

ARTIGO ORIGINAL

Reconhecimento de padrões sazonais em colônias de abelhas *Apis mellifera* via clusterização

Felipe Anderson O. Maciel¹, Antonio Rafael Braga^{1,3}, Ticiania L. Coelho da Silva³, Breno M. Freitas² and Danielo G. Gomes¹

¹Programa de Pós-Graduação em Engenharia de Teleinformática, Grupo de Redes de Computadores, Engenharia de Software e Sistemas (GREat) - Universidade Federal do Ceará (UFC) - Campus do Pici and

²Setor de Abelhas - Departamento de Zootecnia - UFC - Campus do Pici and ³UFC - Campus Quixadá

{felipemaciel,antoniobraga}@great.ufc.br; {ticiania,freitas,danielo}@ufc.br

Submetido: 23/10/2018. Revisado: 05/11/2018. Aceito: 05/11/2018.

Abstract

As the main pollinating agent, bees are essential for the food production for humans and the ecosystems maintenance. Among the crops used for human consumption, 75% depend on pollination. In line with a current concern with bee survival, here we propose to identify patterns of *Apis mellifera* colonies to assist the beekeeper in the management and maintenance of their hives. Our method applies a clustering technique in two real datasets of hives in temperate climate with sensor data of temperature, humidity and mass. We used three datasets from the HiveTool.net portal; two of them divided in cold (autumn and winter) and hot (spring and summer) periods, and the third, for comparative purposes, divided into periods mixing cold and hot seasons: winter and spring, and summer and fall. From the application of the Calinski-Harabasz index and the K-means algorithm, we have identified coherent patterns associated to the transitions between the seasons. In addition, we can conclude that the strongest colony is most efficient in trying to maintain the microclimate of the hive during the winter.

Key words: *Apis mellifera*; Clustering; Data mining; Honey bees; Pattern recognition; Precision beekeeping.

Resumo

Na qualidade de principal agente polinizador, as abelhas são essenciais à produção de alimentos para o ser humano e para manutenção dos ecossistemas. Entre as culturas agrícolas utilizadas para o consumo humano, 75% dependem de polinização. Alinhando-se a uma preocupação atual com a sobrevivência das abelhas, este artigo visa identificar padrões de colônias de *Apis mellifera* a fim de auxiliar o apicultor no manejo e na manutenção de suas colmeias. Nosso método consistiu na aplicação de uma técnica de clusterização em dois *datasets* reais de colmeias em clima temperado com dados de temperatura, umidade e massa. Foram utilizados três *datasets* do portal HiveTool.net; dois deles divididos em período frio (outono e inverno) e período quente (primavera e verão) e o terceiro, para efeito comparativo, dividido em períodos mesclando estações frias e quentes: inverno e primavera, e verão e outono. A partir da aplicação do índice Calinski-Harabasz e do algoritmo K-means, identificamos padrões coerentes e associados às transições entre as estações do ano. Além disso, pudemos concluir que a colônia mais forte é mais eficiente ao tentar manter o microclima da colmeia durante o inverno.

Palavras-Chave: Abelhas; Apicultura de precisão; Clusterização; Mineração de dados; Padrões.

1 Introdução

As abelhas são consideradas os principais agentes polinizadores da maioria das espécies de plantas silvestres e culturas polinizadas por insetos (Kevan and Phillips; 2001; Klein et al.; 2007; Ollerton et al.; 2011). Cerca de 75% das culturas agrícolas utilizadas diretamente para o consumo humano dependem da polinização (Potts et al.; 2016). Na qualidade de principal agente polinizador, as abelhas são essenciais à produção de alimentos para o ser humano e para a manutenção dos ecossistemas.

No Brasil, 85 das 141 espécies de plantas cultivadas para uso na alimentação humana, produção animal, biodiesel e fibras dependem em certo grau da polinização animal (Giannini et al.; 2015). Em relação à produção agrícola de 2012, estimou-se o valor econômico da polinização para 44 culturas, que apresentam ganhos variados com a polinização animal, em aproximadamente 30% da produção total de 45 bilhões de dólares (Giannini et al.; 2015).

Abelhas *Apis mellifera* são as mais utilizadas em todo o mundo para a polinização de plantas cultivadas. A preferência por essa espécie se justifica pelo seu fácil manejo, tamanho de suas colônias, sua abundância em diferentes ecossistemas e pelo seu perfil generalista na busca de recursos (Potts et al.; 2010). Além disso a *Apis mellifera* desempenha uma importante função de produtora de mel e outros produtos apícolas. Em 2013, o valor da produção de mel no Brasil foi de 316 milhões de reais (Pires et al.; 2016).

Trabalhos recentes têm registrado reduções no número de espécies polinizadoras silvestres e domesticadas e nos serviços de polinização em todo o mundo (Potts et al.; 2010). Nos últimos anos, populações de abelhas da espécie *Apis mellifera* têm sofrido mortes em massa em regiões da Europa e da América do Norte devido ao fenômeno da desordem do colapso da colônia (*colony collapse disorder* – CCD) (Barron; 2015; Chauzat et al.; 2014; Gil-Lebrero et al.; 2016). Esse fenômeno se caracteriza pela rápida perda de abelhas operárias, evidenciada pelo enfraquecimento ou morte da colônia com excesso de crias, em comparação ao número de abelhas adultas; ausência de crias e abelhas adultas mortas dentro ou fora da colmeia; e invasão da colmeia por pragas (Pires et al.; 2016).

Ainda não se sabe ao certo ao que se deve a ocorrência do CCD, mas vários estudos têm sido conduzidos para discutir as causas e consequências desse fenômeno (Biesmeijer et al.; 2006; Oldroyd; 2007). Sabe-se, por exemplo, que há relação entre o CCD e fatores como mudanças climáticas, estresse, uso de pesticidas, exposição a patógenos, desmatamento, intensificação da agricultura e a perda associada de recursos florais (Barron; 2015; Goulson et al.; 2015; Potts et al.; 2010; Tscharrntke et al.; 2012), o que, consequentemente, pode comprometer seriamente os serviços de polinização (Bommarco et al.; 2011; Deguines et al.; 2014; Kremen et al.; 2002).

Apesar do CCD ter sido registrada em regiões da Europa e da América do Norte, também há registros de perdas de colônias de *Apis mellifera* africanizada no Brasil. Em 2012, por exemplo, uma longa seca atingiu a região nordeste, fazendo com que as

abelhas abandonassem parte das colmeias a procura de um novo ambiente com melhores condições. Consequentemente muitas colônias foram perdidas (Kridi et al.; 2014, 2016).

Para examinar as colônias em estados anormais, ou mesmo para uma vistoria de rotina, os apicultores recorrem à inspeção visual. Para isso, normalmente é necessário abrir as colmeias, remover os quadros e averiguá-los. Além de ser um processo invasivo, uma inspeção minuciosa demanda tempo, o que pode comprometer as funções de polinização e produção de mel. Há também o risco de abelhas serem esmagadas durante a movimentação dos quadros das colmeias. Além disso, muitas colônias são mantidas em apiários remotos ou rurais, e inspeções em tais locais exigem longos deslocamentos.

De maneira geral, o ciclo anual de colônias de abelhas em climas temperados pode ser dividido em dois períodos: a época das estações mais frias do ano, quando as abelhas ficam menos ativas; e a época das estações mais quentes do ano, quando há alta atividade das abelhas (Kviesis and Zacepins; 2016). Durante esses períodos, diversos estágios das colônias podem ser observados. Ter o conhecimento sobre qual estado uma colônia está em determinado momento, sem abrir a colmeia, possibilita ao apicultor administrar melhor seu apiário (Kviesis and Zacepins; 2016).

Existe um grande interesse no desenvolvimento de novos métodos não invasivos que possam contribuir para avaliar o estado de uma colônia de abelhas (Zacepins et al.; 2016). O monitoramento remoto de apiários pode auxiliar apicultores agregando valiosas informações sobre o estado e o comportamento das abelhas, além de poupar as abelhas de estresses desnecessários ou outras atividades não-produtivas (Zacepins and Karasha; 2012).

A literatura recente mostra a aplicação das redes de sensores sem fio, dos sistemas ciberfísicos e da Internet das Coisas (IoT) no monitoramento de colônias de abelhas (Sanchez et al.; 2015; Kridi et al.; 2014, 2016; Murphy et al.; 2016; Zacepins et al.; 2017) para o desenvolvimento da chamada *apicultura de precisão*, cujo objetivo central é maximizar a produtividade das colmeias e diminuir o consumo de recursos (Zacepins et al.; 2015). Os parâmetros das colmeias mais utilizados nesses sistemas são temperatura, umidade, massa, gases, áudio, vibrações, imagem e vídeo (Meikle and Holst; 2015).

Entende-se que o comportamento das abelhas precisa ser estudado de forma mais intensa, com coletas frequentes e análise dos dados, de modo que seja possível descrever o seu comportamento enquanto protagonistas da polinização. É necessário mais empenho no desenvolvimento e implementação de sistemas que possam identificar e alertar os apicultores sobre distúrbios no desenvolvimento de colônias de abelhas (Kviesis and Zacepins; 2016). Apesar do monitoramento remoto de colônias via redes de sensores ser uma estratégia bem difundida no âmbito da apicultura de precisão, nota-se uma certa carência de soluções automatizadas para extração do conhecimento dos dados das colmeias (Zacepins et al.; 2015). Uma direção interessante a seguir é a combinação de diferentes atributos no processo de análise dos dados, e a aplicação de

softwares de mineração de dados poderia resultar em conclusões inesperadas (Zacepins et al.; 2015).

A mineração de dados, também conhecida como descoberta de conhecimento em banco de dados, análise avançada de dados, ou como aprendizado de máquina, já produziu aplicações práticas em áreas como análise de exames médicos, detecção de fraude em cartões de crédito, predição do comportamento de consumo e de interesses pessoais dos usuários da Internet, e otimização de processos de fabricação (Chen et al.; 1996; Mitchell; 1999). Este método também é comumente utilizado na pesquisa científica moderna, inclusive na ciência biológica (Xu and Tian; 2015). A clusterização, ou agrupamento, como um elemento básico na composição da análise e mineração de dados, desempenha um papel expressivo nesse contexto (Xu and Tian; 2015).

