

ISSN 0104-1347

Investigações sobre o uso de dados hiperespectrais na classificação de espécies agrícolas com respostas espectrais muito semelhantes¹

On the use of hyperspectral image data for classification of crops with nearly equal spectral response

Mauro Erbert², Vitor Haertel³ e Moacir Antônio Berlato⁴

Resumo - A investigação trata do emprego de dados de imagens hiperespectrais no processo de classificação de espécies agrícolas que apresentam curvas de resposta espectral semelhantes. Neste contexto, a utilização de dados com baixa dimensionalidade, provenientes de sensores tradicionais como LandSat-TM com 6 bandas espectrais refletivas, normalmente produzem resultados pouco precisos. A utilização de dados hiperespectrais e de metodologias adequadas ao tratamento destes, são presentemente assuntos de grande interesse entre a comunidade científica. Neste trabalho são investigadas as diferentes técnicas de regularização da matriz covariância tendo em vista a utilização de dados em alta dimensionalidade. São explorados métodos que utilizam dois parâmetros de regularização (**I, g**). Os dados coletados utilizados são provenientes do sensor AVIRIS com 224 bandas espectrais. Os experimentos foram realizados em espécies de Milho e Soja, onde ocorrem situações em que as classes de interesse apresentam curvas de resposta espectral semelhantes sendo portanto de difícil classificação. Os resultados dos experimentos mostram que, com pequenos ajustes no parâmetro de regularização **I**, é possível melhorar a acurácia do classificador quadrático em »30%, minimizando o problema conhecido como “fenômeno de Hughes”. Contudo, constatou-se um comportamento diferente com o segundo parâmetro de regularização **g** onde sua utilidade é praticamente nula no contexto dos experimentos (dados reais), diferentemente do constatado nos resultados obtidos com dados simulados conforme reportado na literatura. Conclui-se que os dados hiperespectrais são muito úteis ao setor agrícola, principalmente para separação de alvos que possuam curvas de resposta espectral muito semelhantes.

Palavras-chave: sensoriamento remoto, hiperespectral, regularização, matriz covariância, classificação, agricultura

Abstract - This study deals with the use of high-dimensional image data applied to the classification of crops that present nearly equal spectral responses. In such cases, the use of the more traditional low-dimensional image data such as the data provided by the sensor system LandSat either yield low accuracy results or fail completely. This problem can be solved by employing very high-dimensional image data, such as the data provided by the AVIRIS system. The problem that arises in this case lies in the estimation of the parameters required by the classifier, particularly the covariance matrix. In this study it is investigated a possible approach to alleviate this problem, namely the methods of regularization of the covariance matrix. The proposed methodology implements approaches that make use of two regularization parameters (**I, g**) as proposed by different authors. Tests were performed on soybean and corn crops having very similar spectral responses. The results showed that the first parameter (**I**) plays an important role in improving the accuracy in the classification process whereas the second parameter (**g**) proved to be of little help to improve the classification accuracy. The results of this investigation showed the usefulness of high-dimensional image data in the classification of agricultural fields.

Key words: remote sensing, hyperspectral, regularization, covariance matrix, classification, agriculture

¹Parte da Dissertação de mestrado do primeiro autor apresentada ao Programa de Pós-Graduação em Sensoriamento Remoto, Centro Estadual de Pesquisas em Sensoriamento Remoto e Meteorologia CEP SRM/UFRGS. Av. Bento Gonçalves 9500, Campus do Vale, CP 15044 – CEP 91501-970, Porto Alegre – RS, Brasil – <http://www.ufrgs.br/srm>;

²Professor, MSc, Sensoriamento Remoto, Instituto de Informática - Universidade Luterana do Brasil, (merbert@uol.com.br);

³Professor, PhD, Centro de Sensoriamento Remoto e Meteorologia - Universidade Federal do Rio Grande do Sul, (victor.haertel@ufrgs.br);

⁴Professor, Dr., Centro de Sensoriamento e Meteorologia - Universidade Federal do Rio Grande do Sul, (moacir.berlato@ufrgs.br);

Introdução

Métodos para classificação de imagens digitais multiespectrais em sensoriamento remoto, foram e continuam sendo objeto de estudo de um grande número de pesquisadores (JIA & RICHARDS (1994), HAERTEL & LANDGREBE (1999), CORTIJO *et al.* (1999), HAERTEL (2000)). A utilização destas metodologias para fins de classificação de espécies vegetais, constitui-se em uma importante aplicação das técnicas de sensoriamento remoto. Neste caso, um problema especialmente desafiador ocorre quando duas ou mais espécies distintas apresentam características espectrais muito próximas, isto é, quando as curvas de resposta espectral das espécies individuais, e portanto os vetores média de suas respostas espectrais, são muito semelhantes entre si. Nestas condições torna-se difícil separar as mesmas com uma acurácia aceitável pela maioria dos usuários.

Utilizando dados coletados pelos sistemas sensores tradicionais, como por exemplo os sistemas LandSat-TM e SPOT (7 e 5 bandas espectrais, respectivamente), este problema de baixa acurácia torna-se de difícil solução. Uma possível alternativa para este problema, recentemente explorada, consiste na utilização de uma nova geração de sistemas sensores que coletam imagens em um número muito maior de bandas espectrais. Estes sistemas, denominados genericamente de *hiperespectrais* (por exemplo AVIRIS, com 224 bandas espectrais), coletam dados de uma mesma área, em um número muito maior de bandas espectrais, fornecendo desta forma dados em alta dimensionalidade.

