

Mapeamento da vegetação por meio de séries de imagens NDVI usando mineração de dados: comparação dos sensores AVHRR/NOAA e MODIS/Terra

Rachel Scrivani^{1,2}
Renata Ribeiro do Valle Gonçalves²
Jurandir Zullo Jr.^{1,2}
Luciana Alvim Santos Romani³

¹ Universidade Estadual de Campinas – Faculdade de Engenharia Agrícola – Feagri
Av. Cândido Rondon, 501 - Barão Geraldo
13083-875 - Campinas - SP, Brasil
rachelscrivani@gmail.com

² Universidade Estadual de Campinas - Centro de Pesquisas Meteorológicas e Climáticas
Aplicadas a Agricultura – Cepagri
Cidade Universitária "Zeferino Vaz"
13083-970 - - Campinas – SP, Brasil
renata@cpa.unicamp.br, jurandir@cpa.unicamp.br

³ Embrapa Informática Agropecuária
Av. André Toselo, 209 - Caixa Postal 6041
13083-886 - Campinas - SP, Brasil
luciana.romani@embrapa.br

Abstract. In recent years, due to an increase in the number of images available from high temporal resolution satellites such as MODIS / Terra and AVHRR / NOAA allow the generation of satellite image time series (SITS) which are widely used in different areas. An important application of SITS is the mapping and monitoring of agricultural crops such as sugarcane. This crop is configured as an important Brazilian commodity with economic impact for the country. As this crop is produced in large fields, we can apply SITS of low spatial resolution to aid at monitoring and evaluating its growing. In this context, this article presents an approach for automatic mapping of use and land cover using sensors with low spatial resolution and high temporal resolution by comparing results obtained from NDVI series of MODIS and AVHRR sensors. For this purpose we have applied time series mining on NDVI series, specifically the clustering method, analyzing comparatively series by cluster of MODIS and AVHRR sensors through the linear regression method. Results of applying the proposed method to three municipalities with high production of sugarcane in São Paulo - Brazil have showed that both sensors have produced groups representing agricultural areas (especially cane sugar), dense vegetation, water and urban areas.

Palavras-chave: remote sensing, time series, regression analysis, clustering, sugarcane, sensoriamento remoto, séries temporais, análise de regressão, agrupamento, cana-de-açúcar.

1. Introdução

Segundo o Ministério da Agricultura Pecuária e Abastecimento, o agronegócio brasileiro é moderno, eficiente e competitivo (MAPA, 2014). O Brasil é um dos líderes mundiais na produção e exportação de vários produtos agropecuários. A produção nacional de cana-de-açúcar estimada para 2014 é de 737 milhões de toneladas sendo que a região Sudeste concentra 65,9% da produção nacional, onde o estado de São Paulo, o líder da produção nacional, é responsável por 54,9% da produção brasileira (IBGE, 2014). Por ser a cana-de-açúcar uma monocultura que ocupa grandes extensões territoriais e ter características vegetativas uniformes, ela se torna apta ao uso do sensoriamento remoto. Diversos trabalhos mostram este potencial de aplicação (Rudorff et al., 2005; Sugawara et. al.,2009; Picoli et. al., 2009).

A utilização de dados dos sensores orbitais MODIS (*Moderate Resolution Imaging Spectro Radiometer*) a bordo do satélite Terra e o sensor AVHRR (*Advanced Very High Resolution Radiometer*) a bordo do satélite NOAA que fornecem dados de NDVI (*Normalized Difference Vegetation Index*) e tem distribuição pública de suas imagens, têm sido utilizadas em estudos da cultura da cana-de-açúcar (Sugawara et. al.,2009; Gonçalves et. al, 2013). Diversos estudos vêm realizando comparações dos valores de NDVI entre os sensores MODIS e AVHRR (Fensholt e Sandholt, 2004; Ji et al., 2008; Fensholt et.al, 2009; Fensholt e Proud, 2012).

