

IDENTIFICANDO PADRÕES DE DADOS DE PRODUTIVIDADE E DE FALHAS EM CANAVIAIS COM TÉCNICA DE AGRUPAMENTO

Gabriel Fogaça ¹; Eudocio Rafael Otavio da Silva ²; José Paulo Molin ³

¹ Graduando em Eng^o Agrônômica, Lab. de Agricultura de Precisão, Depto. Engenharia de Biosistemas, Escola Superior de Agricultura Luiz de Queiroz, ESALQ, Universidade de São Paulo, Piracicaba – SP.

² Eng. Agrônomo, Doutorando, Lab. de Agricultura de Precisão, Depto. de Engenharia de Biosistemas, Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo, Piracicaba – SP.

³ Eng. Agrícola, Professor Titular, Lab. de Agricultura de Precisão, Depto. de Engenharia de Biosistemas, Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo, Piracicaba – SP.

Apresentado no
Congresso Brasileiro de Agricultura de Precisão e Digital - ConBAP 2024
Porto Alegre, RS, 2 a 4 de julho de 2024

RESUMO: A cana-de-açúcar é uma cultura de extrema importância econômica no Brasil. Diversas ferramentas de monitoramento têm sido desenvolvidas para aumentar a produtividade e longevidade dos canaviais. O monitoramento de falhas nas fileiras de cana-de-açúcar destaca-se por ser um problema recorrente e que influencia diretamente na produtividade. Entretanto, o padrão de distribuição e comprimento das falhas em áreas com diferentes potenciais produtivos é pouco discutido. Este estudo teve o objetivo de investigar padrões de ocorrência do comprimento de falhas na fileira de cana-de-açúcar em regiões de potenciais produtivos distintos do canavial, utilizando a técnica de agrupamento. Dados de produtividade obtidos de colheita em canaviais comerciais foram submetidos à análise de cluster, sendo agrupados de acordo com o potencial produtivo de cada região. Para cada cluster, foram verificadas as ocorrências e comprimentos de falhas na fileira da cultura. Os resultados mostraram que nem sempre as regiões de menor potencial produtivo apresentaram a maior média do comprimento de falhas, mas sempre apresentaram a maior área de falhas. Os resultados também evidenciaram um comportamento padrão na curva de distribuição das falhas para todos os clusters analisados.

PALAVRAS-CHAVE: Cana-de-açúcar; k-médias; Sensoriamento remoto.

PERFORMANCE OF FILTERING METHODS FOR YIELD DATA IN SUGARCANE FIELDS USING ARTIFICIAL OUTLIERS

ABSTRACT: Sugarcane is a crop of extreme economic importance in Brazil. Several monitoring tools have been developed to increase the yield and longevity of sugarcane fields. Monitoring failures in sugarcane rows stands out as a recurring problem that directly influences the yield of sugarcane fields. However, the distribution pattern and length of gaps in areas with different productive potentials is little discussed. This study aimed to investigate patterns of occurrence of the length of gaps in the sugarcane row in regions with different production potentials than the sugarcane field, using the clustering technique. Yield data obtained from harvesting in commercial sugarcane fields were subjected to cluster analysis, being grouped according to the productive potential of each region. For each cluster, the occurrences and lengths of gaps in the crop row were verified. The results showed that the regions with the lowest productive potential did not always have the highest average gaps length, but they always presented the largest gaps area. The results also showed a standard behavior in the gap distribution curve for all clusters analyzed.

KEYWORDS: Sugarcane, Remote Sensing.

INTRODUÇÃO: O Brasil é o maior produtor mundial da cultura da cana-de-açúcar. De acordo com a Companhia Nacional de Abastecimento (CONAB), na safra 2022/23 o Brasil produziu um total de 610.131,4 mil toneladas de cana-de-açúcar (CONAB, 2023). Dada a importância da cultura, diversas técnicas de manejo