Duas classes espectralmente idênticas, isto é, que apresentam valores idênticos para o vetor média, podem ser separadas com alta acurácia, desde que suas matrizes covariância difiram substancialmente entre si (FUKUNAGA, 1990). Entretanto, o processo de classificação que utiliza dados em alta dimensionalidade apresenta problemas que não ocorrem, ou são de menor importância, quando se opera com dados em baixa dimensionalidade. O mais relevante destes problemas consiste na estimação dos parâmetros utilizados pelos métodos estatísticos de classificação, como o classificador Bayesiano ou a Máxima Verossimilhança Gaussiana (MVG), que estão entre os processos de classificação mais utilizados (JAIN *et al.*, 2000). A maior dificuldade encontrada reside na estimação da matriz covariância, pois à medida que o número de bandas espectrais cresce

(aumento da dimensionalidade dos dados), aumenta também o número de parâmetros a serem estimados (ver Figura 1), o que é feito a partir de amostras representativas de cada uma das classes envolvidas (amostras de treinamento). Para que estas estimativas sejam confiáveis, o número de amostras de treinamento deve ser compatível com o número de parâmetros a serem estimados (FUKUNAGA & HAYES, 1989). Em situações reais, o número de amostras de treinamento disponível é geralmente limitado. Por esta razão, um esforço considerável vem sendo desenvolvido pela comunidade científica no sentido de encontrar alternativas para essa dificuldade, principalmente no que se refere à estimação da matriz covariância. Entre as possíveis alternativas para minorar este problema encontram-se os métodos de regularização da matriz covariância, que consiste em substituir a estimativa da matriz covariância de cada classe por uma combinação linear das estimativas da matriz covariância da classe e a matriz covariância comum. Esta técnica vem sendo explorada por vários autores, incluindo FRIEDMAN (1989), AEBERHARD *et al.* (1994) e CORTIJO *et al.* (1999).

O presente estudo investiga o comportamento destas técnicas de regularização da matriz covariância no processo de classificação de imagens hiperespectrais de sensoriamento remoto aplicados à agricultura e para o caso de espécies agrícolas com características espectrais muito semelhantes, e portanto separáveis somente com baixa acurácia quando são utilizados dados multiespectrais tradicionais de baixa dimensionalidade.

Material e métodos

Nos processos de classificação de natureza estatística, cada pixel em uma imagem é representado por um vetor cujas componentes são as respostas ou valores fornecidos por cada banda espectral. Nesta abordagem, cada pixel pode ser visto como um ponto em um espaço de n dimensões, sendo n igual à dimensionalidade dos dados, isto é, ao número de bandas espectrais utilizadas. O processo de classificação visa definir regiões neste espaço, cada qual associada a uma classe e caracterizada por uma função de natureza probabilística (função decisão) (TOU & GONZALES, 1974). No caso do classificador máxima verossimilhança (MV), a função decisão $G(x)$, associada à classe ω_i , tem a seguinte forma:

$$G_i(X) = p(X | \mathbf{w}_i)P(\mathbf{w}_i) \quad (1)$$

em que, X é o vetor que representa um pixel individual **a ser classificado**; $p(X/\omega_i)$ representa a função densidade de probabilidade associada à classe ω_i ; e $P(\omega_i)$ representa a probabilidade *a priori* associada à classe ω_i . Neste caso, a regra de classificação fica:

$$X \in \mathbf{w}_i \text{ se } G_i(X) > G_j(X), \text{ para todo } j \neq i \quad (2)$$

A efetiva utilização do classificador MV requer que seja conhecida a forma genérica da função densidade de probabilidade ($p(X/\omega_i)$). A experiência prática utilizando dados multiespectrais em sensoriamento remoto tem mostrado que a função densidade de probabilidade multivariada Gaussiana descreve, com bom desempenho, o comportamento destes dados. Neste caso, a função decisão (1) toma a forma:

$$G_i(X) = (2\pi)^{-n/2} |\Sigma_i|^{-1/2} \exp\left\{-\frac{1}{2}(X - \mathbf{m}_i)^T \Sigma_i^{-1}(X - \mathbf{m}_i)\right\} P(\mathbf{w}_i) \quad (3)$$

e o classificador passa a denominar-se Máxima Verossimilhança Gaussiana (MVG), onde n é a dimensionalidade dos dados, μ_i é o vetor média associada à classe ω_i e Σ_i é a matriz covariância associada à classe ω_i . De acordo com o critério de classificação (2), o resultado do processo de classificação não será alterado se (1) for substituída por uma função monotonicamente crescente de $G_i(x_i)$. Suprimindo-se ainda as parcelas ou fatores comuns à todas as $G_i(x_i)$, obtém-se

$$G_i(X) \propto -\ln|\Sigma_i| - (X - \mathbf{m}_i)^T \Sigma_i^{-1}(X - \mathbf{m}_i) + 2\ln(P(\mathbf{w}_i)) \quad (4)$$

$i=1, \dots$, número de classes (k)

A expressão (4) se constitui na forma mais comumente utilizada para a função decisão no classificador MVG. Esta decisão gera um classificador quadrático, incluindo termos X^2 resultantes do termo

$$(X - \mathbf{m}_i)^T \Sigma_i^{-1}(X - \mathbf{m}_i)$$

Os parâmetros μ_i e Σ_i são normalmente estimados a partir das amostras de treinamento disponíveis, por meio dos estimadores:

$$\hat{\mathbf{m}}_i = \frac{1}{N_i} \sum_{i=1}^{N_i} X_i \quad (5)$$

$$\hat{\Sigma}_i = \frac{1}{N_i - 1} \sum_{i=1}^{N_i} (X_i - \hat{\mathbf{m}}_i)(X_i - \hat{\mathbf{m}}_i)^T$$

sendo,

N_i o número de amostras de treinamento disponíveis na classe ω_i . O processo de classificação que implementa um classificador quadrático é genericamente denominado de Análise Discriminante Quadrática (ADQ).

