



UNIVERSIDADE FEDERAL DO CEARÁ
CENTRO DE TECNOLOGIA
DEPARTAMENTO DE ENGENHARIA DE TELEINFORMÁTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE TELEINFORMÁTICA
MESTRADO ACADÊMICO EM ENGENHARIA DE TELEINFORMÁTICA

FELIPE ANDERSON OLIVEIRA MACIEL

RECONHECIMENTO DE PADRÕES SAZONAIS EM
COLÔNIAS DE ABELHAS APIS MELLIFERA

FORTALEZA

2018

FELIPE ANDERSON OLIVEIRA MACIEL

RECONHECIMENTO DE PADRÕES SAZONAIS EM
COLÔNIAS DE ABELHAS APIS MELLIFERA

Dissertação apresentada ao Curso de Mestrado Acadêmico em Engenharia de Teleinformática do Programa de Pós-Graduação em Engenharia de Teleinformática do Centro de Tecnologia da Universidade Federal do Ceará, como requisito parcial à obtenção do título de mestre em Engenharia de Teleinformática. Área de Concentração: Sinais e Sistemas

Orientador: Prof. Dr. Danielo Gonçalves Gomes

FORTALEZA

2018

Dados Internacionais de Catalogação na Publicação
Universidade Federal do Ceará
Biblioteca Universitária
Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

M138r Maciel, Felipe Anderson Oliveira.
Reconhecimento de Padrões Sazonais em Colônias de Abelhas *Apis mellifera* / Felipe Anderson Oliveira Maciel. – 2018.
50 f. : il. color.

Dissertação (mestrado) – Universidade Federal do Ceará, Centro de Tecnologia, Programa de Pós-Graduação em Engenharia de Teleinformática, Fortaleza, 2018.
Orientação: Prof. Dr. Danielo Gonçalves Gomes.

1. Mineração de dados. 2. Reconhecimento de padrões. 3. Clusterização. 4. Apicultura de precisão. 5. Abelhas. I. Título.

CDD 621.38

FELIPE ANDERSON OLIVEIRA MACIEL

RECONHECIMENTO DE PADRÕES SAZONAIS EM
COLÔNIAS DE ABELHAS APIS MELLIFERA

Dissertação apresentada ao Curso de Mestrado Acadêmico em Engenharia de Teleinformática do Programa de Pós-Graduação em Engenharia de Teleinformática do Centro de Tecnologia da Universidade Federal do Ceará, como requisito parcial à obtenção do título de mestre em Engenharia de Teleinformática. Área de Concentração: Sinais e Sistemas

Aprovada em: 30 de julho de 2018

BANCA EXAMINADORA

Prof. Dr. Danielo Gonçalves Gomes (Orientador)
Universidade Federal do Ceará (UFC)

Prof. Dr. Breno Magalhães Freitas
Universidade Federal do Ceará (UFC)

Prof.^a Dr.^a Ticiano Linhares Coelho da Silva
Universidade Federal do Ceará (UFC)

Prof. Dr. Gustavo Pessin
Universidade do Vale do Rio dos Sinos (UNISINOS)

À minha família, Márcia, Antienes, Andréa e
Pedro. Amo vocês.

AGRADECIMENTOS

À Deus pela força, pelo ânimo, disposição e capacidade para seguir e concluir esta empreitada.

Aos meus pais, Márcia e Antienes, irmã Andréa, e namorada, Larissa, pelo apoio, incentivo e paciência nos momentos difíceis.

Ao Prof. Dr. Danielo Gonçalves Gomes por me orientar em minha dissertação de mestrado.

Ao doutorando em Engenharia de Teleinformática Rafael Braga pela orientação e amizade, e a todos os integrantes do projeto *Sm@rtBee* pela assistência, em especial: Alisson Lima, Sara Lonngren, Lucas Esteves, Rhaniel Xavier e Gustavo Cignachi.

A todos os professores pelos ensinamentos e lições.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) e ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq)/Universal #432585/2016-8 pelo apoio financeiro.

Ao doutorando em Engenharia Elétrica Ednardo Moreira Rodrigues e seu assistente Alan Batista de Oliveira, aluno de graduação em Engenharia Elétrica, pela adequação do *template* utilizado neste trabalho para que o mesmo ficasse de acordo com as normas da biblioteca da Universidade Federal do Ceará (UFC).

“A noite é mais escura logo antes do amanhecer.”

(Harvey Dent)

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, esta dissertação visa identificar padrões de colônias de *Apis mellifera* a fim de auxiliar o apicultor no manejo e na manutenção de suas colmeias. Nossa metodologia consistiu na aplicação de uma técnica de clusterização em dois *datasets* reais de colmeias com dados de temperatura, umidade e massa. A partir da aplicação do índice Calinski-Harabasz e do algoritmo K-means, foram identificados padrões coerentes e associados às transições entre as estações do ano. Além disso, foi possível concluir que a colônia mais forte é mais eficiente ao tentar manter o microclima da colmeia durante o inverno.

Palavras-chave: Mineração de dados, Reconhecimento de padrões, Clusterização, Abelhas, *Apis mellifera*, Apicultura de precisão.

ABSTRACT

As the main pollinating agent, bees are essential to the production of food for mankind and to the maintenance of the ecosystem. Among the crops used for human consumption, 75% rely on pollination. Aligning to a current concern with bees survival, this dissertation aims to find out patterns of *Apis mellifera* colonies in order to assist the beekeeper in the management and maintenance of his hives. Our methodology consisted in the application of a clustering technique in two real datasets of hives with data of temperature, humidity and weight. From the application of the Calinski-Harabasz index and the K-means algorithm, we have identified coherent patterns associated with the transitions between the seasons. In addition, we conclude that the strongest colony is most efficient in trying to maintain the microclimate of the hive during the winter.

Keywords: Data mining, Pattern recognition, Clustering, Honey bees, *Apis mellifera*, Precision beekeeping.

LISTA DE FIGURAS

Figura 1 – Diagrama que ilustra as atividades da metodologia proposta.	16
Figura 2 – Primeiro protótipo do nó sensor em detalhe e implantado em uma colmeia. .	29
Figura 3 – Topologia utilizada na tentativa de implantação do primeiro protótipo. . . .	29
Figura 4 – Segundo protótipo do nó sensor em detalhe.	30
Figura 5 – Módulo SIM800L em detalhe.	31
Figura 6 – Representação dos quartis. Fonte: http://www.texample.net/tikz/examples/box-and-whisker-plot	33
Figura 7 – <i>Clusters</i> do primeiro período de Arnas <i>I</i>	38
Figura 8 – <i>Clusters</i> do segundo período de Arnas <i>I</i>	38
Figura 9 – Médias mensais da temperatura e umidade ambientes - Arnas <i>II</i>	39
Figura 10 – <i>Clusters</i> do primeiro período de Arnas <i>II</i>	40
Figura 11 – <i>Clusters</i> do segundo período de Arnas <i>II</i>	41
Figura 12 – <i>Clusters</i> do primeiro período de Emil.	43
Figura 13 – <i>Clusters</i> do segundo período de Emil.	43
Figura 14 – Médias mensais da temperatura e umidade ambientes - Emil.	44

LISTA DE TABELAS

Tabela 1 – Principais aspectos dos trabalhos relacionados.	27
Tabela 2 – Informações sobre os conjuntos de dados.	32
Tabela 3 – Períodos nos quais os conjuntos de dados foram divididos.	35
Tabela 4 – Centroides dos <i>clusters</i> obtidos para o 1º período de Arnas I.	37
Tabela 5 – Centroides dos <i>clusters</i> obtidos para o 2º período de Arnas I.	37
Tabela 6 – Centroides dos <i>clusters</i> obtidos para o 1º período de Arnas II.	39
Tabela 7 – Centroides dos <i>clusters</i> obtidos para o 2º período de Arnas II.	40
Tabela 8 – Centroides dos <i>clusters</i> obtidos para o 1º período de Emil.	42
Tabela 9 – Centroides dos <i>clusters</i> obtidos para o 2º período de Emil.	42

SUMÁRIO

1	INTRODUÇÃO	13
1.1	Contextualização e motivação	13
1.2	Objetivos	16
1.3	Produção científica	17
1.4	Organização da dissertação	17
2	REFERENCIAL TEÓRICO	18
2.1	Apicultura	18
<i>2.1.1</i>	<i>Apicultura de precisão</i>	<i>18</i>
<i>2.1.2</i>	<i>Termorregulação</i>	<i>19</i>
<i>2.1.3</i>	<i>Atributos do monitoramento na apicultura de precisão</i>	<i>19</i>
2.2	Mineração de dados	21
<i>2.2.1</i>	<i>Algoritmo K-means</i>	<i>22</i>
2.3	Trabalhos relacionados	23
3	MATERIAL E MÉTODOS	28
3.1	Ensaio para coleta de dados	28
<i>3.1.1</i>	<i>Protótipo 1</i>	<i>28</i>
<i>3.1.2</i>	<i>Protótipo 2</i>	<i>30</i>
3.2	Conjuntos de dados	31
3.3	Pré-processamento dos dados	32
<i>3.3.1</i>	<i>Remoção de anomalias</i>	<i>32</i>
<i>3.3.2</i>	<i>Normalização dos dados</i>	<i>33</i>
<i>3.3.3</i>	<i>Validação da clusterização</i>	<i>33</i>
3.4	Avaliação experimental	34
4	RESULTADOS	37
4.1	Arnas I	37
4.2	Arnas II	39
4.3	Emil	42
5	CONCLUSÃO	45
5.1	Contribuição	45
5.2	Limitações e trabalhos futuros	46

REFERÊNCIAS	47
--------------------------	----

1 INTRODUÇÃO

Esta dissertação apresenta uma proposta de metodologia para identificar e caracterizar os padrões comportamentais sazonais de colônias de abelhas *Apis mellifera* através da aplicação de técnicas de mineração de dados.

