

MÉTODOS EXPLORATÓRIOS MULTIVARIADOS DE AGRUPAMENTO EM DADOS DE INFESTAÇÃO DE PLANTAS DANINHAS

CASTRO, M. J. F. (FCAV – UNESP, Jaboticabal/SP – marina-junqueira@bol.com.br), FERRAUDO, A. S. (DCE/FCAV – UNESP, Jaboticabal/SP – fsajago@gmail.com), ALVES, P. L. C. A. (DCE/FCAV – UNESP, Jaboticabal/SP – plalves@fcav.unesp.br), KUVA, M. A. (Herbae C. P. Agrícolas Ltda. - mkuva@herbae.com.br), FERRAUDO, G. M. (Pesquisador Associado da CanaVialis, grupo Monsanto, doutorando FCAV – UNESP, Jaboticabal/SP - guiferraud@gmail.com)

RESUMO: O objetivo deste trabalho foi avaliar a estrutura de grupos por análise multivariada de agrupamento por métodos hierárquico e não hierárquico bem como compará-las ao comportamento das redes neurais artificiais em dados de comunidades de plantas daninhas visando aperfeiçoar o processo de escolha de tratamentos químicos adequados para reduzir custos e preservar o meio ambiente, uma vez que causam grandes prejuízos à cultura da cana-de-açúcar. O conjunto de dados utilizado é constituído por 782 registros de plantas daninhas contidos em talhões de cana, obtidos junto a usinas de açúcar e álcool localizadas na região norte do Estado de São Paulo. Os registros levam em conta o número de táxons de plantas daninhas por categorias, com participação mínima de 3% na infestação. A análise de agrupamento por método hierárquico resultou em um dendrograma contendo uma estrutura de grupos caracterizada inicialmente por dois grandes grupos subdivididos em outros 12 grupos, a análise de agrupamento por método não hierárquico possibilitou explorar a influência das variáveis na classificação das comunidades de plantas daninhas nesses grupos. Os resultados da organização das amostras nos grupos obtidos pela rede de Kohonen foi um pouco diferente da organização em grupos obtida com as análises de agrupamento por método hierárquico.

Palavras-chave: Plantas daninhas, Cana-de-açúcar, Análise estatística multivariada de agrupamento

INTRODUÇÃO

A análise de agrupamento tem por finalidade reunir, por algum critério de classificação as unidades amostrais em grupos, de tal forma que exista homogeneidade dentro do grupo e heterogeneidade entre grupos (Johnson & Wichern, 1992; Cruz & Regazzi, 1994).

A análise de agrupamento por método hierárquico é aplicada inicialmente quando não se conhece nenhuma informação “a priori” sobre o número de grupos a serem obtidos. É uma

técnica que permite classificar (indivíduos ou objetos) em subgrupos excludentes. A representação dos grupos é feita num gráfico com uma estrutura de árvore denominado dendrograma (Sneath & Sokal, 1973) e quando se tem uma definição do número k de grupos aplica-se análise de agrupamento por método não hierárquico processada com o algoritmo k-médias que busca a melhor solução na formação dos k grupos (HAIR et al., 2005).

Nos últimos anos técnicas estatísticas multivariadas, tem sido aplicadas utilizando registros de dados obtidos em comunidades de plantas daninhas. Pereira et al. (2010) realizaram um ensaio com oito diferentes tratamentos químicos com o objetivo de estudar a viabilidade da utilização de técnicas da estatística multivariada para avaliar as alterações da composição florística das plantas daninhas. Dentre as conclusões, informam que as técnicas estatísticas utilizadas foram capazes de avaliar as alterações da composição florística pelos tratamentos empregados, sendo que as espécies de plantas daninhas mais discriminadas foram *Alternanthera tenella* (ALRTE) e *Commelina benghalensis* (COMBE), seguidas das gramíneas *Panicum maximum* (PANMA), *Cenchrus echinatus* (CCEC) e *Eleusine indica* (ELEIN).

Outra abordagem utilizada como análise de agrupamentos são as redes neurais artificiais e, em especial, as redes competitivas de Kohonen ou Mapas Auto-Organizáveis (*Self-Organizing Maps – SOM*) que são modelos neurocomputacionais desenvolvidos pelo professor Teuvo Kohonen no início dos anos 80, apropriadas quando a não linearidade está presente nos dados (Kohonen, 1990).

