

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL - UFRGS
CENTRO ESTADUAL DE PESQUISAS EM SENSORIAMENTO REMOTO E
METEREOLOGIA – CEP SRM
PROGRAMA DE PÓS-GRADUAÇÃO EM SENSORIAMENTO REMOTO – PPGSRM
MESTRADO

UTILIZAÇÃO DE SEMIVARIOGRAMAS COMO REDUTOR DE
DIMENSIONALIDADE NO RECONHECIMENTO DE PADRÕES EM
IMAGENS DIGITAIS HIPERESPECTRAIS

Dissertação submetida como requisito
parcial para a obtenção do grau de MESTRE
EM SENSORIAMENTO REMOTO

JÉFERSON DANIEL DE MATOS

Orientador: Prof^o Dr^o Robin T. Clarke

Co-orientador: Prof^o PhD Vitor Haertel

Porto Alegre, março de 2003.

AGRADECIMENTOS

Em primeiro lugar, gostaria de agradecer aos meus orientadores Prof^o Robin T. Clarke e Vitor Haertel pela dedicação dispensada na elaboração deste trabalho e pela grandiosa colaboração na minha formação acadêmica. Desejo agradecer também ao Prof^o Jorge Ricardo Ducati pela sua imensa disponibilidade e compreensão.

Aos meus queridos pais pelo constante incentivo ao longo de toda a minha vida. Aos Coordenadores da PED-RMPA, Roberto Wiltgen e Lúcia Garcia pelo companheirismo e compreensão, principalmente nos momentos finais de elaboração deste trabalho, assim como agradeço a todos os demais pesquisadores, auxiliares e estagiários da FEE, DIEESE e FGTAS.

Aos meus colegas e amigos Diego, Rafael e Sérgio pelo companheirismo tanto nas atividades de trabalho como nas atividades de lazer.

Agradeço também a Camila pela leitura atenta do texto e por aturar meu mau humor. Por fim agradeço a todos meus familiares, amigos e colegas que me deram força e torceram por mim e a Deus que tornou este momento possível.

SUMÁRIO

1.INTRODUÇÃO	1
1.1 Conceitos Básicos	1
1.2 Sistemas Sensores	5
1.3 Objetivos do Trabalho	10
2. MÉTODOS EM RECONHECIMENTO DE PADRÕES	12
2.1 Tipologia em Reconhecimento de Padrões	12
2.1.1 Template Matching	12
2.1.2 Abordagem Sintática	13
2.1.3 Redes Neurais	14
2.1.4 Métodos que utilizam Lógica Nebulosa	16
2.2 Abordagem Estatística em Reconhecimento de Padrões	17
2.3 Análise de Componentes Principais como Redutor de Dimensionalidade	23
2.4 Classificadores na Abordagem Estatística	30
2.5 Análise de Dados Espaciais (Estatística Espacial)	33
2.5.1 Geoestatística (Análise de Dados no Espaço Contínuo)	35
2.5.2 Krigeagem	37
2.5.3 Semivariograma e Variograma – Ferramentas da Krigeagem	39
2.5.4 Principais Modelos Teóricos e Forma Geral do Semivariograma	42
3. METODOLOGIA PROPOSTA	45
3.1 Posicionamento do trabalho quanto às diversas áreas do conhecimento	45
3.2 Problemas na Utilização de Dados com Alta Dimensionalidade	46
3.3. Utilizando ferramentas da estatística espacial para reduzir dimensionalidade	48
3.4 Exemplo de aplicação da metodologia proposta	53

4. ESTUDO DE CASO - APLICAÇÃO EM DADOS REAIS DO SENSOR AVIRIS	64
4.1 Seleção de Bandas espectrais	64
4.2 Definição do número de partições e critérios dos locais de partições	64
4.3 Descrição das classes	66
4.4 Resultados dos Experimentos e Análise	67
4.4.1 Experimentos utilizando duas classes	68
4.4.1.1 Resultados utilizando matriz de covariâncias distintas para cada classe	69
4.4.1.2 Resultados assumindo matriz de covariâncias iguais para todas as classes	71
4.4.1.3 Análise dos resultados de classificação utilizando duas classes	73
4.4.2 Experimentos utilizando três classes	80
4.4.2.1 Resultados utilizando matriz de covariâncias distintas para cada classe	80
4.4.2.2 Resultados assumindo matriz de covariâncias iguais para todas as classes	81
4.4.2.3 Análise dos resultados de classificação utilizando três classes	83
4.4.3 Experimentos utilizando seis classes	85
4.4.3.1 Resultados utilizando matriz de covariâncias distintas para cada classe	85
4.4.3.2 Resultados utilizando matriz de covariâncias iguais para todas as classes	86
4.4.3.3 Análise dos resultados de classificação utilizando seis classes	87
5. CONCLUSÕES E RECOMENDAÇÕES	89
6. BIBLIOGRAFIA	93
7. ANEXOS	98

LISTA DE FIGURAS

Figura 1 - Espectro eletromagnético	3
Figura 2 - Assinatura espectral característica da água, solo e vegetação, localização dos canais do sensor LANDSAT-TM	5
Figura 3 - Comparação entre a informação espectral obtida nos sensores AVIRIS e LANDSAT	8
Figura 4 - A técnica dos sensores hiperspectrais	9
Figura 5 - Distribuição subjacente de 3 classes distintas - Exemplo unidimensional de regiões de decisão conexas	30
Figura 6 – Exemplo simulado de 3 classes normalmente distribuídas	31
Figura 7 – Regiões de decisão para o exemplo das 3 classes normalmente distribuídas	31
Figura 8 – Representação espacial das coordenadas “x” e “y”	41
Figura 9- Principais modelos de semivariograma, com respectivos parâmetros	43
Figura 10 - Simulação de um problema da Análise de Componentes Principais na discriminação de classes	48
Figura 11 – (a) Imagem do sensor AVIRIS bandas 17,27 e 50 (b) Imagem temática da verdade terrestre	63
Figura 12- Média amostral espectral para as diferentes classes	67
Figura 13- Função distribuição de probabilidade para duas classes	78

