Introducing Pollenyzer: An app for automatic determination of colour diversity for corbicular pollen loads

Parzival Borlinghaus [a,*], Jakob Jung [a], Richard Odemer [b]

[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*

ARTICLE INFO

ABSTRACT

Pollen is known to be the only source of proteins and fats for honey bees. Therefore, it is an important component of nutrition, essential for brood care and a good indicator for the availability of resources in a landscape. It is also known that a diverse diet is beneficial for bee health, also in relation to winter losses. In this work, an app is presented that allows to quantify the pollen from a pollen trap and to determine its colour diversity in an automatic way. The colour diversity is closely related to the actual plant diversity. This correlation allows conclusions to be drawn on the apicultural importance of a landscape and on biodiversity in general. In this way, the app provides beekeepers with important information about the nutritional condition of their colonies, while scientists can benefit from aggregated information about local biodiversity. The app is free of use and available as a web app on all devices.

1. Introduction

Pollen plays a crucial role in the nutrition of honey bees as it is the sole source of proteins and fats [1]. It is also an essential component for brood care and serves as a reliable indicator of the quality of a foraging location. Furthermore, a varied diet of pollen helps to increase the strength of colonies, which is particularly important for their survival during the winter months [1].

Given the importance of a diverse pollen supply, the question arises how pollen diversity can be determined. Methods of laboratory analysis include the analysis of pollen grains in honey, bee bread or pollen from pollen traps [2].

An alternative, which is less costly and requires neither expert knowledge nor laboratories, is an approximation of local and temporal pollen diversity using the colour diversity of pollen from pollen traps. Pollen traps force bees to pass through a grid when entering the hive. The grid is designed in a way that the bees lose their pollen load when passing the pollen trap, which accumulates in a container underneath. The pollen can then be manually sorted by colour to determine colour diversity, a process known as (manual) chromatic assessment. Colour diversity, in turn, can be used to indirectly estimate species diversity in flowering plants, an approach carried out by Conti et al. [3] for a particularly species-rich area in Italy. At intervals of 3 weeks, a total of 19 pollen samples were collected from three hives using pollen traps from April to September for 3 days each. In order to exclude different colour perceptions, a 10 g aliquot was subjected to a manual chromatic assessment by the very same expert. A self-made colour chart with 30 typical pollen colours served as a reference for sorting. The actual pollen diversity was then examined in the laboratory. Under these conditions, the researchers figured that each new colour found corresponded to 1.52 new species in the sample. It was also investigated whether a diversity index (Shannon) obtained by chromatic assessment was significantly different from that based on laboratory analyses. Contradictory results were obtained, as there were significant differences in

---

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

a direct comparison of two differently determined diversity indices, but not when looking across all samples using the Wilcoxon matched pair test. Kirk [4] summarises his long-term observations as follows: "Does colour diversity reflect phylogenetic diversity (...)? The broad phylogenetic association with pollen colour found here indicates that colour diversity will also broadly reflect phylogenetic diversity." [4]

Besides the different colour perception of humans (cf. Asano et al. [5]) and the influence of the light situation on the observed colours [4], the limitations are due to the properties of the pollen: It is known that pollen can change its colour depending on the degree of dryness [6], that different species can produce the same pollen colours [7] and, conversely, different pollen colours can be assigned to the same plant species in rare cases [6]. These effects result in fewer colours being found in chromatic assessment than species in laboratory analysis [3,2].

Assuming this connection, the spatial and temporal pollen (colour) diversity for many European countries was surveyed in 2014 and 2015 in a large-scale citizen science study named 'Citizen Science Investigation on Pollen' (CSI Pollen) [8]. Here, with the help of 750 beekeepers (citizen scientists), almost 18 000 pollen samples were collected and undergone manual chromatic assessment [2]. It is evident that a study that relies on hundreds of volunteers has to accept some limitations. For example, the citizen scientists could not be expected to determine the exact size of the colour clusters. Instead, beekeepers were asked to categorise pollen colours according to their frequency as very rare, rare or abundant. The study design also made it inevitable that different people were used in the visual evaluation. In the context of differences in human colour perception [5], this poses risks that seemed unavoidable at the time. The evaluation of the CSI Pollen study showed that the most significant random effect was attributed to the beekeeper. This means that different beekeepers would likely find different numbers of pollen colours in the same pollen sample. This is not surprising, since each of the 18 000 samples contained up to 20 g or 2 000 pollen, which made the manual chromatic assessment very tedious. Dimou and Thrasyvoulou [9] report that sorting a 20 g sample takes an average of 84 minutes. Depending on patience and concentration, it was up to the individual participants to sort the pollen carefully (i.e. to form many and therefore homogeneous clusters) or to work fast (i.e. to form fewer and more heterogeneous clusters). The authors, unlike Conti et al. [3] and Kirk [4], decided not to provide the participating beekeepers with fixed reference colour charts. Without a reference, the decision on whether to classify colours as different (or not) was left exclusively to the beekeeper. Practical reasons against reference colours are the difficulties in providing suitable high-quality prints (cf. Kirk [4,10]) and possible problems when pollen cannot be clearly assigned to one of two reference colours. Hence, no colour name could be assigned to the manually sorted pollen in the CSI study. This information and an exact cluster size could not be obtained for practical reasons and were not included in the study's scope.

Nevertheless, the CSI pollen study is of great impact. Through the involvement of citizen scientists, the project managed to expand across Europe, making it "much larger than individual researchers could ever achieve" [2]. While this holds true, further automation is the logical consequence of this proven idea.

In this work, we seek to objectify, accelerate and extend the manual chromatic assessment carried out before. By presenting an app that allows quantifying pollen load from pollen traps and determining their colour diversity objectively and automatically, the aforementioned shortcomings of the chromatic assessment can be solved. (1) Compared to the human based chromatic assessment approach, the proposed solution allows calibrated cluster colours to be determined, making the colours comparable across time and place. With the additionally obtained calibrated pollen colours, conversion factors from colour diversity to species diversity can be found at the level of individual colours, as suggested by Conti et al. [3]. This means that samples that show rare colours are also likely to contain fewer species in a laboratory analysis and vice versa. The automated chromatic assessment offers the possi-

bility to use the general abundance of pollen colours for the prediction of local species diversity, which is expected to strengthen the correlation and to improve the prediction accuracy. (2) In addition, colour frequency categories can be replaced by real count values easily. Not until the cluster sizes are known is it possible to determine popular biodiversity indices. (3) By making less of an imposition on citizen scientists, it is likely that even more participants can be reached for even longer periods of time with fewer dropouts. Digital support for the beekeeper is also called precision beekeeping and has already been used successfully elsewhere [11,12].

