

## MÉTODOS DE AGRUPAMENTO FUZZY APLICADOS ÀS SISTEMAS DE GALPÃO DE COMPOSTAGEM

VANIA C. MOTA<sup>1</sup>, ALESSANDRO T. CAMPOS<sup>2</sup>, FLÁVIO A. DAMASCENO<sup>3</sup>, EDUARDO A. SOARES<sup>4</sup>, DANIEL F. LEITE<sup>5</sup>

<sup>1</sup>Matemática, Mestre em Estatística e Experimentação Agropecuária, Doutoranda em Engenharia Agrícola, Depto. de Engenharia, Universidade Federal de Lavras (UFLA). Lavras, MG. Fone: (35)984622727. E-mail: vaniamota33@gmail.com

<sup>2</sup>Eng. Agrícola, Professor Associado, Depto. de Engenharia, Universidade Federal de Lavras (UFLA). Lavras, MG.

<sup>3</sup>Eng. Agrícola, Professor Adjunto, Depto. de Engenharia, Universidade Federal de Lavras (UFLA). Lavras, MG.

<sup>4</sup>Eng. de Sistema de Informação Mestrando em Engenharia de Sistemas e Automação, Depto. de Engenharia, Universidade Federal de Lavras (UFLA). Lavras, MG.

<sup>5</sup>Engenheiro de Controle e Automação, Professor Adjunto, Depto. de Engenharia, Universidade Federal de Lavras (UFLA). Lavras, MG.

**RESUMO:** Este artigo trata da aplicação de métodos de agrupamento *fuzzy* na avaliação de confinamentos *Compost Barn* (CB). Agrupamentos *fuzzy* são desenvolvidos para auxiliar a tomada de decisão no que diz respeito ao controle de variáveis, como a umidade, temperatura e aeração da cama de compostagem. A ideia é promover o bem-estar de bovinos leiteiros e melhorar os índices de produtividade. Os dados foram obtidos de fazendas no estado do Kentucky. Foram utilizados para a análise os métodos de agrupamento *K-Means* (KM), *Fuzzy C-Means* (FCM), *Gustafson-Kessel* (GK) e *Gath-Geva* (GG). A eficiência dos métodos foi determinada pelos índices de validação *Xie-Beni*, *Partition Coefficient*, *Partition Index* e *Dunn Index*. Seis classes foram identificadas nos dados. As classes se referem ao grau de estabilidade, maturidade e atividade do composto, e a concentração bacteriana. O agrupamento GG mostrou-se ser o mais adequado de acordo com a maioria dos índices, seguido pelo GK. A principal razão para os melhores resultados é o uso de medidas de distância baseadas na máxima verossimilhança e métrica de Mahalanobis. Os resultados de modelagem *fuzzy* e as informações linguísticas mostraram-se úteis para auxiliar na tomada de decisão em fazendas que adotam CBP como sistema de confinamento para bovinos leiteiros.

**PALAVRAS-CHAVE:** Agrupamento *Fuzzy*, Bovinos Leiteiros, Galpão de Cama de Compostagem.

### FUZZY CLUSTERING METHODS APPLIED TO COMPOST BARN SYSTEMS

**ABSTRACT:** This paper concerns the application of fuzzy clustering methods in the evaluation of compost bedded pack (CBP) barns. Fuzzy classifiers are developed to assist decision making regarding the control of variables such as bed moisture, temperature and aeration. The idea is to promote dairy cattle welfare and therefore improve productivity indices. The data was obtained in the state of Kentucky. Clustering methods, namely, K-Means, Fuzzy C-Means, Gustafson-Kessel (GK), and Gath-Geva (GG), are considered for data analysis. The efficiency of the methods is determined by validation indices such as the Xie-Beni criterion, Partition Coefficient, and Partition and Dunn indices. Six classes were identified in the data. The classes refer to degrees of stability, maturity and activeness of the compost, and to the ideal bacterial concentration of the bed. The GG method showed to be the most accurate method according to the majority of the validation indices, followed by the GK method. The main reason for the best results is the use of maximum-likelihood-based and Mahalanobis distance measures instead of Euclidean measure. Fuzzy modeling results and linguistic information have shown to be useful to help decision making in farms that adopt CBP barns as containment systems for dairy cattle.

**KEYWORDS:** Fuzzy Clustering, Dairy Cattle, Compost Bedded Pack Barn.

## INTRODUÇÃO

*Compost Barn* (CB) é um sistema de confinamento onde os animais têm mais liberdade de movimento dentro da instalação e são capazes de se deitar de forma mais natural e confortável. A ideia é melhorar a saúde geral e a longevidade do gado leiteiro e, portanto, melhorar os índices de produtividade (BLACK et al. 2013). O CB é semelhante ao sistema de confinamento *Loose Housing*, exceto pelo fato da cama ser revolvida diariamente. Os materiais comumente utilizados para compor a cama são serragem, aparas de madeira e palhada de milho. No CB, a cama deve ser arejada por meio de seu revolvimento no mínimo duas vezes ao dia para secar a superfície e estimular a atividade microbiana.