são adotadas nas lavouras de cana, dentre elas, a colheita mecanizada. Apesar dos seus diversos benefícios, ainda prevalecem problemas como a compactação do solo e degradação das plantas nas fileiras ocasionados pelo tráfego de máquinas (LUZ et al., 2023). Os danos causados pelo tráfego de máquinas na colheita em conjunto com fatores genéticos, fisiológicos e climáticos podem degradar as plantas, afetando a brotação (SERAFIM et al., 2013). O não desenvolvimento da cana-planta ou cana-soca geram falhas nas fileiras de cana-de-açúcar. Stolf (1986) realizou um dos primeiros estudos referentes a falhas e propôs que elas seriam equivalentes à distância entre duas plantas na fileira. A partir de experimentações, verificou que quanto maior a porcentagem de falhas no canavial, maior a redução da produtividade de cana. Décadas depois, foi verificado em Maldaner et al. (2024) que a densidade de plantas afeta a produtividade da cana, considerando regiões de produtividades distintas e que a densidade populacional diminuiu ao longo do tempo à medida que o povoamento se deteriorava. Logo, verifica-se que são relativamente novos na literatura os estudos que apresentam as relações da falha e da produtividade em alta resolução espacial. A principal decisão derivada da ocorrência de falhas é a de reforma do canavial. Entretanto, não se faz distinção, dentro de um mesmo talhão, entre os padrões de distribuição e comprimento de falhas em regiões com diferentes potenciais produtivos, a fim de compreender se a decisão de gestão dessas falhas deve ser a mesma independente da produção observada/esperada em uma área. Dessa forma, se faz necessário identificar padrões espaciais das falhas em diferentes contextos e cenários na lavoura, considerando o potencial produtivo da região onde elas estão presentes.

OBJETIVOS: O objetivo deste estudo foi investigar padrões de ocorrência do comprimento de falhas na fileira em regiões de potenciais produtivos distintos do canavial, utilizando a técnica de agrupamento.

MATERIAL E MÉTODOS: O estudo foi realizado em área comercial de cana-de-açúcar, localizada no estado de São Paulo, Brasil. A região possui clima tropical de inverno seco, caracterizada como Aw, de acordo com a classificação climática de Köppen-Geiger. A área total da lavoura abrange 42,82 ha, em que os talhões 1 e 2 possuem 23,03 e 23,79 ha, respectivamente. A variedade de cana-de-açúcar cultivada nos talhões é a RB966928, colhida em terceiro corte. Os dados de produtividade foram coletados a partir de um sistema de sensores comerciais (Solinftec, Araçatuba, São Paulo, Brasil) que toma como base a pressão de propulsão necessária para picar a cana-de-açúcar no motor do sistema picador (MALDANER et al., 2022). A colhedora, de uma fileira, é equipada com um receptor Sistema Global de Navegação por Satélite (GNSS) de posicionamento acurado. Após a obtenção dos dados de colheita, os conjuntos de dados foram submetidos a filtragem, visando remover dados errôneos e discrepantes. A coleta de dados de falhas na fileira da cana por meio de Veículo Aéreo Não Tripulado (UAV) foi realizada por uma empresa de serviços contratada para esse fim. Nesse método, o mosaico de imagens foi gerado a partir de voo do UAV, modelo eBee equipado com um sensor RGB S.O.D.A com resolução espacial de 0,03 m e as geometrias de linha, correspondente às fileiras de cana, foram reconstituídas. O voo foi realizado baseado na altura da planta, quando a planta atingiu entre 0,60 a 1,20 m de crescimento no início da safra, na visibilidade da entrelinha e fase de perfilhamento da cana soca. Posteriormente, foram obtidas as falhas na fileira de cana, utilizando-se como critério a identificação de espaçamentos maiores do que 0,58 m entre as plantas, com base nos dados extraídos do mosaico de imagens gerado pelo UAV. Para realizar a análise dos padrões de dados de produtividade e de falhas, foram gerados três agrupamentos (do inglês, clusters) classificados em alta, média e baixa produtividade. A análise de agrupamento tem o intuito de agrupar dados semelhantes em clusters, tornando-se uma ferramenta altamente eficaz para a análise de dados (GÉRON, 2021). Nesse estudo, os clusters foram gerados por meio do método fuzzy k-médias. O método k-médias classifica os dados em grupos. A partir do momento em que é informado o número de classes desejadas, são calculados os centróides de cada grupo. Posteriormente, os dados são realocados conforme a similaridade entre eles, através do cálculo entre a distância de cada dado em relação aos centróides definidos. O processo se repete até que os clusters não mudem entre as iterações. O método fuzzy k-médias possui processo semelhante ao k-médias. No entanto, no método fuzzy k-médias todos os dados afetam o cálculo dos centróides, pois no algoritmo, os dados possuem um grau de pertencimento, que é inversamente proporcional à distância desse ponto ao centróide do cluster (HOT & POPOVI?-BUGARIN, 2015). Para isto, utilizou-se os dados de produtividade filtrados de cada talhão através do plugin Smart-map (PEREIRA et al., 2022) inserido no Sistema de Informação Geográfico (SIG) QGIS v. 3.28.14. Foram definidos