Este artigo é uma extensão de trabalhos prévios dos autores (Maciel, Braga, Xavier, da Silva, Freitas and Gomes; 2018; Maciel, Braga, Silva, da Silva, Freitas and Gomes; 2018), com melhorias nos resultados e na contribuição e aprofundamento do referencial teórico e da metodologia.

1.1 Objetivo

O objetivo central deste artigo é identificar e caracterizar padrões de colônias de abelhas *Apis mellifera* de acordo com os dois períodos normalmente encontrados em um ciclo anual das colônias em clima temperado. Para cumpri-lo, estabelecemos as seguintes atividades, ilustradas no fluxograma da Figura 1:

- i. obter conjuntos de dados de temperatura, umidade e massa de colônias de abelhas *Apis mellifera* que compreendam um ciclo anual completo;
- ii. fragmentar os conjuntos de dados nos dois períodos normalmente encontrados no ciclo anual das colônias;
- iii. detectar e remover as anomalias dos dados;
- iv. normalizar os dados;
- v. reconhecer a quantidade ideal de padrões para cada período;
- vi. caracterizar e interpretar cada padrão encontrado.

2 Referencial Teórico

Nesta seção desenvolve-se a fundamentação teórica dos temas tratados neste trabalho. A Subseção 2.1 apresenta os conceitos referentes à apicultura, enquanto a Subseção 2.2 aborda os conceitos mais relevantes referentes à mineração de dados para o contexto deste trabalho. Na Subseção 3 são discutidos alguns trabalhos relacionados ao assunto.

2.1 Apicultura

A prática do homem de criar abelhas é tão antiga que não se sabe ao certo quando ocorreu o princípio da apicultura (Zogovic et al.; 2017). Nas primeiras abordagens do cultivo de abelhas, o homem chegava a destruir as colônias para poder ter acesso aos produtos gerados pelos insetos. Com o passar do tempo e a aquisição de conhecimento, as técnicas

de manejo melhoraram, e a apicultura passou a ter outros objetivos, como tirar proveito da função de polinização desempenhada pelas abelhas (Zogovic et al.; 2017).

Nos últimos 200 anos, a prática do cultivo de abelhas sofreu um desenvolvimento mais intenso, com o reconhecimento de propriedades importantes das colônias por parte dos apicultores e a modernização das colmeias (Zogovic et al.; 2017). Com isso, a apicultura se tornou mais eficiente, e a prática contemporânea passou a ser descrita como apicultura racional, na qual o apicultor retira os produtos das colmeias incomodando minimamente as abelhas (Zogovic et al.; 2017).

2.1.1 Apicultura de precisão

Ainda no século 20 surgiu o interesse de se monitorar e coletar dados de colônias de abelhas. Há registro de coleta de dados de temperatura e massa de uma colmeia por diversos dias entre 1907 e 1908 (Meikle and Holst; 2015). Atualmente, com a modernização da comunicação e dos sensores, e com a tecnologia da informação cada vez mais presente em todas as áreas do conhecimento, apicultores e pesquisadores podem monitorar remotamente vários aspectos físicos das colônias de abelhas (Meikle and Holst; 2015; Zogovic et al.; 2017). Essa é a premissa da apicultura de precisão.

A apicultura de precisão envolve basicamente as seguintes etapas: coletar individualmente os dados das colmeias, analisar a informação obtida e dar suporte à tomada de decisão para o gerenciamento das atividades em um apiário. Zacepins et al. (2012) definiram a apicultura de precisão como uma estratégia de gerenciamento de apiários baseada no monitoramento de colônias de abelhas para minimizar o consumo de recursos e maximizar a produtividade das abelhas. Essa abordagem visa reduzir o desperdício de recursos e o estresse das abelhas causado por atividades desnecessárias (Zacepins et al.; 2015). Uma vez que os sensores são instalados, as colmeias podem ser monitoradas sem perturbação, inclusive durante os períodos quando as inspeções invasivas são contraindicadas, como durante o inverno (Meikle and Holst; 2015).

2.1.2 Termorregulação

A preocupação com as inspeções invasivas em períodos como o inverno se dá pelo fato desse processo ser prejudicial à termorregulação. A termorregulação é um recurso utilizado pelas abelhas para controlar o microclima no interior da colmeia, principalmente no ninho (Almeida et al.; 2006; Kridi et al.; 2014). A faixa normal de temperatura dentro de uma colônia de abelhas vai de 33 °C a 36 °C (Kleinhenz et al.; 2003; Petz et al.; 2004), e as crias são bastante sensíveis a qualquer variação fora deste intervalo (Stabentheiner et al.; 2010).

Quando a temperatura está abaixo de 33 °C, as abelhas operárias se concentram em volta da área de cria e aumentam a temperatura corporal através da endotermia sob demanda (Abou-Shaara et al.; 2017; Stabentheiner et al.; 2010). Elas também utilizam a estratégia do aquecimento ativo, por meio da movimentação dos músculos de voo torácicos (Abou-Shaara et al.; 2017). Já quando a temperatura está acima de 36 °C, as abelhas operárias amenizam o calor

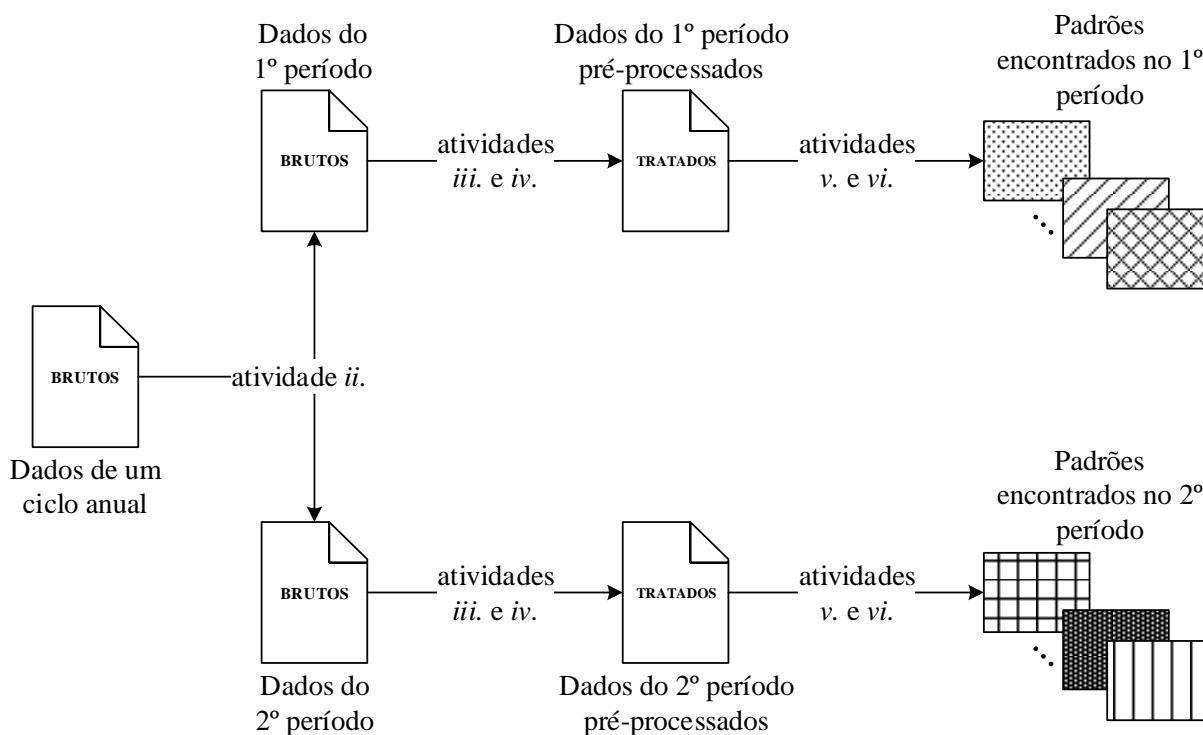


Figura 1: Diagrama do método proposto.

vibrando rapidamente as asas em posições específicas da colmeia para gerar uma ventilação efetiva (Abou-Shaara et al.; 2017; Southwick and Moritz; 1987) e coletando água para realizar-se o resfriamento por evaporação (Nicolson; 2009). Além disso, algumas abelhas também podem sair da colmeia para evitar o superaquecimento (Abou-Shaara et al.; 2017).

2.1.3 Atributos do monitoramento na apicultura de precisão

Nota-se a influência que a temperatura tem sobre o estado de uma colônia de abelhas. Através da informação deste atributo, apicultores e pesquisadores têm tentado detectar estados de colônias como o aumento do consumo de alimento, o início da produção de crias, o estado pré-emigratório e a morte da colônia (Zacepins et al.; 2015).

Juntamente com a temperatura, a umidade relativa do ar é outro fator relevante monitorado na apicultura de precisão. Dentro da colmeia a umidade é controlada em grande parte pelas abelhas operárias (Li et al.; 2016), uma vez que este atributo é importante para o desenvolvimento das crias, mais especificamente para a eclosão dos ovos. O intervalo ótimo da umidade interna para eclosão normal dos ovos vai de 90% a 95% (Abou-Shaara et al.; 2017). Este parâmetro também está fortemente associado ao resfriamento por evaporação, um recurso fundamental para as abelhas controlarem a hipertermia da colmeia (Ostwald et al.; 2016). Já a umidade externa à colmeia desempenha um papel fundamental para as abelhas quando em climas quentes (Abou-Shaara et al.; 2017). As abelhas resistem melhor às condições de forte calor

quando a umidade relativa está aproximadamente em 75%; abaixo de 50% a sobrevivência das abelhas é impactada negativamente, e chega em um nível crítico quando a umidade fica abaixo de 15% (Abou-Shaara et al.; 2012).

A massa da colmeia também é um atributo comumente levado em consideração na apicultura de precisão. Ela é composta pela soma da massa da caixa da colmeia, dos favos com o estoque de alimento e das abelhas, e pode ser utilizada para monitorar diferentes estados da colônia ou para identificar diferentes eventos (Zacepins et al.; 2015). Entre esses eventos estão: a ocorrência do fluxo de néctar durante as épocas de forrageamento, o consumo do estoque de alimento durante as épocas sem forrageamento, a ocorrência de abandono da colmeia (devido à diminuição da massa), além de ser possível estimar o número de abelhas envolvidas na atividade de forrageamento (Zacepins et al.; 2015).

