

Application of Data Layering in Precision Beekeeping: The Concept

Nikolajs Bumanis, Olvija Komasilova, Vitalijs Komasilovs, Armands Kviesis, Aleksejs Zacepins

Faculty of Information Technologies
Latvia University of Life Sciences and Technologies
Jelgava, Latvia
Corresponding e-mail: nikolajs.bumanis@gmail.com

Abstract—The monitoring and predictions of various multi-level states of honeybee colonies are performed using emerging Internet of Things technologies and data processing methods. It is become common to use multiple sensors and devices providing multi-modal data to monitor a single activity. Modern data analysis and data processing procedures include a step of data fusion in order to provide more accurate input data. This, however, requires implementation of machine learning and large data sets, whereas gathering large data sets of real time and observation data is a common problem for small to medium size apiaries. This why there are no real implementation of data fusion method in precision beekeeping field. The aim of this paper was to introduce the concept of data layering, which aims to solve the global precision beekeeping problems without implementation of machine learning. The concept was demonstrated within the scope of foraging optimization problem using three data sets: flowering calendar data, rainfall precipitation data and bee activity data.

Keywords—*precision beekeeping, data fusion, data layering*

I. INTRODUCTION

Along with development of Internet of Things (IoT) novel approaches for optimization of beekeeping related activities were introduced [1]. Application of IoT solutions in the framework of beekeeping lead to introduction of a new subcategory of precision agriculture called precision beekeeping [2]. The precision beekeeping (PB) in essence is an apiary management strategy that uses the principles of data collection and data processing in order to minimize resource consumption and maximize productivity of bees [3].

The main objectives of precision beekeeping are to monitor, detect and predict such colony states as queenlessness, broodlessness, pre-swarming, swarming and after swarming. There are many studies [4]–[7] aimed to achieve these objectives. The common point of these studies relates to the processing of a particular time series data related to colony state, e.g. using temperature to detect swarming or brooding.

In modern apiaries, data are collected through use of wireless network technologies [8], [9]. Firstly, this introduces the limitations of sensory data itself, e.g. data imperfections, data granularity, data inconsistency. Secondly, integration of sensors may pose a technical challenge when environmental variables, such as terrain, internet or mobile network coverage, become the deciding factors. The solution to both of these problems is to use

data from multiple sources, including multi sensor data, and non-sensory data from third party organizations, i.e. meteorological stations, map services. However, any additional data source may provide the data of different format, or in different time intervals, thus leading to the challenge of processing this data as a singular unit. This leads to the problem of organizing raw data before it is processed. There are various methods to organize raw data, the most common is concatenation of the two time series data sets of same format. The advanced methods are built around data fusion that use neural networks at its core [10], [11]. There are multiple advantages of using data fusion methods to prepare raw data sets, e.g. elimination of imperfections and achieving data consistency. The biggest advantage of data fusion is possibility of fuse multi modular and heterogeneous data sets. According to Data-Information-Knowledge-Wisdom [12] hierarchy data are the part of the first step of achieving a complete understanding of particular topic. However, data fusion methods require detailed data sets, typically spatial-temporal data, for training a neural network and, therefore, are scarcely used [13] in the framework of PB. One of the steps of a complete data fusion method is data layering – organizing the data in the form of the layers, where each layer corresponds to particular time series, e.g. one layer per month/week/day/hour/minute/second. As a part of data fusion method, it should be possible for data sets, processed using data layering approach to achieve similar objectives as data fusion. In the field of PB these topics include global objectives like monitoring the overall colony health, determining the level of stress, optimizing harvesting procedures, selection of the most effective apiary location.

The aim of this research was to propose the concept of data layering in precision beekeeping field, in order to achieve understanding about beekeeping related global objective.

This research was performed in the framework of HIVEOPOLIS project, that aims to make technologies available to honeybees that are naturally inaccessible for them (internet, databases, satellite data, robots) and to feed information collected by bees through these channels back to us researchers and also to other hives.

II. DATA ORGANIZATION IN PRECISION BEEKEEPING

Modern precision beekeeping incorporates the use of multisensory system and decision support systems [3], [14] to

provide real-time information and expert grade support for beekeepers. There are various studies [15], [16] depicting the basic precision beekeeping – monitoring one or multiple particular physical variable, and use mathematical models to determine the current colony state.

