping systems and dates, which ranged between 14.5 % and 17 %. The human eye completed the evaluation of a sampling quadrat in 1–2 min, with the expert observer taking less time to complete the task compared to the novice. Manually annotating the images with the program implied a minimum time investment of 5 min, up to 20 min. However, the annotator became more efficient as they used the application, reducing annotation times.

The weed cover model was improved by the inclusion of the interaction between observer and sampling date (Fig. 6A) and between observer and cropping system (Fig. 6B). The range of values 15 days after crop emergence (at lentil third true leaf stage) was narrow, resulting in similar estimates between observer types. Five weeks after emergence (at lentil branching stage), the novice observer seemed to over-estimate the weed cover (15.3 % ± 0.49) compared to both the expert observer (7.2 % ± 0.43) and the AI-tool (4.9 % ± 0.44).

The novice observer gave generally higher values compared to both the expert observer (+ 90 %) and the AI-tool (+150 %). The expert observer and the program gave comparable values, and the biggest discrepancies were observed in the MIX50 plots (AI-tool: 1.3 % ± 0.9; expert: 2.7 % ± 0.4) and in the pure buckwheat plots (AI-tool: 1.8 % ± 0.6; expert: 3.4 % ± 0.5) in which the values given by the expert were double compared to the AI-tool.

The buckwheat, lentil and soil cover models were improved when the three-way interaction between sampling date, cropping system and observer was included (Fig. D.1- Appendix).

At 15 days after crop emergence (third leaf stage of lentil) the observation ranges were comparable regarding crop cover values, but the greatest discrepancies were found in the pure plots. The expert and novice observer gave, respectively, 36 % and 58 % higher lentil cover values (pure plots) compared to the program, whereas the novice observer seemed to under-estimate buckwheat cover (19.1 % ± 1.2) in the pure buckwheat plots by 35 % compared to the expert (25.9 % ± 1.2) and by 45 % compared to the AI-tool (27.8 % ± 1.2).

Regarding bare ground cover, observers were aligned, except for a slight over-estimation of bare soil from the novice in the pure buckwheat plots (79.9 % ± 0.8) compared to the AI-tool (73.2 % ± 0.8) and the expert (70.7 % ± 0.9) (Fig. D.1C).

At lentil branching, 5 weeks after crop emergence, greater variability between observers was detected. Observers were most aligned in the ALT plots in estimating both crops' cover, whereas the MIX plots displayed the most discrepancies. The human eye was inclined to under-estimate buckwheat cover and over-estimate lentil cover in the MIX50 plots compared to the program. This can be seen to a lesser extent also in the pure buckwheat plots, in which both humans gave lower buckwheat cover values compared to the AI-tool. The greatest variability was observed in the MIX25 (for both crops) (Fig. D.1A) and pure lentil plots in which the novice underestimated crop cover in the field, but values were more aligned with the expert after using the program (Fig. D.1B).

Regarding bare soil cover, the highest variability between observers was found in the intercropping plots (Fig. D.1C), where the values obtained with the program were generally lower compared to the ones estimated by the human eye. In the pure buckwheat plots the expert observer seemed to over-estimate ground cover (52.5 % ± 0.9) compared to both the novice (43.7 % ± 0.9) and the AI-tool (42.1 % ± 0.8).

Annotating an image using the AI tool took between 5 and 22 min in T1 and between 7 and 22 min in T2. In T1 the user experienced a learning effect (Fig. 7A), taking less time to annotate the images as they learned how to use the tool and going from an average of 16 min to 10 min. The same is true for the images annotated during T2. The annotation time was initially of 18 min on average per image, possibly because of the long time passed between the previous use of the tool (over three weeks) and because of the different appearance of the crops. As the novice continued to use the tool, the annotation time decreased to 14 min per image on average, confirming that annotation time can be reduced with practice.