Outros métodos utilizados na apicultura de precisão consistem no monitoramento de áudio, vídeo, concentração de gases e das vibrações das colmeias. Diversas técnicas de processamento de áudio têm sido desenvolvidas para analisar o comportamento das abelhas, mas somente os pesquisadores têm conseguido aplicá-las com sucesso (Zacepins et al.; 2015). A falta de aplicação no âmbito industrial pode ser explicada pela complexidade na interpretação dos sons devido ao componente estocástico do zumbido emitido por uma colônia (Zacepins et al.; 2015). A partir do vídeo monitoramento do fluxo de abelhas na entrada da colmeia é possível ter um indicador da condição da colônia (Zacepins et al.; 2015). O monitoramento de

gases, como oxigênio e dióxido de carbono, pode gerar informações importantes a respeito do metabolismo das abelhas. No entanto, sensores de gás têm alto custo e precisam de fluxo de ar controlado, o que pode influenciar o microclima da colmeia (Zacepins et al.; 2015). Já através dos sinais de vibrações das colmeias é possível, por exemplo, prever o abandono da colmeia com vários dias de antecedência (Bencsik et al.; 2011).

A etapa da análise dos dados dos atributos coletados geralmente é o principal obstáculo na abordagem da apicultura de precisão e ainda está em um estágio inicial de desenvolvimento (McBratney et al.; 2005; Zacepins et al.; 2015). Alguns resultados já foram obtidos sobre as possíveis relações entre dados capturados e eventos em colônias, mas ainda não foram amplamente testados em diferentes subespécies de *Apis mellifera* ou em circunstâncias climáticas variadas. Uma direção interessante a seguir é a combinação de diferentes atributos no processo de análise dos dados, e a aplicação de softwares de mineração de dados poderia resultar em conclusões inesperadas (Zacepins et al.; 2015).

2.2 Mineração de dados

A mineração de dados se refere ao processo não trivial de extração de informação implícita, previamente desconhecida e potencialmente útil de um conjunto de dados (Piateski and Frawley; 1991). Ela pode ser baseada em diversas abordagens. Há, por exemplo, a mineração baseada em generalização, a mineração baseada em reconhecimento de padrões, a mineração baseada em teorias estatísticas e matemáticas, e as abordagens integradas (Chen et al.; 1996).

O propósito da abordagem baseada em reconhecimento de padrões é realizar a classificação do padrão ou de maneira supervisionada ou de maneira não-supervisionada. Na classificação supervisionada o padrão de entrada é identificado como membro de uma classe predefinida, enquanto que na classificação não-supervisionada o padrão é atribuído a uma classe desconhecida (Jain et al.; 2000). Neste caso, como não há informação prévia sobre as classes às quais os dados pertencem, utiliza-se apenas as informações extraídas dos próprios dados para tentar agrupá-los por similaridade. Na classificação não-supervisionada, ou aprendizagem não-supervisionada, a clusterização é considerada a abordagem mais importante (Xu and Tian; 2015).

A clusterização tem como objetivo separar um conjunto finito de dados não rotulados em um conjunto finito e discreto de estruturas de dados “naturais” ocultas (Xu and Wunsch; 2005). Não há consenso sobre uma definição completa de clusterização (Xu and Tian; 2015), mas a dada por Jain and Dubes (1988) é considerada clássica e está descrita a seguir:

- instâncias (ou amostras) no mesmo *cluster* devem ser tão semelhantes quanto possível;
- instâncias em *clusters* diferentes devem ser tão diferentes quanto possível;
- a medida de similaridade e dissimilaridade deve ser clara e ter significado prático.

Entre os algoritmos de clusterização mais

tradicionais estão os algoritmos baseados em partições, que têm como ideia básica considerar o centro dos dados como o centro do *cluster* correspondente (Xu and Tian; 2015).

O algoritmo de clusterização baseado em partições mais utilizado na prática é o *K-means* (Wu et al.; 2008), cujo conceito é recalculando o centro do *cluster* por um processo iterativo que irá continuar até que um critério de convergência seja atendido (Xu and Tian; 2015). Este algoritmo é eficaz em conjuntos de dados de larga escala (Wu et al.; 2008), e sua complexidade aumenta de forma aproximadamente linear com o número de amostras do conjunto de dados (Xu and Wunsch; 2005).

As principais limitações do *K-means* são sua sensibilidade à inicialização, já que os centros iniciais são atribuídos aleatoriamente, e às anomalias do conjunto de dados, já que os centroides obtidos pelo algoritmo são gerados a partir da média das amostras do conjunto de dados, e esta não é uma medida estatística robusta (Wu et al.; 2008). Isso torna indispensável a adoção da estratégia de remoção de anomalias como uma etapa de pré-processamento dos dados. Além disso, o *K-means* não é eficaz ao manipular conjuntos de dados com muitos atributos (Xu and Wunsch; 2005).

A seguir detalha-se o funcionamento do algoritmo *K-means*.

2.2.1 Algoritmo *K-means*

O objetivo do *K-means* (MacQueen; 1967) é dividir o conjunto de N vetores de dados em K partições (ou *clusters*) não-superpostas ($K \ll N$), com o auxílio de K protótipos, também chamados de centroides, devidamente posicionados no espaço dos dados. Então cada vetor de dados é associado a um centroide por critério de similaridade, como o de menor distância, por exemplo.

O conjunto W de K centroides é representado por (1), em que p é a dimensão do vetor de dados.

$$W = \{w_i\}_{i=1}^K \mid w_i \in \mathbb{R}^p. \quad (1)$$

O *cluster* V associado a cada centroide w é definido por (2), em que x é um vetor de atributos, e $\|x - w\|$ indica a distância euclidiana.

$$V_i = \{x \in \mathbb{R}^p \mid \|x - w_i\| < \|x - w_j\|, \forall j \neq i\}. \quad (2)$$

De maneira geral, o *K-means* se comporta da seguinte forma: inicialmente são selecionados K vetores aleatórios do espaço de dados como centroides; o *cluster* de cada protótipo w_i é determinado por (2); então calcula-se a nova posição de cada centroide w_i como a média dos N_i objetos do *cluster* V_i (3).

$$w_i = \frac{1}{N_i} \sum_{x \in V_i} x. \quad (3)$$

O *cluster* V_i (2) e a nova posição do centroide w_i (3) são repetidamente recalculados até a convergência do algoritmo. Isso acontece quando a posição do centroide w_i não muda mais ou quando atinge-se um número máximo de iterações, por exemplo. Para avaliar quantitativamente o posicionamento

dos centroides, calcula-se a soma das distâncias quadráticas (*sum of squared distance* – SSD) de um vetor de dados ao centroide mais próximo (4). Este erro ajuda a avaliar a qualidade dos clusters gerados.

$$SSD = \sum_{\forall x \in V_i} \|x - w_i\|^2. \quad (4)$$

3 Trabalhos correlatos

A literatura apresenta diversos trabalhos referentes a sistemas de monitoramento de apiários e apicultura de precisão. Alguns tem focado no aspecto ferramental dos sistemas (Zacepins and Karasha; 2012; Zacepins et al.; 2016, 2017), enquanto outros também exploram a análise dos dados coletados (Kridi et al.; 2014, 2016; Kvisis and Zacepins; 2016), nos quais a temperatura é a métrica-base para reconhecimento de padrões das colmeias.

A massa das colmeias é outra boa métrica de monitoramento da produtividade de uma colmeia com perspectivas de correlação entre a produção de mel e diferentes parâmetros de condições meteorológicas. (Fitzgerald et al.; 2015; Ruan et al.; 2017). De forma mais abrangente, temperatura, umidade relativa e massa podem ser combinados para checagem da termorregulação e evolução da colônia na época de floração (Gil-Lebrero et al.; 2016)

As técnicas de análise de componentes principais (PCA – *Principal Component Analysis*) e de análise de função discriminante (DFA – *Discriminant Function Analysis*) tem sido aplicadas em dados de vibração das colmeias para identificação de abandono iminente (Bencsik et al.; 2011) e ciclo de produção das crias (Bencsik et al.; 2015).

As abelhas rainhas emitem um ruído agudo específico logo antes de abandonar a colmeia, enquanto o ruído produzido pelo restante da colônia se torna mais alto. Portanto, sensores de áudio ajudam a detectar o estado iminente de abandono de uma colmeia (Murphy et al.; 2015a,b).

Além de temperatura, umidade, massa, vibração e áudio, os gases CO₂ (dióxido de carbono) e O₂ (oxigênio), gases poluentes, dados climáticos (temperatura externa, índice pluviométrico e intensidade de luz solar) podem ser usados para classificar possíveis estados da colônia (Murphy et al.; 2016).

A imagem também pode ser uma métrica útil na tomada de decisão em um apiário. Por exemplo, um sistema de monitoramento com câmeras posicionadas na entrada das colmeias captura informações do fluxo de abelhas portanto do nível de atividade das abelhas ao redor da colmeia e pode ser um bom indicador da saúde da colônia (Tashakkori and Ghadiri; 2015). No tocante a soluções para combate a infestações, algoritmos de processamento de imagens podem ser aplicados em vídeos internos de uma colmeia, por exemplo, para detectar e mesmo eliminar com lasers o ácaro *Varroa destructor* (Chazette et al.; 2016).

Nota-se que a maioria dos trabalhos tem como atributo básico de monitoramento a temperatura do interior das colmeias, seja ela monitorada isoladamente ou em conjunto com outros fatores. Apesar da viabilidade, em alguns casos, de se coletar a umidade com o mesmo equipamento da temperatura,

esse atributo não é tão recorrente no âmbito da apicultura de precisão e precisa ser mais explorado.

A análise dos dados de massa e a vibração das colmeias podem apresentar perspectivas promissoras e também precisam ser mais exploradas. Já o monitoramento e a análise dos dados de atributos como áudio, vídeo e gases podem ser muito custosos, uma vez que os sinais de áudio e vídeo requerem equipamentos mais robustos para serem processados, e sensores de gás tem alto custo financeiro. Especificamente para a captação do áudio, é necessário o uso de um procedimento sistemático na localização dos microfones dentro da colmeia bem como na filtragem de outros ruídos que não sejam emitidos pelas abelhas.