There are currently various data sources in the field of PB. These data sources can be divided into three groups, based on the layer of observation: [4], [17]: apiary, colony and individual bee-related levels.

Apiary level data includes meteorological and video observation data. Main meteorological parameters are wind and precipitations. Apiary management software tend to use [13] third party weather stations to acquire these parameters. Spatial observations allow identifying the type of fields and crops [18], [19] that are usable for bee foraging. The sources of apiary level data are wide angle video cameras, local apiary weather stations, public weather stations and satellite imagery services

Colony level data includes temperature, humidity, weight, sound, vibration and video data. Temperature, weight and humidity are the most popular parameters [6], [20], whereas swarming and colony death are the most popular [5], [21] monitor objectives. Researchers use these parameters to determine such beehive states as broodlessness, intensive brood rearing, swarming, pre-swarming and after swarming, overheating, as well as colony death. Sound and video data are also used to determine air and noise pollution. Researcher use sound and vibrations [7] to determine such beehive states as queenlessness, broodlessness, swarming (including prior and after swarming periods), beehive overpopulation and colony death. The sources of colony level data are temperature sensors, humidity sensors, weight sensors, noise and sound receivers, mono and multispectral video cameras.

The individual bee-related monitoring addresses such objectives as bee counting, i.e. bees going in/out of hive [22], number of infested bees [23], and bee activity [24]. The sources of individual bee-related data are mainly mono and multispectral video cameras.

In general, colony state can be determined and predicted using such physical variables as temperature inside hive, ambient temperature, humidity, weight and audio data. In addition, such parameters as difference between temperature inside hive and ambient temperature, and month that represents a season. These variables create data sets that, although may be imperfect, are enough to get information related to basic hive and colony states. Implementation of video monitoring lead to additional observatory data.

Introducing fuzzy logic helps in prediction of colony states in short term. Kviesis et al. [14] implemented fuzzy logic principles in early identification of honeybee colony states such as colony death and swarming using temperature inside hive, ambient temperature and a season.

Komasilova et al. [25] performed the study aimed to help beekeepers to select the optimal place for apiary location. Authors proposed the heat map generation model that calculates optimal apiary location based on the potential amount of resources that can be harvested by honeybees (*Apis mellifera*). The following data sets were used as layers during the study:

satellite image of a map, foraging area, roads, and fields. In addition, parameters such as minimal amount of honey required for local needs and average amount of honey collected by an apiary were taken into calculation.

One of the scarcely available application of data fusion in precision beekeeping can be seen in the recent study performed by Braga et al. [13]. Authors trained, validated, and tested three well-known and distinguished classification algorithms (k-Nearest Neighbors, Random Forest, and Neural Networks) and used real datasets from 6 apiaries, 27 Western honeybee (*Apis mellifera*) beehives monitored over three years (2016, 2017 and 2018). The data sets consisted of internal temperature and beehive weight as well as weather data, provided by 6 meteorological stations, including temperature, dew point, wind direction, wind speed, rainfall, and daylight. As a step of processing the raw data, Braga et al. used simple data fusion techniques to fuse data from internal and external sensors using variable timestamp.

III. CONCEPT OF DATA LAYERING

The basis for data layering derives from the ability to correlate different time series data if these sets share a common parameter, like timestamp. Data layering is applied in accordance to combination of data sets, either related or unrelated with each other, whereas selection of data sets is determined by potential distinct objectives that can be achieved, e.g. bee foraging optimization, colony health prediction, automation of management procedures.

The concept of data layering is further described around bee foraging optimization objective.

A. Preparation of data sets

The goal of data preparation is to create a distinct relation between previously unrelated data sets. Data sets are created by pre-processing raw data using data analytics procedures, e.g. removing outliers, normalizing values. Data sets for global objectives are created mainly using apiary and colony level data, whereas internal hive management objectives also implement individual bee-related data.

The important step of preparation is identifying a common parameter or longer period time series identifier, e.g. month, week, and day. As PB application is dependent on location that can influence the flexibility of monitor/prediction system, data sets may contain seasonal data. The data sets that cover longer periods are flowering calendars, precipitation calendar, and bee activity calendar.