os parâmetros Fuzzy Performance Index (FPI), Normalized Classification Entropy (NCE) e k igual a três (PEREIRA et al., 2022). Esse agrupamento permitiu a separação em três regiões de diferentes potenciais produtivos da cana. Após a criação dos clusters, foi usada a função de 'Interseção' do SIG QGIS, para identificar os dados de falhas correspondentes a cada cluster. Para cada cluster em cada talhão, foram obtidas a quantidade de falhas (espaçamento entre plantas), os valores do comprimento de falha na fileira e a porcentagem de falhas no campo (MOLIN & VEIGA, 2016). Os dados de produtividade e de falhas dos clusters em cada talhão foram submetidos à estatística descritiva, em que foram obtidos os seus valores máximos, mínimos, médios, desvio padrão (DP) e coeficiente de variação (CV%). A comparação da produtividade e do comprimento das falhas relacionadas aos clusters foi avaliada por meio da análise de variância (ANOVA) e, quando encontradas diferenças significativas, os valores médios foram comparados pelo teste Tukey a 5% de probabilidade ($p < 0,05$).

RESULTADOS E DISCUSSÃO: A Figura 1 exibe os clusters gerados de acordo com os valores de produtividade de cada talhão, sendo classificados em alta, média e baixa produtividade (Figura 1).

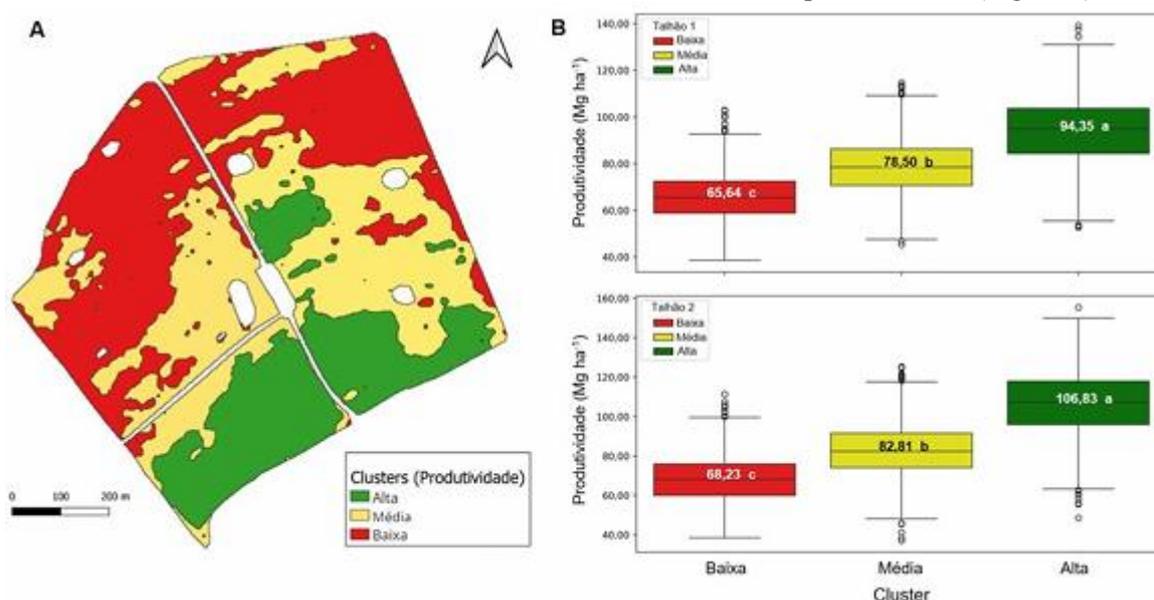


FIGURA 1. Clusters de produtividade obtidos nos talhões 1 e 2 (A) e comparação da produtividade (em Mg ha⁻¹) em cada cluster (B). Para cada talhão, letras minúsculas diferentes dentro da caixa indicam diferença significativa entre a produtividade dos clusters (teste de Tukey; $y; p < 0,05$).