A Seção 1.1 contextualiza a importância das abelhas na natureza e justifica a realização deste trabalho, cujos objetivos geral e atividades (objetivos específicos) são abordados na Seção 1.2. A Seção 1.3 apresenta os artigos científicos resultantes da pesquisa desenvolvida. Por fim, a Seção 1.4 apresenta a organização deste documento.

1.1 Contextualização e motivação

As abelhas são consideradas os principais agentes polinizadores da maioria das espécies de plantas silvestres e culturas polinizadas por insetos (KEVAN; 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; TSCHARNTKE *et al.*, 2012), o que, conseqüentemente, 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. Conseqüentemente muitas colônias foram perdidas (KRIDI *et al.*, 2014; KRIDI *et al.*, 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; 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; 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; 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; KRIDI *et al.*, 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; 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; 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; 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; TIAN, 2015).

1.2 Objetivos

Este trabalho tem como objetivo geral a proposição de uma metodologia para, através da aplicação de técnicas de mineração de dados, 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, foram estabelecidas as seguintes atividades, ilustradas na Figura 1:

- i. obter conjuntos de dados de temperatura, umidade e massa de colônias de abelhas *Apis mellifera* que compreendam um ciclo anual completo (Seção 3.2);
- ii. fragmentar os conjuntos de dados nos dois períodos normalmente encontrados no ciclo anual das colônias (Seção 3.4);
- iii. detectar e remover as anomalias dos dados (explicação do algoritmo na Subseção 3.3.1 e execução na Seção 3.4);
- iv. normalizar os dados (explicação do algoritmo na Subseção 3.3.2 e execução na Seção 3.4);
- v. reconhecer a quantidade ideal de padrões para cada período (explicação do algoritmo na Subseção 3.3.3 e execução na Seção 3.4);
- vi. caracterizar e interpretar cada padrão encontrado (explicação do algoritmo na Subseção 2.2.1 e execução no Capítulo 4).

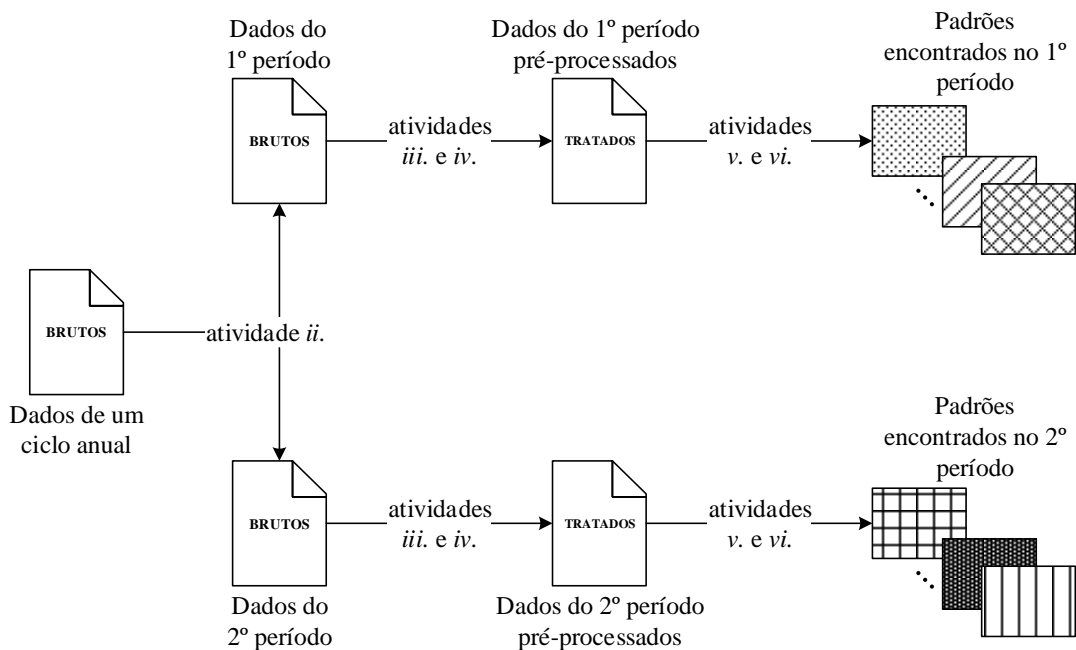


Figura 1 – Diagrama que ilustra as atividades da metodologia proposta.

1.3 Produção científica

Até o presente momento, as publicações referentes a esta dissertação foram dois artigos em conferências:

1. **MACIEL, F. A. O.**; BRAGA, A. R.; XAVIER, R.; SILVA, T. L. C. da; FREITAS, B. M.; GOMES, D. G. *Minerando dados para caracterizar padrões sazonais de colônias de abelhas *Apis mellifera**. In: SBC. SBSI'18: XIV Simpósio Brasileiro de Sistemas de Informação. Caxias do Sul/RS, Brasil: UCS, 2018 (MACIEL *et al.*, 2018b).
2. **MACIEL, F. A. O.**; BRAGA, A. R.; SILVA, A. L.; SILVA, T. L. C. da; FREITAS, B. M.; GOMES, D. G. *Reconhecimento de padrões de colônias de abelhas *Apis mellifera* segundo mudanças das estações do ano*. In: SBC. Anais do IX Workshop de Computação Aplicada à Gestão do Meio Ambiente e Recursos Naturais / XXXVIII Congresso da Sociedade Brasileira de Computação. Natal/RN, Brasil: UERN, 2018 (MACIEL *et al.*, 2018a).

1.4 Organização da dissertação

A estrutura desta dissertação está organizada da seguinte forma:

- **Capítulo 2:** apresenta a fundamentação teórica sobre os principais temas abordados nesta dissertação e discute os trabalhos relacionados mais relevantes;
- **Capítulo 3:** descreve os aspectos metodológicos da pesquisa realizada em relação ao ferramental utilizado nos ensaios para coleta de dados e no pré-processamento e análise dos dados. Além disso, descreve os experimentos realizados;
- **Capítulo 4:** apresenta e discute os resultados obtidos com a análise dos dados, interpretando-os de acordo com as características sazonais de cada período observado;
- **Capítulo 5:** sintetiza as principais conclusões e discute suas limitações e perspectivas de melhorias para trabalhos futuros.

2 REFERENCIAL TEÓRICO

Neste capítulo desenvolve-se a fundamentação teórica dos temas tratados nesta dissertação. A Seção 2.1 apresenta os conceitos referentes à apicultura, enquanto a Seção 2.2 aborda os conceitos mais relevantes referentes à mineração de dados para o contexto deste trabalho. A Seção 2.3 discute 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 (ZOGOVIĆ *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 (ZOGOVIĆ *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 (ZOGOVIĆ *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 (ZOGOVIĆ *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; 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; HOLST, 2015; ZOGOVIĆ *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; 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 vibrando rapidamente as asas em posições específicas da colmeia para gerar uma ventilação efetiva (ABOU-SHAARA *et al.*, 2017; SOUTHWICK; 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; 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; 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; WUNSCH, 2005). Não há consenso sobre uma definição completa de clusterização (XU; TIAN, 2015), mas a dada por Jain e 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; 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; 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; 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; WUNSCH, 2005).

A Subseção 2.2.1 detalha 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 (2.1), em que p é a dimensão do vetor de dados.

$$W = \{w_i\}_{i=1}^K \mid w_i \in \mathbb{R}^p. \quad (2.1)$$

O *cluster* V associado a cada centroide w é definido por (2.2), em que x é um vetor de atributos, e $\|x - w\|$ indica a distância euclidiana.

$$V_i = \{x \in \mathbb{R}^p \mid \|x - w_i\| < \|x - w_j\|, \forall j \neq i\}. \quad (2.2)$$

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 (2.2); então calcula-se a nova posição de cada centroide w_i como a média dos N_i objetos do *cluster* V_i (2.3).

$$w_i = \frac{1}{N_i} \sum_{x \in V_i} x. \quad (2.3)$$

O *cluster* V_i (2.2) e a nova posição do centroide w_i (2.3) 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 (2.4). Este erro ajuda a avaliar a qualidade dos clusters gerados.

$$SSD = \sum_{\forall x \in V_i} \|x - w_i\|^2. \quad (2.4)$$

2.3 Trabalhos relacionados

Diversos trabalhos podem ser encontrados na literatura referentes a sistemas de monitoramento de apiários e apicultura de precisão. Determinados autores abordam somente o aspecto ferramental dos sistemas, enquanto outros também exploram a análise realizada com os dados coletados. Alguns trabalhos são discutidos a seguir.

Zacepins e Karasha (2012) descreveram o desenvolvimento de um sistema *web* para o monitoramento remoto em tempo real da temperatura de colmeias. No entanto, este trabalho focou na parte do desenvolvimento ferramental do sistema, e não houve contribuição em relação à análise dos dados monitorados. Em um trabalho posterior, Zacepins *et al.* (2016) propuseram um algoritmo de suporte à decisão que pode ser usado para detecção automática do momento do abandono da colmeia. Para coletar os dados de temperatura, os autores monitoraram durante quatro meses, dez colmeias com um sistema fundamentado na plataforma Raspberry Pi.

