Computers and Electronics in Agriculture 197 (2022) 106957

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

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

Honey bee counter evaluation – Introducing a novel protocol for measuring daily loss accuracy

Parzival Borlinghaus[a,*], Richard Odemer[b], Frederic Tausch[c], Katharina Schmidt[c], Oliver Grothe[a]

[a] Institute for Operations Research, Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany
[b] Julius Kühn-Institut (JKI) – Federal Research Centre for Cultivated Plants, Institute for Bee Protection, Braunschweig, Germany
[c] apic.ai GmbH, Karlsruhe, Germany

ARTICLE INFO

Keywords:
Robbers test
Automated bee counting device
Regulatory risk assessment methodology
Harmonized validation protocol
Precision beekeeping

ABSTRACT

Automated bee counters advanced over the last hundred years and became increasingly diverse. However, to date, there is no method for standardized validation of counting accuracy and thus no reliable data on daily bee losses, or background mortality in colonies. However, such data are in urgent need by regulators to establish future guidelines for pesticide risk assessment. In this work, existing approaches were combined to form a novel protocol for validating bee counters. In a case study, we demonstrated that the protocol is sufficiently feasible to determine the measurement accuracy of a commercial counting system. Measurement accuracy was modeled by the difficulty of specific measurement conditions. Daily loss, i.e., the difference between incoming and outgoing bees, can be used to assess colony health, environmental impacts, and infer the effect of pesticides on bee colonies. The developed protocol makes innovations in this field measurable and creates a foundation for the benchmarking of different types of bee counting systems. We discuss how it can be utilized in an effort to move the sector forward in the future.

1. Introduction

Bee pollinators are not only of great importance for ecosystems (Ollerton et al., 2011) but also make a major economic contribution to us through their ecosystem services (Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services (IPBES), 2016). This pollinator group, however, has been suffering from a decline for quite some time, affecting already 37% of European bee species (Biesmeijer et al., 2006; National Research Council, 2007; European Commission et al., 2016) and, posing a threat to global food security and nutrition. Honey bees have not been affected by this decline, and the number of bee colonies worldwide has almost doubled since they were recorded in the 1960s (FAOSTAT, 2021). Regardless, they represent an essential part of pollination ecology and, in part, have become an important model organism for investigating the drivers of bee decline (National Research Council, 2007).

Because pesticides are suspected of being one of these drivers, the European Food Safety Authority (EFSA) was tasked with developing more stringent guidelines for the risk assessment of pesticides (EFSA Panel on Plant Protection Products and their Residues (PPR), 2012; European Food Safety Authority, 2013). Since "[t]he viability of each colony, the pollination services it provides, and its yield of hive products all depend on the colony's strength and, in particular, on the number of individuals it contains" (European Food Safety Authority, 2013), EFSA has formulated specific protection goals (SPGs) aligned to colony size and mortality. Thus, characteristics worth protecting can be collectively assessed by evaluating daily bee losses. However, high-quality data on background mortality, i.e. the natural daily loss of (forager) bees are scarce.

To monitor even small changes in bee mortality, tools are needed that can accurately count bees entering and leaving the hive. For the vast majority of electronic bee counters, determining daily losses is, thus, of primary interest (Odemer, 2021). Since foragers usually fly only during the day and return to the hive at dawn, observation intervals of 24 h are well suited and suggested by EFSA as a reference (European Food Safety Authority, 2013).

The ability to accurately count bees is not only important from a regulatory perspective, but also relevant to scientists, beekeepers and

* Corresponding author.
E-mail address: parzival.borlinghaus@kit.edu (P. Borlinghaus).

https://doi.org/10.1016/j.compag.2022.106957
Received 20 February 2022; Received in revised form 6 April 2022; Accepted 9 April 2022

other ecotoxicological issues. Beekeepers are provided, for example, with the means to evaluate the quality of sites, to observe swarm behaviour and colony development. If abnormalities occur, counting data helps to find the underlying causes. Pham-Delégue et al. (2002) call for sublethal effects, such as sensory impairment and associated effects on foraging behavior, to be routinely recorded in addition to lethal effects of pesticides and considered in risk assessment, e.g., through the use of newly developed bee counting systems (see Chmiel et al. (2020) on sublethal effects). Schuhmann et al. (2022) point out that ecotoxicological studies of individual plant protection products in the laboratory can be rather artificial, as field studies often involve mixtures of pesticides and their interactions are not well understood. However, field studies take place under more realistic conditions and lethal effects can be accurately determined with bee counters. Bermig et al. (2020) consider the determination of activity and background mortality in a control population a good way to quantify both sublethal and long-term effects. The EFSA guidance document contains more than a hundred, rather unreliable, publications on background mortality, many of which could have been greatly simplified by the use of automated bee counters (European Food Safety Authority, 2013). Accurate determination of bee losses thus goes hand in hand with accurate determination of bee activity, which also has wide-ranging applications. For example, Danka and Beaman (2007) used electric bee counters to compare the pollination performance of different bee species based on the number of flights recorded.

Over the last hundred years, several technical devices have been tested on beehives, which today form the basis of precision beekeeping (Zacepins et al., 2012). Scales, thermometers, hygrometers, anemometers, microphones, radar, photoelectric sensors, capacitive sensors, and RFID transmitters have been used to monitor bee health (for a description of some methods, see Zacepins et al. (2015), Marchal et al. (2020), Odemer (2021)). Within precision beekeeping, these technologies are divided into three categories: (1) Sensors that collect data at the apiary level, (2) at the colony level and (3) at the level of single individuals (Zacepins et al., 2015). Bumanis et al. (2020) describe how, even with sensors of different categories, multi-modal data from bee-related sources can be merged through data fusion.

However, bee counters, which belong to category (3), form only a small part of the sensors found on hives. Two of the first devices registered counts based on the weight of the bees, however, technical development allowed the implementation of novel methods. From (infrared) photoelectric sensors to capacitive sensors and camera-based methods, many ideas have emerged. In recent years, the trend has been towards camera-based methods (Odemer, 2021). The reasons are lower maintenance, greater potential, and the possibility of less invasive data collection. Camera-based methods do not require special tunnels that could get dirty or blocked. Further potential lies in the additional classification of tracked objects (e.g., workers, drones, and intruders) (Marstaller et al., 2019), classification of tracked behavior (e.g., thermoregulation and guarding) (Kridi et al., 2016), and detection of parasites (Schurischuster et al., 2018) and corbicular pollen loads (Sledevič, 2018; Marstaller et al., 2019). A comprehensive historical overview of the development of automatic bee counters was presented by Odemer (2021).

Despite a large number of counting systems, Odemer (2021) criticizes the lack of standardized methods for determining the accuracy of bee counters. His review of 38 different bee counters shows that 63% of them were not validated at all, while of those that included information on precision, 17% did not include information on the methodology used for validation. Without proper evaluation of counting systems, comparison of different counters is not possible and interpretation of counting results is limited. Therefore, Odemer calls for a validation standard for bee counters (Odemer, 2021).

This paper is intended to make a scientific contribution in two ways. It (1) reviews previous assessment approaches, highlights the difficulties of meaningful assessment, and derives a novel protocol for determining the accuracy of daily losses. (2) A commercial video-based bee counter is used to test the protocol for practicality in a case study. The protocol combines three existing assessment methods found in the literature which are subsequently presented.

2. Literature review on bee counter evaluation methods

As an example of why validation is important, Fig. 1 illustrates three possible measurement results concerning the evaluation of bee counters. Fig. 1A shows two measurement curves (solid lines) of faulty bee counters. One gives only zero values, the other measures noise. Fig. 1B shows the trace of a systematically biased bee counter (solid line). In this case, the counter misses a portion of the returning bees. The systematic bias results in an unrealistically high daily loss indicated by the high number of bees that are supposedly still outside the hive at the end of the day. Fig. 1C shows the trace of a stochastically biased counting system (solid line). The counting system provides values that are sometimes too large and sometimes too small, with no preferred direction.