De acordo com LESO et al.(2013), o material de cama deve ser cuidadosamente manejado para fornecer uma superfície suave e seca em que as vacas podem caminhar. Esse material quando misturados com esterco bovino produz um adubo com uma quantidade adequada de matéria orgânica para melhorar a fertilidade do solo (GALAMA, 2011).

O sucesso do sistema CB depende principalmente do manejo e revolvimento da cama. Quando esta é manuseada e tratada corretamente a compostagem aumentará a temperatura da cama para reduzir populações microbianas patogênicas. Esse processo também diminui a umidade da cama e aumenta a superfície seca (BARBERG et al. 2007).

Um material de cama seco e macio também melhora a limpeza do animal e, portanto, reduz os problemas associados, tais como a diminuição da claudicação animal (DAMASCENO, 2012). O objetivo do desenvolvimento de um sistema de monitoramento e classificação baseado em dados do sistema *Compost Barn* é analisar variáveis relacionadas ao confinamento de bovinos de raça leiteira, a fim de inferir sobre um índice específico denominado Índice do Processo de Compostagem (CPI).

Existem diferentes algoritmos de agrupamento de dados com diferentes propriedades na literatura de aprendizagem de máquina (HOPNER et al., 1999). Eles são úteis para a construção de um sistema de monitoramento e classificação. Muitas vezes, esses algoritmos fornecem soluções diferentes. Não existe um método de agrupamento “melhor” para todos os possíveis conjuntos de dados e de classes de problemas. Um ponto chave é selecionar um algoritmo de agrupamento apropriado para um conjunto de dados (WANG & PENG, 2009). Os métodos de agrupamento considerados neste trabalho para construir um sistema de classificação são baseados em clusterização fuzzy (ZADEH, 1965; HOPNER et al., 1999).

Estudos sobre agrupamento fuzzy são motivados pela necessidade de resultados interpretáveis em problemas complexos de classificação e agrupamento de dados (GIUSTI & LIBELLI, 2010). O objetivo desse artigo é desenvolver um modelo classificador fuzzy do Índice do Processo de Compostagem - CPI, na avaliação de confinamentos *Compost Barn* (CB). O modelo é útil para determinar funções de fronteira capazes de realizar mapeamento entre amostras de dados e classes. As classes referem-se à descrições do estado de compostagem da cama. O modelo de classificação resultante é útil para auxiliar na tomada de decisões em fazendas que adotam o sistema CB. As decisões visam proporcionar conforto e bem-estar animal.

## MATERIAL E MÉTODOS

### ➤ Origem dos dados

Os dados foram coletados de diferentes propriedades no estado de Kentucky, Estados Unidos. A localização das propriedades pode ser vista na Fig. 1. Foram consideradas quarenta e duas (42) fazendas para análises. Os dados foram coletados de outubro de 2010 a março de 2011.