Percebe-se ainda que muitos trabalhos focam apenas no desenvolvimento ferramental dos sistemas de monitoramento, ignorando a semântica intrínseca dos dados. Isso comprova que a etapa de análise dos dados é o principal obstáculo na abordagem da apicultura de precisão e precisa de maior atenção. A Tabela 1 sintetiza os principais aspectos de cada trabalho relacionado discutido.

4 Material e métodos

Esta seção descreve os aspectos metodológicos da pesquisa realizada em relação ao ferramental utilizado, obtenção e pré-processamento dos dados, estratégias de aprendizado, bem como a análise e detecção dos estados da colônia de abelhas sob estudo.

4.1 Conjuntos de dados

Os dados utilizados neste artigo foram obtidos no portal HiveTool.net, um projeto de código aberto que visa produzir ferramentas para monitoramento e gerenciamento de colmeias de abelhas. Os dados referem-se à massa, umidade e temperatura interna de colmeias de *Apis mellifera*, além dos dados de temperatura e umidade ambiente. Contudo, para a análise por agrupamento, foram considerados apenas os dados de massa, umidade e temperatura interna de colmeias.

Foram analisados três conjuntos de dados. Dois deles são referentes a colmeias do apiário Arnas (56° 53' 17.8656" N, 9° 50' 39.5340" L), localizado em Rebild, Dinamarca. O terceiro refere-se a uma colmeia do apiário Emil (58° 19' 9.1596" N, 8° 32' 9.8124" L), localizado em Grimstad, Noruega.

O primeiro conjunto de dados do apiário Arnas possui 94.552 amostras coletadas entre 01/12/2016 e 30/11/2017. O segundo conjunto de dados do apiário Arnas e o conjunto de dados do apiário Emil possuem, respectivamente, 92.749 e 88.181 amostras coletadas entre 01/03/2017 a 28/02/2018. O intervalo de medição entre as amostras foi de cinco minutos para todos os conjuntos de dados.

4.2 Pré-processamento dos dados

Em um problema de clusterização, as estratégias de pré e pós-processamento dos dados, como a remoção de anomalias, a normalização e a validação da clusterização, são tão importantes quanto a execução

Tabela 1: Visão geral dos trabalhos correlatos.

Referência	Métrica(s) monitorada(s)	Ferramental computacional para monitoramento	Técnica de análise
Zacepins and Karasha (2012)	Temperatura	PC	-
Zacepins et al. (2016)	Temperatura	Raspberry Pi	Algoritmo de suporte à decisão
Zacepins et al. (2017)	Temperatura	PIC	-
Kridi et al. (2014)	Temperatura	Arduino	Rec. de padrões (<i>K-means</i>)
Kridi et al. (2016)	Temperatura	Arduino	Rec. de padrões (<i>K-means</i>)
Kviesis and Zacepins (2016)	Temperatura	Raspberry Pi	Redes neurais
Fitzgerald et al. (2015)	Massa	ATmega 1281	-
Ruan et al. (2017)	Massa	Arduino	-
Gil-Lebrero et al. (2016)	Temperatura, umidade e massa	ATmega 1281	-
Bencsik et al. (2011)	Vibração	PC	PCA
Bencsik et al. (2015)	Vibração	-	DFA
Murphy et al. (2015a)	Áudio	ATmega 1281	Processamento de sinais
Murphy et al. (2015b)	Áudio, vídeo e vibrações	ATmega 1281 e Raspberry Pi	Processamento de sinais
Murphy et al. (2016)	Temperatura, umidade e gases	ATmega 1281	Árvores de decisão
Chazette et al. (2016)	Temp., massa, áudio e vídeo	Raspberry Pi	Processamento de imagens
Tashakkori and Ghadiri (2015)	Vídeo	-	Processamento de imagens

do algoritmo em si (Xu and Wunsch; 2005).

4.2.1 Remoção de anomalias

A remoção de anomalias foi realizada pelo método de Tukey (Tukey; 1977), no qual define-se um intervalo com limites inferior e superior dos atributos do conjunto de dados. As amostras que estiverem fora do intervalo são consideradas anomalias. O limite inferior é calculado por (5), e o superior por 6, em que Q_1 e Q_3 são respectivamente o primeiro e o terceiro quartis de um atributo dos dados.

$$Q_1 - 1.5 \times (Q_3 - Q_1). \quad (5)$$

$$Q_3 + 1.5 \times (Q_3 - Q_1). \quad (6)$$

4.2.2 Normalização dos dados

A etapa de normalização se faz necessária para que os dados estejam na mesma escala nas etapas posteriores da análise. Foi utilizado o método da normalização estatística. Neste método, cada atributo X do conjunto de dados teve cada valor x normalizado por (7), em que μ e σ são respectivamente a média e o desvio padrão de X .

$$Z = \frac{x - \mu}{\sigma}. \quad (7)$$

4.2.3 Validação da clusterização

O processo de clusterização particiona o conjunto de dados em um número ideal de K subconjuntos. Entretanto, em certas circunstâncias, o valor de K é desconhecido e precisa ser estimado a partir dos próprios dados (Xu and Wunsch; 2005). Diversos algoritmos de clusterização têm o valor de K como parâmetro de entrada, tornando, conseqüentemente,

a qualidade dos resultados fortemente dependente da estimativa de K . Um particionamento em muitos subconjuntos dificulta a análise e interpretação dos resultados, enquanto um particionamento em poucos subconjuntos causa perda de informação (Xu and Wunsch; 2005).

Uma maneira de encontrar o valor ideal de K é por meio dos índices de validação de clusterização. Normalmente esses índices avaliam dois aspectos da clusterização: (i) a coesão interna, baseada nas distâncias entre os vetores de dados de um mesmo *cluster* e que indica o quanto um agrupamento é compacto; (ii) e a separação externa, baseada nas distâncias entre os centroides dos *clusters* e que indica a distinção dos *clusters*.

Milligan and Cooper (1985) avaliaram e compararam 30 índices de acordo com seus desempenhos em relação a uma série de conjuntos de dados artificiais, e o índice Calinski-Harabasz (Calinski and Harabasz; 1974) obteve os melhores resultados. O cálculo deste índice é definido por (8), em que B_K e W_K são, respectivamente, as matrizes de dispersão entre grupos e intragrupo, das quais são calculados os traços (operador $tr()$).

$$CH(K) = \frac{tr(B_K)/(K-1)}{tr(W_K)/(N-K)}. \quad (8)$$

O índice CH deve ser calculado para vários valores de K , e como é desejado um valor alto de B_K (*clusters* bem distintos) e um valor baixo de W_K (*clusters* compactos), o maior valor de CH indica o K mais adequado para o conjunto de dados.

4.2.4 Avaliação experimental

Cada conjunto de dados descrito na Subseção 4.1 foi dividido em dois períodos de seis meses.

Para o conjunto de dados Arnas I, os períodos foram de (i) dezembro de 2016 a maio de 2017, e (ii) junho a novembro de 2017. O primeiro período corresponde às estações inverno e primavera, e o segundo corresponde às estações verão e outono no hemisfério norte.

Para os conjuntos de dados Arnas II e Emil, os períodos foram de (i) março a agosto de 2017, e (ii) de setembro de 2017 a fevereiro de 2018. O primeiro período corresponde às estações primavera e verão, e o segundo corresponde às estações outono e inverno no hemisfério norte.

É importante destacar que a diferença das estações dos períodos entre os conjuntos de dados Arnas I e os demais foi utilizada para se avaliar o comportamento do método em dois casos: (i) quando cada período contém uma estação fria e uma estação quente, e (ii) quando cada período contém somente estações frias ou somente estações quentes. O último caso refere-se aos dois períodos nos quais o ciclo anual de colônias de abelhas em clima temperado geralmente é dividido, conforme discutido na Seção 1.

O procedimento para os experimentos se deu pela execução das etapas de pré-processamento dos dados e pela aplicação do algoritmo *K-means* a cada período dos conjuntos de dados.

Após a limpeza e normalização dos dados, investigou-se qual seria a quantidade de partições mais adequada para cada período de cada conjunto de dados. Para isso, o índice Calinski-Harabasz foi calculado 100 vezes para valores de *K* entre 2 e 15. Tomou-se como valor ideal o *K* que apresentou o maior índice Calinski-Harabasz por mais vezes.

Em seguida, o algoritmo *K-means* foi aplicado no conjunto de dados referente a cada período, tendo como parâmetro o respectivo *K* calculado pelo índice Calinski-Harabasz. Como os centroides iniciais são escolhidos aleatoriamente, a posição final dos mesmos pode variar a cada execução do algoritmo. Para se definir os melhores centroides, o SSD foi utilizado como critério de decisão. Um menor valor de SSD significa menores distâncias entre os dados dos *clusters* e os seus respectivos centroides. Dessa forma, o algoritmo foi executado 20 vezes para a escolha dos centroides que produziram o menor SSD.

Após a execução do algoritmo *K-means* com os dados normalizados, cada amostra recebeu um rótulo referente ao *cluster* do qual ela faz parte. Esses rótulos foram então aplicados às amostras dos dados não-normalizados, de forma que a correspondência entre as posições das amostras nos dois conjuntos de dados fosse obedecida. Isso foi feito para se obter as estatísticas descritivas dos atributos isolados de cada *cluster* na escala dos dados originais.

Para a implementação da avaliação experimental, foram utilizados os algoritmos *K-means* e Calinski-Harabasz disponíveis na biblioteca de aprendizagem de máquina *scikit-learn* para a linguagem de programação *Python*. Já o método de Tukey foi implementado por código próprio. Além disso, foram utilizadas como suporte as bibliotecas *numpy*¹ e *pandas*², para manipulação e análise de estruturas de dados.

5 Resultados

Nesta seção, os resultados encontrados pela aplicação do método descrito na Seção 4 aos conjuntos de dados são apresentados e discutidos.

5.1 Arnas I

Tanto para o primeiro período, de dezembro de 2016 a maio de 2017, quanto para o segundo período de Arnas I, de junho a novembro de 2017, o método proposto retornou 2 *clusters* como resultado. Seus centroides são apresentados respectivamente na Tabela 2 e na Tabela 3.

Tabela 2: Centroides dos *clusters* obtidos para o 1º período de Arnas I.

