High accuracy monitoring of honey bee colony development by a quantitative method

Nuno Capela, Yoko L. Dupont, Agnès Rortais, Artur Sarmento, Alexandra Papanikolaou, Christopher J. Topping, Gérard Arnold, M. Alice Pinto, Pedro J. Rodrigues, Simon J. More, Simone Tosi, Thiago S. Alves & José Paulo Sousa

View supplementary material

Published online: 19 Jul 2022.

Submit your article to this journal

View related articles

View Crossmark data

IBRA
INTERNATIONAL BEE
RESEARCH ASSOCIATION

Taylor & Francis
Taylor & Francis Group

ORIGINAL RESEARCH ARTICLE

Check for updates

High accuracy monitoring of honey bee colony development by a quantitative method

Nuno Capela[a]* (iD), Yoko L. Dupont[b], Agnès Rortais[c], Artur Sarmento[a], Alexandra Papanikolaou[c], Christopher J. Topping[b], Gérard Arnold[d], M. Alice Pinto[e] (iD), Pedro J. Rodrigues[f], Simon J. More[g], Simone Tosi[h] (iD), Thiago S. Alves[f] and José Paulo Sousa[a] (iD)

[a]Center for Functional Ecology, University of Coimbra, Coimbra, Portugal; [b]Department of Ecoscience, Aarhus University, Aarhus, Denmark; [c]European Food Safety Authority, Parma, Italy; [d]Evolution, Genome, Behavior and Ecology laboratory, National Centre for Scientific Research, Paris, France; [e]Centro de Investigação de Montanha (CIMO), Instituto Politécnico de Bragança, Bragança, Portugal; [f]Research Centre in Digitalization and Intelligent Robotics, Instituto Politécnico de Bragança, Bragança, Portugal; [g]UCD Centre for Veterinary Epidemiology and Risk Analysis, School of Veterinary Medicine, University College Dublin, Dublin, Ireland; [h]Department of Agricultural, Forest, and Food Sciences, University of Turin, Turin, Italy

ABSTRACT

Honey bees are key insect pollinators, providing important economic and ecological value for human beings and ecosystems. This has triggered the development of several monitoring methods for assessing the temporal development of colony size, food storage, brood and pathogens. Nonetheless, most of these methods are based on visual assessments that are observer-dependent and prone to bias. Furthermore, the impact on colony development (invasiveness), as well as accuracy, were rarely considered when implementing new methods. In this study, we present and test a novel accurate and observer-independent method for honey bee colony assessment, capable of being fully standardized. Honey bee colony size is quantified by assessing the weight of adult bees, while brood and provision are assessed by taking photos and conducting image analysis of the combs with the image analysis software Deepbee®. The invasiveness and accuracy of the method were investigated using field data from two experimental apiaries in Portugal, comparing results from test and control colonies. At the end of each field experiment, most of the tested colonies had the same colony size, brood levels and honey production as the control colonies. Nonetheless, continuous weight data indicated some disturbance in tested colonies in the first year of monitoring. The overall accuracy of the image analysis software was improved by training, indicating that it is possible to adapt the software to local conditions. We conclude that the use of this fully quantitative method offers a more accurate alternative to classic visual colony assessments, with negligible impact on colony development.

Introduction

Pollination is vital for the functioning and sustainability of terrestrial ecosystems and is considered one of the most important regulating ecosystem services (IPBES, 2016). Pollinators are responsible for the maintenance of many terrestrial ecosystems, since the service they provide allows, directly or indirectly, for other species to co-exist and develop. It is estimated that 87.5% of all flowering plant species are to some extent dependent on animal pollination (Ollerton et al., 2011). Amongst pollinators, the Western honey bee (*Apis mellifera*) has been introduced worldwide and is a key species in crop pollination (Garibaldi et al., 2013; Klein et al., 2007). Furthermore, honey bees are complementary pollinators in natural habitats, as they are the most frequent visitor in 13% of plant species and the only

flower-visitor observed in 5% of the plants (Hung et al., 2018).

The increasing use of honey bees for crop pollination, derived from the rapid expansion of pollinator-dependent crops (Aizen et al., 2009), has led to an increase in the number of colonies worldwide (FAO, 2020). Even the decreasing trend in the number of colonies due to sudden colony losses and winter losses in Europe and North America from the 1990s onwards, has been inverted since 2008 (FAO, 2020; Osterman et al., 2021). Nonetheless, these figures only cover the total number of colonies, without considering their health status or strength, and high winter losses (>15%) have been reported since the occurrence of the ectoparasitic mite *Varroa destructor* outside of its native range (Potts et al., 2010; 2016). In addition to Varroa, multiple drivers of honey bee colony losses have been identified,

including historical land use changes leading to scarcity of flower resources, use of pesticides, detrimental beekeeping practices, and increased pressure from other pests and parasites (Steinhauer et al., 2018; Vanengelsdorp et al., 2009). Due to the diversity and interaction of stressors, the main challenge is to understand their impacts, in isolation and in combination, in addition to confounding factors such as landscape context and climate (More et al., 2021).

The economic and ecological value of honey bees and the urgent need to understand colony losses have motivated the implementation of research methods to assess the temporal development of colony strength and health (Delaplane et al., 2013; Human et al., 2013; EFSA AHAW Panel (EFSA Panel on Animal Health & Welfare), 2016). These methods provided researchers with numerous useful tools for colony assessment. However, universal methods and protocols for the field assessment of honey bee colonies have not been established yet, due to local variation and constraints imposed by climate, honey bee genetic diversity (subspecies and ecotypes), beekeeping practices and landscape. Therefore, the results obtained using different methods are not necessarily directly comparable. As one example, honey production may be estimated by calculating the number of honey cells in a comb or by weighting the honey frames, however, no simple conversion exists between these two measures.

Most protocols for colony strength assessment are based on the Liebefeld method or adaptations from it (Delaplane et al., 2013). The Liebefeld method consists of a visual estimate of the number of adult bees on each side of the frame, in addition to a visual estimate of the comb surface area (dm^2) containing open brood, capped brood and provision (Dainat et al., 2020). The method has been enhanced by training observers using images of combs with known cell content (e.g., Dainat et al., 2020; Hernandez et al., 2020), or by the use of a grid or a measuring tape to estimate the brood ellipse (e.g., Odoux et al., 2014). These adaptations improve the accuracy of the visual estimates, although estimates are still observer-dependent.

