

# INVESTIGAÇÃO DE SURTOS DE DOENÇAS EM APIÁRIOS NO ESTADO DO RIO DE JANEIRO UTILIZANDO PESQUISA OPERACIONAL

Soline Maria Gonçalves Ikeda<sup>1</sup>, Marcelo Dib Cruz<sup>2</sup>, Wagner de Souza Tassinari<sup>3</sup>, Angel Ramón Sanchez Delgado<sup>4</sup>, Maria Cristina Lorenzon<sup>5</sup>

**RESUMO** – Até o momento não existe uma resposta confiável a respeito do misterioso desaparecimento das abelhas pelo mundo. Nos diversos fatores apontados como responsáveis pelo sumiço das abelhas, o excesso de parasitas que afetam esses insetos está entre os principais. Dentre os desenhos epidemiológicos utilizados neste contexto, destacam-se os estudos de agregados (ou *clusters*, na língua inglesa). De maneira geral, aglomerados espaciais de doenças podem ser atribuídos aos fatores demográficos, genéticos, ambientais ou, socioculturais superpostos geograficamente ao padrão de ocorrência observado. Na atualidade, o estudo de técnicas para a detecção de aglomerados espaciais no campo da epidemiologia, continua recebendo importantes contribuições científicas. O presente estudo objetiva identificar aglomerados espaciais de doenças em apiários no Estado do Rio de Janeiro utilizando pesquisa operacional. Isto será feito de duas maneiras diferentes: através da programação linear inteira e utilizando uma metaheurística. Na primeira, o número de grupos é conhecido e o que se deseja é encontrar os grupos. No segundo caso, nem os grupos e nem o número de grupos é conhecido. Os dados foram colhidos em 38 municípios do estado do Rio de Janeiro (Almeida et al., 2013).

Palavras chave: aglomerados espaciais, apicultura sustentável, meta-heurística, sanidade de apiários.

## INVESTIGATION OF DISEASE OUTBREAKS IN APIARIES IN THE STATE OF RIO DE JANEIRO USING OPERATIONAL RESEARCH

**ABSTRACT** - So far, there is no reliable answer regarding the mysterious disappearance of bees around the world. In the various factors considered responsible for the disappearance of the bees, the excess of parasites that affect these insects is among the main ones. Among the epidemiological designs used in this context, we highlight the studies of aggregates (or *clusters*, in english). In general, spatial clusters of diseases can be attributed to demographic, genetic, environmental, or socio-cultural factors superimposed geographically on the observed pattern of occurrence. At present, the study of techniques for the detection of spatial clusters in the field of epidemiology, continues to receive important scientific contributions. The present study aims to identify spatial clusters of diseases in apiaries in the State of Rio de Janeiro using operational research. This will be done in two different ways: through integer linear programming and using a metaheuristic. In the first, the number of groups is known and what is desired is to find the groups. In the second case, neither the groups nor the number of groups is known. Data were collected from 38 municipalities in the state of Rio de Janeiro (RJ-Brazil).

Keywords: metaheuristics, sanity of apiaries, space clusters, sustainable beekeeping.

<sup>1</sup> Mestre em Modelagem Matemática e Computacional (Bolsista Capes) - Universidade Federal Rural do Rio de Janeiro. Seropédica, RJ. solika\_sol@yahoo.com.br.

<sup>2</sup> Professor Doutor do Departamento de Sistemas de Informação - Universidade Federal Rural do Rio de Janeiro. Seropédica, RJ. madibcruz@gmail.com.

<sup>3</sup> Professor Doutor do Departamento de Matemática com área de concentração em Estatística - Universidade Federal Rural do Rio de Janeiro. Seropédica, RJ. wtassinari@gmail.com.

<sup>4</sup> Professor Doutor do Departamento de Matemática com área de concentração em Matemática Aplicada / Matemática Discreta e Combinatória - Universidade Federal Rural do Rio de Janeiro. Seropédica, RJ. asanchez@ufrj.br.

<sup>5</sup> Professora Doutora do Departamento de Zootecnia - Universidade Federal Rural do Rio de Janeiro. Seropédica, RJ. lorenzon\_ufrj@yahoo.com.br.



## INTRODUÇÃO

Existem hoje perto de 30 mil espécies de abelhas que em consonância com a alta abundância, são responsáveis por mais de 70% da produção agrícola mundial (Gallai et al., 2008; Neumann & Carreck, 2010), além da produção de mel e outros produtos apícolas. Nos últimos dez anos, colônias de abelhas começaram a desaparecer em diferentes países do mundo. Nos Estados Unidos, perto de 30% e 70% das abelhas sumiram nos estados da Califórnia e Texas respectivamente. Em algumas regiões da Europa, a diminuição chega a 80% (Steinhauer et al., 2014).

O desaparecimento das colmeias tem acarretado impacto direto sobre a produção de alimentos (Kevan et al., 2010). Na apicultura, as perdas anuais de colmeias acima de 20%, traduzem a relevância do diagnóstico e do controle de doenças, parasitoses e intoxicações. Sabendo-se que as doenças que matam abelhas adultas podem ser: “mal de outono”, nosemose, acariose, amebíase e varroatose, são necessários exames laboratoriais para sua confirmação. (Lorenzon et al., 2012). Como ação para um retorno rápido de um diagnóstico coletivo a nível populacional, o inquérito epidemiológico é uma importante ferramenta analítica para a investigação dos fatores de risco presentes na causalidade das doenças e na caracterização de sua distribuição, no espaço e no tempo no âmbito populacional.

Desde a década de 80, surgiu um renovado interesse nos estudos de padrões espaciais e temporais de doenças, conforme salienta a extensa literatura publicada em periódicos de diferentes áreas (Knox, 1991; Werneck & Struchiner, 1997). Dentre os desenhos epidemiológicos utilizados neste contexto, destacam-se os estudos de agregados ou *clusters*. De maneira geral, aglomerados espaciais de doenças podem ser atribuídos aos fatores demográficos, genéticos, ambientais ou, socioculturais superpostos geograficamente ao padrão de ocorrência observado. O estudo de técnicas para detecção de aglomerados espaciais no campo da epidemiologia tem recebido importantes contribuições científicas (Werneck & Struchiner, 1997).