Em alguns casos, é aceitável introduzir a simplificação de que todas as classes compartilham uma mesma matriz covariância, comum a todas elas ($\Sigma_1 = \Sigma_2 = \dots = \Sigma$). Neste caso, a matriz covariância comum pode ser estimada por:

$$\hat{\Sigma} = \frac{1}{N - 1} \sum_{i=1}^k N_i \hat{\Sigma}_i \quad (6)$$

$$N = \sum_{i=1}^k N_i$$

Neste caso, o termo quadrático em (4) torna-se comum para todas as classes e a função decisão (4) simplifica-se, tornando-se uma função linear em X :

$$G_i(X) = -\frac{1}{2}(X - \hat{\mathbf{m}}_i)^T \hat{\Sigma}^{-1}(X - \mathbf{m}_i) + \ln(P(\mathbf{w}_i)) \quad (7)$$

O processo de classificação que implementa um classificador linear é genericamente denominado de Análise Discriminante Linear (ADL).

A estimação dos parâmetros (μ_i, Σ_i) em dados de alta dimensionalidade apresenta novos desafios ao processo de classificação (HSIEH & LANDGREBE, 1998). O número de parâmetros a serem estimados, particularmente na matriz covariância, cresce rapidamente à medida que a dimensionalidade dos dados aumenta (Figura 1 e Tabela 1). Se o número das amostras de treinamento não aumentar, enquanto o número de parâmetros a serem estimados cresce, a confiabilidade na estimação destes decresce, ou seja, o acréscimo de bandas espectrais resulta no aumento da variância, isto é, da incerteza na estimativa dos parâmetros do classificador.

O número total de parâmetros (np) a serem estimados pode ser obtido pela expressão

$$np = \left(n + \frac{n^2 - n}{2}\right)$$

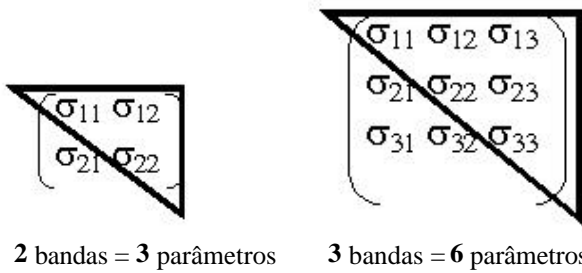


Figura 1. Exemplo do número de parâmetros a serem estimados na matriz covariância dependendo do número de bandas a serem utilizadas.

(conforme Figura 1), onde n representa a dimensionalidade dos dados. Portanto, para o caso de um sensor hiperespectral como o sistema AVIRIS (224 bandas), o número de parâmetros necessários, seria 900 vezes maior do que no caso de um sistema convencional como o LandSat-TM (7 bandas).

A degradação na acurácia dos resultados da classificação com o aumento da dimensionalidade dos dados, mantendo constante o número de amostras de treinamento, é conhecida como *fenômeno de Hughes* (HUGHES, 1968) ou a *maldição da dimensionalidade* (TADJUDIN & LANDGREBE, 1998) (Figura 2). Em um primeiro momento, o acréscimo de novas bandas espectrais, faz com que a acurácia da classificação tenda a aumentar, mas em um determinado ponto, entretanto, um máximo para a acurácia é atingido e a partir daí, um acréscimo no número de bandas espectrais passa a causar uma diminuição na acurácia do processo de classificação. Isto significa que o acréscimo de novos dados, ao invés de incrementar a acurácia do processo de classificação, tende a produzir o efeito inverso a partir de um determinado momento. Isto deve-se à degradação na estimativa dos parâmetros, especialmente na matriz covariância.

Tabela 1. Cálculo dos parâmetros necessários (np) para o sensor LandSat-TM e o sensor AVIRIS. Equação gerada a partir do número de parâmetros da matriz covariância (ver Figura 1).

LANDSAT-TM	AVIRIS
$n = 7$	$n = 224$
$np = (7 + \frac{7^2 - 7}{2})$	$np = (224 + \frac{224^2 - 224}{2})$
$np = 28$	$np = 25200$

Este problema, bem como possíveis alternativas para minimizá-lo, vem sendo objeto de estudo de vários pesquisadores, como HAERTEL & LANDGREBE (1999), CORTIJO & DE LA BLANCA (1999), HSIEH & LANDGREBE (1998), TADJUDIN & LANDGREBE (1998), SHAHSHAHANI & LANDGREBE (1994), AEBEHARD et al. (1994) e FRIEDMAN (1989).

Entre as possíveis soluções ressaltam-se: (1) o uso de amostras de treinamento não-rotuladas para a estimação de parâmetros, utilizando o algoritmo EM (*Expectation Maximization*); (2) a redução na dimensionalidade dos dados, como por exemplo no método de Componentes Principais, e (3) regularização da matriz covariância, que introduz outros estimadores para a matriz covariância de cada classe. Processos de classificação que fazem uso deste último método recebem a designação genérica de *Análise Discriminante Regularizada* (ADR).

Diferentes abordagens para fins de regularização da matriz covariância estão relatadas na literatura. Por exemplo, TITTERINGTON (1985), O'SULLIVAN (1986), FRIEDMAN (1989), mostram que as técnicas de regularização obtiveram alto sucesso na solução de problemas de estimativas não confiáveis da matriz covariância, devido ao número insuficiente de amostras de treinamento. Baseando-se ainda em CORNFIELD (1967), ao empregar um método de regularização, procura-se obter estimativas mais adequadas para os parâmetros, por meio da introdução de uma tendenciosidade adequada em sua

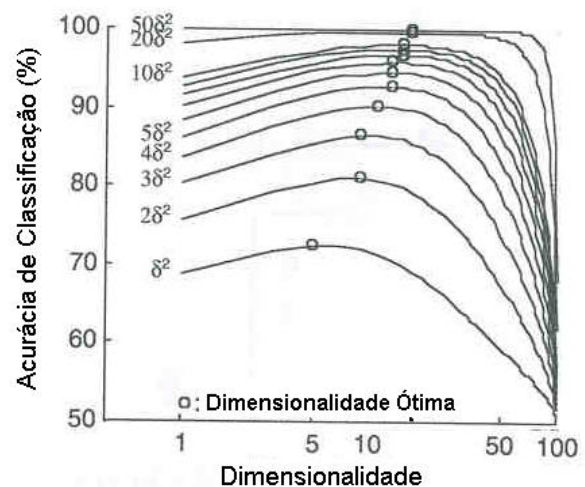


Figura 2. Fenômeno de Hughes. (HSIEH & LANDGREBE, 1998).

estimação, de forma que as estimativas sejam mais “plausíveis fisicamente” (FRIEDMAN, 1989), isto é, que a tendenciosidade introduzida tenda a corrigir as distorções que sabidamente ocorrem nestas situações.