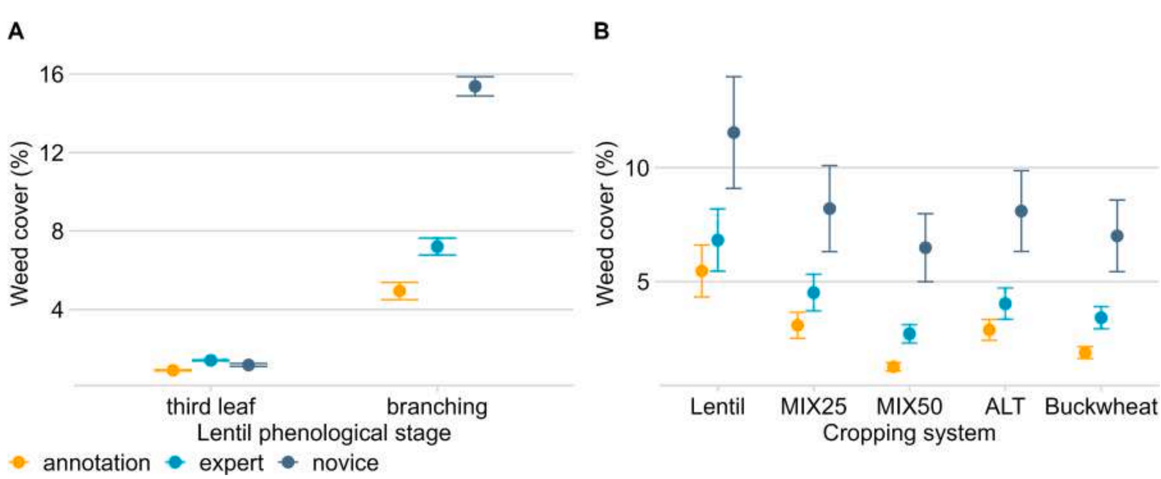

Fig. 6. Effect of observer in estimating weed cover in each A) lentil growth stage and B) cropping system.

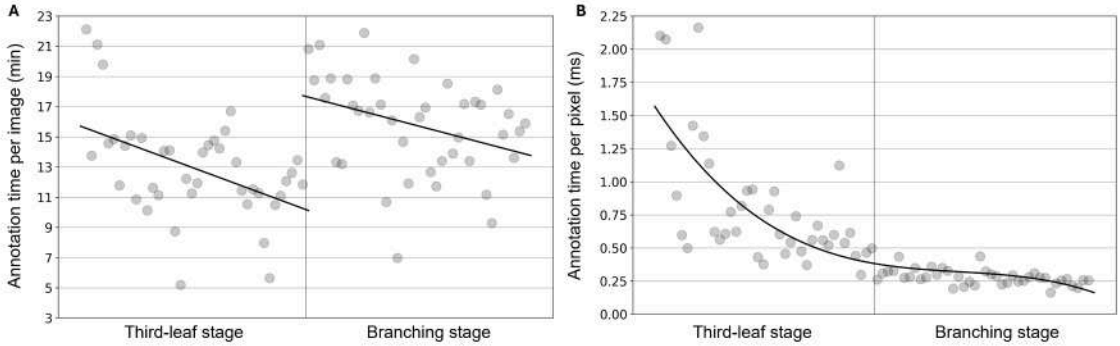

Fig. 7. A) Annotation times per image in minutes (min) using the AI tool. B) Annotation times per pixel over the final annotated area in milliseconds (ms) using the AI tool.

The ratio between the annotation time and the number of annotated pixels decreased sharply as the user gained more annotation experience (Fig. 7B), going from 2.25 ms/pixel to 0.25 ms/pixel in T1, showing that the user became faster at annotating the images as a consequence of the improved efficiency in annotating large areas of pixels. This is even more evident in T2, where the annotation speed started at 0.25 ms/pixel and reached up to 0.15 ms/pixel. Therefore, although images collected in T2 took more time to annotate (Fig. 7A), the annotation process was more efficient compared to T1 (Fig. 7B).

4. Discussion

The present study aimed to investigate the sources of inter-observer error in vegetation surveys and whether the error linked to experience could be reduced using an AI-based annotation tool. The main hypotheses were that: 1) observer error would increase with higher vegetation complexity, and 2) the benefits of using the AI-tool would be greater at lower vegetation cover, due to absence of stratification.

Overall, the observation gap between the expert and novice observer was reduced with respect to weed cover (−44 %), while the gap slightly increased concerning crop and soil cover. Morrison et al. (2020) report greater inter-observer error in small plots (≤ 1 m^2), compared to large plots (10 m^2) and found that species were assigned to different cover classes 26.2 % of the time. The present study used continuous values, which may have helped reduced the error related to the size of the surveyed area. The average value ranges were limited: 14 % - 17 % for lentil, 22 % - 28 % for buckwheat, 3 % - 8 % for weeds, and 62 % - 67 % for soil.

