Addressing RFID Misreadings to Better Infer Hive Activity

Ferry Susanto1,2,3, Thomas Gillard1,4, Paulo de Souza Jr.1, Benita Vincent1, Setia Budi1,5,6, Auro Almeida7, Gustavo Pessin8, Helder Arruda8, Raymond N. Williams1,5, Ulrich Engelke1, Peter Marendy1,5, Pascal Hirsch1,5, Jing He2,3

1Data61, Commonwealth Scientific and Industrial Research Organisation, Sandy Bay, TAS 7005, Australia
2Nanjing University of Finance and Economics, Nanjing Shi, Jiangsu Sheng, China
3Swinburne University of Technology, Hawthorn, VIC 3122, Australia
4School of Life and Environmental Sciences, University of Sydney, NSW 2006, Australia
5School of Engineering and ICT, University of Tasmania, Sandy Bay, TAS 7005, Australia
6Faculty of Information Technology, Maranatha Christian University, Bandung 40164, Indonesia
7Land and Water, Commonwealth Scientific and Industrial Research Organisation, Sandy Bay, TAS 7005, Australia
8Swinburne University of Technology, Hawthorn, VIC 3122, Australia

Corresponding author: Jing He (e-mail: lotusjing@gmail.com).

ABSTRACT This paper proposes a method to address misreadings and consequent inadequacy of Radio-Frequency Identification (RFID) data for social insect monitoring. Six months worth of field experiment data were collected to demonstrate the application of the method. The data is transformed into a linear combination of Gaussian model and curve-fitted using evolutionary algorithm. The results show that the proposed method allows us to improve the quality of data that infer honey bee behaviour at the colony level.

INDEX TERMS Apis mellifera, RFID, Optimisation, Genetic algorithm, Curve fitting, Data quality

I. INTRODUCTION

Pollinators play an integral role in food production, responsible for 1/3 of all commercial crop pollination, and with an estimated value of up to USD$ 200 billion per annum [1], [2]. Honey bees, bumble bees, and some European stingless bees are exploited commercially as they are easily managed, and live in colonies with large numbers of individuals. Yet European honey bee (Apis mellifera) colony numbers in Europe and North America especially have been reported to decline over the last century [3]. Bee population decline has caused by the short-range of reach of the antennas, and through gates (e.g. hives or feeders). This is likely to be known to fail when collecting information on bees passing excessively human intervention. However, RFID technology is quickly. Readings can be be recorded constantly and without identifiable, by a small number of operators and relatively of bees in a single hive can be tagged so as to be individually. The main advantages of this technology are that thousands of bees in a single hive can be tagged so as to be individually identifiable, by a small number of operators and relatively quickly. Readings can be be recorded constantly and without excessive human intervention. However, RFID technology is known to fail when collecting information on bees passing through gates (e.g. hives or feeders). This is likely to be caused by the short-range of reach of the antennas, and
bees flying too fast in and out of the hives and feeders. Furthermore, the orientation and spatial positioning of bees as they enter hives could also cause misreadings due to the polarisation of some RFID tags. A number of studies have reported the success rate of such systems when reading insect RFID tags, with results varying depending on, for example, experimental setup, reader capabilities, RFID tag (i.e. antenna) size, duty cycle, and the power of the electromagnetic signal. Here, we categorise the performance of the reader as: low (i.e. less than 80% success readings) [26], medium (i.e. between 80% and 90%) [21], [27], or high (i.e. greater than 90%) [12], [18] read efficiency. As a consequence of a low reading success rate, it becomes difficult to interpret what behaviour an individual bee was involved in and the duration of that bee being inside and/or outside the hive.

Other concerns associated with tagging small insects are the additional weight of the tag, and the alterations necessary to the hive entry to accommodate readers and antennas. Because most social insects have a short adult lifespan, typically from a few days to a couple of months, studying them with the assistance of tags requires regular visits to colonies. As a consequence hives are opened regularly, changing the internal environment, stressing the colony as a whole and killing some of the insects in the process. Some insects are extremely sensitive to odours emitted by glues, and therefore insects fitted with tags may be attacked by other individuals.

Despite these limitations, RFID tagging of insects is becoming the most practical available tool to investigate individual behaviour in a colony on a large scale. Using RFIDs is now more popular and, despite missing some readings, the behaviour of the colony can be reasonably well captured in those experiments using electronic tags.

In this paper, we estimate the number of bees engaged in three different behaviours (e.g. by the entry, short mission, and foraging) by applying a classification method to the entire colony’s activity data as tagged bee pass readers mounted at the hive entrance. This work addresses the key limitation of one popular method in the electronic tagging of social insects: lost readings do not allow accurate interpretation of individual behaviour.