Uma das propostas mais interessantes é aquela apresentada por FRIEDMAN (1989), com uma técnica que faz uso de dois parâmetros (λ e γ) variando no intervalo $[0,1]$. Através da utilização de diferentes valores numéricos para λ e γ , obtém-se situações intermediárias entre os métodos *ADQ* e *ADL*. Estas combinações têm por objetivo combinar, em diferentes graus, tanto a estimativa da matriz covariância comum, obtida a partir de todas as amostras disponíveis para todas as classes, quanto da matriz covariância de cada classe individual, resultando em uma estimativa mais confiável, denominada de *matriz covariância regularizada* ($\hat{\Sigma}_i(\lambda, \gamma)$) para cada classe individual. O primeiro passo para a obtenção da matriz covariância regularizada é dado por:

$$\hat{\Sigma}_i(\mathbf{I}) = \frac{(1-\mathbf{I})\hat{Q}_i + \mathbf{I}\hat{Q}}{N_i(\mathbf{I})}, \quad 0 \leq \mathbf{I} \leq 1 \quad (8)$$

onde:

$$N_i(\mathbf{I}) = (1-\mathbf{I})N_i + \mathbf{I}N \quad (9)$$

$$\hat{Q} = \sum_{i=1}^k \hat{Q}_i \quad \hat{Q}_i = N_i \hat{\Sigma}_i \quad (10)$$

em que $\hat{\Sigma}_i$ representa a matriz covariância própria de cada classe individual, estimada em (5), $\hat{\Sigma}_i \mathbf{I}$ representa a matriz covariância regularizada pelo parâmetro λ ; N é o somatório das amostras (N_i) de todas as classes; \hat{Q}_i representa a matriz covariância de cada classe $\hat{\Sigma}_i$, ponderada pelo tamanho da amostra correspondente e \hat{Q} é o somatório das estimativas \hat{Q}_i .

O parâmetro λ controla a contribuição da estimativa da matriz covariância $\hat{\Sigma}_i$ na estimativa da matriz covariância comum $\hat{\Sigma}$. Este primeiro procedimento de regularização da matriz covariância resulta em um classificador intermediário entre os métodos *ADL* e *ADQ*, e normalmente produz uma melhora significativa no resultado do processo de classificação, especialmente quando o tamanho de amostras das

classes individuais é pequeno. A regularização introduzida pela equação (8) pode, entretanto, ser insuficiente. Para minimizar este problema, FRIEDMAN (1989) sugere um segundo passo no processo, o qual tem por objetivo corrigir o fato bem conhecido de que quando o número de amostras de treinamento é pequeno, os valores estimados para os autovalores da matriz covariância tornam-se tendenciosos, ou seja, os autovalores menores são estimados ainda menores e os maiores ainda maiores. Este segundo passo de regularização é implementado pela equação (11). O efeito deste processo será amenizar a tendenciosidade provocada pelo parâmetro, diminuindo os autovalores maiores e aumentando os autovalores menores.

$$\hat{\Sigma}_i(\mathbf{I}, \mathbf{g}) = (1-\mathbf{g})\hat{\Sigma}_i(\mathbf{I}) + \frac{\mathbf{g}}{n} \text{tr}[\hat{\Sigma}_i(\mathbf{I})] \cdot \mathbf{I} \quad (11)$$

$$0 \leq \mathbf{I} \leq 1$$

$$0 \leq \mathbf{g} \leq 1$$

em que $\text{tr}[\hat{\Sigma}_i(\lambda)]$ é o traço de $\hat{\Sigma}_i(\lambda)$; e \mathbf{I} é a matriz identidade. Nota-se que $\text{tr}[\hat{\Sigma}_i(\lambda)]$ é igual à soma dos autovalores de $\hat{\Sigma}_i(\lambda)$, e portanto o quociente $\text{tr}[\hat{\Sigma}_i(\lambda)]/M$ representa a média dos autovalores de $\hat{\Sigma}_i(\lambda)$.

Empregando $\hat{\Sigma}_i(\lambda, \gamma)$ no lugar da $\hat{\Sigma}_i$ na função decisão (4), a função decisão regularizada torna-se:

$$G_i(X) = -(X - m_i)^T \hat{\Sigma}_i^{-1}(\mathbf{I}, \mathbf{g})(X - m_i) - \ln |\hat{\Sigma}_i(\mathbf{I}, \mathbf{g})| + 2 \ln(P(w_i)) \quad (12)$$

AEBERHARD et al. (1994), salientam não existir um processo que permita definir quais as melhores configurações de parâmetros de λ e γ , para que se tenha uma probabilidade de classificação correta (*PCC*) maximizada. A avaliação dos resultados do classificador é realizada experimentalmente, com um conjunto de pares de valores numéricos (λ, γ), e então são selecionados os valores onde o resultado tenha fornecido a melhor acurácia.