Fensholt e Sandholt (2004) realizaram a comparação do NDVI dos sensores AVHRR/NOAA e MODIS/Terra utilizando métodos de regressão para uma região semiárida no Senegal e concluíram que o NDVI foi melhor emulado pelo MODIS do que pelo AVHRR. Ji et. al (2008) também compararam o NDVI dos sensores AVHRR/NOAA e MODIS/Terra utilizando métodos de regressão e análises de concordância. Neste caso, a área de estudo foi os Estados Unidos e os autores concluíram que a continuidade de séries temporais NDVI de AVHRR para MODIS é satisfatória, mas uma transformação linear entre os dois conjuntos é recomendada. Fensholt et al. (2009) compararam dados de NDVI do AVHRR/GIMMS e MODIS/Terra para a região do Sahel, buscando padrões por meio de análises de consistência, concluindo que o resultado dos padrões encontrados são diferentes para áreas semiáridas e áreas mais úmidas. Fensholt e Proud (2012) analisaram dados de NDVI do AVHRR/GIMMS e MODIS/Terra em escala global, utilizando métodos de regressão linear. Os autores encontraram as maiores diferenças no hemisfério sul (América do Sul e Austrália) e no norte do Ártico, e concluíram que há compatibilidade global entre MODIS e GIMMS NDVI, mas com variações específicas da classe de cobertura do solo (dentro da classe e entre classes).

Neste contexto surgem algumas questões: dados de NDVI dos sensores MODIS e AVHRR podem ser utilizados da mesma forma, obtendo resultados similares para o monitoramento agrícola da cultura da cana-de-açúcar no Brasil? É possível beneficiar-se de imagens de baixa resolução espacial, mas alta resolução temporal para identificar de forma automática a cobertura e uso da terra?

Neste cenário, este artigo propõe o mapeamento automático do uso/cobertura da terra por meio de sensores de baixa resolução espacial e alta resolução temporal comparando os resultados obtidos com uso de séries de NDVI dos sensores MODIS e AVHRR. Neste artigo é apresentada, então, uma abordagem baseada na mineração das séries de NDVI pelo método de agrupamento, analisando de forma comparativa os resultados tanto para as séries provenientes dos sensores MODIS/Terra quanto do AVHRR/NOAA por meio do método de regressão linear. Os resultados obtidos da aplicação do método proposto para três municípios com alta produção de cana-de-açúcar no estado de São Paulo, Brasil mostra que ambos os sensores produziram grupos representativos de áreas agrícolas (especialmente a cana-de-açúcar), vegetação densa, água e área urbana. O restante deste artigo está organizado da seguinte forma: na Seção 2 é apresentado o Material e Métodos, em seguida são discutidos os

resultados e finalmente a Seção 4 conclui o trabalho.

2. Metodologia de Trabalho

A área de estudo definida para este artigo são os municípios de Barretos, Piracicaba e Ribeirão Preto, localizados entre as coordenadas geográficas 20°33'26"S 48°34'04"W, 21°10'42"S 47°48'24"W e 21°10'42"S 47°48'24"W respectivamente. Estes municípios são grandes produtores de cana-de-açúcar no estado de São Paulo, Brasil, conforme dados adaptados do site Canasat na Tabela 1.

| Ano | Barretos – SP | Piracicaba – SP | Ribeirão Preto – SP |
|------|---------------|-----------------|---------------------|
| 2003 | 24843 | 40397 | 30576 |
| 2004 | 29343 | 42940 | 30500 |
| 2005 | 37358 | 46978 | 31098 |
| 2006 | 43394 | 47821 | 32001 |
| 2007 | 52513 | 50767 | 31266 |
| 2008 | 68192 | 51141 | 32272 |
| 2009 | 70670 | 52440 | 32486 |
| 2010 | 72035 | 51238 | 32227 |
| 2011 | 73668 | 50858 | 32080 |
| 2012 | 75490 | 49969 | 31759 |
| 2013 | 76787 | 49866 | 30895 |

Neste trabalho foram utilizados dois conjuntos de imagens. O primeiro conjunto é do sensor MODIS a bordo do satélite Terra, produto MOD13A2, com resolução de 16 dias, composição e resolução espacial de 1km, de 2003 a 2009, disponibilizadas pelo *Earth Observing System* (EOS/NASA) por meio do *Land Processes Distributed Active Archive Center* (LP-DAAC). O segundo conjunto de imagens é do sensor AVHRR a bordo do satélite NOAA-17, disponibilizadas pelo Centro de Pesquisas Meteorológicas e Climáticas Aplicadas à Agricultura, da Universidade Estadual de Campinas (Cepagri/Unicamp). As imagens MODIS adquiridas por meio da ferramenta *MODIS Reprojection Tool Web Interface* (MRTWeb) do LP-DAAC inclui correções radiométrica, geométrica, atmosférica, projeção cartográfica, valores de NDVI e conversão do formato, no recorte de *tiles*, sendo neste trabalho usado o *tile* h13v11. As imagens AVHRR/NOAA, também já processadas disponibilizam as mesmas correções das imagens MODIS e também foi utilizado o recorte do estado de São Paulo.