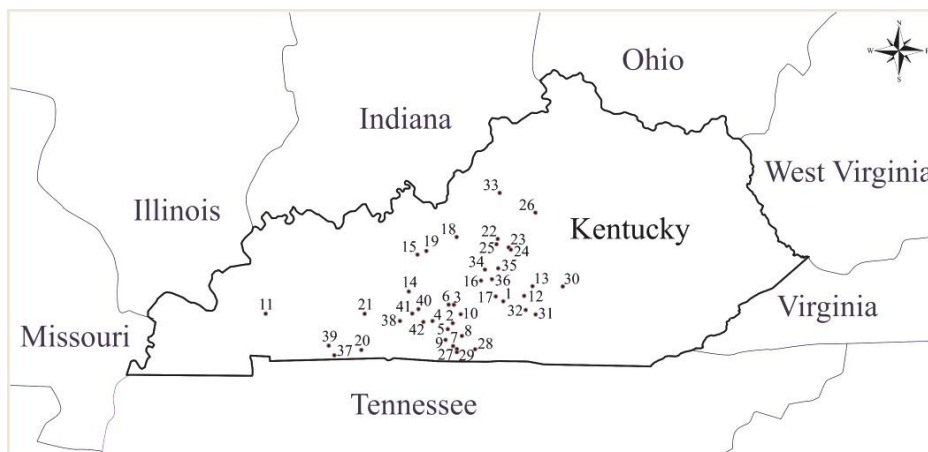


FIGURA 1. Localização das 42 propriedades CB no Estado do Kentucky

Considere  $x$  o vetor de entrada de um classificador, e  $y$  sua saída correspondente. As variáveis de entrada são as seguintes:  $x_{(1)}$  é a umidade da cama (%);  $x_{(2)}$  é a temperatura da cama ( $^{\circ}$  C);  $x_{(3)}$  é o carbono total (%);  $x_{(4)}$  é o nitrogênio total (%);  $x_{(5)}$  é a relação Carbono/Nitrogênio (%);  $x_{(6)}$  é o score de limpeza;  $x_{(7)}$  é o score de claudicação; e  $x_{(8)}$  refere-se à densidade animal (vaca/m<sup>2</sup>). A variável de saída é o índice de processo de compostagem. De acordo com especialistas, seis classes do Índice do Processo de Compostagem, CPI, são possíveis. Os graus relacionados às classes variam de 0 a 100; uma descrição de seus significados é mostrada na Tabela 1. Naturalmente, um número menor de classes leva à descrições mais rudimentares.

TABELA 1. Descrições e graus de compostagem

Classes	Grau	Descrição da eficiência
1	15	Matéria prima muito instável, concentração bacteriana leve
2	30	Composto da cama muito ativo, concentração bacteriana moderada
3	45	Composto de cama imaturo, ainda ativo
4	60	Ainda maturando, a cama continua em possível decomposição
5	75	Maturada, compostagem da cama finalizada
6	90	Bem-maturada, concentração leve de bactérias

As medições foram coletadas nos celeiros CB nos pontos localizados dentro da área da cama, ver esquema na Fig. 2A. A temperatura do ar foi medida uma vez por dia, dentro e fora dos galpões, usando um medidor de temperatura (precisão de 1 $^{\circ}$ C; Kestrel<sup>®</sup>, modelo 4000, Sylvan Lake, MI, EUA). Além disso, alturas diferentes em relação ao solo, medindo 0,05m e 1,2m, foram consideradas. Os pontos de amostragem (A1 a A9) representam o centro de nove espaços de grade dentro da cama (serragem, aparas de madeira etc.) que misturado com os dejetos bovinos se transformam em um composto de esterco. Os pontos foram estabelecidos com base no espaçamento de pilares ao longo da área da cama e geralmente ao longo do centro da cama.

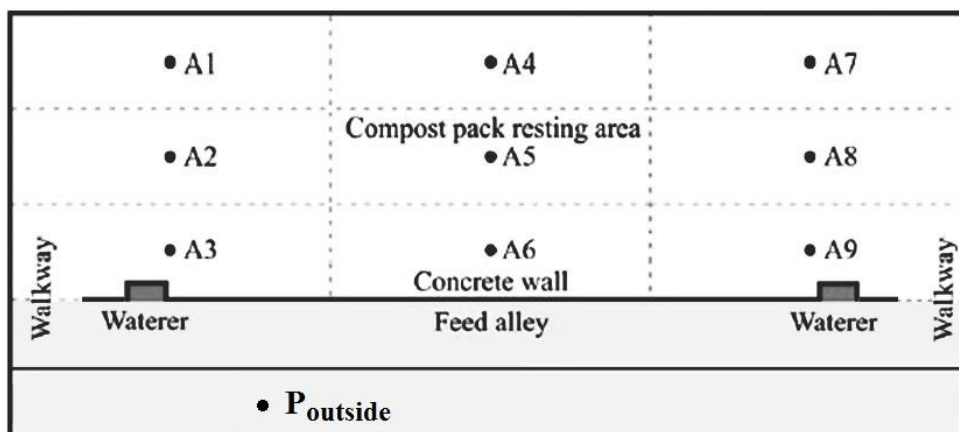


FIGURA 2. Nove áreas (A1 - A9) e um ponto fora da cama (P<sub>outside</sub>) considerado para medir a temperatura

Amostras da cama foram coletadas de uma mistura de composto a 0,1m nos nove locais diferentes dentro dos galpões, usando sacos de plástico (Ziploc®, Double Zipper, Racine, WI, US). Amostras foram coletadas usando uma enxada de ferro com uma alça de madeira e um trado de solo. As amostras foram armazenadas em gelo após a sua aquisição e refrigerado a 1°C até chegar ao laboratório. O estudo da umidade da cama ao longo do galpão de compostagem foram medidos a 105°C em estufa durante aproximadamente 24 horas. As análises dos nutrientes do material de cama foram realizadas pela Universidade de Kentucky - Serviços Regulatórios (Lexington). A relação C:N foi calculada para todos os galpões CB.