Mais recentemente, Zacepins *et al.* (2017) descreveram a inclusão dos princípios do paradigma da Internet das Coisas (IoT - *Internet of Things*) na apicultura de precisão, com a utilização de componentes energeticamente eficientes e conexão através das redes de telefonia móvel. Similar ao trabalho de (ZACEPINS; KARASHA, 2012), não houve contribuição em relação à análise dos dados.

Kridi *et al.* (2014) e Kridi *et al.* (2016) apresentam um sistema de monitoramento da temperatura do interior de colmeias baseado em redes de sensores sem fio capaz de detectar o

comportamento de pré-abandono da colônia. Os autores utilizaram a plataforma de prototipagem Arduino para desenvolver as redes de sensores sem fio. Já na etapa de análise dos dados, foi aplicado o *K-means* para o reconhecimento de padrões que pudessem identificar o aumento anormal da temperatura que resulta no abandono da colmeia.

Também através de dados da temperatura do interior de colmeias, Kviešis e Zacepins (2016) identificaram os padrões dos estados de abandono da colmeia e de produção das crias com o uso de redes neurais. Para a obtenção dos dados, os autores utilizaram termômetros digitais conectados à plataforma Raspberry Pi.

Fitzgerald *et al.* (2015) descreveram o desenvolvimento de uma balança sem fio fundamentada no microcontrolador ATmega1281 para monitorar a massa de colmeias. Os autores afirmam que a massa é um fator chave para se avaliar a força e a produtividade de uma colmeia, mas nenhuma análise é feita nesse sentido. Em uma abordagem semelhante, Ruan *et al.* (2017) apresentaram um sistema de monitoramento de massa de colmeias baseado no paradigma da IoT e fundamentado na plataforma de prototipagem Arduino. Os autores afirmam que no futuro os dados de massa podem ser utilizados para se obter a correlação entre a produção de mel e diferentes parâmetros ecológicos, como as condições meteorológicas.

Gil-Lebrero *et al.* (2016) projetaram um sistema de monitoramento de colmeias, também baseado em redes de sensores sem fio, para coleta dos dados de temperatura, umidade relativa e massa. Segundo os autores, a diferença deste trabalho para os outros é a aquisição sincronizada das amostras de todas as colmeias de um apiário, o que é fundamental para comparar as análises futuras dos dados de colmeias diferentes. Apesar de não apresentar uma contribuição relacionada à análise dos dados, os autores constataram a habilidade das abelhas de regular a temperatura e a umidade da colmeia em situações de teste. Já a partir da observação dos dados da massa, os autores puderam verificar a evolução da colônia durante a época de floração.

Também é possível identificar os sinais do abandono da colmeia através dos sinais de vibração de uma colmeia. Bencsik *et al.* (2011) introduziram acelerômetros nas paredes de colmeias para, através da técnica de análise de componentes principais (PCA - *Principal Component Analysis*), identificar uma frequência de vibração altamente específica para o processo de abandono do enxame. Em uma abordagem semelhante, mas com os acelerômetros posicionados no centro da colmeia, Bencsik *et al.* (2015) utilizaram a técnica de análise de função discriminante (DFA - *Discriminant Function Analysis*) para verificar a correlação entre a amplitude vibracional e o ciclo de produção de crias.

O som é outro fator que pode indicar quando uma colônia irá abandonar a colmeia. Murphy *et al.* (2015a) desenvolveram uma rede de sensores sem fio fundamentada no microcontrolador ATmega1281 para monitorar e processar os sinais dos sons de colônias de abelhas e identificar os sinais de um iminente abandono da colmeia. Segundo os autores, 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. Além do som, Murphy *et al.* (2015b) também utilizaram acelerômetros e imagens em infravermelho do interior de colmeias para projetar um sistema de monitoramento autônomo que emite alertas em casos de situações críticas. Para processar os sinais de vídeo, os autores utilizaram a plataforma Raspberry Pi.

Em um trabalho mais recente, Murphy *et al.* (2016) utilizaram uma rede de sensores sem fio para descrever as condições internas de uma colônia. Os dados de diversos parâmetros foram coletados: gases CO_2 (dióxido de carbono) e O_2 (oxigênio), gases poluentes, temperatura e umidade relativa. Além disso, os dados climáticos (temperatura externa, índice pluviométrico e intensidade de luz solar) também foram monitorados para fornecer uma dimensão complementar para a análise. Um estudo biológico dos dados foi realizado para classificar dez possíveis estados da colônia, e a partir desse estudo os autores implementaram um algoritmo de classificação baseado em árvores de decisão de 95,38% de precisão.

Chazette *et al.* (2016) implementaram um sistema de monitoramento de colmeias fundamentado na plataforma Raspberry Pi para coleta de dados de temperatura, massa, áudio e vídeo. A análise foi realizada através da aplicação de um algoritmo de reconhecimento de imagens para observar a infestação da colônia pelo ácaro *Varroa destructor*. Além do sistema de monitoramento, os autores descreveram a prototipagem de um sistema de atuação para matar os ácaros detectados com o uso de *lasers*.

Tashakkori e Ghadiri (2015) apresentaram um sistema de monitoramento que utiliza câmeras posicionadas na entrada das colmeias e estima o fluxo de abelhas através de técnicas de processamento de imagens. Os autores afirmam que o nível de atividade das abelhas ao redor da colmeia pode ser um bom indicador da saúde da colônia. No entanto nenhuma conclusão mais categórica foi feita nesse sentido.

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.

Tabela 1 – Principais aspectos dos trabalhos relacionados.

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

3 MATERIAL E MÉTODOS

Este capítulo descreve os aspectos metodológicos da pesquisa realizada em relação ao ferramental utilizado nos ensaios para coleta de dados, obtenção dos conjuntos de dados, estratégias de pré-processamento e análise dos dados e a avaliação experimental.

3.1 Ensaio para coleta de dados

Foram implementados dois protótipos de nós sensores e estação base que seriam instalados nas colmeias para o monitoramento e coleta dos dados.

Os ensaios para implantação dos protótipos foram conduzidos no apiário do Grupo de Pesquisa com Abelhas do Setor de Abelhas¹ do Departamento de Zootecnia da Universidade Federal do Ceará (344°33.7”S, 3834°46.5”O). As colmeias são do tipo Langstroth americana, de madeira e com dimensões externas 508 x 416 x 241 mm e dimensões internas 464 x 372 x 232 mm. As abelhas são da espécie *Apis mellifera*.

3.1.1 Protótipo 1

O primeiro protótipo foi fundamentado na plataforma Arduino Mega, que utiliza o microcontrolador ATmega2560 e tem um consumo mínimo de aproximadamente 43 mA, de acordo com o site² do fabricante. Nele foram integrados sensores de temperatura, umidade, ruídos e gás dióxido de carbono (CO₂). A Figura 2 mostra registros do nó sensor em detalhes, antes de ser implantado, e após ser implantado em uma colmeia.

Para as medições de temperatura e umidade foi utilizado o sensor DHT22. Ele apresenta uma faixa de medição de 0 a 100% com resolução de 0.1% para umidade relativa e de -40 °C a 80 °C com resolução de 0.1 °C para temperatura.

Para as medições do nível de ruído no interior da colmeia foi utilizado o sensor de som KY-038. Seu princípio de funcionamento baseia-se na variação da tensão de saída de acordo com a intensidade sonora, conforme a regulagem do potenciômetro presente no módulo.

Para a coleta dos dados a respeito da concentração de CO₂ utilizou-se o sensor MQ-135, cuja sensibilidade também pode ser ajustada através de um potenciômetro.

Por terem um tamanho reduzido e, no caso do DHT22, ser capaz de coletar duas

¹ <http://www.abelhas.ufc.br>

² <http://www.arduino.cc>



Figura 2 – Primeiro protótipo do nó sensor em detalhe e implantado em uma colmeia.

grandezas com somente um módulo, esses sensores obedeciam o requisito de serem minimamente invasivos para as colônias.

O rádio escolhido para comunicação foi o NRF24L01+ com antena externa de 2 dB, que opera na banda de 2.4 GHz e tem potência máxima de 20 dBmW. De acordo com seu *datasheet*, o alcance para a comunicação varia entre 520 m a 1000 m dependendo da taxa de transmissão dos dados e da quantidade e natureza dos obstáculos entre os módulos comunicantes. No caso do apiário onde os ensaios foram conduzidos, as colmeias são cercadas por uma vegetação densa, o que impediu a comunicação direta entre o nó sensor e a estação base (*gateway*). Para contornar esse problema, foram utilizados dois nós repetidores, compostos apenas por um Arduino e um rádio, que possibilitavam a transmissão do sinal por uma rota livre. A Figura 3 ilustra o cenário.