A partir da Tabela 1, é possível observar os resultados das análises dos dados de produtividade de cada cluster.

TABELA 1. Análise estatística dos dados de produtividade nos clusters para os talhões 1 e 2.

Talhão	Cluster	Produtividade média (Mg ha ⁻¹)	Valor mínimo (Mg ha ⁻¹)	Valor máximo (Mg ha ⁻¹)	Desvio Padrão	Coeficiente de variação (%)
1	Alta	94,35	52,64	139,08	14,65	15,53
	Média	78,50	45,21	114,65	11,39	14,51
	Baixa	65,64	38,52	102,99	9,89	15,07
2	Alta	106,83	48,61	155,25	16,36	15,31
	Média	82,81	37,12	125,24	12,89	15,56
	Baixa	68,23	38,42	111,27	11,15	16,34

Os resultados da análise estatística da produtividade mostram as diferenças entre os clusters dos talhões 1 e 2. Por meio do teste de comparação de médias de Tukey, observou-se que os valores médios de produtividade entre os clusters foram significativamente diferentes ($p < 0,05$). Dessa maneira, o agrupamento se mostrou eficaz na distinção de zonas de produção de cana.

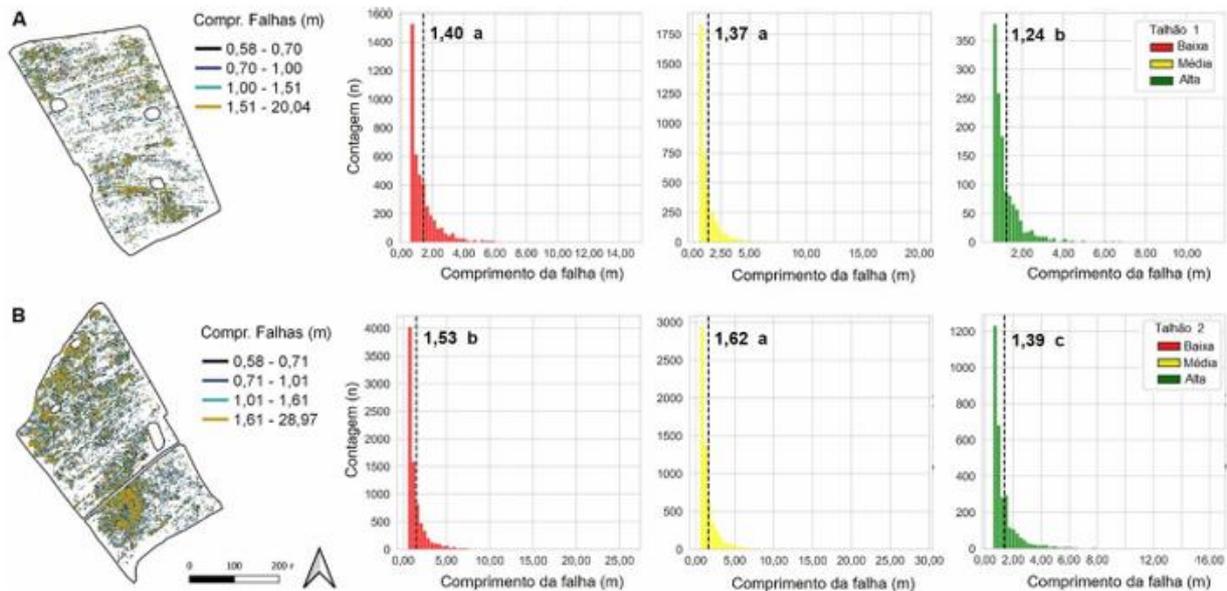


FIGURA 2. Distribuição espacial dos dados de falhas na fileira de cana coletados obtidos por UAV e comparação do comprimento de falhas (em m) em cada cluster nos talhões 1 (A) e 2 (B). A linha vertical tracejada indica o valor médio do comprimento de falha no histograma. Para cada talhão, letras minúsculas diferentes indicam diferença significativa entre o comprimento de falhas dos clusters (teste de Tukey; $p < 0,05$).