The need for quantitative data, which are both accurate and observer-independent, has been increasing in recent years. As a quantitative measure of the honey bee adult population, the weight of the frames with and without bees has been used as an alternative to visual assessments (e.g., Meikle & Weiss, 2017; Odoux et al., 2014). Semi-automatic or automatic analysis of comb images has also been evaluated as an alternative to visual assessment of comb cell content. However, a range of challenges has hindered the use of image analysis methods.

Some software/algorithms only reliably identified capped brood cells (Rodrigues et al., 2016; Yoshiyama et al., 2011), while others could detect different cell content but with a low accuracy level (Liew et al., 2010), or extensive time was needed for the analysis (Meikle & Weiss, 2017). A recent development, DeepBeeⒸ (Alves et al., 2020), is an open-source software capable of distinguishing different comb cell contents (eggs, larvae, capped brood, bee bread, nectar, honey and others) with high accuracy (94.3% overall accuracy according to Alves et al., 2020).

Despite these technological advancements, which enable quantitative assessments of honey bee colony strength and provision, no standard protocols are available which provide reliable and accurate colony analysis. Such an analysis should include a detailed assessment of the most important indicators of colony development and health, based on accurate, observer-independent data. Furthermore, most of the existing protocols have, in general, not been assessed for their invasiveness, i.e., their potential impact on colony development. If the monitoring method adversely affects colony development, the resulting data will not reflect normal colony growth. Hence, a standard protocol should be based on high quality quantitative data, using monitoring methods, which do not impact colony development.

The Animal Health and Welfare (AHAW) Panel of the European Food Safety Authority (EFSA) published a scientific opinion that mapped existing colony indicators and assessment methods, known as the "HEALTHY-B" toolbox (EFSA AHAW Panel (EFSA Panel on Animal Health & Welfare), 2016). Based on the most important colony health status indicators from this document, relevant methods were selected and used in a large-scale field study, in order to develop a field protocol, which involved collecting accurate quantitative empirical data on colony size (adult bees), brood development, provision, and health (diseases and parasites loads) (Dupont et al., 2021; Supplemental material A).

In the current study, we tested a new quantitative method, as an alternative to the widely used Liebefeld method. In two apiaries located in Portugal, we assessed colony strength and development using weight and image analysis of cell combs to quantify colony size, brood and provision and their dynamics across two field seasons. To evaluate the image analysis usefulness and adaptation to local conditions, we tested the accuracy of the cell detection made by the software DeepBeeⒸ. Furthermore, we tested whether disturbance induced by frequent and invasive monitoring had a measurable impact on colony development by comparing

monitored (test) colonies and non-monitored (control) colonies.

Materials and methods

Experimental set-up

Experimental apiaries were installed in two distinct landscapes: Serra da Lousã (40°02′53.6″N 8°14′.9″W) and Idanha-a-Nova (39°51′33.0″N 7°09′.7″W), Portugal. The landscape in Lousã was dominated by forested areas and scrubs, with a high diversity of nectar and pollen resources available from March to October, with a peak in May. In Idanha, the landscape was dominated by pastures, cork-oak forest and cereal crops for fodder, containing floral resources from March to July, with a peak in April/May (Dupont et al., 2021).

Each apiary included seven colonies of local honey bee populations, i.e., *Apis mellifera iberiensis* in Langstroth hives. The colonies were established in the autumn of 2018 using Varroa-treated package bees (provided by a local professional Beekeeper). To minimize variability among colonies due to genetics, colonies originating from sister queens produced in 2018, were used in both apiaries. All colonies were managed using local standard beekeeping practices. Colonies were treated against Varroa mites with Apivar (amitraz) in January 2019, with Apiguard (thymol) in August 2019 and again with Apivar in February and August 2020 and visually screened for disease symptoms at every visit.

In each apiary, five colonies (hereafter denoted *test colonies*) were subjected to regular colony assessments (see Supplemental material A for details), throughout two field seasons, between March and September of 2019 and 2020 guaranteeing low levels of Varroa mites. Test colonies were assessed approximately every ±19 days, to guarantee a snapshot of all the brood cycles within the worker bee brood development cycle of 21 days. Two colonies (hereafter denoted *control colonies*) were only subjected to colony assessments at the beginning and the end of the field season, in both years and to regular beekeeping practices during the season. In 2020, monitoring started later due to COVID-19 restrictions and cold weather in Lousã.

Continuous monitoring of hive weight

All hives were equipped with a Beeyard stand-alone hive scale for continuous logging of the hive weight. Hive weight was monitored continuously throughout two field seasons, from 05 April 2019 to 31 December 2020 in Lousã and from 30 March 2019 to 31 December 2020 in Idanha. Data on hive weight was logged continuously and automatically once per hour. However, to reduce diurnal fluctuations in weight due to foraging, the hive weight at midnight was used as the daily measure of total hive weight. For both years, cumulative weight (increase/decrease) from the beginning of each field season until honey harvesting was used to compare the development of test *vs*. control colonies, by calculating the weight difference between the cumulative mean weight of control colonies and the cumulative mean weight of test colonies.

Colony assessment

During colony assessment, smoke was applied to keep the bees on the frame. Afterwards, each comb frame was hanged on a fixed hanging scale, to ensure the scale stability and weighed with adult bees, and set aside in a separate box. In a second step, bees were gently removed from the comb by brushing them off into the original hive, and the empty comb was weighed again. Brood and food resources were assessed through image analysis of photos taken from both sides of each comb frame. To provide homogenous light conditions, images were obtained using a digital camera (DSRL Nikon D3300, 24.2 MP) installed inside a photography tunnel (for further details see Alves et al., 2020). This procedure was repeated for all the frames in the colony, before returning them to the original hive. To avoid heat loss during monitoring conducted early in the season, colony assessments were carried out when weather conditions were favourable, i.e. >14 °C, with week wind and dry weather. Monitoring was carried out as quickly as possible, to minimize disturbance of the colony. Furthermore, care was taken to cage the queen during monitoring, to avoid physical damage or exposure to cold or hot external temperatures.

The number of adult bees was calculated based on the mean weight per bee. The mean bee weight was estimated by weighing 50 individual bees, randomly collected after applying smoke to the colony (*Supplemental material B*).