Pitelli et al. (2009), utilizaram, para agrupar populações de macrófitas aquáticas que colonizaram o reservatório de Santana, Piraí-RJ, durante o ano de 2004, análises de agrupamento por método hierárquico e redes neurais artificiais.

Este estudo objetivou avaliar a estrutura de grupos por análise multivariada de agrupamento por métodos hierárquico e não hierárquico bem como compará-las ao comportamento das redes neurais artificiais em registros de infestação de plantas daninhas contida em talhões de cana de açúcar. Espera-se que os resultados obtidos contribuam para o aperfeiçoamento do processo de escolha de tratamentos químicos adequados para reduzir custos e preservar o meio ambiente.

MATERIAL E MÉTODOS

As informações de comunidades de plantas daninhas contidas em 782 registros de talhões de cana soca referentes à safra 2008/2009 e 2009/2010 com número de táxons de plantas daninhas especificados por dez categorias, com participação mínima de 3% na infestação, foram fornecidas por usinas de açúcar e álcool pioneiras no monitoramento de

plantas daninhas e em controles por tratamentos com herbicidas localizadas no Norte do Estado de São Paulo, Brasil.

Levando-se em conta a mínima participação no conjunto de talhões considerados bem como os aspectos agrônômicos, os 34 táxons inicialmente contidos no banco de dados foram reduzidos a 10 categorias: POROL (indivíduos da espécie *Portulaca oleraceae*); BRASS (indivíduos do gênero *Brachiaria*); DIGSS (indivíduos do gênero *Digitaria*); AMASS (indivíduos do gênero *Amaranthus*); CIPOS (*Ipomoea spp.*, *Merremia spp.*, *Momordica charantia* entre outras); PANMA (indivíduos da espécie *Panicum maximum*); SIDSS (os indivíduos do gênero *Sida*); COMBE (indivíduos da espécie *Commelina benghalensis*); OFE (outras folhas estreitas, família das Poaceas); OFL (outras folhas largas). As categorias OFE e OFL contemplaram todas as espécies que não atingiram 3% de participação considerando todos os registros que compuseram a pesquisa. Esse critério de redução foi utilizado anteriormente por Kuva et al. (2010).

Inicialmente os dados foram submetidos às análises de agrupamento por método hierárquico e não hierárquico, a primeira técnica foi processada utilizando como coeficiente de semelhança a distância euclidiana simples e como método de ligação o método de Ward. A análise de agrupamento por método não hierárquico foi processada utilizando o algoritmo k-médias (Hair et al., 2005). Finalmente os dados foram submetidos à análise de agrupamento por redes neurais artificiais de Kohonen. Esta rede possui duas camadas de neurônios artificiais: uma para receber os padrões de entrada e outra para gerar os padrões de saída, a classificação das amostras contidas nos grupos obtidos por análise de agrupamento por método hierárquico foi feita com uma rede cuja topologia consta de 10 neurônios (Kohonen, 1990).

Todas as análises foram processadas com o ambiente R (2010), versão 2.11.0. Além dos pacotes da distribuição base utilizaram-se os pacotes “RODBC”, “MASS”, “cluster” e “vegan”.

RESULTADOS E DISCUSSÕES

A Figura 1 contém o dendrograma contendo a estrutura de grupos das comunidades de plantas daninhas construído com a análise de agrupamento por método hierárquico. Observa-se uma estrutura contendo inicialmente dois grandes grupos que podem ser subdivididos em outros 12 grupos. Esta divisão em 12 grupos foi adotada no processamento da análise de agrupamento por método não hierárquico que possibilita explorar a influência das variáveis na classificação das comunidades de plantas daninhas nesses grupos.



Figura 1. Dendrograma resultante da análise de agrupamento por método hierárquico, processada com a distância euclidiana simples e o método de ligação de Ward.

Ao interpretar os padrões dos grupos a estrutura final ficou assim descrita: Grupo POROL (G_1); Grupo CIPOS (G_2); Grupo OFE (G_3); Grupo PANMA (G_4); Grupo SIDSS (G_5); Grupo BI - Infestação baixa (G_6); Grupo MBI - Infestação muito baixa (G_7); Grupo BRASS (G_8); Grupo PANMA (Monoespecífico) (G_9); Grupo OFL (G_10); Grupo DIGSS (G_11) e Grupo COMBE (Monoespecífico) (G_12). Os grupos Monoespecíficos são aqueles caracterizados especificamente por uma categoria de planta daninha, enquanto que os demais são grupos nos quais tem-se uma planta daninha predominante convivendo com outras categorias de baixas infestações.