LISTA DE TABELAS

Tabela 1- Matriz de dados original para 1600 pixels nas 224 bandas do sensor AVIRIS	54
Tabela 2- Matriz de dados contendo as semivariâncias até lag 20 para 1600 pixels em uma subdivisão de 4 faixas de bandas do sensor AVIRIS	56
Tabela 3- Matriz de dados contendo os parâmetros do modelo de semivariograma linear em cada uma das quatro faixas de bandas para os 1600 pixels da amostra de treinamento	57
Tabela 4- Número de partições adotadas no estudo e características de cada partição	65
Tabela 5- Transecto de bandas por número de partições	65
Tabela 6-Quantidade de pixels por classe em amostra de treinamento e teste	66
Tabela 7- Acurácia obtida na classificação por classe em cada combinação por partição	70
Tabela 8- Acurácia ponderada observada em cada combinação por partição	71
Tabela 9- Acurácia média observada em cada classe independente da combinação, conforme número de partições	71
Tabela 10- Acurácia obtida na classificação por classe em cada combinação por partição,supondo matriz de covariâncias iguais	72
Tabela 11- Acurácia ponderada observada em cada combinação por partição, supondo matriz de covariâncias iguais	73

Tabela 12- Acurácia média observada em cada classe independente da combinação, conforme número de partições, supondo matriz de covariâncias iguais	73
Tabela 13- Valores observados para o determinante da matriz de covariâncias para cada classe em cada partição	76
Tabela 14- Número de vezes que o termo determinante da matriz de covariâncias para cada classe é maior em relação à classe corn min	76
Tabela 15- Matriz de confusão para as três classes de soja e acurácias por partição	81
Tabela 16- Matriz de confusão para as três classes de milho e acurácias por partição	81
Tabela 17- Matriz de confusão para as três classes de soja e acurácias por partição supondo matriz de covariâncias iguais	82
Tabela 18- Matriz de confusão para as três classes de milho e acurácias por partição, supondo matriz de covariâncias iguais	82
Tabela 19- Matriz de confusão para as seis classes estudadas e acurácias por partição	86
Tabela 20- Matriz de confusão para as seis classes estudadas e acurácias por partição, supondo matriz de covariâncias iguais	87

RESUMO

Com o advento dos sensores hiperespectrais se tornou possível em sensoriamento remoto, uma serie de diferentes aplicações. Uma delas, é a possibilidade de se discriminar classes com comportamentos espectrais quase idênticas. Porém um dos principais problemas encontrados quando se trabalha com dados de alta dimensionalidade, é a dificuldade em estimar os inúmeros parâmetros que se fazem necessários. Em situações reais é comum não se ter disponibilidade de tamanho de amostra suficiente, por exemplo, para se estimar a matriz de covariâncias de forma confiável. O sensor AVIRIS fornece uma riqueza de informações sobre os alvos, são 224 bandas cobrindo o espectro eletromagnético, o que permite a observação do comportamento espectral dos alvos de forma muito detalhada. No entanto surge a dificuldade de se contar com uma amostra suficiente para se estimar a matriz de covariâncias de uma determinada classe quando trabalhamos com dados do sensor AVIRIS, para se ter uma idéia é preciso estimar 25.200 parâmetros somente na matriz de covariâncias, o que necessitaria de uma amostra praticamente impraticável na realidade. Surge então a necessidade de se buscar formas de redução da dimensionalidade, sem que haja perda significativa de informação. Esse tipo de problema vem sendo alvo de inúmeros estudos na comunidade acadêmica internacional. Em nosso trabalho pretendemos sugerir a redução da dimensionalidade através do uso de uma ferramenta da geoestatística denominada semivariograma. Investigaremos se os parâmetros calculados para determinadas partições do transecto de bandas do sensor AVIRIS são capazes de gerar valores médios distintos para classes com comportamentos espectrais muito semelhantes, o que por sua vez, facilitaria a classificação/discriminação destas classes.

ABSTRACT

The advent of hyper-spectral sensors greatly extends the range of possible applications for remote-sensing methods. However one problem encountered is that of discriminating between classes having spectral signatures which are almost identical, as the existence of many spectral bands greatly increases the number of parameters to be estimated, and it commonly happens that the available samples of training data from the different classes are insufficient to allow reliable estimation of within-class dispersion matrices. The AVIRIS sensor supplies a wealth of information with 224 bands covering the electromagnetic spectrum, from which it should be possible to discriminate between classes of pixel having near-identical spectral behavior. However when using data from the AVIRIS sensor the difficulty arises of how to obtain a sufficiently large sample sample to estimate the covariance matrix of a given class: the covariance matrix of just one class requires the estimation of 25,200 parameters, requiring a very large sample which in practice is not possible. It is therefore necessary to look for ways of reducing the dimensionality of the problem without significant loss of information. In this work, we explore the possibility of reducing the dimensionality of the problem by using geostatistical methods to parameterize the covariance matrix in terms of semi-variograms. We investigate the semi-variogram parameters that result from different partitions of the sequence of AVIRIS spectral bands, and show that they are capable of distinguishing between classes with very similar spectral behaviors.

1.INTRODUÇÃO

1.1 Conceitos Básicos

Sensoriamento remoto pode ser entendido como a tecnologia que permite a captação, armazenamento e interpretação de informações sobre características de objetos na superfície terrestre, sem que seja necessário o contato físico com eles, com o intuito de estudar as interações entre atmosfera, solo e oceanos. Atualmente séries de sensores a bordo de satélites estão constantemente captando informações de alvos localizados na superfície terrestre.