The evaluation of a bee counter aims at excluding malfunctioning counters (Fig. 1A), quantifying possible biases (Fig. 1B, C), and determining the accuracy of daily loss measurements. The procedure in cases (B) and (C) is not straightforward because the true 24-h bee count is neither known nor determinable: it is impossible to obtain the true bee count by manually counting the number of incoming and outgoing bees in real-time (Odemer, 2021). Bermig et al. (2020) noted that manual counting of bees in short video sequences was only possible at 0.3 times the speed and had to be done separately for each of the 24 entrance and

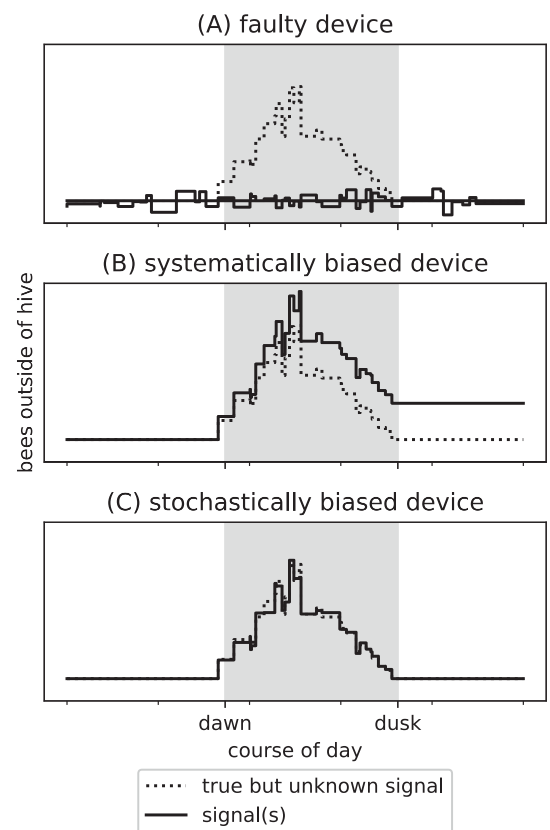

Fig. 1. Example illustrations of measurement time series' of malfunctioning bee counters. The true but unknown measurement series is illustrated by dotted lines. The grey area marks the period when flight activity is expected. (A) Measurement series of two faulty counting devices, one measuring zero-values, one measuring noise. (B) Measurement series of a systematically biased counting device. A certain percentage of incoming bees are missed, leading to high measured losses at the end of the day. (C) Measurement series of stochastically biased counting device.

exit tubes of the system. This indication is also consistent with the authors' experience that it requires about four hours to correctly capture all the bees in a one-minute video at a very high bee density and 40 frames per second. Meikle and Holst's assessment that "the use of human observation, while probably accurate, clearly limits the time in which the hive can be observed due to fatigue" (Meikle and Holst, 2015) is shared by many other researchers (Rosenquist, 2019; Tausch et al., 2020; Bermig et al., 2020).

Despite these difficulties, there are approaches to partial evaluation of bee counters.

2.1. Evaluation by observation

Evaluation by manual observation is the most widely used evaluation technique (Odemer, 2021) and has been used in numerous studies (Campbell et al., 2008; Jiang et al., 2016; Ngo et al., 2019; Bermig et al., 2020; Kulyukin et al., 2021). Experimenters note short periods, rarely longer than 3 min, and compare their observations with their measurements. Odemer (2021) describes various ways to make such comparisons. The annotation is time-consuming and the benefit is small. A brief example illustrates the problem. For 60-s samples, the smallest detectable error is one bee per minute. Therefore, a counting system is either error-free (to the best of our knowledge, no such system exists today) or has a sampling error of at least one bee per minute. At this point, there is no reason to consider this error as stochastic, but we must assume that the errors add up and do not level off. It follows that the error in a 24-h time interval could be (up to) 1440 bees. The EFSA (European Food Safety Authority (EFSA), 2020) states the general natural background mortality rate at 3.75% after reviewing the literature. To determine a 10% increase in mortality, it is necessary to distinguish the loss of 375 bees from 413 bees in a colony of 10,000 bees. Therefore, the resolution of such an evaluation method cannot meet the requirements. This problem does not arise with the robbers test described below because the need to annotate the data is conveniently avoided. However, the sample evaluation has a key advantage. Unlike the robbers test, a bad bee counter in a given sample will always give bad results, while a good counter will always give good results.

2.2. Robbers test evaluation

To evaluate the BeeSCAN bee counter reliably and realistically, Struye (2001) described a novel experimental setup, which the author calls the 'robbers test'. In this test, the counting device replaces the bottom of an empty hive containing only a food source (robber hive). This food attracts robbers from nearby colonies that can only reach the food by passing through the counter and vice versa. At the end of the day, the difference of incoming and outgoing bees must be zero as bees will abandon the food source before sunset (corrected for the number of (dead) bees remaining in the hive) (Struye, 2001; Bermig et al., 2020; Odemer, 2021). Although the underlying count events can be in the order of hundreds of thousands, the robbers test provides only one data point per day, which is the deviation from zero.

While it is true that a good bee counter should achieve a difference close to zero, the opposite conclusion is false. A poor counter may produce a good result (see Fig. 1A). An extreme example is a counter that gives only zero counts and thus achieves the best possible result in the robbers test. Beyond this limiting case, however, the robbers test also fails in more realistic situations. For example, a perfect counter is indistinguishable from a counter with arbitrarily strong symmetric error (see Fig. 1A, C). Another problem not named by Struye (2001) is the observation that the movement patterns of the bees change due to the setup of the robbers test. The bees move through the sensor passages more goal-oriented, and long dwell times in the passages are rather the exception (Rosenquist, 2019, p. 39). This behavioral variation tends to underestimate the error rate, as long dwell times generally complicate the counting problem (Liu et al., 1990). In addition, no activity is expected at the feeder in the darkness that would occur under normal conditions (Bermig et al., 2020; Odemer, 2021). Because recording in darkness is particularly difficult for camera-based bee counters, the robbers test for such systems may further underestimate the error.

2.3. Evaluation by literature comparison

Evaluation methods based on comparison with literature or expertise work differently. For example, Ngo et al. (2019) and Gonsior et al. (2020) evaluated their counters through ecotoxicological studies. For this purpose, a group of bee colonies was exposed to insecticides known to have negative effects on the flight behavior and homing ability of bees. The authors demonstrated that their counter was able to detect a significant group difference compared to untreated colonies and concluded that the bee counter was functional.

Other methods include the correlation of bee counts with temperature, humidity, and solar radiation (Rickli et al., 1989; Liu et al., 1990; Chen et al., 2015; Jiang et al., 2016; Ngo et al., 2019). These factors are known to affect flight activity, and therefore the observation of strong correlations demonstrates the principle operability of bee counters. Liu et al. (1990), Chen et al. (2015), Jiang et al. (2016) also compared the measured data with extreme weather events and found, for example, a decrease in flight activity associated with rain. Rickli et al. (1989) correlated the measured activity with different colony sizes. A positive correlation at this point is consistent with expectation and can again be interpreted as an indication of a functioning bee counter.

All methods that attempt to confirm correlations known from the literature help to identify faulty bee counters. However, they are not suitable for quantifying the accuracy of daily loss measurements.

2.4. Rationale for a new evaluation protocol

At the current state of understanding, two things are noteworthy. First, the challenge of determining daily loss in an automated manner with accuracy suitable for regulatory risk assessment has not yet been solved. Second, there is a lack of sufficient evaluation of existing systems (Rickli et al., 1989; Odemer, 2021). Marchal et al. (2020) attribute the difficulty of bee counting to the large number of similar insects that must be observed noninvasively under turbulent conditions.

A large number of counting systems, the variety of sensors used, and the need for an accurate bee counting system require a robust evaluation protocol to determine daily losses. None of the evaluation procedures found in the literature can rule out malfunctioning bee counters and quantify the remaining errors as daily loss uncertainties. Table 1 shows that the robbers test alone is not capable of detecting malfunctioning counters, as shown in Fig. 1A. However, the advantage of the test is that it reveals errors that build up over long periods (Fig. 1B). In particular, the conclusion that a good robbers test result indicates a well-working bee counter is incorrect. In contrast, evaluation by observation is a more accurate and reliable version of literature comparison. Both methods indicate the degree of correlation between measurements and

Table 1
Capabilities of different evaluation approaches. None of the methods is suitable for determining the uncertainties of daily loss measurements.

Capability/ Evaluation approach	Evaluation by observation	Robbers test	Literature comparison
Find malfunctioning bee counters by ensuring a minimum correlation between measurements and actual flight counts.	yes	no	to some extent
Distinguish between stochastic and systemic errors.	no	yes	to some extent
Quantify daily loss uncertainties.	no	to some extent	no

actual flight counts. However, their results are not directly transferable to the accuracy of daily losses.