The main contribution of this paper is the development of a method that allows the estimation of how many active bees in a hive will likely be foraging at a given moment (e.g. 3:30pm) or within a given time period (e.g. between 10 am and 2 pm). Here, “foraging” refers to those activities bees undertake in order to search for and collect resources for the colony, mainly food resources such as nectar and pollen, as well as hive resources including water and resin. The activities include searching for new food sources as well as exploiting current ones that the bee knows about, either as a result of previously visiting the source or by observing a waggle dance performed by another bee that has visited the source.

The paper is structured in the following manner: Section 2 describes the methodology and Section 3 presents the experimental results obtained. A discussion and conclusion will be provided in Section 4 and Section 5, respectively.

II. MATERIALS AND METHODS

A. FIELD EXPERIMENT

The dataset utilised for this work was obtained from a field experiment conducted in Tasmania, Australia. We set up four beehives with a radio-frequency identification (RFID) reader installed at each hive’s entrance, as illustrated in Figure 1a. We visited the hives on a regular basis (e.g. once or twice a week) to tag bees with RFID tags (Section II-B). Using this setup, the bee passes through the hive entrance and is detected at a particular reader (Figure 1b); the individual detected, and the date and time of detection are recorded. The data is organised in individual daily CSV files, based on UTC time.

B. BEE TAGGING

Adult worker bees were tagged at the hive using 2.5 × 2.5 × 0.4 mm RFID tags (Hitachi Chemical, Japan) secured to the thorax using cyanoacrylate super glue (Cyberbond LLC, Batavia, Illinois, USA). Each tag weights 2.4mg, one-third of an adult honey bee’s maximum foraging weigh. Live bees were restrained against the honeycomb using a modified dissection probe, by applying gentle pressure between the thorax and abdomen. The tag was applied on the thorax between the wings, ensuring that both wing pairs were free before releasing the bee. The bee was then observed to be able to fly prior to proceeding to the next bee. Each tag is coded in hexadecimal format with a unique bee identification number, consisting of a range of parameters including the country in which the experiment is taking place, the hive number within an apiary the bee originated from, and the bee’s species, strain, and caste.

C. CLASSIFICATION OF BEE BEHAVIOURS

Bee activity is, for the purpose of this paper, the detection of a bee fitted with an RFID tag passing through the entry. Bee behaviour is the interpretation of what the bee was actually doing. The assessment of what behaviour a bee was exhibiting usually needs to take into account several successive RFID detections.

The daily activity of bees is greatly affected by external factors, especially the weather, and in particular temperature, precipitation, solar radiation, and wind speed [28]. If, for example, the hive becomes too hot, bees can leave the hive and “beard” at the entrance, or use their wings to ventilate at the entry of the hive. If the temperature is too low (i.e. typically below 10°C) bees will not leave the hive, and instead form a cluster on the comb of the brood nest to maintain the optimal brood temperature of 34.5°C. Bees do not fly during storms or during high wind speeds. If the external temperature is mild and on a sunny, calm day, bees will likely be found foraging in large numbers.

Either by observation or experience, honey bee activity data can be classified here into four behaviour categories:
By The Entry (BTE): Bees classified as being "by the entry" are those with successive detections of the same bee by an RFID antenna at a maximum time interval between successive readings of less than three minutes. For example, a bee could be by the entry for 30 minutes, and it will be classified as BTE provided successive readings are within three minutes or less. This behaviour is usually associated with hive maintenance, including cleaning and control of hive temperature, defense, or after returning from a foraging trip [29]–[31].

Short Mission (SM): Bees engaged in short missions are those with successive detections intervals between three and six minutes. This means the bees left the entry for a period of time no longer than six minutes. Bees engaged in short missions are those believed to be making short orientation or defecation flights, inspecting the surroundings, or engaged in defense activity [32].

Foraging (FG): Bees will be classified as foraging when the gap between successive detections is longer than six minutes. During the day, a bee may be detected many times and, in most cases, the time intervening between the first and last detections of the day will be considered foraging, except when successive readings indicate the bee is by the entry or on short missions. Bees with recorded first and last detections will only be considered as foraging between sunrise and sunset. Foraging periods almost certainly incorporate periods of time when bees return from the field and stay inside the hive before going out again. Foraging is a crucial behaviour of bees and can be associated with different roles: scouts, which spontaneously search for new food/water sources; exploiters and water carriers are individuals that make repetitive flights to food and water sources, respectively; recruits are individuals that search for food sources with a prior awareness of the approximate location of the source after observing a waggle dance [32]–[34]. In principle, using RFID technology, bees should be detected at every instance of leaving and returning to the hive and this would provide some insight into the duration bees spent foraging and how long they stay inside the hive before leaving again. However, misreadings of RFID systems make this task practically impossible. A way to overcome this difficulty is to associate the sporadic detection of bees to a foraging behaviour. Therefore, the bee will be engaged in foraging activities for a long period of time, comprising several missions. Rather than recording each mission as a discrete event, the overall behaviour is defined as a foraging role.