Os principais produtos do sensoriamento remoto são imagens multispectrais, ou seja, de uma mesma área são obtidas várias imagens em diferentes faixas do espectro eletromagnético. Quando se fala em sensoriamento remoto e sistema de sensores, logo se tem em mente a alta tecnologia associada e a constante evolução dessas tecnologias.

As principais propriedades de um sistema sensor são: resolução espectral, resolução espacial, resolução radiométrica e resolução temporal. Essas três propriedades servem como critérios para a escolha de qual sistema sensor melhor se adapta a um determinado tipo de problema e objetivo que se tem no monitoramento de áreas.

Resolução Espectral: O Espectro eletromagnético é o conjunto dos diferentes tipos de radiação eletromagnética. O espectro eletromagnético abrange um amplo intervalo de comprimentos de onda (e conseqüentemente de freqüências) indo desde ondas muito longas, e conseqüentemente baixas freqüências (rádio), passando pelas ondas infravermelho e ondas de luz visível até chegar a freqüências muito altas (o que corresponde a comprimentos de ondas curta) dos raios gama e raios X. Os comprimentos de onda no intervalo da luz visíveis têm uma cor específica associada a cada um deles quando eles passam através de um prisma. As freqüências mais baixas produzem a cor vermelha enquanto que as freqüências mais altas, comprimentos de ondas mais curtas, produzem a cor violeta. Aqueles comprimentos de onda entre estes dois pontos produzem as cores laranja, amarelo, verde, e azul.

Em sensoriamento remoto é comum definir intervalos do espectro eletromagnético nos quais os sensores são sensíveis. Quanto maior for o número de intervalos e menor a largura deles, mais rica e refinada serão as informações sobre os alvos imageados e, portanto, maior será a resolução espectral. Essas faixas, em sensoriamento remoto são chamadas de **bandas espectrais**.

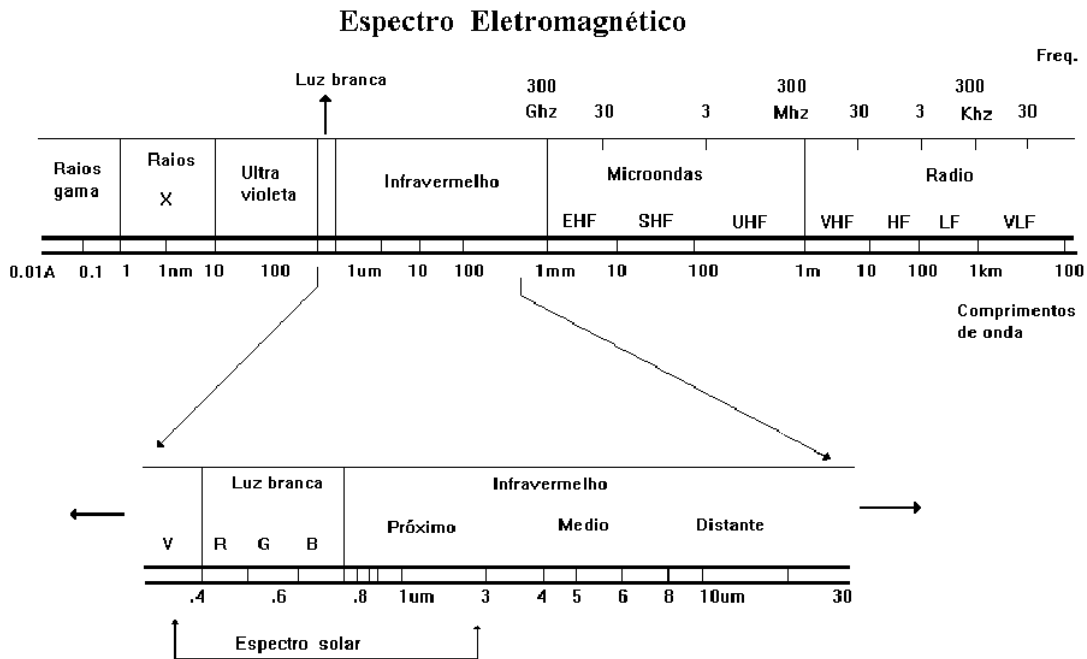


Figura 1- Espectro eletromagnético

Resolução Espacial: A resolução espacial de uma imagem digital se relaciona com o poder de definição, ou seja, o detalhe de observação passível de detecção. A resolução espacial está atingindo a ordem do metro e tornando possível uma série de aplicações antes inimagináveis.

Resolução Radiométrica: A resolução radiométrica de um sistema sensor esta baseada na sensibilidade de distinguir dois diferentes níveis de intensidade do sinal. Quanto maior for a capacidade de distinção, mais alta será a resolução radiométrica, por exemplo, uma resolução de 10 bits (1024 níveis) é mais alta do que uma resolução de 8 bits (256 níveis).

Resolução Temporal: A resolução temporal se refere à frequência com que são obtidas imagens de um determinado local por um determinado sistema sensor. Quanto

menor for o espaço de tempo entre o monitoramento de uma mesma área, maior será a resolução temporal.

Os diferentes sistemas sensores hoje disponíveis apresentam diferentes propriedades quanto às resoluções explicitadas acima. Dependendo do tipo de problema associado, um determinado tipo de sistema será adequado. Em muitas situações não será preciso e nem será adequado utilizar aquele sistema que possua melhores resoluções.

Os sensores captam, medem e registram a magnitude da **radiância** eletromagnética proveniente do sol, refletida e/ou emitida por um alvo na superfície da Terra, em cada faixa do espectro eletromagnético referente a cada banda, posteriormente é realizada a discretização da magnitude da radiância em níveis de contadores digitais (CD) e em termos espaciais a discretização se dá na forma de pixel na imagem, que se refere a uma determinada unidade de área na superfície terrestre.