O problema aqui tratado pode ser definido na área da pesquisa operacional como um problema de Clusterização ou Agrupamento (Cruz & Ochi, 2011; Bastos et al., 2014). Clusterização é o termo genérico para um processo que une objetos similares em um

mesmo grupo. Cada grupo é denominado de cluster. O número de clusters ( $k$ ) pode ser conhecido a priori ou não. Quando o número de clusters é conhecido (a priori), tratamos com um problema de  $k$ -clusterização ou, simplesmente problema de clusterização (PC). Caso contrário, ou seja, quando não se conhece o número de grupos, o problema é denominado problema de clusterização automática (PCA).

O objetivo deste trabalho é investigar a existência de aglomerados espaciais de doenças em apiários no estado do Rio de Janeiro utilizando pesquisa operacional, com o intuito de verificar quais são as regiões de maior prioridade para a intervenção na cadeia apícola. O problema é tratado de duas maneiras diferentes. Na primeira dividimos os casos de doenças em um número pré-definido de grupos (PC). Este processo pode ser utilizado, por exemplo, no caso da existência de um número pré-definido de equipes que investigarão os aglomerados espaciais das doenças, uma equipe para cada aglomerado. Neste caso, dado o número de grupos, o que se deseja é encontrar as configurações dos grupos.

Outra interpretação possível é fazer uma simulação dos diâmetros médios dos clusters, pois a proximidade geográfica dos apiários é um fator de grande importância na análise dos surtos de doenças. A diminuição ou aumento do diâmetro médio dos clusters é alcançado com o aumento ou diminuição do número de clusters. Na segunda maneira, o que se deseja é encontrar possíveis grupos, sem saber a priori, quantos e quais são (PCA), para investigar os grupos naturalmente formados, e com isso verificar possíveis surtos de doenças.

## MATERIAL E MÉTODOS

O presente estudo utilizou dados de 38 municípios do estado do Rio de Janeiro (22°54'23"S, 43°10'21"W). Este grupo de municípios contempla unidades de todas as mesorregiões do estado do Rio de Janeiro (Figura 1).

A base de dados utilizada neste estudo é proveniente do levantamento dos déficits da apicultura do estado do Rio de Janeiro, e foram coletados no período de 2009 a 2010, através da aplicação de questionários realizados por técnicos dos Núcleos de Defesa Agropecuária (NDA's), vinculados a Secretaria de Estado de Agricultura e Pecuária do Estado do Rio de Janeiro (SEAPEC) (Lorenzon, 2008). Ao todo, 401 apicultores participaram deste levantamento de informações, mas



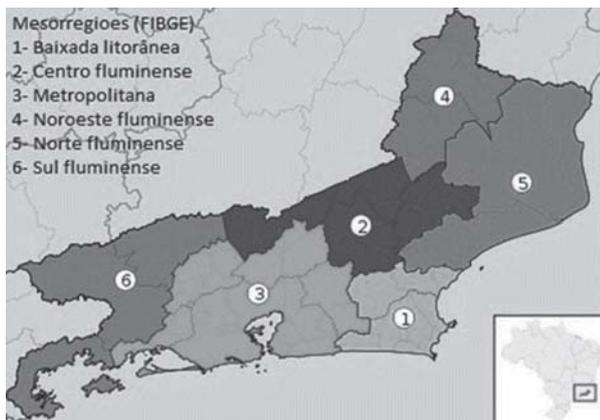


Figura 1 - Distribuição geográfica das mesorregiões do estado do Rio de Janeiro (RJ).

para o presente estudo foram selecionados aqueles que responderam sobre a ocorrência de doenças em seus apiários (269 apicultores). As coordenadas geográficas dos apiários foram identificadas via GPS, no momento da aplicação dos questionários e utilizadas em formato *Universal Transversa de Mercator (UTM)*.

A primeira modelagem do problema a considerar é o modelo linear inteiro 0-1 apresentado em Rao (1971). O objetivo deste modelo é minimizar a maior distância interna de todos os clusters, ou ainda, minimizar a maior diagonal de todos os clusters. Cada ponto só pode pertencer a um cluster e o número de clusters  $k$  é um dado de entrada. A grande dificuldade deste modelo é que o número de restrições cresce a medida que as quantidades de pontos e de clusters crescem. A continuação o modelo.

Definamos as variáveis  $X_{ij} = 1$  se o ponto  $i$  pertence ao cluster  $j$  e  $x_{ij} = 0$  caso contrário. Sejam  $d_{ij}$  a distância entre o ponto  $i$  e o ponto  $j$ ,  $D_l$  o diâmetro do cluster  $l$  e  $D_{max}$  o maior diâmetro de todos os clusters. Então o problema de programação 0-1 é:

$$\text{Minimizar } Z = D_{max} \quad (1)$$

$$\text{Sujeito a: } D_l \geq d_{ij} (x_{il} + x_{jl} - 1) \quad i, j = 1, 2, \dots, m; l = 1, 2, \dots, k \quad (2)$$

$$\sum_{l=1}^k x_{il} = 1 \quad (3)$$

$$D_{max} \geq D_l \quad l = 1, 2, \dots, k \quad (4)$$

$$x_{ij} \in \{0, 1\}$$

A restrição (2) garante que a diagonal de cada cluster é maior ou igual à distância máxima entre dois pontos dentro do cluster. A restrição (3) garante que cada ponto só pode ser alocado a um cluster e, finalmente, a restrição (4) garante que  $D_{max}$  seja maior que todas as outras diagonais de todos os clusters. Já no segundo modelo, o objetivo é selecionar  $k$  pontos médios do conjunto de pontos considerados  $X$ , um ponto médio para cada cluster, onde a soma das distâncias dos outros pontos do cluster a este ponto médio é mínima. Cada ponto só pode pertencer a um cluster e o número de clusters  $k$  é um dado de entrada. O ponto médio de um cluster é definido como sendo aquele ponto que pertence ao cluster e que está mais próximo a todos os outros pontos do cluster. Seja  $p$  o ponto está associado ao ponto médio e caso contrário, significa que o ponto  $p$  é o ponto médio. Agora o problema das  $k$ -medanas é:

$$\text{Minimizar } \sum_{i=1}^n \sum_{j=1}^n d_{ij} x_{ij} \quad (5)$$

$$\text{Sujeito a: } \sum_{j=1}^n x_{ij} = 1 \quad i = 1, 2, \dots, n \quad (6)$$

$$\sum_{j=1}^n x_{ij} = k \quad (7)$$

$$x_{ij} \leq x_{jj} \quad i, j = 1, 2, \dots, n \quad (8)$$

$$x_{ij} \in \{0, 1\} \quad i, j = 1, 2, \dots, n$$

A restrição (6) assegura que cada ponto será associado a somente um ponto médio. A restrição (7) assegura que serão escolhidos exatamente  $k$  pontos médios e a restrição (8) que cada ponto  $i$  está associado a um ponto  $j$  se e somente se o ponto  $j$  é um ponto médio. Seguidamente procuraremos identificar aglomerados espaciais de doenças em apiários no Estado do Rio de Janeiro, utilizando uma metaheurística. Lembremos que uma metaheurística é uma heurísticas genérica para a solução aproximada de problemas, principalmente problemas de otimização combinatória de elevada complexidade. Entre elas, podemos citar: Algoritmos Evolutivos, GRASP, ILS, Busca Tabu, VNS, entre outras.

A metaheurística ILS (Iterated Local Search), descrita em Glover & Kochenberger, (2003) e Dib (2010), consiste, essencialmente, na aplicação iterativa de um procedimento de busca local desde uma solução inicial, que é

previamente obtida a partir da utilização de um procedimento aleatório de construção ou considerando uma heurística de construção. O procedimento de busca local tem por finalidade melhorar a solução inicial e aquelas produzidas após perturbações das soluções ótimas locais. O Algoritmo ILS para o PCA (ILS\_PCA), aqui proposto, é uma heurística baseada na metaheurística ILS composta por quatro componentes principais: a fase de construção, as buscas locais, a perturbação e o critério de aceitação. A Fase de construção do ILS\_PCA é uma etapa inicial que tem por objetivos tentar reduzir a cardinalidade dos dados de entrada do problema, bem como gerar uma solução inicial de boa qualidade. A ideia é substituir grupos de objetos da base de dados cuja similaridade é considerada alta, por um único objeto (*cluster inicial*) que represente o grupo. A continuação o pseudocódigo do procedimento construtivo.

#### PROCEDIMENTO CONSTRUTIVO ( $X, u$ )

1. Para  $i = 1$  até  $n$  fazer
2.  $d_{min}(x_i) = \min\{\|x_i - x_j\| : i \neq j; i, j = 1, 2, \dots, n\}$
3. Fim Para
4.  $d_{medio} = \frac{1}{n} \sum_{i=1}^n d_{min}(x_i)$
5.  $r = u * d_{medio}$
6. Para  $i = 1$  até  $n$  fazer
7.  $N_i = circulo(x_i, r)$
8.  $T = T \cup N_i$
9. Fim Para
10. Ordenar  $T$  em ordem decrescente
11.  $i = 1$
12. Enquanto  $T \in \emptyset$  fazer
13.  $B_i = \text{Próximo}(N_i \in T)$
14.  $T = T - N_i$
15.  $i = i + 1$
16. Fim Enquanto
17. Retornar  $B = \{B_1, B_2, \dots, B_i\}$ , os  $t$  clusters parciais
18. Fim Construtivo

A redução do tamanho da entrada (pré-processamento) é realizada agrupando-se em um mesmo cluster os pontos pertencentes a uma região densa.

Inicialmente, nas linhas 1, 2 e 3, para cada ponto é definido a menor distância a outro ponto qualquer. Depois, na linha 4, é calculada a média destas distâncias, denominada  $d_{medio}$ . Então, cada ponto  $x_i \in X$  é considerado o centro de um círculo cujo valor do raio é  $r = u * d_{medio}$ , onde  $u$  é um parâmetro de entrada. Logo após, na linha 7 é calculado o conjunto de pontos contidos no círculo de centro  $x_i$  e raio  $r$ . Estes valores são colocados em uma lista que é ordenada em ordem decrescente. Os elementos de  $T$  são considerados os clusters parciais  $B = \{B_1, B_2, \dots, B_i\}$ . Para que os clusters não possuam elementos em comum, toda vez que um círculo é selecionado, todos os pontos do seu interior não podem mais entrar em outro círculo. Com este procedimento as regiões mais densas são selecionadas. Para que os clusters não possuam elementos em comum, toda vez que um círculo é selecionado, todos os seus pontos não podem mais entrar em outro círculo. Com este procedimento as regiões mais densas são selecionadas.

Considere os clusters iniciais gerados como  $B = \{B_1, B_2, \dots, B_m\}$  e seja  $v_i (i = 1, 2, \dots, m)$  o centroide de cada cluster  $B_i$ . Para representar uma solução é utilizada uma cadeia binária de posições. Por exemplo, se  $m = 7$ , então uma cadeia binária poderia ser  $\{0110010\}$ . Se o valor correspondente ao  $B_i$  na cadeia binária for igual a 1, isso significa que o cluster inicial  $B_i$  faz parte da solução como cluster pai. Se o valor correspondente ao  $B_i$  na cadeia binária for igual a 0,  $B_i$  é considerado um cluster filho. Os clusters filhos são unidos aos clusters pais utilizando o critério de menor distância entre os centroides. A cada união, o valor do centroide do cluster pai é recalculado. No final, todos os clusters filhos são unidos aos clusters pais para gerar uma solução completa. Portanto, nesta representação, a cada nova solução pode-se ter um número distinto de clusters pais, que não se alteram depois deste processo. Os clusters gerados após esse processo são denominados clusters finais  $C = \{C_1, C_2, \dots, C_k\}$ .

