

Acompanhamento de safras de cana-de-açúcar por meio de técnicas de agrupamento em séries temporais de NDVI

Luciana Alvim Santos Romani ^{1,2}
Renata Ribeiro do Valle Gonçalves ³
Bruno Ferraz do Amaral ¹
Jurandir Zullo Jr. ³
Caetano Traina Jr. ¹
Elaine Parros Machado de Sousa ¹
Agma Juci Machado Traina ¹

¹ Universidade de São Paulo – USP/ICMC
Av. Trabalhador são-carlense, 400 - 13566-590 - São Carlos - SP, Brasil
brunoslash@grad.icmc.usp.br
{caetano, parros, agma}@icmc.usp.br

² Embrapa Informática Agropecuária
Caixa Postal 6041 – 13083-886 - Campinas - SP, Brasil
luciana@cnptia.embrapa.br

³ Cepagri/Feagri - Unicamp
Cidade Universitária – 13083-970 - Campinas - SP, Brasil
{renata, jurandir}@cpa.unicamp.br

Abstract. Sugar cane is one of the most important crops for the Brazilian economy, especially because it is the main source for ethanol production, replacing fossil fuels. The recent development of bi-fuel engines has further increased demand for ethanol, which is one of the most important derivatives of sugar cane. Recently, remote sensing imagery has been widely used to monitor sugar cane harvests and to support scientific research. However, with the advancement of remote sensing technology, large amounts of data have been generated and therefore the analysis and discovery of useful knowledge from these data have become essential tasks to assist research works. In this context, this paper proposes a methodology based on data clustering to analyze NDVI time series in order to monitor the growing cycles of sugar cane crops throughout the year. NDVI images were obtained from AVHRR/NOAA satellites that have low spectral resolution, but can be used to monitor agricultural crops grown in large areas such as sugar cane. The experiments conducted on a productive region of the São Paulo state show that the proposed approach can identify areas with similar development patterns also considering different growing seasons of the crops. The results confirm that AVHRR/NOAA images can be used satisfactorily in crop monitoring.

Palavras-chave: remote sensing, data mining, agriculture, sensoriamento remoto, mineração de dados, agricultura.

1. Introdução

Nos últimos anos, a cana-de-açúcar tem se tornado cada vez mais importante economicamente para o Brasil, por se configurar em uma alternativa viável para auxiliar na diminuição da emissão de gases do efeito estufa, bem como principal fonte de substituição de combustíveis fósseis. O aumento nas vendas de automóveis com motores bi-combustível (gasolina e etanol) também tem impulsionado a produção de etanol, um dos principais derivados da cana (Nass et al. 2007, Goldemberg et al., 2008). Dessa maneira, a cana passa a ser cada vez mais estratégica para o Brasil, que é o maior produtor mundial.

Embora pesquisas científicas tenham propiciado o aumento da produtividade das cultivares de cana-de-açúcar, o aumento na produção deve-se em grande parte à expansão da cana principalmente no estado de São Paulo, maior produtor brasileiro (Martinelli e Filoso, 2008). Recentemente, imagens de sensoriamento remoto têm sido amplamente utilizadas para

avaliar a substituição de outras culturas, como café e citros, por lavouras de cana-de-açúcar. Rudorff et al. (2009) têm usado séries de imagens EVI do MODIS para identificar o uso da terra e o avanço da cana em todo o Brasil. Em um estudo mais recente, Rudorff et al. (2010) confirmaram que imagens de satélite são eficientes para auxiliar na avaliação de características importantes do cultivo da cana, proporcionando resultados relevantes para o debate sobre a produção sustentável de etanol.

Outro fator importante para a indústria sucroalcooleira é a manutenção da colheita o ano todo para garantir a produção do etanol e outros derivados. Com isso, embora as datas de plantio não variem tanto, em algumas regiões a data de corte pode ser estendida gerando canas: soca (ano) e planta (ano e meio) (Scarpari e Beauclair, 2004). Embora haja um registro das plantações com ciclos vegetativos diferentes (de ano, e de ano e meio), em imagens de satélites de baixa resolução essa diferença não é evidente, ocorrendo mistura espectral. No entanto, como a cana-de-açúcar é cultivada em grandes extensões, satélites de baixa resolução espacial, como o AVHRR/NOAA, têm sido usados para estimar a produtividade da cana (Nascimento et al. 2009) e avaliar a influência das condições climáticas para a cultura (Gonçalves et al. 2009).