A Figura 2 ilustra a distribuição espacial das falhas nos talhões 1 e 2 e compara os comprimentos de falhas em cada cluster. Observa-se que os comprimentos de falhas variaram entre 0,58 e 20,04 m no talhão 1 e entre 0,58 e 28,97 m no talhão 2. A análise das curvas de distribuição mostrou um padrão de comportamento para as falhas em que mesmo separando-as em clusters dentro do talhão, apresentam uma curva exponencial decrescente, independente da ocorrência de falha ser em uma região de alto ou baixo potencial produtivo. Sendo assim, observou-se que à medida que o comprimento de falha aumenta, sua ocorrência tende a diminuir. Este comportamento da curva corrobora com o que foi discutido em Stolf (1986), que em seu experimento para avaliar falhas nas fileiras de cana também foi observada a diminuição da ocorrência de falhas à medida que os seus comprimentos aumentaram.

A Tabela 2 mostra a análise estatística dos dados de comprimento de falhas nos clusters para os talhões 1 e 2. A análise dos resultados revela que o comprimento médio das falhas é consistente, independentemente do potencial produtivo da área. Isso sugere que o tamanho médio das falhas é uniforme em diferentes zonas de produção. O fator que mais influenciou a produtividade foi a quantidade total de falhas, ou seja, falhas longas são raras, mas seu impacto pode ser substancial ao totalizar o seu comprimento na zona de produção gerada (cluster). Portanto, além do número de falhas e do comprimento individual de cada uma, é crucial para os gestores o conhecimento da zona de produção em que estão localizadas estas falhas. Por exemplo, as áreas de baixo potencial produtivo tiveram um comprimento total de falhas 300% maior do que as áreas de alto potencial, mesmo com um comprimento médio de falhas semelhante, variando entre 1,24 m e 1,62 m por falha.

TABELA 2. Análise estatística dos dados de comprimento de falhas nos clusters para os talhões 1 e 2.

Talhão	Cluster	Média do comprimento de falha (m)	Valor mínimo (m)	Valor máximo (m)	Desvio Padrão	Coefficiente de variação (%)	Soma total de falhas (m)	Área de falhas (ha)	Área com falhas (%)
1	Alta	1,24	0,58	11,31	0,91	73,39	1.607,09	0,24	1,05
	Média	1,37	0,58	20,04	1,23	89,78	5.390,20	0,81	3,51
	Baixa	1,40	0,59	14,70	1,09	77,86	5.896,65	0,88	3,84
2	Alta	1,39	0,58	16,30	1,26	90,65	4.389,40	0,66	2,76
	Média	1,62	0,58	28,97	1,68	104,35	10.011,91	1,50	6,31
	Baixa	1,53	0,58	26,13	1,55	101,31	12.469,62	1,87	7,86

No talhão 2, o cluster de alta produtividade também apresentou a menor média de comprimento de falhas entre os demais clusters. No entanto, o cluster que apresentou a menor média de comprimento de falhas do talhão 2 foi o cluster de média produtividade, evidenciando que nem sempre o cluster de baixa produtividade apresentou a menor média de comprimento de falhas. Isto também é evidenciado pelo teste de comparação de médias de Tukey, no qual foi observado que o comprimento médio de falhas para os clusters de média e baixa produtividade não apresentaram diferenças significativas a 95% de confiança ($p < 0,05$). Entretanto, é necessário verificar mais do que os valores médios obtidos pelas variáveis em cada cluster. Logo, ao analisar os valores da soma total de falhas (m), área de falhas (ha) e área com falhas (%) para cada cluster em cada talhão, foi observado que o cluster de alta produtividade de cana-de-açúcar, sempre apresentou os menores valores desses parâmetros enquanto o cluster de baixa produtividade apresentaram os maiores valores.

É importante ressaltar que a relação entre a quantidade total de falhas e a produtividade não é linear. Uma falha de mesmo comprimento em áreas com diferentes potenciais produtivos resulta em perdas de produtividade variadas, sendo mais pronunciadas em áreas de maior potencial. Embora a distribuição espacial do comprimento das falhas seja semelhante em áreas de alto e baixo potencial produtivo, a ocorrência total de falhas difere. Isso suscita a questão: a maior ocorrência de falhas é consequência de uma região de baixo potencial produtivo, ou uma região de baixo potencial produtivo é consequência da maior ocorrência de falhas? Pesquisas futuras podem investigar essa questão, pois a resposta tem implicações significativas na interpretação dos dados de ocorrência de falhas e sua relação com diferentes áreas de produção, influenciando (e possivelmente alterando) as decisões dos gestores a partir disso.

