

## Mapeamento de características de solos tropicais utilizando Self-Organizing Map aplicado à dados hiperespectrais

Luiz Eduardo Vicente <sup>1</sup>

Michael J. Friedel <sup>2</sup>

Fabio Iwashita <sup>3</sup>

Andrea Koga-Vicente <sup>4</sup>

<sup>1</sup> Embrapa Monitoramento por Satélite  
CEP 13070-115- Campinas-SP, Brasil  
vicente@cnpm.embrapa.br

<sup>2</sup> United States Geological Survey (USGS)  
Box 25046 - Lakewood - CO, EUA  
mfriedel@usgs.gov

<sup>3</sup> Griffith University – Australian Rivers Institute  
170 – Kessels Road – Brisbane, AU  
budafbi@gmail.com

<sup>4</sup> Centro de Pesquisas Meteorológicas e Climáticas Aplicadas a Agricultura  
(CEPAGRI)/UNICAMP  
Caixa Postal 13083-970 - Campinas - SP, Brasil  
andrea.kvicente@gmail.com

**Abstract.** The sensor Hyperion was launched in late 2000 and represents an advance to hyperspectral remote sensing. However, due from technical aspects has always presented a large amount of nonlinear noise among the bands. Our goal was apply self-organizing map (SOM) technic on Hyperion data without noise correction for mapping tropical soil characteristics. For these purposes, spectral signatures (0.4-2.5 $\mu$ m) of tropical soils were used as input data. We mapped Photosynthetically Active Vegetation as background of Hyperion scene, Bare Soil and Clayey Soil, with range of  $R^2$  between 0.90-0.93. SOM was efficient in both visualization and analysis procedures, and in the ability identifying nonlinear topological relation in order to classify targets, to correct noise and clustering spectral soil characteristics.

**Palavras-chave:** hyperspectral remote sensing, Self-Organizing Map, Reflectance Spectroscopy, Sensoriamento Remoto hiperespectral, Self-Organizing Map, Espectroscopia de Reflectância.

### 1. Introdução

O sensor Hyperion foi lançado no final de 2000 e opera embarcado numa órbita de 705 km de altitude na plataforma *Earth Observing - 1* da NASA (*National Aeronautics and Space Administration*), e até hoje continua sendo o único sensor hiperespectral orbital em operação, o que o tornou a mais democrática e abrangente fonte de dados hiperespectrais, mesmo que ainda em caráter experimental. Por esse motivo, mesmo após o término do período planejado de um ano de imageamento houve grande interesse por parte da comunidade de usuários e pesquisadores na continuidade do programa EO-1, o que ocorre até hoje através de um acordo de suporte com a USGS (*United States Geological Survey*). As bandas do Hyperion possuem largura de 10 nm totalizando 242 bandas no intervalo espectral do VNIR (*Visible and Near Infrared*)/SWIR (*Shortwave Infrared*) (0.356 – 0.2577 $\mu$ m), com resolução espacial de 30 m e uma faixa de imageamento de 20 x 7,5 Km (Jupp et al., 2004)

A busca por uma resolução espacial robusta aliada ao registro da radiação eletromagnética (REM) num espectro contínuo em centenas de comprimentos de onda é implementada no Hyperion pelo sistema de imageamento de varredura (*pushbroom*). Porém, apesar de representar um avanço significativo na aquisição de dados hiperespectrais, esse sistema ainda apresenta entraves oriundos da capacidade limitada dos seus detectores em linha de registrar mínimas variações da REM considerando a relação tempo de imageamento vs área (*Ground Instantaneous Field of View*) de exposição dos alvos na superfície, sobretudo quando relacionados ao mapeamento de alvos de relativo baixo albedo como solos, notoriamente um alvo que apresenta mais ruído, principalmente em condições de alta umidade relacionados a presença de solos argilosos (Tits, 2013), com ênfase em regiões de baixa energia (Radiação Eletromagnética – REM – segundo a curva de espectro solar de irradiação) como o SWIR. Tais premissas tornam o Hyperion o sensor ideal no que tange aos objetivos desse trabalho, voltado para a aplicação de procedimentos de espectroscopia de reflectância na medida da utilização de bibliotecas espectrais, associadas a técnicas de soft computing com ênfase no algoritmo SOM (*Self-Organizing Map*) para o mapeamento de características de solos. Espera-se contribuir na avaliação desse método de classificação apto a identificar relações não lineares em conjuntos de dados complexos relativos a sensores hiperespectrais.