Recorded images were assessed for comb cell utilization using an upgraded version of the DeepBee© software, adapted to local conditions (see next section – DeepBee© analysis training). The software automatically detected cells in the comb images and classified them into the following categories: eggs, larvae, capped brood, pollen, nectar, honey, and others (Alves et al., 2020). Although the number of honey and nectar cells was quantified by the DeepBee© software, the weight of honey or nectar varied with cell depth. Therefore, we estimated honey/nectar provision (honey production) for each comb frame by subtracting the weight of the
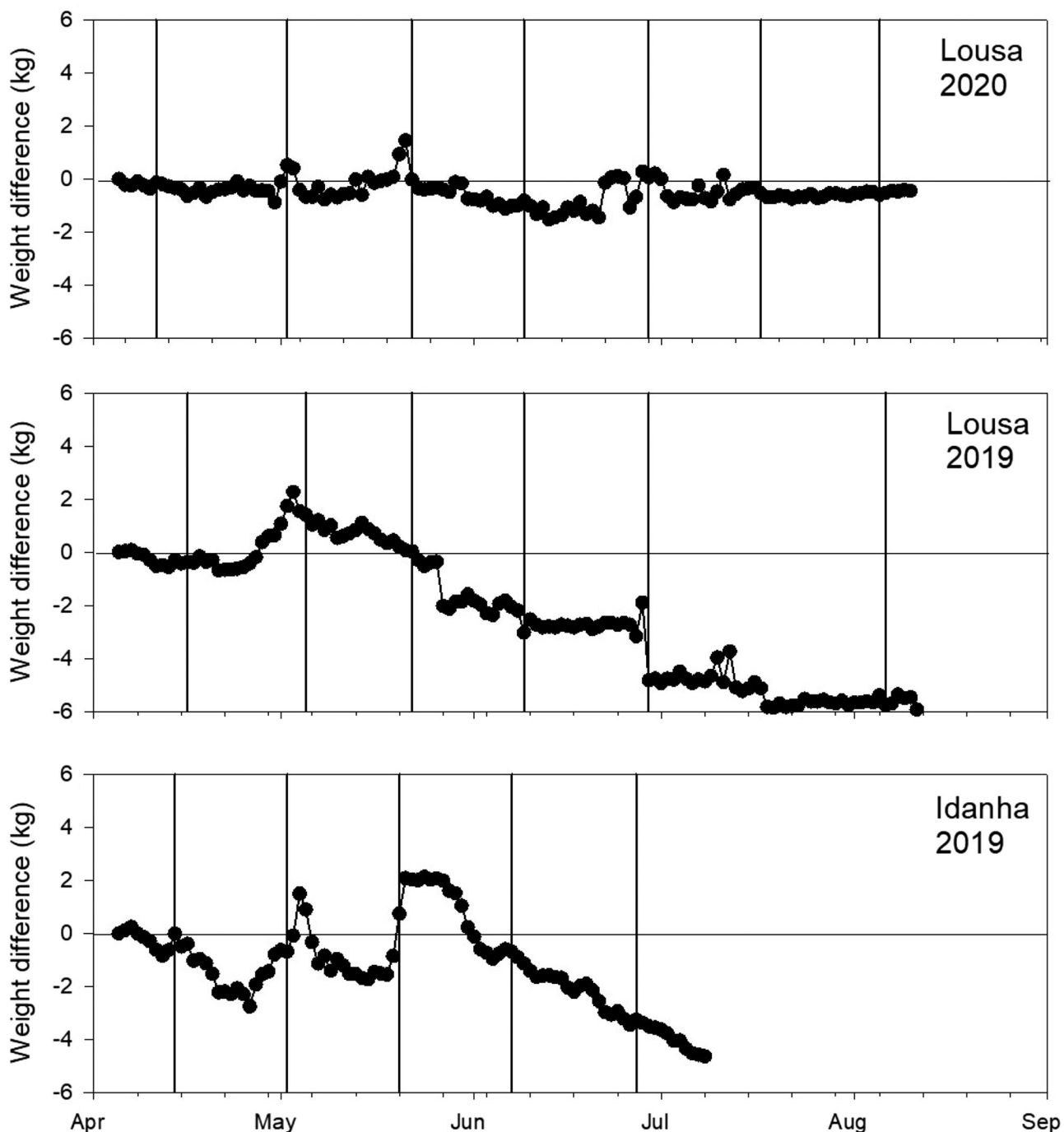
Figure 1. Seasonal change in the mean weight difference between the mean weight of test colonies and mean weight of control colonies, from the beginning of the experiment until honey collection in the Lousã apiary in 2020, Lousã apiary in 2019, and Idanha apiary in 2019. The honey collection in Idanha is earlier than in Lousã.

foundation and other components (capped brood, larva and beebread) from the weight of the frame comb without bees. The mean weight of empty frames was calculated by weighing 50 nest, and 50 honey super, Langstroth comb foundation frames. Beebread mean weight was calculated by individually weighing 100 beebread cells on a precision scale, while capped brood and larva mean weight was calculated based on Zółtowska et al. (2011), in which the body weight of the successive developmental stages (for both larvae and pupae) was determined (*Supplemental material B*).

To detect potential disturbance of colony development due to handling during the detailed colony assessment, the number of adults, the number of cells containing brood and the amount of provision (honey/nectar) were used as proxies of colony performance. The number of beebread cells was not used for comparison since honey bees have a preference to consume fresh pollen and ignore old beebread (Carroll et al., 2017). During the experiment, care was taken to verify that all the colonies had enough beebread cells in the colony.