Há diferenças na formulação dos valores do NDVI entre o MODIS e o AVHRR/NOAA. São elas: um ajuste da influência dos aerossóis aplicado apenas no MODIS (Gallo et. al., 2005), as bandas do vermelho e do NIR do AVHRR são mais amplas do que no MODIS, as imagens AVHRR são redefinidas a partir de dados com uma resolução de 1,1km no nadir, enquanto as imagens MODIS são agregadas a partir dos dados originais com resolução de 250m, diferença no horário da captação da imagem e uma pequena diferença nas técnicas de composição do NDVI (Ji et. al, 2008). Na composição no MODIS há a determinação dos dois maiores valores de NDVI para cada pixel por intervalo composto de 16 dias, sendo o valor NDVI observado como o mais próximo ao estado crítico, o selecionado para inclusão no produto composto. E a técnica de composição no AVHRR simplesmente seleciona o maior valor NDVI observado para cada pixel por intervalo composto de 16 dias (Ji et al., 2008).

A metodologia deste trabalho compreende duas etapas. A primeira etapa consistiu na

extração das séries temporais de NDVI de ambos os sensores gerando um conjunto de 3.173 séries para o AVHRR e 3.701 séries para o MODIS. Em seguida foi realizada a mineração das séries temporais por meio do método de agrupamento, usando os valores de NDVI dos dois conjuntos de imagens com o software SatImagExplorer (Chino et al., 2010). O método de agrupamento utilizado foi o *K-means*, um método de particionamento não supervisionado, bem conhecido que tem o k como parâmetro de entrada e particiona n observações nos k grupos, sendo a semelhança do grupo medida como o valor médio dos objetos no grupo (Hans e Kamber, 2006). A métrica de distância utilizada foi a Euclidiana. O número k definido para este trabalho foi 6, de acordo com a sugestão de especialistas. Tanto a extração das séries temporais como os agrupamentos foram executados com o número total de pixels da imagem. Foram extraídas as séries e realizados os agrupamentos para cada ano de imagem individualmente, de 2003, 2004, 2005 até 2009 e também para o período todo de 2003 à 2009. Para validação do número k foi usado o coeficiente de Silhueta (do inglês, *Silhouette coefficient*), que é um método de interpretação e validação de resultados de agrupamentos que mede o quanto um objeto é similar aos outros objetos de seu grupo e o quanto ele está separado dos outros grupos (Rousseeuw, 1987). O valor de Silhueta varia de -1 a 1. Quanto mais próximo do valor um, melhor é o agrupamento.

A segunda etapa do trabalho consistiu nas análises de regressão linear (Equação 1) para comparar as séries de NDVI dos sensores MODIS e AVHRR. As regressões foram realizadas para os três municípios estudados. Inicialmente foram calculados os valores médios de NDVI para cada grupo (cluster de 0 a 5) gerado pelo método de agrupamento, resultando em 6 séries de NDVI para o MODIS e AVHRR por município. As regressões foram então calculadas para cada par de séries de NDVI a fim de comparar os dois sensores. Todas as regressões foram realizadas com a média mensal do NDVI.

$$\text{Regressão Linear: } y = \alpha x + \beta \quad (1)$$

3. Resultados e Discussão

O k igual a 6 selecionado para este trabalho pela sugestão de especialistas foi considerado adequado após analisados os resultados da avaliação do coeficiente de Silhueta, que apresentou resultados positivos para os grupos gerados, que significa que os grupos estão coesos e bem separados uns dos outros, conforme Tabela 2.