B. Data layering

Data required to optimize bee foraging includes information and knowledge about a region – location, terrain, climate, local nectar and pollen plants, coverage of internet or mobile network; apiary size, bee species and their activities; weather conditions, especially precipitation and wind. There are also environmental conditions that affect a plant in terms of nectar production – air temperature, relative air humidity, rain, wind, sun lightning and it is intensity [26], [27]. There are typically regional beekeeping organizations, like Latvian Beekeeping Association and The

British Beekeepers Association, responsible for delivering an overview about local plants, their flowering periods and productivity. This information is aggregated into flowering calendars. However, there are multiple challenges regarding the use of data provided by flowering calendars: there are no flowering calendar that would cover all plants, therefore a local flowering calendar must be chosen; calendars provide different data, i.e. flowering per month, flowering per season, amount of nectar and pollen in abstract scale (low, medium, high) or real values. In essence, there are no flowering calendar that could be used without modification.

For concept demonstration purposes the flowering calendar [28] created in the framework of SAMS project was selected. Calendar provides the information about tree flowering in Ethiopia, depicting the tree species, plant family, growth form, amount of nectar and pollen using four values – none, “~” low quantities, “o” average quantities and “+” high quantities, and flowering status per month using three values – none, “o” flowering and “+” peak flowering period. In order to create a data set a normalization was applied:

- Nectar, n :
none: 0; low quantities: 0.5; average quantities: 1; high quantities: 1.5.
- Pollen, p :
none: 0; low quantities: 0.5; average quantities: 1; high quantities: 1.5.
- Flowering, fl :
none: 0; flowering is happening: 1; peak flower period: 2.

Using normalized values, the plant richness was calculated:

- PR – plant richness, percentile;
- $MaxPR$ – maximal plant richness, decimal;
- t – time period, month/week/day/hour/minute.

$$PR_t = \frac{fl_t \times (1+n_t) \times (1+p_t)}{MaxPR} \times 100 \quad (1)$$

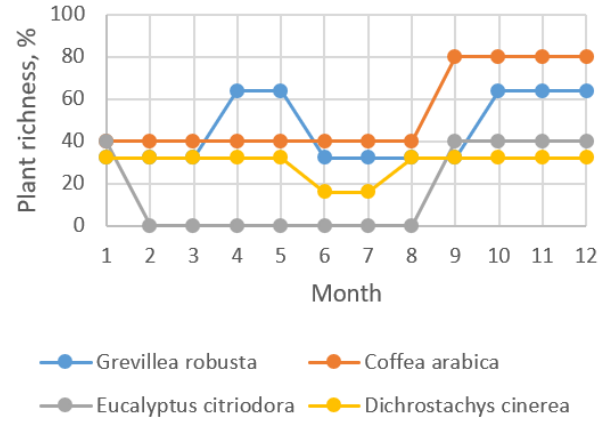
In the following example t equal to one month and maximal plant richness is calculated by using the highest values: $n = 1.5$, $p = 1.5$ and $fl = 2$. In the framework of flowering calendar for Ethiopian nectar plants $MaxPR$ equals to 12.5.

For demonstration purposes four plants with distinct flowering and production features were selected: *grevillea robusta*, *coffea Arabica*, *eucalyptus citriodora* and *dichrostachys cinerea*. The PR of the selected plants (Fig. 1) constitute a one data set. The graph represents the dynamic of plant richness value during a year. As can be seen on Fig. 1. silky oak (*grevillea robusta*) has two periods of having plant richness value equal to 65% in April to May and October to December, and having plant richness value of 30% in other months. In comparison, Arabian coffee (*coffea Arabica*) has a stable value of 40% from January to August, and spiking to 80% from September to December. We propose to assume each particular

data entry in data set as separate data layer. This includes each particular plant.

Fig. 1. Plant richness data for four selected plants.

Representing part of this data set, depicting two of the more distinct parameters using area chart, two overlapping data layers



of this data set can be seen on Fig. 2. Based on this information alone it can be assumed, that fields with Arabian coffee will yield more honey in total; therefore, it would be economically efficient to change apiary location in proximity of Arabian coffee fields.