A série de satélites NOAA gera duas imagens diárias para cada satélite em órbita e, portanto, o volume de dados disponível para estudos tem crescido muito nos últimos anos. São necessários então, métodos computacionais que automatizem desde a correção geométrica das imagens até a análise da enorme quantidade de dados extraídos das imagens. Nesse contexto, a abordagem de descoberta do conhecimento, mais comumente denominada *Mineração de Dados*, tem sido considerada uma alternativa adequada para a exploração e a descoberta de padrões relevantes e novos nesse grande volume de dados (Han e Kamber, 2001; Romani et al. 2008; Romani et al. 2009).

Nesse sentido, este artigo propõe o uso da técnica de agrupamento de dados (*clustering*) para auxiliar no acompanhamento de safras de cana-de-açúcar a partir de séries temporais de NDVI obtidas do satélite AVHRR/NOAA para o estado de São Paulo, Brasil. Os experimentos realizados em uma região do estado que concentra áreas com produção alta de cana mostraram que é possível acompanhar a evolução da cultura ao longo da safra identificando regiões com padrões semelhantes. Além disso, pode-se obter uma classificação mensal dos valores de NDVI por região, o que pode servir de subsídio para pesquisas futuras.

O restante desse artigo descreve a Metodologia do Trabalho na Seção 2, discute os Resultados na Seção 3 e apresenta as Conclusões na Seção 4.

2. Metodologia do trabalho

A região nordeste do estado de São Paulo, situada entre as coordenadas geográficas 48°50' e 46°50' de longitude oeste e 22°30' e 20°30' de latitude sul, como ilustrado na Figura 1 foi escolhida como área de estudo para este trabalho. Responsável pela maior produção de cana-de-açúcar no estado, essa região corresponde à cena Landsat/TM órbita/ponto 220/75. Com a imagem Landsat-7/ETM+, adquirida em 03/10/2002, foi feito o mapeamento dos talhões de cana-de-açúcar. Esse mapeamento teve por finalidade extrair apenas os valores de NDVI da cana-de-açúcar, separando, assim, a área urbana, os solos expostos e outras vegetações.

Inicialmente, foram selecionados os talhões de cana-de-açúcar por meio de uma máscara produzida a partir da cena do Landsat 7. Concomitantemente, foram processadas imagens AVHRR/NOAA para o ano/safra 2002/2003 e foram geradas composições mensais das imagens NDVI. Em seguida, as séries de valores NDVI foram extraídas das imagens e finalmente as técnicas de agrupamento foram aplicadas. Um fluxograma da metodologia empregada neste trabalho é apresentado na Figura 2.