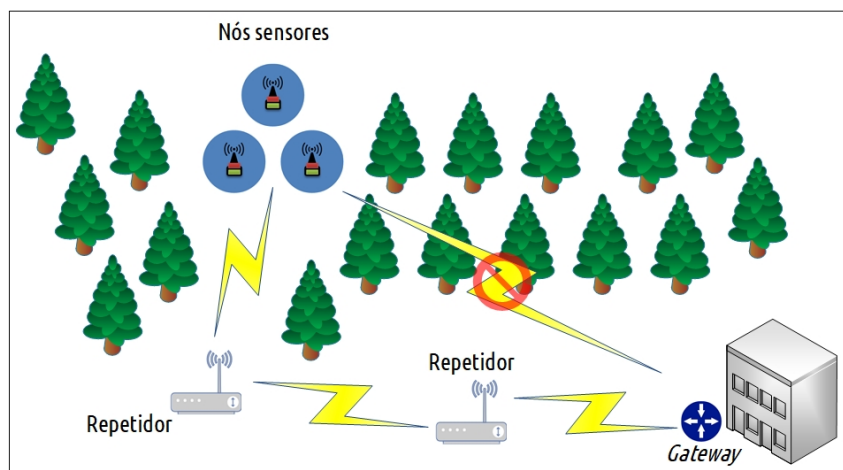


Figura 3 – Topologia utilizada na tentativa de implantação do primeiro protótipo.

A estação base era composta por um Arduino, que recebia os dados monitorados pelos nós sensores através do módulo de rádio NRF24L01+, e por um BeagleBone, que possibilitava o acesso à Internet para que os dados fossem salvos em uma nuvem computacional. Ela foi instalada no bloco didático mais próximo ao apiário, a cerca de 50 metros, com fornecimento de energia e acesso à Internet permanentes.

Este protótipo, entretanto, apresentou problemas em relação ao consumo energético dos nós sensores e à estabilidade de conexão entre os componentes da rede, mesmo com o uso dos nós repetidores. Além disso, os sensores KY-038 e MQ-135 não apresentaram um funcionamento satisfatório, e nem mesmo em testes em ambientes controlados foi possível obter dados válidos.

3.1.2 Protótipo 2

O segundo protótipo foi implementado com a perspectiva de se solucionar os problemas apresentados pelo primeiro protótipo.

A primeira intervenção para a diminuição do consumo energético dos nós sensores foi a substituição do Arduino Mega pelo Arduino Nano. Este utiliza o microcontrolador ATmega328 e tem um consumo mínimo de aproximadamente 19 mA. Ademais, os sensores KY-038 e MQ-135 foram removidos da arquitetura, reduzindo a quantidade de componentes para serem alimentados. A Figura 4 mostra um registro do nó sensor do segundo protótipo em detalhes.

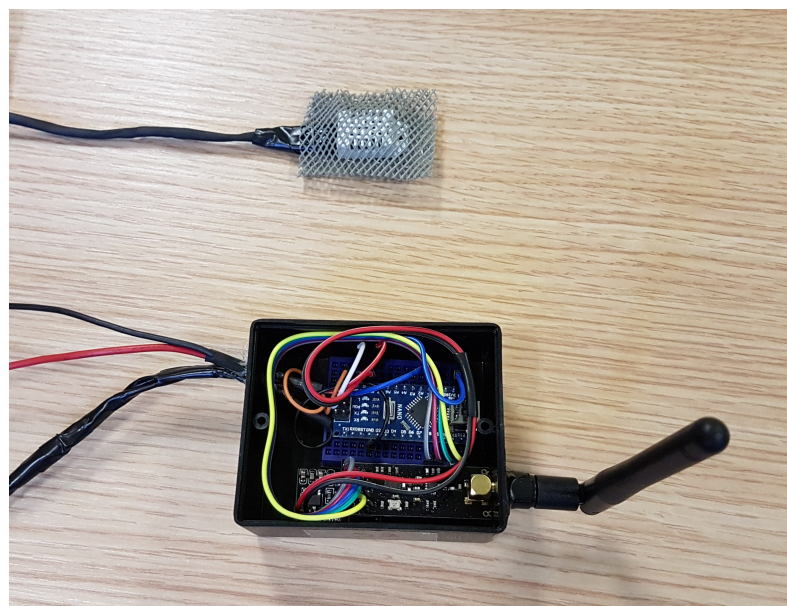


Figura 4 – Segundo protótipo do nó sensor em detalhe.

Para contornar o problema de comunicação entre o nó sensor e a estação base, a topologia foi modificada com o deslocamento do *gateway* para dentro do próprio apiário. Para manter a conexão com a nuvem computacional na qual os dados estavam sendo salvos, o módulo SIM800L, apresentado na Figura 5, foi integrado ao *gateway*. Esse módulo permite o envio e recebimento de dados através da rede de telefonia celular. O SIM800L foi integrado diretamente no Arduino do *gateway*, dispensando o uso do BeagleBone. Assim, a estação base passou a ser composta por um Arduino, um módulo de rádio NRF24L01+ e um módulo SIM800L.



Figura 5 – Módulo SIM800L em detalhe.

A alteração realizada na topologia da rede sanou o problema de conexão entre os nós sensores e a estação base, e, conseqüentemente, possibilitou a remoção dos nós repetidores. No entanto, as intervenções feitas visando a diminuição do consumo energético não foram capazes de proporcionar ao segundo protótipo uma autonomia energética satisfatória.

Recentemente, outras intervenções a nível de *software*, como a utilização de rotinas focadas na economia de energia, foram efetuadas. Os resultados dos testes executados em laboratório foram promissores, e a autonomia do protótipo aumentou consideravelmente. No entanto, essas melhorias ainda não foram testadas no apiário.

3.2 Conjuntos de dados

Após as tentativas frustradas de se obter dados próprios, buscou-se como alternativa utilizar dados disponibilizados na Internet. Os conjuntos de dados mais completos foram obtidos no portal HiveTool.net³. HiveTool é um projeto de código aberto cujo o objetivo é produzir

³ <http://www.hivetool.net/>

ferramentas de software e hardware 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. No entanto, 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 na cidade de Rebild, na Dinamarca. O terceiro conjunto de dados é referente a uma colmeia do apiário Emil (58° 19' 9.1596" N, 8° 32' 9.8124" L), localizado na cidade de Grimstad, na 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. A Tabela 2 sintetiza as informações a respeito dos conjuntos de dados.

Tabela 2 – Informações sobre os conjuntos de dados.

Apiário	País	Número de amostras	Início	Fim
Arnas I	Dinamarca	94.552	01/12/2016	30/11/2017
Arnas II	Dinamarca	92.749	01/03/2017	28/02/2018
Emil	Noruega	88.181	01/03/2017	28/02/2018

3.3 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 do algoritmo em si (XU; WUNSCH, 2005).

3.3.1 Remoção de anomalias

A remoção de anomalias foi realizada pelo método de Tukey (TUKEY, 1977). Neste método é definido um intervalo com limites inferior e superior dos atributos do conjunto de dados, e as amostras que estiverem fora desse intervalo são consideradas anomalias. O limite inferior é calculado por (3.1), e o limite superior por 3.2, em que Q_1 e Q_3 são respectivamente

o primeiro e o terceiro quartis de um atributo do conjunto de dados. Uma representação do intervalo e dos quartis pode ser observada na Figura 6.

$$Q_1 - 1.5 \times (Q_3 - Q_1). \quad (3.1)$$

$$Q_3 + 1.5 \times (Q_3 - Q_1). \quad (3.2)$$

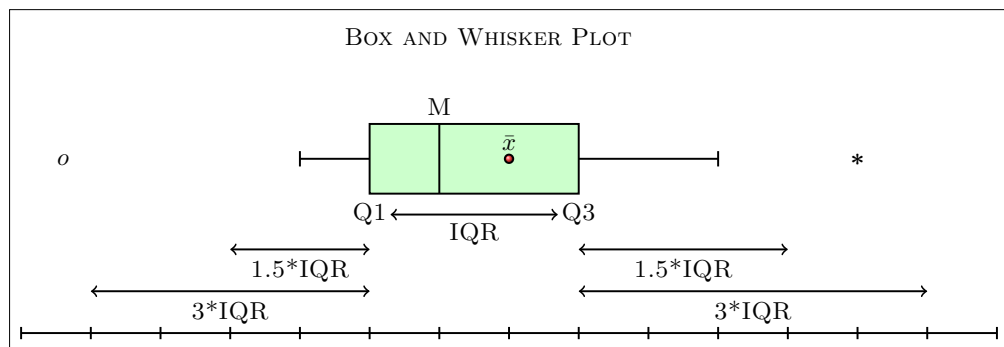


Figura 6 – Representação dos quartis. Fonte: <http://www.texample.net/tikz/examples/box-and-whisker-plot>

3.3.2 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 (3.3), em que μ e σ são respectivamente a média e o desvio padrão de X .

$$Z = \frac{x - \mu}{\sigma}. \quad (3.3)$$

3.3.3 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; 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; 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 e 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; HARABASZ, 1974) obteve os melhores resultados. O cálculo deste índice é definido por (3.4), em que B_K e W_K são, respectivamente, as matrizes de dispersão entre grupos e intragrupo, das quais são calculados os traços (operador $tr()$).

$$CH(K) = \frac{tr(B_K)/(K-1)}{tr(W_K)/(N-K)}. \quad (3.4)$$

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.

3.4 Avaliação experimental

Cada conjunto de dados descrito na Seção 3.2 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.

A Tabela 3 sintetiza as informações a respeito dos períodos nos quais os conjuntos de dados foram divididos. 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 da metodologia 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 no Capítulo 1.

Tabela 3 – Períodos nos quais os conjuntos de dados foram divididos.

Apiário	1º período	Estações 1º período	2º período	Estações 2º período
Arnas I	12/16 a 05/17	Inverno e primavera	06/17 a 11/17	Verão e outono
Arnas II	03/17 a 08/17	Primavera e verão	09/17 a 02/18	Outono e inverno
Emil	03/17 a 08/17	Primavera e verão	09/17 a 02/18	Outono e inverno

O procedimento para os experimentos se deu pela execução das etapas de pré-processamento dos dados, discutidas na Seção 3.3, e pela aplicação do algoritmo *K-means* discutido no Capítulo 2, 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*⁴ e *pandas*⁵, para

⁴ <http://www.numpy.org/>

⁵ <https://pandas.pydata.org/>

manipulação e análise de estruturas de dados.

