



Data Article

Weight, temperature and humidity sensor data of honey bee colonies in Germany, 2019–2022



Diren Senger^{a,*}, Clemens Gruber^b, Thorsten Kluss^a,
Carolin Johannsen^a

^aAG Cognitive Neuroinformatics, University of Bremen, Enrique-Schmidt-Str. 5, 28359 Bremen

^bHiveeyes, Berlin

ARTICLE INFO

Article history:

Received 23 October 2023

Revised 20 December 2023

Accepted 22 December 2023

Available online 27 December 2023

Dataset link: [Weight, Temperature and Humidity Sensor Data of Honey Bee Colonies in Germany, 2019 - 2022 \(Original data\)](#)

Keywords:

Insects

Internet of Things (IoT)

Citizen science

Time series

ABSTRACT

Humans have kept honeybees as livestock to harvest honey, wax and other products for thousands of years and still continue doing so. Today however, beekeepers in many parts of the world report unprecedented high numbers of colony losses. Sensor data from honey bee colonies can contribute to new insights about development and health factors for honey bee colonies. The data can be incorporated in smart decision support systems and warning tools for beekeepers. In this paper, we present sensor data from 78 honey bee colonies in Germany collected as part of a citizen science project. Each honey bee hive was equipped with five temperature sensors within the hive, one temperature sensor for outside measurements, a combined sensor for temperature, ambient air pressure and humidity, and a scale to measure the weight. During the data acquisition period, beekeepers used a web app to report their observations and beekeeping activities. We provide the raw data with a measurement interval of up to 5 s as well as aggregated data, with per minute, hourly or daily average values. Furthermore, we performed several preprocessing steps, removing outliers with a threshold based approach, excluding changes in weight that were induced by beekeeping activities and combining the sensor data with the most important meta-data from the beekeepers' observations. The

* Corresponding author.

E-mail address: diren@uni-bremen.de (D. Senger).

data is organised in directories based on the year of recording. Alternatively, we provide subsets of the data structured based on the occurrence or non-occurrence of a swarming event or the death of a colony. The data can be analysed using methods from time series analysis, time series classification or other data science approaches to form a better understanding of specifics in the development of honey bee colonies.

© 2023 The Author(s). Published by Elsevier Inc.