Os experimentos foram desenvolvidos utilizando-se dados do sensor hiperespectral AVIRIS que dispõe de 224 bandas espectrais, utiliza uma plataforma aerotransportada e é considerado um sensor experimental. A cena⁵ utilizada cobre uma região experimental, localizada no Estado de Indiana, Estados Unidos, na qual são realizados testes com várias es-

⁵É um segmento de imagem (435x435=189.225 pixels) de junho de 1992, no período inicial de crescimento, com grande quantidade de solo exposto. Possui 220 bandas espectrais das 224 bandas do sensor AVIRIS.

pécies agrícolas que apresentam uma resposta espectral muito semelhante, sendo utilizadas diferentes técnicas de plantio (plantio tradicional, plantio direto e plantio em cultivo mínimo). Entre as várias espécies agrícolas existentes na região, foram selecionadas cinco, como as classes a serem utilizadas neste estudo. São elas: milho plantio tradicional (MPT), milho cultivo mínimo (MCM), soja plantio tradicional (SPT), soja plantio direto (SPD) e floresta (F). Uma característica adicional que torna esta área adequada ao estudo proposto consiste na disponibilidade de dados de campo detalhados, os quais permitem uma análise precisa dos resultados dos testes desenvolvidos.

Estas classes, à exceção da floresta, apresentam respostas espectrais muito semelhantes entre si, o que torna o processo de classificação muito difícil, principalmente quando utilizados dados tradicionais de baixa dimensionalidade (LandSat-TM). Desta forma, estas classes servem muito bem para fins de teste dos métodos de classificação utilizando dados espectrais em alta dimensionalidade, como os fornecidos pelo sistema AVIRIS. A inserção da classe floresta nos experimentos deve-se ao fato da sua curva de resposta espectral ser distinta das demais classes, e portanto facilmente separável a partir de dados em baixa dimensionalidade, servindo como base de comparação para classificações utilizando dados em alta dimensionalidade. As curvas de resposta espectral das classes utilizadas neste estudo são exibidas na Figura 3.

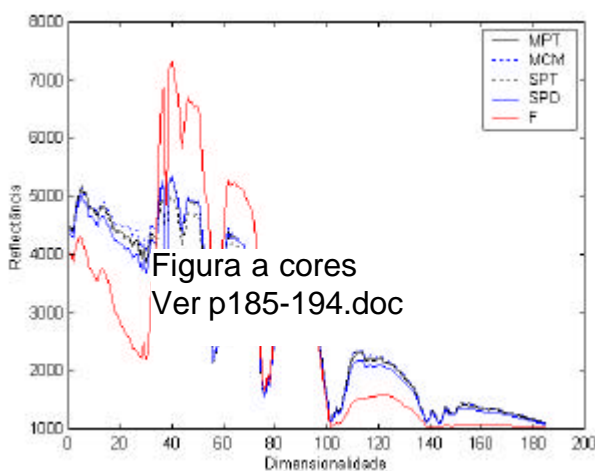


Figura 3. Curva de resposta espectral de cada classe, onde MPT (Milho Plantio Tradicional), MCM (Milho Cultivo Mínimo), SPT (Soja Plantio Tradicional), SPD (Soja Plantio Direto) e F (Floresta).

Objetivando fazer uso apenas das informações úteis para a separação das classes, foram desconsideradas as bandas ruidosas, isto é, bandas com o sinal afetado pelo efeito da absorção da atmosfera, especialmente pelo vapor d'água. Por esta razão, foram utilizadas neste estudo um total de 185 bandas espectrais (bandas: 3-102, 111-148, 166-220), das 220 disponíveis da cena. Foi executada uma série de testes visando determinar a acurácia obtida no processo de classificação utilizando os dados selecionados. Neste processo, tomou-se a dimensionalidade destes dados como variável independente e a resultante acurácia na classificação como variável dependente. O valor da dimensionalidade dos dados variou desde 5 até 185 (bandas espectrais), a intervalos de 5 bandas espectrais. Desta forma, é possível identificar os efeitos provocados pelo fenômeno de Hughes, pois, na medida em que a dimensionalidade aumenta pela adição de bandas espectrais, no classificador quadrático pode-se visualizar, detalhadamente, o comportamento da acurácia. As amostras disponíveis para cada classe foram divididas em dois grupos, as amostras para fins de treinamento (estimação dos parâmetros) e as amostras de teste (para testar a acurácia do processo de classificação). Para fins desta análise, selecionou-se um número pequeno de amostras de treinamento (200 amostras por classe). Esta quantidade de amostras é pequena se comparada à dimensionalidade dos dados, e tem por objetivo criar as condições adequadas à análise da eficiência das técnicas de regularização da matriz covariância em estudo, e que ocorre com frequência em situações reais. Para fins de estimar a acurácia do processo de classificação, foram usadas todas as amostras de teste disponíveis para cada classe (Tabela 2).

Os experimentos são realizados para vários pares de valores numéricos dos parâmetros de regularização da matriz covariância, onde λ e γ variam de 0 a 1, em intervalos de 0,1.

Tabela 2. Número de amostras de treinamento e teste das classes envolvidas no estudo.

Nome da classe	Amostras/ Treinamento	Amostras/ Teste
Milho Plantio Tradicional	200	1821
Milho Cultivo Mínimo	200	1221
Soja Plantio Tradicional	200	2327
Soja Plantio Direto	200	1132
Floresta	200	2332

Resultados e discussão

A classe floresta não foi considerada nos resultados, apesar de ter sido processada conjuntamente com as outras classes, pois alcançou uma acurácia de classificação superior a 99% em todos os experimentos.

Algumas considerações podem ser feitas ao atribuir os valores $\lambda = 0,0$ e $\gamma = 0,0$ para os parâmetros de regularização. Neste caso, o classificador assume a forma do classificador quadrático convencional (*ADQ*), isto é, a matriz covariância regularizada $\Sigma_i(\lambda, \gamma)$ (equações (8) e (11)), se reduz à matriz covariância de cada classe individual. A situação oposta ocorre para o par de valores numéricos $\lambda = 1,0$ e $\gamma = 0,0$. Neste caso, a matriz covariância regularizada $\Sigma_i(\lambda, \gamma)$ torna-se idêntica à matriz covariância comum Σ_i e o classificador assume a forma de um classificador linear (*ADL*). As tabelas de contingência (Tabela 3) ilustram os resultados alcançados em ambos processos.