Seja um ponto pertencente ao cluster  $C_w \in C$  com  $|C_w| = M > 1$ . A distância entre os objetos  $x_i$  e  $x_j$  é definida por  $d_{ij}$ . A similaridade média de  $x_i$  em relação a todos os pontos de  $C_j \in C_w$  é dada por

$$a(x_i) = \frac{1}{M-1} \sum_{j=1}^m d_{ij} \quad (i = 1, 2, \dots, m) \text{ com } x_i \neq x_j$$

$\in C_w$ . Nos casos em que  $C_w$  possuir um único elemento, definimos  $a(x_i) = 0$ . Considere ainda, cada um dos clusters  $C_t \in C$  com  $t \neq w$  e  $|C_t| = T > 1$ . A similaridade média



do ponto  $x_i$  em relação a todos os pontos de  $C_i$  é

$$d(x_i, C_i) = \frac{1}{T} \sum_{j=1}^m d_{ij} \text{ para todo } x_j \in C_j, (j=1,2,\dots,m).$$

Seja  $b(x_i) = \text{Mínimo } d(x_i, C_j) \text{ (} i = 1,2,\dots,m \text{)}$ .

O valor de silhueta do ponto  $x_i \in X$  é dado por

$$s(x_i) = \frac{b(x_i) - a(x_i)}{\max(a(x_i), b(x_i))} \text{ e a função } F, \text{ denominada}$$

índice silhueta, é definida como:  $F(x_i) = \frac{1}{n} \sum_{i=1}^n s(x_i)$ .

Para avaliar a qualidade da solução encontrada, é utilizada proposta por Kaufman & Rousseeum (1990). Esta função admite valores dentro do intervalo  $[-1, 1]$  e não é necessário definir o valor de  $k$ , pois o valor mais adequado de  $k$  é atingido ao maximizar esta função. Para gerar uma solução, inicialmente é realizado um procedimento para construir uma lista restrita de candidatos (LRC). Nesta etapa, várias soluções são geradas, e o número de clusters pais das melhores soluções é armazenado na LRC. A continuação o pseudocódigo que gera uma solução inicial.

#### PROCEDIMENTO GERAÇÃO DE SOLUÇÃO INICIAL ( $B, MaxIter$ )

1. Seja  $B = \{B_1, B_2, \dots, B_m\}$  o conjunto de clusters iniciais
2. Para  $i = 1$  até  $MaxIter$  fazer
3. Para  $k = 2$  até  $m-1$  fazer
4.  $s^0 = Gerar\ Solução(k, B)$
5. Atualizar LRC ( $s^0$ )
6. Se  $s^0 > s^*$  então fazer  $s^* = s^0$
7. Fim Se
8. Fim Para
9. Fim Para
10. Retornar  $s^*$  e LRC

Na linha 1, os clusters iniciais são dados e são armazenados no conjunto  $B$ . Depois é feito um processo iterativo, que a cada momento, gera uma solução  $s^0$  com  $k$  clusters pais ( $s^0 = Gerar\ Solução(k, B)$ ),  $K = 2, \dots, m-1$ . Cada solução é avaliada e o número de clusters pais das melhores é armazenado na lista LRC (Atualizar LRC ( $s^0$ )). Isto está indicado entre as linhas 3 e 5. Depois, nas linhas 6 e 7, a melhor solução  $s^*$  até o momento é atualizada. Este processo iterativo é repetido por  $MaxIter$  vezes, como indica a linha 2. O retorno deste procedimento é a solução inicial  $s^0$

e a LRC. A lista LRC define o limite inferior  $l_i$  e o limite superior  $l_s$  de clusters pais que as soluções do algoritmo podem ter. Neste trabalho é utilizado um conjunto denominado  $CE$  (conjunto ELITE), que armazena a melhor solução de cada iteração do algoritmo. As soluções armazenadas são diferentes entre si. O  $CE$  é utilizado no final do algoritmo, para efetuar a busca local Inversão Individual, que tem como objetivo tentar melhorar as melhores soluções encontradas ao longo das iterações do algoritmo e armazenadas no  $CE$ .

A primeira busca local proposta, denominada “troca entre pares”, é uma busca intensiva que troca o status de dois elementos da solução com valores diferentes. Por exemplo, suponha que a solução corrente  $\{10111010\}$  é Primeiro é trocado o primeiro e o segundo elemento da solução. Então a nova solução é  $\{01111010\}$ . Se este novo solução melhorar o valor da função, então ele será aceito e será a nova solução corrente. A próxima troca é feita entre o primeiro e o terceiro elemento da solução corrente. A busca local termina quando todas as trocas entre dois elementos com valores distintos são testadas. O objetivo da busca local “troca entre pares” é investigar possíveis soluções diferentes com o mesmo número de clusters pais.

A ideia básica desta segunda busca local proposta, denominada “inversão individual”, é tentar melhorar a solução corrente analisando soluções próximas a ela. Para isso, essa busca permuta o valor de cada elemento da solução (1 por 0, ou 0 por 1), um por vez, e calcula o novo valor da função. Porém, o algoritmo só aceita a mudança, se o novo valor da função for maior que o valor anterior. Por exemplo, imagine que a solução corrente é  $\{0101101\}$ . Primeiramente, é trocado o primeiro elemento da solução. Então a nova solução é  $\{1101101\}$ . Se esta solução tem valor da função maior que o anterior, então ele será a nova solução, e assim é trocado o segundo elemento. A nova solução agora é  $\{1101101\}$ . Se este possuir valor da função maior que o anterior, então a mudança é aceita. Caso contrária, a solução é mantida. A busca acaba quando todos os elementos da solução são testados. A busca local “inversão individual” se justifica, pois encontrar o número ideal de clusters é um dos objetivos do problema, e a inclusão ou retirada de um cluster pai pode melhorar uma determinada solução.