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

Specifications Table

Subject	Biostatistics																																																																																																																
Specific subject area	The sensor data can be used in honey bee research or to implement a sensor-guided warning system for beekeepers																																																																																																																
Data format	Raw, Filtered, Aggregated																																																																																																																
Type of data	Table																																																																																																																
Data collection	From each hive we collected 5 temperature values for inside temperature and one outside temperature value (DS18B20 sensors), humidity and pressure values (BME280 sensor) and the overall weight of the colony (from a Bosche H30/H40 loadcell). We only included data from calibrated setups. Beekeeper reported their observations and activities in a web app, which is used as meta-data.																																																																																																																
Data source location	We provide the raw data with a measurement interval of 5–10 s, but also aggregated and preprocessed data with an interval of one minute, one hour or one day. The data was collected at apiaries in Germany. In the following we list the colony locations, rounded to one digit for anonymisation. <i>In the key column, we specify the unique identifier of the colony.</i>																																																																																																																
	<table border="1"> <thead> <tr> <th></th> <th><i>lat</i></th> <th><i>lon</i></th> <th><i>key</i></th> </tr> </thead> <tbody> <tr><td>1</td><td>50.9</td><td>7.0</td><td>5</td></tr> <tr><td>2</td><td>52.5</td><td>13.4</td><td>6</td></tr> <tr><td>3</td><td>53.1</td><td>8.8</td><td>11</td></tr> <tr><td>4</td><td>50.9</td><td>7.0</td><td>22</td></tr> <tr><td>5</td><td>52.5</td><td>13.4</td><td>25</td></tr> <tr><td>6</td><td>50.9</td><td>7.0</td><td>26</td></tr> <tr><td>7</td><td>53.1</td><td>8.8</td><td>27</td></tr> <tr><td>8</td><td>53.1</td><td>8.8</td><td>43</td></tr> <tr><td>9</td><td>52.5</td><td>13.4</td><td>46</td></tr> <tr><td>10</td><td>53.1</td><td>8.8</td><td>48</td></tr> <tr><td>11</td><td>50.9</td><td>7.0</td><td>49</td></tr> <tr><td>12</td><td>50.9</td><td>7.0</td><td>56</td></tr> <tr><td>13</td><td>51.9</td><td>8.6</td><td>58</td></tr> <tr><td>14</td><td>52.0</td><td>8.5</td><td>62</td></tr> <tr><td>15</td><td>51.9</td><td>8.9</td><td>0</td></tr> <tr><td>16</td><td>53.4</td><td>7.2</td><td>66</td></tr> <tr><td>17</td><td>51.9</td><td>8.9</td><td>67</td></tr> <tr><td>18</td><td>53.1</td><td>8.8</td><td>3</td></tr> <tr><td>19</td><td>51.3</td><td>7.2</td><td>68</td></tr> <tr><td>20</td><td>53.5</td><td>9.6</td><td>69</td></tr> <tr><td>21</td><td>50.0</td><td>9.1</td><td>72</td></tr> <tr><td>22</td><td>50.8</td><td>6.9</td><td>73</td></tr> <tr><td>23</td><td>51.3</td><td>7.2</td><td>74</td></tr> <tr><td>24</td><td>52.0</td><td>8.6</td><td>75</td></tr> <tr><td>25</td><td>51.9</td><td>8.9</td><td>76</td></tr> <tr><td>26</td><td>50.8</td><td>6.9</td><td>77</td></tr> <tr><td>27</td><td>53.4</td><td>7.2</td><td>78</td></tr> </tbody> </table>		<i>lat</i>	<i>lon</i>	<i>key</i>	1	50.9	7.0	5	2	52.5	13.4	6	3	53.1	8.8	11	4	50.9	7.0	22	5	52.5	13.4	25	6	50.9	7.0	26	7	53.1	8.8	27	8	53.1	8.8	43	9	52.5	13.4	46	10	53.1	8.8	48	11	50.9	7.0	49	12	50.9	7.0	56	13	51.9	8.6	58	14	52.0	8.5	62	15	51.9	8.9	0	16	53.4	7.2	66	17	51.9	8.9	67	18	53.1	8.8	3	19	51.3	7.2	68	20	53.5	9.6	69	21	50.0	9.1	72	22	50.8	6.9	73	23	51.3	7.2	74	24	52.0	8.6	75	25	51.9	8.9	76	26	50.8	6.9	77	27	53.4	7.2	78
	<i>lat</i>	<i>lon</i>	<i>key</i>																																																																																																														
1	50.9	7.0	5																																																																																																														
2	52.5	13.4	6																																																																																																														
3	53.1	8.8	11																																																																																																														
4	50.9	7.0	22																																																																																																														
5	52.5	13.4	25																																																																																																														
6	50.9	7.0	26																																																																																																														
7	53.1	8.8	27																																																																																																														
8	53.1	8.8	43																																																																																																														
9	52.5	13.4	46																																																																																																														
10	53.1	8.8	48																																																																																																														
11	50.9	7.0	49																																																																																																														
12	50.9	7.0	56																																																																																																														
13	51.9	8.6	58																																																																																																														
14	52.0	8.5	62																																																																																																														
15	51.9	8.9	0																																																																																																														
16	53.4	7.2	66																																																																																																														
17	51.9	8.9	67																																																																																																														
18	53.1	8.8	3																																																																																																														
19	51.3	7.2	68																																																																																																														
20	53.5	9.6	69																																																																																																														
21	50.0	9.1	72																																																																																																														
22	50.8	6.9	73																																																																																																														
23	51.3	7.2	74																																																																																																														
24	52.0	8.6	75																																																																																																														
25	51.9	8.9	76																																																																																																														
26	50.8	6.9	77																																																																																																														
27	53.4	7.2	78																																																																																																														

(continued on next page)