Os resultados da organização das amostras nos grupos obtidos pela rede de Kohonen foi um pouco diferente da organização em grupos obtida com as análises de agrupamento por método hierárquico. A infestação da comunidade DIGSS (G_11) teve um padrão diferenciado, pois em único neurônio ficaram contidas suas amostras e somente elas. Assim, a infestação dessa comunidade pode ser considerada diferenciada e merece um tratamento químico diferenciado. A infestação da comunidade PANMA neste estudo ficou distribuída em duas populações: uma distribuição contendo PANMA que convive com outras espécies (G_4) e a outra em que somente ela está presente e que foi denominada de PANMA monoespecífica (G_9). A rede de Kohonen fez uma distinção diferente ainda, isolou num único neurônio uma parte da infestação de PANMA (G_4) e em outro neurônio as demais amostras de PANMA (G_4) juntamente com a infestação da comunidade PANMA monoespecífica (G_9). As infestações das comunidades POROL (G_1) e COMBE (G_12) apresentaram mesmo padrão de infestação, pois ficaram contidas num único neurônio podendo ambas serem controladas como o mesmo tratamento. As comunidades OFE e OFL se dividiram em dois grupos, onde dois neurônios apresentaram aproximadamente metade das amostras de OFE e metade de OFL podendo estas infestações serem controladas como tratamentos semelhantes. A comunidade SIDSS não apresentou um padrão único de infestação, pois praticamente metade das amostras ficou em um neurônio e metade em outro. As comunidades CIPOS (G_2) e BRASS (G_8) apresentaram mesmo padrão de

infestação, uma vez que suas amostras ficaram contidas num mesmo neurônio o que se sugere um único tratamento de controle.

CONCLUSÕES

O agrupamento dos talhões de acordo com a composição potencial de plantas daninhas poderá aperfeiçoar o processo de escolha de um tratamento químico adequado resultando em uma maior eficácia do processo, reduzindo a necessidade de uma segunda intervenção para repasse e promovendo assim redução dos custos e preservação do meio ambiente.

Estas abordagens são importantes ferramentas que devem ser inseridas no rol das análises utilizadas em estudos exploratórios na área biológica devido à qualidade dos resultados gerados na classificação dos objetos em estudo, facilitando a interpretação das complexas relações entre o grande número de variáveis dependentes envolvidas.

REFERÊNCIAS BIBLIOGRÁFICAS

- CRUZ, C. D.; REGAZZI, A. J. Divergência genética. In: CRUZ, C. D.; REGAZZI, A. J. **Métodos biométricos aplicados ao melhoramento genético**. Viçosa, UFV: Imprensa Universitária. 1994, cap. 6, p. 287-323.
- HAIR, J. F. et al. **Análise Multivariada de dados**. 5. ed. Porto Alegre: Bookman Ed., 2005. v. 1.
- JOHNSON, R. A.; WICHERN, D. W. **Applied multivariate statistical analysis**. 3 ed. New Jersey: Prantice Hall, 1992. 642p.
- KOHONEN, T. The self-organizing map. **Proceedings IEEE**, v. 78, n. 9, p. 1464-1480, 1990.
- KUVA M. A. et al. Organização das comunidades infestantes de plantas daninhas na cultura da cana-de-açúcar em agrupamentos padrões. In: CONGRESSO BRASILEIRO DA CIÊNCIA DAS PLANTAS DANINHAS, 27., 2010, Ribeirão Preto. **Anais...** p. 2190-2194.
- PEREIRA, F. C. M. et al. Estudo de seleção de flora de plantas daninhas por herbicidas utilizando técnicas de análise multivariada – Índice de valor de importância. In: CONGRESSO BRASILEIRO DA CIÊNCIA DAS PLANTAS DANINHAS, 27º, 2010 Ribeirão Preto. **Anais...** p. 154-158.
- PITELLI, R. L. C. M. et al. Utilização de análise multivariada e redes neurais artificiais na determinação do comportamento de colonização de populações de macrófitas aquáticas no reservatório de Santana. **Planta Daninha**, Viçosa, v. 27, n. 3, p. 429–439, 2009.
- R Development Core Team. **R: A language and environment for statistical computing**. Vienna: R Foundation for Statistical Computing, 2010. Disponível em: <<http://www.R-project.org>>. Acesso em: 20 abril 2011.
- SNEATH, P. H. A.; SOKAL, R. R. **Numerical taxonomy**. San Francisco: W. H. Freeman, 1973.