3. Harmonised evaluation protocol for daily loss measurements of honey bee colonies

Various studies have shown that the accuracy of a bee counter depends to some extent on external factors. One important factor affecting the performance of all bee counters is the activity of the bees themselves. Counters that use tunnels for separation, for example, are unable to distinguish bees that are moving closely together (Liu et al., 1990; Bermig et al., 2020). Crowding of bees causes Spangler (1969) to recommend his bee counter only for very small colonies, and Struye (2001) to guarantee correct measurements only when bees are no closer than one millimeter. Such factors may vary from system to system. Campbell et al. (2008) report that shade, debris, and intense bee movement negatively affect the performance of their system. Jiang et al. (2016) reported reflections and strong 'nervosity' of the bees as additional sources of error.

To ensure that a counting system works well in all relevant situations (scenarios), it is useful to determine such performance factors. For each scenario, an evaluation by observation is proposed (see Section 2.1). If it has been determined that only small deviations occur for all scenarios, the next step is to check whether these small deviations add up over one day. For this purpose, several robbers tests are performed (see Section 2.2). A large deviation from the target value of the robbers test of zero indicates either a very large stochastic error or a (small) systematic bias of the counting system. However, larger errors could already be excluded due to the sufficiently good performance in the evaluation by observation. Any systematic biases need to be further investigated and corrected. To estimate the uncertainties for any day, it is necessary to relate the deviations of the robbers tests to the difficulty of the measurement conditions, i.e., the performance factors.

The difficulty of the measurement conditions results directly from the scenarios that comprise the day being studied. It is to be expected that a robbers test that consists to a large extent of difficult scenarios, e.g. due to flight-friendly weather, will lead to measurement errors more frequently than a rainy day that may be composed of particularly simple scenarios. By modeling the performance of the robbers test based on the measurement conditions, the expected measurement errors for a given day can be predicted based on their difficulty. The following nine-step procedure implements these ideas. This section is followed by the case study, which provides further guidance for practical implementation.

3.1. Evaluation step 1: Performance factors

An operator expert for the bee counter being evaluated determines all relevant factors that affect the performance of the counter. For practicality, it must be ensured that each performance factor can be determined robustly and automatically. This means that for non-camera-based bee counters, without further ado, there is only one performance factor at hand, the activity determined with the counter itself. Caution is required as in this case the counter itself is involved in the evaluation. Apart from this particularity, there are no differences for other bee counting approaches. Example: A large number of bees and infrared lighting have a negative effect on counting accuracy and represent the performance factors.

3.2. Evaluation step 2: Deriving the scenarios

All (reasonable) combinations of performance factors are created and corresponding scenarios are derived. Example: "Few bees move slowly under infrared lighting" is considered as a scenario.

3.3. Evaluation step 3: Scenario annotation

For each scenario, corresponding samples must be collected and manually labeled. These annotations are considered as ground truth in the next step. Example: For each scenario, 10, 20, or 30-s videos are annotated.

3.4. Evaluation step 4: Scenario evaluation-by-observation

All annotated videos are processed with the counting system and compared with the ground truth. However, the metric for a good match is open and a suitable metric must be found. The most intuitive metric for the performance of a bee counter is the deviation of the inputs and outputs from their true values. However, due to the short duration, some scenarios contain bees but no counting events. A metric based solely on counting events is bound to be poorly resolved, especially for short duration, and does not take advantage of the information underlying the count. It is advisable to use metrics that can handle all types of scenarios. For example, the Multiple Object Tracking Accuracy metric (Bernardin et al., 2006) is determined by the number of false-positive detections, false-negative detections, and id switches. Since many more data points are included, more accurate performance information is obtained. It is suitable for any video-based counting device and any scenario that contains at least one annotated object. Counters for which no information is available other than the number of counting events (e.g. counters that solely use photoelectric or capacitive sensors) must fall back on the deviation of ingoing and outgoing bees.

3.5. Evaluation step 5: Scenario difficulty rating

All scenarios are rated on a difficulty scale from zero to one. The scenario that achieved the best result in the previous step is assigned difficulty level zero, and the scenario with the worst result is assigned difficulty level one. All other scenarios are rated accordingly between the easiest and the most difficult scenario. When rating, the relative differences are maintained. This transformation serves to make the results easier to interpret and independent from the metric chosen in the previous step.

3.6. Evaluation step 6: Robbers tests

Robbers tests are carried out during several days and, if possible, with different colonies if a tunnel tent is used (Odemer, 2021). Since flight activity in a robbers test scenario is usually lower than in full-sized colonies, a reduction in bee traffic should be avoided where possible. This means that the food source in the robber hive should be highly attractive (e.g., honeycombs) and should last for at least several days to generate as much flight activity as possible. In addition, weather conditions must stimulate bees to forage, and bee density should ideally be high within 2 km of the robber hive. If the trials are conducted in a tunnel tent to prevent the spread of robbing to neighboring colonies or to prevent disease, the experiments should be conducted with strong colonies with more than 20,000 bees.

3.7. Evaluation step 7: Difficulty of robbers tests' measurement conditions

For each robbers test, the measurement conditions are evaluated. The difficulty of the measurement conditions of a day is the weighted difficulty of the scenarios that comprise the studied day. For this purpose, the time intervals of the day must be matched with the scenario that is most analogous to it.

3.8. Evaluation step 8: Modelling accuracy based on the difficulty of the measurement condition

The influence of the measurement conditions ('difficulty') on the performance of the bee counter is determined using an appropriate model class (e.g. linear or polynomial models). There should be a positive correlation between the difficulty of the test day and the accuracy of the bee counter.

3.9. Evaluation step 9 (optional): Plausibility check

It might be useful to additionally compare the measured loss with the loss expected from the literature or to show correlations with light intensity or temperature. Rosenquist wrote: "Long-term observations over several weeks under regular conditions do not provide accurate reference values. Nevertheless, they were very important for functional verification" (Rosenquist, 2019).

4. Case study

4.1. Materials & method

The range of commercial bee monitoring solutions capable of detecting daily losses is limited. US-based Keltronix Inc. offers Eyesonhives, a camera-based monitoring solution (Keltronix, 2022). However, as the system determines the level of flight activity in front of the hive without direction, it is unable to determine numbers for incoming and outgoing bees. Similarly, BeeScanning (2022), BeeAndme (2022), Arnia (2022) offer different tools and sensors, that also do not allow direct conlusions to be drawn about daily losses. BeeCheck, which was developed for the specific purpose of counting bees by the Federal Research Centre for Cultivated Plants (JKI) (2022), has not yet reached product status. Hiverize (2022) provides building instructions for monitoring bees, but these are limited to weather data and weight measurements. Of particular note is the BeeSCAN system, but development stopped more than 18 years ago (Lowland Electronics, 2022; Struye, 2001). In addition, Beehivemonitoring (2022) offers a non-camera-based module for counting bees.

However, the monitoring technology of the company apic.ai was chosen because it is a self-designed, state-of-the-art, and commercially available bee counter. It is not a simple counting device, since among other things the corbicular pollen loads can also be quantified visually. Regardless, the system will be referred to as a bee counter for this paper. It consists of a camera unit attached to the entrance of the hives and the software BRAT (Bee Recognition and Tracking), which analyzes the collected image data. The camera is the Raspberry Pi Camera (2.1) with an IMX219 sensor and a resolution of 0.168 millimetres per pixel. An Nvidia Jetson Nano acts as the controller, which is battery-buffered and powered by a solar cell. True-colour LEDs with a high CRI value serve as the light source.

All bees pass through the camera's field of view as they enter and leave the hive. In the camera unit, a plastic glass prevents the bees from crawling over each other, but grouping is still possible. The camera unit is shown in Fig. 2. Fig. 3 shows how the video data was processed. The exact evaluation procedure is described below and follows exactly the protocol proposed in Section 3.

In addition, the "Hohenheimer Einfachbeute" with Zander frames and queenright colonies of different and local A. mellifera subspecies were used. Bees had shown no visual signs of diseases and were maintained by a professional beekeeper. To attract enough robber bees, we used honeycombs with stored honey (approx. 2 kgs per comb). The feeding regime followed a weekly or bi-weekly interval and is displayed together with the case study timeline in the appendix (Figure A.1).