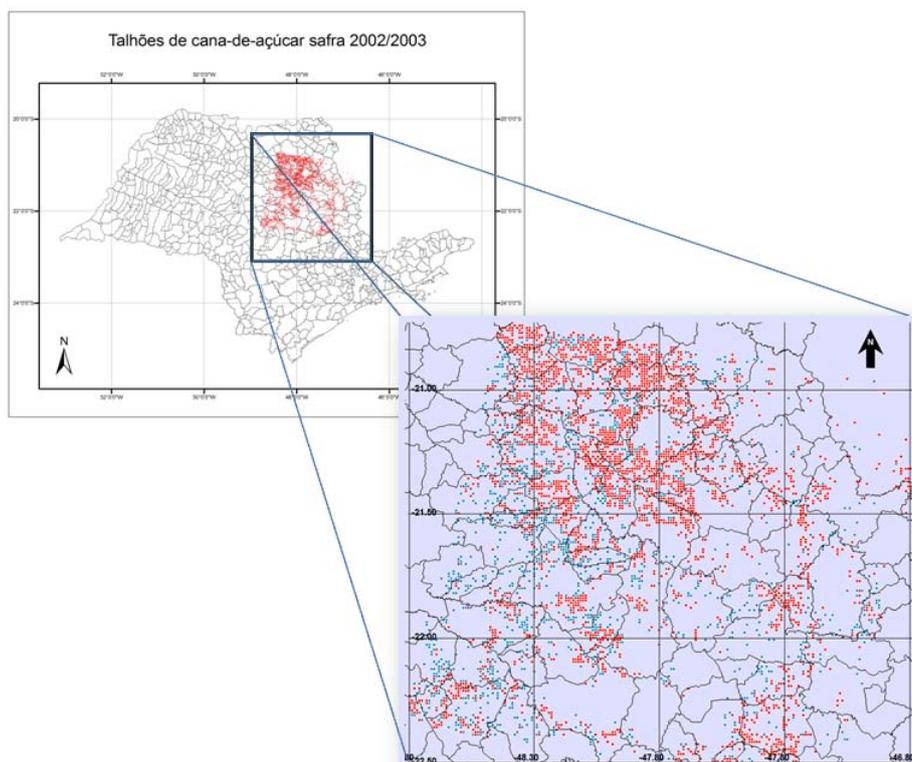


Figura 1. Área de estudo mostrando o mapeamento dos talhões de cana-de-açúcar.

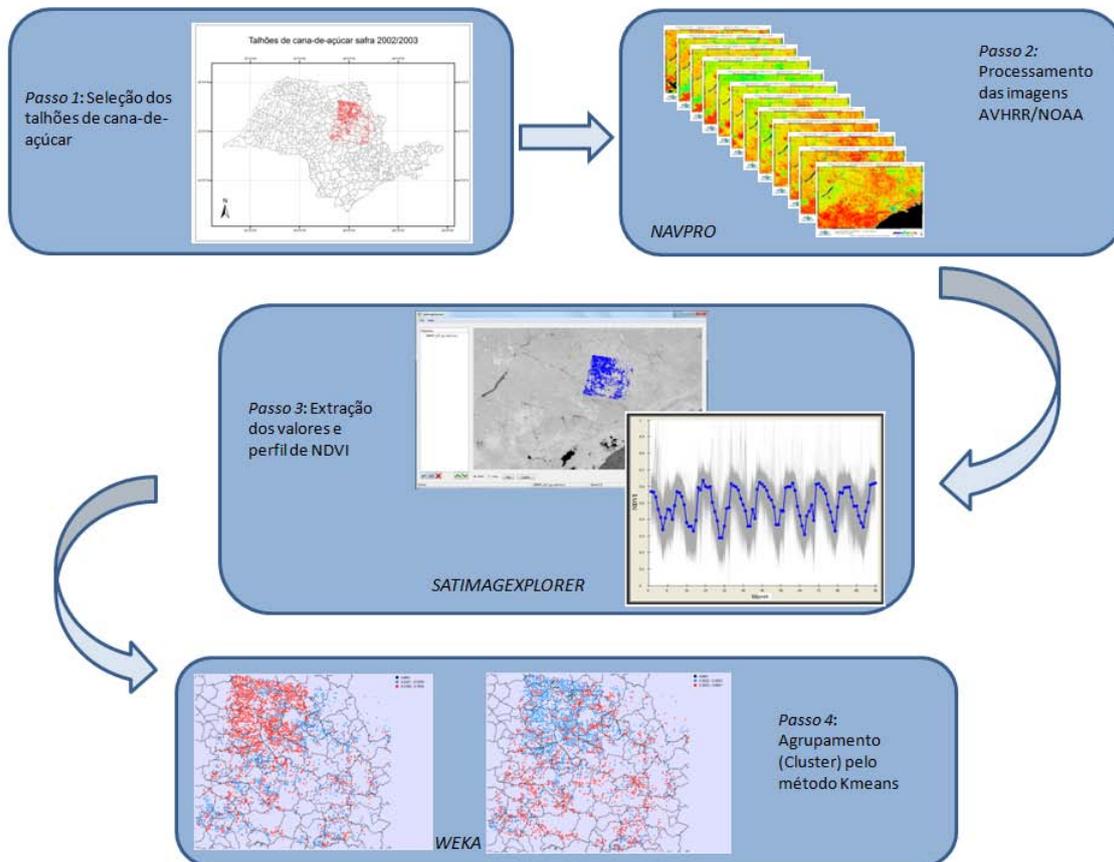


Figura 2. Fluxograma da metodologia utilizada neste trabalho.