Os scores de locomoção e higiene foram coletadas para cada fazenda usando o galpão CB como a instalação de alojamento principal. Um mínimo de 50 vacas foi pontuado em cada fazenda. Neste caso, todas as vacas foram marcadas. As vacas foram selecionadas aleatoriamente usando o último dígito do número da marca auricular (por exemplo, número par, número ímpar ou múltiplo de 3) e obteve-se pontuações tanto para a locomoção quanto para a higiene do ponto de vista de um mesmo observador em cada visita às fazendas. A higiene foi avaliada utilizando um valor entre 1 e 4 (COOK & REINEMANN, 2008), onde 1 significa 'limpo' e 4 significa 'sujo', marcando o úbere e as patas traseiras

A claudicação foi avaliada observando-se as vacas caminhando em concreto e usando o sistema de pontuação de locomoção (SPRECHERET et al. 1997), onde 1 significa 'normal', 2 significa 'ligeiramente coxo', 3 significa 'moderadamente coxo', 4 significa 'coxo' e 5 refere-se a 'Severamente coxo'. A observação da locomoção foi realizada encorajando o animal a se mover e avaliar as pernas e as costas. As vacas com valor maior ou igual 3 para locomoção foram classificadas como clinicamente coxo.

#### ➤ Modelagem Fuzzy

Inicialmente, algoritmos de agrupamento foram utilizados para encontrar padrões nos dados. Os métodos de agrupamento, *K-Means* (KM), *Fuzzy C-Means* (FCM), *Gustafson-Kessel* (GK) e *Gath-Geva* (GG) foram aplicados. Todos os algoritmos requerem que o número de clusters (grupos) sejam definidos previamente. O número de clusters no conjunto de dados é um parâmetro crucial para a eficácia do classificador. Portanto, foram escolhidas diferentes quantidades de clusters e diferentes modelos foram obtidos.

A análise de cluster refere-se a um amplo espectro de algoritmos que tentam subdividir um conjunto de dados  $X$  em subconjuntos disjuntos e não-vazios. Os subconjuntos reproduzem  $X$  por meio de união (BEZDEK, 1981). Um cluster é normalmente formado por amostras de dados semelhantes, ao redor de um centro, chamado protótipo ou centróide (CHATTOPADHYAY, 2012).

Métodos de agrupamento fuzzy são originados da análise de dados e de reconhecimento de padrões (BABUSKA & VERBRUGEM, 2003). O conceito fuzzy é relevante por ser usado para representar o grau que uma dada amostra de dados é semelhante a outros elementos de um grupo. O grau de semelhança entre uma amostra de dados e um centróide pode ser calculado usando uma medida de distância ou de similaridade adequada.

As principais medidas de similaridade utilizadas em algoritmos de agrupamento são baseadas nas distâncias de *Manhattan* (norma-1), Euclidiana (norma-2), *Chebyshev* (norma-infinito) e distâncias de Mahalanobis. Em particular, a distância de Mahalanobis contorna algumas das limitações de  $p$ -normas para espaços vetoriais  $n$ -dimensionais (com  $n$  finito) já que leva automaticamente em conta a escala dos eixos coordenados e correlações entre as variáveis.

No entanto, as matrizes de covariância requerem cálculos adicionais, isto é, o tempo computacional cresce quadraticamente com a quantidade de variáveis. Na prática, quanto maior o valor de  $n$ , maior o número de desvios padrões que uma amostra de dados está afastada do centro de um cluster, e menor sua chance de ser um elemento do cluster. Além disso, para  $p$ -normas,

$$\|x\|_{p+a} \leq \|x\|_p \text{ para todo } x \in \mathbb{R}^n, p \geq 1, a \geq 0, \quad (1)$$

isto é, a  $p$ -norma  $\|x\|_p$  de qualquer vetor  $x$  não cresce com  $p$ ; qualquer outra norma é limitada inferiormente pela norma-1. Os algoritmos KM, FCM, GK e GG, foram utilizados nas análises, e seguem procedimentos semelhantes: (i) estabelecem parâmetros iniciais e centros de agrupamentos (clusters); (ii) calculam as distâncias entre amostras e centros de clusters; e (iii) atualizam uma matriz de partição, centros de clusters e parâmetros associados. Estes últimos passos são repetidos até que um critério de parada seja satisfeito. Como esses algoritmos são amplamente discutidos na literatura, as formulações detalhadas podem ser vistas por exemplo em (BABUSKA, 1998; BALASKO et al., 2005).