## 2. Material e método

A área de estudo situa-se no município de Ibiraci, porção nordeste do estado de São Paulo, no quadrilátero formado pelas coordenadas de: latitude  $-20^{\circ} 19' 08''$ , longitude  $-47^{\circ} 08' 23''$  e latitude  $-20^{\circ} 32' 35''$ , longitude  $-47^{\circ} 06' 37''$ . O local foi escolhido por apresentar condições de extrema diversidade de alvos relacionados à agricultura, tais como pastagens, soja e milho, além de florestas nativas e pastagens naturais (IBGE, 2004). Soma-se a essa característica um relevo movimentado e massiva presença de Latossolos e Neossolos Litólicos (Embrapa, 2011), ambas características desafiadoras ao sensoriamento remoto, seja pelo potencial de grande número de locais com baixos valores de albedo devido a presença de sombra e sesquióxidos de ferro no solo, ou pela alta heterogeneidade dos alvos agrícolas associados a cultivos de pequena extensão envolvendo restos de cultivo (e.g. matéria seca), mistura de solo e vegetação, etc.

### 2.1. Processamento de dados Hyperion

A imagem Hyperion foi adquirida no nível 1Gst, com data de passagem de 09 de Outubro de 2004. Foram aplicados procedimentos de calibração radiométrica, considerando: (i) supressão de valores de radiância 0 devido a problemas de geração do dado original relativo a baixa resposta espectral de algumas bandas; (ii) supressão de bandas de absorção atmosférica. Dessa forma nós trabalhamos com o total de 158 bandas cobrindo a região do VNIR-SWIR (0.436 - 0.235  $\mu\text{m}$ ), as quais foram aplicadas procedimentos de supressão de efeitos da atmosfera e conversão para reflectância de superfície considerando parâmetros locais de: (i) vapor d'água e aerossóis, (ii) modelo de atmosfera (MODTRAN) com base em equivalência da coluna de vapor d'água, (iii) ângulo de inclinação terrestre sobre a área de estudo; (iv) visibilidade inicial; (v) altura de escala do aerossol e (vi) razão de mistura de  $\text{CO}_2$ . Apesar da inegável necessidade de correção atmosférica e conversão à reflectância para fins de espectroscopia de reflectância, tais procedimentos geram artefatos de ocorrência não linear na imagem próprios de sensores *pushbroom* (Jupp et al., 2004). Nesse trabalho tais ruídos não foram corrigidos propositalmente com o intuito já definido de avaliar o desempenho do algoritmo SOM na superação de tal problema.

## 2.2. Self-Organizing Maps (SOM)

A técnica SOM é uma alternativa de modelagem adequada para tratamento de dados que apresentam relações não lineares (Kohonen, 2001), pois trata-se de um método não supervisionado que não requer conhecimento *a priori* das relações subjacentes entre as variáveis investigadas ou designação de padrões de saída (Friedel, 2011).

O SOM pode ser combinado com a técnica de agrupamento k-means e outras técnicas estatísticas multivariadas para avaliar e explorar relações complexas e não-lineares entre variáveis esparsas e com alto nível de ruído. Neste estudo, o resultado consiste em neurônios organizados em uma grade retangular bi-dimensional com células hexagonais (mapa).

O processo de treinamento do SOM fornece uma maneira de representar dados de alta dimensionalidade em um espaço de dimensionalidade reduzida, muito menor que o conjunto de dados original (Kohonen, 1984), formando uma grade de neurônios competitivos chamada de camada de saída (Kohonen, 2001). Neste estudo, cada dado de entrada  $x \in X$  é considerado um vetor no conjunto  $X = [x_1, x_2, \dots, x_N]^t$  com  $N$  sendo a dimensão do espaço dos dados de entrada. Um número fixado de  $t$  neurônios indexado por  $i$  é arranjado em uma grade regular  $G$  com cada neurônio associado com um vetor de peso  $w_i$  no conjunto  $W = [w_1, w_2, \dots, w_N]$ , o qual tem a mesma dimensionalidade dos vetores de entrada. Estes vetores de peso conectam cada vetor de entrada  $x$  em paralelo com todos os neurônios de  $G$ . Os neurônios são conectados uns aos outros, e neste caso, a interconexão é definida usando um torróide.