Analysing the annotations performed by the novice observer showed that about 80 % of the pixels were annotated automatically by the AI component (89.3 % in T1, 70.2 % in T2), and only about 20 % were annotated manually by the novice (10.7 % in T1, 29.8 % in T2). An example of this is shown in Fig. 8, representing a sample from T2. The original image is shown on the left, with the trace of the annotation process on the right. In the latter, black pixels are not annotated (ground, rocks, residues), grey pixels are AI-annotated, and white pixels are human-annotated. During the second data collection period, there are several major changes in crops which may create confusion to the AI component, requiring increased human intervention. First, as the crops grow, they shoot branches, which may seem unrelated to the plant, as in the upper part of the picture. Also, as the leaves grow large and cover more pixels then during the first annotation period, pixel with similar colors are close although coming from different plants. This is especially the case of buckwheat, which grows large leaves that cover most of the image and, due to this, the AI component tends to have smaller confidence in annotating the borders (e.g. especially on the right side of the image). Moreover, some plants may have flowers, as in the case of buckwheat, and the AI component may fail at recognizing them, mistaking them for rocks. In all these cases, the annotator decided to manually intervene to fix the suggested annotations, as can be seen in the trace (Fig. 8, right).

The benefit of using the AI-tool was notable when estimating higher weed covers, across all cropping systems, as it reduced the differences between the expert and novice by 67 %. Andujar et al. (2010) reported that observers had the tendency to overestimate weed cover at low weed densities and to underestimate values at high densities but found no significant differences between the estimations performed by four different observers. The present findings do not support this statement, since the novice observer over-estimated weed cover at the higher cover. Andujar et al. [9] referred to “true” weed cover values, which were not

Fig. 8. A sample from annotation time 2. Left: Clean image. Right: Trace of the annotation process (black: no annotation, grey: AI annotation, white: human annotation).

used in the present study, and compared equally experienced observers.

The present findings support those of Morrison [14], who states that high coefficients of variation (CV) (>100 %) are common at low vegetation covers, while the coefficient of variation is reduced (25–50 %) at high vegetation covers. At low weed cover, in-field training successfully reduced inter-observer error; even so, the AI-tool reduced the coefficient of variation related to the novice's observations by ~30 %. The coefficients of variation were more similar between observers in the second sampling date, though again the precision of the estimations was improved after using the AI-tool. The present findings related to weed cover allow to conclude that it may be worth dedicating 10–15 min to annotate an image since it leads to improved observation precision and related CV even at low vegetation covers, thus increasing the validity of cover estimation data.

Based on these findings, the authors expected crop and soil cover to also be associated with high coefficients of variation; on the contrary, the values were in the order of 25–50 % even at low covers. Buckwheat cover values were in line with actual planting densities, while this was not the case for lentil despite having a constant planting density. Before using the program, the novice underestimated buckwheat cover by 45 % in the sole plots (compared to the expert), and consequently overestimated soil cover (confirming [9]). Soil cover seemed to be precisely (CV=5 %) overestimated in 3 out of 5 cases, but the presence of lentil in the frame seemed to help reduce inter-observer error for both buckwheat and soil. Although no true cover values were used in this study, the alignment between expert and tool provides reliable evidence on whether the novice over- or under-estimated cover values. The tool proved useful to decrease inter-observer error for buckwheat and soil cover in the early stages of the cropping season.

In the second survey, buckwheat cover values calculated with the AI-tool were generally higher compared to the field observations from the novice. The highest discrepancies with the expert were found in the MIX50 (37 %) and in the pure buckwheat (27 %) plots, the lowest inter-observer error was recorded in the alternate row plots while the tool proved useful in reducing inter-observer error in the MIX25 plots. Buckwheat is an erect plant and was almost 1 m tall at the time of sampling: its height may have been an issue for observers in the field, leading to more conservative cover estimations. More conservative cover values were given in the MIX arrangement compared to the ALT arrangement, at the same buckwheat densities, but this difference was less pronounced in the field observations compared to the tool, leading to lower intra-personal error. To the best of the authors' knowledge there are no studies on the effect of eye-vegetation distance or of specific plant height on the precision of cover estimation.

Lentil cover was subject to both higher inter- and intra-observer error already at low vegetation cover. The expert provided values in between those of the novice and the tool, except in the MIX25 layout (25 % higher). Although inter-observer error was reduced in both the pure lentil and MIX50 plots, the cover values provided by the tool were still 36 % and 38 % lower compared to the expert, respectively. The greatest reduction in inter-observer error was observed in the alternate row plots, in which the differences between expert and novice went from 20 % to 6 % after using the tool.