Departed bees (DB): Bees that leave the hive and never return, either because they die or because they swarm (including absconding) [35], [36]. Swarming was not observed in our hives during the experiment.

Table 1 presents a summary of bee behaviours as described above. For the purpose of this work, we also performed data curation to filter out erroneous data in accordance with empirical study based on field observations and an initial investigation of the data. One of the main issues relates to continuous readings with extremely short time intervals (according to our rules, this is classified as BTE). This can happen when a dead bee, with its RFID still attached, is located within the reading range of the reader. To overcome this, we configured the software in such a way that BTE reads with a duration of more than 30 minutes were discarded. Similarly, FG durations of more than six hours were omitted, as this is most likely attributable to instances of missed readings.

The classification criteria for bee behaviour proposed above can be altered to accommodate other users’ needs, without further limitations to the model implemented in the current work. For example, if a beekeeper or an entomologist understands their bees are at the entry for no longer than two minutes, it is possible to change the model to incorporate this observation. These behaviours are described for the European honey bee, *Apis mellifera*, and are not necessarily the same for other bee species.
The behaviours described above are applied to the recordings of bee activities to identify the behaviour of each individual bee. After this step, a collective distribution for each bee behaviour is generated. The diurnal distribution of all of the inferred behaviour categories over the whole dataset resembles normal distributions (Figure 4). The resulting histogram is least-squared curve fitted using a Gaussian model (Section II-E).

The proposed method considers the overall activity as a linear combination of each type of possible activity (i.e. Gaussian Mixture Model) related to bee behaviour. Therefore, the linear combination of Gaussian curves can be written as:

\[ G_{ALL} = \alpha G_{BTE} + \beta G_{SM} + \theta G_{FG} \]  

where each component of this equation is a Gaussian curve \( G \) expressed as:

\[ G(x, BKG, I, T_{\mu}, T_{\sigma}) = I e^{-\frac{(x-T_{\mu})^2}{2T_{\sigma}^2}} + BKG \]  

where \( x \) is the data point (time of day in this case) to be estimated, \( BKG \) is a given background, \( I \) is the intensity, and \( T_{\mu} \) and \( T_{\sigma} \) are the mean and the standard deviation of the distribution respectively. The parameters \( (\alpha, \beta, \theta) \) represent the relative number of bees involved in different behaviours within the colony, calculated using the area under the curve for different Gaussian Probability Density Functions (PDFs).

For the purpose of this work, Gaussian parameters to be curve-fitted are as follows:

i) **Background-effect (BKG):** The data ‘normalisation’ which ensures that the PDF to be curve-fitted complies with the shape of a distribution. This is needed because brief visits to the hive entry (BTE and SM) occur regularly at night time. They are associated with bee defense or bees working to better climatise the colony (e.g., temperature or moisture control). Such events are considered a background (BKG) activity and are homogeneously distributed during the entire day and night. An example of such phenomena can be observed in Figure 4, where BTE detections occur between the hours of sunset and sunrise.

ii) **Intensity (I):** The height of the PDF indicating the overall probability of a particular activity taking place.

iii) **The average of time in a day (T_{\mu}):** The time of day in which a particular bee activity is most likely to occur (highest \( I \)).

iv) **The standard deviation of time of day (T_{\sigma}):** The spread of the PDF of bee activity in a day.

Curve fitting is performed with the experimental data to determine the parameters of the Gaussian curves and their relative contribution to the overall distribution of behaviour. Once the curve fitting is achieved, two key questions can be answered:

1. How many bees are performing a given behaviour at a given moment of the day? This is determined by the relative intensity of each curve at the moment of interest.
2. How many bees are performing a given behaviour during a given period of the day? This is determined by calculating the area under the curves during the period of interest.

### E. CURVE FITTING USING GENETIC ALGORITHMS

Genetic Algorithm (GA) is a meta-heuristic method to generate a near-optimal solution for an optimisation problem by evolving a pre-defined genetic representation (i.e. chromosome design), using natural selection process (e.g. selection, crossover, and mutation), towards a better solution. The following sub-sections discuss the requirement to perform GA in detail: (i) data notation; (ii) parameter initialisations and constraints; (iii) chromosome design; and (iv) fitness function.