Comparados os resultados das tabelas de contingência *ADQ* e *ADL*, verifica-se que o processo de classificação fornece resultados mais precisos quando se emprega o método *ADL*, obtendo porcentagens finais de acurácia superiores ao método *ADQ*. Este comportamento tende a confirmar o fato de que, em situações nas quais o tamanho das amostras de treinamento é muito pequeno quando comparado com a dimensionalidade dos dados, o classificador linear geralmente, fornece resultados mais precisos do que o quadrático (HOFFBECK & LANDGREBE,

1996; TADJUDIN & LANDGREBE, 1998). Nota-se que estes resultados foram obtidos para uma dimensionalidade igual a 185 (dimensão máxima) e amostras de treinamento de 200 pixels/classe. Observa-se que no método *ADL* o fenômeno de Hughes já é amenizado, porém em alguns casos a acurácia do classificador diminuiu consideravelmente (Figura 4).

Os gráficos na Figura 4 mostram o comportamento dos classificadores *ADQ* e *ADL* durante todo o processo de inclusão de novas bandas espectrais mantendo-se constante o número de amostras de treinamento. Os gráficos ilustram a porcentagem de acerto em cada classe, utilizando-se as 185 bandas espectrais disponíveis. Deve-se mencionar que em alguns casos, uma acurácia maior foi obtida utilizando-se um número menor de bandas espectrais, isto é, dados com dimensionalidade menor e portanto estimativas mais confiáveis para os parâmetros do classificador. Nota-se que o fenômeno de Hughes é evidente quando se utiliza o classificador *ADQ*. O classificador *ADL* fornece uma acurácia superior ao *ADQ* quando se trabalha em alta dimensionalidade (185 bandas) pelo fato de atenuar o fenômeno de Hughes. Este resultado superior obtido com o classificador linear (*ADL*) para dados em alta dimensionalidade claramente ilustra a situação que FRIEDMAN(1989) menciona, com relação à baixa confiabilidade da estimativa para os parâmetros associados as classes individuais quando o número de amostras é pequeno, como é o caso do experimento com 200 amostras de treinamento. Sabe-se que o classificador quadrático (*ADQ*) deveria fornecer

Tabela 3. Tabelas de contingência do processo quadrático (*ADQ*) e do processo linear (*ADL*). Percentagem de acertos dos classificadores em cada uma das classes. Respectivas quantidades de amostras de treinamento e amostras de teste utilizadas. Experimento com todas as bandas disponíveis(185).

	Milho Plantio Tradicional	Milho Cultivo Minimo	Soja Plantio Tradicional	Soja Plantio Direto	Floresta
Milho Plantio Tradicional	921	426	799	303	1
Milho Cultivo Minimo	357	432	267	172	1
Soja Plantio Tradicional	344	164	1055	92	0
Soja Plantio Direto	218	208	205	534	64
Floresta	0	0	0	30	2265
AMOSTRAS:					
Treinamento	200	200	200	200	200
Teste	1820	1220	2326	1131	2331
% Acertos (Classe)	50.5766%	35.3888%	45.3373%	47.1731%	97.1269%

ADQ Bandas → 185
 $\lambda \rightarrow 0$
 $\gamma \rightarrow 0$

	Milho Plantio Tradicional	Milho Cultivo Minimo	Soja Plantio Tradicional	Soja Plantio Direto	Floresta
Milho Plantio Tradicional	946	330	439	79	0
Milho Cultivo Minimo	367	727	108	121	0
Soja Plantio Tradicional	352	56	1744	53	0
Soja Plantio Direto	130	107	30	878	30
Floresta	0	0	0	0	2301
AMOSTRAS:					
Treinamento	200	200	200	200	200
Teste	1820	1220	2326	1131	2331
% Acertos (Classe)	51.9495%	59.5414%	74.9463%	77.5618%	98.6707%

ADL Bandas → 185
 $\lambda \rightarrow 1$
 $\gamma \rightarrow 0$

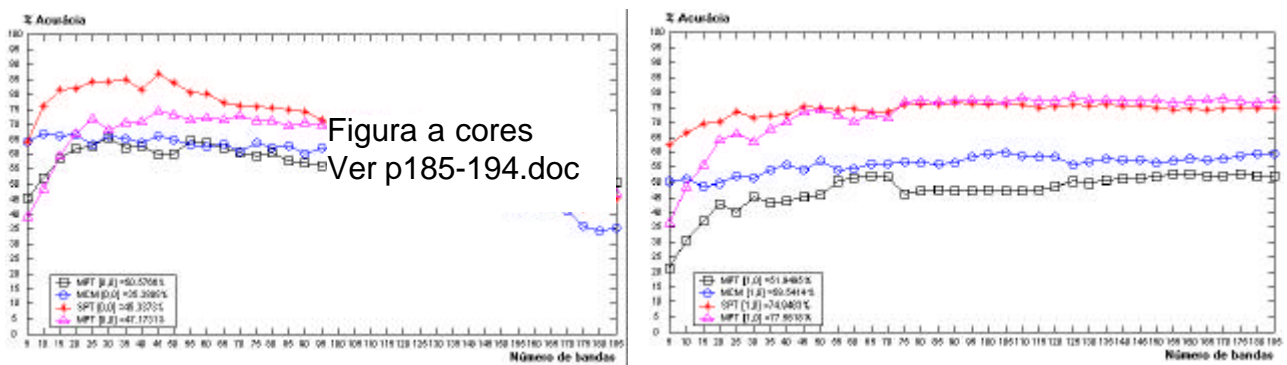


Figura 4. Comportamento da acurácia no aumento da dimensionalidade referente aos modelos *ADQ* e *ADL*, com o total de acurácia alcançada na utilização de todas as 185 bandas nas classes: MPT (Milho Plantio Tradicional), MCM (Milho Cultivo Mínimo), SPT (Soja Plantio Tradicional), SPD (Soja Plantio Direto) e F (Floresta).

resultados superiores desde que os valores estimados para os parâmetros das classes individuais fossem confiáveis. A introdução das técnicas de regularização devem em princípio minimizar o problema da estimação dos parâmetros. Os gráficos da Figura 5, mostram a acurácia de cada classe em diferentes **combinações dos parâmetros** e γ no classificador *ADR*.