	Temperatura interna (°C)	Umidade interna (%)	Massa da colmeia (kg)
Cluster 0	13.1	79.6	27.7
Cluster 1	29.8	55.5	24.0

Os *clusters* com o menor valor da temperatura podem ser entendidos como os estados da colônia durante as épocas frias de cada período. Portanto, o Cluster 0 do primeiro período e o Cluster 1 do segundo período podem ser interpretados, respectivamente, como o estado da colônia durante o inverno e o outono.

Tabela 3: Centroides dos *clusters* obtidos para o 2º período de Arnas I.

	Temperatura interna (°C)	Umidade interna (%)	Massa da colmeia (kg)
Cluster 0	30.3	55.5	15.8
Cluster 1	11.3	91.7	28.5

Complementarmente, os *clusters* com o maior valor da temperatura podem ser entendidos como os estados da colônia durante as épocas quentes de cada período. Logo, o Cluster 1 do primeiro período e o Cluster 0 do segundo período podem ser interpretados, respectivamente, como o estado da colônia durante a primavera e o verão.

Esses resultados podem ser comprovados pela Figura 2 e pela Figura 4, que mostram a proporção da quantidade de amostras de cada *cluster* durante os meses do primeiro e do segundo período, respectivamente. Já a Figura 3 e a Figura 5 mostram a distribuição das amostras dos *clusters* no espaço de dados do primeiro e do segundo período, respectivamente.

Nota-se que, no primeiro período, as amostras do Cluster 0 estão distribuídas entre dezembro de 2016 e fevereiro de 2017, intervalo que corresponde ao inverno. Em março há amostras de ambos os *clusters*, indicando a transição entre o inverno e a primavera, e o restante das amostras do Cluster 1 estão totalmente distribuídas em abril e maio, os outros meses da primavera.

Já no segundo período, as amostras do Cluster 0 estão distribuídas entre junho e agosto, meses

¹<http://www.numpy.org/>

²<https://pandas.pydata.org/>

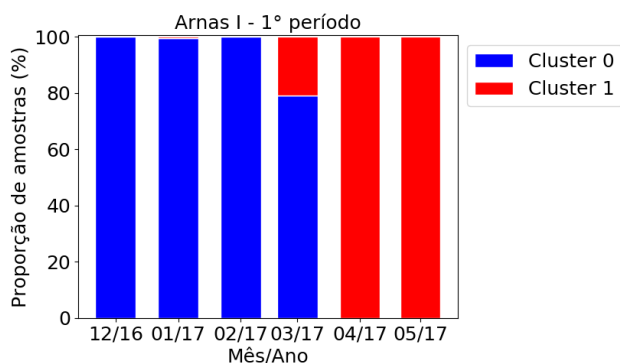


Figura 2: Proporção da quantidade de amostras de cada *cluster* para o 1º período de Arnas I.

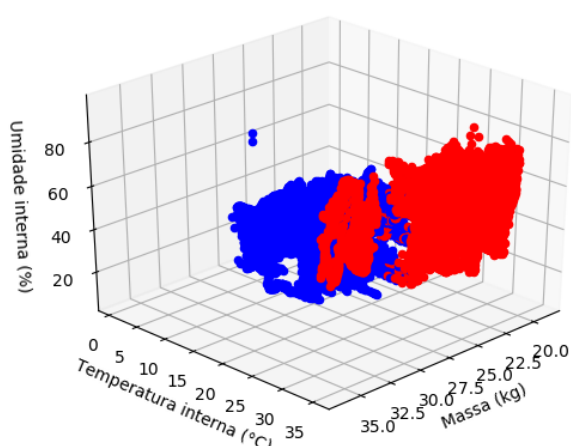


Figura 3: Distribuição das amostras dos *clusters* do 1º período de Arnas I no espaço de dados.

que correspondem ao verão, enquanto as amostras do *Cluster 1* estão distribuídas principalmente entre setembro e novembro, meses que correspondem ao outono.

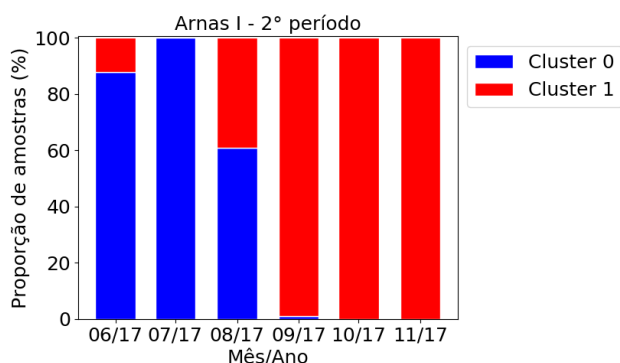


Figura 4: Proporção da quantidade de amostras de cada *cluster* para o 2º período de Arnas I.

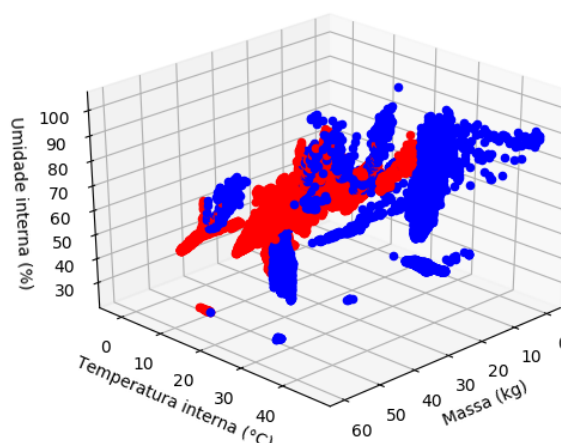


Figura 5: Distribuição das amostras dos *clusters* do 2º período de Arnas I no espaço de dados.

5.2 Arnas II

Para o primeiro período de Arnas II, de março a agosto de 2017, o método proposto retornou como resultado 5 *clusters*. Seus centroides são apresentados na Tabela 4.

Tabela 4: Centroides dos *clusters* obtidos para o 1º período de Arnas II.

	Temperatura interna (°C)	Umidade interna (%)	Massa da colmeia (kg)
<i>Cluster 0</i>	14.3	89.2	25.3
<i>Cluster 1</i>	29.9	53.3	23.3
<i>Cluster 2</i>	30.8	54.0	11.1
<i>Cluster 3</i>	33.6	58.4	37.3
<i>Cluster 4</i>	16.4	85.9	24.6

Dois *clusters* apresentam centroides semelhantes: o *Cluster 0* e o *Cluster 4*. Eles podem ser entendidos como o estado da colônia durante a transição entre as épocas fria e quente do ano (do inverno para a primavera e do verão para o outono). Chega-se a essa conclusão pelos valores da temperatura, os mais baixos, e da umidade, os mais altos. Isso se dá pela influência das épocas frias, que na Dinamarca também são as mais úmidas, sobre as colônias. Os perfis de temperatura e umidade ambientes podem ser observados na Figura 6.

O *Cluster 1* pode ser entendido como o estado da colônia em meados da primavera. Isso é evidenciado principalmente pela baixa umidade, já que nessa estação as precipitações, e consequentemente a umidade, apresentam médias baixas, como podemos observar na Figura 6. O *Cluster 2* pode ser interpretado como o estado da colônia no verão, não só pela baixa umidade (similar à observada no *Cluster 1*), mas também pela pouca massa. O baixo valor da massa pode indicar a coleta pelo apicultor do mel produzido pelas abelhas durante a primavera. O *Cluster 3* pode ser entendido como o estado da colônia no final da primavera. Nesse momento ocorre o ápice da produção de mel, que se reflete no alto valor da massa do centroide.

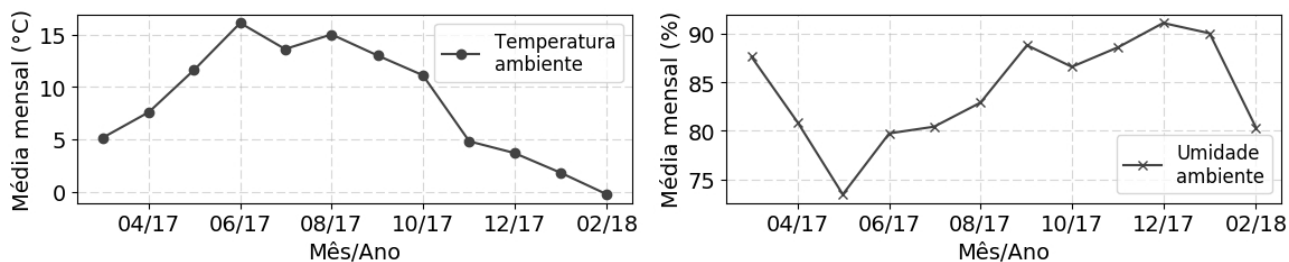


Figura 6: Médias mensais da temperatura e umidade ambientes - Arnas II.

Esses resultados podem ser comprovados pela Figura 7, que mostra a proporção da quantidade de amostras de cada cluster durante os meses do primeiro período. Nota-se que as amostras do Cluster 0 estão distribuídas principalmente no mês de março, e as amostras do Cluster 4 no mês de agosto. Esses meses correspondem, respectivamente, ao início da primavera (após o inverno) e ao final do verão (antes do outono). Já as amostras do Cluster 1 estão distribuídas principalmente entre os meses de março, abril e maio, que correspondem à primavera. As amostras do Cluster 2 estão distribuídas entre os meses de junho, julho e agosto, que correspondem ao verão. Por fim, as amostras do Cluster 3 estão distribuídas principalmente no mês de maio, o último mês da primavera no hemisfério norte. A Figura 8 mostra a distribuição das amostras dos clusters do primeiro período de Arnas II no espaço de dados.

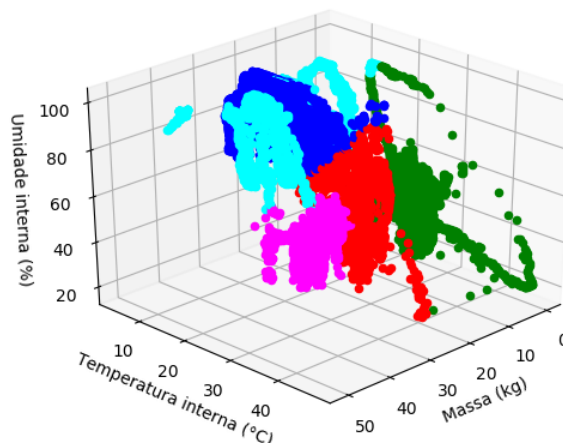


Figura 8: Distribuição das amostras dos clusters do 1º período de Arnas II no espaço de dados.