In this way, the app provides beekeepers with important information about the nutritional status of their colonies, while scientists benefit from aggregated information about local and temporal biodiversity. The web app is freely available on all devices.

2. Materials and methods

The app named 'Pollenyzer' allows users, i.e. beekeepers, to analyse the colour diversity of pollen samples. At the same time, it invites to enrich each pollen analysis with scientifically relevant metadata. The software underlying Pollenyzer is open source and publicly available.[1] The automatic chromatic assessment includes five steps. (1) A pollen trap is mounted and pollen is collected for a typical period of one to three days. (2) The collected pollen is placed on an A5 paper (210 mm x 297 mm) and photographed. (3) The resulting image is uploaded and processed via the app. Optionally, users can provide metadata about their pollen samples. (4) The user is presented with visualisations of the results. (5) Optionally, users can post-process the automatic clustering of pollen colours according to their perception.

The app is implemented as a progressive web app (PWA) that can be accessed and, if desired, installed on all devices via the browser using standard web technologies.[2] In addition to the frontend, the app consists of three backend components, the implementation of which is explained below. Subsequently, we present a study that investigated whether image processing of pollen samples can achieve the same study results as manual chromatic assessment.

2.1. Pollen load detection

The literature already contains examples of image processing of photos with pollen loads. For example, Chica and Campoy [13] and Salazar-González et al. [14] used conventional image processing methods to separate the pollen from the background. However, with both methods it was not possible (and necessary) to separate individual pollen loads. In other words, instead of the positions of individual pollen loads, only the area occupied by pollen could be discerned, a disadvantage that can be remedied with today's AI-supported methods.

Convolutional Neural Networks (CNNs) have been successfully used for the last decade and are considered state of the art in the field of image processing [15]. In this work, a CNN, more specifically a U-net [16], was trained to detect pollen on A5 paper.

The U-net architecture consists of a contracting path that grasps context and a corresponding expanding path that allows precise localisation [16]. All convolutional layers have the same kernel size of (3, 3), followed by a batch normalisation layer and a rectified linear activation. Two such layers form a convolutional block followed by a max-pooling layer with stride size of (2, 2). The contracting path consists of four such convolutional blocks with 16, 32, 64, and 128 features per convolutional layer. The contracting path is similar in structure, but uses deconvolution layers instead of max-pooling. The convolutional blocks

---

[1] https://github.com/pollenyzer/beesypollen, Pollenyzer code base, last accessed on 13.03.2023.

[2] https://pollenyzer.github.io, Pollenyzer web interface, last accessed on 13.03.2023.
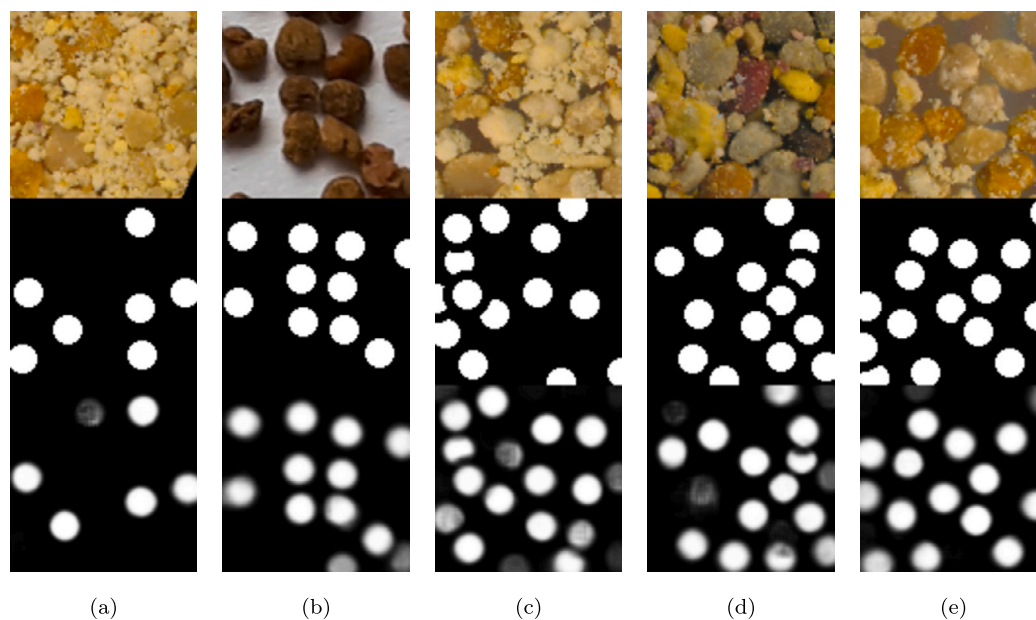
Fig. 1. Five training samples showing the input image data (top), the annotations (middle) and the neural networks output layer (bottom). Pure white pixels reflect a one hundred percent confidence of the neural network that there is a pollen centre at the pixel's location.

of the contracting path are connected to their corresponding counterparts of the expanding path by so called skip connections, which are typical building blocks of the U-net architecture. Since the task of detecting pollen is not overly complicated, the size of the network is comparably small containing less than two million trainable parameters. U-nets were originally designed for image segmentation (e.g. Borlinghaus et al. [12]), but have already been successfully used for object detection [17]. In the present case, the network's output is a single layer segmentation map showing the networks confidence that a pixel is a pollen's midpoint. Therefore, to obtain a list of pollen coordinates, U-nets require additional blob detection on the segmentation map at inference time. Here, blob detection is implemented as a search for local maxima on the thresholded segmentation map. To simplify the pollen detection, the net was trained to detect pollen on white A5 paper. The paper with known dimensions (210 mm x 297 mm) served as a size reference and allows to automatically scale input images according to the expected pollen dimensions. The network was trained on images of shape 96x96x3 and binary masks as labels. Fig. 1 shows five training input images (top), labels (middle) and output segmentation maps (bottom). Since detections at the image borders were especially hard and error-prone, a workaround was implemented. During training, cross-entropy loss was applied only to the 48x48px centre of the image, ignoring potential errors that arise in border regions. In Fig. 1c one special characteristic of the masks can be noticed. Looking closely, one can see that overlapping neighbouring pollen (white circles) were separated by additional black borders. This was to encourage the network to draw clear boundaries between detected pollen on the segmentation map, thereby facilitating subsequent blob detection. In the same figure (bottom) one sees the benefit of this practice.