Tabela 2. Valores gerados pelo método Silhueta

| 2003 a 2009 | Silhueta | | | | | |
|-------------|----------|------|------------|------|----------------|------|
| | Barretos | | Piracicaba | | Ribeirão Preto | |
| | MODIS | NOAA | MODIS | NOAA | MODIS | NOAA |
| 0 | 0,37 | 0,19 | 0,41 | 0,35 | 0,64 | 0,54 |
| 1 | 0,1 | 0,08 | 0,4 | 0,18 | 0,39 | 0,36 |
| 2 | -0,03 | 0,12 | 0,11 | 0,19 | 0,1 | 0,22 |
| 3 | 0,06 | 0,03 | 0,11 | 0,21 | 0,18 | 0,08 |
| 4 | 0,24 | 0,16 | 0,35 | 0,29 | 0,13 | 0,18 |
| 5 | 0,31 | 0,25 | 0,26 | 0,11 | 0,37 | 0,19 |

Os resultados obtidos dos agrupamentos de NDVI para os sensores MODIS/Terra e AVHRR/NOAA, tanto anual, quanto para o período todo analisado (2003 a 2009), apresentaram padrão espacial e temporal similar para os três municípios de estudo. Os valores de NDVI do sensor MODIS são sempre maiores para todos os grupos, porém a distribuição,

quantidade de pixels por grupo, tem valores muito próximos como pode ser visto na Figura 1.

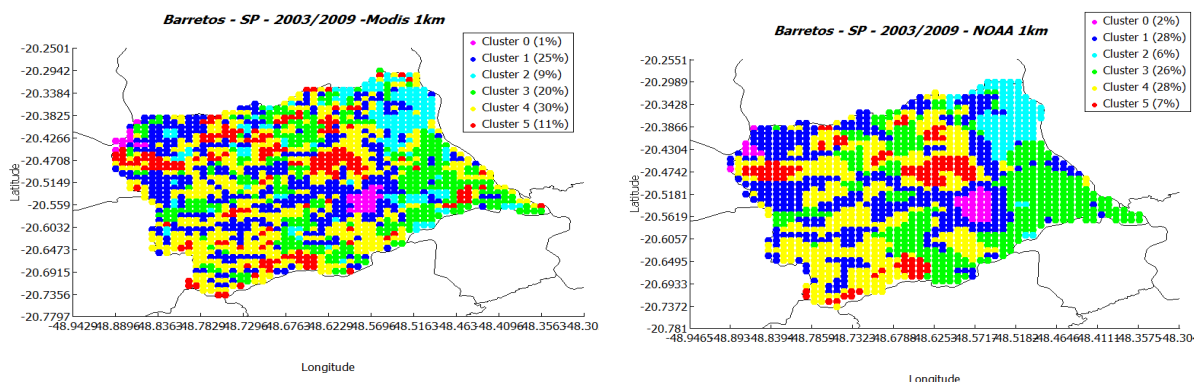


Figura 1. Mapa dos agrupamentos, com a distribuição dos valores de NDVI para cada grupo, obtidos dos dois sensores (MODIS e AVHRR), para o período de 2003 à 2009 para o município de Barretos.

Os grupos de cor lilás sempre tem o menor valor do índice de vegetação. Para o MODIS, os valores dos centroides deste grupo variam entre 0,27 e 0,33 e para o AVHRR os valores dos centroides deste grupo variam entre 0,23 e 0,31, indicando que esta região é de área urbana ou água. Por outro lado, o grupo vermelho apresenta os mais altos valores de NDVI. Para o MODIS, os valores dos centroides deste grupo variam entre 0,68 e 0,69 e para o AVHRR os valores dos centroides deste grupo variam entre 0,50 e 0,52, indicando que esta região é de vegetação densa. O grupo de tom azul escuro, com valores de NDVI para MODIS entre 0,46 e 0,55 e para NOAA entre 0,35 e 0,44 indica vegetação de transição, entorno de área urbana e entorno de corpos d’água. Os grupos de tom azul claro e verde estão num intervalo de valores de 0,56 a 0,60 para o MODIS e de 0,42 a 0,47 para o AVHRR, indicando área de cultivo agrícola, nestes municípios predominantemente cana-de-açúcar. O grupo amarelo indica a transição de vegetação de cultivo de cana-de-açúcar para vegetação densa, com o valor médio de 0,61 para o MODIS e entre os valores 0,47 e 0,48 para o AVHRR; como mostrado nas Figuras 2 a 5.