É utilizado um gradiente estocástico como método de aprendizagem do SOM (Kohonen, 1984). Consiste em um processo de dois passos que é realizado a cada vez que um padrão de entrada é apresentado ao mapa: competição para determinar a *Best Matching Unit* (BMU) e a aprendizagem cooperativa (informação de propagação contida no vetor de entrada ao longo do mapa). Os vetores dos dados de entrada  $X$  são apresentados ao mapa de uma forma aleatória para gerar grupos sem introduzir tendências para uma classe específica. Cada vez que um vetor de entrada é apresentado ao mapa, a função de custo é calculada e o número de vezes que cada neurônio torna-se uma BMU é gravado. No final de cada interação, a função de custo médio  $R$  é calculada. O processo de treinamento para quando  $R$  é menor que uma fração determinada de seu valor inicial, por exemplo,  $R < 10^{-5}$ .

Em última análise, este processo de treinamento resulta em uma topologia onde similaridades entre padrões de dados são mapeados em pesos semelhantes dos neurônios vizinhos, e a função de densidade assintótica local dos pesos neurais aproxima-se daquela do conjunto de treinamento (Ritter e Schulten, 1986).

O treinamento do mapa é conduzido em duas fases: a primeira fase envolve 20 interações usando uma vizinhança gaussiana com um raio inicial e final de 11 unidades e 3 unidades; e a segunda fase envolve 400 interações usando usando vizinhança gaussiana com um raio inicial e final de 3 unidades e 1 unidade. As taxas de aprendizagem inicial e final de 0.5 e 0.05 decaem linearmente para  $10^{-5}$ , e a função de vizinhança gaussiana diminui exponencialmente, fornecendo convergência razoável para o problema.

## 2.3. Procedimentos SOM aplicados a dados hiperespectrais

Na fase de treinamento do SOM foi usado um conjunto de 8 assinaturas espectrais de solo, provenientes de biblioteca espectral de solos tropicais (Vicente, 2007). Existem importantes feições espectrais relacionadas à presença de óxidos de ferro, típicas de solos tropicais, mais especificamente hematita, por volta de 0.530 a 0.800  $\mu\text{m}$ , e uma elevação gradual na reflectância para comprimentos de ondas mais longas, causado principalmente pela variação no tamanho das partículas e presença de quartzo. No SWIR, há uma feição diagnóstica importante entre 2.100 e 2.203  $\mu\text{m}$ , relacionada com a ocorrência de hidróxidos

(OH) presentes na argila, neste caso a caulinita (Vicente, 2007) (Fig. 1). As assinaturas espectrais foram reamostradas para a resolução espectral do Hyperion (158 bandas, 0.350 - 0.235  $\mu\text{m}$ ) e usadas nos procedimentos descritos abaixo.

Devido aos aspectos físicos particulares de cada alvo deseja-se que eventuais espectros puros sejam participes de campo ou laboratório da área de estudo ou extraídos da própria cena, mas nesse caso o objetivo era fornecer ao SOM um range de parâmetros de entrada para a formação de agrupamentos na imagem, considerando características locais possíveis de serem encontradas em solos tropicais e seu *background*, no caso a vegetação, reforçando o caráter não supervisionado de nosso estudo. Para a aplicação do algoritmo SOM foi utilizado o software SiroSOM<sup>®</sup> (CSRIO - *Commonwealth Scientific and Industrial Research Organisation*)