For each study year, the performance of test and control colonies was compared at the beginning and
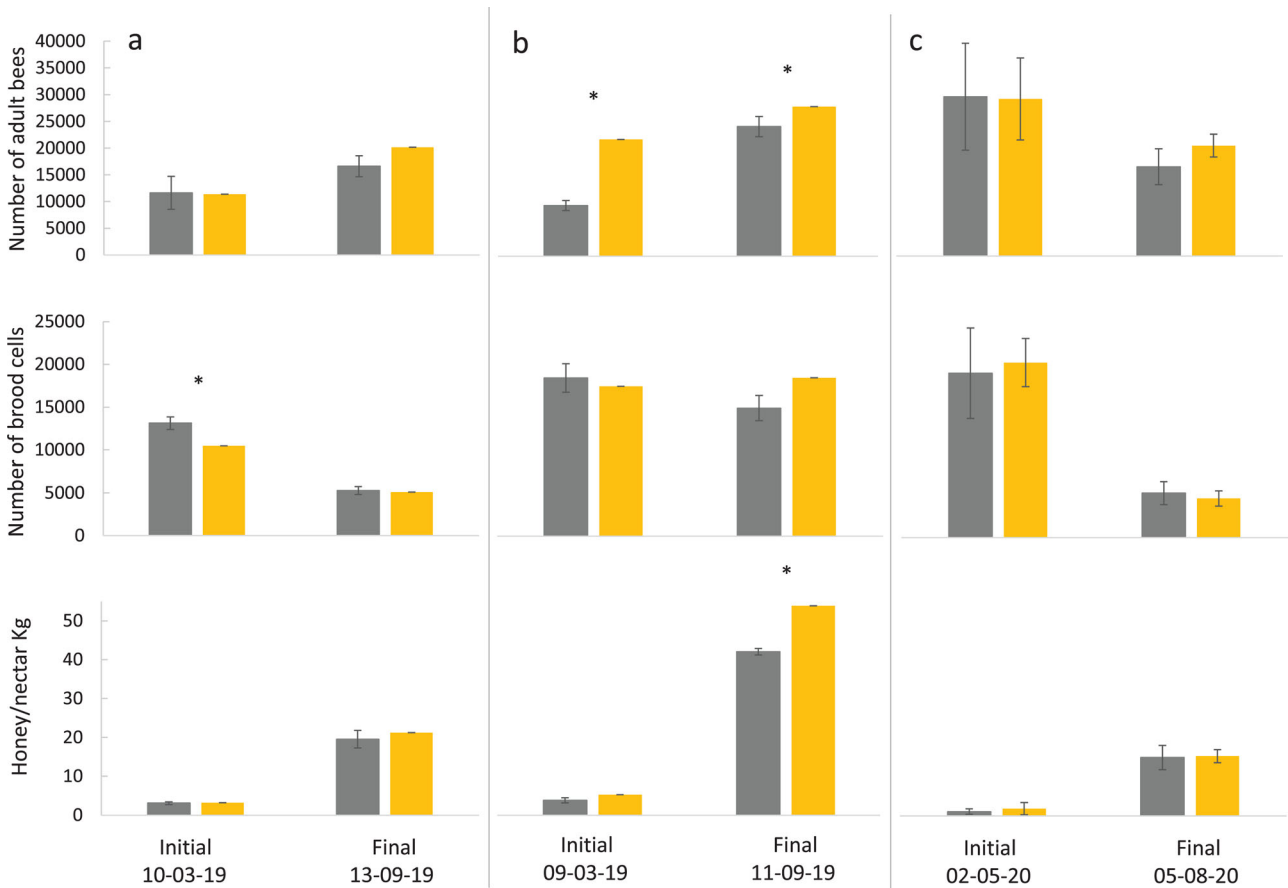
Figure 2. Mean number and SD of adult bees, brood cells (eggs + larvae + pupae) and honey/nectar present in test (grey) and control (yellow) colonies. (a) Lousã apiary in 2019 - 3 test vs. 1 control colony; (b) Idanha apiary in 2019 – 3 test vs. 1 control colony; (c) Lousã apiary in 2020 – 4 test vs 2 control colonies. (*) represents significant statistical differences.

end of the field season. For Lousã and Idanha in 2019, due to the loss by swarming of one control colony, a one-sample *t*-test was carried out to compare the colony size, brood cells, and honey production of test colonies. For Lousã 2020, an independent samples *t*-test was performed to compare the test with control colonies.

Image analysis training

The original version of the DeepBee© software has a high level of accuracy (F1 score of 94%) compared to visual assessments, although error rates varied among the different classes, with the least accurate classes being "eggs" (84% correctly identified cells), followed by larvae (88%) (Alves et al., 2020). Since the software is based on deep learning, it is possible to increase its performance and adapt it to local conditions such as variations in colours, the structure of pollen and wax, or luminosity during image capture. To improve the performance of the software, in particular for the detection of eggs and larvae, we selected 25 images containing these cell classes from our colonies. In this training set of images, the automatic classification of each cell was carefully examined and manually revised and corrected

whenever needed. After the training, the accuracy of the upgraded version was assessed using 40 random images from our colonies by comparing the automatic output with the manually corrected output. First, the error for each class was calculated as:

$$error = \left((manual\ correction - software\ output)/ \right.$$
$$\left. manual\ correction\right) \times 100$$

Secondly, the overall accuracy of each class was calculated as:

$$accuracy = 100 - abs(error)$$

Results

Colonies that swarmed or became queenless during the experiment were removed from the analysis. This resulted in data being available for analysis from three test colonies and one control colony in Lousã and Idanha in 2019, four test colonies and two control colonies in Lousã in 2020, and five test colonies and no control colonies in Idanha in 2020. Therefore, Idanha 2020 data were discarded from the analysis.