1) **Data Notation**

First of all, we discuss the procedure we use to formalise the problem (Figure 4) in a mathematical way, and the notations to be used within the following sub-sections. The dataset is divided into two levels:

i) **Bee behaviour.** Let \( A = \{a_1, a_2, a_3\} = \{BTE, SM, FG\} \) be a list of distinct behaviours with

<table>
<thead>
<tr>
<th>Bee Behaviour</th>
<th>Summary description</th>
</tr>
</thead>
<tbody>
<tr>
<td>By the entry (BTE)</td>
<td>Successive readings of the same bee within less than three minutes indicates that this particular bee is near the RFID reader.</td>
</tr>
<tr>
<td>Short mission (SM)</td>
<td>Individual bee engaged in an activity of short duration (i.e. between three and six minutes). For example, making orientation or defecation flights and inspecting the hive surroundings.</td>
</tr>
<tr>
<td>Foraging (FG)</td>
<td>Individual bee engaged in activities to search for or exploit food/water sources. Example bee roles undertaking such activities are: scout, exploiter, recruit, water carrier.</td>
</tr>
<tr>
<td>Departed bee (DB)</td>
<td>The last detection of a bee in its lifetime.</td>
</tr>
</tbody>
</table>

**TABLE 1:** Categorisation of bee behaviours using RFID data. These behaviours reflect the duration of and time between consecutive readings and are used to report the most relevant results.
Based on the categorisation procedure above, we can now calculate the 'importance' of sections. Such representations will be used in the following subsections.

Since this work is based on a 'data-driven' modelling process, it is necessary to calculate the 'importance' of each datum in order to compute the mean \( \mu \) and standard deviation \( \sigma \) of a particular activity’s occurrence within a day. Therefore, the significance (denoted by the 'weight' \( W = \{w_1, w_2, \cdots, w_j, \cdots, w_J\} \)) of each datum (i.e. time index in day \( T \)) corresponds to the data availability \((d_{i,j})\) and is addressed by utilising the weighted mean \( \mu^* \) and standard deviation \( \sigma^* \) equation as below:

\[
\mu^*(V, W) = \frac{\sum_i^N u_i \cdot v_i}{\sum_i^N u_i} \tag{3}
\]

\[
\sigma^*(V, W) = \left[ \frac{\sum_i^N u_i \cdot (v_i - \mu^*)^2}{\sum_i^N u_i} \right]^{\frac{1}{2}} \tag{4}
\]

where \( V \) is a list of data with each datum denoted using \( v_i \) (i.e. \( V = \{v_1, v_2, \cdots, v_n, \cdots, v_N\} \)).

2) Initial Parameters Estimation and Constraints

The \( BKG \)-effect is estimated using the mean value of data \( D_i \) (i.e. \( i^{th} \) activity \( A \)) that holds the minimum Coefficient of Variation (CV) of the first \( n_{first} \) and last \( n_{last} \) data within time of day:

\[
\arg \min_{n_{first}, n_{last}} CV( \{D_i\}_{n_{first}} \cup \{D_i\}_{n_{last}} ) \tag{5}
\]

where \( n_{first} \) datum \( \{D_i\}_{n_{first}} = \{D_{i,j} : y \in \mathbb{Z} \land y \leq n_{first} \} \) and \( n_{last} \) datum in time-of-day \( \{D_i\}_{n_{last}} = \{D_{i,j} : y \in \mathbb{Z} \land n_{last} \leq y \leq J \} \). Thus, the \( BKG \) at activity \( i \) is obtained by:

\[
BKG_i = \mu( \{D_i\}_{n_{first}} \cup \{D_i\}_{n_{last}} ) \tag{6}
\]

and its constraint:

\[
C(BKG_i) = \sigma( \{D_i\}_{n_{first}} \cup \{D_i\}_{n_{last}} ) \tag{7}
\]

where \( \mu() \) and \( \sigma() \) corresponds to the minimised Equation 5.

Then, let time \( T_{first,last} = \{t_z \in \mathbb{Z} \land n_{first} < y < n_{last} \} \) be a sub-set of \( T \) and its corresponding datum \( U_i \) = \{\{D_{i,j} \} : y \in \mathbb{Z} \land n_{first} < y < n_{last} \}. \) The remaining parameters for individual Gaussian \( G_i \) of distinct activity \( a_i \) are estimated in the following:

\[
\begin{align*}
I_i & = \mu(\{u : u \in U_i \cup u \geq U_{i,unq,(n-2)}\}) \\
(T_\mu)_i & = \mu^*(\{T_{first,last}, U_i\}) \\
(T_\sigma)_i & = \sigma^*(\{T_{first,last}, U_i\})
\end{align*}
\]

\[
\begin{align*}
I_i & = \sigma(\{u : u \in U_i \cup u \geq U_{i,unq,(n-2)}\}) \\
(T_\mu)_i & = 30 \text{ min} \\
(T_\sigma)_i & = \sigma^*(\{T_{first,last}, U_i\}) \div 2
\end{align*}
\]

where \( u \) denotes each datum within \( U_i \) and \( U_{i,unq,(n-2)} \) is the third largest 'unique' value within \( U_i \) (represented using an order statistic). Lastly, note that the lower and higher boundary (search space) for the optimisation is in the form:

\[
C_{lo} = Estimation - Constraint \\
C_{hi} = Estimation + Constraint \tag{8}
\]