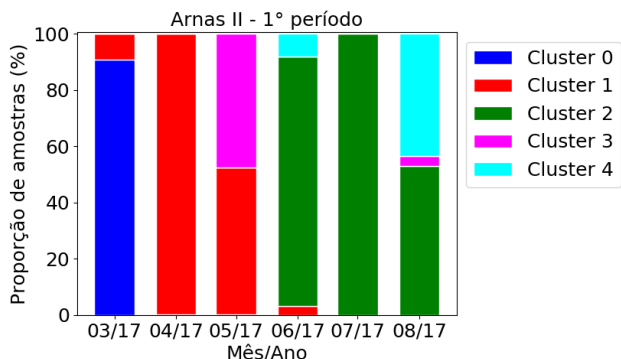


Figura 7: Proporção da quantidade de amostras de cada cluster para o 1º período de Arnas II.

Para o segundo período de Arnas II, de setembro de 2017 a fevereiro de 2018, o método proposto retornou como resultado 3 clusters, cujos centroides são apresentados na Tabela 5.

Tabela 5: Centroides dos clusters obtidos para o 2º período de Arnas II.

	Temperatura interna (°C)	Umidade interna (%)	Massa da colmeia (kg)
Cluster 0	10.3	78.1	31.0
Cluster 1	13.4	93.3	31.2
Cluster 2	7.7	96.0	29.5

Por terem as temperaturas mais altas, o Cluster 0

e o Cluster 1 podem ser entendidos como o estado da colônia durante a transição entre as épocas quente e fria do ano (do verão para o outono e do inverno para a primavera). Como o Cluster 1 possui a umidade mais alta entre os dois, é possível presumir que ele refere-se à transição do verão para o outono, já que a média da umidade é maior no outono, como podemos observar na Figura 6.

De forma complementar, o Cluster 2 pode ser interpretado como o estado da colônia em meados do segundo período, quando se inicia o inverno. Nessa época ocorrem as menores médias de temperatura, vide Figura 6.

Esses resultados podem ser comprovados pela Figura 9. Percebe-se que as amostras do Cluster 0 estão distribuídas principalmente nos últimos meses do segundo período, que correspondem ao final do inverno (antes da primavera). Já as amostras do Cluster 1 estão distribuídas principalmente nos meses de setembro e outubro, que correspondem ao início do outono. Por fim, as amostras do Cluster 2 estão distribuídas principalmente entre os meses de outubro e novembro, quando se encerra o outono e se inicia o inverno. A Figura 10 mostra a distribuição das amostras dos clusters do segundo período de Arnas II no espaço de dados.

É importante destacar que, se a temperatura da colônia seguisse a tendência da temperatura ambiente observada na Figura 6, o valor da temperatura do Cluster 0 deveria ser menor. No entanto, ele é maior do que o valor da temperatura do Cluster 2. Isso indica a capacidade de termorregulação da colmeia,

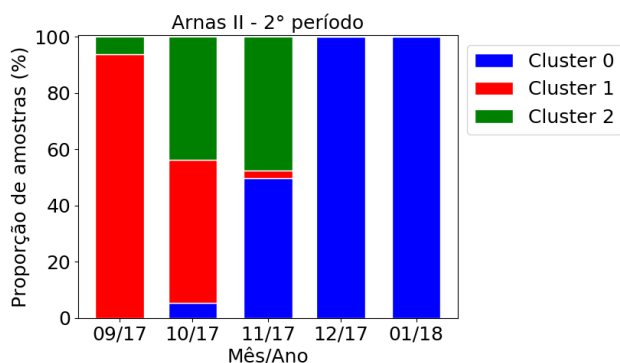


Figura 9: Proporção da quantidade de amostras de cada *cluster* para o 2º período de Arnas II.

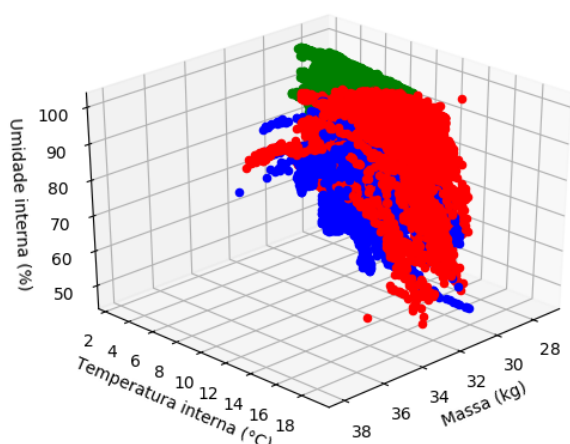


Figura 10: Distribuição das amostras dos *clusters* do 2º período de Arnas II no espaço de dados.

evidenciada também pelo valor alto e regular da massa, já que o controle térmico depende da presença de um maior número de abelhas operárias (Cook and Breed; 2013).

5.3 Emil

Para o primeiro período de Emil, de março a agosto de 2017, o método retornou como resultado 3 *clusters*. Seus centroides são apresentados na Tabela 6.

Tabela 6: Centroides dos *clusters* obtidos para o 1º período de Emil.

	Temperatura interna (°C)	Umidade interna (%)	Massa da colmeia (kg)
Cluster 0	26.5	55.0	12.4
Cluster 1	29.4	69.1	14.3
Cluster 2	10.9	75.8	14.0

Similar à avaliação feita para a colmeia de Arnas, pode-se interpretar o *Cluster 2* como o estado da colônia durante a transição entre as épocas fria e quente do ano devido ao baixo valor da temperatura. Ainda levando em consideração o valor da temperatura, pode-se interpretar o *Cluster 0* como

o estado da colônia na primavera, e o *Cluster 1* no verão, já que este possui o valor mais alto.

Tais resultados podem ser comprovados pela Figura 11. Percebe-se que as amostras do *Cluster 2* estão distribuídas principalmente no mês de março, o primeiro mês após o inverno no hemisfério norte. Já as amostras do *Cluster 0* estão distribuídas principalmente nos meses de abril e maio, que correspondem ao final da primavera. Por fim, as amostras do *Cluster 1* estão distribuídas principalmente nos meses de junho, julho e agosto, que correspondem ao verão. A Figura 12 mostra a distribuição das amostras dos *clusters* do primeiro período de Emil no espaço de dados.

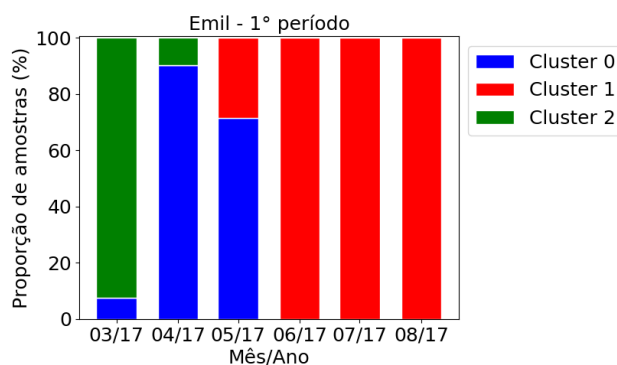


Figura 11: Proporção da quantidade de amostras de cada *cluster* para o 1º período de Emil.

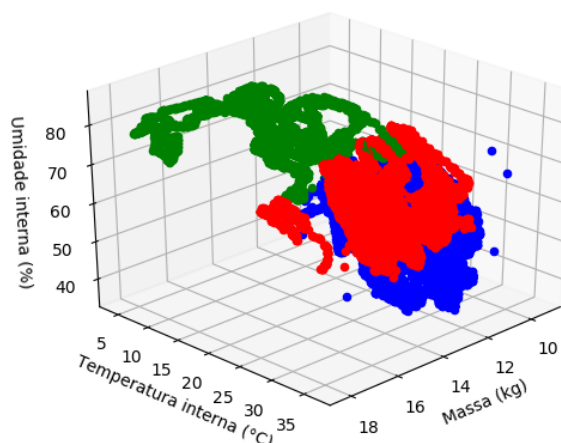


Figura 12: Distribuição das amostras dos *clusters* do 1º período de Emil no espaço de dados.

Para o segundo período, de setembro de 2017 a fevereiro de 2018, o método apresentado retornou como resultado 6 *clusters*, cujos centroides são apresentados na Tabela 7.

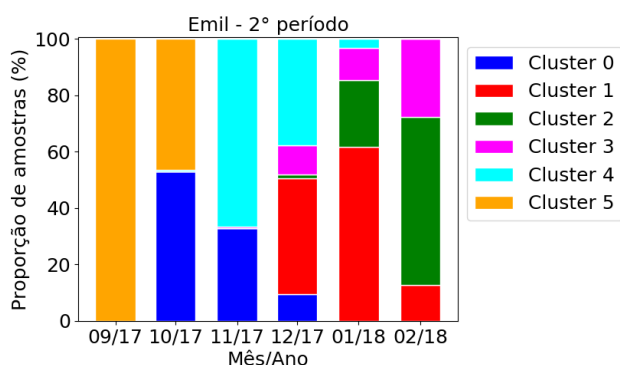
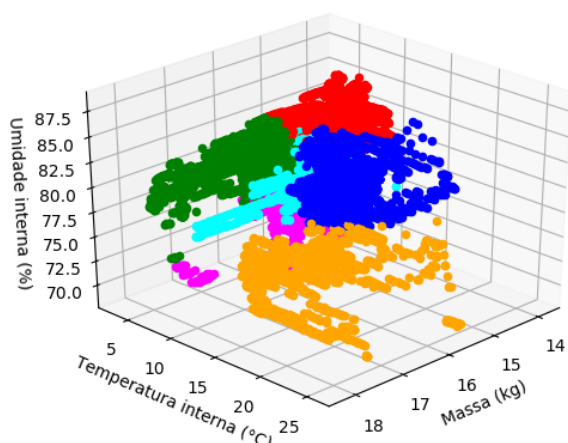
É possível notar que o valor da massa dos centroides é mais baixo se comparado aos de Arnas II em ambos os períodos, com exceção de um *cluster*. Pode-se, então, presumir que o controle térmico da colmeia do apiário Emil está comprometido. Isso fica visível principalmente no segundo período, que

Tabela 7: Centroides dos clusters obtidos para o 2º período de Emil.