Pode-se observar uma considerável melhora na acurácia das classes MCM e SPD, além de praticamente eliminar a degradação da acurácia ocasionada pelo fenômeno de Hughes em todas as classes. As acurácias mais altas obtidas com o uso de todas as bandas espectrais, predominam no classificador *ADR* (Tabela 4).

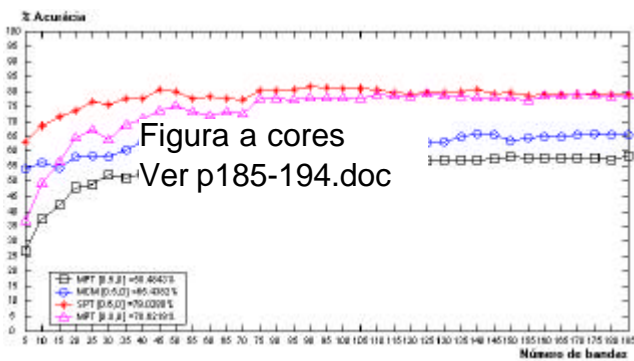


Figura 5. Comportamento da acurácia no aumento da dimensionalidade referente ao modelo *ADR*, e a melhor acurácia alcançada com a combinação de λ e γ nas classes: MPT (Milho Plantio Tradicional), MCM (Milho Cultivo Mínimo), SPT (Soja Plantio Tradicional), SPD (Soja Plantio Direto) e F (Floresta).

As melhores taxas de acerto das classes MPT e SPT foram obtidas com o método *ADQ*, atingindo em 30 bandas, uma acurácia de 65,40%, e com 45 bandas, uma acurácia de 87,06% respectivamente, porém a acurácia cai drasticamente com o aumento da dimensionalidade dos dados (Tabela 5). Já a classe MCM obteve a melhor acurácia com o método *ADR* com 65 bandas de 68,55% com $\lambda=0,1$ e $\gamma=0,0$, e a classe SPD com 125 bandas obteve acurácia de 79,68% com $\lambda=0,7$ e $\gamma=0,0$.

Outra característica importante é que apesar do método *ADL* ser inferior visto que não leva em consideração as informações das matrizes covariância características de cada classe individual, e ter obtido uma acurácia melhor do que a *ADQ* no uso de todas as bandas, o método *ADQ* atinge maior taxa de acerto, mas em um número menor de bandas. Deve-se enfatizar que estes experimentos foram realizados utilizando-se um número muito pequeno de amostras de treinamento por classe, condição esta encontrada com frequência em situações reais.

Os estudos realizados por FRIEDMAN(1989) e aqueles desenvolvidos por CORTIJO & DE LA BLANCA(1999), obtiveram resultados similares aos alcançados neste trabalho. O classificador *ADR* se constitui portanto em uma opção vantajosa ao classificador quadrático (máxima verossimilhança) *ADQ*, sempre que o número de amostras é pequeno quando comparado com a dimensionalidade dos dados.

A visualização da contribuição da *ADR* no processo de classificação pode ser analisado na Figura 6, onde selecionou-se a classe SPT, por exemplo, e a

Tabela 4. Acurácia máxima das classes nos classificadores com dimensionalidade final de 185 bandas.

Nome da classe	<i>ADQ</i>	<i>ADL</i>	<i>ADR</i>	λ	γ	<i>n= todas as bandas</i>
Milho Plantio Tradicional	50,57	51,94	58,48	0,5	0,0	185
Milho Cultivo Mínimo	35,38	59,54	65,43	0,5	0,0	185
Soja Plantio Tradicional	45,33	74,94	79,02	0,6	0,0	185
Soja Plantio Direto	47,17	77,56	78,62	0,8	0,0	185

partir da configuração quadrática ($\lambda=0,0;\gamma=0,0$), nota-se a mudança na melhora da acurácia e na diminuição dos feitos do fenômeno de Hughes, a medida que o nível de regularização de λ aumenta.

Conclusões

O método de regularização (*ADR*) é eficiente no processo de classificação de dados de sensoriamento remoto em alta dimensionalidade (imagens hiperespectrais) quando o número de amostras de treinamento é limitada. Em todos os experimentos realizados o método que implementa a regularização da matriz covariância (*ADR*) suaviza consideravelmente, os efeitos do fenômeno de Hughes presente no método *ADQ*, além de resultar em uma acurácia mais alta do que aquela obtida utilizando o classificador linear (*ADL*). O processo de regularização da matriz covariância mostra-se capaz de fornecer uma estimativa mais confiável do que aquela obtida pela equação (5), sempre que o número de amostras de treinamento é pequeno, quando comparado com a dimensionalidade dos dados. Desta forma, a *ADR* viabiliza a utilização de dados de sensoriamento remoto em alta dimensionalidade, mesmo em situações comumente encontradas na prática, isto é, quando o número de amostras de treinamento é limitado.

Como sugestões para futuros trabalhos nesta área, pode-se sugerir:

- buscar um melhor entendimento com relação à contribuição do parâmetro γ para a separação de

Tabela 5. Acurácia máxima alcançada de cada classe com "n" dimensionalidade.