O objetivo da perturbação é alterar o número de clusters pais da solução corrente. Dado uma solução

inicial, esta pode ter o número de clusters pais maior ou menor que o ideal. Portanto, pode ser necessário aumentar ou diminuir este número. A perturbação deve aumentar ou diminuir este valor em uma unidade, caso o número de clusters pais esteja bem próximo ao ideal, ou ainda, aumentar ou diminuir em  $n$  unidades, para procurar soluções afastadas da solução inicial.

A perturbação começa incrementando o número de clusters pais em uma unidade. Se a solução obtida após a perturbação e a busca local for melhor que a anterior, o processo é repetido, ou seja, o número de clusters pais desta solução é incrementado em uma unidade de novo. Caso a solução obtida não melhore a anterior, então a Perturbação aumenta o número de clusters pais em duas unidades. O processo continua enquanto a solução possuir o número de clusters pais dentro do limite superior  $l_s$  definido pela LRC. Depois, o número de clusters pais da solução é decrementado até atingir o limite inferior  $l_i$  da LRC. A perturbação não permite que um determinado número de clusters pais seja utilizado mais de uma vez, para efetuar a busca local “troca entre pares”. O critério de aceitação utilizado é que a solução gerada é aceita como a solução corrente, se esta melhorar a melhor solução encontrada até o momento. Em geral, o algoritmo ILS\_PCA gera uma solução inicial e aplica a busca local “troca entre pares”. Após cada perturbação, o algoritmo também aplica a busca local “troca entre pares”. No final, o algoritmo aplica a busca local “inversão individual”, no conjunto CE gerado.

Para verificar qual modelo linear inteiro Rao (1971) e k-Mediana deve ser utilizado na resolução do problema proposto, foram realizados alguns testes. Tais testes visam verificar o tempo gasto em cada modelo e o comportamento dos modelos à medida que o valor de  $k$  cresce. Para isso, foram avaliadas as seguintes instâncias conhecidas na literatura e referenciadas em Dib (2010): *Ruspini*, *200 DATA* e *300 p4c1*. A Tabela 1 descreve as instâncias. A Tabela 2 os resultados dos testes realizados com o modelo Rao e na Tabela 3 os obtidos com o modelo k-Mediana. Estas Tabelas mostram a variação do  $k$ , o valor ótimo para cada

Tabela 1 - Descrição das instâncias testadas

Nome	Nº pontos	Dimensão
Ruspini	75	R <sup>2</sup>
200data	200	R <sup>2</sup>
300p4c1	300	R <sup>2</sup>

$k$  e o tempo de execução. Os modelos foram executados utilizando o software XPRESS. Nos testes de contagem de tempo foram utilizados computadores com

Tabela 2 - Testes realizados com o modelo RAO

Nome	k	Ótimo	t (seg.)
<i>Ruspini</i>	7	32,984	14.163
	6	36,619	1.345
	5	40,249	658
	4	47,634	60
	3	88,588	18
	2	102,078	9
	<i>200 DATA</i>	4	9,257
3		10,308	344
2		16,669	243
<i>300 p4c1</i>	4	132.966	8.1356
	3	109,59	1.349
	2	365.322	153

Tabela 3 - Testes realizados com o modelo k-Mediana

Nome	k	Ótimo	t (s)
<i>Ruspini</i>	k = 2	2395.8	2s
	k = 3	1619.47	2s
	k = 4	861.47	2s
	k = 5	779.68	2s
	k = 6	714.65	2s
	k = 7	650.84	2s
	k = 8	600.00	2s
	k = 9	555.34	2s
	k = 10	512.81	2s
	<i>200 DATA</i>	k = 2	575.67
k = 3		337.70	4s
k = 4		280.21	4s
k = 5		240.42	3s
k = 6		217.48	3s
k = 7		204.00	3s
k = 8		191.24	3s
k = 9		179.61	3s
k = 10		168.46	3s
<i>300p4c1</i>		k = 2	16063.8
	k = 3	11001.2	18s
	k = 4	8160.05	17s
	k = 5	7344.33	16s
	k = 6	6563.90	16s
	k = 7	5988.01	16s
	k = 8	5579.9	15s
	k = 9	5286.93	15s
	k = 10	5019.77	15s
	k = 20	3412.00	15s
k = 30	2706.24	15s	
k = 40	2274.11	15s	
k = 50	1963.32	15s	



processadores Intel Xeon Quad core 3.00 Ghz com 16G de memória RAM.

No modelo de Rao, a partir do  $k$  acima dos valores definidos na Tabela 3, o software demorou mais de um dia e não convergiu para um resultado, e por isso o software foi interrompido. Portanto, os valores de  $k$  da tabela, são os únicos valores possíveis executáveis. Em relação ao modelo  $k$ -medianas, os valores de  $k$  poderiam ser aumentados, porém não seriam relevantes para a comparação. Pelos testes realizados, verificou-se que o modelo  $k$ -medianas é mais rápido e consegue executar instâncias de maior cardinalidade, sendo o escolhido na condução desta investigação.

## RESULTADOS E DISCUSSÃO

No Brasil, a sanidade dos apiários tem sido pouco investigada e os apicultores encontram dificuldades para atender seus chamados devido ao despreparo dos técnicos da defesa agropecuária em obter e encaminhar amostras, assim como a falta de laboratórios de patologia apícola. Desta maneira é incerto o indicador de perdas de colmeias das regiões brasileiras. Em outros países, um patamar acima de 20% é considerado preocupante (Neumann & Carreck, 2010; Steinhauer et al., 2014).