Os resultados obtidos neste estudo indicam que a identificação de padrões de produtividade e de falhas, através do método fuzzy k-médias, possui potencial para ser utilizada como uma camada de informação na tomada de decisões, como, por exemplo, na identificação de regiões no canavial que possam ser reformadas, visto que os parâmetros geralmente considerados para isso são os valores de produtividade obtidos na safra e a porcentagem de falhas determinadas através de pequenas amostras aleatórias coletadas dentro do campo, dificultando a realização de uma caracterização confiável da variabilidade das falhas existentes nos canaviais (MALDANER et al. 2021). Assim, a geração desses agrupamentos se mostra promissora para a gestão do canavial

CONCLUSÃO: O agrupamento fuzzy k-médias mostrou que as falhas de maior comprimento não estão necessariamente situadas nos locais de menor produtividade de cana-de-açúcar. Ao contrário, o fator que se observou mais afetar a produtividade dos talhões analisados foi o comprimento total de falhas, que foi mais observado em locais de menor produtividade, apesar do comprimento médio das falhas ter sido similar ao das zonas de média e alta produtividade. Os resultados também mostram um padrão no comportamento das curvas de distribuição das falhas, sendo que em todos os clusters a curva de distribuição se apresentou de forma exponencial decrescente. Conclui-se que a identificação de padrões de falhas nas regiões de potenciais produtivos distintos por meio da técnica de agrupamento é uma potencial abordagem para a tomada de decisão no canavial

AGRADECIMENTO: À empresa Solinftec pelo suporte na obtenção dos dados.

REFERÊNCIAS

COMPANHIA NACIONAL DE ABASTECIMENTO (CONAB). **Acompanhamento da safra brasileira de cana-de-açúcar**, v. 10, safra 2022/23, n. 4, quarto levantamento, abril 2023.

GÉRON, A.; **Mãos à obra: aprendizado de máquina com Scikit-Learn, Keras & TensorFlow: Conceitos, ferramentas e técnicas para a construção de sistemas inteligentes**. 2^a ed., Alta Books, 2021.

HOT, E; POPOVI?-BUGARIN, V. Soil data clustering by using K-means and fuzzy K-means algorithm. **23rd Telecommunications Forum Telfor (TELFOR) IEEE**, p. 890-893, 2015.

LUZ, F. B.; GONZAGA, L. C.; CASTIONI, G. A. F.; LIMA, R. P., CARVALHO, J. L. N.; CHERUBIN, M. R. Controlled traffic farming maintains soil physical functionality in sugarcane fields. **Geoderma**, v. 432, 116427, 2023.

MALDANER, L. F.; CANATA, T. F.; MOLIN, J. P. An approach to sugarcane yield estimation using sensors in the harvester and zigbee technology. **Sugar Tech**, v. 24, n. 3, p. 813-821, 2022.

MALDANER, L. F.; MOLIN, J. P.; SILVA, E. R. O. Spatial-Temporal Analysis to Investigate the Influence of in-Row Plant Spacing on the Sugarcane Yield. **Sugar Tech**, v. 26, p. 194-206, 2024.

MOLIN, J. P.; VEIGA, J. P. S. Spatial variability of sugarcane row gaps: measurement and mapping. **Ciência e Agrotecnologia**, v. 40, p. 347-355, 2016.

PEREIRA, G. W.; VALENTE, D. S. M.; QUEIROZ, D. M. D.; COELHO, A. L. D. F.; COSTA, M. M.; GRIFT, T. Smart map: An open-source QGIS plugin for digital mapping using machine learning techniques and ordinary kriging. **Agronomy**, v. 12, n. 6, 1350, 2022.

SERAFIM, L. G. F.; STOLF, R.; SILVA, J. R.; SILVA, L. C. F.; MANIERO, M. A.; BASSINELLO, A. I. Influência do plantio mecanizado no índice de brotação da cana de-açúcar. **STAB**, v. 31, p. 22-25, 2013.

STOLF, R. Metodologia de avaliação de falhas nas linhas de cana-de-açúcar. **STAB**, v. 4, n. 6, p. 22-36, 1986