Uma imagem digital no formato “raster” ou “matricial” é formada por uma série de pixels (n), para cada existe um número (p) de atributos medidos, no caso os contadores digitais (CD) que expressam as radiâncias observadas e discretizadas nas diferentes bandas ou faixas do espectro eletromagnético.

A reflectância de um alvo observada ao longo do espectro eletromagnético é conhecida como sendo sua **assinatura espectral** ou **curva de resposta espectral**. Na prática dispõe-se somente de uma aproximação dessa curva, formada pelas respostas obtidas em cada uma das bandas espectrais disponíveis. Quanto maior for o número de bandas e por conseqüência menor a sua largura, mais detalhadas serão as informações espectrais sobre o alvo, aproximando-se da sua verdadeira assinatura espectral.

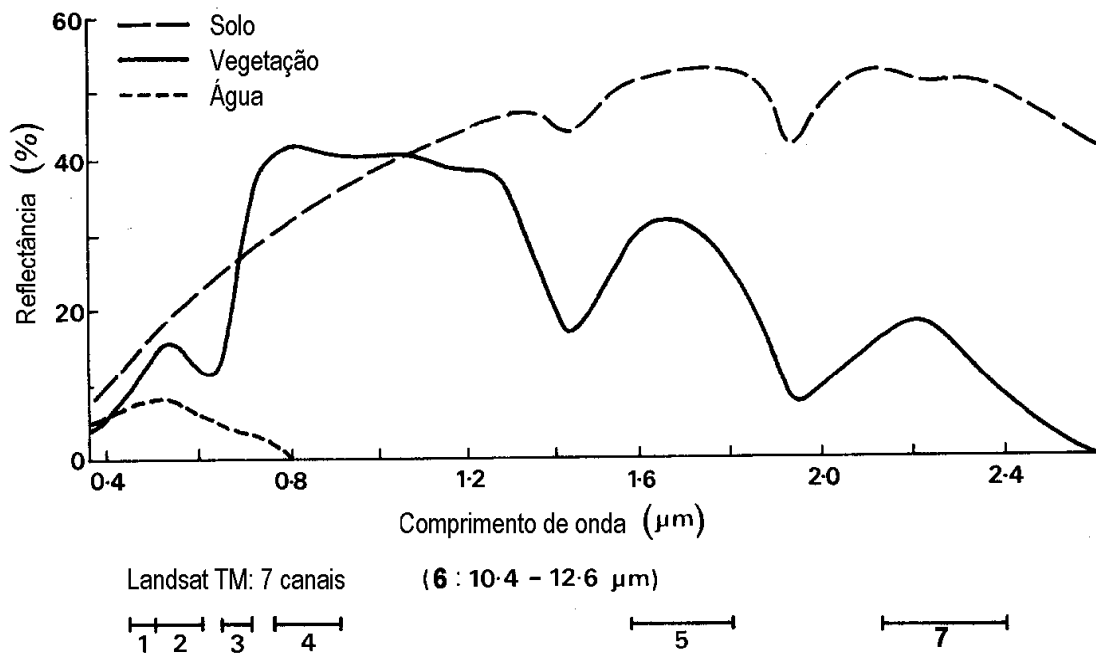


Figura 2- Assinatura espectral característica da água, solo e vegetação, localização dos canais do sensor LANDSAT-TM.

1.2 Sistemas Sensores

Como foi mencionado anteriormente, existem diversos **sistemas sensores** em órbita extraíndo informação da superfície terrestre. Estes sensores apresentam especificidades próprias nas quais se diferenciam dos demais. Um dos sistemas sensores mais conhecidos é o **LANDSAT**. Foram lançados até hoje sete satélites do programa LANDSAT, desde 1972. Atualmente ativos se encontram o LANDSAT 5 e o LANDSAT 7. O LANDSAT 5 também conhecido como TM (Thematic Mapper) dispõe de um sistema sensor que cobre o espectro eletromagnético de 0,45 até 12,6 μm sendo esta cobertura não contínua, dividida em 7 faixas (bandas), 3 delas na faixa do visível e 4 na faixa do infravermelho, a resolução espacial é de 30m exceto para a banda 6 (região do infravermelho termal) que é de 120m.

A faixa de varredura do sistema é de 185 x 185Km e sua resolução radiométrica é de 8 bits, o que permite a detecção de 256 níveis de intensidade de radiação, o satélite demora 16 dias para dar uma volta completa na órbita terrestre e então novamente captar informação de uma mesma área.

O sistema LANDSAT-TM, bem como vários outros em uso, obtém um número pequeno de bandas espectrais, não fornecendo desta forma dados em alta resolução espectral.

Mais recentemente, com o avanço da tecnologia tornou-se viável a disponibilidade de sistemas sensores de muito alta resolução espectral, isto é, sistemas com um número muito alto de sensores, operando em faixas estreitas e contíguas do espectro. Estes sistemas, denominados de “hiperspectrais” abriram novas possibilidades em sensoriamento remoto.

Um dos sistemas hiperspectrais mais conhecidos, ainda do tipo aero-transportado, é o sistema **AVIRIS** (Airborne Visible/Infrared Imaging Spectrometer), possuindo 224 bandas que cobrem o espectro eletromagnético de 0,41 a 2,45 μm . Este sistema sensor apresenta uma resolução espectral mais alta do que o LANDSAT-TM, pois é composto de bandas mais estreitas, o que permite maior aquisição de informação dos alvos de forma mais detalhada do ponto de vista espectral. Também possui maior resolução espacial 20m, porém o total de área varrida é de aproximadamente 11Km de extensão. Sua resolução radiométrica atualmente é de 12 bits.