Ao se realizar o diagnóstico apícola do estado do Rio de Janeiro, verifica-se alta ocorrência de perdas por doenças de colmeias, dentro de um perfil de incertezas quanto ao seu diagnóstico. Observa-se que 33,83% dos apicultores apresentam relato positivo para doença. Esta cifra pode estar subestimada, ao se analisar as informações sobre as perdas com relatos negativos para doença. O grupo, que representa 49,36%, apresentaram respostas incertas sobre as fugas dos enxames (Almeida et al., 2013).

O modelo  $K$ -medianas foi executado e o resultado é mostrado na Figura 1. O modelo foi executado com vários valores de  $k$  (como verificado na Tabela 3), ou seja, os dados foram divididos em grupos variando de 2 a 9. No caso do algoritmo ILS\_PCA, os objetivos do algoritmo são encontrar o número natural de clusters e sua configuração. O ILS\_PCA foi executado e o número de clusters encontrado foi 6. Os clusters e a sua configuração são mostrados na Figura 2. Cada cor ilustrada nos mapas dos resultados indicam os diferentes tipos de clusters que o algoritmo encontrou de acordo com os padrões.

Analisando os dados gerados no procedimento de geração de solução inicial, outras informações podem ser obtidas. Primeiro, os dados podem ser plotados e associados ao mapa das mesorregiões do Rio de Janeiro, como mostra a Figura 3. Neste estudo, os aglomerados foram apresentados segundo um ranqueamento, ou seja, quanto maior a proximidade entre os apiários dentro do cluster estabelecido pelo modelo matemático, maior será sua prioridade por atenção. Tal definição foi estabelecida devido a possibilidade das doenças apícolas serem transmissíveis, já que a maioria dos apicultores desconhecem os sintomas e há casos de doenças de etiologia duvidosa como, da cria ensacada brasileira (Pacheco et al., 2009). Na primeira posição está o cluster 1 que engloba a região Sul Fluminense, onde a distância média entre os apiários (representados pelos pontos) estão mais próximos em comparação com os outros. Em segundo lugar está a região Metropolitana e Baixada que como pode ser observado na figura abaixo, está com distância média dos apiários entre 1699,7~1699,9 metros. Na terceira posição consta o cluster de número 3, referente à região Noroeste Fluminense com distância média entre os apiários de 1882,0~1882,2 metros. Em quarto lugar está o cluster de número quatro, como consta na figura abaixo, que se refere a região Centro Fluminense e Metropolitana com distância média entre os apiários de 2361,5~2361,7 metros. Na quinta posição encontra-se o cluster de número 5, que engloba a região Sul Fluminense e Metropolitana, com distância média de 2470,9~2471,1 metros. Por último está o cluster número 6 que engloba a região Centro Fluminense com 2543,6~2543,8 metros de distância média entre os apiários. Apesar do ranqueamento, todos os grupos são importantes para que medidas de prevenção e cuidados sejam tomadas pelos apicultores em seus apiários.

Há ainda fatores que podem favorecer a disseminação de doenças como, a migração de colmeias, a revenda de mel entre regiões e a proximidade entre apiários. Verifica-se que a localização dos aglomerados com indicativo de doença está generalizada, ou seja, ocorre em todas as mesorregiões do estado do Rio de Janeiro. Nas regiões de estudo existem 13 municípios sujeitos à alta prevalência de doenças (acima de 50%): Sapucaia, Mendes, Paracambi, Silva Jardim, Rio Bonito, São José do Vale do Rio Preto, Barra Mansa, Rio Claro, Engenheiro Paulo de Frontin, Casimiro de Abreu, Petrópolis, Araruama e Porciúncula. Esta informação corrobora com Almeida et al. (2013).