At high vegetation cover, the expert observer had the lowest coefficient of variation between cropping systems. Although the plots seem to display high inter-personal error, the differences in lentil cover between the AI-tool and the expert were limited in the pure lentil and ALT plots, while cover values provided by the program were much lower in the MIX plots. The variability regarding lentil cover estimation could be due to both vegetation stratification and to the amount of bare soil in the sampling frame. Both were confirmed by the alignment between expert and tool in the ALT plots. Lentil leaves are tiny, and large soil patches between plants may complicate isolating lentil leaves for the human eye, leading to the overestimation of cover at low covers in the pure lentil plots, confirming the findings from Andujar et al. [9]. Stratification is ignored in ecological vegetation surveys to provide a cover/density value for each species which is needed to calculate their weight in the vegetation [2]. Lentil was always covered to some degree by buckwheat already at the initial growing stages, leaving only the leaves in the first layer to be annotated in the 2D images processed by the program. To a lower degree, stratification posed some issues in-field as well, explaining the intra-observer error between cropping systems for both lentil and soil values. The inter-observer error related to soil values in the second sampling date is the most perplexing but explainable when considering the spatial arrangement of the plants, their height, the stratification and possibly the angle at which observations were carried out.

The authors acknowledge that the cover values given in the field and by the tool are calculated following different objectives, which represents a fundamental limitation. Despite this, the intra-observer error was constant amongst the three observer types, allowing for a possible correction of the estimated values. This aspect should be further investigated and may become an additional implementation of the AI-tool in reducing the error linked to vegetation complexity in vegetation surveys. A further limitation may be given by the image resolution. The photo quality depends both on the smartphone camera and on the user's attentiveness. This can lead to partially blurry photos, especially when capturing small plants. For the purposes of the present study, the camera quality was deemed sufficient, and a higher-quality camera would not have resolved the issue of plant stratification. The smallest plants, mainly weeds, were also consistently distinguishable from crops. For future work, especially if focused on identifying individual weed species, a higher image resolution could be beneficial for accurately classifying small individuals. Nevertheless, a novel aspect of this study lies in the decision to use devices that are inexpensive, widely accessible, and require no post-processing (e.g., corrections from drone imagery). A current and valid application of the tool is related to double-checking soil cover values or values of vegetation in the first layer of the image. If general vegetation cover is of interest, then the program as of now has a great potential of reducing inaccuracy of observations with a reasonable time investment (10–15 min). Moreover, it is worth underlining that the tool is currently not trained for the task at hand: specifically, while VegAnn integrates in-domain knowledge and is capable of separating vegetation and soil, the same does not hold for SAM. Although SAM Base effectively assists the user, a segmentation tool trained with more specific knowledge (e.g. segmentation masks per species) likely represents a further step in making the annotation process more efficient, reducing the annotation times and enabling quicker cover estimates. Future developments of the program include the training of segmentation models for automatic annotation, which the authors are already working on with promising results.

5. Conclusions

The present study investigated the effect of observer experience and vegetation complexity on inter-and intra-observer error and whether inter-observer error could be reduced with an AI-tool, developed *ad hoc* for this study. The tool was able to effectively correct data from the novice observer, who had the highest intra-personal error. At low vegetation cover the tool reliably improved estimation of all categories, while further tests are needed to verify its performance at high vegetation covers. Overall, these findings provide a promising foundation for more precise and standardized vegetation mapping in both ecological and precision agriculture contexts. Although the current version of the tool still requires a time investment that may not be justified for expert observers, it already represents a valuable aid for reducing variability among less-experienced users. With further development, particularly through automation and the expansion of the image database, the tool could ultimately replace in situ visual assessments, requiring operators only to capture field photographs. Future work will focus on training the tool to recognize individual arable weed species and on integrating hyperspectral imagery to address species-specific identification challenges that cannot be resolved through standard 2D imagery (e.g., *Sorghum halepense* and *Echinochloa crus-galli*).

Funding

This study was carried out within the Agritech National Research Center and received funding from the European Union Next-Generation EU (PIANO NAZIONALE DI RIPRESA E RESILIENZA (PNRR) – MISSIONE 4 COMPONENTE 2, INVESTIMENTO 1.4 – D.D. 1032 17/06/2022, CN00000022), in particular Task 4.2.1, CUP number G23C22001100007.

CRediT authorship contribution statement

Alessandra Virili: Writing – original draft, Investigation, Formal analysis, Data curation. **Alex Falcon:** Writing – original draft, Visualization, Software, Formal analysis, Data curation. **Beatrice Portelli:** Writing – original draft, Visualization, Software, Formal analysis, Data curation. **Alessandro Peressotti:** Writing – review & editing, Funding acquisition. **Giuseppe Serra:** Writing – review & editing, Supervision, Methodology, Conceptualization. **Elisa Marraccini:** Writing – review & editing, Supervision, Methodology, Conceptualization.

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

The authors thank the bachelor student who patiently annotated all the images used for the analyses.