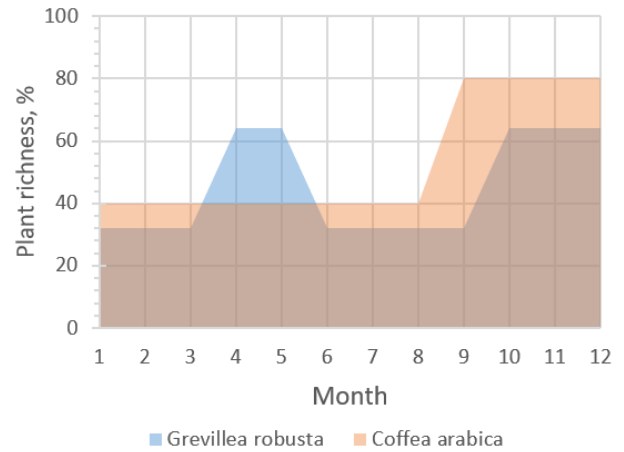


Fig. 2. Data layers of two selected plants.

The second data layer to showcase the importance of layering is rainfall precipitation layer. Is it proven [27] that rain is one of the determining factors of foraging efficiency. Similarly to flowering calendars, rainfall precipitation data is regional. The line graph representing monthly rainfall precipitation data typically has the form of one or multiple parabolas with highest points during rainfall seasons. On the basic level the data layer of rainfall precipitation can be based on historical data of previous years, whereas accurate calculations will require the use of rainfall prediction algorithms and forecasts. For demonstration purposes a monthly rainfall precipitation data in the form close to parabola was obtained

(Fig. 3). In order to correlate plant flowering and rainfall precipitation data sets, a converting was applied to represent rainfall precipitation data in percentage values, instead of absolute numbers of precipitation amount (otherwise represented in centimeters or inches). From this moment onwards, the value of each data layer is considered a general “data layer value” in range of 0-100.

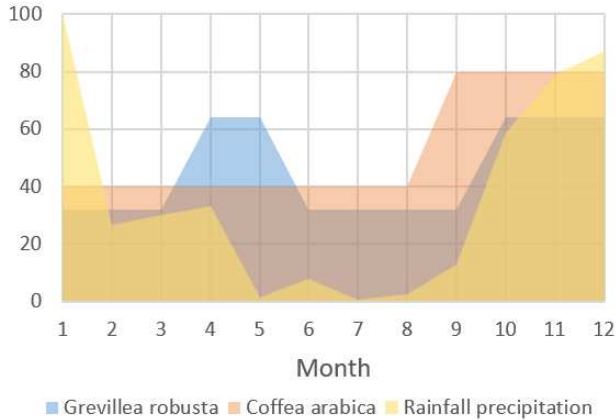


Fig. 3. Rainfall precipitation data layer.

After adding rainfall precipitation layer to this demonstration, it became harder to determine, if Silky oak or Arabian coffee yields more. Arabian coffee yields more on average, however, heavy rainfall during October to December may result in overall reductions of its fields. In practice, though, rainfall precipitation must be analyzed with the focus on particular region, as even small countries have different rainfall distribution; therefore, the data layers must be based on location.

Another important data layer is bee activity. Bee activity is seasonal, however, it differs depending on the average summer and winter temperatures of the country. The main activity period includes brood rearing, swarming, active foraging, etc. It is hard to evaluate each activity with a value, thus for demonstration purposes scale from 1 to 10 was selected. This scale represents the hypothetical bee activity during a particular month, including the probability of previously mentioned events (Fig. 4.).

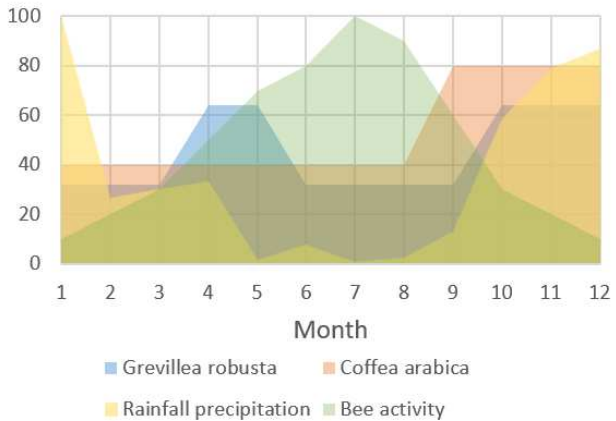


Fig. 4. Bee activity data layer.