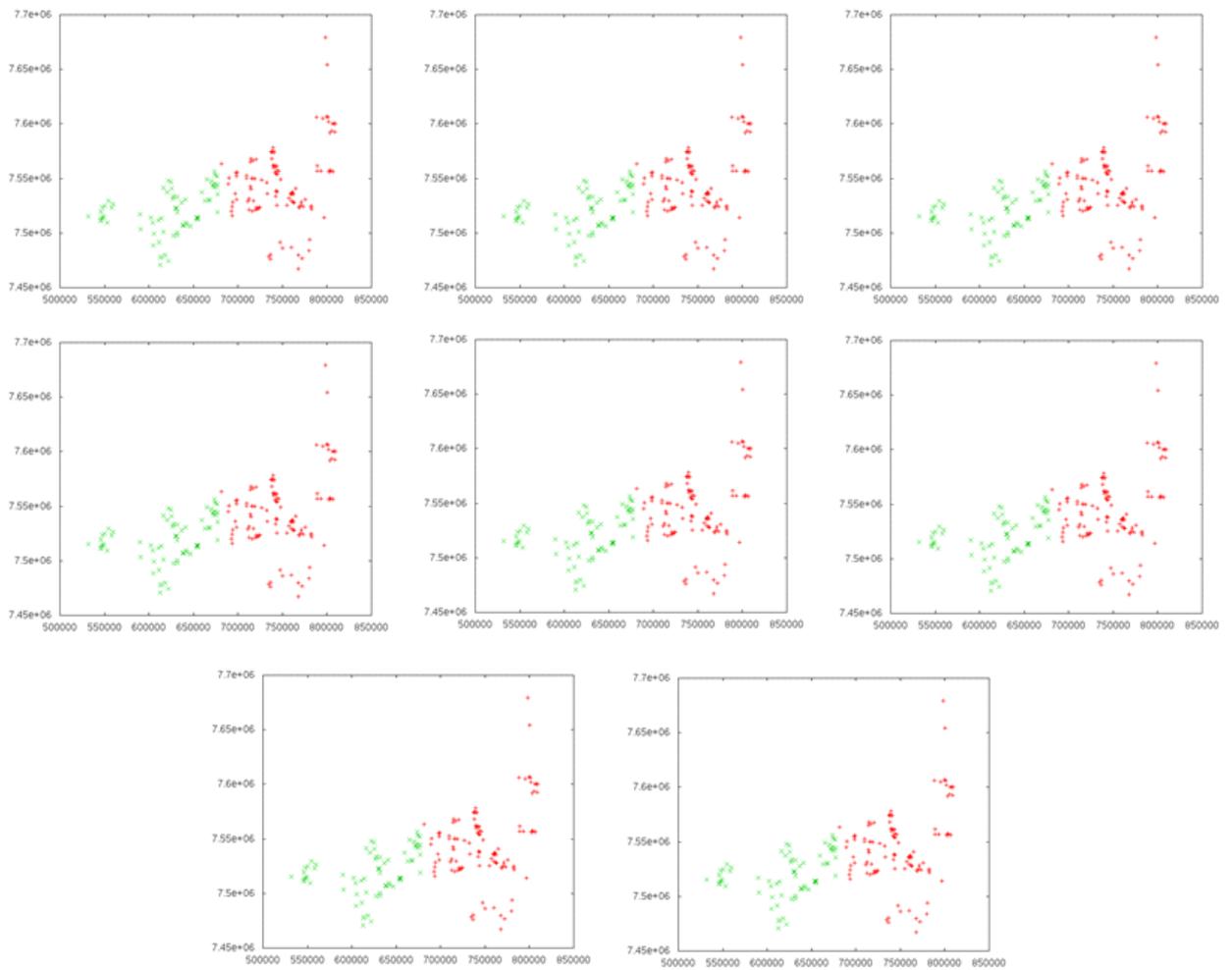


Figura 1 - Representação da distribuição espacial dos clusters, com k variando de 2 a 9.

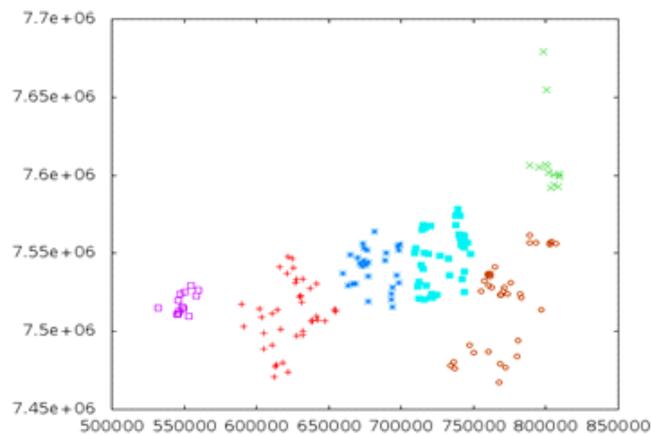


Figura 2 - Representação da distribuição espacial dos clusters encontrado pelo ILS\_PCA.



Na Tabela 4 apresentam-se as mesorregiões seguidas dos municípios acometidos pelas doenças em apiários por ordem de prioridade de atenção, sendo o ranqueamento definido segundo a

proximidade dos apiários, conforme expostos na Tabela 5.

Nesta análise verifica-se que as regiões que se deve priorizar atenção, são elas: Metropolitana, Baixada,

Tabela 5 - Identificação de alguns atributos dos aglomerados detectados com doenças em apiários no período de 2009 a 2010

Aglomerado	Produção anual mel/colmeia (kg)	Prevalência de apiários doentes (%)	Raio médio do aglomerado (km)	Distâncias médias dos apiários (km)
1	13.723	13,3	17,61	0,82
2	10.662	66,6	27,01	1,70
3	13.317	8,6	46,79	1,88
4	7.105	40,0	42,50	2,36
5	9.407	28,7	40,50	2,47
6	7.175	52,0	11,59	2,54

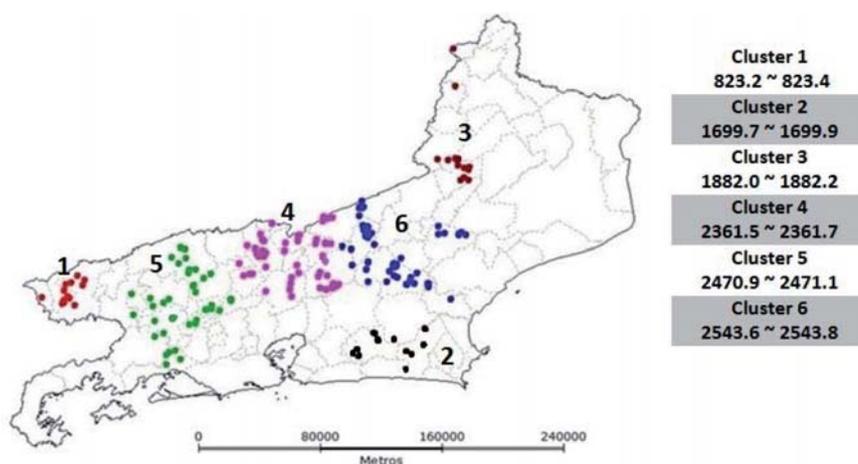


Figura 3 - Representação da distribuição espacial dos clusters, indicando os possíveis surtos de doenças em apiários no estado do Rio de Janeiro.

Tabela 4 - Identificação dos municípios do estado do Rio de Janeiro e suas respectivas mesorregiões ranqueadas pelos aglomerados surtos de doenças em apiários no período de 2009 a 2010

Aglomerado	Mesorregiões	Número e municípios mais afetados
1	Sul Fluminense	(2) Itatiaia e Resende
2	Metropolitana e Baixada	(4) Araruama, Rio Bonito, Silva Jardim e Tanguá
3	Noroeste Fluminense	(3) Aperibé, Itaocara e Porciúncula
4	Centro Fluminense e Metropolitana	(9) Com. Levy Gasparian, Paraíba do Sul, Paty dos Alferes, Petrópolis, São José do Vale do Rio Preto, Sapucaia, Teresópolis, Três Rios e Vassouras
5	Sul Fluminense e Metropolitana	(12) Engº Paulo de Frontin, Miguel Pereira, Paracambi, Pinheiral, Pirai, Rio Claro, Barra do Pirai, Barra Mansa, Valença, Vassouras, Volta Redonda e Itaguaí
6	Centro Fluminense	(7) Sapucaia, Casimiro de Abreu, Nova Friburgo, São José do Vale do Rio Preto, Sumidouro, Teresópolis e Trajano de Moraes.