The pollen came partly from a commercial pollen mix, partly from collected pollen over the course of an entire season and represented a high colour diversity. In total there were 19 996 corbicular pollen loads in the training data set and 2 629 pollen in the test data set. The pollen were taken from a total of 75 images with a resolution of up to 70 MP. The annotation was done under certain guidelines. (1) If stacked, only pollen mainly located in the uppermost layer and (2) only intact pollen parts large enough to extract robust pollen colours were annotated, excluding pollen fragments. Fig. 1a illustrates the problem annotators faced when the transition from pollen to pollen fragments was seamless. Thus, it was hardly possible to maintain a very strict annotation style

across the entire dataset, as in some cases the correct annotation was ambiguous. This inconsistency is reflected in the fact that borderline cases (cf. Fig. 1a) are assigned a mediocre confidence on the segmentation map and are included or excluded from the results by the choice of confidence threshold. The annotation could be sped up significantly as no real pollen segmentations were done. The much faster annotation of the pollen midpoints was sufficient to generate mask labels afterwards.

A total of 550 000 96x96px image parts were randomly cropped from the annotated dataset and augmented (50k of which were for validation purposes). Augmentation included slight scaling, shearing, random rotation, slight alteration of contrast, brightness and colour, and various blurring algorithms. This data set of over half a million images, masks and pollen coordinates was published as part of this work [18]. The training was stopped after 2 h on an Nvidia RTX 3090 GPU (TM), as an increasing overfitting was observed.

Due to the seamless transition from pollen to pollen parts, not only labelling but also the determination of performance metrics has become more difficult. As the labels were sometimes ambiguous or inaccurate, the calculation of intersection over union (IoU), a typical metric for recognition and localisation problems, is not well suited. In this case, the IoU would mainly quantify the inconsistency of the annotators rather than the quality of the detections. Therefore, pollen detection was performed on separate test images. An annotator then assessed whether a detection fulfilled the previously mentioned criteria. In this way, the false-positive, false-negative and true-positive detections were determined and more appropriate metrics of precision, recall and f1 were calculated and are reported in section 3.1.

2.2. Colour extraction

After the pollens' midpoints have been located, their colours were determined. To obtain a robust colour value, the colour values of multiple pixels needed to be taken into account. However, the naive approach of averaging the pixels in a certain radius around the pollen centre led to poor results. This was due to the observation that pollen was usually composed of two colours, the primary colour and the secondary colour. The secondary colour (or shade) is inherent to the three-dimensionality of the pollen and occurred in any typical lighting situation. For this reason, two alternative approaches were tested to separate the primary and secondary colour: the Gaussian mixture model

(GMM) and the k-means algorithm. While the GMM assumes that the colours in the sample are normally distributed around two mean values, k-means makes no assumptions about the distribution. The k-means algorithm is not robust to outliers, but it is much faster, often produces similar results and is therefore preferable in all cases where users are waiting for a timely response. Separating the secondary colour provides additional reliability in cases where the detected midpoints are not perfectly centred. In such cases the largest cluster's centroid was considered the primary colour and is thus purged of possible deviants.

Results of the colour extraction can be found in section 3.2.

2.3. Colour calibration

Due to the distributed nature of citizen science studies, we can not expect beekeepers to adhere to specific experimental constraints like matching imaging devices or accurately reproducible illumination. Hence, we expected images to be taken under vastly different lighting conditions and with varying spectral sensor sensitivities of the capturing devices. As a result, different, so-called device-dependent RGB values are assigned to the same pollen colour on different images. This not only impairs the comparability of clustering results between images of distinct participants but also between photographs of a single beekeeper, as there exists no known relation between the respective device-dependent colour spaces caused by differences in illumination and sensor sensitivities. In order to overcome this issue, our goal was to find translations of these colour values from device-dependent colour spaces to a standardised device-independent colour space. This process is commonly referred to as colour calibration [19].

A widely used method for colour calibration is the use of colour calibration cards. Such cards are of well-monitored production quality, correspondingly expensive and consist of a certain number of colour patches for which reference colour values are known in an absolute colour space based on the CIE Standard Observer. If placed within an image, they allow for a comparison between the observed colour of each patch and the respective reference colour. Thereby, a relation between the device-dependent colour space and the reference colour space can be inferred. Using this relationship, image colours can be transformed from device-dependent RGB values to an absolute colour space, i.e. CIEXYZ, CIELAB or sRGB. In case of a citizen science study, however, each participant would have to obtain their individual colour card, which would be a significant expense. This places an additional burden on the motivation of voluntary participation in such studies. Hence, a suitable and severely less expensive alternative needed to be found.

Examples are provided in the literature, where established colour checkers were substituted with custom devices that fulfil the requirements of a specific domain. Bautista et al. [20] constructed custom colour patches small enough to be used in whole slide scanning, while Zhang et al. [21] designed a colour card with reference colours that are particularly prevalent in diagnostic imaging of tongues. Salazar-González et al. [14] chose to use designated photo equipment under laboratory conditions to ensure pollen colour fastness in a food context. Chica and Campoy [13] circumvented the calibration by using a special computer vision device to detect non-local pollen loads by colour. Our primary constraints, however, were acquisition cost and unrestricted availability of the device to participants. We also wanted to avoid the expensive production and distribution of self-made colour cards. Hence, the object we were searching for should already be a commonplace item. Moreover, the replacement needed to comprise a collection of standardised colours suited for colour calibration.