28	53.1	8.8	79
29	49.2	6.9	80
30	53.0	8.7	84
31	52.9	9.2	17
32	49.0	12.1	85
33	49.5	7.5	87
34	50.8	6.9	89
35	52.0	8.6	21
36	47.7	11.8	90
37	51.3	7.2	93
38	52.6	12.7	95
39	53.4	9.6	96
40	53.1	8.8	97
41	53.4	9.7	98
42	48.2	11.3	100
43	53.1	8.8	36
44	51.9	7.7	101
45	50.8	6.9	102
46	53.4	7.2	103
47	53.2	8.5	104
48	53.1	8.8	105
49	53.1	8.8	107
50	51.9	8.9	109
51	52.0	8.5	47
52	51.5	6.9	111
53	53.1	8.8	112
54	47.7	11.8	115
55	52.4	13.3	57
56	53.1	8.8	118
57	53.1	8.8	119
58	53.4	7.2	120
59	48.9	8.3	123
60	52.3	9.3	60
61	53.1	8.8	64
62	52.6	12.7	128
63	50.9	7.0	129
64	53.7	10.3	70
65	50.9	7.0	134
66	52.0	8.5	82
67	53.1	8.8	86
68	51.9	8.9	135
69	53.8	10.2	88
70	51.2	6.8	99
71	53.7	10.3	136
72	50.8	6.7	108
73	50.7	7.3	141
74	53.8	10.2	126
75	50.7	7.3	139
76	48.2	11.3	151
77	53.6	10.2	152
78	50.9	7.0	132

Data accessibility

Repository name: Zenodo

Data identification number: 10.5281/zenodo.8389137

Direct URL to data [1]: <https://zenodo.org/records/10407693>

1. Value of the Data

- Honey bees contribute to the pollination of flowering plants, including flowers of agricultural crops [2–4] and natural habitats [5]. It is estimated that 68 % of leading single crops depend on insect pollination [2]. However, beekeepers all over the world report high number of colony losses and often it is difficult to identify one specific cause for the dying of a colony [6–9].

- The data can enable researchers to form a better understanding of the behaviour of honey bees and of factors that impact their development and health.
- The data can be incorporated into a consultation and warning system for beekeepers, supporting them in decision making and beekeeping activity planning and notifying them about changes in the hives that might require immediate actions.
- The data benefits bee researchers and beekeepers.
- The data can also be used by software engineers, who want to build a sensor-supported warning and recommendation system for beekeepers.
- Researchers could for example use the data to study correlations, apply methods from time series classification and anomaly detection.

2. Data Description

The data set, which is available from Zenodo, consists of two types of data: sensor data from honey bee hives and meta-data about the health and development states of the bee colonies as observed by the beekeepers. We provide the original form of the data, in which both information types are stored in separate data sets, but also provide a merged set, combining the sensor data and information about the most important events in the development of a bee colony in one frame [1].

2.1. Sensor data

Fig. 1 shows the positioning of the sensors during data collection. The sensor data includes 5 temperature values from every second clearance between the frames inside the hive, humidity, pressure and an additional temperature measurement of a separate sensor also positioned inside the hive, the outside temperature and the weight of the hive.

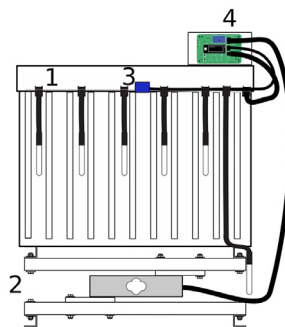


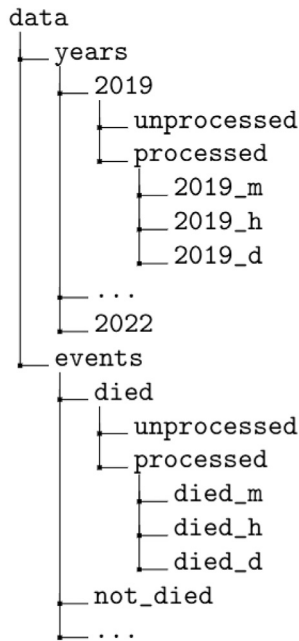
Fig. 1. Sensor setup in honey bee hive: DS18B20 temperature sensors (1), Bosch H30/H40 load cell for weight measurements (2), BME280 sensor for humidity, pressure and temperature (3), and micro-controller (4). (First published in [10]).