For scaling purposes, these values were converted to percentages, and the data layer of bee activity was added to demonstration. It is common to refer to a local beekeeping calendar as a guideline, as they tend to provide both information about bee activities and necessary beekeeping activities [29].

C. Analysis

The current stage of the concept does not include particular formulas to calculate the final value of overlapping layers. However, the result is assumed to be in the range of 0 to 1, and be usable for decision support systems. The evaluation of the result must be problem specific. It is assumed that neighborhood statistics may be used for calculations and fuzzy logic for analysis.

The focal areas of demonstrated plant foraging represented on Fig. 5 and Fig. 6.

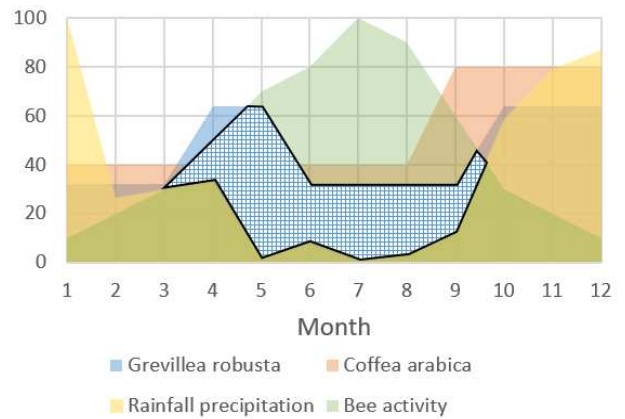


Fig. 5. Focal area for Silky oak foraging.

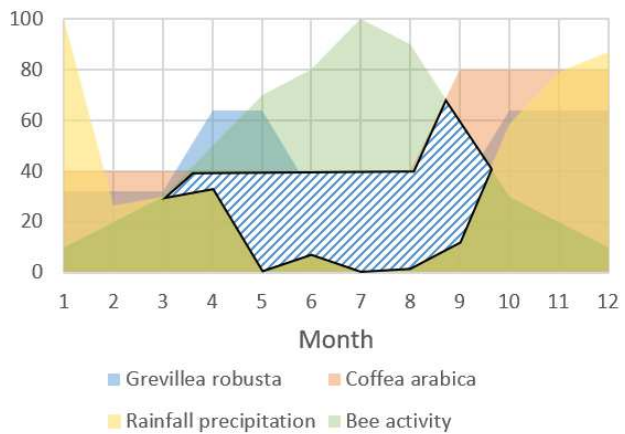


Fig. 6. Focal area for Arabian coffee foraging.

Based on the visual analysis and selected data layers it can be concluded that three scenarios may be implemented in descending order based on foraging efficiency:

- To maximize foraging efficiency place apiary in the reach of Silky oak for period from March to May. After May change apiary location to forage Arabian coffee.

- In case beekeeper is incapable of changing location of apiary mid-season, place apiary in the reach of Arabian coffee. It will provide more yield overall, and will allow bees to store more honey for winter seasons.
- For convenience purposes only, apiary may be placed in the reach of Silky oak for the duration of all season. This will yield the least amount of honey, and may endanger bees in winter season if left unattended.

IV. CONCLUSION

The trends in precision beekeeping are moving towards detection and prediction of global, sophisticated problems, such as overall apiary health, foraging optimization, apiary location optimization. Other industries resolve such problems using machine learning and application of data fusion methods. Data fusion approach and methodology is a novel development turn for applications in precision beekeeping, as there are close to none studies using it. The low interest of data fusion methods for precision beekeeping in particular lies in the necessity of large sensory and observation data sets, that is not common for low to medium large apiaries. The proposed concept is based on the ideology of data fusion method and does not require implementation of machine learning, but still allows to find answer to such global problems as foraging optimization.

The limitation of proposed concept lies in the need to normalize or transform data sets to similar data types, in order to stack and correlate these data sets. The removal of this limitation, if possible at all, is currently under research.