4 RESULTADOS

Neste capítulo, os resultados encontrados pela aplicação da metodologia descrita no Capítulo 3 aos conjuntos de dados são apresentados e discutidos.

4.1 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, a metodologia utilizada retornou 2 *clusters* como resultado. Seus centroides são apresentados respectivamente na Tabela 4 e na Tabela 5.

Tabela 4 – Centroides dos *clusters* obtidos para o 1º período de Arnas I.

	Temperatura interna (°C)	Umidade interna (%)	Massa da colmeia (kg)
<i>Cluster 0</i>	13.1	79.6	27.7
<i>Cluster 1</i>	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 5 – Centroides dos *clusters* obtidos para o 2º período de Arnas I.

	Temperatura interna (°C)	Umidade interna (%)	Massa da colmeia (kg)
<i>Cluster 0</i>	30.3	55.5	15.8
<i>Cluster 1</i>	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 7a e pela Figura 8a, que mostram a proporção da quantidade de amostras de cada *cluster* durante os meses do primeiro e do segundo período, respectivamente. Já a Figura 7b e a Figura 8b mostram a distribuição das

amostras dos *clusters* no espaço de dados do primeiro e do segundo período, respectivamente.

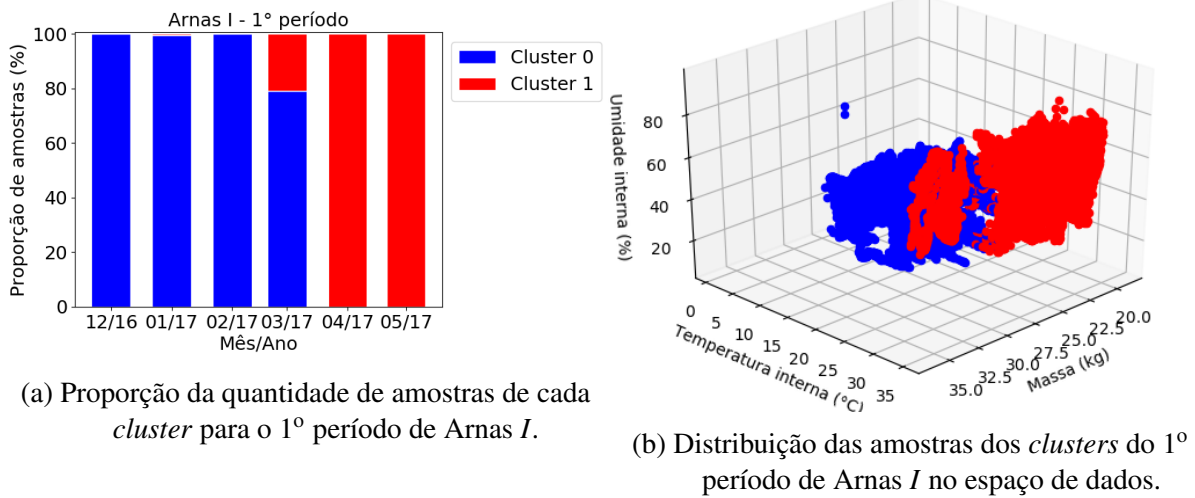


Figura 7 – *Clusters* do primeiro período de Arnas I.

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

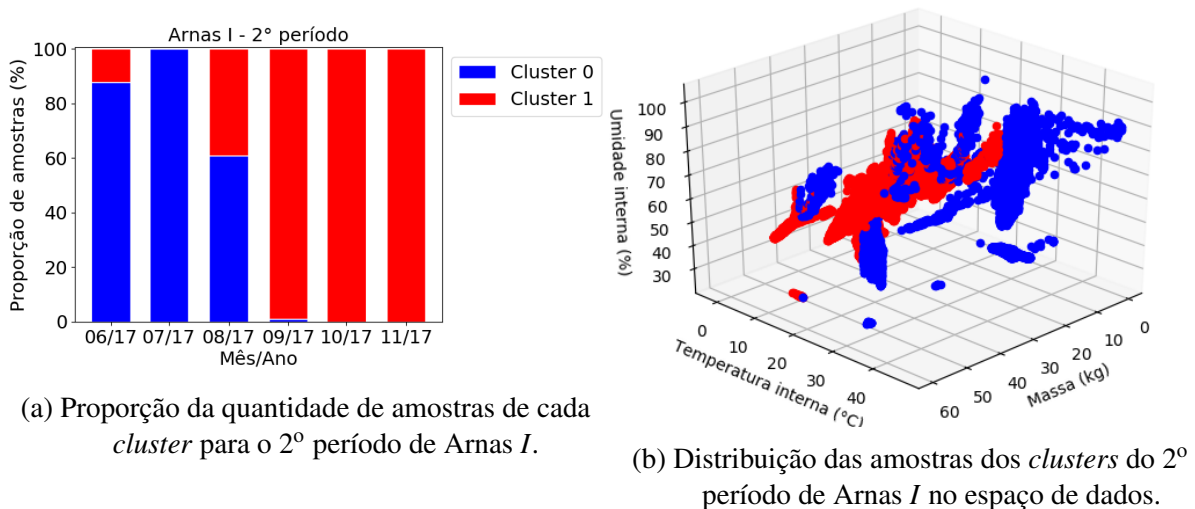


Figura 8 – *Clusters* do segundo período de Arnas I.

4.2 Arnas II

Para o primeiro período de Arnas II, de março a agosto de 2017, a metodologia utilizada retornou como resultado 5 *clusters*. Seus centroides são apresentados na Tabela 6.

Tabela 6 – Centroides dos *clusters* obtidos para o 1º período de Arnas II.

	Temperatura interna (°C)	Umidade interna (%)	Massa da colmeia (kg)
Cluster 0	14.3	89.2	25.3
Cluster 1	29.9	53.3	23.3
Cluster 2	30.8	54.0	11.1
Cluster 3	33.6	58.4	37.3
Cluster 4	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 9.

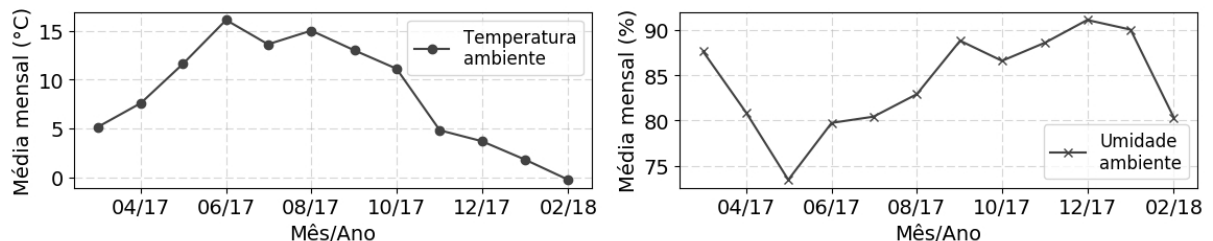
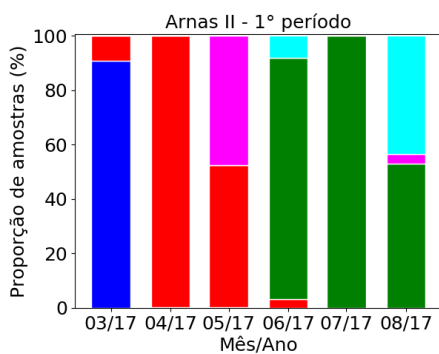


Figura 9 – Médias mensais da temperatura e umidade ambientes - Arnas II.

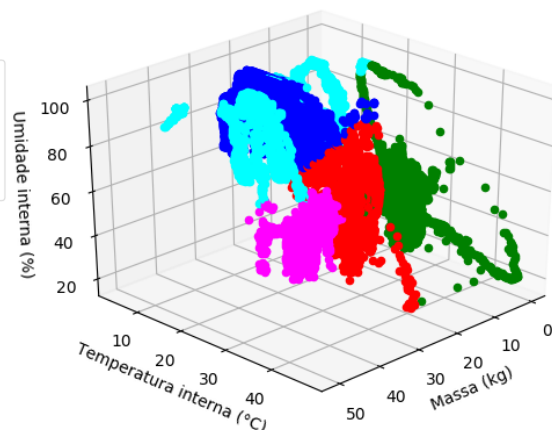
O *Cluster 1* pode ser entendido como o estado da colônia em meados da primavera. Isso é evidenciado principalmente pela baixa umidade, já que nessa estação as precipitações, e consequentemente a umidade, apresentam médias baixas, como podemos observar na Figura 9. O *Cluster 2* pode ser interpretado 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.

Esses resultados podem ser comprovados pela Figura 10a, 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 10b mostra a distribuição das amostras dos *clusters* do primeiro período de Arnas II no espaço de dados.



(a) Proporção da quantidade de amostras de cada *cluster* para o 1º período de Arnas II.



(b) Distribuição das amostras dos *clusters* do 1º período de Arnas II no espaço de dados.

Figura 10 – *Clusters* do primeiro período de Arnas II.

Para o segundo período de Arnas II, de setembro de 2017 a fevereiro de 2018, a metodologia utilizada retornou como resultado 3 *clusters*, cujos centroides são apresentados na Tabela 7.