Após a aplicação dos algoritmos, uma medida de validação interna é usualmente usada para identificar as melhores partições obtidas pelos métodos de agrupamento (WANG et al. 2009). Diferentes medidas de validação foram propostas na literatura. Algumas das medidas mais usuais, que foram consideradas para análise neste artigo, são descritas a seguir.

a) Critério *Xie-Beni* (XB) - o número ideal de clusters pode ser determinado pela procura do valor mínimo para o critério XB (XIE&BENI, 1991):

$$XB = \frac{\sum_{i=1}^c \sum_{k=1}^N (\mu_{ki})^m \|x_k - v_i\|_2}{N \min_{i,k} \|x_k - v_i\|_2}, \quad (2)$$

Pequenos valores de XB são esperados para uma boa compactação dos clusters. Os valores de XB dependem da constante  $m$ , que é definida como sendo 2 geralmente.

b) *Partition Coeficient* (PC) - mede a quantidade de sobreposição entre os clusters (BEZDEK, 1981). Sua formula é dada por:

$$PC = \frac{1}{N} \sum_{i=1}^c \sum_{k=1}^N (\mu_{ki})^2 \quad (3)$$

Os valores de PC variam em  $\left[\frac{1}{c}; 1\right]$ . Quanto mais próximo da unidade é o coeficiente, o agrupamento é dito mais *crisp*. Um valor próximo de  $\frac{1}{c}$  indica que não há tendência de

agrupamento, ou o algoritmo não revelou isto. O número ideal de clusters fornece o maior valor de PC. Normalmente, o PC não é aplicável a algoritmos não-*fuzzy*.

c) *Partition Index*(SC) - consiste na relação da soma de capacidade e separação de clusters (BENSAID et al., 1996). O índice é dado por:

$$SC = \sum_{i=1}^c \frac{\sum_{k=1}^N (\mu_{ki})^m \|x_k - v_i\|_2}{N_i \sum_{j=1}^c \|x_j - v_i\|_2} \quad (4)$$

onde,  $N_i$  é o número de amostras no  $i$ -ésimo cluster. SC é útil quando se comparam partições com um número similar de clusters. Um valor menor de SC sugere uma melhor partição.

d) *Dunn Index*(DI) - é definido por:

$$DI = \frac{d_{\min}}{d_{\max}} \quad (5)$$

em que,  $d_{\min}$  é a distância mínima entre amostras de diferentes clusters, e  $d_{\max}$  é a maior distância dentro do cluster. O número ideal de clusters é aquele que fornece o maior valor de DI (DUNN, 1974). Um DI maior indica a existência de clusters compactos e bem-separados.

Os índices de validação XB, PC, SC e DI indicam a melhor partição alcançada. Finalmente, um grau de CPI e uma descrição linguística estão associados com o resultado de clusterização.

## RESULTADOS E DISCUSSÃO

A teoria dos sistemas fuzzy fornece ferramentas que podem ser usadas para transformar informações matematicamente imprecisas ou incertas, coletadas em situações do nosso cotidiano em dados passíveis de tratamento computacional. O objetivo é tratar fenômenos naturais e/ou situações reais, a fim de obter modelos computacionais que estão mais próximos do que é real ou verdadeiro (ZADEH, 1965). Em sistemas fuzzy, uma função característica generalizada está associada a um universo de discurso que possui um valor num intervalo de 0 e 1. Este valor indica o grau de pertinência ou associação de um elemento a um conjunto (ZADEH, 1965).

Normalmente, um conjunto de dados tem algum tipo de organização que pode ser explicado por meio de um conjunto de partições (clusters). O estudo destas partições é uma tarefa de análise de dados, baseada em medida de similaridade e um algoritmo de aprendizagem não supervisionado.

Os índices XB, PC, SC e DI são usados para avaliar a qualidade das partições finais fornecidas pelos algoritmos KM, FCM, GK e GG. Valores mínimos de XB e SC e valores máximos de PC e DI sugerem a melhor partição. O critério de parada de todos os algoritmos,  $\epsilon$ , foi definido como 0,01. Os dados foram normalizados entre 0 e 1 para considerar as variáveis em proporção. Além disso,  $c$  amostras aleatórias do conjunto de dados são consideradas como médias iniciais para todos os algoritmos. A Tabela 2 mostra os resultados.

Na Tabela 2, os valores em negrito indicam o número ideal de clusters de acordo com um índice de validação para um classificador. Por exemplo, os índices XB e DI aconselham que 6 é o número ideal de clusters KM, enquanto o índice SC recomenda 7 clusters KM. Em particular, o índice PC não é aplicável (n/a) à KM, uma vez que depende de valores de pertinência. Note-se que, em geral (em 13 dos 15 casos), 6 é o número ideal de partições para os dados de galpões CB, independente do algoritmo utilizado. O tempo é dado em segundos.

Os valores sublinhados na Tabela 2 indicam o melhor classificador e sua estrutura de acordo com um índice. De acordo com XB e PC, o classificador GG com 6 partições forneceu os melhores resultados; SC e DI apontam que GK e FCM, respectivamente, são os melhores classificadores usando 6 partições. KM é um particionamento não- fuzzy. Embora simples, em

geral, mostrou ser o menos preciso entre os métodos avaliados. FCM normalmente apresentou o menor tempo de treinamento. Embora o número de iterações de GK seja cerca de cinco vezes maior que o de GG (21 iterações contra 4 para a estrutura com 6 clusters), o tempo de treinamento total de ambos os classificadores é de cerca de 300 milissegundos.