4.2. Procedure & results

4.2.1. Evaluation step 1: Performance factors

Three performance factors based on experience and literature were considered. These were (i) the number of bees in the field of view of the camera, (ii) the type of lighting (infrared/white light), and (iii) the degree of crowding (Spangler, 1969; Liu et al., 1990; Bermig et al., 2020). Looking more closely at video samples, the total number of bees was less relevant and is already included in their degree of crowding. Since it is assumed that infrared light has no effect on bee behaviour (Barker, 1972), the use of infrared light is generally recommended between dusk and dawn. However, in this case study, the counting device was not attached to the beehive itself, but an "empty" robber hive. Therefore, for convenience, it was decided not to switch from white light to infrared light since the bees were not in the robber hive during the night hours. This also means that lighting did not have to be considered as a factor in the evaluation.

The remaining crowdedness factor was determined as presented in Fig. 4. The crowdedness score ranges from zero to one and expresses the proportion of the image area covered by densely crowded bees. The

Fig. 2. The camera unit of the apic.ai monitoring technology in use. The system is mounted to the entrance in front of the hive and can be easily removed for maintenance.

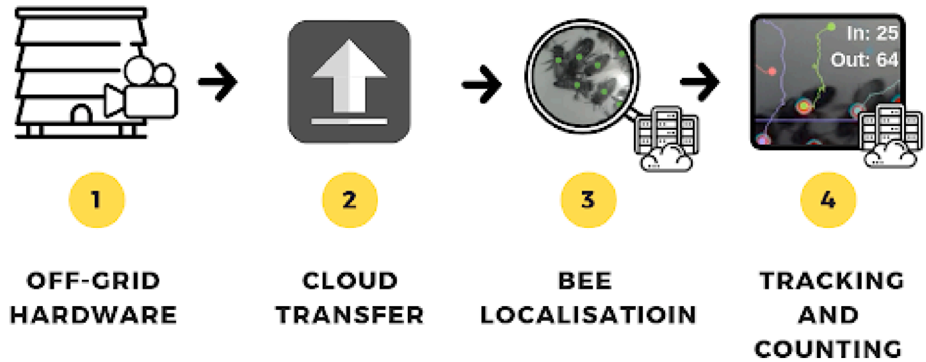

Fig. 3. The video data is first stored locally and then transferred to the cloud where CNN based localisation, tracking and counting is carried out.

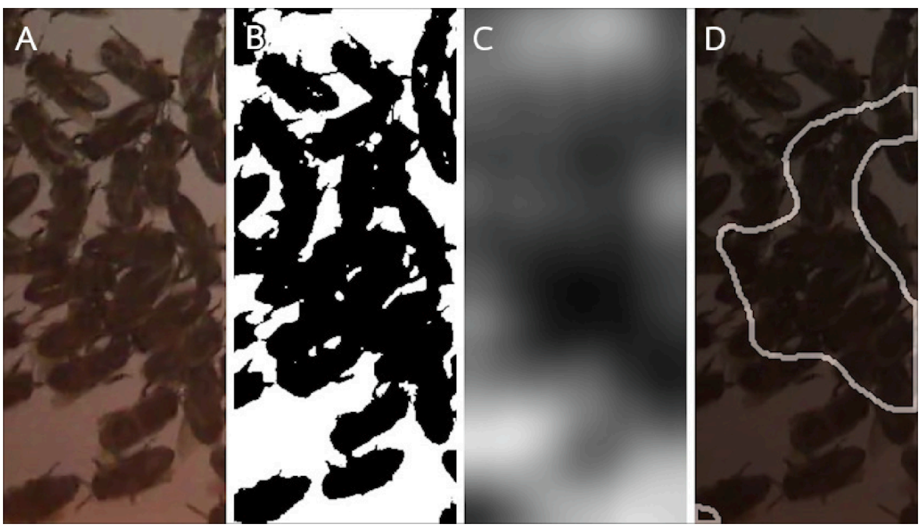

Fig. 4. Determination of the crowdedness factor in input images. From left to right: (A) A section of the raw camera image. (B) Foreground-background separation with simple thresholds. (C) A (Gaussian) blur version from image (B), revealing areas of densely crowded bees. (D) Result of thresholding for image (C). The contours show the portion of the input image where densely crowded bees were located. The crowdedness value indicates the proportion of the highlighted contour to the total area. In the present case, the crowdedness is approximately 34%.

choice of performance factors was then confirmed, as a strong and significant negative correlation was found between the quality measure chosen in step (4) and the crowdedness factor described here (details on the correlation are found in step (4)). It can therefore be assumed that no relevant performance factors were overlooked in this case study.

4.2.2. Evaluation step 2: Deriving the scenarios

Since only one relevant performance factor was found, the compilation of the scenarios was not very complex. A total of 59 video clips with very different occupancy were selected. Each clip was ten seconds long and had a frame rate of 40 fps. Fig. 5 shows example frames from three scenarios that differ greatly in their crowdedness score. Note that the crowdedness score is only indirectly related to the number of bees.

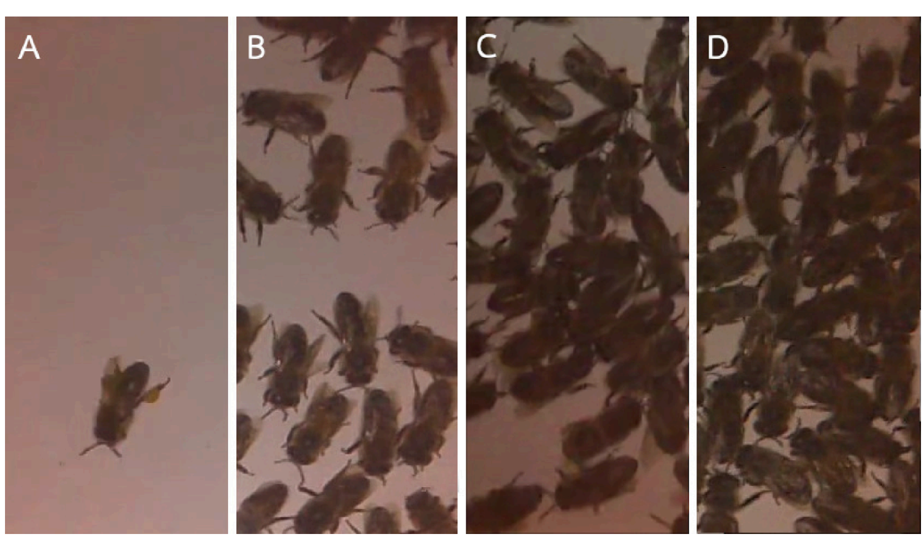

Fig. 5. Sample images from three 10-s scenarios with varying degrees of crowding (A: 0%, B: 4%, C: 34%, D: 91%).

For the performance of the bee counter, it is more important how close the bees are. While counting the bees in (A) and (B) does not cause any problems for man and machine, the same is much more error-prone in (C) and (D).

4.2.3. Evaluation step 3: Scenario annotation

All 59 scenarios in the video clips were annotated. The center of each bee was carefully marked on each clip in each frame to obtain the bees' trajectories. With a large number of bees, annotation is extremely time-consuming, taking up to 40 min per ten-second video clip.

The labels must be double-checked, as errors due to operator fatigue cannot be ruled out. It was found that even small annotation mistakes have strong effects on further analyses (e.g., a video with hardly any bees was wrongly classified as the most difficult because a single bee was forgotten to be annotated). This must be avoided at all events.

4.2.4. Evaluation step 4: Scenario evaluation-by-observation

The general functionality of the bee counter was demonstrated using the previously annotated scenarios. The 'Multiple Object Tracking Accuracy' (MOTA) (Bernardin et al., 2006) metric was used as a quality measure. The MOTA metric takes into account not only the interrupted paths of the tracker (#id-switches) but also errors of the detector (#false-positive, #false-negative) and is defined for each scenario as:

$$\mathrm{MOTA} = 1 - \frac{\#\text{false-positives} + \#\text{false-negatives} + \#\text{id-switches}}{\#\text{labeled objects}} \tag{1}$$

Since some scenarios ($N = 9$) contained empty video clips ($\#\text{labeled objects} = 0$), no MOTA score could be calculated in these cases. However, since there were no false-positive detections in all these cases, the quality measure was manually set to 1.0.