Appendices

A. Typical explanation of visual cover estimations in the field

Looking at the sampling frame (which represents 100 % cover), we practice dividing it in two (50 %), then in four (25 %) smaller areas. The 25 % and 10 % reference categories are commonly used thresholds in field-based vegetation surveys and provide a practical visual scale that balances estimation accuracy with observer consistency. These percentages are large enough to be visually meaningful and comparable among observers, while still allowing finer differentiation when estimating the cover of dominant or subdominant species. We know that 10 % of the specific frame (70 cm x 50 cm) is equivalent to either a 7 cm width (about four fingers) by the total height of the frame or to about a 15 cm x 20 cm area (corresponding to an open hand, palm down, with the thumb pointed out). Following the same logic, we know that 1 % of that specific frame is 5 cm x 7 cm. The smallest possible value we can reasonably estimate is 0.01 %, about 0.5 cm by 0.7 cm (~1 cm) or the size of a fingernail. We then practice assigning the vegetation to different cover categories: for example, if a species has a high coverage, we know we can use areas >10 % and even >25 % without having to deal with picturing tiny "pixels" in our frame. If we encounter 1 or few small individuals, we can gauge the cover using our fingernail as a reference and multiply that number (0.01 %) by the number of individuals that we see. We practice "moving" the vegetation around in our frame so it is all clustered in an area to which we can give a cover value. Though this may seem difficult, it is easier to mentally shift the vegetation prior to giving a value rather than adding up single values throughout the sampling area which also requires moving our eyes more frequently and potentially getting distracted or fatigued.

B. AI model size and granularity effect on precision and recall

SAM Base with fine-grained masks (Base-High: Table B.1, row Base, column High) is the best in terms of P across both times (T1 and T2). SAM Large-High is the second-best combination, with a small decrease in performance. The high precision (>90 % in most cases) means that the masks suggested by these models contain for the most part correct pixels, i.e. when annotating for a specific class, they do not include pixels belonging to other classes. This is advantageous when using the annotation tool, as annotators do not need to use the manual brush tools to remove unwanted pixels. The usage of fine-grained masks (High) leads to higher P, and the difference in performance with coarse masks (Low) is especially striking for the classes with smaller leaves (Weeds and Lentil).

The result on P is also explained by the Recall values (Table B.2). In fact, using fine-grained masks (High) leads to much lower R than variants using coarser masks (Low), on both annotation times. For instance, for ALL, the three High variants achieve 41.9–50.9 % R in T1 and 61.3–66.4 % R in T2. This is significantly lower than the Low granularity variants, which reach 88.8–89.3 % R in T1 and 95.4–96.5 % in T2. This means that, although SAM Base and Large using fine-grained masks are quite precise in selecting correct pixels, they do so at the expense of annotating fewer pixels. Although this means fewer manual brush strokes for the annotators to remove unwanted pixels, it may require additional clicks or manual annotations for them to cover all the desired pixels for the class.

SAM Huge records the highest Recall values for Buckwheat and Lentil at T1, and the second highest R values for all classes at T2. However, the increase in model size, memory requirements, and computational resources needed to run SAM Huge, does not justify its use instead of SAM Large or SAM Base.

Table B.1
Comparison of nine combinations for the SAM component of our annotation tool. Three SAM variants (Base, Large, Huge) with three granularities (Low, Mid, High) are tested on both annotation times (T1 and T2). Performance is reported using Precision both in aggregated form (ALL) and per class (Weeds, Buckwheat, Lentil). The best results for each time and class are bolded and underlined, the second-best results are bolded.

P		ALL			Weeds			Buckwheat			Lentil		
		Low	Mid	High	Low	Mid	High	Low	Mid	High	Low	Mid	High
T1	Base	97.3	98.7	<u>**99.2**</u>	61.8	83.5	<u>**88.1**</u>	86.0	96.0	<u>**98.1**</u>	68.2	83.3	<u>**91.1**</u>
	Large	97.4	98.2	**98.9**	67.4	84.4	**85.0**	85.8	93.8	**97.8**	65.8	74.9	**87.2**
	Huge	97.3	98.1	98.8	69.4	83.5	83.9	87.6	94.1	97.2	69.3	75.3	86.7
T2	Base	96.4	98.6	<u>**99.0**</u>	65.3	88.5	<u>**94.4**</u>	83.7	97.0	<u>**98.0**</u>	74.4	88.4	<u>**94.5**</u>
	Large	96.8	98.4	**98.9**	74.8	87.5	**93.9**	83.6	96.4	**97.7**	68.2	78.5	**93.3**
	Huge	96.5	98.3	**98.9**	76.5	89.6	92.8	83.0	96.0	**97.7**	70.4	75.5	92.8

Table B.2
Comparison of nine combinations for the SAM component of our annotation tool. Three SAM variants (Base, Large, Huge) with three granularities (Low, Mid, High) are tested on both annotation times (T1 and T2). Performance is reported using Recall both in aggregated form (ALL) and per class (Weeds, Buckwheat, Lentil). The best results for each time and class are bolded and underlined, the second-best results are bolded.