a

b
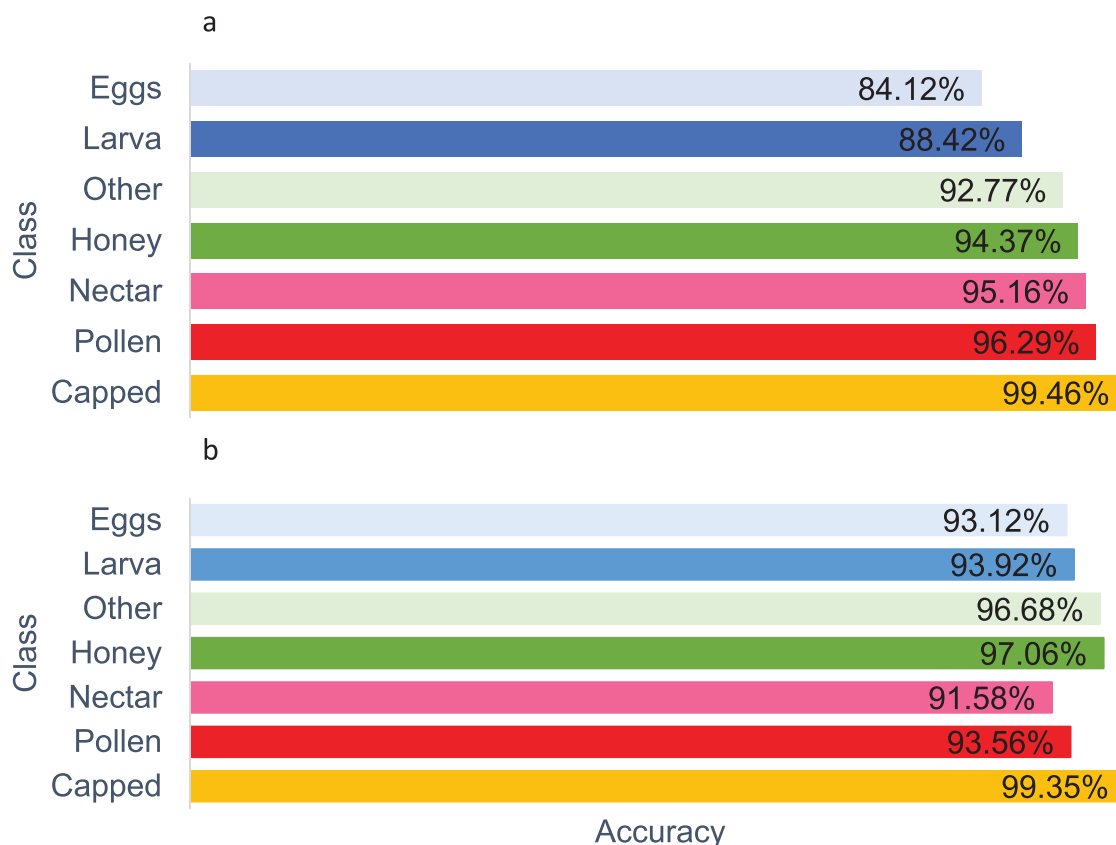
Accuracy

Figure 3. DeepBee© accuracy (a) in Alves et al. (2020) and (b) after training with a set of comb images recorded in the current study. Eggs, larvae, other and honey classification was improved while pollen and nectar were reduced. Capped cell classification remained above 99% accuracy. In the original DeepBee© version, young larvae and eggs are the main source of inaccuracy. Nevertheless, the upgraded version increased the detection of eggs and larvae above 93%.

Hive scale data

Seasonal change in hive weight (Figure 1) was calculated using the automatic hive scale data. These data reflect diurnal changes in weight due to nectar flow and pollen collection, consumption of provision, and changes in adult and brood populations. A decrease in the weight of the test colonies was observed in Lousã 2019 and Idanha 2019 compared to the control colonies by the end of the season. The difference in weight between control and test colonies in the end of the season in Lousã 2019 was approximately 6 kg, which represents approximately 9% of the total colony weight. In Idanha 2019, a similar weight difference (approximately 5 kg) representing approximately 5.5% of the total colony weight was found. In Lousã 2020, the weight patterns of the test colonies were within the range of the control ones.

Colony assessment

Colony performance parameters (colony size, number of brood cells, kg of honey/nectar at the beginning and end of the season) showed no significant differences (independent samples t-test) in Lousã in 2020 (Figure 2c). Only the number of initial brood cells in test colonies was significantly higher than in control

(one sample t-test, p = 0.036) in Lousã in 2019 (Figure 2a). In Idanha in 2019, the initial and final numbers of adult bees were lower in test colonies compared to control colonies (one sample t-test, p < 0.01 in both cases). Similarly, the final honey production was lower in test colonies compared to control ones (one sample t-test, p < 0.01; Figure 2b).

Image analysis software accuracy

Several training sessions of DeepBee© were carried out to improve the accuracy of the identification of egg cells and larvae. Comparing the original (Figure 3a with data from Alves et al., 2020) and the upgraded version after the training session in the current study (Figure 3b, from our own set of 40 random pictures after training), the image classification resulted in the preservation of a near-perfect capped brood detection, and in a better performance in detecting eggs, larvae and honey, albeit a poorer performance in pollen, and nectar classification (Figure 3b). With an overall increase in the software accuracy, the "other" class accuracy, that usually is associated with empty cells, also increased. The upgraded version of DeepBee© allowed for a more accurate assessment of brood development with a

Table 1. Comparison of available methods for honey bee colony assessment, from visual assessments (Liebefeld) to quantitative ones (the proposed method).

Parameter	Liebefeld	Coleval (Hernandez et al., 2020)	ECOBEE (Odoux et al., 2014)	Quantitative method
Data accuracy	Low	Intermediate	Intermediate	High
Distinguish eggs/larvae/pupae	Yes	No (open vs. closed brood)	No (open vs. closed)	Yes
Identification of pollen, nectar and honey	Yes	Yes	No	Yes
Colony size assessment	Yes	Yes	Yes	Yes
Observer dependent	Yes	Yes	No	No
Time spent on assessment	Slow	Fast	Fast	Slow
Experience setting	Simple	Simple	Simple	Laborious
Observer training	Laborious	Laborious	Simple	Simple
Adequate for large experiments	Yes	Yes	Yes	No

minor impact on provision quantification as honey/nectar were estimated using weight.

Discussion

In this study, we gathered detailed data on colony size, brood, and provision, monitored at regular intervals during the field season from test colonies and only at the beginning and end of the season in control colonies. This was combined with continuous daily colony weight measured by automatic hive scales. Both sets of data were used to assess if the disturbance induced by frequent and invasive monitoring had a measurable impact on colony development. Also, the image method accuracy and adaptation to local conditions were evaluated by measuring the image analysis software accuracy.

In general, comparing the performance of test and control colonies at the beginning and end of the season only indicates minor impacts due to the implementation of the protocol every ±19 days in some scenarios. In Idanha 2019, test and control colonies differed in the final colony size and honey production. However, this may be related to the initial status (the control colony had a higher population in spring 2019) and not with any impaired development. Nonetheless, the small number of colonies and high variability in colony development may mask subtle effects of monitoring mainly when comparing only two data points in time. To overcome this concern, we used scale data to assess variation during the season.