3) GA’s Chromosome Design

Also called chromosome encoding and decoding, is a crucial step required to quantify the problem into an 'individual' for the optimisation process. In this work, we designed a single individual using the following approach [37], [38]:

Based on Equation 1, a complete distribution of the data consisting of \( I = 3 \) activities (BTE, SM, and FG) and 4 parameters \( (BKG, I, T_\mu \text{ and } T_\sigma) \) as in Equation 2) is required to generate one single Gaussian distribution. Therefore, in our case, one individual will consist of 12 elements \( (3 \text{ activities } \times 4 \text{ parameters}) \) with values between 0 and 1. In order to decode the value of a particular element within the individual, the following equation is employed:

\[
p(x, C_{lo}, C_{hi}) = C_{lo} + x \cdot (C_{hi} - C_{lo}) \tag{9}
\]

where \( x \) and \( p \) are the encoded and the decoded value of a particular Gaussian parameter respectively; \( C_{lo} \) and \( C_{hi} \) are the constraint values calculated from the previous section (Equation 8). Figure 2 demonstrates the design of a single individual with a decoding example for a parameter \( I \) of the Gaussian distribution \( G_i \).
FIGURE 2: An illustration of the chromosome design utilised in this work, where each element within the individual holds a value between 0 and 1, and B denotes the background (BKG). An encoding and decoding example for the intensity (I) parameter of distribution GI is also presented. In this case, assume that we have element I1 with value 0.7615 (encoded) within the individual which is equivalent to 205.32 (decoded) after applying Equation 9.

4) Fitness function
The quality of a particular individual is assessed by minimising the sum of the chi-square (χ²) function [37] for different bee activities:

$$fitness = \sum_{i} \chi_{i}^{2}$$

(10)

$$\chi_{i}^{2} = \frac{1}{J - N_{p}} \sum_{j} \frac{(d_{i,j} - G_{i,j})^{2}}{d_{i,j} + 1}$$

(11)

where J is the number of elements in a day (see Table 2); Np is the number of parameters to be optimised (four in this case); and \(G_{i,j}\) is the estimated value using \(G_{i}\) at time \(t_{j}\). Note that the + 1 within the denominator on the right-side of Equation 11 is employed to avoid a divide-by-zero error, which could occur if the number of data points was extremely low.

III. EXPERIMENTAL RESULTS
This section provides the results obtained from the experiment. Figure 3 depicts the overview throughout the entire experiment, commenced on April 2nd and ended on November 11th, 2014. During the period, a total of 2,425 RFID tags were deployed; however, only 1,101 bees fitted with RFIDs were detected at least once after being tagged. Such a phenomenon can be explained by: (i) misreadings of the RFID system; (ii) the tag was not fitted properly so that the bee was able to remove it; and (iii) tags lost during the tagging process resulting from environmental conditions (e.g., on a windy day). The number of bees alive increases on the day our team members make a field visit to tag bees, and it is shown that approximately 30% of the tagged bees will be detected at least once on the following days.

Although data was collected between April and November, the analyses in the following sections were only undertaken on data collected between May and October to accommodate for the build up and decline of tagged bee numbers in the instrumented hives as shown in Figure 3, allowing for more robust results.

A. BEE ACTIVITY
Eight months of experimental data from 1,101 bees were recorded in CSV files. Bee detections were classified using the previously established rules (Section II-C) in order to assess each bee’s behaviour at a given time. Once the behaviour was determined, the data were grouped according to the time of day the activity occurred.

The daily distribution of bee behaviour over the period of eight months is shown in Figure 4. The figure shows that bee activity starts to increase at approximately 7 am, and declines at approximately 8 pm. During that period, bees are most active between 12 pm and 1 pm. It is also observed that there are some detections before and after 7 am and 8 pm respectively which have been classified as by the entry. The analysis in this work removes Australian daylight savings time for consistency across data.

Figure 5 shows when bee activity is assigned to the category ‘departed bee’. Departed bees are those that left the hive and never returned. Occurrences after sunset and before sunrise are likely to be associated with bees that died inside the hive and were transported out by worker bees.

B. CURVE FITTING
Figure 6 shows the result for the curve fitting of each behaviour (by the entry, short missions, foraging) for the entire period of the experiment. The sum of each behaviour overlaps relatively well with the overall bee activity (black dots). It suggests that bees start foraging at approximately 7 am and finish at 8 pm. Around noon (between approximately 12 pm and 1 pm), most bees are involved in foraging role (e.g. exploiter, recruit, scout, water carrier); followed by by-the-entry activities (e.g. hive defense, temperature control); and lastly, on short missions (e.g. orientation flights, wandering around the nest).