Nome da classe	Maior Acurácia	Modo	λ	γ	<i>n</i>
Milho Plantio Tradicional	65,40	<i>ADQ</i>	0,0	0,0	30
Milho Cultivo Mínimo	68,55	<i>ADR</i>	0,1	0,0	65
Soja Plantio Tradicional	87,06	<i>ADQ</i>	0,0	0,0	45
Soja Plantio Direto	79,68	<i>ADR</i>	0,7	0,0	125

classes curvas espectrais muito semelhantes usando dados de sensoriamento remoto em alta dimensionalidade de (imagens hiperespectrais);

- estudar um método mais eficiente visando a obtenção das melhores configurações para os parâmetros de regularização (λ, γ), com a finalidade de tornar a aplicação do método *ADR* mais acessível ao usuário.

Referências bibliográficas

- AEBERHARD, S.; COOMANS, D.; DE VEL, O. Comparative Analysis of Statistical Pattern Recognition Methods in High Dimensional Settings, **Pattern Recognition**, v. 27, n. 8, p. 1065-1077, 1994.
- CORNFIELD, J., Discriminant Functions, **Review of the International Statistical Institute**, New York, v. 35, p. 142-153, 1967.
- CORTIJO F.J.; DE LA BLANCA N.P., The Performance of Regularized Discriminant Analysis Versus Non-parametric Classifiers Applied to High-Dimensional Image Classification, **International Journal of Remote Sensing**, v. 20, n. 17, p. 3345-3365, 1999.
- FRIEDMAN, J.H., Regularized Discrimination Analysis, **Journal of the American Statistical Association**, v. 84, n. 405, p. 165-175, 1989.
- FUKUNAGA, K.; HAYES R.R., Effects of Sample Size in Classifier Design, **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 11, n. 8, p. 873-885, 1989.
- FUKUNAGA, K., **Introduction to Statistical Pattern Recognition**, 2. ed., Boston: Academic Press, 1990.
- HAERTEL V. ; LANDGREBE, D.A., On the Classification of Classes with Nearly equal Spectral Response in Remote Sensing Hyperspectral Image Data, **IEEE Transactions on Geoscience and Remote Sensing**, v. 37, n. 5, p. 2374-2386, 1999.

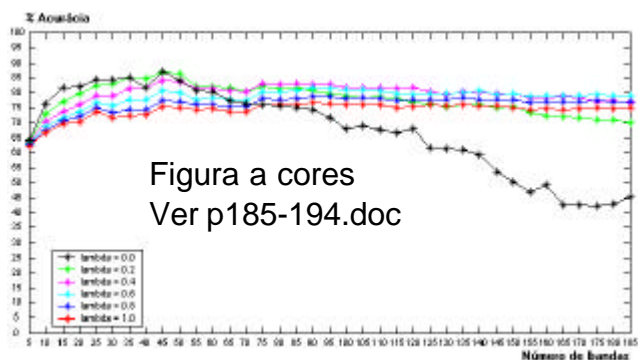


Figura 6. Minimização do Fenômeno de Hughes com a contribuição do parâmetro λ (lambda).

HAERTEL V., Considerations about Classification of Remote Sensing Digital Image Data in High Dimensional Settings. In: SIMPOSIO LATINOAMERICANO DE PERCEPTION REMOTA, 9., 2000. Puerto Iguazú, Misiones, Argentina. **Anais...**, Puerto Iguazú: Sociedad de Especialistas Latinoamericanos em Percepcion Romota y Sistemas de Informacion Espacial - SELPER, não paginado.

HOFFBECK J.P.; LANDGREBE D.A., Covariance Matrix Estimation and Classification With Limited Training Data, **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 18, n. 7, p. 763-767, 1996.

HSIEH, P.; LANDGREBE, D.A., Classification of High Dimensional Data. In: **School of Electrical And Computer Engineering**, Purdue: Purdue University, 1998. 121 p. (TR-ECE 98-4).

HUGHES, G.F. On the Mean Accuracy of Statistical Pattern Recognizers, **IEEE – Transactions an Information Theory**, v. II-14, n. 1, p. 55-65, 1968.

JAIN, A., DUIN, R.P.W. and MAO, J., Statistical Pattern Recognition: A Review, **IEEE Transactions on**

Pattern Analysis and Machine Intelligence, v. 22, n. 1, p. 4-37, 2000.

JIA, X.; RICHARDS, J., Efficient Maximum Likelihood Classification for Imaging Spectrometer Data Sets, **IEEE Transactions on Geoscience and Remote Sensing**, v. 32, n. 2, p. 274-280, 1994.

JIA, X.; RICHARDS, J., Segmented Principal Components Transformation for Efficient Hyperspectral Remote-Sensing Image Display and Classification, **IEEE Transactions on Geoscience and Remote Sensing**, v. 37, n. 1, p. 538-542, 1999.

JIMENEZ, L. O., LANDGREBE, D.A., Hyperspectral data Analysis and Supervised Feature Reduction Via Projection Pursuit, **IEEE Transactions on Geoscience and Remote Sensing**, v. 37, n. 6, p. 2653-2667, 1999.

O'SULLIVAN, F. A Statistical Perspective on Ill-Posed **Inverse Problems**, **Statistical Science**, v. 1, p. 502-527, 1986.

SHAHSHAHANI, B. M.; LANDGREBE D. A., The Effect of Unlabeled Samples in Reducing the Small Sample Size Problem and Mitigating the Hughes Phenomenon, **IEEE Transactions on Geoscience and Remote Sensing**, v. 32, n. 5, p. 1087-1095, 1994

TADJUDIN, S.; LANDGREBE, D.A., Classification of High Dimensional Data with Limited Training Samples. In: **School of Electrical And Computer Engineering**, Purdue: Purdue University, 1998. 123 p. (TR-ECE 98-8).

TITTERINGTON, D.M., Common Structure of Smoothing Techniques in Statistics, **International Statistical Review**, v. 53, p. 141-170, 1985.

TOU, J.; GONZALES R.C., **Pattern Recognition Principles**, 3. ed., 1979, Massachusetts, Reading: Addison-Wesley Publishing Company Inc. 1974. 377p. (Advanced Book Program).