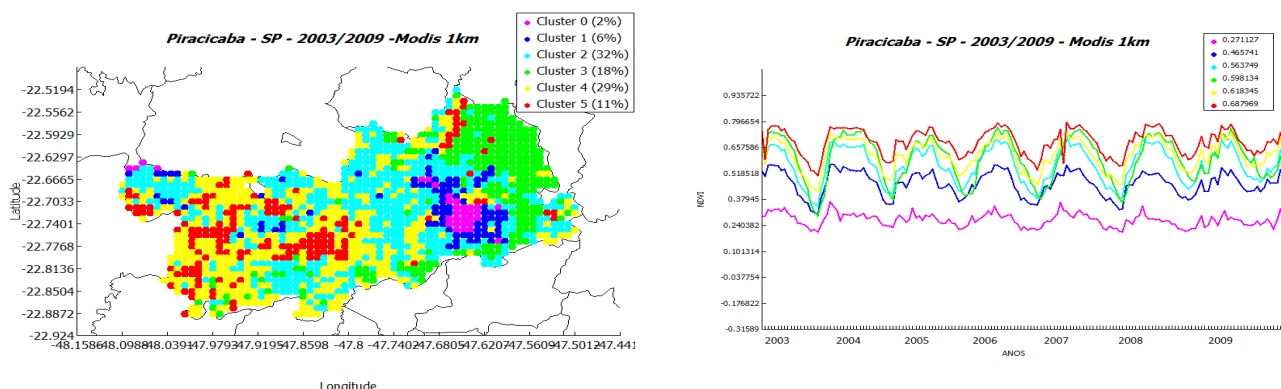


Figura 2. Mapa dos agrupamentos e gráfico dos centroides resultantes da análise para as séries do MODIS, no período de 2003 a 2009, para o município de Piracicaba.

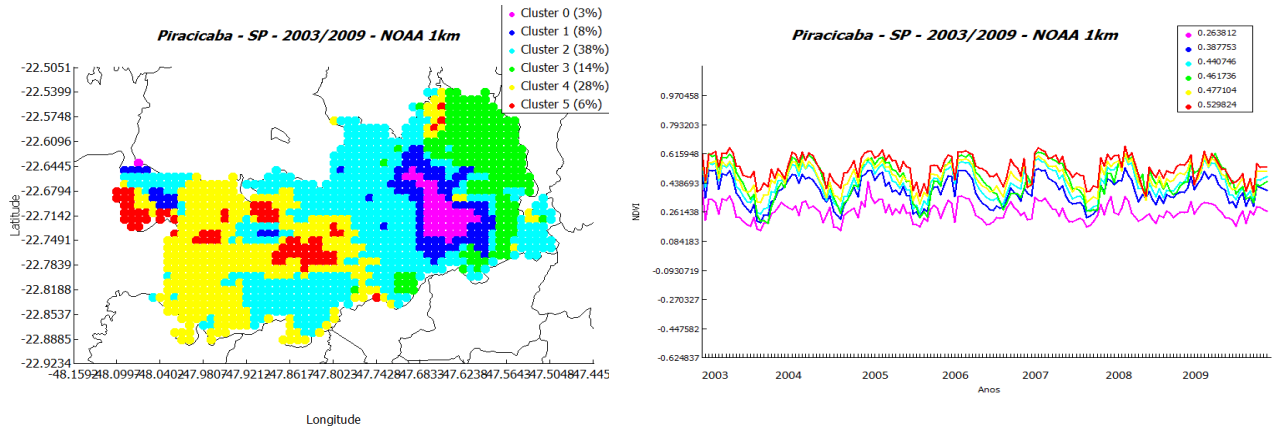


Figura 3. Mapa dos agrupamentos e gráfico dos centroides resultantes da análise para as séries do NOAA, no período de 2003 a 2009, para o município de Piracicaba.