	Temperatura interna (°C)	Umidade interna (%)	Massa da colmeia (kg)
Cluster 0	13.5	79.9	15.6
Cluster 1	7.2	82.1	14.5
Cluster 2	7.0	80.6	16.5
Cluster 3	5.9	74.4	14.8
Cluster 4	7.9	78.4	15.2
Cluster 5	16.3	74.2	16.9

corresponde ao outono e inverno.

Percebe-se pela Figura 13 que a proporção dos clusters distribuídos pelos meses do segundo período segue a ordem decrescente da temperatura, o que corresponde à tendência da temperatura ambiente, mostrada na Figura 15. Em setembro de 2017, primeiro mês após o verão, há amostras somente do Cluster 5, que possui o maior valor de temperatura. Já em fevereiro de 2018 as amostras dos Clusters 2 e 3, que possuem os menores valores de temperatura, são predominantes. A Figura 14 mostra a distribuição das amostras dos clusters do segundo período de Emil no espaço de dados.

**Figura 13:** Proporção da quantidade de amostras de cada cluster para o 2º período de Emil.**Figura 14:** Distribuição das amostras dos clusters do 2º período de Emil no espaço de dados.

Nota-se que a aplicação do nosso método sobre os conjuntos de dados Arnas II e Emil resultou em identificação mais detalhada dos padrões comportamentais das colônias relacionados às estações dos períodos frios e quentes. Já em relação ao conjunto de dados Arnas I, quando as estações dos períodos foram mescladas, resultou na identificação de padrões que refletiam apenas as condições externas à colmeia.

6 Conclusão

Este artigo propõe um método para identificar e caracterizar padrões sazonais de colônias de abelhas *Apis mellifera* através da aplicação de técnicas de mineração de dados. A literatura de apicultura de precisão tem normalmente considerado apenas uma variável física, geralmente a temperatura, para o monitoramento e análise; em alguns casos são consideradas outras variáveis (e.g. umidade), mas a análise dos dados monitorados tem sido normalmente desconsiderada. O método aqui proposto leva em conta três variáveis (temperatura interna, umidade interna e massa da colmeia) bem como sua respectiva análise.

Para o caso em que os datasets foram divididos de acordo com o ciclo anual referido na literatura, os padrões foram reconhecidos como coerentes, já que correspondem ao que se observa em campo. Podemos concluir também que a colmeia aparentemente mais forte consegue desempenhar com mais eficiência o controle térmico do interior da colmeia durante período frio (subseção 5.3). Para o caso em que os datasets foram divididos mesclando-se estações frias e quentes, a aplicação da método proposto resultou na identificação de padrões que refletiam apenas as condições externas às colmeias (subseção 5.1). Assim, podemos concluir que a divisão do ciclo anual das colônias de abelhas referida na literatura está adequada.

Na qualidade de estudos futuros a curto e médio prazos, pretendemos aplicar o método aqui proposto em dados de colônias de *Apis mellifera* africanizada, um polihíbrido criado no Brasil. Para isso, estamos desenvolvendo um sistema de monitoramento dentro do projeto Universal Sm@rtBee³.

Outra perspectiva de trabalho futuro é a aplicação de técnicas de evolução de clusters Spiliopoulou et al. (2006); da Silva et al. (2014), no sentido de se prever com acurácia um estado indesejado iminente de uma determinada colônia bem como um eventual reconhecimento de novos padrões no comportamento das abelhas.

Agradecimentos

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001. Breno M. Freitas e Danielo G. Gomes agradecem ao CNPq (processos 302934/2010-3, 311878/2016-4 e 432585/2016-8). Os autores agradecem Rhaniel Xavier e Alisson Silva pelo apoio técnico.

³<http://smartbee.great.ufc.br/>

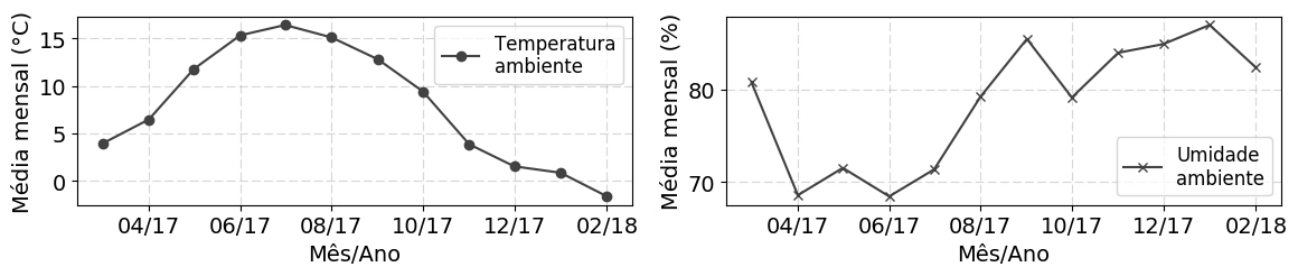


Figura 15: Médias mensais da temperatura e umidade ambientes - Emil.

Referências

- Abou-Shaara, H. F., Al-Ghamdi, A. A. and Mohamed, A. A. (2012). Tolerance of two honey bee races to various temperature and relative humidity gradients, *Environmental and experimental Biology* 10(4): 133-138.
- Abou-Shaara, H., Owayss, A., Ibrahim, Y. and Basuny, N. (2017). A review of impacts of temperature and relative humidity on various activities of honey bees, *Insectes Sociaux* 64(4): 455-463.
- Almeida, G. F., Grassi, M. L., Gramacho, K. P. and Gonçalves, L. S. (2006). Comportamento enxameatório em abelhas africanizadas induzido por temperatura com auxílio de câmara de controle de temperatura, sensores e apidômetros, XVI Congresso Brasileiro de Apicultura e II Congresso Brasileiro de Meliponicultura, Aracaju.
- Barron, A. B. (2015). Death of the bee hive: understanding the failure of an insect society, *Current Opinion in Insect Science* 10: 45-50.
- Bencsik, M., Bencsik, J., Baxter, M., Lucian, A., Romieu, J. and Millet, M. (2011). Identification of the honey bee swarming process by analysing the time course of hive vibrations, *Computers and Electronics in Agriculture* 76(1): 44-50.
- Bencsik, M., Le Conte, Y., Reyes, M., Pioz, M., Whittaker, D., Crauser, D., Delso, N. S. and Newton, M. I. (2015). Honeybee colony vibrational measurements to highlight the brood cycle, *PloS One* 10(11): e0141926.
- Biesmeijer, J. C., Roberts, S. P., Reemer, M., Ohlemüller, R., Edwards, M., Peeters, T., Schaffers, A., Potts, S. G., Kleukers, R., Thomas, C. et al. (2006). Parallel declines in pollinators and insect-pollinated plants in Britain and the Netherlands, *Science* 313(5785): 351-354.
- Bommarco, R., Lundin, O., Smith, H. G. and Rundlöf, M. (2011). Drastic historic shifts in bumble-bee community composition in Sweden, *Proceedings of the Royal Society of London B: Biological Sciences*.
- Calinski, T. and Harabasz, J. (1974). A dendrite method for cluster analysis, *Communications in Statistics* 3(1): 1-27.
- Chauzat, P., Laurent, M., Rivière, M.-P., Saugeon, C., Hendrikx, P., Ribière-Chabert, M. et al. (2014). A pan-European epidemiological study on honey bee colony losses 2012-2013, *European Union Reference Laboratory for Honeybee Health, Brussels, Rapport technique*.
- Chazette, L., Becker, M. and Szczerbicka, H. (2016). Basic algorithms for bee hive monitoring and laser-based mite control, *2016 IEEE Symposium Series on Computational Intelligence (SSCI)*, IEEE, pp. 1-8.
- Chen, M.-S., Han, J. and Yu, P. S. (1996). Data mining: an overview from a database perspective, *IEEE Transactions on Knowledge and Data Engineering* 8(6): 866-883.
- Cook, C. N. and Breed, M. D. (2013). Social context influences the initiation and threshold of thermoregulatory behaviour in honeybees, *Animal Behaviour* 86(2): 323-329.
- da Silva, T. L. C., de Macêdo, J. A. F. and Casanova, M. A. (2014). Discovering frequent mobility patterns on moving object data, *Proceedings of the Third ACM SIGSPATIAL International Workshop on Mobile Geographic Information Systems, MobiGIS '14*, ACM, New York, NY, USA, pp. 60-67.
URL: <http://doi.acm.org/10.1145/2675316.2675325>
- Deguines, N., Jono, C., Baude, M., Henry, M., Julliard, R. and Fontaine, C. (2014). Large-scale trade-off between agricultural intensification and crop pollination services, *Frontiers in Ecology and the Environment* 12(4): 212-217.
- Fitzgerald, D. W., Murphy, F. E., Wright, W. M. D., Whelan, P. M. and Popovici, E. M. (2015). Design and development of a smart weighing scale for beehive monitoring, *Signals and Systems Conference (ISSC), 2015 26th Irish*, IEEE, pp. 1-6.
- Giannini, T. C., Cordeiro, G. D., Freitas, B. M., Saraiva, A. M. and Imperatriz-Fonseca, V. L. (2015). The dependence of crops for pollinators and the economic value of pollination in Brazil, *Journal of Economic Entomology* 108(3): 849-857.
- Gil-Lebrero, S., Quiles-Latorre, F. J., Ortiz-López, M., Sánchez-Ruiz, V., Gámiz-López, V. and Luna-Rodríguez, J. J. (2016). Honey bee colonies remote monitoring system, *Sensors* 17(1): 55.
- Goulson, D., Nicholls, E., Botías, C. and Rotheray, E. L. (2015). Bee declines driven by combined stress from parasites, pesticides, and lack of flowers, *Science* 347(6229): 1255957.
- Jain, A. K. and Dubes, R. C. (1988). *Algorithms for clustering data*, Prentice-Hall, Inc.