Portanto, o sensor AVIRIS é classificado como hiperspectral, enquanto que o LANDSAT é multispectral. Outra diferença essencial entre esses dois sistemas sensores é

que o AVIRIS utiliza como plataforma uma aeronave enquanto que o LANDSAT é instalado em um satélite a 705 Km de altitude.

Destacam-se ainda os sensores da série **SPOT** (Satellite Pour l'Observation de la Terre) que apresenta atualmente resolução espacial de 2,5 m e que se encontra na sua versão de número 5, outro sistema importante e muito recente é chamado **IKONOS**, a bordo de um satélite a uma altitude de aproximadamente 680Km, é capaz de adquirir imagens com resolução espacial de até 1m no pancromático e 4m nas bandas multispectrais.

No que se refere à resolução espectral é correto afirmar que quando um sistema sensor apresenta faixas espectrais consideravelmente largas, estamos perdendo ou deixando de captar uma série de informações relevantes sobre o alvo. Os valores observados em faixas largas, são na verdade uma aproximação da média da radiância na faixa do espectro correspondente a banda.

Em reconhecimento de padrões, por exemplo, muitas vezes trabalhamos com duas ou mais classes que possuem comportamentos espectrais muito semelhantes, nesse caso, é praticamente impossível uma classificação satisfatória quando utilizamos poucas bandas que cobrem faixas largas do espectro eletromagnético, como LANDSAT-TM; Para esses casos é necessária a utilização de dados provenientes dos sensores hiperspectrais, como AVIRIS.

Como podemos observar na figura 3, os dados provenientes do sensor AVIRIS (hiperspectral) fornecem uma quantidade maior de informação sobre o alvo imageado do

que os dados obtidos pelo sensor LANDSAT (multispectral), devido à quantidade de bandas existentes em cada sistema sensor.

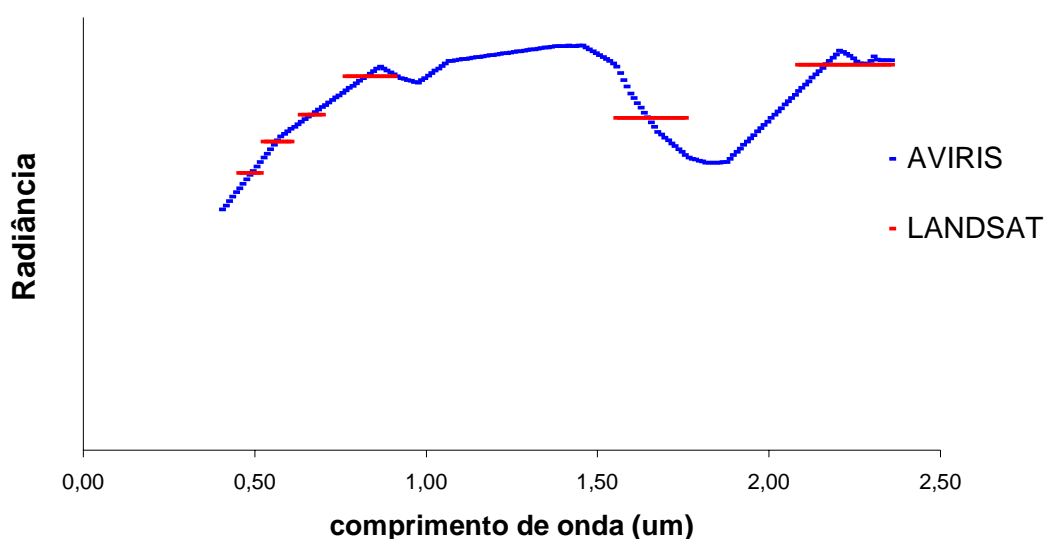


Figura 3 - Comparação entre a informação espectral obtida nos sensores AVIRIS e LANDSAT

A informação espectral de uma cena pode ser representada em cada banda individualmente por uma imagem em níveis de cinza, onde a posição do pixel é dada pelas coordenadas “x” e “y” (linha e coluna respectivamente) juntamente com a variável “z” que representa a magnitude da radiância refletida e/ou emitida discretizada em contadores digitais, pela correspondente porção no terreno. No caso de imagens multispectrais, a variável “z” aumenta sua dimensionalidade, assumindo a forma de um vetor, no qual cada componente está associado a uma determinada faixa do espectro eletromagnético em particular. A representação gráfica ao longo dos diferentes intervalos de comprimento de onda (bandas) pode ser entendida como uma aproximação da curva de assinatura espectral do pixel, cada alvo em particular apresenta uma curva própria de assinatura espectral.