Empirical MOTA scores ranged from 0.9401 to 1.0 ($Median = 1.0$, $IQR = 0.0003$), indicating a well-performing system. The vast majority of scenarios did not pose a challenge to the system studied, and only cases of extreme overcrowding resulted in errors (see Fig. 5 right). In the most extreme scenario, $N = 102$ individuals and a total of $N = 29\,846$ positions were annotated. In this case, the number of errors was 17.52 per path (individual). A correlation analysis showed that the MOTA scores of the scenarios and the crowdedness factors were highly correlated (Pearson's $r = -0.8953, p = 1.1 \times 10^{-21}$.

Despite the good result, it cannot yet be ruled out whether the remaining errors accumulate over the day (i.e. are systematic) or not. At this point, information from the robbers test is missing to be able to make statements about longer periods.

4.2.5. Evaluation step 5: Scenario difficulty rating

The performance score MOTA of each scenario is transformed according to the protocol and represents the difficulty of the measurement condition in the following. That is, the scenario with the best score ($MOTA_{best} = 1.0$) in the previous step is assigned difficulty zero, and the scenario with the worst score ($MOTA_{worst} = 0.9401$) is assigned difficulty one. Thus, for each scenario i its difficulty_i can be calculated from its performance score p_i:

$$\text{difficulty}_i = \frac{p_{best} - p_i}{p_{best} - p_{worst}} \tag{2}$$

4.2.6. Evaluation step 6: Robbers tests

The robbers tests were carried out in Braunschweig, Germany in the period from 23/09/2021 to 14/10/2021. Additional information on the timeline and conduct of the robbers tests can be found in the appendix in figures A.1, A.2 and A.3. Due to bad weather, the opening of boxes to restock honeycombs, maintenance of the system, and missing data, most days had to be discarded, as even the shortest periods without data can have fatal implications for the evaluation. The remaining four complete measuring days are shown in Fig. 6 and were made publicly available (Borlinghaus et al., 2022). A typical activity pattern of entries and exits can be observed, which is strongly related to the diurnal pattern (Crailsheim et al., 1996) although the measurements took place towards the end of the season. While in occupied hives activity can usually be measured at night, too (Crailsheim et al., 1996), in robber hives there is neither thermoregulation nor guard bees or regular activity after dark as the box is empty.

From previous experiences, we knew that the general flight activity in robbers tests is lower than in a full-sized colony. For this reason, the experimental design of (Bermig et al., 2020) was modified and tunnels were omitted. At the site, about 60 colonies were situated in the flight radius, which could potentially visit the robber hive. As a result, the bees sometimes fought violently at the hive entrance. These behavioral changes, however, had a positive effect on crowding and complicated the measurement conditions, contradicting Rosenquist's (Rosenquist, 2019) observation that bees move in and out of the hive quickly rather than dwell at the entrance in the robbers test.

Subtracting outgoing bees from incoming bees yields daily losses of −1871 bees (01/10/2021, D1), −828 bees (02/10/2021, D2), −113 bees (03/10/2021, D3), and +56 bees (04/10/2021, D4). Since the actual 'loss' or difference in any robbers test is expected to be zero, the values here represent the daily measurement error (Odemer, 2021). These errors are net errors since the deviations balance out in both directions.

4.2.7. Evaluation step 7: Difficulty of robbers tests' measurement conditions

To determine the difficulty of the measurement conditions of an entire robbers test, the video data were divided into 30-s time intervals and the 'crowding' performance factor was determined for each. These factors were used to assign each interval the most similar scenario, and

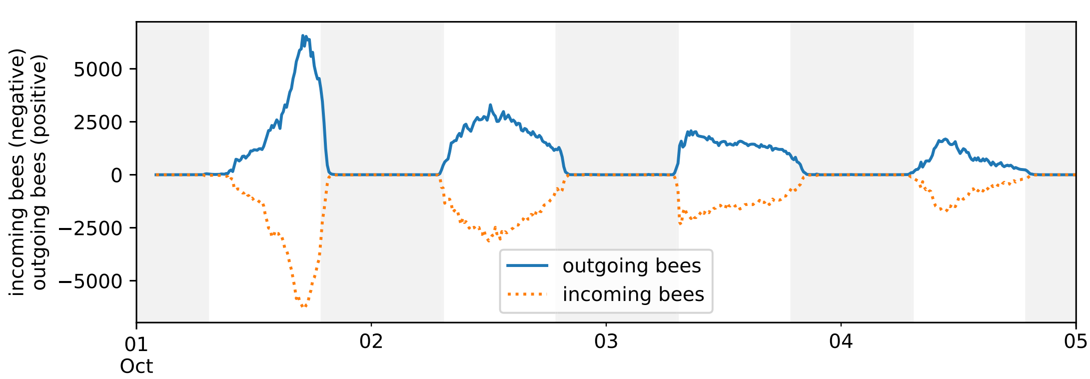

Fig. 6. Bidirectional bee movement aggregated in ten-minute intervals as determined by the bee counter. For example, 6 573 bees left the robber hive between 5 pm and 5.10 pm on 01/10/21. The first day was preceded by the renewal of the food source. The amount of forage (and the bees' interest) steadily decreased over the next four days. The shaded intervals mark the periods between dusk and dawn.

the difficulties of the scenarios were used for a weighted daily difficulty average.

Fig. 7 shows the mean difficulty in 10-min time frames for all robbers tests. Note the parallels to the activity in Fig. 6. A day consisting of only the most difficult scenario would have a daily difficulty of one. The mean daily difficulties seen here from D1 to D4 are 0.1329 (±0.3049), 0.0745 (±0.0994), 0.0383 (±0.0831), and 0.0149 (±0.0525), respectively.

Reconciling scenarios and robbers tests is more difficult when multiple performance factors are used. In some cases, it may be advisable to z-transform the performance factors to achieve comparable scaling of the characteristics.

4.2.8. Evaluation step 8: Modelling accuracy based on the difficulty of the measurement condition

The expected deviation of the bee counter from the true value as a function of the measurement condition is required. This condition is described by the performance factors and was previously referred to as 'difficulty'. Since the true value is generally not known for bee colonies, the robbers test provides a solution. Here the true value is equal to zero and the deviation from zero can be measured accurately. To determine the accuracy of the system for an arbitrary daily loss measurement, the difficulty of this measurement would first have to be defined and then a robbers test with the same difficulty would have to be performed. Since this is not possible, it is necessary to infer the difficulty from known robbers-test-error-difficulty combinations.

Thus, the mean deviation from the true measurement requires modeling based on the difficulty of the measurement. This mean deviation is equal to the standard deviation and can be easily estimated. However, since all robbers tests are generally of varying difficulty, only samples of size one are available. By linking the mean deviations of different trials, a robust statement can however still be made. The estimator for the standard deviation with known expected value μ is given by $\hat{\sigma} = \sqrt{1/N\sum_{n=1}^{N}(y_i - \mu)^2}$ where y_i is the measurement result of the i-th robbers test. For $\mu = 0$ and $N = 1$, the equation can be simplified to $\hat{\sigma} = \sqrt{y_i^2} = abs(y_i)$.

Model selection is difficult due to the small sample size ($N = 4$). Since the data points offer little guidance, the conclusion from the previous steps is used. In step (4), a strong linear correlation at the scenario level was found between MOTA and difficulty. The correlation coefficient of $r = -0.8953$ justifies the assumption of a linear relationship not only for the accuracy of the scenarios (MOTA) and their difficulty but also for the accuracy of the daily loss measures (standard deviation) and their difficulty. In addition, standard deviation, rather than variance, is modeled because it has the unit 'bees', as do the errors underlying the MOTA score. An intercept is omitted because the modeled standard deviation cannot be less than zero. In summary, the following applies to the standard deviation of the bee counter:

$$abs(y_i) = \beta_1 * x_i + \epsilon_i, \; i = 1, \ldots, n \tag{3}$$

where x_i denotes the difficulty of the robbers test. Linear regression yields $\hat{\beta} = 12\,673$ ($N = 4, r^2 = 0.95$). Note: Since the variance of the residuals increases with difficulty, heteroskedasticity is the result. However, heteroscedasticity only affects the efficiency of the estimator, not its unbiasedness, so it is not a concern. The estimated $\hat{\beta}$ can now be used to determine the standard deviation to the daily loss measurements obtained with the bee counter.