TABELA 2. Validação dos algoritmos KM, FCM, GK e GG.

-		KM			
#Clusters	XB	SC	PC	DI	Time
2	4.2808	1.3477	n/a	0.0170	1.1613
3	3.1336	0.5161	n/a	0.1398	1.2717
4	6.4833	0.9074	n/a	0.0947	0.8981
5	4.3469	0.7834	n/a	0.0607	1.0503
6	<b>2.9845</b>	0.6153	n/a	<b>0.1460</b>	1.4029
7	5.7141	<b>0.5319</b>	n/a	0.1019	1.1848
8	6.5851	1.1006	n/a	0.0716	0.6750
-		FCM			
#Clusters	XB	SC	PC	DI	Time
2	5.1576	1.4498	0.6047	0.1463	0.1013
3	3.2810	1.3918	0.6867	0.1497	0.0609
4	6.2318	0.9511	0.6466	0.1588	1.08191
5	3.3109	0.9490	0.6251	0.1807	0.0851
6	<b>2.2784</b>	<b>0.5018</b>	<b>0.8136</b>	<b>0.1993</b>	0.0634
7	2.7849	0.8718	0.6080	0.1902	0.0693
8	2.7602	0.8748	0.5779	0.1880	0.0977
-		GK			
#Clusters	XB	SC	PC	DI	Time
2	3.2111	1.4282	0.8168	0.1381	0.1491
3	2.3771	1.3679	0.6978	0.1351	0.3416
4	5.8738	0.9044	0.7393	0.1570	0.5805
5	3.3392	0.6429	0.7972	0.1659	0.2342
6	<b>2.2428</b>	<b>0.4661</b>	<b>0.8386</b>	<b>0.1696</b>	0.3456
7	2.5160	0.5518	0.8237	0.1272	0.7081
8	2.5985	0.5406	0.8125	0.1389	0.6729
-		GG			
#Clusters	XB	SC	PC	DI	Time
2	2.0344	3.1814	0.9295	<b>0.1752</b>	0.6870
3	1.7304	3.0563	0.9411	0.1667	0.2260
4	2.0565	1.2786	0.9684	0.1547	0.1820
5	1.7349	1.2182	0.9729	0.1510	0.2050
6	<b>1.3156</b>	<b>1.1139</b>	<b>0.9740</b>	0.1505	0.2603
7	1.5572	1.2330	0.9710	0.1485	0.3101
8	1.5543	2.6780	0.9394	0.1500	0.3461

A Figura 3 ilustra fronteiras de decisão suaves fornecidas pelo uso de clusters com geometrias diferentes. Enquanto o método FCM é baseado em clusters esféricos sobrepostos, GK e GG permitem a formação de elipsóides no espaço dos dados devido ao uso das funções de distância Manhattan e Gauss, respectivamente. Os classificadores FCM e GK de 6 partições são considerados na figura. Os limites de decisão para o classificador GG são fundamentalmente semelhantes aos do classificador GK e foram, portanto, omitidos. É

importante notar na Fig. 3 que foram escolhidos para serem mostrados apenas os atributos mais representativos no controle de um processo de compostagem da cama, isto é,  $x_{(1)}$  (umidade da cama) e  $x_{(2)}$  (temperatura da cama). Observe também que os clusters GK são alongados para capturar a orientação dos dados. Cada um dos agrupamentos está associado à um rótulo de classe, um grau e uma descrição do grau de eficiência da compostagem da cama, conforme listado na Tabela 1.

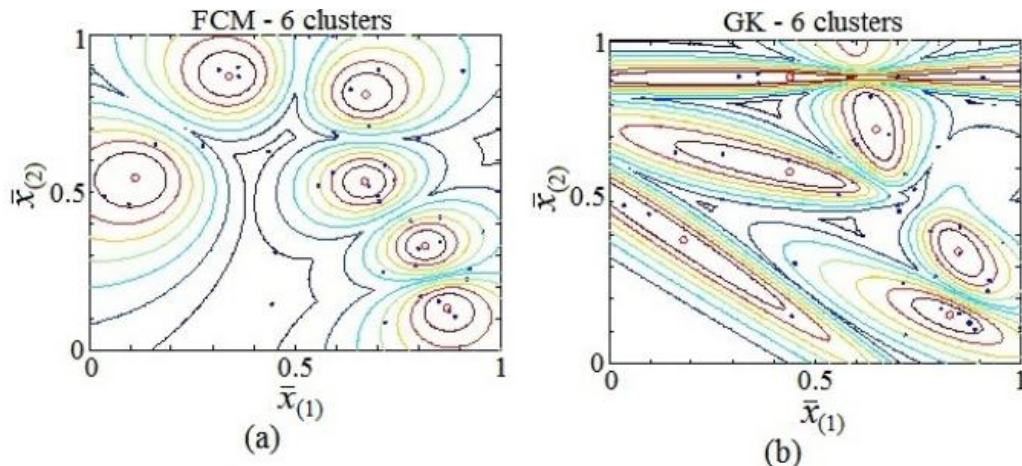


FIGURA 3. (a) Clusters FCM (b) Clusters GK. Pontos vermelhos são centros; pontos azuis são dados amostrais