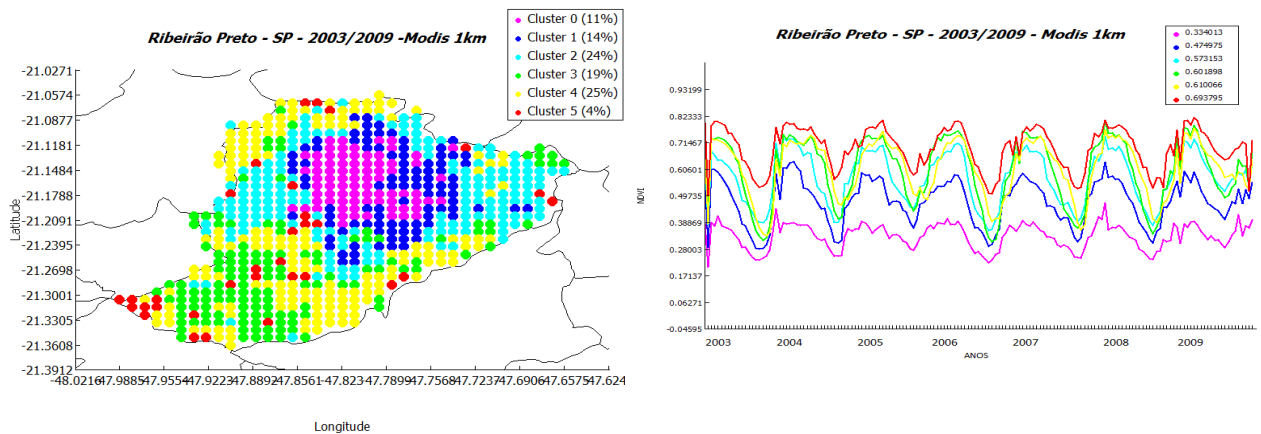


Figura 4. Mapa dos agrupamentos e gráfico dos centroides resultantes da análise para as séries do MODIS, no período de 2003 a 2009, para o município de Ribeirão Preto.

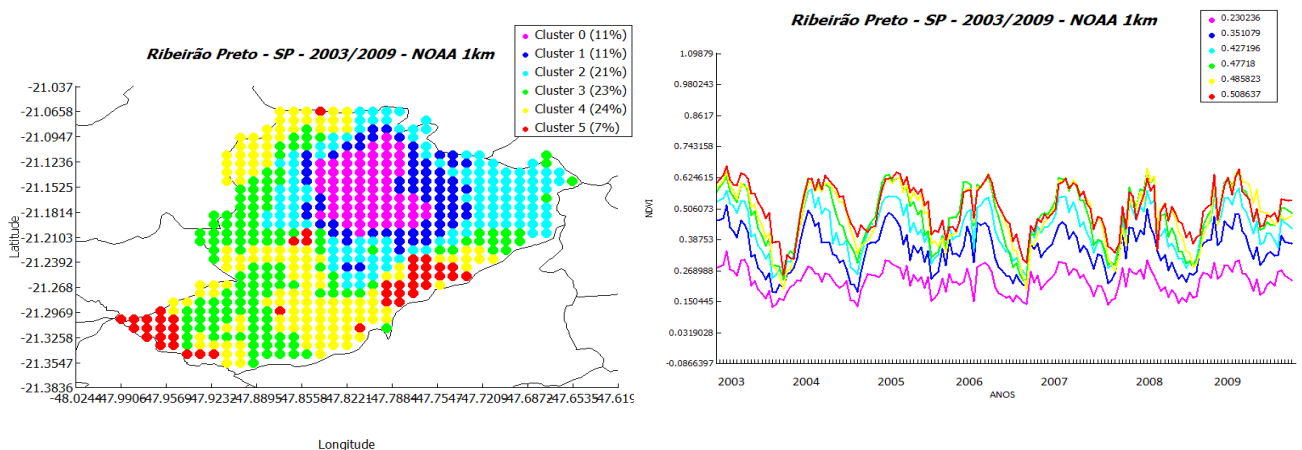


Figura 5. Mapa dos agrupamentos e gráfico dos centroides resultantes da análise para as séries do NOAA, no período de 2003 a 2009, para o município de Ribeirão Preto.

As regressões realizadas com as séries de valores médios de NDVI por grupo do MODIS e do AVHRR apresentaram resultado significativo (r^2 acima de 0,7) para os três municípios nos grupos 1 a 4, onde estão localizadas as áreas de cultivo agrícola, as áreas de cana-de-açúcar (grupos 2 e 3) e transição de vegetação (grupos 1 e 4). Para as demais áreas, como área

urbana e água (grupo 1) e vegetação densa (grupo 5), os resultados foram menos significativos, principalmente devido à mistura de alvos que pode ser encontrada num mesmo pixel nestas regiões, conforme Tabela 3.