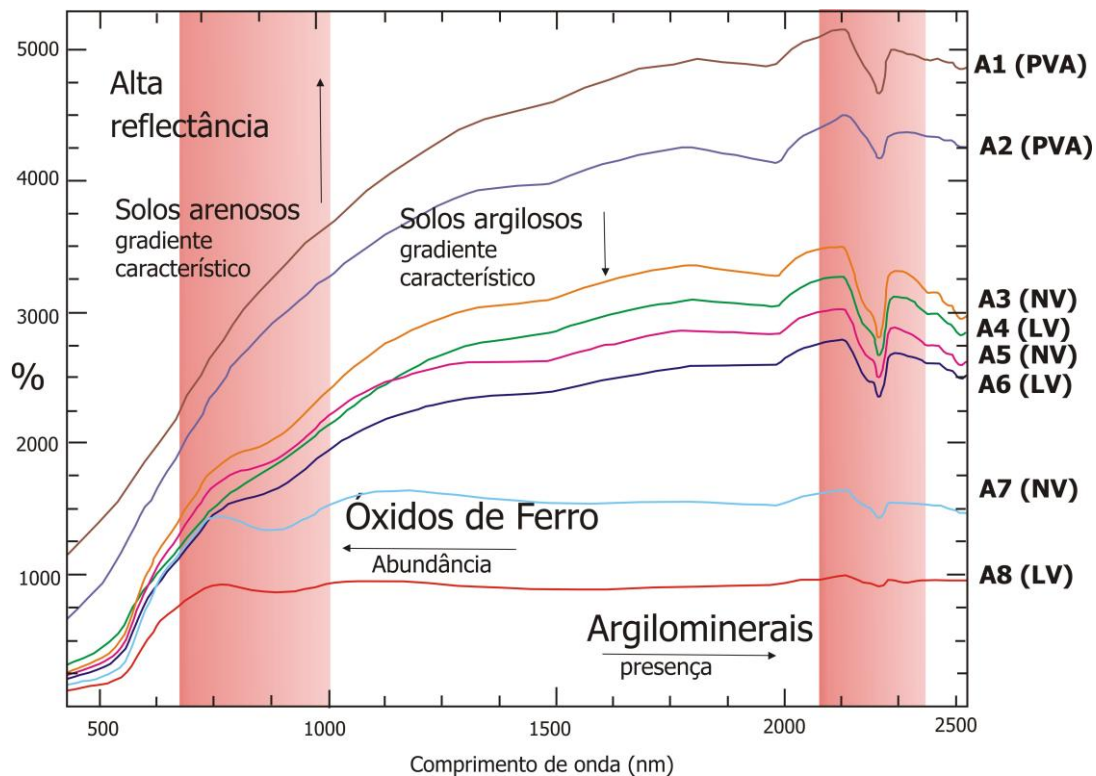


Figura 1. Biblioteca espectral de solos tropicais utilizada como entrada de dados no processamento SOM: Argissolo Vermelho Amarelo (PVA), Nitossolo Vermelho (NV), Latossolo Vermelho (LV).

### 3. Resultados e discussão

A análise dos planos components revela (Fig. 2b) as relações não evidentes entre as bandas do Hyperion e as amostras das bibliotecas espectrais sumarizadas pela Matriz-U (Fig. 2a). Observa-se três agrupamentos da Matriz-U relacionados a: VFA (Vegetação Fotossinteticamente Ativa), Solo Exposto e Solo Argiloso (Fig. 2a). Os planos components fornecem contribuição de cada banda para a formação da Matriz-U (Fig. 2b). Exemplo, a banda RefB38, localizada em 0.803  $\mu\text{m}$ , contribui com baixos valores de reflectância (cores frias - azul) associados a feição espectral diagnóstica de óxidos de ferro na mesma posição da Matriz-U para o cluster 2, referente a Solos Argilosos. Dessa forma, é possível analisar e validar a contribuição de cada banda para a classificação.

Na Matriz-U é possível analisar o agrupamento das três coberturas de solo detectadas, sendo que o azul escuro representa alta similaridade entre as variáveis agrupadas e o

vermelho, alta dissimilaridade. Desta forma, ficam visíveis os agrupamentos distintos com bordas transicionais. Para o reconhecimento entre as coberturas representadas pelo processamento e a cobertura correspondente *in situ* foi realizada classificação visual, utilizando a própria imagem Hyperion (composição colorida cor verdadeira) em contraposição a imagem de alta resolução espacial do mesmo período.

Destaca-se o grupo 2, como solos com alto teor de argila. Nota-se a participação de muitos neurônios agrupados em uma mesma célula (representado pelo tamanho do hexágono branco), significando que o processamento foi efetivo em detectar e reunir as características da cobertura representada. Em contrapartida, na classificação da vegetação (grupo 1) é possível observar a disposição mais disseminada dos neurônios na matriz, refletindo a maior diversidade de características que compõe o alvo. Este fato encontra correspondência na realidade, em virtude da heterogeneidade dos tipos de vegetação encontrados na área, tais como: matas, soja, milho e pastagens. Padrão semelhante ocorre em relação ao grupo 3, onde foi identificado solo exposto com presença de matéria seca.