Um ponto a ser enfatizado é que o índice de validação *Dunn Index* atesta que o classificador FCM de 6 clusters fornecem os melhores resultados, o que se opõe aos outros índices. Estes apontam GK e GG como os mais precisos. Da equação (5), DI depende da distância mínima entre amostras de diferentes clusters,  $d_{min}$  e a maior distância intra-cluster,  $d_{max}$ . Como os grupos fuzzy geralmente se cruzam, o valor de  $d_{min}$  aparentemente não muda muito com geometrias de agrupamento diferentes. Por outro lado, o valor de  $d_{max}$  é maior se clusters com geometria esparsa forem permitidos. A Figura 4 mostra claramente que  $d_{max}$  tende a ser maior para elipses com semi-eixos maiores ligeiramente maiores do que o raio de um círculo. A seta tracejada na Fig. 4(b) é o valor de  $d_{max}$  na Fig. 4(a). A partir desta análise particular, DI parece não ser apropriado para comparar FCM à GK e GG.

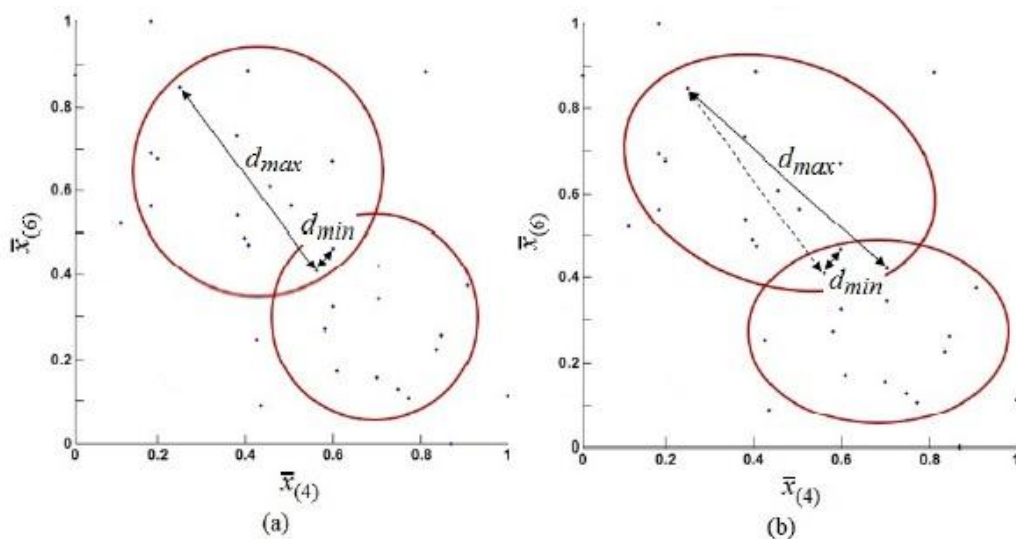


FIGURA 4. Ilustração do cálculo do *Dunn Index* para clusters com geometrias diferentes: (a) esfera; (b) elipse

O objetivo dos classificadores avaliados é auxiliar os produtores e agricultores na tomada de decisões relacionadas à sua propriedade. O objetivo final dos sistemas de galpões CB é que o gado leiteiro permaneça mais saudável e mais limpo com uma baixa incidência de problemas de locomoção e, portanto, tenha uma alta produção. Modelagem *fuzzy*, graus de pertinência, e informações linguísticas associadas a clusters são importantes para melhorar a compreensão do processo de compostagem e para decidir e controlar variáveis como a aeração, umidade e temperatura da cama.

## CONCLUSÕES

Os índices de validação *Xie-Beni*, *Partition Coefficient*, *Partition Index* e *Dunn Index* indicaram a existência de 6 classes nos dados. As classes estão relacionadas à um Índice do Processo de Compostagem, CPI, ou grau de eficiência do processo de compostagem. O grau de eficiência refere-se à estabilidade, maturidade e atividade do composto, e a baixa concentração bacteriana na cama.

O agrupamento GG mostrou-se ser o mais adequado de acordo com a maioria dos índices, seguido pelo GK. Estes forneceram consistentemente as melhores partições. A principal razão para os melhores resultados é o uso de medidas de distância baseadas na máxima verossimilhança e métrica de Mahalanobis. Essas distâncias produziram clusters mais flexíveis para os dados e, portanto, foram decisivas para a obtenção de melhores resultados. Os resultados da classificação *fuzzy* e as informações linguísticas mostraram-se úteis para auxiliar a tomada de decisão em fazendas que adotam CB como sistema de confinamento de bovinos leiteiros.