R		**ALL**			**Weeds**			**Buckwheat**			**Lentil**		
		Low	**Mid**	**High**	**Low**	**Mid**	**High**	**Low**	**Mid**	**High**	**Low**	**Mid**	**High**
T1	**Base**	**89.3**	66.2	41.9	25.2	29.4	17.5	**81.6**	63.4	42.3	62.8	61.6	38.0
	Large	**89.1**	78.3	48.1	27.8	**30.8**	20.3	80.0	73.8	46.0	60.0	**65.5**	47.3
	Huge	88.8	77.1	50.9	28.7	**30.6**	22.1	**83.4**	71.8	47.1	61.9	**66.6**	51.5
T2	**Base**	**96.5**	80.4	61.3	23.5	45.9	31.0	**86.4**	72.3	56.8	76.5	**85.4**	64.4
	Large	95.4	82.6	65.2	25.6	**50.0**	34.6	77.2	69.0	60.0	74.6	84.3	68.8
	Huge	**95.9**	83.2	66.4	28.6	**47.5**	35.7	**78.9**	67.2	60.1	75.8	**85.0**	71.8

C. Overall observer effect

Fig. C.1 shows the overall effect of observer types on the cover estimation of soil, buckwheat, lentil, and weed. In general, observer type can be discerned for all cover categories and the values estimated by the two human observers are close.

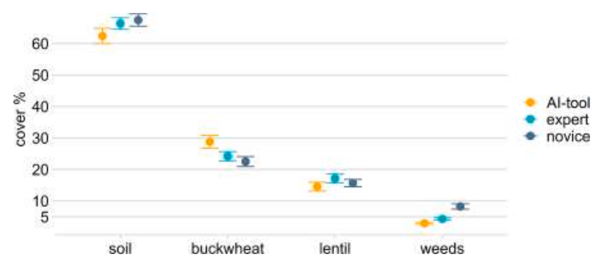

Fig. C.1. Overall effect of observer type on soil, buckwheat, lentil and weed cover (%).

D. Effect of observer type on the estimation of cover type averaged across cropping system and growth stage

Fig. D.1 provides a more in-depth analysis of the effect of the cover type, cropping system, and growth stage on cover estimation.