Neste estudo foi utilizado o banco de imagens dos satélites AVHRR/NOAA do Centro de Pesquisas Meteorológicas e Climáticas Aplicadas à Agricultura, da Universidade Estadual de Campinas (Cepagri/Unicamp). Foram selecionadas imagens do satélite NOAA-16 para os meses de abril a junho de 2002 e imagens do satélite NOAA-17 de julho de 2002 a março de 2003. O processamento das imagens do AVHRR/NOAA foi realizado pelo sistema automático de navegação NAVPRO, que é um sistema para processamento e geração automática de produtos de imagens AVHRR/NOAA, desenvolvido por pesquisadores do *Colorado Center for Astrodynamics Research (CCAR), Aerospace Engineering Sciences*, da Universidade do Colorado (Boulder/EUA), em parceria com Cepagri/Unicamp (Esquerdo et al. 2006).

O sistema NAVPRO é compatível com a plataforma Linux e executado em “script Shell”, o qual realiza todo o processamento das imagens de forma automática, fazendo a conversão do formato, calibração radiométrica, correção geométrica (georreferenciamento preciso) e a geração de produtos, como o NDVI.

O cálculo do NDVI foi feito a partir de imagens diárias do AVHRR/NOAA já pré-processadas, isto é, imagens com a calibração radiométrica, georreferenciamento preciso e recorte de São Paulo. O NDVI foi obtido pelas bandas 1 e 2, que correspondem ao vermelho e infravermelho próximo, respectivamente. Nesse processo, foram excluídos os pixels com ângulo zenital solar maior que 70° e ângulos de varredura maiores que 42° . Esses pixels receberam o valor -1,1. Para amenizar o efeito da atmosfera nas imagens, foram geradas Composições de Valor Máximo (MVC) de NDVI mensal.

Com essas composições mensais de NDVI, foram definidos perfis temporais com valores de NDVI no Sistema *SatImagExplorer*, que permite que o especialista interaja com as imagens e especifique regiões de interesse (ROI), utilizando como base de entrada uma imagem da série (Chino et al. 2010). A seguir, o *SatImagExplorer* extrapola a indicação de região para todas as imagens da sequência, gerando séries temporais das ROIs correspondentes à indicada para todas as imagens disponíveis. Essa ferramenta permite que o especialista possa focar sua análise em pontos estratégicos e de interesse, além de facilitar a análise de longas séries de dados. A Figura 3 apresenta uma tela do sistema *SatImagExplorer* com a máscara para a região de estudo e o gráfico da série temporal de valores NDVI extraídos.

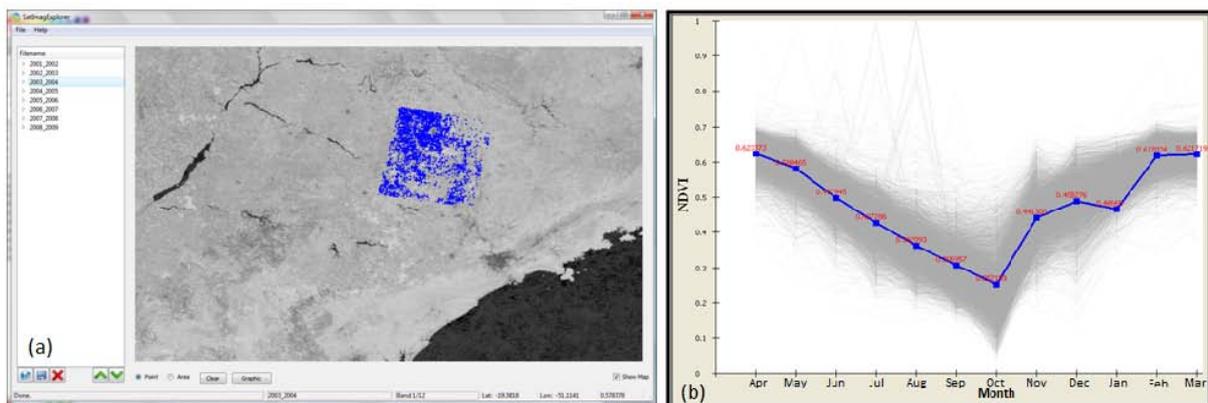


Figura 3. Telas do software *SatImagExplorer* utilizado para extração automática de séries temporais de imagens de satélite, ilustrando em (a) seleção dos pixels por arquivo de coordenadas e em (b) gráfico dos valores NDVI extraídos para todos os pixels da máscara (adaptado de Chino et al., 2010).