The proportion of bees involved in different behaviours varies relative to time of day. Therefore, a normalisation
FIGURE 3: Summary information of field visits for tagging bees and daily number of bees alive throughout the experiment. In this case, the lifespan of an individual bee (referred as ‘alive’ bee) is estimated from the first day it was tagged until the very last day of its detection.

FIGURE 4: Cumulative plot of six months experimental data, illustrating bee behaviour distributions throughout the day.

FIGURE 5: Daily distribution of events associated with departed bees. These are events where bees are detected for the last time. The events occurring between dusk and dawn are likely to be of bees that died inside the hive and were transported out of the hive by other workers.
FIGURE 6: A visualisation produced by the curve-fitting program developed for this work that demonstrates the Gaussian PDF of distinct bee activities in a day. The x-axis shows the time of day and the y-axis is the frequency/count of activities. The dots (in different colours corresponding to the histogram values in Figure 4) represent the data to be curve fitted with BTE, SM, FG denoted in blue, green, red respectively. The solid lines are the curve fitted Gaussian PDFs for isolated activities ($G_{BTE}$, $G_{SM}$, $G_{FG}$) and the combined ones ($G_{ALL}$ in black dashed-line) where the curve fitting used the optimisation approach proposed in this work (Section II-E).

FIGURE 7: The proportion of bee behaviours relative to time of day.

of the curve fitted Gaussian PDF (Figure 6) is depicted in Figure 7. The normalised curve reveals that approximately 80% of the workers within the colony are engaged in BTE and ≈ 20% are in SM during early morning (before sunrise) and late at night (after sunset).

C. BEE BEHAVIOUR

Once each individual bee behaviour has been initially interpreted and following the curve fitting process, we can determine the proportion of bees performing a specific task at a given moment of the day or during a specific period of the day. If we consider the cohort of bees fitted with electronic tags to be representative of the entire bee population in a hive, it is possible to estimate how many bees would be engaging in, for example, foraging activities.

The result shows the proposed method is reasonable and the area under each curve should represent the number of bees undertaking distinct behaviours throughout the day. Monthly proportions (from May to October) of bees foraging, in short missions or by the entry are shown in Figure 8 and values of the Gaussian parameters summarising this data are given in Table 3. Figure 8 shows an overall increase of foraging behaviour in the long term, and a decrease in by the entry behaviour and short missions. Also, it shows that the data availability dropped significantly in July, most probably due to the markedly decreased temperature during the winter
FIGURE 8: The proportion of bee colony behaviours for different months. The histogram (in yellow) presents the data availability for each month.

<table>
<thead>
<tr>
<th>Month</th>
<th>BTE</th>
<th>SM</th>
<th>FG</th>
<th>BKG (Counts)</th>
<th>I (Counts)</th>
<th>$T_{\mu}$ (hh:mm)</th>
<th>$T_{\sigma}$ (minutes)</th>
<th>Area (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>May</td>
<td>0.27</td>
<td>0.00</td>
<td>0.00</td>
<td>61.85</td>
<td>18.93</td>
<td>12:50</td>
<td>87</td>
<td>23.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.00</td>
<td>0.27</td>
<td>0.00</td>
<td>114.60</td>
<td>207.54</td>
<td>13:05</td>
<td>90</td>
<td>70.5</td>
</tr>
<tr>
<td>Jun</td>
<td>0.21</td>
<td>0.00</td>
<td>0.00</td>
<td>53.98</td>
<td>32.98</td>
<td>12:55</td>
<td>73</td>
<td>22.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.00</td>
<td>0.19</td>
<td>0.00</td>
<td>44.62</td>
<td>10.74</td>
<td>13:05</td>
<td>77</td>
<td>71.9</td>
</tr>
<tr>
<td>Jul</td>
<td>0.21</td>
<td>0.00</td>
<td>0.00</td>
<td>53.98</td>
<td>32.98</td>
<td>12:55</td>
<td>73</td>
<td>20.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.00</td>
<td>0.07</td>
<td>0.00</td>
<td>80.03</td>
<td>184.88</td>
<td>12:53</td>
<td>69</td>
<td>75.1</td>
</tr>
<tr>
<td>Aug</td>
<td>0.21</td>
<td>0.00</td>
<td>0.00</td>
<td>53.98</td>
<td>32.98</td>
<td>12:55</td>
<td>73</td>
<td>12.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.00</td>
<td>0.07</td>
<td>0.00</td>
<td>80.03</td>
<td>184.88</td>
<td>12:53</td>
<td>69</td>
<td>84.7</td>
</tr>
<tr>
<td>Sept</td>
<td>0.28</td>
<td>0.07</td>
<td>0.00</td>
<td>37.08</td>
<td>8.03</td>
<td>12:38</td>
<td>140</td>
<td>18.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.37</td>
<td>0.04</td>
<td>0.00</td>
<td>13.49</td>
<td>4.11</td>
<td>13:21</td>
<td>125</td>
<td>9.2</td>
</tr>
<tr>
<td>Oct</td>
<td>0.37</td>
<td>0.04</td>
<td>0.00</td>
<td>13.49</td>
<td>4.11</td>
<td>13:21</td>
<td>125</td>
<td>88.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.00</td>
<td>0.04</td>
<td>0.00</td>
<td>147.31</td>
<td>124.85</td>
<td>12:48</td>
<td>124</td>
<td>88.7</td>
</tr>
</tbody>
</table>