ACKNOWLEDGMENT

Scientific research, publication and presentation are supported by HIVEOPOLIS project that is funded by the Horizon 2020 FET Programme of the European Union under grant agreement No. 824069.

REFERENCES

- [1] A. Zacepins, A. Kviesis, A. Pecka, and V. Osadcuks, "Development of Internet of Things concept for Precision Beekeeping," *2017 18th Int. Carpathian Control Conf. ICCCC 2017*, pp. 23–27, 2017, doi: 10.1109/CarpathianCC.2017.7970365.
- [2] J. Zacepins, A. Stalidzans, E. Meitalovs, "Application of Information Technologies in Precision Agriculture A. Zacepins, E. Stalidzans," *Proc. 13th Int. Conf. Precis. Agric. (ICPA 2012)*, 2012.
- [3] A. Zacepins, V. Brusbardis, J. Meitalovs, and E. Stalidzans, "Challenges in the development of Precision Beekeeping," *Biosyst. Eng.*, vol. 130, pp. 60–71, 2015, doi: 10.1016/j.biosystemseng.2014.12.001.
- [4] H. Human *et al.*, "Miscellaneous standard methods for *Apis mellifera* research," *J. Apic. Res.*, vol. 52, no. 4, 2013, doi: 10.3896/IBRA.1.52.4.10.
- [5] S. Ferrari, M. Silva, M. Guarino, and D. Berckmans, "Monitoring of swarming sounds in bee hives for early detection of the swarming period," *Comput. Electron. Agric.*, vol. 64, no. 1, pp. 72–77, Nov. 2008, doi: 10.1016/j.compag.2008.05.010.
- [6] E. Stalidzans and A. Berzonis, "Temperature changes above the upper hive body reveal the annual development periods of honey bee colonies," *Comput. Electron. Agric.*, vol. 90, pp. 1–6, 2013, doi: 10.1016/j.compag.2012.10.003.
- [7] M. Bencsik *et al.*, "Honeybee colony vibrational measurements to highlight the brood cycle," *PLoS One*, vol. 10, no. 11, Nov. 2015, doi: 10.1371/journal.pone.0141926.
- [8] A. Kviesis, A. Zacepins, M. Durgun, and S. Tekin, "Application of wireless sensor networks in precision apiculture," *Eng. Rural Dev.*, vol. 14, no. January, pp. 440–445, 2015.
- [9] D. S. Kridi, C. G. N. de Carvalho, and D. G. Gomes, "Application of wireless sensor networks for beehive monitoring and in-hive thermal patterns detection," *Comput. Electron. Agric.*, vol. 127, pp. 221–235, Sep. 2016, doi: 10.1016/j.compag.2016.05.013.
- [10] S. Beddar-Wiesing and M. Bieshaar, "Multi-Sensor Data and Knowledge Fusion - A Proposal for a Terminology Definition," *ArXiv*, Jan. 2020, [Online]. Available: <http://arxiv.org/abs/2001.04171>.
- [11] F. Castanedo, "A review of data fusion techniques," *Sci. World J.*, vol. 2013, 2013, doi: 10.1155/2013/704504.
- [12] G. D. Allen, "Hierarchy of Knowledge – from Data to Wisdom," *Int. J. Curr. Res. Multidiscip.*, vol. 2, no. 1, pp. 15–23, 2004.
- [13] A. R. Braga, D. G. Gomes, R. Rogers, E. E. Hassler, B. M. Freitas, and J. A. Cazier, "A method for mining combined data from in-hive sensors, weather and apiary inspections to forecast the health status of honey bee colonies," *Comput. Electron. Agric.*, vol. 169, no. September 2019, p. 105161, 2020, doi: 10.1016/j.compag.2019.105161.
- [14] A. Kviesis, V. Komasilovs, and O. Komasilova, "Application of fuzzy logic for honey bee colony state detection based on temperature data," *Biosyst. Eng.*, vol. 193, pp. 90–100, 2020, doi: 10.1016/j.biosystemseng.2020.02.010.