4.2.9. Evaluation step 9: Plausibility check

In this context, reference should be made to the study by Gonsior et al. (2020). There, a previous model of the bee counter was assessed as part of an ecotoxicological trial. In an experiment, four bee colonies were fed over ten days with a neonicotinoid-spiked sugar solution, known to have sublethal effects on the flight behavior of bees. When comparing the number of foraging flights with those of four untreated control colonies, significant group differences were found. The result demonstrates that the system was able to show the expected effect and represents a general plausibility test.

5. Discussion

5.1. Case study

Using scenarios is a logical consequence of the observation that the performance of bee counters varies with changing measurement conditions. One main factor was found to have a strong effect on bee counter performance. A total of 59 scenarios with varying degrees of 'crowding' were drawn and commented on. Since all scenarios achieved MOTA scores greater than 94.01%, evaluation by observation confirmed the basic functionality of the bee counter. The strong correlation between scenarios' MOTA scores and difficulties indicates that the crowdedness performance factor is highly suitable for inferring the accuracy of the bee counter. Although the MOTA score requires special treatment of scenarios for which the number of labeled objects equals zero, the advantage outweighs the disadvantage. Rather than using the sparse count events, the metric uses the underlying sources of detector and tracker error.

This promising pre-evaluation justified the implementation of several robbers tests. Despite the long experimental period of 21 days, most of the robbers tests had to be rejected. Reasons for this included the late season and the resulting increase in rainy and cold days, but also the maintenance and restocking of the feed. The robbers test also shows the importance of the operating time of the bee counter. If the system had

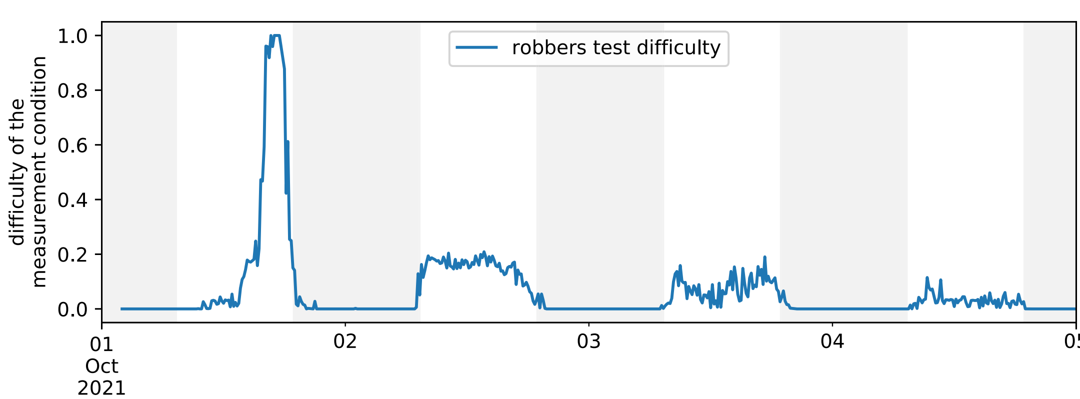

Fig. 7. To find out the difficulty of the robbers test's measurement condition, the robbers test video data was divided into 30-s intervals. For each such interval, the most similar scenario was determined based on the measurable performance factors, and its difficulty was assigned. The average difficulty across all 30-s intervals of a robbers test trial resulted in the daily difficulty.

been down for ten minutes during the most active period on D1, which would correspond to an operating time of 99.31%, 6573 outgoing bees and a similar number of incoming bees would not have been recorded. The impact on daily loss, which is likely to be between 200 and 2000 bees depending on season and colony size, is considerable. Thus, not including days with missing data is important, but reduces the number of valid trials.

The utilization of the original robbers test described by Struye (2001), in which bees from multiple colonies rob the same robber hive, could solve the problem of consistent and goal-oriented bee behavior described by Rosenquist (2019). The robber experiments contain moments of extreme crowding and thus simulate a more realistic flight behavior at the robber hive.

A total of four robbers tests were available for further analysis and showed that despite the high MOTA values of the scenarios, the measurement of the 24-h intervals is prone to error. This is reflected in a mean absolute deviation of 717 ($\pm$846) bees compared to a target value of zero. Due to limited data availability, it was not possible to make a reliable statement whether the bee counter unilaterally counts more bees or not. The observation of a bias has been reported by several researchers for other bee counters and would lead to skewed counts in and out, distorting the loss measurements (Ngo et al., 2019; Bermig et al., 2020). Consequently, the reason for a preferential direction should be investigated and eliminated. The data further corroborate findings from other authors that on days with higher flight traffic, the error of the counter increases (Odemer, 2021).

Since bees react to visible light, it is reasonable to use infrared light that is invisible to bees at night (Barker, 1972). Nevertheless, the robbers tests were performed here exclusively with white light and consequently, the performance factor 'lighting' was omitted. This had no consequences for the robbers test since no bees are expected at the feeder during the infrared hours anyway. This also means that the influence of the factor could not have been determined at all. However, for conducting experiments in the realistic scenario, the lighting factor is potentially important, since under certain circumstances bees may stay near the entrance after dark (Odemer, 2021). One way out would be to conduct additional robbers testing experiments with all-day infrared lighting. However, this would have required modifications to the system that were beyond the scope of this case study. Therefore, the regression model only applies to the trials that do not require infrared light.

Due to the small number of valid robbers tests, the model had to be derived from considerations and could not be checked directly on the data. The results are plausible and consistent with expectations.

To determine whether the tested bee counter is suitable for regulatory risk assessment and the implementation under Good Laboratory Practice (GLP), the error of the counter must be comprehensible and considered. Assuming that the robbers tests have a similar level of difficulty to that of full-sized bee colonies, the expected error in difficult cases may exceed 1500 bees. Thus, at the current stage of development (September 2021), it is not possible to determine accurate daily losses or bee mortality as defined by regulatory requirements. The estimated absolute error of the system under field conditions is too large to detect small differences in colony population dynamics. In addition to improving the bee counter, it would be possible to simplify the measurement conditions. For example, limiting the number of colonies to smaller sizes (see Spangler (1969)) or by making structural changes to the flight board that forms the camera's field of view to reduce crowding.

Further robbers tests are planned for 2022. An improved camera system promises to yield further insight into the feasibility of producing loss assessments suitable for regulatory purposes.

5.2. General

To date, there is no method to determine daily honey bee loss or background mortality with the accuracy required by the 2013 EFSA bee guidance document (European Food Safety Authority, 2013). Existing bee counters are not sufficiently suitable for this purpose, especially because data on accuracy and its evaluation under field conditions are not given. Not least because a suitable standardized evaluation protocol for the determination of accuracy has not been available (Odemer, 2021).

Although there were approaches to evaluate bee counters, the results could not be used to determine the accuracy of daily loss measurements. It was argued that (1) evaluation procedures based only on sample evaluation do not reveal bias or provide an indication of the accuracy of daily loss measurements and that (2) evaluation procedures based only on robbers tests cannot distinguish with confidence between inoperative and operative counters and do not provide accuracy measurements under realistic conditions.

A combination of the two approaches could, however, solve these problems. Hence, the protocol presented here balances specificity with adaptability to other types of designs. It was tested by being applied to a commercial counting device where it became obvious that requirements for accurate loss measurement are high. Even minimal deviations per count event accumulate because the reference quantity (activity) is two orders of magnitude larger than the quantity of interest (loss).

Apart from that, small inaccuracies in the determination of individual bee movements can have serious consequences for the calculation of the daily loss. In Struye's BeeSCAN, the activity was about a hundred times the measured daily loss (Struye, 2001). Given this factor and the relatively small reference size, effects such as those in Rickli et al. (1989), Bermig et al. (2020) can quickly occur. There, promising bee counters led to unrealistic results in the determination of daily losses. The first study reported losses that exceeded expected results by a factor of five, while the second reported a gain of over 14,000 bees in just one day. In both cases, small deviations, not measurable in short periods, added up over the day due to the sheer volume of entries and exits.