We had expected that the higher number of bees and brood cells at the beginning of the experiment would result in a larger population of foragers for collecting resources during peak flowering in spring. Nevertheless, the 2019 scale data showed a decrease in the weight of test colonies compared to control ones in both apiaries, although the decrease in weight started only after spring. This possibly means that the initial status did not play a role in the final production nor in colony size and there is a measurable impact caused by the regular colony inspections. Nonetheless, the negative impacts of monitoring on test colonies could only be detected after several assessments. This tendency was not observed in 2020, as the seasonal patterns of weight change of the test colonies were within the same range as the control ones. Furthermore, no differences were detected between the initial and end-of-season parameters when comparing the test and control colonies (Figure 2a). This suggests that the level of disturbance due to monitoring is not affected by the methodology per se, but on the handling by the observer. Unexperienced observers spend more time on each assessment, increasing the colony stress by preventing the colony from a faster return to their original state (before the stress), and by increasing the brood temperature fluctuations, which, above certain levels, cannot allow the brood to recover from the stressor (Ramirez et al., 2021). Also, the gentleness used in frame handling and brushing the bees can play a role in decreasing these stressors. Possibly, our experience in colony assessment acquired during the 2019 field season resulted in a more swift and effective monitoring and induced less disturbance to the test colonies in 2020. The registered colony weight loss in 2019 that was not registered in 2020 could therefore be explained by the higher amount of energy used to restore the colony after the suffered stressor (e.g., Schott et al., 2021).

We hence conclude that conducting detailed colony assessments every ±19 days during the field season is likely to induce some stress, although effects are subtle when comparing colony performance due to colony feedback mechanisms and to inter-colony variability. Such feedback mechanisms may allow stressed colonies to spend more energy and resources on recovering (e.g., Schott et al., 2021). Reducing the number of visits/colony assessments can compromise the temporal resolution of data points, but will decrease the induced stress, increasing the data reliability on the specific days of monitoring. Likewise, the colonies only seem to be impacted after a few colony assessments. Therefore, the use of the method during a short timeframe (i.e., less than 3 months) would not compromise the quality of the data. Finally, we recommend training sessions on non-experimental colonies before an experiment is carried out, to improve the observers' skills in colony handling.

The proposed protocol allows the user to assess colony size, brood development and provision using quantitative assessments that are independent of the observer. Compared to other existing methods (Table 1), the proposed protocol requires the construction of a photography tunnel for recording comb images (see construction details in Alves et al., 2020), and an investment in the tunnel and a digital camera. We believe that these constraints are easily overcome by researchers, but financial and logistic challenges may limit the implementation by beekeepers while doing regular colony assessments. When compared to the Liebefeld-based methods, the quantitative method allows the acquisition of accurate cell count data on all brood stages (from eggs to pupae), different food reserves (nectar/honey and pollen) and number of adult bees. The method can be applied anywhere in the world without any previous training for comb content extrapolations (observer-independent), and the comb images can be stored, and hence re-analysed and/or used in future training of DeepBee©. However, although much more accurate data are obtained by image analysis, the recording of images is more labour intensive than conducting a Liebefeld assessment. Disregarding the time required to set-up the photography tunnel (10 minutes, if the tunnel is installed in the apiary during the season), monitoring takes two people approximately 25 (±5) minutes, 32 (±5) minutes and 49 (±2) minutes for a colony with one, two boxes or three boxes, respectively. Our team achieved similar assessment times in another project in which an enhanced Liebefeld method (using a grid) was used. However, the time spent on both assessments is not directly comparable, as we had more experience in the quantitative method.

In addition to being capable of recognizing and distinguishing the different cell contents, DeepBee© is an open-access and user-friendly software. Moreover, it can be upgraded and adapted for different image acquisition conditions, including different comb frame dimensions, colours, or photographic light conditions. For instance, the software could be adapted to different pollen colours and structure of each cell originating from different bee subspecies and landscapes, as reported by Dupont et al. (2021). However, for optimizing the performance of the software, we recommend following the recommendations of the DeepBee© developers for the acquisition of images, taking into consideration lux intensity, LED positioning (see Alves et al., 2020 for details) and adjusting the tunnel dimensions to fit the frame dimensions (Dupont et al., 2021). Previously developed methods reported that images acquired under field conditions often suffer from poor and variable light conditions. For instance, in the study by Meikle and Weiss (2017), only capped brood was identified with high certainty. Furthermore, DeepBee© detects and identifies the content of every single cell, which is more accurate than extrapolating areas, and hence avoids over-estimating e.g., the number of capped brood cells, due to the empty cells in the middle of the brood area (i.e., Bargen et al., 2020). One drawback of image analysis compared to visual assessment methods is the time required for software training and processing. On the other hand, the training allowed us to improve the accuracy of DeepBee© above 93% for all cell classes, even for the low frequency cells (i.e., eggs and larvae) and the improved version performed well in the remaining images acquired during this two-year study. Also, DeepBee© can analyse a comb in less than 30 seconds and the software can run a batch of images automatically overnight, whereas, in Bargen et al. (2020), 20 minutes were required to analyse one frame comb using the HiveAnalyzer© software.

The logistic constraints of the protocol (e.g., using a tunnel for image acquisition) can also be a challenge in studies with many colonies and apiaries. Therefore, we recommend the use of this protocol in studies with <10 colonies per apiary and a limited number of apiaries. To avoid robbing events we suggest monitoring colonies far from each other and to clean and/or smoking the materials to remove any pheromones that can trigger defensive behaviours.

Future directions and developments

Automatic and semi-automatic, non-invasive, real-time and accurate data are optimal for keeping track of the health status and development of a honey bee colony. New technologies including scales and other sensors (e.g., temperature, humidity, sound, and vibration (Eouzan et al., 2019; Ramsey et al., 2020)) are attractive, as they are associated with minimal colony disturbance. These technologies can be used for colony evaluation by beekeepers or researchers, and are promising tools for obtaining standardized, large-scale, and long-term data on colony development. However, the output from these sensors needs to be validated to be transformed into colony development and health status. If these sensors are calibrated with low accuracy data, a high error rate will be associated with the output, and the user will not get accurate information on the colony development and events. In a similar vein, computer simulation models which predict colony development, are highly dependent on accurate field data for validation and calibration. EFSA proposed the development of a honey bee model as a predictive tool to assess the impact of multiple stressors on

honey bee colony development (European Food Safety Authority, 2016). The model (ApisRAM) is composed of several modules that represent the honey bee colony, the landscape management, resources and pesticide fate, and other stressors (e.g., infectious agents, pests and predators) that affect colony health (European Food Safety Authority, 2016). The combination of field data, which is collected at fixed points in time, with agent-based models having high predictive power, will represent a huge step forward allowing early detection and prevention of colony mortality (Requier, 2019). The proposed protocol was developed to gather accurate data for calibrating the ApisRAM model (Dupont et al., 2021). However, we envisage the use of the protocol in other studies across Europe and elsewhere, as it can encompass heterogeneity with regard to bee subspecies, climate, landscape, etc. Despite the possible downsides associated with a higher workload, we believe the protocol has the potential to provide reliable data, guaranteeing sound knowledge to help future decision-making.