- [15] A. Kviesis and A. Zacepins, "System architectures for real-time bee colony temperature monitoring," *Procedia Comput. Sci.*, vol. 43, no. C, pp. 86–94, 2015, doi: 10.1016/j.procs.2014.12.012.
- [16] A. Zacepins, A. Pecka, V. Osadcuks, A. Kviesis, and S. Engel, "Solution for automated bee colony weight monitoring," *Agron. Res.*, vol. 15, no. 2, pp. 585–593, 2017.
- [17] A. Zacepins and E. Stalidzans, "Information processing for remote recognition of the state of bee colonies and apiaries in precision beekeeping (apiculture)," *Biosyst. Inf. Technol.*, vol. 2, no. 1, pp. 6–10, 2013, doi: 10.11592/bit.130502.
- [18] G. Atluri, A. Karpatne, and V. Kumar, "Spatio-temporal data mining: A survey of problems and methods," *ACM Comput. Surv.*, vol. 51, no. 4, pp. 1–37, 2018, doi: 10.1145/3161602.
- [19] P. Calatayud-Vernich, F. Calatayud, E. Simó, J. A. Pascual Aguilar, and Y. Picó, "A two-year monitoring of pesticide hazard in-hive: High honey bee mortality rates during insecticide poisoning episodes in apiaries located near agricultural settings," *Chemosphere*, vol. 232, pp. 471–480, 2019, doi: 10.1016/j.chemosphere.2019.05.170.
- [20] W. G. Meikle and N. Holst, "Application of continuous monitoring of honeybee colonies," *Apidologie*, vol. 46, no. 1, pp. 10–22, 2015, doi: 10.1007/s13592-014-0298-x.
- [21] D. S. Kridi, C. G. N. De Carvalho, and D. G. Gomes, "A predictive algorithm for mitigate swarming bees through proactive monitoring via wireless sensor networks," in *PE-WASUN 2014 - Proceedings of the 11th ACM Symposium on Performance Evaluation of Wireless Ad Hoc, Sensor, and Ubiquitous Networks*, 2014, pp. 41–47, doi: 10.1145/2653481.2653482.
- [22] A. E. Souza Cunha, J. Rose, J. Prior, H. M. Aumann, N. W. Emanetoglu, and F. A. Drummond, "A novel non-invasive radar to monitor honey bee colony health," *Comput. Electron. Agric.*, vol. 170, 2020, doi: 10.1016/j.compag.2020.105241.
- [23] K. Bjerge, C. E. Frigaard, P. Høgh Mikkelsen, T. Holm Nielsen, and M. Misbi, "A computer vision system to monitor the infestation level of Varroa destructor in a honeybee colony," *Comput. Electron. Agric.*, vol. 164, 2019, doi: 10.1016/j.compag.2019.104898.
- [24] T. N. Ngo, K. C. Wu, E. C. Yang, and T. Te Lin, "A real-time imaging system for multiple honey bee tracking and activity monitoring," *Comput. Electron. Agric.*, vol. 163, no. 1, 2019, doi: 10.1016/j.compag.2019.05.050.
- [25] O. Komasilova, V. Komasilovs, A. Kviesis, N. Bumanis, H. Mellmann, and A. Zacepins, "Model for the bee apiary location evaluation," *Agron. Res.*, vol. 18, no. Special Issue 2, pp. 1350–1358, 2020, doi: 10.15159/AR.20.090.
- [26] G. Hennessy *et al.*, "Gone with the wind: effects of wind on honey bee visit rate and foraging behaviour," *Anim. Behav.*, vol. 161, pp. 23–31, Mar. 2020, doi: 10.1016/j.anbehav.2019.12.018.

- [27] X. J. He, L. Q. Tian, X. B. Wu, and Z. J. Zeng, "RFID monitoring indicates honeybees work harder before a rainy day," *Insect Science*, vol. 23, no. 1. Blackwell Publishing Ltd, pp. 157–159, Feb. 01, 2016, doi: 10.1111/1744-7917.12298.
- [28] "Tree Flowering Calendar" https://wiki.sams-project.eu/index.php/Tree_flowering_calendar_Ethiopia (accessed Jun. 14, 2020).
- [29] "Beekeeping Calendar" <https://beekeepers.amazingbees.com.au/beekeeping-calendar.html> (accessed Jun. 18, 2020).