Sobre os valores de NDVI extraídos foi aplicada a técnica de agrupamento de dados utilizando o algoritmo *K-Means* (Han et al. 2001). O *K-Means* é um algoritmo baseado em particionamento em que, após a obtenção do agrupamento (classe), cada elemento da série

pertence a uma, e apenas uma classe. Com base em critérios de similaridade, os elementos são atribuídos às classes de modo que, nas séries de NDVI, as áreas agrupadas em uma mesma classe são mais semelhantes se comparadas a áreas de outras classes.

Nos experimentos, foi utilizada a implementação do *K-Means* disponível no Weka (Hall et al. 2009), uma ferramenta que dispõe de várias técnicas de mineração de dados. Foram realizados os testes do algoritmo para diferentes números de classes, além de pré-processamentos nas séries como, por exemplo, remoção de valores considerados *outliers*, para amenizar os ruídos nos resultados.

A ferramenta Weka fornece o recurso de exibir, ao final da execução do algoritmo, a qual classe cada área foi atribuída, assim como as imagens dessa atribuição, o que é muito útil, dado que se trata da análise de pontos num espaço geográfico. Os resultados obtidos com o *K-Means* são discutidos na próxima seção.

3. Resultados e Discussão

Analisando as imagens MVC de NDVI na região nordeste de São Paulo pode-se perceber a evolução do crescimento vegetativo da cana-de-açúcar (Figura 4). O plantio inicia-se em agosto apresentando imagens com tons de verde e azul na região nordeste do estado. Essas cores representam o NDVI baixo (em torno de 0,2) caracterizando regiões com solo exposto e pouca vegetação. Isso vale também para os meses de setembro a novembro. A partir de dezembro, quando a cana-de-açúcar começa a se desenvolver e adquirir mais biomassa, essas regiões apresentam tons amarelo, laranja e vermelho. Os meses de fevereiro a maio apresentam um tom vermelho mais forte, período em que a cana-de-açúcar está no maior estágio de desenvolvimento e onde o NDVI é máximo (valores entre 0,7 e 0,8). As áreas escuras nas imagens representam pixels contaminados por nuvens e água.

A partir das imagens MVC de NDVI, foram gerados os perfis temporais de NDVI e extraídos seus valores mensais. Sobre esses valores foi aplicado o algoritmo *K-Means* considerando duas classes. A Figura 4 ilustra as imagens MVC e as classes resultantes do processo de agrupamento na região de interesse, onde a classe 1 em azul representa o NDVI baixo e médio, e a classe 2 em vermelho indica NDVI mais alto. É possível observar que nos meses de dezembro a maio a classe 2 é predominante, correspondendo ao NDVI alto que nas imagens MVC também aparecem em vermelho. Já nos meses de junho a novembro, quando o NDVI é baixo (tons de verde nas imagens MVC), a classe 1 é predominante. O agrupamento dos valores de NDVI coincide, portanto, com a evolução do crescimento vegetativo da cana-de-açúcar na área de estudo apresentada nas imagens MVC.

A Tabela 1 apresenta em mais detalhes os intervalos mensais de NDVI para cada classe obtidos com o *K-Means*. Como pode ser observado, os valores destacados em azul correspondem a intervalos de NDVI baixo e médio, entre 0,2 e 0,6. Os intervalos superiores, entre 0,6 e 0,8, estão relacionados com valores mais altos de NDVI e aparecem destacados em vermelho. Nos meses de julho a novembro, os valores para ambas as classes são mais baixos, correspondendo ao período em que a cana está sendo colhida. No entanto, como mostrado na Figura 4, mesmo nesses meses não aparece apenas a classe 1, o que indica que existe cana de ano e cana de ano e meio. Assim, embora o NOAA seja um satélite de baixa resolução, foi possível, por meio da técnica de agrupamento, perceber que na região estudada existem plantações de cana com diferentes ciclos de desenvolvimento, o que é uma alternativa encontrada para manter a produção de etanol durante o ano todo.