TABLE 3: Gaussian parameter values for each month of data reported in Figure 8. The ‘Area(%)’ column indicates the percentage of detections recorded of bees undertaking particular behaviours, relative to the total detections for that period.
period in Tasmania (as further discussed in the Discussion section).

Furthermore, based on Table 3, the standard deviation of the Gaussian parameter \( T_r \) during winter period (i.e. June to August) is lower compared to other months (e.g. \( T_r < 80\text{min} \) overall). This was probably caused by (i) lower temperatures during the winter months that reduces bee activity, and (ii) the fact that the sun rises later and sets much earlier than in other seasons [28], [39].

### IV. DISCUSSION

RFID systems where readers are installed in the field with limited power availability, operating with high reading frequency to capture every potential bee tagged leaving or returning the hive, in confined spaces like a beehive entry and with tags small enough to fit on bees are operationally challenging. Missing readings were inevitable and this fact makes the interpretation of each individual bee behaviour very difficult. This work addresses this problem by developing a method that assigns a behaviour for each bee based on roles and extrapolates that behaviour for a cohort of bees doing the same activity.

Foraging behaviours were restricted to daylight hours, typically between 5 am and 8 pm; our data correlates well with nature, as bees will not forage when ambient temperature or solar radiation levels are too low. Additionally, the proportion of bees undertaking various behaviours varies over the course of the day. An increased probability of bees undertaking short missions and remaining by the entry is seen during the hours when bees are not actively engaged in foraging activities. This is likely due to forager role plasticity resulting in the reallocation of foragers to defensive or hygienic roles, or simply a matter of proportions altering as numbers of bees engaged in BTE and SM increases relative to FG (Figure 7).

Not only was activity variable over the daily cycle, but our six months of data include a distinct shift in behaviour over the long term (Figure 8). The bee tagging period was commenced in April and terminated in October. In April and November tagged bee numbers within the hives were much lower than in the intermediate months as a result of tagged populations becoming established and dying out respectively (Figure 3). As a result of this, only data from May to October was included in the analysis. A significant decrease in readings was observed in July (Figure 8) due to predominantly cold weather, higher rainfall, and decreased solar radiation when compared to the months of September through April. Furthermore, a proportional increase in foraging behaviours was observed over the course of the experiment. This may be due to an improvement in operator skill over time, resulting in more efficient tagging and a reduction in tagging-associated mortality, or adjustment of the colony to the colder winter temperature after the initial shock in July. Increased activity is expected, and observed, concurrent with the increase in temperatures into the spring months.

Bee behaviour classification can be interpreted on the basis of the frequency of readings. By considering issue of misreadings in the RFID system, we are able to define the levels of certainty for the behavioural characterisations in our proposed model (Table 4).

High frequency reads are associated with the constant presence of a bee by the colony entry. This leads to a high degree of certainty about the assignment of a behaviour of a bee to be ‘by the entry’ or ‘short mission’. A departed bee is also very clear as the last recording of a bee could be confidently assigned as a bee that never returned to the hive.

The foraging behaviour of bees is, by its very nature, complex. Bees could leave and return to the hive in missions lasting as long as an hour (and possibly longer) several times in a day. If the RFID system does not miss any readings, we would be able to confirm with absolute certainty when the bee left the hive, and when the same bee returned. We would also be able to say how long the bee was inside the hive between outdoor missions and for how long each mission lasted. When a single reading is missing, however, it becomes almost impossible to determine the bee’s behaviour at a given time with absolute certainty. With our approach we are able to estimate with some degree of confidence when the bees were engaged in foraging activities. This is possible because our rules are defined in such a way that single readings exclude BTE behaviours, and long durations

<table>
<thead>
<tr>
<th>Bee Behaviour</th>
<th>Activities</th>
<th>Duration</th>
<th>Certainty</th>
</tr>
</thead>
<tbody>
<tr>
<td>By the entry</td>
<td>High frequency readings, defense, air conditioning of hive</td>
<td>( x \leq 3\text{min} )</td>
<td>High</td>
</tr>
<tr>
<td>Short mission</td>
<td>Orientation flights, walking around the hive</td>
<td>( 3\text{min} &lt; x \leq 6\text{min} )</td>
<td>Medium</td>
</tr>
<tr>
<td>Foraging</td>
<td>Collecting/depositing food/water, scouting for new food resources</td>
<td>( x &gt; 6\text{min} )</td>
<td>Low</td>
</tr>
<tr>
<td>Departed bee</td>
<td>Last detection of an individual, e.g. dead bee</td>
<td>–</td>
<td>High</td>
</tr>
</tbody>
</table>