In Table 1 we list the column headings (first column) associated with sensor measurements and we give a description of the specific measurements (second column), name the sensor used for the measurements (third column) and specify its location at the hive (fourth column). Furthermore, each set of sensor measurements (one row in the data) is linked to the timestamp of its recording, and the unique key of the hive.

Table 1

Overview of columns with sensor values.

Column(s)	Description	Sensor type	Location
t_i_1, ..., t_i_5	temperature (°C) inside of the hive	DS18B20	located in every second clearance between frames, t_i_3 being in the centre and t_i_1 and t_i_5 at the margin of the hive
t_o	outside temperature (°C)	DS18B20	hanging outside at the hive
weight_kg	overall weight (kg)	Bosche H30/ H40 and HX711	underneath the hive
h	relative humidity (%)	BME280	in the clearance between two frames
t	temperature (°C) inside of the hive	BME280	in the clearance between two frames
p	pressure (hPa)	BME280	in the clearance between two frames

**Fig. 2.** Directory tree of provided data.

2.2. Organisation of data directories

We provide the complete sensor data organized according to two different paradigms: years and events (Fig. 2). In the years directory the data is divided based on the year of recording, starting in 2019 and ending with 2022. In the events directory we provide 3-month periods of data from bee colonies before a certain event was observed by a beekeeper, and intervals of the same length from colonies during similar periods of the year, where the specific event was not observed. As events, swarming and the death of a colony were chosen. For each year and event, there are two sub-directories: unprocessed and processed. The first one contains the unprocessed data, as it was obtained by a call to our database. The second one contains the data after first processing steps and merged with the most important information about the

state of the bee colony obtained from the inspection data. The processed data is available with three different intervals for the sensor data: one-minute (directory ending with `_m`), one-hour (directory ending with `_h`) and one-day (directory ending with `_d`). For each honey bee colony assigned to the sub-directories there is its own file in csv format. The file name contains the unique key of the hive and, for the event specific data, the timestamp of the event.

2.3. Beekeepers' inspection data

Beekeepers in the BeeObserver project used a web app to take note of their observations and activities at the apiary, giving important information about the development and health state of their colonies. We give further details about this app in the method section . We supply the original inspection data, but since the inspection lists can contain more than 100 different items, we decided to select the most important events and states in the development of a bee colony to merge with the sensor data: The occurrence of a **queen cell**, in which potentially a new queen bee is developing; the occurrence of a **swarm**, when a new queen bee hatched and about a third of the honey bee colony leaves the hive with its old queen; information about any additional **food**, e.g. sugar syrup, supplied by the beekeeper; the amount of **honey** harvested; **treatments** undertaken by the beekeeper, e.g. to decrease the level of varroa-infestation of their colony; the **death** of a colony. From these states and events, swarming events and the death of colonies are of special interest for beekeepers and we provide additional data files sorted based on the occurrence or non-occurrence of these events. The merged data contains, additionally to the columns with sensor data summarised in Table 1, four columns for each of the mentioned events specifying the timestamp and the distance in time for the last and next occurrence of each event (Table 2). For example, a value of 600 (seconds) in the column `feeding.last` implies that the colony was fed 10 min before the measurements were taken that are presented in the same row of the data set. We added two additional columns to the data which we prepared based on the occurrence or non-occurrence of an event: the timestamp of the event of interest and the distance in time to this event from the point in time when the measurements presented in the row were taken. Furthermore, the merged data contains the longitude and latitude of the apiary location, rounded to the first digit.

In Germany, the use of bee hives comprising of several stackable boxes is most common, with a box for brood at the bottom and additional boxes called supers stacked on top for honey comb building. Beekeepers frequently add or remove supers, adapting to a colony's fluctuating size and for harvesting. Sometimes beekeepers use additional stones to secure their hives against wind. This leads to sudden changes in the weight data, that are not related to the colonies' de-

Table 2

Overview of columns with inspection data and apiary information.