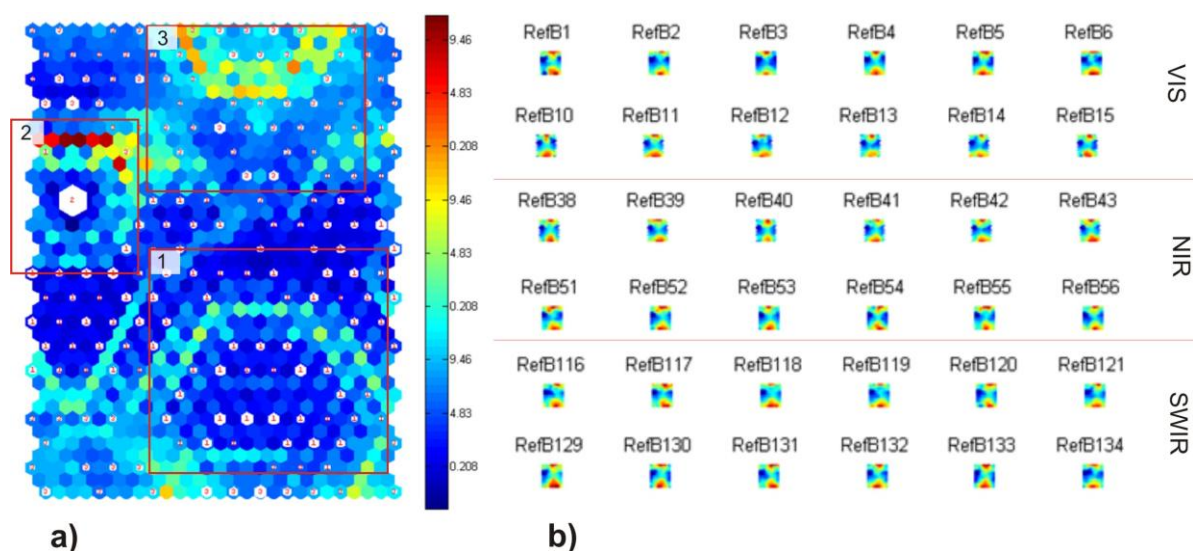


Figura 2. Resultados SOM. Agrupamento dos dados Hyperion na Matriz-U (a); exemplos de Planos Componentes das bandas Hyperion por região do espectro eletromagnético (b).

Os agrupamentos da classificação SOM: VFA, Solo Exposto e Solo Argiloso foram espacializados e validados na área de estudo através de imagens de alta resolução do mesmo período (Fig. 3a). A origem da vegetação não fotossinteticamente ativa (VNFA) sobre solos é extremamente comum em ambientes agrícolas, e é atribuída principalmente a matéria seca oriunda de restos de cultivo e gramínea de pastagem (Serbin, 2009), ambos abundantes no local. O período da imagem corresponde a época de plantio e preparo do solo para os cultivos locais, propiciando a melhor exposição de áreas de solos e favorecendo a identificação espectrorradimétrica de seus componentes.

Nosso objeto de análise foi a gradação da presença de Solo Exposto para Solo Argiloso nas áreas não vegetadas mapeadas, nesse caso a VFA serviu de *background* de referência. Na figura 3b é possível avaliar a presença da abundância relativa das classes Solo Exposto e Solo Argiloso, em contraposição a imagem Hyperion original através de composição colorida cor verdadeira. Nesse sentido, mesmo com a alta reflectância associada a matéria seca (VNFA) foi possível a separação de áreas de solo exposto com maior ou menor presença de Solos Argilosos, excluindo áreas de pastagem com mistura de solos, vegetação rasteira, bem como alvos associados a baixos valores de albedo, tais como sombras e umidade

no solo/matéria orgânica, ambos podendo ser classificados como solos com alto teor de ferro (Vicente e Souza Filho, 2011).