To detect field-relevant changes in daily losses, bee counters must have no more than 1 error per 1000 entries or exits, depending on colony size and requirements. These numbers assume 105,000 flights, a colony size of 30,000 bees, and a natural mortality rate of 3.75% (European Food Safety Authority (EFSA), 2020). If a bee counter fails the pre-evaluation (sample evaluation), the system must be redesigned to incorporate learnings and improvements. No generalized threshold has been established to indicate whether or not the pre-evaluation has failed. Instead, it is recommended to report the entire evaluation process. Whether a bee counter is useful or not depends on the intended application. The result of the evaluation shows which errors are to be expected under which measuring conditions. Whether these errors are acceptable or not is ultimately decided by the potential user: If there is a bee counter where the evaluation has shown that daily loss measurements with an accuracy of $\pm$100 bees can be expected under the existing measuring conditions, the inaccuracy corresponds to about 10% of the expected background mortality in the previous example.

Future designs of counters must be technically sound and capable of operating efficiently and autonomously under field conditions (Odemer, 2021). In addition, it is necessary to generate validated data with a standardized protocol that meets scientific requirements and allows accurate conclusions to be drawn about the daily loss of foragers. Without this standardization, no progress in this field will be possible. However, with the technological advancements that exist today and will exist in the future, such standardization should be readily implementable.

6. Conclusion

High-quality data on honey bee background mortality are currently unavailable due to a lack of methodology to generate them. With the here presented evaluation for daily loss measurements, a protocol was introduced that should be suitable for determining the accuracy of electronic bee counters under field conditions in a standardized way.

The protocol combines existing approaches into a new, harmonized method that can be performed regardless of how the bee counter operates. The thorough evaluation is time-consuming but only needs to be done once for a bee counter system. The work thus makes innovations in practice measurable and creates the basis for comparability of bee counting systems, enabling faster progression of the sector. Hence, it should be possible to advance the field by developing counters that meet or even exceed scientific and regulatory requirements.

Data availability statement

The data that support the findings of this study are openly available in figshare at http://doi.org/10.6084/m9.figshare.18670754.

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Katharina Schmidt, reports financial support was provided by Federal Ministry of Education and Research. Richard Odemer reports financial support was provided by Federal Ministry of Food and Agriculture. Richard Odemer reports financial support was provided by German Development Agency for Agriculture. Parzival Borlinghaus reports financial support was provided by German Federal Environmental Foundation. Frederic Tausch reports financial support was provided by Federal Ministry of Education and Research. Katharina Schmidt reports a relationship with apic.ai GmbH that includes: board membership and employment. Frederic Tausch reports a relationship with apic.ai GmbH that includes: board membership and employment. Parzival Borlinghaus was a former employee of apic.ai GmbH, which manufactures and distributes the tested bee counter.

Acknowledgements

Funding: This work was supported by the German Federal Ministry of Education and Research [Grant No. 033D004]; the Federal Ministry of Food and Agriculture (BMEL) based on a decision of the Parliament of Germany via the Federal Office for Agriculture and Food (BLE) under the innovation support program; the German Government's Special Purpose Fund held at Landwirtschaftliche Rentenbank [Grant No. 28RZ3IP092]; P. Borlinghaus receives a scholarship from the German Federal Environmental Foundation (DBU). None of the funders was involved in the study design, the collection, analysis, and interpretation of data, the writing of the report, or the decision to submit the article for publication.

We would like to thank the reviewers for their valuable comments and efforts towards improving our manuscript.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.compag.2022.106957.

References

Arnia, 2022. Arnia. https://www.arnia.co (accessed 01. April 2022).

Barker, Roy J., 1972. Is There Any Evidence that Honeybees are Attracted by Infrared? In: Bee World 53.2, pp. 66–68. doi: https://doi.org/10.1080/0005772X.1972.11097407. issn: 0005-772X, 2376–7618.

BeeAndme, 2022. BeeAndme. http://beeandme.com/ (accessed 01. April 2022).

Beehivemonitoring, 2022. Beehivemonitoring. https://beehivemonitoring.com/ (accessed 01. April 2022).

BeeScanning, 2022. BeeScanning. https://beescanning.com/ (accessed 01. April 2022).

Bermig, Sven, Odemer, Richard, Gombert, Alina J., Frommberger, Malte, Rosenquist, Ralf, Pistorius, Jens, 2020. Experimental validation of an electronic counting device to determine flight activity of honey bees (Apis mellifera L.) en. In: Journal of Cultivated Plants, pp. 132–140. doi: https://doi.org/10.5073/JFK.2020.05.03.

Bernardin, Keni, Elbs, Alexander, Stiefelhagen, Rainer, 2006. Multiple Ob ject Tracking Performance Metrics and Evaluation in a Smart Room Environment. In: Sixth IEEE International Workshop on Visual Surveil lance, in conjunction with ECCV, vol. 90. Citeseer.

Biesmeijer, J.C., Roberts, S.P.M., Reemer, M., Ohlemüller, R., Edwards, M., Peeters, T., Schaffers, A.P., Potts, S.G., Kleukers, R., Thomas, C.D., Settele, J., Kunin, W.E., 2006. Parallel Declines in Pollinators and Insect-Pollinated Plants in Britain and the Netherlands. In: Science 313.5785, pp. 351–354. doi: https://doi.org/10.1126/science.1127863. issn: 0036–8075, 1095–9203.

Borlinghaus, Parzival, Schmidt, Katharina, Tausch, Frederic, Odemer, Richard, 2022. Honey bee counter dataset. figshare. https://doi.org/10.6084/m9.figshare.18670754.v1.

Bumanis, Nikola js, Komasilova, Olvija, Komasilovs, Vitalijs, Kviesis, Armands, Zacepins, Aleksejs, 2020. Application of Data Layering in Precision Beekeeping: The Concept. In: 2020 IEEE 14th International Conference on Application of Information and Communication Technologies (AICT). IEEE, pp. 1–6. doi: https://doi.org/10.1109/AICT50176.2020.9368733.

Campbell, Jason, Mummert, Lily, Sukthankar, Rahul, 2008. Video Monitoring of Honey Bee Colonies at the Hive Entrance. In: Visual observation & analysis of animal & insect behavior, ICPR 8, pp. 1–4.

Chen, Wei-Sheng, Wang, Chien-Hao, Jiang, Joe-Air, Yang, En-Cheng, 2015. Development of a monitoring system for honeybee activities. In: 2015 9th International Conference on Sensing Technology (ICST). IEEE, Auckland, New Zealand, pp. 745–750. doi: 10.1109/ICSensT.2015.7438495. isbn: 978-1-4799-6314-0.

Chmiel, John A., Daisley, Brendan A., Pitek, Andrew P., Thompson, Graham J., Reid, Gregor, 2020. Understanding the Effects of Sublethal Pesticide Exposure on Honey Bees: A Role for Probiotics as Mediators of Environmental Stress. In: Frontiers in Ecology and Evolution 8, p. 22. doi: 10.3389/fevo.2020.00022. issn: 2296–701X.

Crailsheim, K., Hrassnigg, N., Stabentheiner, A., 1996. Diurnal behavioural differences in forager and nurse honey bees (Apis mellifera carnica Pollm). In: Apidologie 27.4, pp. 235–244. doi: 10.1051/apido:19960406. issn: 0044–8435.

Danka, Robert G, Beaman, Lorraine D, 2007. Flight Activity of USDA–ARS Russian Honey Bees (Hymenoptera: Apidae) During Pollination of Lowbush Blueberries in Maine. en. J. Econ. Entomol. 100 (2), 6.

EFSA Panel on Plant Protection Products and their Residues (PPR), 2012. Scientific Opinion on the science behind the development of a risk assessment of Plant Protection Products on bees (Apis mellifera, Bombus spp. and solitary bees). en. In: EFSA Journal 10.5, p. 2668. doi: 10.2903/j.efsa.2012.2668. issn: 1831–4732.

European Commission et al., 2016. European red list of bees. en. LU: Publications Office.

European Food Safety Authority (EFSA). Alessio Ippolito, Monica del Aguila, Elisa Aiassa, Irene Muñoz Gua jardo, Franco Maria Neri, Fernando Alvarez, Olaf Mosbach-Schulz, and Csaba Szentes, 2020. Review of the evidence on bee background mortality. en. In: EFSA Supporting Publications. doi: 10.2903/sp.efsa.2020.EN-1880. issn: 23978325.