Com isso, é possível utilizar a técnica de agrupamento de dados para acompanhar o desenvolvimento da cultura sem auxílio de mapeamento dos talhões. Além disso, é possível obter uma classificação mensal dos valores de NDVI por região, mesmo havendo algumas regiões com datas de plantio e de corte diferentes, isto é, ciclos vegetativos diferentes (de ano, e de ano e meio). Essa diferença fica evidente ao analisar a distribuição no mapa das classes ao longo dos meses, pois em nenhum mês existem apenas áreas de uma única classe. Isso

confirma o atraso no corte em algumas regiões para garantir a colheita por um período mais longo, que equivale a quase o ano todo.

Tabela 1. Intervalos mensais dos agrupamentos de NDVI (safra 2002/2003).

Mês	Classe	Valor mínimo de NDVI	Valor médio de NDVI	Valor máximo de NDVI
Abril	1	0,2667	0,5032	0,5473
	2	0,5475	0,5918	0,7431
Maio	1	0,2066	0,4740	0,5238
	2	0,5243	0,5743	0,7641
Junho	1	0,2136	0,4265	0,4730
	2	0,4731	0,5195	0,7409
Julho	1	0,2023	0,3288	0,3828
	2	0,3829	0,4369	0,6829
Agosto	1	0,2022	0,3078	0,3623
	2	0,3625	0,4171	0,6961
Setembro	1	0,2020	0,3012	0,3642
	2	0,3644	0,4271	0,6935
Outubro	1	0,2014	0,2838	0,3378
	2	0,3380	0,3921	0,6226
Novembro	1	0,2016	0,3272	0,3948
	2	0,3950	0,4627	0,6835
Dezembro	1	0,2441	0,5255	0,5765
	2	0,5768	0,6276	0,7654
Janeiro	1	0,2165	0,4978	0,5547
	2	0,5556	0,6130	0,7966
Fevereiro	1	0,2271	0,5851	0,6233
	2	0,6234	0,6615	0,7778
Março	1	0,2113	0,5461	0,5837
	2	0,5837	0,6213	0,7346

4. Conclusão

Este trabalho apresentou a análise de agrupamento de dados utilizada para auxiliar no monitoramento de safras agrícolas de cana-de-açúcar no estado de São Paulo por meio de imagens de NDVI obtidas do sensor AVHRR dos satélites NOAA. O algoritmo *K-Means* mostrou que ao serem gerados agrupamentos para duas classes já é possível identificar variação de valores de NDVI de níveis mais altos nos períodos de dezembro a maio para valores mais baixos nos demais meses.

Além disso, ao usar satélites de baixa resolução, é mais difícil evidenciar a diferença entre canas de ano e de ano e meio. Mas neste estudo realizado com a técnica de agrupamento de dados, mesmo ocorrendo mistura espectral, foi possível acompanhar a evolução da cultura ao longo da safra identificando regiões com padrões semelhantes. Observando os resultados, é possível perceber que apesar do predomínio de uma classe em relação a outra durante os meses do ano, em nenhum momento existe apenas uma delas. Isso indica a diferença de períodos de corte da cana, o que garante a colheita ao longo do ano todo, fator importante para manter o fornecimento de derivados da cana, especialmente o etanol.

Os valores de NDVI para as duas classes possuem uma variação ao longo dos meses, indicando em quais meses a cana atinge seus valores máximos e mínimos. Essa classificação mensal pode ser usada como parâmetro de entrada para outros algoritmos de mineração a fim de permitir a extração de padrões e conhecimentos novos a partir de imagens de baixa resolução, como é o caso das imagens NOAA.