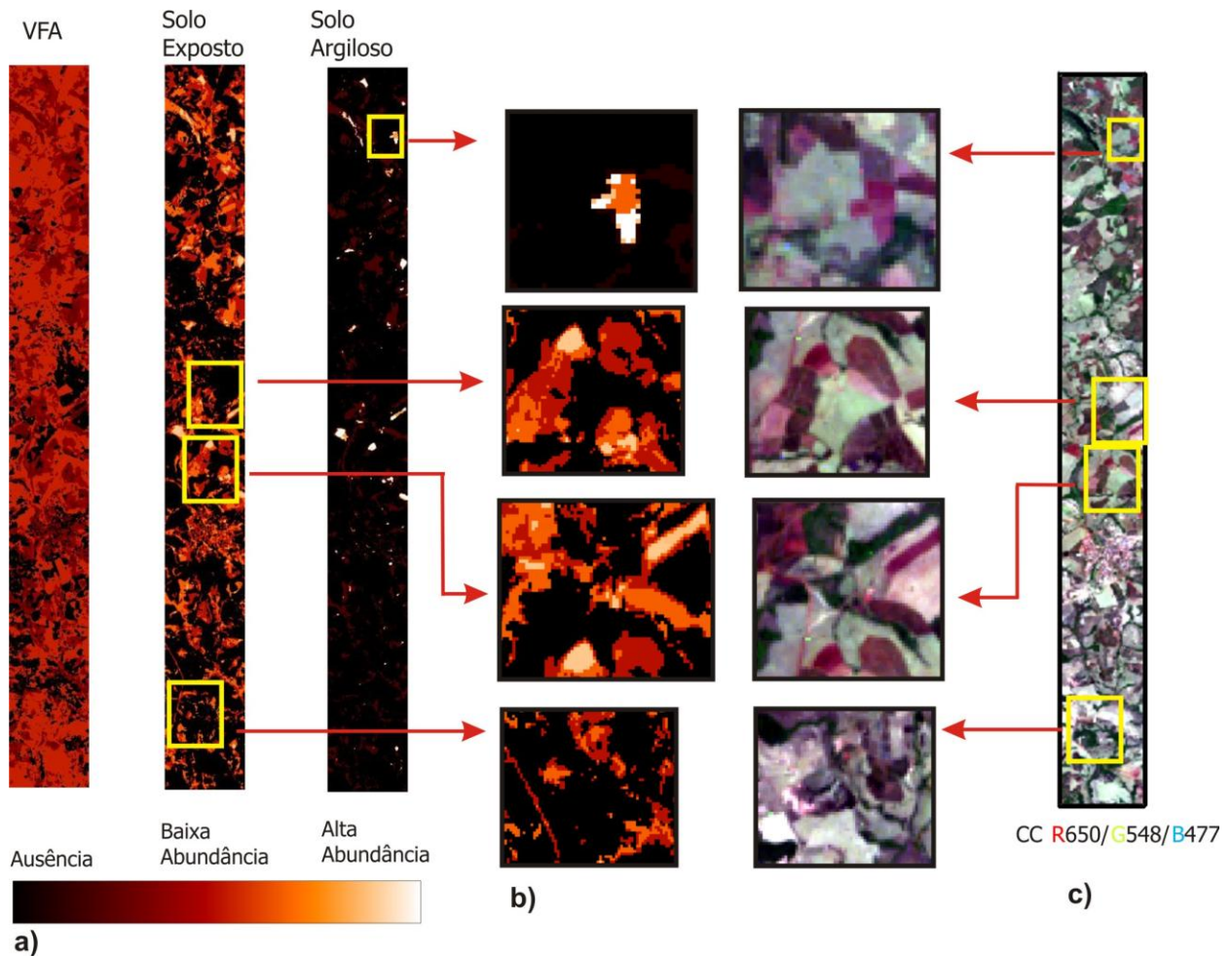


Figura 3. Imagens resultantes da classificação SOM para: VFA, Solo Exposto e Solo Argiloso (a). Detalhes da classificação SOM (b) em comparação com imagem Hyperion CC R650/G548/B477 (c).

Foram realizadas validações banda a banda da classificação SOM/Hyperion para diferentes áreas teste, algumas das quais são mostradas como exemplo na figura 4b, obedecendo o critério de uma hipotética gradação de abundância de solos arenosos à argilosos (figuras 4b'1' - 4b'4'). Todos os resultados foram satisfatórios ( $R^2$  - 0.90-0.93), com destaque para os solos argilosos, corroborado pelo mapeamento de solos da área com predominância de Latossolos e Nitossolos. Ademais, locais classificados como Solo Exposto são afetados com a provável influência da VNFA enquanto matéria seca, tendendo a um aumento dos valores de reflectância (Serbin, 2009). Em todos os locais de validação foram coletados espectros da imagem original Hyperion (figura 4c), onde pode-se observar os ruídos ao longo do conjunto de bandas (círculo preto). Também foi implementada uma validação para o VFA com locais de vegetação verde (figura 4b'2') em contraposição ao espectro de pastagem (figura 4c'2'), onde os valores de reflectância muito se assemelham aos de solos arenosos (figura 4c'4'), demonstrando a dificuldade de separação espectral entre esses alvos.