Tabela 7 – Centroides dos *clusters* obtidos para o 2º período de Arnas II.

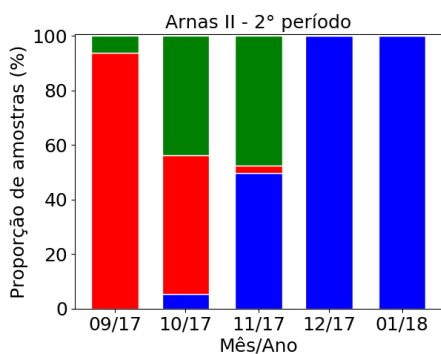
	Temperatura interna (°C)	Umidade interna (%)	Massa da colmeia (kg)
Cluster 0	10.3	78.1	31.0
Cluster 1	13.4	93.3	31.2
Cluster 2	7.7	96.0	29.5

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

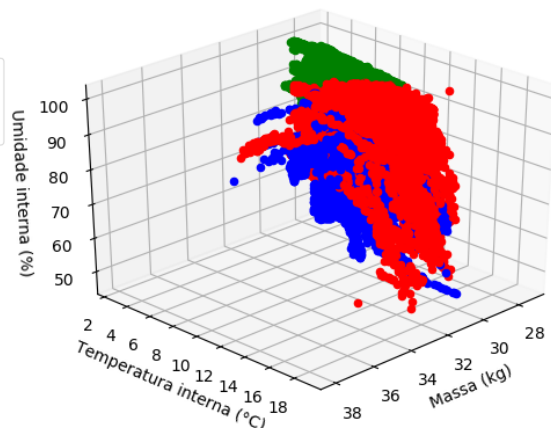
dois, é possível presumir que ele refere-se à transição do verão para o outono, já que a média da umidade é maior no outono, como podemos observar na Figura 9.

Complementarmente, 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 9.

Esses resultados podem ser comprovados pela Figura 11a. 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 11b mostra a distribuição das amostras dos *clusters* do segundo período de Arnas II no espaço de dados.



(a) Proporção da quantidade de amostras de cada *cluster* para o 2º período de Arnas II.



(b) Distribuição das amostras dos *clusters* do 2º período de Arnas II no espaço de dados.

Figura 11 – *Clusters* do segundo período de Arnas II.

É importante destacar que, se a temperatura da colônia seguisse a tendência da temperatura ambiente observada na Figura 9, 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, 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; BREED, 2013).

4.3 Emil

Para o primeiro período de Emil, de março a agosto de 2017, a metodologia retornou como resultado 3 *clusters*. Seus centroides são apresentados na Tabela 8.

Tabela 8 – Centroides dos *clusters* obtidos para o 1º período de Emil.

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

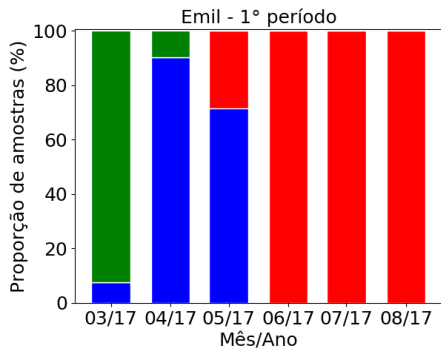
Similar à avaliação feita para a colmeia de Arnas, pode-se interpretar o *Cluster 2* como o estado da colônia durante a transição entre as épocas fria e quente do ano devido ao baixo valor da temperatura. Ainda levando em consideração o valor da temperatura, pode-se interpretar o *Cluster 0* como o estado 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 12a. 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 12b mostra a distribuição das amostras dos *clusters* do primeiro período de Emil no espaço de dados.

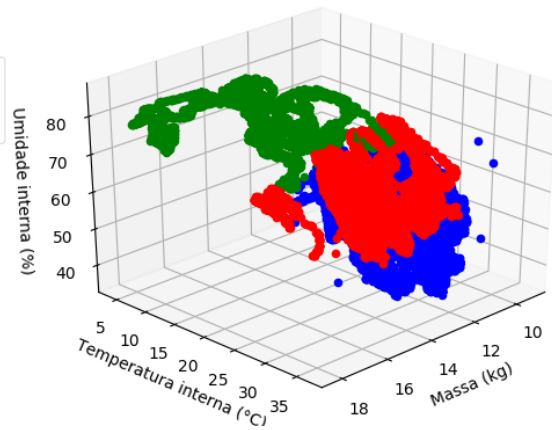
Para o segundo período, de setembro de 2017 a fevereiro de 2018, a metodologia apresentada retornou como resultado 6 *clusters*, cujos centroides são apresentados na Tabela 9.

Tabela 9 – Centroides dos *clusters* obtidos para o 2º período de Emil.

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

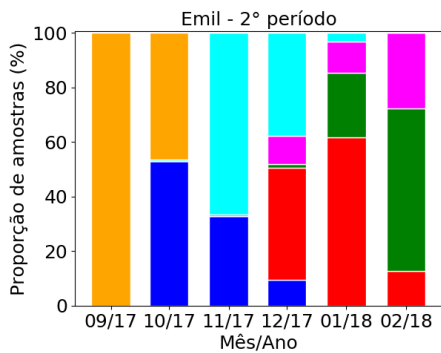


(a) Proporção da quantidade de amostras de cada *cluster* para o 1º período de Emil.

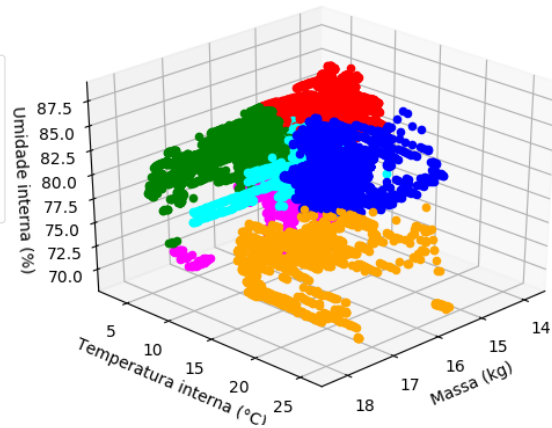


(b) Distribuição das amostras dos *clusters* do 1º período de Emil no espaço de dados.

Figura 12 – *Clusters* do primeiro período de Emil.



(a) Proporção da quantidade de amostras de cada *cluster* para o 2º período de Emil.



(b) Distribuição das amostras dos *clusters* do 2º período de Emil no espaço de dados.

Figura 13 – *Clusters* do segundo período de Emil.

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 corresponde ao outono e inverno.

Percebe-se pela Figura 13a 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 14. 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 13b mostra a distribuição das amostras dos *clusters* do segundo período de Emil no espaço de dados.

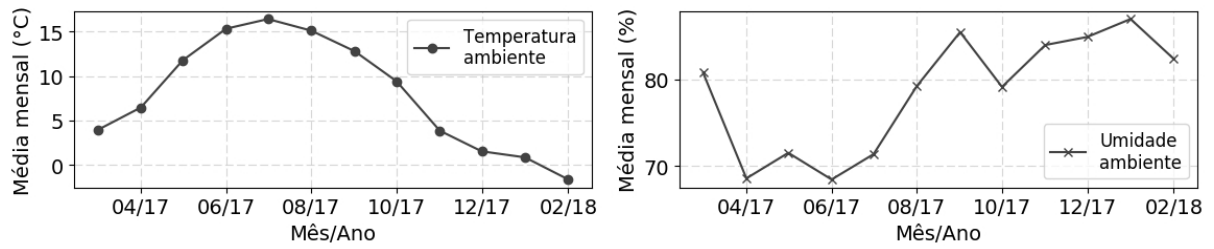


Figura 14 – Médias mensais da temperatura e umidade ambientes - Emil.

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

5 CONCLUSÃO

Esta dissertação apresenta uma metodologia para identificar e caracterizar padrões sazonais de colônias de abelhas *Apis mellifera* através da aplicação de técnicas de mineração de dados. Foram utilizados três conjuntos de dados do portal HiveTool.net para a aplicação da metodologia. Dois conjuntos de dados, um referente a uma colônia na Dinamarca e outro referente a uma colônia na Noruega, foram divididos de acordo com os dois períodos do ciclo anual de colônias de abelhas referido na literatura para clima temperado: período frio, composto por outono e inverno, e período quente, composto por primavera e verão. Para efeito comparativo, o terceiro conjunto de dados, também referente a uma colônia na Dinamarca, foi dividido em períodos que mesclavam estações frias e quentes: inverno e primavera, e verão e outono.

Para o caso em que os conjuntos de dados foram divididos de acordo com o ciclo anual referido na literatura, a aplicação da metodologia proposta resultou na identificação detalhada de padrões das colônias de abelhas *Apis mellifera* relacionados às estações de cada um dos dois períodos do ciclo (Seções 4.2 e 4.3). Os padrões foram reconhecidos como coerentes, já que correspondem ao que se observa em campo. Além disso, é possível concluir que a colmeia aparentemente mais forte consegue desempenhar com mais eficiência o controle térmico do interior da colmeia durante período frio (Seção 4.3).

Para o caso em que os conjuntos de dados foram divididos mesclando-se estações frias e quentes, a aplicação da metodologia proposta resultou na identificação de padrões que refletiam apenas as condições externas às colmeias (Seção 4.1). Assim, os resultados foram considerados inexpressivos, mostrando que a divisão do ciclo anual das colônias de abelhas referida na literatura é, de fato, apropriada.