Sul e Noroeste Fluminense. Tais regiões somam mais que 60% da produção total de mel no estado do Rio de Janeiro. Embora a área do surto não seja tão expressiva, a distância média entre os apiários com doenças é de 1,46 km. A prevalência em si em cada região, não foi tratada como o problema majoritário, mas sim a proximidade geográfica dos apiários que é um fator de grande importância e que é levado em consideração em estudos epidemiológicos (Almeida et al., 2013), haja à vista, que as abelhas migram por abandono e por enxameação (divisão natural) e seu raio de forrageamento médio de 5 km, pode alcançar perto de 25 km, em situação de escassez de alimento.

### CONCLUSÕES

1.O trabalho apresentou modelos exatos de Programação Linear Inteira e uma Metaheurística para a investigação da existência de aglomerados espaciais em apiários, com o intuito de verificar quais são as regiões com maior surto de doenças apícolas.

2. Os resultados obtidos neste estudo podem auxiliar no estabelecimento de estratégias de prevenção e controle fundamentadas na análise dos fatores associados à sanidade apícola no estado do Rio de Janeiro e na prevalência de doenças nos municípios estudados.

3. As abelhas, como todos os organismos vivos, são susceptíveis a várias doenças, parasitas e predadores, cuja ação pode ter um efeito prejudicial no seu normal desenvolvimento, e mais importante na sua produtividade. Os apicultores, como criadores de gado e na procura de uma apicultura sustentável têm a responsabilidade de promover o desenvolvimento de colônias fortes e saudáveis, em especial nas épocas de maior produção.

4. Com o apoio dos algoritmos utilizados foi possível identificar áreas com maior surto de doenças em apiários no Estado do Rio de Janeiro. Dessa forma, com base nas patologias determinadas, os apicultores podem tomar decisões sobre quais medidas preventivas aplicar.

### LITERATURACITADA

ALLARDICE, P; BONE, K; HUTCHISON, F. **Segredos e virtudes das plantas medicinais: um guia com centenas de plantas nativas e exóticas e seus poderes curativos.** Reader's Digest. 1.ed. 1999.

ALMEIDA, C.T.; TASSINARI, W.S.; LORENZON M.C.A. Identificação de fatores associados à ocorrência de doenças de abelhas africanizadas (*Apis mellifera* L.) em apiários do estado do Rio de Janeiro. **Revista Brasileira de Medicina Veterinária**, v.35, p. 33-40, 2013.

BASTOS, L.; OCHI, L. S.; PROTTI, F.; SUBRAMANIAN, A.; MARTINS, I.; PINHEIRO, R. Efficient Algorithms for Cluster Editing. **Journal of Combinatorial Optimization**, v.28, p.1-26, 2014.

BAZARAA, M. S.; JARVIS, J. J.; SHERALI, H. D. **Linear programming and network flows.** Wiley, 1990. 638p.

CRUZ, M. D.; OCHI, L. S. Um Algoritmo Evolutivo com Memória Adaptativa para o Problema de Clusterização Automática. **Learning and Nonlinear Models**, v.8, n.4, p.227-239, 2011.

GLOVER, F.; KOCHENBERGER, G.A. **Handbook of metaheuristics.** Kluwer Academic Publishers, 2003. 641p.

GALLAI, N.; SALLES, J. M.; SETTELE, J.; VAISSIÈRE, B. Economic evaluation of the vulnerability of world agriculture confronted with pollinator decline. **Ecological Economics**, v.68, n.3, p.810-821, 2008.

KAUFMAN, L.; ROUSSEUM, P. J. **Finding groups in data: an introduction to cluster analysis.** A Wiley-Intercience publication, 1990.

KEVAN, P. G.; PHILLIPS, T. P. The economic impacts of pollinator declines: an approach to assessing the consequences. **Conservation Ecology**, v.5, n.1, p.1-17, 2010.

KNOX, E. G. **Spatial and temporal studies in epidemiology.** Oxford textbook of public health: methods of public health, v.2, p. 95-105, 1991.

LORENZON, M.C; TASSINARI, W. S.; KOSHIYAMA, A. S.; ALMEIDA, C. T. **Indicadores e desafios da apicultura fluminense - um retrato brasileiro.** 1. Ed. Espírito Santo: ABOVE, v. 1000, 2012. 232p.



NEUMANN, P; CARRECK, N. L. Honey bee colony losses. **Journal of Apicultural Research**, v.49, n.1, p.1-6, 2010.

PACHECO, M. R.; BARTH, O. M.; LORENZON, M. C. Tipos polínicos encontrados em colônias de abelhas africanizadas sujeitas à doença criam ensacadas brasileiras. **Ciência Rural**, v.39, n.7, p. 2141-214, 2009.

RAO, M. R. Cluster analysis and mathematical programming. **Journal of the American Statistical Association**, v.66, p.622-626, 1971.

STEINHAUER, N. A. et al. A national survey of managed honeybee 2012-2013 annual colony losses in the USA: results from the Bee Informed Partnership. **Journal of Apicultural Research**, v.53, n.1, p.1-18, 2014.

VINOD, H. D. Integer programming and theory of groups. **Journal of the American Statistical Association**, v.64, n.326, p.506-519, 1969.

WERNECK, G. L.; STRUCHINER, C. J. Estudos de agregados de doença no espaço-tempo: conceitos, técnicas e desafios. **Cadernos de Saúde Pública**, v.13, n.4, p.611-624, 1997..

Recebido para publicação em 23/6/2018 e aprovado em 9/12/2018.