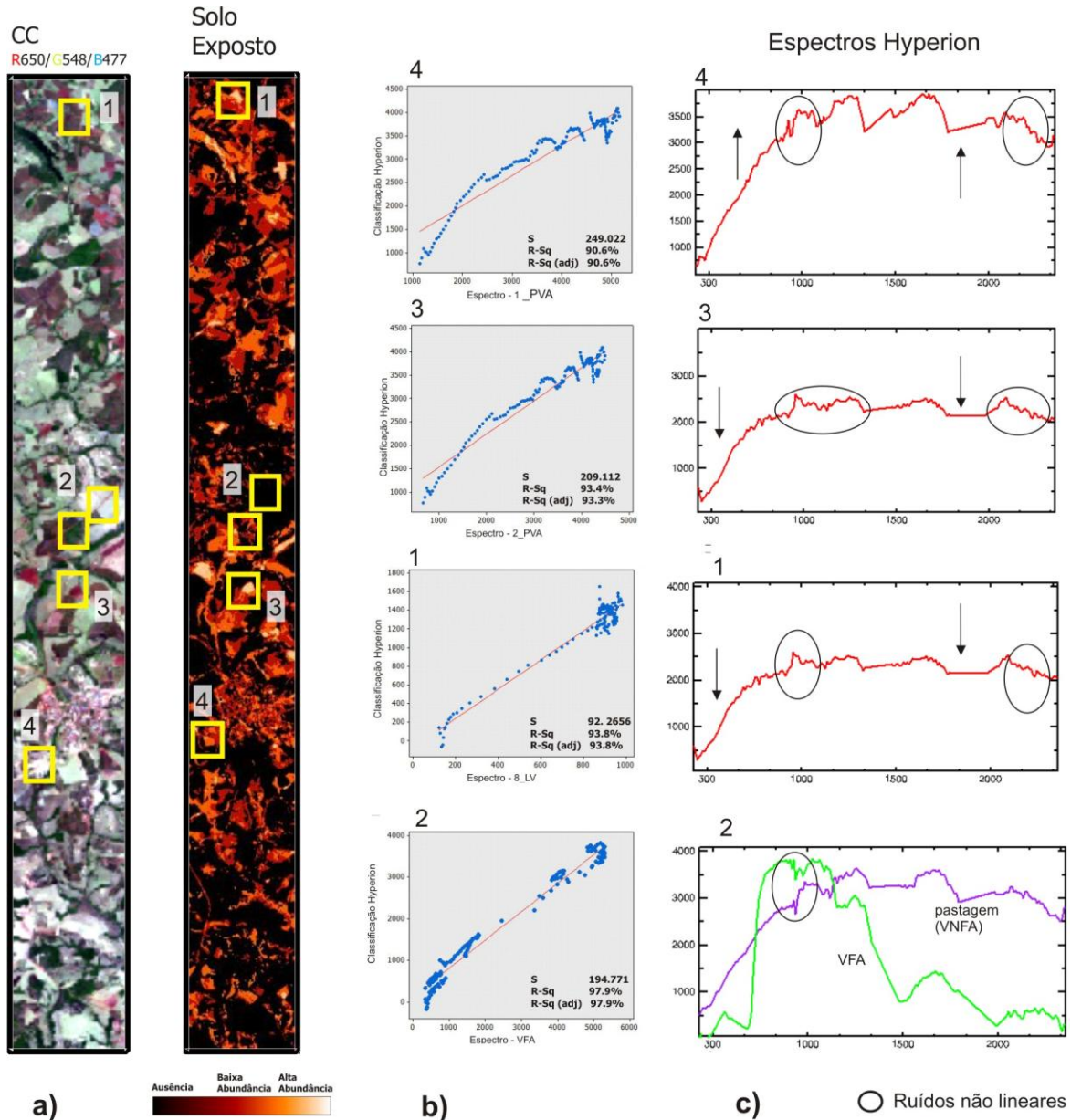


Figura 4. Comparação entre imagem Hyperion CC R650/G548/B477 e classificação SOM Solo Exposto (a); validação das áreas testes (b); espectros Hyperion coletados nas áreas teste (c). Os círculos pretos indicam ruídos não lineares nos dados Hyperion.