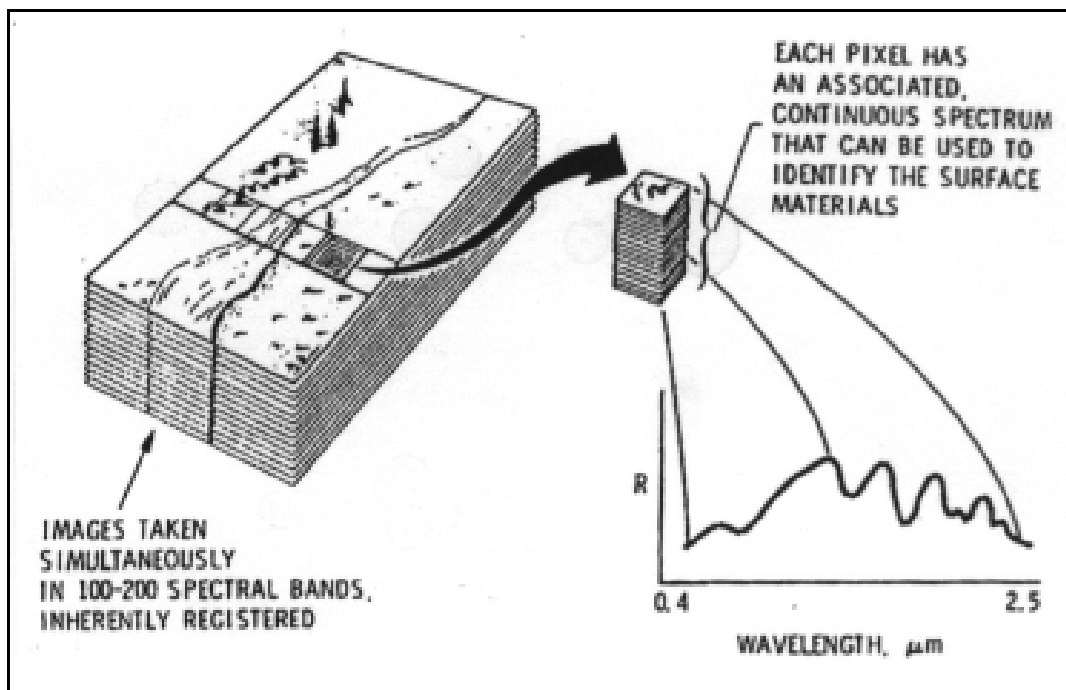


Figura 4- A técnica dos sensores hiperspectrais.

Através da análise da assinatura espectral é possível discriminar do ponto de vista espectral alvos distintos, segundo o comportamento dos fenômenos de absorvância, reflectância e transmitância na interação do alvo com a radiação eletromagnética.

Do ponto de vista estatístico podemos entender cada pixel em uma imagem como sendo uma observação de uma determinada área localizada na superfície terrestre. As radiâncias medidas nas diferentes bandas do espectro eletromagnético são entendidas como sendo as variáveis aleatórias que foram observadas e mensuradas nas correspondentes áreas (pixels).

Os sensores registram a quantidade de fótons refletidos e emitidos por uma área na determinada faixa do espectro eletromagnético que se refere à banda. Essas quantidades

expressam uma estimativa da radiância média da área para a determinada faixa do espectro eletromagnético.

Temos que salientar que os valores observados pelos sensores devem ser interpretados como uma estimativa da radiância verdadeira, e não como a própria. Os valores observados pelos sensores são influenciados por uma série de fatores aleatórios como vento, umidade, tipo de solo, provocando assim uma dispersão nos valores associados à pixels de uma mesma classe. Sendo assim, podemos evidenciar que existem componentes aleatórios associados a esses valores, surgindo a necessidade da utilização de uma abordagem estatística na interpretação e análise dessas imagens.

1.3 Objetivos do Trabalho

O nosso trabalho se encontra inserido dentro de uma grande área denominada reconhecimento de padrões em imagens digitais. Mais precisamente, estaremos utilizando técnicas envolvidas na abordagem estatística. Nosso problema principal é a classificação de pixels em uma imagem hiperespectral. Iremos sugerir a utilização de técnicas e ferramentas da geoestatística como um possível redutor da dimensionalidade de problemas quando não é possível a utilização das técnicas já existentes ou quando estas não trazem resultados satisfatórios.

O problema da redução da dimensionalidade dos dados pode ser abordado de duas formas distintas: seleção de feições e extração de feições. No primeiro caso é explorada a alta correlação existente entre as bandas espectrais e empregam-se critérios que selecionam as bandas que mais contribuem para a separabilidade entre as classes. O segundo caso, implementa métodos para combinação (linear ou não) entre as bandas, criando desta forma componentes “sintéticas”. Este processo permite então a utilização de um número mais

reduzido de variáveis com perda mínima de informação. O nosso trabalho estará inserido dentro do segundo caso, extração de feições.

Enumeraremos os principais objetivos do nosso trabalho através de uma série de questões preliminares:

- 1) É possível utilizar ferramentas da geoestatística com o intuito de reduzir a dimensionalidade em dados provenientes do sensor AVIRIS?
- 2) Os parâmetros obtidos através da modelagem de semivariogramas teóricos são eficazes para discriminar classes com comportamento espectral semelhantes?
- 3) Se os parâmetros do semivariograma calculados para cada pixel se mostrarem regidos por uma distribuição conhecida, é possível utilizar algum classificador já consagrado e obter resultados satisfatórios?
- 4) Qual modelo de semivariograma se ajusta melhor aos dados de contadores digitais observados pelas bandas do sensor AVIRIS?

2. MÉTODOS EM RECONHECIMENTO DE PADRÕES

2.1 Tipologia em Reconhecimento de Padrões

De forma geral, podemos entender o reconhecimento de padrões como um conjunto de técnicas capazes de identificar e classificar características em categorias a partir de informações. Categorias podem ser entendidas como entidades ou padrões de informações que apresentam similaridades (classes).

Existem diferentes tipos de abordagens em reconhecimento de padrões, segundo Jain et. al.,2000, os diferentes problemas envolvidos em reconhecimento de padrões podem ser classificados em: *Template Matching*; Abordagem Sintática; Redes Neurais; Métodos que utilizam Lógica Nebulosa e Abordagem Estatística. Este trabalho está compreendido unicamente dentro da abordagem estatística, mesmo assim optamos por definir as outras abordagens e métodos de uma forma geral, explicitando os diferentes tipos de problemas e de aplicações envolvidos em cada uma dessas outras abordagens.

2.1.1 Template Matching

A mais antiga das abordagens em reconhecimento de padrões, e dessa forma também a mais simples. É baseado em modelos de protótipos que contenham os padrões a serem reconhecidos, dispõe-se de padrões que são submetidos a uma operação genérica, e então é medida a similaridade entre os padrões e os protótipos.

O padrão a ser reconhecido é comparado contra um modelo armazenado, sendo considerado na similaridade qualquer tipo de variação possível na estrutura do padrão, seja de rotação, translação ou até diferença de escala e deformação.