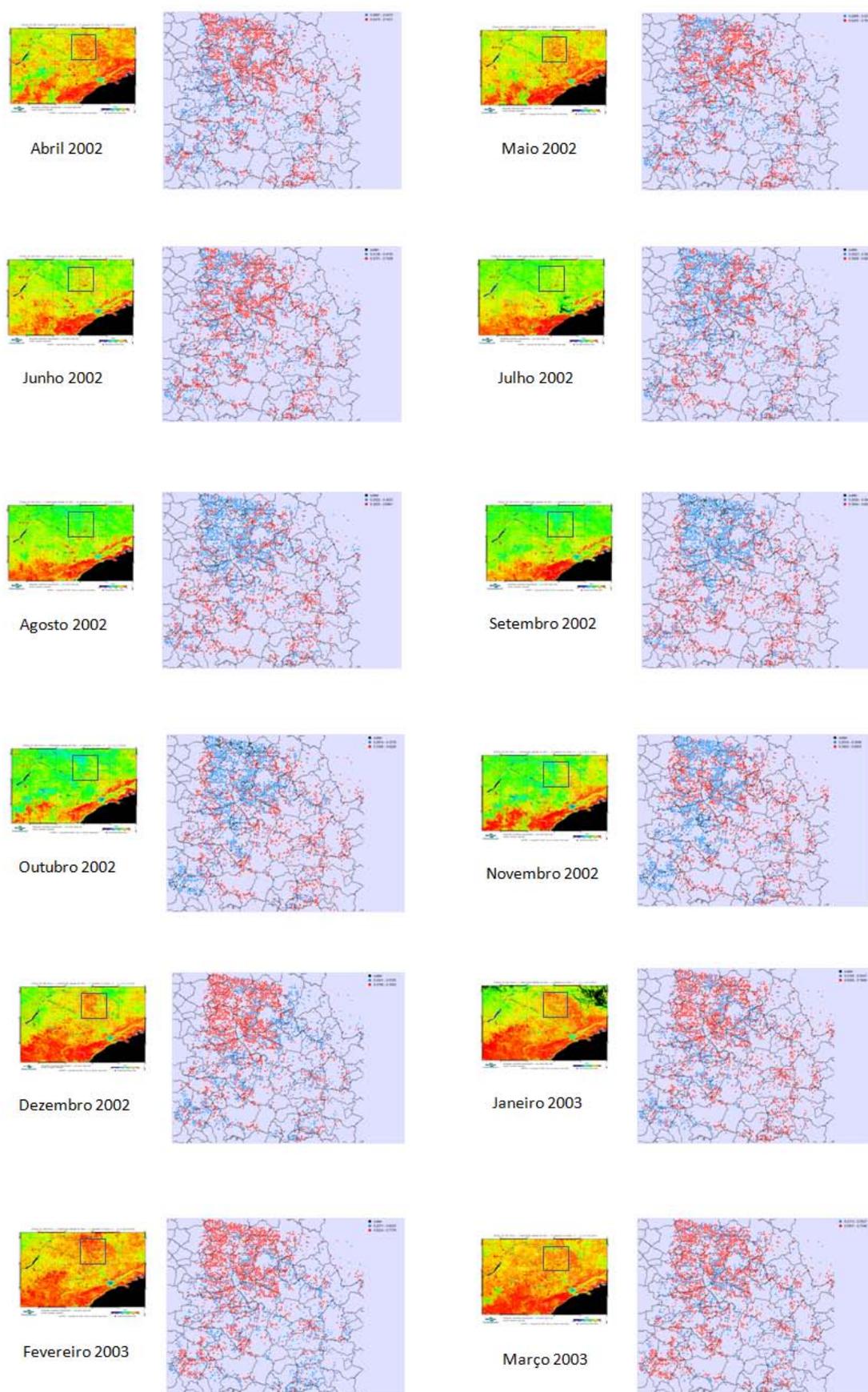


Figura 4. Imagens MVC mensal de NDVI e agrupamento de NDVI (2 classes) da área de estudo para os meses de Abril de 2002 a Março de 2003.

Agradecimentos

Agradecemos as agências Fapesp, CNPq, Capes, Microsoft Research, Santander e Embrapa pelo apoio financeiro e ao Cepagri/Unicamp pela base de imagens de satélite AVHRR/NOAA.

Referências Bibliográficas

Chino, D. Y. T.; Romani, L. A. S.; Traina, A. J. M. Construindo séries temporais de imagens de satélite para sumarização de dados climáticos e monitoramento de safras agrícolas. **Revista Eletrônica de Iniciação Científica**, v.10, p.1-16, 2010.