Column(s)	Description
<code>event.last</code> for event $\in \{queencell, feeding, honey, treatment, died, swarming\}$	timestamp of last listing of event in inspection data
<code>event.next</code> for event $\in \{queencell, feeding, honey, treatment, died, swarming\}$	timestamp of next listing of event in inspection data
<code>event.last.dif</code> for event $\in \{queencell, feeding, honey, treatment, died, swarming\}$	time distance to last listing of event in inspection data (seconds)
<code>event.next.dif</code> for event $\in \{queencell, feeding, honey, treatment, died, swarming\}$	time distance to next listing of event in inspection data (seconds)
<code>lon</code>	longitude of apiary, rounded to first digit, corresponding to a precision of 5–7 km for central Europe
<code>lat</code>	latitude of apiary, rounded to first digit
<code>event</code>	timestamp of event, e.g. swarming, <i>only for data frames sorted based on events in inspection data</i>
<code>time_dist_event</code>	time distance to event (seconds), <i>only for data frames sorted based on events in inspection data</i>

Table 3

Overview of columns with weight processing information.

Column(s)	Description	Example
weight_delta	difference to previous weight measurement, the measurement from one minute ago (kg)	In the sequence [1.0 NA 1.0 1.2 2.0] of weight data, <i>weight_delta</i> would be: [0.0 NA NA 0.2 0.8]
no_jump	Boolean variable, <i>true</i> if preceding measurement (one-minute interval) is available	In the sequence [1.0 NA 1.0 1.2 2.0] of weight data, <i>no_jump</i> would be [NA TRUE TRUE TRUE TRUE]
outlier_lim	Boolean variable, <i>true</i> if <i>weight_delta</i> is larger than a limit, for this publication 0.3 kg per minute	In the sequence [1.0 NA 1.0 1.2 2.0] of weight data, <i>outlier_lim</i> would be [FALSE FALSE FALSE FALSE TRUE]
weight_delta_noOutlier	<i>weight_delta</i> with outliers set to 0 (based on <i>outlier_lim</i>)	In the sequence [1.0 NA 1.0 1.2 2.0] of weight data, <i>weight_delta_noOutlier</i> would be [0.0 NA NA 0.2 0.0]
weight_kg_noOutlier	cumulative sum of <i>weight_delta_noOutlier</i>	In the sequence [1.0 NA 1.0 1.2 2.0] of weight data, <i>weight_kg_noOutlier</i> would be [0.0 NA 0.0 0.2 0.2]
with_0	<i>weight_delta_noOutlier</i> with all NAs replaced with 0	In the sequence [1.0 NA 1.0 1.2 2.0] of weight data, <i>with_0</i> would be [0.0 0.0 0.0 0.2 0.0]

velopment. We therefore provide a column with weight information corrected for these changes. Our exact algorithm is described in the methods section. The corrected weight is stored in the column *weight_kg_noOutlier*. The cleaned change in weight per time interval might also be interesting to the user, as it is a more comparable measure of the activity of different colonies. It is stored in the column *weight_delta_noOutlier*. Furthermore, we added some columns related to intermediate steps of the weight processing, which might be relevant for advanced use cases. Note, that the aggregated data (one-hour or one-day interval) contains the averaged values. We provide an overview of all columns related to the weight processing in [Table 3](#).

The original inspection data with rounded coordinates is stored in the file *dump_june_2023_rounded_coordinates.csv*. The data in wide data format is available in *wide-June23.csv*. Additionally, we provide the starting and end dates we used to exclude test measurements in *all_info.csv*.

3. Experimental Design, Materials and Methods

In this section we describe the data collection process given our two data sources: the sensors at the honey bee hives and the web application to be used by beekeepers for the documentation of observations and beekeeping activities. We also describe the preprocessing steps we performed on the data to offer a ready-to-use version of the data set, as well as the raw data.