#### 4. Conclusões

A capacidade do SOM em identificar relações não lineares foi satisfatoriamente testada, considerando que mesmo sem a aplicação de procedimentos para redução dos ruídos originais dos dados Hyperion (SNR – *Signal to Noise* ~ 100:1) foi possível identificar diversas características espectrais dos alvos sob investigação, com destaque para o mapeamento semi-quantitativo de Solo Exposto e seu gradiente para solos com abundância em componentes argilominerálicos, sendo esse o principal objetivo de mapeamento do trabalho.

O SOM mostrou-se computacionalmente robusto frente a grande quantidade de dados da cena Hyperion (158 bandas x 91 x 818), sobretudo por ter a capacidade de manter as

relações topológicas durante o processamento, em detrimento de informações periféricas de localização, por exemplo, que onerariam o tempo de processamento. A presença massiva de vegetação, típica de áreas tropicais, não representou problemas de separabilidade dos alvos principais, entretanto, pastagem e solo exposto mostraram acentuada ambiguidade espectral, provavelmente devido a presença de matéria seca (VNFA) sobre o solo.

O uso de bibliotecas espectrais genéricas de alvos próprios de ambiente tropical reforça novas possibilidades de análise em espectroscopia de reflectância, reduzindo tempo de processamento e reiterando a importância da elaboração e validação de base de dados espectrorradiométricos. Espera-se que com uma maior quantidade de dados de campo, a aplicação do SOM seja cada vez mais efetiva para o sensoriamento remoto hiperespectral.

## 5. Agradecimentos

À equipe de pesquisa e desenvolvimento do projeto Agspec da Embrapa Monitoramento por Satélite (CNPM) e respectivos parceiros da *United Geological Survey* (USGS), Griffith University e do Centro de Estudos Meteorológicos e Climatológicos Aplicados a Agricultura (CEPAGRI/UNICAMP), pelo e apoio às atividades desenvolvidas nesse trabalho.

## 6. Referências Bibliográficas

Embrapa. **Mapa de Solos do Brasil, 1 : 5 0000**. Embrapa Solos (CNPS), 2011. Acesso em <<http://www.cnps.embrapa.br/>>

Friedel, M.J., 2011a. **Modeling hydrologic and geomorphic hazards across post-fire landscapes using a self-organizing map approach**. *Environmental Modelling & Software* 26 (12), 1660-1674

Instituto Brasileiro de Geografia e Estatística (IBGE). **Pesquisa Agrícola Municipal**. Acesso em 2012 <<http://www.ibge.gov.br/home/estatistica/indicadores/agropecuaria/lspa/default.shtm>>

Jupp, D. L. B.; Datt, B.; Lovell, J.; Campbell, S.; King, E. **Discussions around Hyperion Data: Background Notes for the Hyperion Data Users Workshop**. CSIRO, Office of Space Science and Applications Earth Observation Centre, 46 p. 2004.

Kohonen, T. **Self-organization and associative memory**, Berlin: Springer-Verlag, 1984. 312p.

Kohonen, T., 2001. **Self-organizing maps**, Third Extended Edition, Springer Series in: Information Sciences, V. 30, Springer, Berlin, Heidelberg, New York, 253 pp.

Ritter, H. Schulten, K., 1986. **On the stationary state of Kohonen's self-organizing sensory mapping**. *Biological Cybernetics* 54 (1), 99-106.

Serbin, G.; Daughtry, G. S. T.; Hunt, E. R.; Jr., Reeves, J. B., III.; Brown, D. J. **Effects of soil composition and mineralogy on remote sensing of crop residue cover**. *Remote Sensing of Environment*, v. 13, 224 – 238, 2009.

Tits, L.; Stuckens, J.; Farifteh, J.; Coppin, P. **Integration of in situ measured soil status and remotely sensed hyperspectral data to improve plant production system monitoring: Concept, perspectives and limitations**. v. 115, p. 1824-1836, 2013.

Vicente, L. E. **Caracterização de Sistemas Ambientais Tropicais Complexos utilizando análise Sistêmica e Classificação Hiperespectral de dados do sensor ASTER (Advanced Spaceborne Thermal Emission and Reflection Radiometer)**. 2007. 192 f. Tese (Doutorado) - Instituto de Geociências, Universidade Estadual de Campinas, Campinas.

Vicente, L. E.; Souza Filho, C. R. **Identification of mineral components in tropical soils using reflectance spectroscopy and advanced spaceborne thermal emission and reflection radiometer (ASTER) data**. v. 115, p. 1824-1836, 2011.