An item that met these requirements can be found on the packaging of Kellogg's (TM) products: the PrintSpec colour strip by Mellow Colour (TM), placed on the bottom of most Kellogg's cereal products. It is designed to monitor colour consistency of printers in packaging plants and consists of 24 individual colour patches, see Fig. 2. Participants may easily obtain their copies from a wide array of grocery stores at a significantly lower price than conventional colour checkers. There

Fig. 2. Test image containing (from top to bottom): the Calibr8 ColorChart SG XS, a cutout of the PrintSpec colour strip, and a print of representative pollen colours.

are a number of factors that can influence the appearance of colours during the printing process of such packaging. These include the type and colour of the substrate and the specifics of the printing method used. For this reason, we obtained the reference colour values for the PrintSpec colour strip from the corresponding MediaStandard Print - a guideline for standardised printing processes - on which the Kellogg's printing process is based.

To allow colour calibration, the user can roughly cut out the PrintSpec colour strip and place it anywhere in the image. Similarly to the pollen detection procedure in section 2.1, we employed a U-Net for the automatic detection of PrintSpec colour strips in images. Once the PrintSpec was identified by the network, we performed a perspective correction, extracted the colour patches and used these colours in conjunction with the known CIELAB reference colours for the given packaging material to perform colour calibration. For this task, we employed the respective algorithms implemented in the colour correction module of the OpenCV-contrib python package (version 4.6.0.66). These allowed us to find a linear transformation from the linearised device-dependent colour space to some absolute colour space - in our case linear sRGB - that minimised the CIEDE2000 colour distance between reference colours and transformed image colours. Finally, the colour correction, represented by a colour correction matrix, was performed by simple matrix multiplication. The CIEDE2000 colour difference equation is the latest evolution of the CIE76 colour distance first proposed in 1976 (see Luo et al. [22] on the development of the CIEDE2000 formula and Robertson [23] on the CIE76 colour distance). Both formulae calculate distances between colours in the CIELAB colour space and as such aim to represent the human perception of colour. In experiments by Mokrzycki and Tatol [24] the following was found: CIE76 differences of $0 < \Delta E \leq 1$ are not noticed by a human observer, $1 < \Delta E \leq 2$ are noticed only by experienced observers, $2 < \Delta E \leq 3.5$ are also noticed by inexperienced observers, $3.5 < \Delta E \leq 5$ are noticed as a clear differences in colour while values above 5 are perceived as distinct colours.

To answer the question to what extent an almost free colour correction with everyday objects can replace a professional colour correction, we tested our method on 10 pictures taken at intervals of one hour over the course of a day. This test setup was intended to cover the natural light changes in our recommended photo setup using indirect daylight. As shown in Fig. 2, these images contained a professional Calibr8 ColorChart SG XS, a PrintSpec colour strip taken from a com-
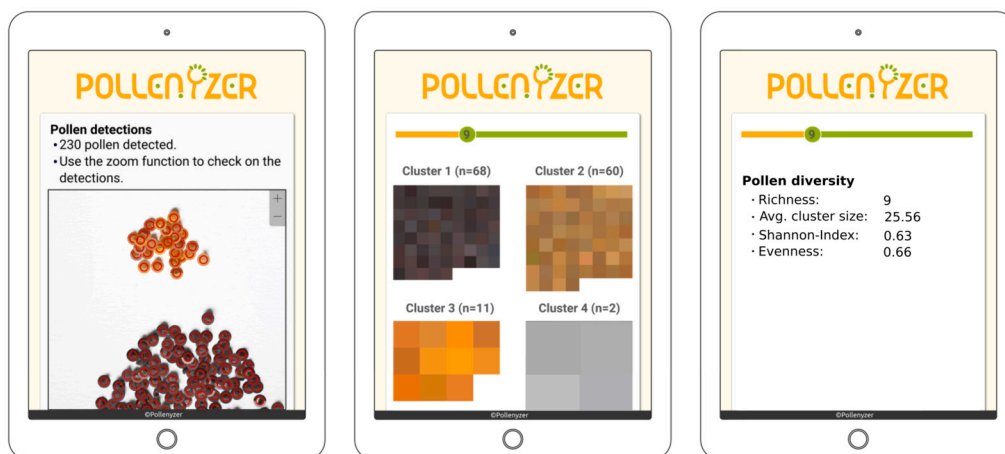
Fig. 3. Illustration of the app's user interface. Left: Red circles on the input image highlight detected pollen. Middle: All extracted pollen colours are displayed sorted by cluster. A slider allows to change the number of clusters and thus their homogeneity. Right: Pollen diversity indices are calculated according to the number of clusters.

mon packaging and a print of 16 circular patches of representative pollen colours. We performed colour calibration with both the Color-Chart and the PrintSpec colour strip (our method) and compared the results, which are shown in section 3.3.

2.4. Pollen clustering

The clustering of the extracted pollen colours is an essential part of the automatic chromatic assessment and can be performed with [3] and without [2] reference colours. Due to a lack of suitable reference colours, a hierarchical clustering of the pollen colours was performed and flattened at a inter-centroid distance of 20. With the calibration being optional, the interpretation of this threshold cannot be more precise than that it has been chosen to roughly mimic the clustering of people in a typical daylight situation.

It is known that people perceive colours differently and take different care when sorting pollen, suggesting that humans do not serve well as a universal benchmarks. Therefore, in order to meet the expectations of individual users, the possibility is offered to manually adjust the predicted number of clusters. Although the threshold may seem arbitrary, it produces consistent results if a colour calibration strip is provided.

2.5. User interface

The user interface allows interaction with the app on all devices. There are two simple user flows for data submission and data analysis. The first flow consists of an explainer, a page with the purpose of collecting metadata and the possibility to take or upload a photo. The type of metadata inquired is based on the CSI pollen survey mentioned earlier and includes information on location, landscape and collection duration. Namely, the user is asked to provide the coordinates of the apiary, the day of opening the trap, the number of days that the trap was closed and to choose the most appropriate habitat type [25]. After receiving the server's response, results are processed and displayed to the user as shown in Fig. 3. All detected pollen loads are first highlighted on the input image so that users can convince themselves of the detection accuracy (left image). The next display illustrates all extracted pollen colours for each cluster found. A slider allows to increase or decrease the number of clusters if the colours seem too heterogeneous or homogeneous. Depending on the number of clusters, the calculated biodiversity indices are displayed. In future versions it will be possible to export the results and display chronological data.