5.1 Contribuição

A principal contribuição desta dissertação é uma metodologia para identificar e caracterizar os padrões sazonais de colônias de abelhas *Apis mellifera* através da aplicação de técnicas de mineração de dados. A maioria da literatura de apicultura de precisão considera 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. Nossa metodologia considera três variáveis (temperatura interna, umidade interna e massa da colmeia) bem como sua respectiva análise.

5.2 Limitações e trabalhos futuros

Tem-se como perspectiva a aplicação da metodologia apresentada em dados de colônias de *Apis mellifera* africanizada, um polihíbrido criado no Brasil e que carece de estudos nesse sentido. Para isso, pretende-se criar uma infraestrutura própria de monitoramento em diversas colmeias, para que se possa, inclusive, caracterizar outros padrões além daqueles relacionados com as estações do ano.

Em termos das técnicas de mineração de dados, uma possibilidade de investigação futura é a utilização da evolução de *clusters* (SPILIOPOULOU *et al.*, 2006; SILVA *et al.*, 2014). Com essa abordagem, seria possível observar a transformação dos *clusters* ao longo do tempo e, conseqüentemente, prever com maior precisão quando uma colmeia estaria por entrar em um estado não desejado. Outra possível vantagem dessa técnica seria a observação mais detalhada dos *clusters* e suas variações, permitindo a identificação de outros padrões não observados com a abordagem utilizada nesse trabalho.

REFERÊNCIAS

- ABOU-SHAARA, H.; OWAYSS, A.; IBRAHIM, Y.; BASUNY, N. A review of impacts of temperature and relative humidity on various activities of honey bees. **Insectes Sociaux**, Springer, v. 64, n. 4, p. 455–463, 2017.
- ABOU-SHAARA, H. F.; AL-GHAMDI, A. A.; MOHAMED, A. A. Tolerance of two honey bee races to various temperature and relative humidity gradients. **Environmental and experimental Biology**, v. 10, n. 4, p. 133–138, 2012.
- ALMEIDA, G. F.; GRASSI, M. L.; GRAMACHO, K. P.; GONÇALVES, L. S. Comportamento enxameatório em abelhas africanizadas induzido por temperatura com auxílio de câmara de controle de temperatura, sensores e apidômetros. In: **XVI Congresso Brasileiro de Apicultura e II Congresso Brasileiro de Meliponicultura**. Aracaju: [s.n.], 2006.
- BARRON, A. B. Death of the bee hive: understanding the failure of an insect society. **Current Opinion in Insect Science**, Elsevier, v. 10, p. 45–50, 2015.
- BENCSIK, M.; BENCSIK, J.; BAXTER, M.; LUCIAN, A.; ROMIEU, J.; MILLET, M. Identification of the honey bee swarming process by analysing the time course of hive vibrations. **Computers and Electronics in Agriculture**, Elsevier, v. 76, n. 1, p. 44–50, 2011.
- BENCSIK, M.; CONTE, Y. L.; REYES, M.; PIOZ, M.; WHITTAKER, D.; CRAUSER, D.; DELSO, N. S.; NEWTON, M. I. Honeybee colony vibrational measurements to highlight the brood cycle. **PloS One**, Public Library of Science, v. 10, n. 11, p. e0141926, 2015.
- 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.* Parallel declines in pollinators and insect-pollinated plants in britain and the netherlands. **Science**, American Association for the Advancement of Science, v. 313, n. 5785, p. 351–354, 2006.
- BOMMARCO, R.; LUNDIN, O.; SMITH, H. G.; RUNDLÖF, M. Drastic historic shifts in bumble-bee community composition in sweden. **Proceedings of the Royal Society of London B: Biological Sciences**, The Royal Society, 2011.
- CALINSKI, T.; HARABASZ, J. A dendrite method for cluster analysis. **Communications in Statistics**, Taylor & Francis, v. 3, n. 1, p. 1–27, 1974.
- CHAUZAT, P.; LAURENT, M.; RIVIÈRE, M.-P.; SAUGEON, C.; HENDRIKX, P.; RIBIÈRE-CHABERT, M. *et al.* A pan-european epidemiological study on honey bee colony losses 2012–2013. **European Union Reference Laboratory for Honeybee Health, Brussels, Rapport technique**, 2014.
- CHAZETTE, L.; BECKER, M.; SZCZEBICKA, H. Basic algorithms for bee hive monitoring and laser-based mite control. In: **IEEE. 2016 IEEE Symposium Series on Computational Intelligence (SSCI)**. [S.l.], 2016. p. 1–8.
- CHEN, M.-S.; HAN, J.; YU, P. S. Data mining: an overview from a database perspective. **IEEE Transactions on Knowledge and data Engineering**, IEEE, v. 8, n. 6, p. 866–883, 1996.
- COOK, C. N.; BREED, M. D. Social context influences the initiation and threshold of thermoregulatory behaviour in honeybees. **Animal behaviour**, Elsevier, v. 86, n. 2, p. 323–329, 2013.

- DEGUINES, N.; JONO, C.; BAUDE, M.; HENRY, M.; JULLIARD, R.; FONTAINE, C. Large-scale trade-off between agricultural intensification and crop pollination services. **Frontiers in Ecology and the Environment**, Wiley Online Library, v. 12, n. 4, p. 212–217, 2014.
- FITZGERALD, D. W.; MURPHY, F. E.; WRIGHT, W. M. D.; WHELAN, P. M.; POPOVICI, E. M. Design and development of a smart weighing scale for beehive monitoring. In: **IEEE. Signals and Systems Conference (ISSC), 2015 26th Irish**. [S.l.], 2015. p. 1–6.
- GIANNINI, T. C.; CORDEIRO, G. D.; FREITAS, B. M.; SARAIVA, A. M.; IMPERATRIZ-FONSECA, V. L. The dependence of crops for pollinators and the economic value of pollination in brazil. **Journal of Economic Entomology**, Oxford University Press, v. 108, n. 3, p. 849–857, 2015.
- GIL-LEBRERO, S.; QUILES-LATORRE, F. J.; ORTIZ-LÓPEZ, M.; SÁNCHEZ-RUIZ, V.; GÁMIZ-LÓPEZ, V.; LUNA-RODRÍGUEZ, J. J. Honey bee colonies remote monitoring system. **Sensors**, Multidisciplinary Digital Publishing Institute, v. 17, n. 1, p. 55, 2016.
- GOULSON, D.; NICHOLLS, E.; BOTÍAS, C.; ROTHERAY, E. L. Bee declines driven by combined stress from parasites, pesticides, and lack of flowers. **Science**, American Association for the Advancement of Science, v. 347, n. 6229, p. 1255957, 2015.
- JAIN, A. K.; DUBES, R. C. **Algorithms for clustering data**. [S.l.]: Prentice-Hall, Inc., 1988.
- JAIN, A. K.; DUIN, R. P. W.; MAO, J. Statistical pattern recognition: A review. **IEEE Transactions on pattern analysis and machine intelligence**, Ieee, v. 22, n. 1, p. 4–37, 2000.
- KEVAN, P.; PHILLIPS, T. The economic impacts of pollinator declines: an approach to assessing the consequences. **Conservation Ecology**, The Resilience Alliance, v. 5, n. 1, 2001.
- KLEIN, A.-M.; VAISSIERE, B. E.; CANE, J. H.; STEFFAN-DEWENTER, I.; CUNNINGHAM, S. A.; KREMEN, C.; TSCHARNTKE, T. Importance of pollinators in changing landscapes for world crops. **Proceedings of the Royal Society of London B: Biological Sciences**, The Royal Society, v. 274, n. 1608, p. 303–313, 2007.
- KLEINHENZ, M.; BUJOK, B.; FUCHS, S.; TAUTZ, J. Hot bees in empty broodnest cells: heating from within. **Journal of Experimental Biology**, The Company of Biologists Ltd, v. 206, n. 23, p. 4217–4231, 2003.
- KREMEN, C.; WILLIAMS, N. M.; THORP, R. W. Crop pollination from native bees at risk from agricultural intensification. **Proceedings of the National Academy of Sciences**, National Acad Sciences, v. 99, n. 26, p. 16812–16816, 2002.
- KRIDI, D. S.; CARVALHO, C. G. N. d.; GOMES, D. G. A predictive algorithm for mitigate swarming bees through proactive monitoring via wireless sensor networks. In: **ACM. Proceedings of the 11th ACM symposium on Performance evaluation of wireless ad hoc, sensor, & ubiquitous networks**. [S.l.], 2014. p. 41–47.
- KRIDI, D. S.; CARVALHO, C. G. N. de; GOMES, D. G. Application of wireless sensor networks for beehive monitoring and in-hive thermal patterns detection. **Computers and Electronics in Agriculture**, Elsevier, v. 127, p. 221–235, 2016.

KVIESIS, A.; ZACEPINS, A. Application of neural networks for honey bee colony state identification. In: IEEE. **Carpathian Control Conference (ICCC), 2016 17th International**. [S.l.], 2016. p. 413–417.

LI, Z.; HUANG, Z. Y.; SHARMA, D. B.; XUE, Y.; WANG, Z.; REN, B. Drone and worker brood microclimates are regulated differentially in honey bees, *apis mellifera*. **PLoS One**, Public Library of Science, v. 11, n. 2, p. e0148740, 2016.