3.1. Sensor setup

We use a Do-It-Yourself (DIY) sensor setup, which is aimed to be low-cost and at the same time allowing for different hive types, data transmission options and the use of multiple temperature sensors. As depicted above in [Fig. 1](#), it consists of 6 one-wire digital temperature sensors (DS18B20), a combined sensor for humidity, pressure and temperature (BME280) and a single point load cell (Bosche H30 or Bosche H40). All sensors are connected to a circuit board with an ESP32 microcontroller (Fipy by Pycom), positioned in a water-proof box outside of the hive. We employ the robust version of DS18B20 temperature sensors with a stainless steel tube and a waterproof cable. The temperature sensors are connected to each other through a ribbon cable and ribbon cable connectors, which fits in the space between combs and the lid

for most bee hive types. The temperature sensors are positioned in every second clearance between the combs. We prefer cable connectors over approaches requiring soldering skills, so that the setup can be installed and repaired by beekeepers with little technical experience. Therefore we use ribbon cable connectors and pins to connect the temperature sensors to the ribbon cable. To connect the ribbon cable with the microcontroller outside the hive, we use the offcuts of the robust DS18B20 cable. The single point load cell is mounted within a frame of aluminium u-profiles and square-type profiles under the hive and connected to a HX711 multiplier positioned on the circuit board in the waterproof box. The BME280 sensor is also connected to this circuit board. It can be positioned in the clearance between two frames, or in the bee spaces underneath or above the frames. The sensor measurements are taken every 5–10 s. Our default data transmission protocol is wireless LAN, but the FiPy microcontroller also supports Bluetooth, LoraWan, Sigfox, and LTE, which we might support in future. The data is transferred to a central server and stored in an Influx database, a database optimised for time series. Manuals for the installation of the sensors are available as well as the source code for the microcontroller [11,12].

3.2. Beekeepers' inspection data

Beekeepers were asked to use a webapp to document their observations and their activities. We use a fork of the open source beekeeping platform BEEP [13]. This platform also provides visualisations of sensor measurements for the beekeepers and supports collaboration groups between beekeepers. Beekeepers can create a digital hive in BEEP for every hive at their apiary and unique keys are being used to connect sensor devices. To document their observations and activities for a certain bee hive, beekeepers can fill out inspection lists. In the original BEEP webapp, beekeepers can use default inspection lists or choose to create their own custom lists. In our BOB fork, beekeepers are asked to use our specific inspection list designed for the scope of the Bee Observer project. It includes information about the development state of the colony, e.g. the presence of brood in different stages, the presence of queen cells, the size of the colony and the activity and behaviour of bees, information about food supply and honey harvest and any signs of diseases. Beekeepers can also write down custom comments, for which most of them use the German language. The data is stored in a MySQL database. The inspection database includes inspections on hives, that have sensor setups, but also hives without any connected sensors. Overall, our data includes information from more than 2600 inspections. The inspection information in the web app are organised in categories. We already integrated the most relevant information from the inspection data with the sensor data. The format of the inspection information is described in Table 2. For the raw inspection data we supply an additional readme file containing all relevant information on the data format.

3.3. Sampling strategy and participation of beekeepers

The data was collected within the citizen science project BeeObserver. Citizens were involved in all project phases: the definition of research goals, hardware and software development, data collection and evaluation. For the data collection, beekeepers played a central role. In 2019, we organised several workshops to teach beekeepers about the sensor setup and the installation process. We also encouraged beekeepers to support each other with the installation and repairs. With the beginning of the covid-pandemic in 2020, workshops had to be canceled and sensors were posted to beekeepers. As in many citizen science project, there was no sampling strategy such as a location based selection of honey bee colonies. However, it was ensured that beekeepers from cities and rural areas participated, that beekeepers using different types of hives could participate in the project and that different beekeeping practices are represented in the data. We gave detailed instructions to the beekeepers explaining how to setup

the sensors in their hives, to ensure reproducibility. However, we did not give any instructions on beekeeping practices, as we wanted to have different beekeeping practices represented in our data.