A medida de similaridade, normalmente relacionada à idéia de correlação, deve ser otimizada com base no conjunto de treinamento. A maior desvantagem de template matching está em seu alto custo computacional.

No caso do uso de templates rígidos, ocorrem falhas ao reconhecer padrões distorcidos ou com grande variação intra-classe. Os modelos de templates deformáveis podem ser usados quando as deformações não podem ser modeladas diretamente.

2.1.2 Abordagem Sintática

Trata-se de uma abordagem em que os padrões são tratados sob uma perspectiva hierárquica, ou seja, um padrão é composto por sub-padrões mais simples, os quais também são compostos por outros mais simples. Os sub-padrões elementares são chamados de primitivas, e um dado padrão complexo é representado em termos de inter-relações entre essas primitivas. Uma analogia formal pode ser traçada com a sintaxe de uma linguagem em que se pode ver os padrões como sentenças pertencentes a uma linguagem, as primitivas como seu alfabeto, e as sentenças são geradas de acordo com a respectiva gramática.

A gramática para cada classe de padrões deve ser inferida a partir dos exemplos de treinamento. Uma vantagem de reconhecimento estrutural de padrões está no fato de que, além de fazer a classificação, essa abordagem também provém uma descrição de como os

padrões são gerados a partir de primitivas que podem ser repetidas ou obedecerem determinadas ordens no sinal.

Um exemplo de modelo de classificação sintática dependente do contexto em que se observa uma ascensão recente é o de *Hidden Markov Models* (HMM). Trata-se de um processo estocástico finito que é composto por uma cadeia de observações, isto é, a seqüência de vetores observados. Uma HMM consiste de um número finito de estados, e a cadeia observada é produzida como o resultado de sucessivas transições de um estado para outro, cada transição ocorre com uma certa probabilidade, que é determinada na fase de treinamento. O problema é que nem sempre dispomos de informação suficiente sobre os estados e com isso surge dificuldade na determinação do número de estados a priori.

Este tipo de abordagem é principalmente utilizado na área de reconhecimento de vozes, mas também tem aplicação importante em reconhecimento de caracteres, e de eventos temporais em vídeo, como gestos e expressões. As desvantagens da abordagem sintática estão na necessidade de, inicialmente, segmentar os padrões de ruídos para detectar primitivas, e na inferência de gramática dos dados de treinamento, pois essa abordagem pode permitir um grande número de elementos combinatórios, das possibilidades a serem investigadas, o que demandaria um grande esforço computacional.

2.1.3 Redes Neurais

Segundo Campos (2001), uma rede neural artificial pode ser entendida como um modelo de grafo orientado disposto em uma rede em que os nós representam neurônios artificiais e as arestas orientadas são as conexões entre as entradas e saídas dos neurônios. Redes neurais podem ser vistos como máquinas paralelas com muitos processadores simples e muitas interconexões.

As redes neurais podem ser hábeis a aprender relações entre entradas e saídas complexas e não lineares. Dentre suas características, inclui-se a possibilidade de usar procedimentos seqüenciais de treinamento, e o poder de adaptação aos dados. Destacamos dentre as várias famílias de redes neurais, a *feed-forward network*, que inclui *perceptron* multicamadas e redes funções de bases radiais (*radial-basis function - RBF*). Essas redes são organizadas em camadas, possuindo conexões unidirecionais entre uma camada e outra. Outra classe popular de redes é a dos mapas auto organizáveis (*self-organizing maps - SOM*), ou redes de Kohonen, as quais são usadas principalmente para agrupamento de dados (*clustering*) e em mapeamento de características.

O processo de aprendizado envolve atualização da arquitetura da rede e dos pesos das conexões, de forma que a rede possa efetuar tarefas específicas de classificação e análise de agrupamentos (*clustering*). A popularidade de redes neurais cresceu devido ao fato que, elas tem uma baixa dependência a um específico domínio, de forma que o mesmo tipo de rede pode ser aplicado em vários problemas, o que não ocorre com as abordagens baseadas em modelos ou em regras. Além disso, existem muitos algoritmos eficientes de aprendizado com redes neurais.

As redes neurais disponibilizam um novo conjunto de algoritmos não lineares para extração de características (usando camadas escondidas) e classificação (p. ex. *mult-layer perceptrons*). Além disso, algoritmos de extração de características podem ser mapeados em arquiteturas de redes neurais para implementação eficiente em hardware. Apesar de possuírem aparentemente princípios diferentes, a maioria dos modelos de redes neurais são implicitamente equivalentes ou similares a métodos clássicos de reconhecimento estatístico

de padrões. Entretanto, redes neurais oferecem várias vantagens, para extração e classificação, e procedimentos flexíveis para encontrar soluções não lineares.

2.1.4 Métodos que utilizam Lógica Nebulosa

A lógica nebulosa foi criada com inspiração no comportamento humano, que se baseia na interpretação do mundo sem precisão e na descrição desse por atributos lingüísticos. A partir de conjuntos não-nebulosos, é possível construir uma técnica de decisão objetiva baseada em critérios subjetivos que é apropriada para reproduzir decisões tomadas por indivíduos nas mais extensas áreas da atividade humana. A partir dos critérios *crisp* é possível definir critérios de decisão em conceitos como “quente”, “frio”, “escuro”, “claro”, etc. Esses são conceitos subjetivos que estão relacionados a estados físico-ambientais em que os sentidos humanos são capazes apenas de atribuir um valor qualitativo. Estes conjuntos de técnicas são popularmente conhecidos como fuzzyficação.