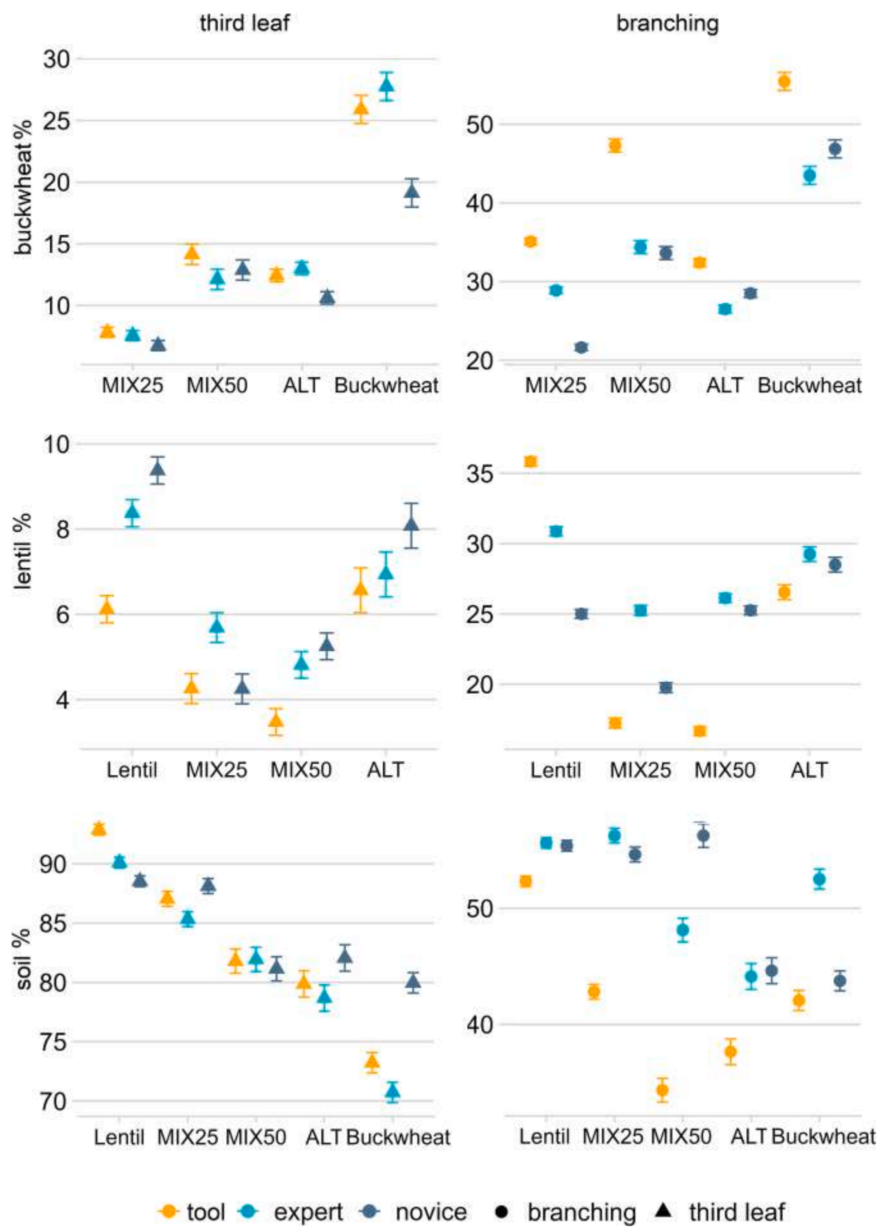

Fig. D.1. Interaction effect of observer, sampling date, and cropping system on A) buckwheat cover (%), B) lentil cover (%) and C) soil cover (%).

Data availability

Data will be made available on request.

References

[1] M.D.K. Owen, Diverse approaches to herbicide-resistant weed management, Weed Sci. 64 (2016) 570–584, https://doi.org/10.1614/WS-D-15-00117.1.
[2] S. Gaba, X. Reboud, G. Fried, Agroecology and conservation of weed diversity in agricultural lands, Bot. Lett. 163 (2016) 351–354, https://doi.org/10.1080/23818107.2016.1236290.
[3] M. Riemens, M. Sønderskov, A.-C. Moonen, J. Storkey, P. Kudsk, An integrated weed Management Framework: a Pan-European perspective, Eur. J. Agron. 133 (2022) 126443, https://doi.org/10.1016/j.eja.2021.126443.
[4] P. Bàrberi, G. Bocci, S. Carlesi, L. Armengot, J.M. Blanco-Moreno, F.X. Sans, Linking species traits to agroecosystem services: a functional analysis of weed communities, Weed Res. 58 (2018) 76–88, https://doi.org/10.1111/wre.12283.
[5] G. Adeux, E. Vieren, S. Carlesi, P. Bàrberi, N. Munier-Jolain, S. Cordeau, Mitigating crop yield losses through weed diversity, Nat. Sustain. 2 (2019) 1018–1026, https://doi.org/10.1038/s41893-019-0415-y.
[6] E. Kazakou, G. Fried, J. Richarte, O. Gimenez, C. Violle, A. Metay, A plant trait-based response-and-effect framework to assess vineyard inter-row soil management, Bot. Lett. 163 (2016) 373–388, https://doi.org/10.1080/23818107.2016.1232205.
[7] C.D. Bonham, Measurements for Terrestial Vegetation, John Wiley & Sons, 2013.
[8] R.L. MacDonald, J.M. Burke, H.Y.H. Chen, E.E. Prepas, Relationship between aboveground biomass and percent cover of ground vegetation in Canadian boreal plain riparian forests, For. Sci. 58 (2012) 47–53, https://doi.org/10.5849/forsci.10-129.
[9] D. Andújar, A. Ribeiro, R. Carmona, C. Fernández-Quintanilla, J. Dorado, An assessment of the accuracy and consistency of Human perception of weed cover: human perception of weed cover, Weed Res 50 (2010) 638–647, https://doi.org/10.1111/j.1365-3180.2010.00809.x.
[10] L.J. Wiles, Software to quantify and map vegetative cover in fallow fields for weed management decisions, Comput. Electron. Agric. 78 (2011) 106–115, https://doi.org/10.1016/j.compag.2011.06.008.
[11] M. Röttgermann, T. Steinlein, W. Beyschlag, H. Dietz, Linear relationships between aboveground biomass and plant cover in low open herbaceous vegetation, J. Veg. Sci. 11 (2000) 145–148, https://doi.org/10.2307/3236786.
[12] R. Gerhards, K. Bezhin, H.-J. Santel, Sugar beet yield loss predicted by relative weed cover, weed biomass and weed density, Plant Prot. Sci. 53 (2017) 118–125, https://doi.org/10.17221/57/2016-PPS.