2.6. Reproduction study

To make sure that the app serves its purpose, the study of Conti et al. [3] already presented at the beginning was reproduced. There, a model was set up to investigate the linear relationship between the number of pollen colours observed and the number of species determined using light microscopy. For both the dependent variable (species count) and the independent variable (colour count), we only considered species or colours that accounted for more than one percent of the sample.

Pollen was collected from two full-sized honey bee colonies using commercial pollen traps (Heinrich Holtermann KG, Brockel, Germany). The traps were placed at the entrance of the hives and activated once a week for 24 hours. Sampling took place from April to September 2022 in Almke near Wolfsburg, Germany. A total of 54 samples were collected, of which 35 samples were palynologically analysed at the Expert Centre for Bees and Beekeeping, Mayen, Germany. The results of the analysis and a sample preparation protocol can be found in the supplementary, the image material was uploaded to a data repository [26]. Mix-ups occurred in two samples and another two samples had to be discarded due to the low quantity, so that a total of 31 pollen samples were available for further evaluation with the Pollenyzer app.

Each pollen sample was placed on a white A5 paper together with a printer test strip and photographed with a smartphone camera (Samsung Galaxy S8) in indirect sunlight at a north-facing window. For a high quality reference, the pollen was also scanned at a resolution of 1200 dpi. Both, photographs and scans were published together with this article. The Pollenyzer app was used to detect the pollen pellets, extract calibrated colours and form clusters. Analogous to Conti et al. [3], the simple linear model $species = b_0 + colours * b_1$ was estimated from the 31 samples and the strength of the linear relationship was compared with the original study.

3. Results

The following section summarises the evaluation results of three app functions, namely detection, colour extraction and colour calibration. The results of the reproduction study are listed at the end of this section.

3.1. Pollen detection

On test images that were not used for training or validation, 1924 pollen were correctly detected. A correct detection is characterised by both annotation guidelines being met. Therefore, the detected pollen loads must be placed mainly in the foreground and its size must allow

Fig. 4. The illustration shows the quality of the colour extraction. The area between the grey circles is filled with the primary colour that was determined with k-means in the vicinity of the detected pollen centres.

Table 1
Testing of pollen detections.

true positives	false-positives	false-negatives	precision	recall	F_1
1924	24	88	98.77%	95.63	97.17%

robust colour extraction. The latter is not the case, for example, if a detection deviates from the pollen centre or pollen fragments of small size have been detected as pollen and consequently no colours can be extracted. Table 1 shows that in 24 cases false positive and in 88 further cases false negative detections were reported. Overall, the precision was 98.77%, the recall (or sensitivity) was 95.63% and the harmonic mean of both values (F_1) was 97.17%. It was observed that both false negative and false positive errors are more likely to occur with dark pollen, as there is a risk of confusion with shadows.

3.2. Colour extraction

Fig. 4 shows exemplary the extracted pollen colours from various pollen detections. The area between grey circles is filled with the extracted primary colour, leaving the inner circle untouched for reference. The extracted pollen colour reproduces the pollen colour so well that it is hardly possible to distinguish it from the pixels in the inner circle. Note the slight shading within the inner circles, that could be isolated by extracting the primary colour, making the colour ring appear slightly brighter and purer. This visual examination shows that the chosen method serves its purpose well.

3.3. Colour calibration

The aim of colour calibration was to ensure that a subject always had the same colours after calibration, regardless of the lighting conditions and hardware used to capture an image. In our particular case, there was the additional objective of achieving the same quality during the calibration process as when using professional equipment.

To test the extent to which both objectives can be fulfilled by our method, a series of images were taken covering typical lighting scenarios. Following the procedure of Kirk [4], we recommend taking the pollen images in indirect daylight, ideally at a north-facing window. This will prevent unnecessary colour variation and give more consistent results. Adhering to this guideline, ten images were taken with the same mobile phone camera at one-hour intervals over the course of a day. The colours recorded in these images changed according to the position of the sun. Each image included a professional colour chart, a

Table 2
Colour distances between corresponding colours before colour calibration.

CIE76			CIEDE2000		
mean	std	max	mean	std	max
5.97	6.44	25.20	3.63	3.85	14.42

Table 3
Colour distances between corresponding colours after colour calibration for both methods.

Calibration Method	CIE76			CIEDE2000		
	mean	std	max	mean	std	max
Calibr8	3.56	3.08	19.15	2.31	2.08	12.58
PrintSpec	3.37	2.76	18.83	2.03	1.52	8.02

PrintSpec colour strip and a high quality print of 16 different typical pollen colours (see Fig. 2).

To determine the colour differences caused by light and camera before any calibration, we calculated the average pairwise distance of all 16 pollen colours to their nine corresponding pollen colours in each of the other nine images. The results are shown in Table 2 and distances are reported for both distance metrics CIE76 and CIEDE2000. With average distances of 5.97 (SD ±6.44) for CIE76 and 3.63 (SD ±3.85) for CIEDE2000, equivalent colour patches can be clearly perceived as two distinct colours by a human observer [24]. If photographs are taken in unfavourable conditions, such as with a flash or candlelight, even greater variations in colour can be expected and the need for calibration becomes even more urgent.

As expected, the colour distances between different images decreased after colour calibration (see Table 3). For the Calibr8 colour checker, the average pairwise colour distances decreased to 3.56 (SD ±3.08) for CIE76 and 2.31 (SD ±2.08) for CIEDE2000. Using our method, the average pairwise colour distances decreased to 3.37 (SD ±2.76) and 2.03 (SD ±1.52) respectively.

However, these figures alone do not show whether our method can replace a professional colour checker, as they only indicate that colours become more consistent across images after calibration, but do not reveal anything about the absolute nature of these colours. Therefore, we also calculated the average colour distance between the calibration results of the PrintSpec and Calibr8 colour checker for each image. The smaller the distance, the more similar the results of the test strip will be to common practice. The results are reported in Table 4. We could

Table 4

Colour distances between calibration results of the Calibr8 ColorChart and the PrintSpec colour strip.