3.4. Data preparation

Additional to the raw data we provide the sensor data in intervals of one minute, one hour and one day. The raw data is stored with intervals between 5 and 10 s (*In Zenodo: bob_raw_data.zip*) and the one-minute interval data is calculated deriving the median of all available measurements for the one-minute periods (*In Zenodo: bob_publication_data.zip* → folders named “unprocessed” and “processed” → “_m” for one-minute interval data). We chose to apply this aggregation, as a one-minute interval in our experience is sufficient for further analysis tasks, and to reduce the size of the data and to simplify its handling, with the positive side-effect that the number of outliers from sensor failures is also reduced. On request, we can provide subsets of the high-frequency raw data. The one-hour and one-day aggregates are later calculated using averages of the data with a one-minute interval (*In Zenodo: bob_publication_data.zip* → “processed folder” → “_h” in file name for hourly interval, “_d” in file name for daily interval). The sensor data is stored in an InfluxDB version 1.7 [14] and we use the R programming language for all further data preparation [15]. We supply the data in an unprocessed form, as it was derived from a call to the InfluxDB, but we also supply a preprocessed version of the data, including following processing steps and routines: First, we manually inspected each single dataset to disregard periods of data for which the scale was not calibrated or the overall kit was not yet installed. We then run a routine to exclude weight values larger than 150 kg and smaller than -50 kg, temperature values larger than 85 °C and smaller than -40 °C, and relative humidity values outside the range between 0 and 100. These values occur due to sensor failures. Next, we calculate a processed weight value, excluding sudden drastic changes in the weight, which are usually induced by activities by the beekeeper. This can be helpful when looking at the overall annual development of honey bee colonies, while for other applications the raw weight information is to be preferred. We exclude sudden drastic weight changes by taking the derivative and excluding values larger than 0.3, corresponding to changes with more than 300 g weight difference per minute. Let us look at following sequence of weight values as an example: [1, 1, 1.2, 2]. The sequence of derivatives would be: [0, 0.2, 0.8]. Assuming that the unit of the sequence is kg, we would exclude the value of 0.8 (C.f. Table 3). By calculating the cumulative sum we obtain the processed absolute weight value, with the first value always being 0 kg.

We combine the sensor data with selected excerpts of the beekeepers' inspection data. For this we decided to focus on six states and events: the observation of queen cells, swarming, additional food supply, honey harvest, treatments, and the death of colonies. For each sensor measurement we add four columns for each of the six events: two columns specifying the timestamp of the last and next occurrence of the event, and two columns specifying the distance of the last and next occurrence of the event in time in seconds (*Processed data can be found in zenodo in the “processed” folder.*).

We organise the data in directories for the years from 2019 to 2022 (*In Zenodo: “years” folder*). Additionally, we provide subsets of data, focusing on two events: swarming or the death of colonies (*In Zenodo: “events” folder*). Therefore, we gather all sensor data collected before and after the certain event in one directory. For comparison with other colonies, we gather sensor data from the same time periods from all colonies for which the event was not observed. We therefore consider all time periods of the data frames when the event happened and obtain all data for colonies for which the same event was not recorded. We obtain the data from the database with a query for the median value of one minute. For the one-hour and one-day aggregated data we use the average of the intervals. Furthermore, we provide the raw data with its original measuring interval, which is usually 5 or 10 s. The unprocessed data, without cleaning routines applied as described above, is available for the one-minute interval data and the raw data only.

Limitations

The exact address of the apiaries is confidential information, as beekeepers might face theft and vandalism. We therefore rounded the coordinates of the locations to their first digits, corresponding to a precision of 5–7 km on the longitude scale for central Europe. Beekeepers cannot constantly monitor their beehives. Only during inspections, when the hive is opened, or when an experienced beekeeper observes the entrance of the hive, an assessment of the state of the colony can be done. The aim of our project was to collect sensor data under real conditions, and to comprise different beekeeping styles. Some beekeepers perform a weekly inspection in the main beekeeping season (in Germany from April to July or August), others limit the number of inspections to a minimum to avoid a disturbance of the colony. Therefore, the interval of inspections in our data is different for all colonies. Even when weekly inspections have been undertaken, it is possible that the beekeeper missed an important event, e.g. a swarming event. As the participation in the project was voluntary, there are a few colonies with no or limited meta-data.