**TABLE 4:** Summary of the proposed bee behavioural model and its level of certainty for insects monitored under the current empirical study. The ‘Threshold’ column gives the cut-off points for two successive readings within the classification procedure for the raw bee detection data; whilst, the ‘Duration’ column indicates the range of valid bee behaviour durations. For instance, BTE durations of more than 30 minutes will be omitted.
FIGURE 9: Screenshot of a visualisation tool developed to analyse individual bee activity using the bee behaviour classification described in Section II-C. Each row represents a day that the bee was active (i.e. detected at least once in that day); and the horizontal axis is the time of day, showing the times at which different bee behaviours occurred. Within each day, the times of sunrise (yellow) and sunset (green) are also indicated. For this example, the coloured bars represent: (i) blue bar in the first row, the day the individual was tagged; (ii) red, single detection in that day; (iii) light blue, by the entry; (iv) orange, short mission; and (v) purple, foraging period.
between readings exclude SM behaviours, leaving only FG.

Figure 9 illustrates an example of an effort logging record throughout the lifetime of an individual bee. In this instance, this bee was tagged on 13\textsuperscript{th} August 2014 at Hive 001 and last detected on 31\textsuperscript{st} August 2014. It is very likely that this bee was tagged when it was very young because: (i) a ‘single detection’ is observed on August 14\textsuperscript{th} and 17\textsuperscript{th}; and (ii) the bee started to forage on 18\textsuperscript{th} August, despite the fact that it was tagged on August 13\textsuperscript{th}. This bee is very likely to be a forager (e.g., scout, recruit, exploiter) throughout its lifetime. Note that the ‘partitioned’ foraging period indicates that there were detections with more than six minutes intervals between successive readings within its foraging period. This could, for example, be explained by assuming that the bee was either exploiting food sources during these intervals or that it was in the hive depositing nectar or pollen before undertaking further foraging activities.

V. CONCLUSIONS

The proposed method allows for a robust use of data from social insect monitoring based on RFID devices. Misreadings, which are common in RFID-based experiments, can be better managed by combining insect behaviour with activity data. The classification proposed in this paper (Section II-C) is based on results reported in the literature and on observation of our bees. However, depending on different bee species or other factors, it is possible to change the software configuration (i.e., model parameters) associated with bee activity.

Under traditional techniques, once the detection of the insect fails, RFID data become useless. The proposed method addresses this problem by assigning a given behaviour for each tagged insect, then combining results for the entire tagged population using curve fitting based on genetic algorithms.

One limitation of this work is the inability to determine the number of data required in order to have a good representation of the results. For example, the curve fitting of monthly data (Section III-C) does not include April and November because those months do not have enough bee activity data for the curve-fitting purposes. Another limitation is the fact we have a small number of hives which limits the replicability of the experiments. And finally, the ideal calibration is to have another independent method to determine the activity of the bees. While we have scales in some of the hives, a camera, and an image processing technique could be used to determine when the bees leave and return as a mean to calibrate our results. Such an approach will be used in the future.

The method proposed in this paper could be used by other research methods using RFID to study social insects to better analyse the RFID data and overcome the issue of missed readings which are commonly experienced with electronic tagging. This has an important and positive implications for those using RFID data in insect behaviour modelling. Study in the design of environmental sensor networks which involves animal-borne instruments as mobile sensor nodes [40] could also benefit from this work.

ACKNOWLEDGEMENT

This work is supported by National Natural Science Foundation of Innovative Research Groups Science Foundation of China (51221004), ARC DECRA and ARC Discovery projects (DE130100911, DP130101327), the NSFC funding (61332013), the International Science and Technology Cooperation Projects (No. 2016D10008, 2013DFG12810, 2013C24207), the Municipal Natural Science Foundation of Ningbo (No.2015A610119), the Guangzhou Science and Technology Project under Grants (2016201604030034), the Major Projects of Natural Science Research in Jiangsu Higher Education Institutions (NO. 14KJA520001), Jiangsu Production and Research Project (NO. BY2015010-05), the Natural Science Foundation of Zhejiang Province (NO. Y16F020002), the Ningbo Natural Science Fund (NO. 2015A610119), International Science & Technology Cooperation Projects of Ningbo (NO. 2016D10008) and the project of research and development of intelligent resource allocation and sharing platform for marine electronic information industry (NO. 2017GY116).

The authors also would like to thank Geoff Allen and Stephen Quarrell from Tasmanian Institute of Agriculture for their support to the project.

REFERENCES