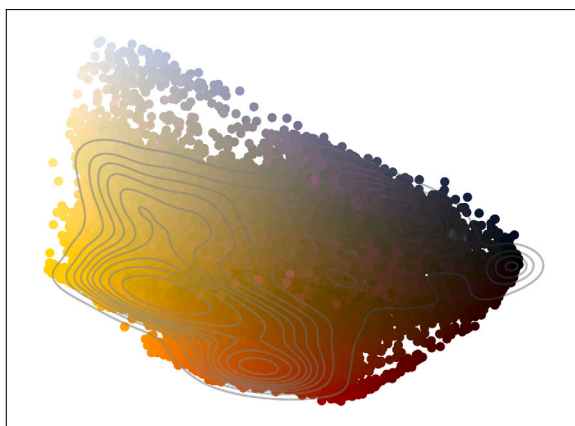| CIE76			CIEDE2000		
mean	std	max	mean	std	max
4.95	1.68	9.63	3.96	1.37	9.50

Fig. 5. The pollen colour spectrum collected from two bee colonies at a site in Almke near Wolfsburg, Germany over the period of one bee year. The colours were calibrated using print test strips and were taken from a mobile phone camera. The contour lines mark the frequency of the total of 40 918 pollen colours shown.
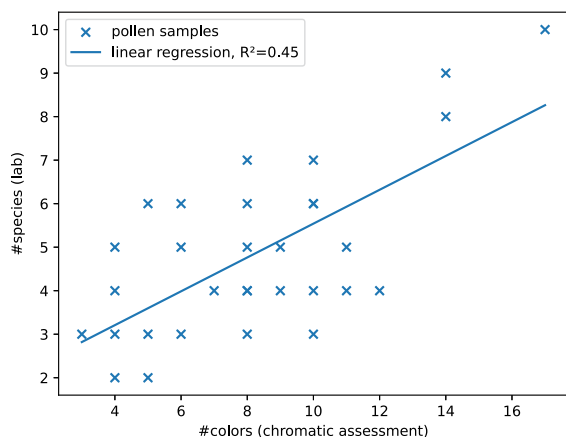

Fig. 6. The regression line shows the correlation of pollen colours and species in a pollen sample ($N = 31$). The coefficient of determination $R^2 = 0.45$ is almost identical to the original study ($R^2 = 0.44$) and shows that manual and automatic chromatic assessment are equally suitable to estimate the number of species from the observed pollen colours of a pollen sample.

observe an average distance of 4.95 (SD ±1.68) for CIE76 and 3.96 (SD ±1.37) for CIEDE2000.

3.4. Reproduction study

Fig. 5 shows the calibrated colours of a total of 40 918 pollen loads detected on 31 samples. A dimensional reduction was carried out to represent the three-dimensional colour space. The contour lines provide information about the frequency of the individual pollen colours. The dark areas around the purple hue most likely indicate the presence of Phacelia (*Phacelia tanacetifolia*), for which this colour is typical. As expected, many different shades of yellow are represented in high abundance. Shades of red are also prominent and, in contrast to light grey, much more common.

Following Conti et al. [3], a linear regression was performed to explain the dependent variable of species by the independent variable of pollen colours. For the samples studied, the regression equation $species = 1.65 + 0.39 * colours$ with $F(1, 29) = 24.10$, $p < .000$ and $R^2 = 0.45$ was obtained. Fig. 6 shows the distribution of the data and the regression line.

4. Discussion

At the beginning of this work, it was argued that automating the chromatic assessment of pollen from pollen traps has the potential to be more objective, reproducible, faster and more comprehensive than the status quo. For example, determining the relative size of pollen clusters allows more accurate prediction of colour diversity, the faster procedure reduces the labour input of study participants and can thus increase willingness to participate, and the uniform application of an algorithm eliminates the subjectivity of the human factor in assessing colour similarity. All this motivates the implementation of the automatic chromatic assessment in the form of an app called Pollenyzer.

The evaluation of the app has shown that pollen loads, if photographed on a white A5 paper, can be detected and counted almost without errors. The data set used to train the pollen detector was published along with this work, allowing others to further improve its accuracy [18]. In addition, the subsequent colour extraction allows the pollen colour to be reproduced in a particularly 'pure' way through the pixel-by-pixel classification of the colours into primary and secondary colours. This procedure also makes the app more robust against imprecise detection of pollen midpoints, as outliers up to half of the pixels can be effectively eliminated.

Until now, one of the obstacles to automating chromatic assessment has been the colour inconsistency of pollen images from mobile phones. We have succeeded in solving this problem with simple means. The use of an everyday object, namely a print test strip, makes both a photograph under studio conditions or the purchase of an expensive, professional colour chart redundant.

Surprisingly, compared to a professional colour chart our method achieves slightly better colour consistency, though imperceptible, between images of the same pollen colours in varying lighting conditions. It produces smaller average distances, shows slightly less variation and has a smaller maximum deviation than the professional counterpart. The unlikely fact that a print test strip leads to better results than a professional colour chart might be due to the slightly more central position of the print test strip in the photo. For both models, the average colour differences after calibration was less pronounced, but still noticeable for inexperienced observers [24]. In addition to the individually achieved colour consistency, the models' interchangeability was also studied. As this requires the comparison of two calibrated colourimetric values which, as noted above, are not 100% consistent themselves, this comparison is limited. All the more surprising is the small CIE76 difference of 4.95 between the two methods and the comparatively small scatter. The calibration results of both methods are, on average, perceived as clearly different but not as different colours [24]. The low variation of 1.68 indicates a tendency towards systematic error, which could be explained by unwanted differences in the target colour spaces. Although the PrintSpec test strip is used to monitor printing processes and therefore warrants for consistent colours by design, industry tolerances exist and must be taken into account. A deviation from the target colours of the print test strip within the respective tolerances could explain the observed shift between the PrintSpec target colour space and the Calibr8 target colour space.

We have therefore shown that, in a typical test setup for this application, a PrintSpec colour strip can provide comparable colour calibration results to the common practice. It should be emphasised that perfect colour consistency could not be achieved in either case and that the differences between the two methods are of the order of the general inaccuracy of colour calibration. In terms of human perception, an in-

experienced observer would still see differences between the calibrated colours on average, but would no longer perceive them as different colours. Given the biological variability of pollen colours and the inherent tolerances of cluster algorithms, it can be claimed that the method serves its purpose.