Tabela 3. Resultado dos valores de NDVI e da regressão linear entre os valores de NDVI do MODIS e NOAA para o período de 2003 a 2009

| Grupos | NDVI | | | | | | r ² | | |
|--------|----------|------------|----------------|----------|------------|----------------|----------------|------------|----------------|
| | MODIS | | | NOAA | | | Barretos | Piracicaba | Ribeirão Preto |
| | Barretos | Piracicaba | Ribeirão Preto | Barretos | Piracicaba | Ribeirão Preto | | | |
| 0 | 0,32 | 0,27 | 0,33 | 0,31 | 0,26 | 0,23 | 0,33 | 0,38 | 0,55 |
| 1 | 0,55 | 0,46 | 0,47 | 0,442 | 0,38 | 0,35 | 0,69 | 0,67 | 0,73 |
| 2 | 0,57 | 0,56 | 0,57 | 0,449 | 0,44 | 0,42 | 0,78 | 0,74 | 0,79 |
| 3 | 0,6 | 0,59 | 0,6 | 0,473 | 0,46 | 0,47 | 0,73 | 0,8 | 0,79 |
| 4 | 0,61 | 0,61 | 0,61 | 0,475 | 0,47 | 0,48 | 0,66 | 0,65 | 0,68 |
| 5 | 0,68 | 0,68 | 0,69 | 0,52 | 0,52 | 0,5 | 0,57 | 0,51 | 0,48 |

A Figura 6 apresenta os gráficos do município de Ribeirão Preto com as funções de regressão e o r² para cada grupo de 0 a 5, como um exemplo dos resultados obtidos neste trabalho. Como pode ser visto na Figura 6, os melhores resultados estão nos agrupamentos 2 e 3 em que são encontradas áreas de cultivo da cana-de-açúcar. Desta forma, comprova-se que séries de NDVI tanto do sensor MODIS quanto do AVHRR podem ser usadas para um mapeamento do uso e cobertura da terra (pré-classificação) em escala regional.

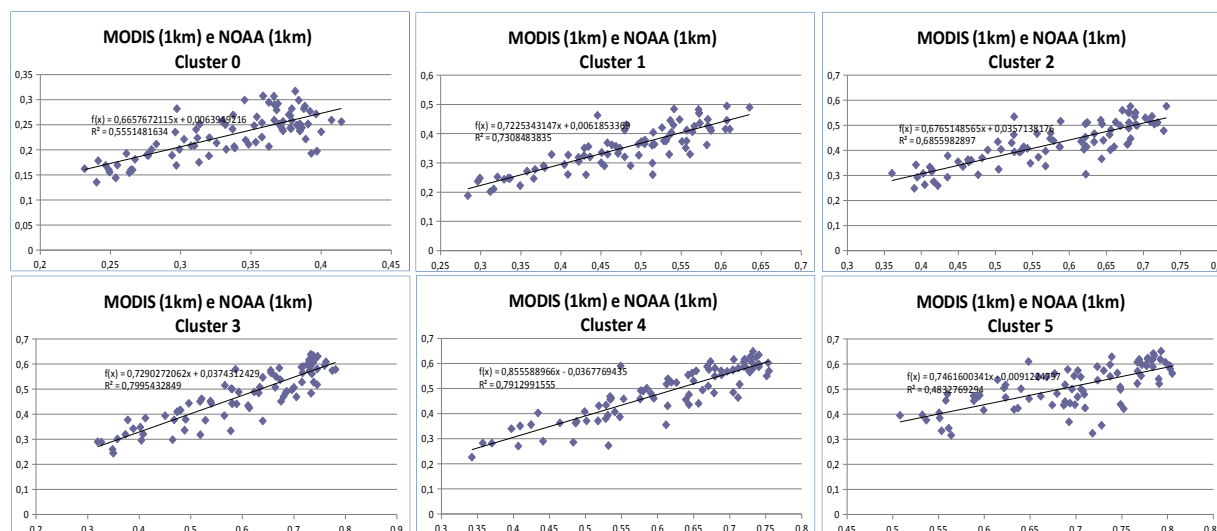


Figura 6. Gráficos resultantes de regressão linear entre os dados das séries de NDVI do MODIS e do AVHRR para o período de 2003 a 2009 para o município de Ribeirão Preto.

4. Conclusões

Neste trabalho foi apresentada uma abordagem para realizar de forma automática o mapeamento do uso/cobertura da terra por meio de sensores de baixa resolução espacial e alta resolução temporal comparando os resultados obtidos com uso de séries de NDVI dos sensores MODIS e AVHRR. A principal contribuição apresentada neste trabalho é a comprovação de que a mineração das séries de NDVI dos sensores MODIS/Terra como do AVHRR/NOAA, de baixa resolução espacial e alta resolução temporal, pelo método de agrupamento, produzem grupos representativos de áreas agrícolas (especialmente a cana-de-açúcar), vegetação densa,

água e área urbana. A partir destes resultados entende-se que o uso de qualquer um dos dois sensores trará resultados compatíveis que poderão ser usados um em substituição ao outro e/ou um como complementar ao outro no mapeamento do uso/cobertura da terra principalmente de áreas agrícolas.