[13] M.N. Dethier, E.S. Graham, S. Cohen, L.M. Tear, Visual versus random-point percent cover estimations: "objective" is not always better, Mar. Ecol. Prog. Ser. 96 (1993) 93–100.
[14] L.W. Morrison, Observer error in vegetation surveys: a review, J. Plant Ecol. 9 (2016) 367–379, https://doi.org/10.1093/jpe/rtv077.
[15] J.L. Moore, C.E. Hauser, J.L. Bear, N.S.G. Williams, M.A. McCarthy, Estimating detection–Effort curves for plants using search experiments, Ecol. Appl. 21 (2011) 601–607, https://doi.org/10.1890/10-0590.1.
[16] L. Klimeš, M. Dančak, M. Hájek, I. Jongepierová, T. Kučera, Scale-dependent biases in species counts in a grassland, J. Veg. Sci. 12 (2001) 699–704, https://doi.org/10.2307/3236910.
[17] S. Burg, C. Rixen, V. Stöckli, S. Wipf, Observation bias and its causes in botanical surveys on high-alpine summits, J. Veg. Sci. 26 (2015) 191–200, https://doi.org/10.1111/jvs.12211.
[18] E.V. Lukina, M.L. Stone, W.R. Raun, Estimating vegetation coverage in wheat using digital images, J. Plant Nutr. 22 (1999) 341–350, https://doi.org/10.1080/01904169909365631.
[19] C. Macfarlane, G.N. Ogden, Automated estimation of foliage cover in forest understorey from Digital Nadir Images, Methods Ecol. Evol. 3 (2012) 405–415, https://doi.org/10.1111/j.2041-210X.2011.00151.x.
[20] J.W. Karl, S.E. McCord, B.C. Hadley, A comparison of cover calculation techniques for relating point-intercept vegetation sampling to remote sensing imagery, Ecol. Indic. 73 (2017) 156–165, https://doi.org/10.1016/j.ecolind.2016.09.034.
[21] N. Nikolić, D. Rizzo, E. Marraccini, A. Ayerdi Gotor, P. Mattivi, P. Saulet, A. Persichetti, R. Masin, Site- and time-specific early weed control is able to reduce herbicide use in Maize - a case study, Ital. J. Agron. 16 (2021) 1780, https://doi.org/10.4081/ija.2021.1780.
[22] A. Kirillov, E. Mintun, N. Ravi, H. Mao, C. Rolland, L. Gustafson, T. Xiao, S. Whitehead, A.C. Berg, W.Y. Lo, et al., Segment anything, in: Proceedings of the Proceedings of the IEEE/CVF International Conference on Computer Vision, 2023.
[23] S. Madec, K. Irfan, K. Velumani, F. Baret, E. David, G. Daubige, L.B. Samatan, M. Serouart, D. Smith, C. James, et al., VegAnn, vegetation annotation of multi-crop RGB images acquired under diverse conditions for segmentation, Sci. Data 10 (2023) 302, https://doi.org/10.1038/s41597-023-02098-y.
[24] R Core Team R: a language and environment for statistical computing. 2021.
[25] D.R. Anderson, K.P. Burnham, Avoiding pitfalls when using information-theoretic methods, J. Wildl. Manag. 66 (2002) 912–918, https://doi.org/10.2307/3803155.
[26] J.B. Johnson, K.S. Omland, Model selection in ecology and evolution, Trends Ecol. Evol. 19 (2004) 101–108, https://doi.org/10.1016/j.tree.2003.10.013.
[27] A. Gałecki, T. Burzykowski, Linear mixed-effects model. *Linear Mixed-Effects Models Using R*; Springer Texts in Statistics, Springer New York, New York, NY, 2013, pp. 245–273. ISBN 978-1-4614-3899-1.
[28] A.F. Zuur, E.N. Ieno, N. Walker, A.A. Saveliev, G.M. Smith, *Mixed Effects Models and Extensions in Ecology with R*; Statistics for Biology and Health (2009). ISBN 978-0-387-87457-9.
[29] A. Gelman, Multilevel (Hierarchical) modeling: what it can and cannot do, Technometrics 48 (2006) 432–435, https://doi.org/10.1198/004017005000000661.
[30] A. Virili, A.-C. Moonen, Reduced weeding shows potential to regulate nutrient leaching in a cabbage (Brassica oleracea, var. Capitata) lysimeter trial, Agric. Ecosyst. Environ. 367 (2024) 108987, https://doi.org/10.1016/j.agee.2024.108987.