Ethics Statement

We did not have a classical experimental setup and did not interfere with the beekeeping practices of the beekeepers. All beekeepers in Germany are required to follow rules and guidelines of the veterinary services.

Where applicable, we use the ARRIVE guidelines [16] to describe the methods for data collection.

An ethical approval was not required.

Data Availability

[Weight, Temperature and Humidity Sensor Data of Honey Bee Colonies in Germany, 2019 - 2022 \(Original data\)](#) (Zenodo).

CRedit Author Statement

Diren Senger: Conceptualization, Methodology, Software, Formal analysis, Investigation, Data curation, Writing – original draft, Writing – review & editing, Visualization, Project administration; **Clemens Gruber:** Conceptualization, Methodology; **Thorsten Kluss:** Conceptualization, Methodology, Visualization, Funding acquisition; **Carolin Johannsen:** Conceptualization, Methodology, Visualization, Funding acquisition.

Acknowledgements

This work would not have been possible without the contribution of hundreds of citizen scientists, e.g. beekeepers and makers.

This research was funded by the German Ministry for Education and Research (BMBF).

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.

References

- [1] D. Senger, C. Gruber, T. Kluss, C. Johannsen, Weight, Temperature and Humidity Sensor Data of Honey Bee Colonies in Germany, 2019-2022 [Data set], Zenodo, 2023, doi:10.5281/zenodo.10407693.
- [2] A.-M. Klein, et al., Importance of pollinators in changing landscapes for world crops, *Proc. R. Soc. B* 274 (1608) (2007) 303–313.
- [3] I.H. Williams, The dependence of crop production within the European Union on pollination by honey bees, *Agric. Zool. Rev.* 6 (1994) 229–257.
- [4] C.J.E. Schulp, S. Lautenbach, P.H. Verburg, Quantifying and mapping ecosystem services: demand and supply of pollination in the European Union, *Ecol. Indic.* 36 (2014) 131–141.
- [5] Keng-Lou James Hung, et al., The worldwide importance of honey bees as pollinators in natural habitats, *Proc. R. Soc. B* 285 (1870) (2018) 20172140.
- [6] S. Bruckner, et al., 2019–2020 Honey Bee Colony Losses in the United States: Preliminary Results, 2020.
- [7] F. Requier, et al., Trends in beekeeping and honey bee colony losses in Latin America, *J. Apic. Res.* 57 (5) (2018) 657–662.
- [8] W.W.P. Christian, et al., A survey of managed honey bee colony losses in the Republic of South Africa–2009 to 2011, *J. Apic. Res.* 53 (1) (2014) 35–42.
- [9] Zee Romée van der, et al., Managed honey bee colony losses in Canada, China, Europe, Israel and Turkey, for the winters of 2008–9 and 2009–10, *J. Apic. Res.* 51 (1) (2012) 100–114.
- [10] C. Johannsen, D. Senger, T. Kluss, A DIY sensor kit, Gaussian processes and a multi-agent system fused into a smart beekeeping assistant, in: 2020 16th International Conference on Intelligent Environments (IE), 2020, pp. 92–99.
- [11] BeeObserver-project, Instruction Manuals, 2023 Available from <https://hiverize.org/english-instructions/> 21-10-2023.
- [12] BeeObserver-project, FiPy Beehive Sensor Code, 2023 21-10-2023. Available from. <https://github.com/Hiverize/FiPy/tree/master>.
- [13] M. Schoneman, et al., BEEP: Digital Tools for Beekeepers, 2022 Retrieved from <https://beep.nl/home-english> 21-10-2023.
- [14] Influx, InfluxDB 1.7 Documentation, 2023.
- [15] R. Core Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, 2021 Retrieved from <https://www.R-project.org/> 21-10-2023.
- [16] C. Kilkenny, W.J. Browne, I.C. Cuthill, M. Emerson, D.G. Altman, Improving bioscience research reporting: the arrive guidelines for reporting animal research, *J. Pharmacol. Pharmacotherapeut.* 1 (2) (2010) 94–99.