Esquerdo, J. C. D. M.; Antunes, J. F. G.; Baldwin, D. G.; Emery, W. J.; Zullo Jr, J. An automatic system for AVHRR land surface product generation. **International Journal of Remote Sensing**, v.27, p.3925-3942, 2006.

Goldemberg, J.; Coelho, S. T.; Guardabassi, P. The sustainability of ethanol production from sugarcane. **Energy Policy**, v. 36, p. 2086-2097, 2008.

Gonçalves, R. R. V.; Nascimento, C. R.; Zullo Jr., J.; Romani, L. A. S. Relationship between the spectral response of sugar cane, based on AVHRR/NOAA satellite images, and the climate condition, in the state of sao paulo (Brazil), from 2001 to 2008. In: International Workshop on the Analysis of Multi-temporal Remote Sensing images (MultiTemp), 5, 2009, Groton. **Proceedings...** 2009, p. 315-322.

Hall, M.; Frank, E.; Holmes, G.; Pfahringer, B.; Reutemann, P.; Witten, I. H. The WEKA Data Mining Software: An Update; **SIGKDD Explorations**, v.11, i.1, 2009.

Han, J.; Kamber, M. **Data Mining - Concepts and Techniques**. New York: Morgan Kaufmann Publishers, 2001, 550p.

Han, J.; Kamber, M.; Tung, A. K. H. Spatial Clustering Methods in Data Mining: A Survey. In: **Geographic Data Mining and Knowledge Discovery**, H. J. Miller and J. Han (Editors), p. 201-230, Taylor and Francis, 2001.

Martinelli, L. A.; Filoso, S. Expansion of sugarcane ethanol production in Brazil: environmental and social challenges. **Ecological Applications**, v. 18, n.4, p.885-898, 2008.

Nascimento, C. R.; Gonçalves, R. R. V.; Zullo Jr., J.; Romani, L. A. S. Estimation of sugar cane productivity using a time series of AVHRR/NOAA-17 images and a phenology-spectral model. In: International Workshop on the Analysis of Multi-temporal Remote Sensing images (MultiTemp), 5, 2009, Groton. **Proceedings...** 2009, p. 365-372.

Nass, L. L.; Pereira, P. A. A.; Ellis, D. Biofuels in Brazil: an overview. **Crop Science**, v. 47, p. 2228-2237, 2007.

Romani, L. A. S.; Sousa, E. P. M.; Traina Jr., C.; Zullo Jr., J.; Traina, A. J. M. Aplicação de Método Baseado em Fractais para Detecção de Correlações entre Imagens AVHRR-NOAA e Dados Climáticos para Regiões Produtoras de Cana-de-açúcar. In: Simpósio Brasileiro de Sensoriamento Remoto (SBSR), 2009, Natal. **Anais...** Editora do INPE, 2009, p. 403-410.

Romani, L. A. S.; Traina, A. J. M.; Ribeiro, M. X.; Sousa, E. P. M.; Zullo Jr., J.; Traina Jr., C. Aplicação de técnicas de mineração em dados climáticos e de satélite para auxiliar no acompanhamento das safras de cana-de-açúcar. In: Workshop em Algoritmos e Aplicações de Mineração de Dados (WAAMD), 4, 2008, Campinas. **Anais...** SBC, 2008, p. 87-92.

Rudorff, B. F. T.; Adami, M.; Aguiar, D. A. d.; Gusso, A.; Silva, W. F. d.; Freitas, R. M. d. Temporal series of evi/modis to identify land converted to sugarcane. In: IEEE International Geoscience and Remote Sensing Symposium (IGARSS), 2009, Cape Town. **Proceedings...** IEEE, 2009, p. 252-255.

Rudorff, B. F. T.; Aguiar, D. A. d.; Silva, W. F. d.; Sugawara, L. M.; Adami, M.; Moreira, M. A. Studies on the rapid expansion of sugarcane for ethanol production in São Paulo state (Brazil) using Landsat data. **Remote sensing**, v. 2, p. 1057-1076, 2010.

Scarpari, M. S.; Beauclair, E. G. F. Sugarcane maturity estimation through edaphicclimatic parameters. **Scientia Agricola**, v.13, n.8, p.8, 2004.