European Food Safety Authority, 2013. Guidance on the risk assessment of plant protection products on bees (Apis mellifera, Bombus spp. and solitary bees). en. In: EFSA Journal 11.7. doi: 10.2903/j.efsa.2013.3295. issn: 18314732, 18314732.

FAOSTAT, 2021. Food and Agriculture Organization of the United Nations. (2021). Stocks of beehives world. http://www.fao.org/faostat/en/#home, visited 2022-01-14.

Federal Research Centre for Cultivated Plants (JKI), 2022. BeeCheck. https://www.dbu.de/projekt_57031643/01_db_2409.htm (accessed 01. April 2022).

Gonsior, Gundula, Tausch, Frederic, Schmidt, Katharina, Knäbe, Silvio, 2020. Impact of an Oomen feeding with a neonicotinoid on daily activity and colony development of honeybees assessed with an AI based monitoring device. In: Pharmacol. Toxicol.. https://doi.org/10.1101/2020.02.04.933556.

Hiverize, 2022. Hiverize. https://hiverize.org (accessed 01. April 2022).

Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services (IPBES), 2016. Assessment Report on Pollinators, Pollination and Food Production. en. In: Zenodo. doi: 10.5281/zenodo.3402857.

Jiang, Joe-Air, Wang, Chien-Hao, Chen, Chi-Hui, Liao, Min-Sheng, Su, Yu-Li, Chen, Wei-Sheng, Huang, Chien-Peng, Yang, En-Cheng, Chuang, Cheng-Long, 2016. A WSN-based automatic monitoring system for the foraging behavior of honey bees and environmental factors of beehives. en. In: Computers and Electronics in Agriculture 123, pp. 304–318. doi: 10.1016/j.compag.2016.03.003. issn: 01681699.

Keltronix, 2022. Eyesonhives. https://www.eyesonhives.com/ (accessed 01. April 2022).

Kridi, Douglas S., de Carvalho, Carlos Giovanni N., Gomes, Danielo G., 2016. Application of wireless sensor networks for beehive monitoring and in-hive thermal patterns detection. en. In: Computers and Electronics in Agriculture 127, pp. 221–235. doi: 10.1016/j.compag.2016.05.013. issn: 01681699.

Kulyukin, Vladimir, Mukherjee, Sarba jit, Minichiello, Angela, Truscott, Tadd, 2021. BeePIV: A Method to Measure Apis Mellifera Traffic with Particle Image Velocimetry in Videos. en. In: Applied Sciences 11.5, p. 2276. doi: 10.3390/app11052276. issn: 2076–3417.

Liu, C., Leonard, J.J., Feddes, J.J., 1990. Automated Monitoring of Flight Activity at a Beehive Entrance using Infrared Light Sensors. en. In: Journal of Apicultural Research 29.1, pp. 20–27. doi: 10.1080/00218839.1990.11101193. issn: 0021–8839, 2078–6913.

Lowland Electronics, 2022. BeeSCAN. http://users.telenet.be/lowland/ (accessed 01. April 2022).

Marchal, Paul, Buatois, Alexis, Kraus, Stéphane, Klein, Simon, Gomez-Moracho, Tamara, Lihoreau, Mathieu, 2020. Automated monitoring of bee behaviour using connected hives: towards a computational apidology. Apidologie 51 (3), 356–368.

Marstaller, Julian, Tausch, Frederic, Stock, Simon, 2019. DeepBees - Building and Scaling Convolutional Neuronal Nets For Fast and Large-Scale Visual Monitoring of Bee Hives. en. In: 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW). IEEE, Seoul, Korea (South), pp. 271–278. doi: 10.1109/ICCVW.2019.00036. isbn: 978-1-72815-023-9.

Meikle, W.G., Holst, N., 2015. Application of continuous monitoring of honeybee colonies. en. In: Apidologie 46.1, pp. 10–22. doi: 10.1007/s13592- 014- 0298- x. issn: 0044–8435, 1297–9678.

National Research Council, 2007. Status of Pol linators in North America. en. Washington, DC: The National Academies Press. doi: 10.17226/11761. isbn: 978-0-309-10289-6.

Ngo, Thi Nha, Wu, Kung-Chin, Yang, En-Cheng, Lin, Ta-Te, 2019. A real-time imaging system for multiple honey bee tracking and activity monitoring. en. In: Computers and Electronics in Agriculture 163, p. 104841. doi: https://doi.org/10.1016/j.compag.2019.05.050. issn: 01681699.

Odemer, Richard, 2021. Approaches, challenges and recent advances in automated bee counting devices: A review. en. In: Annals of Applied Biology, aab.12727. doi: 10.1111/aab.12727. issn: 0003-4746, 1744-7348.

Ollerton, Jeff, Winfree, Rachael, Tarrant, Sam, 2011. How many flowering plants are pollinated by animals? en. In: Oikos 120.3, pp. 321–326. doi: https://doi.org/10.1111/j.1600-0706.2010.18644.x. issn: 1600–0706.

Pham-Delégue, Minh-Há, Decourtye, Axel, Kaiser, Laure, Devillers, James, 2002. Behavioural methods to assess the effects of pesticides on honey bees. en. In: Apidologie 33.5, pp. 425–432. doi: https://doi.org/10.1051/apido:2002033. issn: 614 0044–8435, 1297–9678.

Rickli, M., Bühlmann, G., Gerig, L., Herren, H., Schürch, H.J., Zeier, W., Imdorf, A., 1989. Zur Anwendung eines elektronischen Bienenzählgerätes am Flugloch eines Bienenvolkes. de. In: Apidologie 20.4, pp. 305–315. doi: https://doi.org/10.1051/apido:19890403. issn: 0044–8435.

Rosenquist, Ralf, 2019. Sensorsystem zur Erfassung und Einschätzung von Einflüssen auf die Gesundheit, Bestäubungsleistung und Vitalität von Bienenvölkern. de. Abschlussbericht. GERO Meßsysteme GmbH, p. 61.

Schuhmann, Antonia, Schmid, Anna Paulina, Manzer, Sarah, Schulte, Janna, Scheiner, Ricarda, 2022. Interaction of Insecticides and Fungicides in Bees. In: Frontiers in Insect Science 1. doi: https://doi.org/10.3389/finsc.2021.808335. issn: 2673–8600.

Schurischuster, Stefan, Remeseiro, Beatriz, Radeva, Petia, Kampel, Martin, 2018. A Preliminary Study of Image Analysis for Parasite Detection on Honey Bees. In: Campilho, Aurélio, Karray, Fakhri, ter Haar Romeny, Bart (Eds.), Image Analysis and Recognition. vol. 10882. Series Title: Lecture Notes in Computer Science, Springer International Publishing, Cham, pp. 465–473. doi: https://doi.org/10.1007/978-3-319-93000-8_52.

Sledevič, Tomyslav, 2018. The Application of Convolutional Neural Network for Pollen Bearing Bee Classification. en. In: 2018 IEEE 6th Workshop on Advances in Information, Electronic and Electrical Engineering (AIEEE). doi: https://doi.org/10.1109/AIEEE.2018.8592464.

Spangler, Hayward G., 1969. Photoelectrical Counting of Outgoing and Incoming Honey Bees. In: Journal of Economic Entomology 62.5, pp. 1183–1184. doi: https://doi.org/10.1093/634jee/62.5.1183. issn: 1938–291X, 0022–0493.

Struye, M.H., 2001. Possibilities and limitations of monitoring the flight activity of honeybees by means of BeeSCAN bee counters. In: COLLOQUES-INRA, pp. 269–278.

Tausch, Frederic, Schmidt, Katharina, Diehl, Matthias, 2020. Current achievements and future developments of a novel AI based visual monitoring of beehives in ecotoxicology and for the monitoring of landscape structures. Anim. Behav. Cogn. doi: https://doi.org/10.1101/2020.02.04.933580.

Zacepins, Aleksejs, Stalidzans, E., Meitalovs, J., 2012. Application of information technologies in precision apiculture. In: Proceedings of the 13th International Conference on Precision Agriculture (ICPA 2012).

Zacepins, Aleksejs, Brusbardis, Valters, Meitalovs, Jurijs, Stalidzans, Egils, 2015. Challenges in the development of Precision Beekeeping. In: Biosystems Engineering 130, pp. 60–71. doi: https://doi.org/10.1016/j.biosystemseng.2014.12.001. issn: 15375110.