When clustering pollen colours, the use of reference colours instead of hierarchical clustering would allow the simple and informative representation of the development of pollen colours in measurement series over time. However, this procedure would not only require the curation of a suitable number of calibrated reference colours, but also make calibration for any input image mandatory. When curating, it is important to note that the selected colours should not only have similar distances to each other, but that the distances should also incorporate tolerances that reflect the expected colour degradation due to ambient light, sensor capabilities and calibration. The number of reference colours and their similarity should adapt to the precision available today. If better and better cameras become available in the future, the number of reference colours can be higher and their colours more similar. Whereas until now species diversity has been estimated by colour diversity, the introduction of distinct reference colours would allow species diversity to be estimated more accurately by the occurrence of specific pollen colours, as proposed by Conti et al. [3]. This notion is based on the following argument: At a hypothetical apiary, ten plant species are available. Few of them produce rare pollen colours (e.g. purple in Germany), many of them produce common colours (e.g. yellow in Germany). Given a pollen sample from this site that shows only one rare colour (e.g. purple), one can already assume that even a laboratory analysis will reveal only a few or even a single species, because it would be very unlikely to find two species at the same location that produce a rare pollen colour. Exactly the opposite is true for a sample that shows a single common colour (e.g. yellow). Now, one can expect that many species will be found under the microscope. Thus, observing a rare colour should have a small impact on the prediction of species diversity, whereas observing a common colour should have a large impact. However, up to now, only the number and proportion of colours have been taken into account but not which ones. Please note that in practice a pure yellow pollen sample is more likely than a pure purple sample, but the rationale holds true regardless.

Yet there are also disadvantages to the reference-colour approach and reasons that ultimately lead to the choice of implementing a hierarchical clustering. Firstly, it should be noted that reference colours can only show their full strength in the context of colour calibration, which is currently offered optionally and is not required by users. Secondly, when bees bring in pollen of a species that falls exactly on a colour boundary, each pollen of this species will be randomly assigned to one cluster or the other according to its natural colour dispersion. If such a case occurs, biodiversity indices based on it will be unrealistically high. Thirdly, the compilation of pollen colours requires a large database of calibrated pollen colours, preferably acquired through the Pollenyzer app itself. However, the compilation of such a database will only begin with the release of the Pollenyzer app and will not be available in the short term. For these reasons, the implementation of a pollen clustering based on reference colours is considered future work. By then, a hierarchical clustering of the pollen was performed and flattened at a threshold that approximately mimics the clustering of humans.

The ultimate goal of estimating biodiversity based on observed colour diversity is subject to known limitations, regardless of the method used. These include the fact that different plant species can produce very similar pollen colours and, in rare cases, the same plant species can produce different pollen colours [4]. Moreover, pollen colours can change with fluctuating humidity and composition [7]. Despite these general limitations, a linear relationship between colours and species with an R^2 of 0.44 has been established by Conti et al. [3]. On the one hand, replacing manual colour assessment with automatic assessment introduces two additional sources of error: the need for pollen detection and colour calibration. We have shown that pollen detection is extremely accurate, even with overlapping and damaged pollen. However, colour calibration is more difficult to evaluate as PrintSpec colour strips are subject to limited but unknown colour variations. In our test case, calibration with the PrintSpec colour strip yielded slightly different but more consistent colours than the professional alternative at a fraction of the cost.

On the other hand, automating colour assessment also eliminates an important source of error: the human factor. Differences in colour perception, patience and thoroughness lead to a high degree of subjectivity in pollen sorting. As shown in the CSI pollen study, the random factor 'beekeeper' is the most important factor influencing the number of pollen colours found.

The application of the app on real samples was conducted to show whether it also yields plausible results in practice. For this purpose, the study on the possibilities of (manual) chromatic assessment by Conti et al. [3] was reproduced. When comparing the results, it is noticeable that both the intercept (1.65 vs. 3.49) and the coefficient (0.39 vs. 1.52) are lower for Pollenyzer. The reason for this is partly due to general differences in the environment of the apiaries and partly due to the fact that Conti et al. [3] used about 72 hour collection intervals and the samples used here were collected in only 24 hours. Shorter collection intervals lead to a generally lower number of species and correspondingly lower intercept. Instead of also comparing the coefficients, it is more important to compare the coefficients of determination R^2 (0.45 vs. 0.44), which shows that the manual and the chromatic evaluation are equally suitable for determining the species diversity in a pollen sample and only differ in the time required.

Therefore, the increased objectivity and the enormous time savings make automatic assessment an attractive alternative, despite the potential introduction of new sources of error. Nevertheless, the accuracy of the method should not be overestimated. This also means that the temptation to choose cluster thresholds lower than justified by the natural variability of pollen colour, lighting and colour extraction should be resisted.

5. Conclusion

Biodiversity monitoring is a task that attracts the attention of researchers, policy makers and other stakeholders worldwide. With the Pollenyzer app, beekeepers aren't the only ones with access to an automated tool that can determine pollen colour, and thus plant diversity, in the landscapes where their bees forage. In other words: We can see the landscape through the eyes of bees and identify what is beneficial to them. The app is able to calibrate images and make them comparable across time and place. Databases can be created with metadata and georeferences to map plant diversity in a standardised way. Since we have shown that automatic pollen colour identification can facilitate and replace all aspects of manual colour identification, future citizen science projects to create pollen databases are easily possible. Nationwide projects and eventually PAN European Networks could provide an important monitoring tool for the development of bee-relevant plant diversity in the EU.

CRediT authorship contribution statement

Parzival Borlinghaus: Conceptualisation, Formal analysis, Investigation (lead), Methodology, Project administration, Software (lead), Writing - original draft (lead). Jakob Jung: Investigation (supporting), Software (supporting), Writing - original draft (supporting). Richard Odemer: Resources, Writing - review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The source code was published on github and the pollen image data was uploaded to figshare together with the results of the light microscopic analysis.

Acknowledgements

Appendix A. Supplementary material

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.atech.2023.100263.