Estudos posteriores sobre a comunidade microbiana do composto podem ter efeitos benéficos aos produtores. Compostos derivados de camas do sistema CB podem agregar valor à propriedade, com a formação de material orgânico que pode ser utilizado como adubo fertilizante. O desenvolvimento de modelos *fuzzy* funcionais do tipo Takagi-Sugeno para tratar este problema está atualmente em andamento. Diferentes variações dos métodos de agrupamento GK e GG também serão considerados em trabalhos futuros.

## REFERÊNCIAS

BABUSKA, R.; VERBRUGGEN, H. Neuro-fuzzy Methods for Nonlinear System Identification, **Annu. Rev. Control.** v.27, n. 1, p.73–8, 2003.

BABUSKA, R. **Fuzzy Modeling for Control**, Springer, Boston, 1998.

BALASKO, B.; ABONY, J.; FEIL, B. Fuzzy Clustering and Data Analysis Toolbox, **Department of Process Engineering**, University of Veszprem, Veszprem (Report), p.74. 2005.

BARBERG, A. E.; ENDRES, M. I.; SALFER, J. A.; RENEAU, J. K. Performance and Welfare of Dairy Cows in an Alternative Housing System in Minnesota, **Journal of Dairy Science**, v. 90, p.1575-1583, 2007.

BENSAID, A. M.; HALL, L. O.; BEZDEK, J. C.; CLARKE, L. P.; SILBINGER, M. L.; ARRINGTON, J. A., MURTAGH, R. F. Validity-Guided (Re) Clustering with Applications to Image Segmentation. In: IEEE TRANS. FUZZY SYST. n.4, p.112–123, 1996.

BEZDEK, J. C., **Pattern Recognition with Fuzzy Objective Function Algorithms**, 1st ed. Springer US, 1981.

BLACK, R. A.; TARABA, J. L.; DAY, G. B.; DAMASCENO, F. A.; BEWLEY, J. M. Compost Bedded Pack Dairy Barn Management, Performance, and Producer Satisfaction, **J. Dairy Sci.** v.96, p.8060–8074, 2013.

CHATTOPADHYAY, S.; PRATIHAR, D.; SARKAR, S. A Comparative Study of Fuzzy C-Means Algorithm and Entropy-Based Fuzzy Clustering Algorithms, **Comput. Informatics.** v.30, p. 701–720, 2012.

COOK, N. B.; REINEMANN, D. J., Week ONE: A Tool Box for Assessing Cow, Udder and Teat Hygiene, In: 46TH ANNU. MTG. OF THE NATL. MASTITIS COUNC. San Antonio, Madison, W, p.31– 43, 2008.

DAMASCENO, F. A. **Compost Bedded Pack Barns system and computational simulation of airflow through naturally ventilated reduced model.** 2012. 391f. Tese (Doutorado). Curso de Pós-Graduação em Engenharia Agrícola. Universidade Federal de Viçosa - UFV, Viçosa, MG, 2012.

GALAMA, P. J. Prospects for Bedded Pack Barns for Dairy Cattle. **Wageningen UR Livestock Research**, Lelystad - The Netherlands (Report), 2011.

GIUSTI, E.; MARSILI-LIBELLI, S. Fuzzy Modelling of the Composting Process, **Environ. Model. Softw.** v.25, p. 641–647, 2010.

HOPNER, F.; KLAWONN, F.; KRUSE, R.; RUNKLER, T. **Fuzzy Cluster Analysis: Methods for Classification, Data Analysis and Image Recognition.** Wiley IBMPC Series, 1st ed, 1999.

LESO, L.; UBERTI, M.; MORSHED, W.; BARBARI, M. A Survey of Italian Compost Dairy Barns, **J. Agric. Eng.** v.44, n.3, p.120-124, 2013.

SPRECHER, D. J.; HOSTETLER, D. E.; KANEENE, J. B. A Lameness Scoring System that Uses Posture and Gait to Predict Dairy Cattle Reproductive Performance, **Theriogenology.** v.47, p.1179–87, 1997.

WANG, K.; WANG, B.; PENG, L. CVAP: Validation for Cluster Analyses, **Data Sci.J.** v.8, n.1, p. 88–93, 2009.

ZADEH, L. A. **Fuzzy sets, Inf. Control.** v.8, p.338–353. 1965.

XIE, X. L.; BENI, G., A Validity Measure for Fuzzy Clustering, In: IEEE TRANS. PATTERN ANAL. Mach. Intell. n.3, 841–847, 1991.