MACIEL, F. A. O.; BRAGA, A. R.; SILVA, A. L.; SILVA, T. L. C. da; FREITAS, B. M.; GOMES, D. G. Reconhecimento de padrões de colônias de abelhas *Apis mellifera* segundo mudanças das estações do ano. In: SBC. **Anais do IX Workshop de Computação Aplicada à Gestão do Meio Ambiente e Recursos Naturais / XXXVIII Congresso da Sociedade Brasileira de Computação**. Natal/RN, Brasil: UERN, 2018.

MACIEL, F. A. O.; BRAGA, A. R.; XAVIER, R.; SILVA, T. L. C. da; FREITAS, B. M.; GOMES, D. G. Minerando dados para caracterizar padrões sazonais de colônias de abelhas *Apis mellifera*. In: SBC. **SBSI'18: XIV Simpósio Brasileiro de Sistemas de Informação**. Caxias do Sul/RS, Brasil: UCS, 2018.

MACQUEEN, J. Some methods for classification and analysis of multivariate observations. In: OAKLAND, CA, USA. **Proceedings of the fifth Berkeley symposium on mathematical statistics and probability**. [S.l.], 1967. v. 1, n. 14, p. 281–297.

MCBRATNEY, A.; WHELAN, B.; ANCEV, T.; BOUMA, J. Future directions of precision agriculture. **Precision agriculture**, Springer, v. 6, n. 1, p. 7–23, 2005.

MEIKLE, W. G.; HOLST, N. Application of continuous monitoring of honeybee colonies. **Apidologie**, Springer, v. 46, n. 1, p. 10–22, 2015.

MILLIGAN, G. W.; COOPER, M. C. An examination of procedures for determining the number of clusters in a data set. **Psychometrika**, Springer, v. 50, n. 2, p. 159–179, 1985.

MITCHELL, T. M. Machine learning and data mining. **Communications of the ACM**, ACM, v. 42, n. 11, p. 30–36, 1999.

MURPHY, F. E.; MAGNO, M.; O'LEARY, L.; TROY, K.; WHELAN, P.; POPOVICI, E. M. Big brother for bees (3b)—energy neutral platform for remote monitoring of beehive imagery and sound. In: IEEE. **Advances in Sensors and Interfaces (IWASI), 2015 6th IEEE International Workshop on**. [S.l.], 2015b. p. 106–111.

MURPHY, F. E.; MAGNO, M.; WHELAN, P. M.; O'HALLORAN, J.; POPOVICI, E. M. b+wsn: Smart beehive with preliminary decision tree analysis for agriculture and honey bee health monitoring. **Computers and Electronics in Agriculture**, Elsevier, v. 124, p. 211–219, 2016.

MURPHY, F. E.; SRBINOVSKI, B.; MAGNO, M.; POPOVICI, E. M.; WHELAN, P. M. An automatic, wireless audio recording node for analysis of beehives. In: IEEE. **Signals and Systems Conference (ISSC), 2015 26th Irish**. [S.l.], 2015a. p. 1–6.

NICOLSON, S. W. Water homeostasis in bees, with the emphasis on sociality. **Journal of Experimental Biology**, The Company of Biologists Ltd, v. 212, n. 3, p. 429–434, 2009.

OLDROYD, B. P. What's killing american honey bees? **PLoS Biol**, Public Library of Science, v. 5, n. 6, p. e168, 2007.

OLLERTON, J.; WINFREE, R.; TARRANT, S. How many flowering plants are pollinated by animals? *Oikos*, Wiley Online Library, v. 120, n. 3, p. 321–326, 2011.

OSTWALD, M. M.; SMITH, M. L.; SEELEY, T. D. The behavioral regulation of thirst, water collection and water storage in honey bee colonies. *Journal of Experimental Biology*, The Company of Biologists Ltd, v. 219, n. 14, p. 2156–2165, 2016.

PETZ, M.; STABENTHEINER, A.; CRAILSHEIM, K. Respiration of individual honeybee larvae in relation to age and ambient temperature. *Journal of comparative Physiology B*, Springer, v. 174, n. 7, p. 511–518, 2004.

PIATESKI, G.; FRAWLEY, W. **Knowledge Discovery in Databases**. Cambridge, MA, USA: MIT Press, 1991. ISBN 0262660709.

PIRES, C. S. S.; PEREIRA, F. de M.; LOPES, M. T. do R.; NOCELLI, R. C. F.; MALASPINA, O.; PETTIS, J. S.; TEIXEIRA, E. W. Enfraquecimento e perda de colônias de abelhas no Brasil: há casos de CCD? *Pesquisa Agropecuária Brasileira*, v. 51, n. 5, p. 422–442, 2016.

POTTS, S. G.; BIESMEIJER, J. C.; KREMEN, C.; NEUMANN, P.; SCHWEIGER, O.; KUNIN, W. E. Global pollinator declines: trends, impacts and drivers. *Trends in Ecology & Evolution*, Elsevier, v. 25, n. 6, p. 345–353, 2010.

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.; VANBERGEN, A. J. **Summary for policymakers of the assessment report of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services on pollinators, pollination and food production**. Bonn, Germany, 2016.

RUAN, Z.-Y.; WANG, C.-H.; LIN, H.-J.; HUANG, C.-P.; CHEN, Y.-H.; YANG, E.-C.; TSENG, C.-L.; JIANG, J.-A. 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*, v. 11, n. 6, p. 478–482, 2017.

SANCHEZ, V.; GIL, S.; FLORES, J. M.; QUILES, F. J.; ORTIZ, M. A.; LUNA, J. J. 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*, Elsevier, v. 119, p. 209–216, 2015.

SILVA, T. L. C. da; MACÊDO, J. A. F. de; CASANOVA, M. A. Discovering frequent mobility patterns on moving object data. In: **Proceedings of the Third ACM SIGSPATIAL International Workshop on Mobile Geographic Information Systems**. New York, NY, USA: ACM, 2014. (MobiGIS '14), p. 60–67. ISBN 978-1-4503-3142-5. Disponível em: <<http://doi.acm.org/10.1145/2675316.2675325>>.

SOUTHWICK, E. E.; MORITZ, R. F. Social control of air ventilation in colonies of honey bees, *apis mellifera*. *Journal of insect physiology*, Elsevier, v. 33, n. 9, p. 623–626, 1987.

SPILIOPOULOU, M.; NTOUTSI, I.; THEODORIDIS, Y.; SCHULT, R. Monic: Modeling and monitoring cluster transitions. In: **Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining**. New York, NY, USA: ACM, 2006. (KDD '06), p. 706–711. ISBN 1-59593-339-5. Disponível em: <<http://doi.acm.org/10.1145/1150402.1150491>>.

STABENTHEINER, A.; KOVAC, H.; BRODSCHNEIDER, R. Honeybee colony thermoregulation–regulatory mechanisms and contribution of individuals in dependence on age, location and thermal stress. **PloS One**, Public Library of Science, v. 5, n. 1, p. e8967, 2010.

TASHAKKORI, R.; GHADIRI, A. Image processing for honey bee hive health monitoring. In: IEEE. **SoutheastCon 2015**. [S.l.], 2015. p. 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.* Landscape moderation of biodiversity patterns and processes-eight hypotheses. **Biological Reviews**, Wiley Online Library, v. 87, n. 3, p. 661–685, 2012.

TUKEY, J. W. **Exploratory Data Analysis**. [S.l.]: Addison-Wesley Publishing Company, 1977.

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.* Top 10 algorithms in data mining. **Knowledge and information systems**, Springer, v. 14, n. 1, p. 1–37, 2008.

XU, D.; TIAN, Y. A comprehensive survey of clustering algorithms. **Annals of Data Science**, v. 2, n. 2, p. 165–193, Jun 2015. ISSN 2198-5812. Disponível em: <<https://doi.org/10.1007/s40745-015-0040-1>>.

XU, R.; WUNSCH, D. Survey of clustering algorithms. **IEEE Transactions on Neural Networks**, Ieee, v. 16, n. 3, p. 645–678, 2005.

ZACEPINS, A.; BRUSBARDIS, V.; MEITALOVS, J.; STALIDZANS, E. Challenges in the development of precision beekeeping. **Biosystems Engineering**, Elsevier, v. 130, p. 60–71, 2015.

ZACEPINS, A.; KARASHA, T. Web based system for the bee colony remote monitoring. In: IEEE. **Application of Information and Communication Technologies (AICT), 2012 6th International Conference on**. [S.l.], 2012. p. 1–4.

ZACEPINS, A.; KVIESIS, A.; PECKA, A.; OSADCUKS, V. Development of internet of things concept for precision beekeeping. In: IEEE. **Carpathian Control Conference (ICCC), 2017 18th International**. [S.l.], 2017. p. 23–27.

ZACEPINS, A.; KVIESIS, A.; STALIDZANS, E.; LIEPNIECE, M.; MEITALOVS, J. Remote detection of the swarming of honey bee colonies by single-point temperature monitoring. **Biosystems Engineering**, Elsevier, v. 148, p. 76–80, 2016.

ZACEPINS, A.; STALIDZANS, E.; MEITALOVS, J. Application of information technologies in precision apiculture. In: **Proceedings of the 13th International Conference on Precision Agriculture (ICPA 2012)**. [S.l.: s.n.], 2012.

ZOGOVIĆ, N.; MLADENOVIC, M.; RASIC, S. From primitive to cyber-physical beekeeping. In: **Zdravkovic, M., Konjovic, Z., Trajanovic, M. (Eds.) ICIST 2017 Proceedings Vol.1**. [S.l.: s.n.], 2017. p. 38–43.