Acknowledgements

Disclosure statement

The views or positions expressed in this article do not necessarily represent in legal terms the official position of the European Food Safety Authority (EFSA). EFSA assumes no responsibility or liability for any errors or inaccuracies that may appear. This article does not disclose any confidential information or data. Mention of proprietary products is solely for the purpose of providing specific information and does not constitute an endorsement or a recommendation by EFSA for their use.

Funding

ORCID

Nuno Capela http://orcid.org/0000-0001-7645-3207
M. Alice Pinto http://orcid.org/0000-0001-9663-8399
Simone Tosi http://orcid.org/0000-0001-8193-016X
José Paulo Sousa http://orcid.org/0000-0001-8045-4296

Data availability statement

The data that support the findings of this study are openly available in zenodo at http://doi.org/10.5281/zenodo.4953761, reference number 4953761.

References

Aizen, M. A., Garibaldi, L. A., Cunningham, S. A., & Klein, A. M. (2009). How much does agriculture depend on pollinators? Lessons from long-term trends in crop production. *Annals of Botany*, 103(9), 1579–1588.

Alves, T. S., Pinto, M. A., Ventura, P., Neves, C. J., Biron, D. G., Junior, A. C., De Paula Filho, P. L., & Rodrigues, P. J. (2020). Automatic detection and classification of honey bee comb cells using deep learning. *Computers and Electronics in Agriculture*, 170, 105244. https://doi.org/10.1016/j.compag.2020.105244

Bargen, H., Fauser, A., Gaetschenberger, H., Gonsior, G., & Knaebe, S. (2020). Bee colony assessments with the Liebefeld method: How do individual beekeepers influence results and are photo assessments an option to reduce variability? *Julius-Kühn-Archiv*, (465), 100–105.

Carroll, M. J., Brown, N., Goodall, C., Downs, A. M., Sheenan, T. H., & Anderson, K. E. (2017). Honey bees preferentially consume freshly-stored pollen. *PLoS One*, 12(4), e0175933.

Dainat, B., Dietemann, V., Imdorf, A., & Charrière, J. D. (2020). A scientific note on the 'Liebefeld Method' to estimate honey bee colony strength: its history, use, and translation. *Apidologie*, 51(3), 422–426. https://doi.org/10.1007/s13592-019-00728-2

Delaplane, K. S., Dag, A., Danka, R. G., Freitas, B. M., Garibaldi, L. A., Goodwin, R. M., & Hormaza, J. I. (2013). Standard methods for estimating strength parameters of *Apis mellifera* colonies. *Journal of Apicultural Research*, 52(4), 1–12. https://doi.org/10.3896/IBRA.1.52.4.12

Dupont, Y. L., Capela, N., Kryger, P., Alves, J., Axelsen, J. A., Balslev, M. G., Bruus, M., Castro, S., Frederiksen, J., Groom, G. B., Jeppesen, A. S., Lichtenberg-Kraag, B., Lopes, S., Pinto, A., Alves da Silva, A., Strandberg, B., Sørensen, P. B., & Sousa, J. P. (2021). Research project on field data collection for honey bee colony model evaluation. EFSA supporting publication 2021:EN-6695. 106 pp.

EFSA AHAW Panel (EFSA Panel on Animal Health and Welfare). (2016). Scientific opinion on assessing the health status of managed honey bee colonies (HEALTHY-B): A toolbox to facilitate harmonised data collection. *EFSA Journal*, 14(10), 4578.

Eouzan, I., Garnery, L., Pinto, M. A., Delalande, D., Neves, C. J., Fabre, F., Lesobre, J., Houte, S., Estonba, A., Montes, I., Sime-Ngando, T., & Biron, D. G. (2019). Hygroregulation, a key ability for eusocial insects: Native Western European honeybees as a case study. *PLoS One*, 14(2), e0200048. https://doi.org/10.1371/journal.pone.0200048

European Food Safety Authority. (2016). A mechanistic model to assess risks to honey bee colonies from exposure to pesticides under different scenarios of combined stressors and factors (Vol. 13, No. 7, p. 1069E).

FAO. (2020). FAOSTAT [WWW Document]. Retrieved 2021, November 30, from http://www.fao.org/faostat/en/#data/beehives.

Garibaldi, L. A., Steffan-Dewenter, I., Winfree, R., Aizen, M. A., Bommarco, R., Cunningham, S. A., Kremen, C., Carvalheiro, L. G., Harder, L. D., Afik, O., Bartomeus, I., Benjamin, F., Boreux, V., Cariveau, D., Chacoff, N. P., Dudenhöffer, J. H., Freitas, B. M., Ghazoul, J., Greenleaf, S., … Klein, A. M. (2013). Wild pollinators enhance fruit set of crops regardless of honey bee abundance. Science (New York, N.Y.), 339(6127), 1608–1611.