Agradecimentos

Os autores agradecem o financiamento das instituições: CNPq, FAPESP, Embrapa e Cepagri.

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. In: **REIC**, v. 10, p. 1-16. 2010.

Gallo, K; Ji L; Reed, B.; Eidenshink, J. C.; Dwyer J..Multi-platform comparisons of MODIS and AVHRR normalized difference vegetation index data. **Remote Sensing of Environment**, 99, pp. 221 – 231, 2005.

Gonçalves, R. R. V.; Zullo Junior, J.; Marquezini, O.; Amaral, B. F., Souza, E. P. M., Romani, L. A. S..Análise da relação entre os perfis de NDVI obtidos dos sensores AVHRR/NOAA e MODIS nas áreas produtoras de cana-de-açúcar em São Paulo. **Anais XVI Simpósio Brasileiro de Sensoriamento Remoto - SBSR**, Foz do Iguaçu, PR, Brasil, 13 a 18 de abril de 2013, INPE, p. 640-647.

Han, J and Kamber, M. **Data Mining: Concepts and Techniques**, 2nd edition, Morgan Kaufmann. 2006.

IBGE. **Produção Agrícola 2014**. Disponível em:

<ftp://ftp.ibge.gov.br/Producao_Agricola/Levantamento_Sistematico_da_Producao_Agricola_%5Bmensal%5D/Comentarios/lspa_201404comentarios.pdf>. Acessado em 31 de outubro de 2014.

Fensholt, R and Sandholt, I.. Evaluation of MODIS and NOAA AVHRR vegetation indices with in situ measurements in a semi-arid environment. **International Journal of Remote Sensing**. Vol. 26, No. 12,pp. 2561-2594, 2005.

Fensholt, R; Rasmussen, K; Nielsen, T. T.; Mbow, C.. Evaluation of earth observation based long term vegetation trends — Intercomparing NDVI time series trend analysis consistency of Sahel from AVHRR GIMMS, Terra MODIS and SPOT VGT data. **Remote Sensing of Environment**. 113, pp. 1886–1898, 2009.

Fensholt, R. and Proud, S. R.. Evaluation of Earth Observation based global long term vegetation trends — Comparing GIMMS and MODIS global NDVI time series. **Remote Sensing of Environment**. 119, pp. 131-147, 2012.

Ji L.; Gallo, K.; Eidenshink, J. C.; Dwyer J.. Agreement evaluation of AVHRR and MODIS 16-day composite NDVI data sets. **International Journal of Remote Sensing**. Vol. 29, n°16, pp. 4839-4861, 2008.

MAPA. Ministério da Agricultura, Pecuária e Abastecimento. **Estatísticas**. Disponível em: <<http://www.agricultura.gov.br/vegetal/estatisticas>>. Acessado em 31 de outubro de 2014.

Picoli, M. C. A.; Rudorff, B. F. T.; Rizzi, R.; Giarolla, A.. Índice de vegetação do sensor MODIS na estimativa da produtividade agrícola da cana-de-açúcar. **Bragantia**, Campinas, v. 68, n. 3, 2009.

Rousseeuw, P. J. **Silhouettes: a Graphical Aid to the Interpretation and Validation of Cluster Analysis**. *Computational and Applied Mathematics*, v. 20, 1987. p. 53–65.

Rudorff, B.F.T. et al. Imagens de satélite no mapeamento e estimativa de área de cana-de-açúcar em São Paulo: ano-safra 2003/04. **Agricultura em São Paulo**, v.52, n.1, p.21-39, 2005.

Sugawara, L. M.; Adami, M.; Rudorff, B. F. T.; Faria, V. G. C.. Avaliação de três métodos de estimativa de índice de área foliar aplicados à cana-de-açúcar. **Anais XIV Simpósio Brasileiro de Sensoriamento Remoto - SBSR**, Natal, RN, Brasil, 25 a 30 de abril de 2009, INPE, p. 499-506.