References

[1] R. Brodschneider, K. Crailsheim, Nutrition and health in honey bees, Apidologie 41 (2010) 278–294, https://doi.org/10.1051/apido/2010012.

[2] R. Brodschneider, E. Kalcher-Sommersguter, S. Kuchling, V. Dietemann, A. Gray, J. Božič, A. Briedis, N.L. Carreck, R. Chlebo, K. Crailsheim, M.F. Coffey, B. Dahle, A.V. González-Porto, J. Filipi, D.C. de Graaf, F. Hatjina, P. Ioannidis, N. Ion, A.S. Jørgensen, P. Kristiansen, A. Lecocq, J.-F. Odoux, A. Özkirim, M. Peterson, B. Podrižnik, S. Rašić, G. Retschnig, A. Schiesser, S. Tosi, F. Vejsnæs, G. Williams, J.J. van der Steen, CSI pollen: diversity of honey bee collected pollen studied by citizen scientists, Insects 12 (2021) 987, https://doi.org/10.3390/insects12110987.

[3] I. Conti, P. Medrzycki, F.V. Grillenzoni, F. Corvucci, S. Tosi, V. Malagnini, M. Spinella, M.G. Mariotti, Floral diversity of pollen collected by honey bees (Apis mellifera L.) – validation of the chromatic assessment method, J. Apic. Sci. 60 (2016) 209–220, https://doi.org/10.1515/jas-2016-0028.

[4] W.D.J. Kirk, The colours of pollen available to honey bees through the year, Bee World 95 (2018) 74–77, https://doi.org/10.1080/0005772X.2018.1449280.

[5] Y. Asano, M.D. Fairchild, L. Blondé, P. Morvan, Color matching experiment for highlighting interobserver variability, Color Res. Appl. 41 (2016) 530–539, https://doi.org/10.1002/col.21975.

[6] W.D.J. Kirk, A Colour Guide to Pollen Loads of the Honey Bee, 2 ed., International Bee Research Association, Cardiff, 2006.

[7] D. Hodges, The Pollen Loads of the Honeybee. A Guide on Their Identification by Colour and Form, Bee Research Association, London, 1952.

[8] J.S. van der Steen, R. Brodschneider, Public participation in bee science: C.S.I. pollen, Bee World 91 (2014) 25–27, https://doi.org/10.1080/0005772X.2014.11417585.

[9] M. Dimou, A. Thrasyvoulou, Efficient use of pollen traps to determine the pollen flora used by honey bees, J. Apic. Res. (2006) 42–46, https://doi.org/10.3896/IBRA.1.45.1.10.

[10] W.D.J. Kirk, Recording the colours of pollen loads, Bee World 75 (1994) 169–180, https://doi.org/10.1080/0005772X.1994.11099224.

[11] P. Borlinghaus, R. Odemer, F. Tausch, K. Schmidt, O. Grothe, Honey bee counter evaluation – introducing a novel protocol for measuring daily loss accuracy, Comput. Electron. Agric. 197 (2022) 106957, https://doi.org/10.1016/j.compag.2022.106957.

[12] P. Borlinghaus, F. Tausch, L. Rettenberger, A purely visual re-ID approach for bumblebees (Bombus terrestris), Smart Agric. Technol. 3 (2023) 100135, https://doi.org/10.1016/j.atech.2022.100135.

[13] M. Chica, P. Campoy, Discernment of bee pollen loads using computer vision and one-class classification techniques, J. Food Eng. 112 (2012) 50–59, https://doi.org/10.1016/j.jfoodeng.2012.03.028.

[14] C.Y. Salazar-González, F.J. Rodríguez-Pulido, A. Terrab, C. Díaz-Moreno, C.A. Fuenmayor, F.J. Heredia, Analysis of multifloral bee pollen pellets by advanced digital imaging applied to functional food ingredients, Plant Foods Hum. Nutr. 73 (2018) 328–335, https://doi.org/10.1007/s11130-018-0695-9.

[15] A. Krizhevsky, I. Sutskever, G.E. Hinton, ImageNet classification with deep convolutional neural networks, Commun. ACM 60 (2017) 84–90, https://doi.org/10.1145/3065386.

[16] O. Ronneberger, P. Fischer, T. Brox, U-net: convolutional networks for biomedical image segmentation, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2015, pp. 234–241.

[17] K. Bozek, L. Hebert, A.S. Mikheyev, G.J. Stephens, Towards dense object tracking in a 2D honeybee hive, in: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, IEEE, Salt Lake City, UT, 2018, pp. 4185–4193.

[18] P. Borlinghaus, Pollen detection dataset, https://doi.org/10.6084/m9.figshare.21679949.v1, 2022, publisher: figshare.

[19] H.-C. Lee, Introduction to Color Imaging Science, Cambridge University Press, 2005.

[20] P.A. Bautista, N. Hashimoto, Y. Yagi, Color standardization in whole slide imaging using a color calibration slide, J. Pathol. Inform. 5 (2014) 4, https://doi.org/10.4103/2153-3539.126153.

[21] B. Zhang, W. Nie, S. Zhao, A novel color rendition chart for digital tongue image calibration, Color Res. Appl. 43 (2018) 749–759, https://doi.org/10.1002/col.22234.

[22] M.R. Luo, G. Cui, B. Rigg, The development of the CIE 2000 colour-difference formula: CIEDE2000, Color Res. Appl. 26 (2001) 340–350, https://doi.org/10.1002/col.1049.

[23] A.R. Robertson, The CIE 1976 color-difference formulae, Color Res. Appl. 2 (1977) 7–11, https://doi.org/10.1002/j.1520-6378.1977.tb00104.x.

[24] W.S. Mokrzycki, M. Tatol, Colour difference delta e - a survey, Mach. Graph. Vis. Int. J. 20 (2011) 383–411.

[25] R. Brodschneider, C.S.I. pollen: anleitung zur untersuchung der für honigbienen erhältlichen pollendiversität, http://bienenstand.at/wp-content/uploads/2014/08/CSIPollen%C3%96.pdf, 2014.

[26] P. Borlinghaus, R. Odemer, Pollenyzer dataset - honey bee pollen samples with scans, photos and laboratory analysis, https://doi.org/10.6084/m9.figshare.23065052, 2023, publisher: figshare.