Hernandez, J., Maisonnasse, A., Cousin, M., Beri, C., Le Quintrec, C., Bouetard, A., Castex, D., Decante, D., Servel, E., Buchwalder, G., Brunet, F., Feschet-Destrella, E., de Bellescize, K., Kairo, G., Frontero, L., Pédehontaa-Hiaa, M., Buisson, R., Pouderoux, T., Aebi, A., & Kretzschmar, A. (2020). ColEval: Honey bee COLony Structure EVALuation for field surveys. Insects, 11(1), 41. https://doi.org/10.3390/insects11010041

Human, H., Brodschneider, R., Dietemann, V., Dively, G., Ellis, J. D., Forsgren, E., Fries, I., Hatjina, F., Hu, F.-L., Jaffé, R., Jensen, A. B., Köhler, A., Magyar, J. P., Özkýrým, A., Pirk, C. W. W., Rose, R., Strauss, U., Tanner, G., Tarpy, D. R., … Zheng, H.-Q. (2013). Miscellaneous standard methods for Apis mellifera research. Journal of Apicultural Research, 52(4), 1–53. https://doi.org/10.3896/IBRA.1.52.4.10

Hung, K. L. J., Kingston, J. M., Albrecht, M., Holway, D. A., & Kohn, J. R. (2018). The worldwide importance of honey bees as pollinators in natural habitats. Proceedings of the Royal Society B: Biological Sciences, 285(1870), 20172140. https://doi.org/10.1098/rspb.2017.2140

IPBES. (2016). The assessment report of the intergovernmental science-policy platform on biodiversity and ecosystem services on pollinators, pollination and food production. In Potts, S. G., Imperatriz-Fonseca, V. L., & Ngo, H. T. (Eds.), Secretariat of the intergovernmental science-policy platform on biodiversity and ecosystem services (pp. 552).

Klein, A.-M., Vaissière, B. E., Cane, J. H., Steffan-Dewenter, I., Cunningham, S. A., Kremen, C., & Tscharntke, T. (2007). Importance of pollinators in changing landscapes for world crops. Proceedings. Biological Sciences, 274(1608), 303–313. https://doi.org/10.1098/rspb.2006.3721

Liew, L. H., Lee, B. Y., & Chan, M. (2010 Cell detection for bee comb images using circular Hough transformation [Paper presentation]. In 2010 International Conference on Science and Social Research (CSSR 2010), December. (pp. 191–195). IEEE.

Meikle, W. G., & Weiss, M. (2017). Monitoring colony-level effects of sublethal pesticide exposure on honey bees. JoVE (Journal of Visualized Experiments), 1(129), e56355.

More, S., Bampidis, V., Benford, D., Bragard, C., Halldorsson, T., & Rortais, A. (2021). A systems-based approach to the environmental risk assessment of multiple stressors in honey bees. EFSA Journal, 19(5), e06607.

Odoux, J. F., Aupinel, P., Gateff, S., Requier, F., Henry, M., & Bretagnolle, V. (2014). ECOBEE: A tool for long-term honey bee colony monitoring at the landscape scale in West European intensive agroecosystems. Journal of Apicultural Research, 53(1), 57–66. https://doi.org/10.3896/IBRA.1.53.1.05

Ollerton, J., Winfree, R., & Tarrant, S. (2011). How many flowering plants are pollinated by animals? Oikos, 120(3), 321–326. https://doi.org/10.1111/j.1600-0706.2010.18644.x

Osterman, J., Aizen, M. A., Biesmeijer, J. C., Bosch, J., Howlett, B. G., Inouye, D. W., Jung, C., Martins, D. J., Medel, R., Pauw, A., Seymour, C. L., & Paxton, R. J. (2021). Global trends in the number and diversity of managed pollinator species. Agriculture, Ecosystems & Environment, 322, 107653. https://doi.org/10.1016/j.agee.2021.107653

Potts, S. G., Imperatriz-Fonseca, V., Ngo, H. T., Aizen, M. A., Biesmeijer, J. C., Breeze, T. D., Dicks, L. V., Garibaldi, L. A., Hill, R., Settele, J., & Vanbergen, A. J. (2016). Safeguarding pollinators and their values to human well-being. Nature, 540(7632), 220–229.

Potts, S. G., Roberts, S. P. M., Dean, R., Marris, G., Brown, M. A., Jones, R., Neumann, P., & Settele, J. (2010). Declines of managed honey bees and beekeepers in Europe. Journal of Apicultural Research, 49(1), 15–22. https://doi.org/10.3896/IBRA.1.49.1.02

Ramirez, L., Luna, F., Mucci, C. A., & Lamattina, L. (2021). Fast weight recovery, metabolic rate adjustment and gene-expression regulation define responses of cold-stressed honey bee brood. Journal of Insect Physiology, 128, 104178.

Ramsey, M.-T., Bencsik, M., Newton, M. I., Reyes, M., Pioz, M., Crauser, D., Delso, N. S., & Le Conte, Y. (2020). The prediction of swarming in honey bee colonies using vibrational spectra. Scientific Reports, 10(1), 1–17. https://doi.org/10.1038/s41598-020-66115-5

Requier, F. (2019). Bee colony health indicators: Synthesis and future directions. CAB Review, 14, 1–13.

Rodrigues, P. J., Neves, C., & Pinto, M. A. (2016 Geometric contrast feature for automatic visual counting of honey bee brood capped cells [Paper presentation]. In EURBEE 2016: 7th European Conference of Apidology.

Schott, M., Sandmann, M., Cresswell, J. E., Becher, M. A., Eichner, G., Brandt, D. T., Halitschke, R., Krueger, S., Morlock, G., During, R.-A., Vilcinskas, A., Meixner, M. D., Büchler, R., & Brandt, A. (2021). Honeybee colonies compensate for pesticide-induced effects on royal jelly composition and brood survival with increased brood production. Scientific Reports, 11(1), 1–15. https://doi.org/10.1038/s41598-020-79660-w

Steinhauer, N., Kulhanek, K., Antúnez, K., Human, H., Chantawannakul, P., Chauzat, M.-P., & vanEngelsdorp, D. (2018). Drivers of colony losses. Current Opinion in Insect Science, 26, 142–148.

Vanengelsdorp, D., Evans, J. D., Saegerman, C., Mullin, C., Haubruge, E., Nguyen, B. K., Frazier, M., Frazier, J., Cox-Foster, D., Chen, Y., Underwood, R., Tarpy, D. R., & Pettis, J. S. (2009). Colony collapse disorder: A descriptive study. PLoS One, 4(8), e6481. https://doi.org/10.1371/journal.pone.0006481

Yoshiyama, M., Kimura, K., Saitoh, K., & Iwata, H. (2011). Measuring colony development in honey bees by simple digital image analysis. Journal of Apicultural Research, 50(2), 170–172. https://doi.org/10.3896/IBRA.1.50.2.10

Zółtowska, K., Fraczek, R., & Lipinski, Z. (2011). Hydrolases of developing worker brood and newly emerged worker of Apis mellifera carnica. Journal of Apicultural Science, 55, 27–36.