- Jain, A. K., Duin, R. P. W. and Mao, J. (2000). Statistical pattern recognition: A review, *IEEE Transactions on pattern analysis and machine intelligence* **22**(1): 4–37.
- Kevan, P. and Phillips, T. (2001). The economic impacts of pollinator declines: an approach to assessing the consequences, *Conservation Ecology* **5**(1).
- Klein, A.-M., Vaissiere, B. E., Cane, J. H., Steffan-Dewenter, I., Cunningham, S. A., Kremen, C. and Tschamntke, T. (2007). Importance of pollinators in changing landscapes for world crops, *Proceedings of the Royal Society of London B: Biological Sciences* **274**(1608): 303–313.
- Kleinhenz, M., Bujok, B., Fuchs, S. and Tautz, J. (2003). Hot bees in empty broodnest cells: heating from within, *Journal of Experimental Biology* **206**(23): 4217–4231.
- Kremen, C., Williams, N. M. and Thorp, R. W. (2002). Crop pollination from native bees at risk from agricultural intensification, *Proceedings of the National Academy of Sciences* **99**(26): 16812–16816.
- Kridi, D. S., Carvalho, C. G. N. d. and Gomes, D. G. (2014). A predictive algorithm for mitigate swarming bees through proactive monitoring via wireless sensor networks, *Proceedings of the 11th ACM symposium on Performance evaluation of wireless ad hoc, sensor, & ubiquitous networks*, ACM, pp. 41–47.
- Kridi, D. S., de Carvalho, C. G. N. and Gomes, D. G. (2016). Application of wireless sensor networks for beehive monitoring and in-hive thermal patterns detection, *Computers and Electronics in Agriculture* **127**: 221–235.
- Kviesis, A. and Zacepins, A. (2016). Application of neural networks for honey bee colony state identification, *Carpathian Control Conference (ICCC), 2016 17th International*, IEEE, pp. 413–417.
- Li, Z., Huang, Z. Y., Sharma, D. B., Xue, Y., Wang, Z. and Ren, B. (2016). Drone and worker brood microclimates are regulated differentially in honey bees, *Apis mellifera*, *PloS One* **11**(2): e0148740.
- Maciel, F. A. O., Braga, A. R., Silva, A. L., da Silva, T. L. C., Freitas, B. M. and Gomes, D. G. (2018). Reconhecimento de padrões de colônias de abelhas *Apis mellifera* segundo mudanças das estações do ano, *Anais do IX Workshop de Computação Aplicada à Gestão do Meio Ambiente e Recursos Naturais / XXXVIII Congresso da Sociedade Brasileira de Computação*, SBC, UERN, Natal/RN, Brasil.
- Maciel, F. A. O., Braga, A. R., Xavier, R., da Silva, T. L. C., Freitas, B. M. and Gomes, D. G. (2018). Minerando dados para caracterizar padrões sazonais de colônias de abelhas *Apis mellifera*, *SBSI'18: XIV Simpósio Brasileiro de Sistemas de Informação*, SBC, UCS, Caxias do Sul/RS, Brasil.
- MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations, *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, Vol. 1, Oakland, CA, USA., pp. 281–297.
- McBratney, A., Whelan, B., Ancev, T. and Bouma, J. (2005). Future directions of precision agriculture, *Precision agriculture* **6**(1): 7–23.
- Meikle, W. G. and Holst, N. (2015). Application of continuous monitoring of honeybee colonies, *Apidologie* **46**(1): 10–22.
- Milligan, G. W. and Cooper, M. C. (1985). An examination of procedures for determining the number of clusters in a data set, *Psychometrika* **50**(2): 159–179.
- Mitchell, T. M. (1999). Machine learning and data mining, *Communications of the ACM* **42**(11): 30–36.
- Murphy, F. E., Magno, M., O’Leary, L., Troy, K., Whelan, P. and Popovici, E. M. (2015b). Big brother for bees (3b)—energy neutral platform for remote monitoring of beehive imagery and sound, *Advances in Sensors and Interfaces (IWASI), 2015 6th IEEE International Workshop on*, IEEE, pp. 106–111.
- Murphy, F. E., Magno, M., Whelan, P. M., O’Halloran, J. and Popovici, E. M. (2016). b+ wsn: Smart beehive with preliminary decision tree analysis for agriculture and honey bee health monitoring, *Computers and Electronics in Agriculture* **124**: 211–219.
- Murphy, F. E., Srbinovski, B., Magno, M., Popovici, E. M. and Whelan, P. M. (2015a). An automatic, wireless audio recording node for analysis of beehives, *Signals and Systems Conference (ISSC), 2015 26th Irish*, IEEE, pp. 1–6.
- Nicolson, S. W. (2009). Water homeostasis in bees, with the emphasis on sociality, *Journal of Experimental Biology* **212**(3): 429–434.
- Oldroyd, B. P. (2007). What’s killing american honey bees?, *PLoS Biol* **5**(6): e168.
- Ollerton, J., Winfree, R. and Tarrant, S. (2011). How many flowering plants are pollinated by animals?, *Oikos* **120**(3): 321–326.
- Ostwald, M. M., Smith, M. L. and Seeley, T. D. (2016). The behavioral regulation of thirst, water collection and water storage in honey bee colonies, *Journal of Experimental Biology* **219**(14): 2156–2165.
- Petz, M., Stabentheiner, A. and Crailsheim, K. (2004). Respiration of individual honeybee larvae in relation to age and ambient temperature, *Journal of comparative Physiology B* **174**(7): 511–518.
- Piateski, G. and Frawley, W. (1991). *Knowledge Discovery in Databases*, MIT Press, Cambridge, MA, USA.
- Pires, C. S. S., de Mello Pereira, F., do Rêgo Lopes, M. T., Nocelli, R. C. F., Malaspina, O., Pettis, J. S. and Teixeira, E. W. (2016). Enfraquecimento e perda de colônias de abelhas no Brasil: há casos de CCD?, *Pesquisa Agropecuária Brasileira* **51**(5): 422–442.
- Potts, S. G., Biesmeijer, J. C., Kremen, C., Neumann, P., Schweiger, O. and Kunin, W. E. (2010). Global pollinator declines: trends, impacts and drivers, *Trends in Ecology & Evolution* **25**(6): 345–353.

- Potts, S. G., Imperatriz-Fonseca, V. L., Ngo, H. T., Biesmeijer, J. C., Breeze, T. D., Dicks, L. V., Garibaldi, L. A., Hill, R., Settele, J. and Vanbergen, A. J. (2016). Summary for policymakers of the assessment report of the intergovernmental science-policy platform on biodiversity and ecosystem services on pollinators, pollination and food production, *Report*, Bonn, Germany.
- Ruan, Z.-Y., Wang, C.-H., Lin, H.-J., Huang, C.-P., Chen, Y.-H., Yang, E.-C., Tseng, C.-L. and Jiang, J.-A. (2017). An internet of things-based weight monitoring system for honey, *World Academy of Science, Engineering and Technology, International Journal of Biological, Biomolecular, Agricultural, Food and Biotechnological Engineering* **11**(6): 478–482.
- Sanchez, V., Gil, S., Flores, J. M., Quiles, F. J., Ortiz, M. A. and Luna, J. J. (2015). Implementation of an electronic system to monitor the thermoregulatory capacity of honeybee colonies in hives with open-screened bottom boards, *Computers and Electronics in Agriculture* **119**: 209–216.
- Southwick, E. E. and Moritz, R. F. (1987). Social control of air ventilation in colonies of honey bees, *Apis mellifera*, *Journal of insect physiology* **33**(9): 623–626.
- Spiliopoulou, M., Ntoutsi, I., Theodoridis, Y. and Schult, R. (2006). Monic: Modeling and monitoring cluster transitions, *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '06*, ACM, New York, NY, USA, pp. 706–711.
URL: <http://doi.acm.org/10.1145/1150402.1150491>
- Stabentheiner, A., Kovac, H. and Brodschneider, R. (2010). Honeybee colony thermoregulation-regulatory mechanisms and contribution of individuals in dependence on age, location and thermal stress, *PLoS One* **5**(1): e8967.
- Tashakkori, R. and Ghadiri, A. (2015). Image processing for honey bee hive health monitoring, *SoutheastCon 2015*, IEEE, pp. 1–7.
- Tscharntke, T., Tylianakis, J. M., Rand, T. A., Didham, R. K., Fahrig, L., Batary, P., Bengtsson, J., Clough, Y., Crist, T. O., Dormann, C. F. et al. (2012). Landscape moderation of biodiversity patterns and processes—eight hypotheses, *Biological Reviews* **87**(3): 661–685.
- Tukey, J. W. (1977). *Exploratory Data Analysis*, Addison-Wesley Publishing Company.
- Wu, X., Kumar, V., Quinlan, J. R., Ghosh, J., Yang, Q., Motoda, H., McLachlan, G. J., Ng, A., Liu, B., Philip, S. Y. et al. (2008). Top 10 algorithms in data mining, *Knowledge and information systems* **14**(1): 1–37.
- Xu, D. and Tian, Y. (2015). A comprehensive survey of clustering algorithms, *Annals of Data Science* **2**(2): 165–193.
URL: <https://doi.org/10.1007/s40745-015-0040-1>
- Xu, R. and Wunsch, D. (2005). Survey of clustering algorithms, *IEEE Transactions on Neural Networks* **16**(3): 645–678.
- Zacepins, A., Brusbardis, V., Meitalovs, J. and Stalidzans, E. (2015). Challenges in the development of precision beekeeping, *Biosystems Engineering* **130**: 60–71.
- Zacepins, A. and Karasha, T. (2012). Web based system for the bee colony remote monitoring, *Application of Information and Communication Technologies (AICT), 2012 6th International Conference on*, IEEE, pp. 1–4.
- Zacepins, A., Kviesis, A., Pecka, A. and Osadcuks, V. (2017). Development of internet of things concept for precision beekeeping, *Carpathian Control Conference (ICCC), 2017 18th International*, IEEE, pp. 23–27.
- Zacepins, A., Kviesis, A., Stalidzans, E., Liepniece, M. and Meitalovs, J. (2016). Remote detection of the swarming of honey bee colonies by single-point temperature monitoring, *Biosystems Engineering* **148**: 76–80.
- Zacepins, A., Stalidzans, E. and Meitalovs, J. (2012). Application of information technologies in precision apiculture, *Proceedings of the 13th International Conference on Precision Agriculture (ICPA 2012)*.
- Zogovic, N., Mladenovic, M. and Rasic, S. (2017). From primitive to cyber-physical beekeeping, *Zdravkovic, M., Konjovic, Z., Trajanovic, M. (Eds.) ICIST 2017 Proceedings Vol.1*, pp. 38–43.