Através da determinação do grau de pertinência de cada um desses elementos aos conjuntos são geradas funções de pertinência que são baseadas na relação entre cada elemento e o(s) suporte(s) dos conjuntos. Essa relação pode ser, por exemplo, inversamente proporcional à distância entre os elementos e os suportes. Um suporte pode ser determinado, pelo centróide do conjunto de treinamento, entre outros. A classificação em si utilizando lógica nebulosa normalmente é feita através de regras nebulosas e métodos de defuzzificação. As regras nebulosas normalmente são regras de inferência que utilizam delimitadores dos valores de pertinência de cada padrão aos conjuntos. Uma maior explicação sobre a lógica fuzzy e suas inúmeras aplicações podem ser vistos em Didier (1997).

2.2 Abordagem Estatística em Reconhecimento de Padrões

Segundo Duda and Hart (1973), um sistema de reconhecimento estatístico de padrões pode ser composto basicamente pelas seguintes partes: um sistema de aquisição de dados (por exemplo: sensores ou câmeras); um sistema de pré-processamento, para eliminar ruídos ou distorções; um extrator de características (ou atributos), que cria um vetor de características com dados extraídos dos objetos adquiridos, e um seletor de características, que analisa o conjunto e elimina as mais redundantes; além de um classificador, que analisa um padrão obtido para tomar uma certa decisão.

Na abordagem estatística, cada padrão é representado em termos de d feições (*features*) ou características, sendo visto como um ponto em um espaço d -dimensional. O objetivo é que tal espaço seja gerado de forma que suas características permitam que vetores de padrões pertencentes a categorias diferentes fiquem dispostos de maneira que o processo de classificação seja simples. É desejável que estes vetores ocupem regiões compactas e disjuntas em um espaço de d dimensões.

Cada padrão é uma observação obtida aleatoriamente segundo uma certa probabilidade condicional a uma determinada classe. Dado um conjunto de treinamento contendo amostras de todas as classes, o objetivo é estabelecer fronteiras de decisão no espaço das características que separem os padrões de classes diferentes.

Na teoria de decisão estatística, a fronteira de decisão é determinada para cada classe, que pode ser especificada (aprendizado supervisionado) ou aprendida (aprendizado não-supervisionado).

Nos métodos de classificação supervisionada, o primeiro passo acontece no sentido de escolher a forma geral da função decisão, que pode ser da forma determinística ou probabilística. Após o analista especificar as classes existentes na imagem são selecionadas amostras de treinamento representativas de cada classe, que servirão para estimar os coeficientes dos parâmetros envolvidos na forma geral escolhida. Fukunaga (1990) destaca que no processo supervisionado, nem sempre se sabe exatamente o número de classes envolvidas em uma cena, daí a necessidade da utilização do *threshold*, que permite a não classificação de alguns pixels a nenhuma classe.

Os métodos não-supervisionados são utilizados quando previamente não temos informação relevante e suficiente sobre a natureza do problema, seja a definição do número de diferentes classes existentes em uma imagem, seja a localização de amostras representativas das classes. No processo não-supervisionado, ainda existe a componente heurística, entre outras determinações, o analista precisa fornecer o número de classes existentes (ex: K-Means) e outros parâmetros (ex: ISODATA).

Nesse tipo de situação como não podemos identificar amostras de treinamento, o classificador ou algoritmo tem a tarefa de identificar, por si só, as classes existentes em uma imagem, estes algoritmos se baseiam principalmente em uma técnica estatística multivariada conhecida como análise de agrupamento (cluster analysis).

Os métodos de classificação em reconhecimento estatístico de padrões, também podem ser divididos em dois grandes grupos: métodos paramétricos e métodos não-paramétricos.

Os métodos paramétricos exigem um determinado conhecimento à priori sobre a distribuição subjacente dos dados das diferentes classes existentes em uma imagem. É necessário conhecer de forma geral a distribuição probabilística que melhor se ajusta aos dados em que se está trabalhando. A partir de amostras de treinamento são estimados os parâmetros do modelo para cada classe, e assim de posse da distribuição supostamente conhecida e dos parâmetros estimados para cada classe, temos uma aproximação da verdadeira função distribuição de probabilidade que rege cada classe.

Quando não temos conhecimento sobre a verdadeira distribuição dos dados e não podemos supor um modelo que represente bem aos mesmos, é mais prudente e razoável utilizarmos técnicas estatísticas não paramétricas, pois estas são robustas e, portanto não necessitam de premissas a cerca da verdadeira distribuição de dados, podendo ser utilizadas para diferentes tipos de distribuições de forma indiscriminada. Por exemplo, método de Parzen e método dos k-vizinhos mais próximos.

Na abordagem estatística em reconhecimento de padrões existe um conjunto de técnicas de classificação já consagradas, que podem ser utilizadas conforme a situação do problema de pesquisa. Para técnicas de classificação supervisionada com métodos paramétricos podemos destacar a grande importância do classificador “Bayesiano” que apresenta soluções ótimas em muitos casos. A estratégia deste classificador é que supor que são conhecidas as distribuições dentro de cada classe $p(x/w_i)$ e que dispomos de informação sobre a probabilidade a priori de cada classe. Com base nesses conhecimentos é possível utilizar o teorema de Bayes para estabelecer a probabilidade à posteriori de um certo padrão x pertencer a classe W_i .

$$\text{Teorema de Bayes} \rightarrow P(w_i / x) = \frac{P(w_i) \cdot P(x / w_i)}{\sum_{j=1}^K P(w_j) \cdot P(x / w_j)} \quad \forall i, j \in (1, K)$$

Onde,

$P(w_i)$ é a probabilidade a priori de um padrão pertencer a classe w_i ;

$P(x / w_i)$ é a função densidade de probabilidade da classe w_i ;

$P(w_i / x)$ é a probabilidade a posteriori do padrão x pertencer a classe w_i e

K é o número de classes distintas.

Os classificadores supervisionados nos métodos paramétricos utilizam amostras de treinamento, que são regiões selecionadas, onde já são conhecidas as verdadeiras classes a que pertencem os pixels da determinada região. A partir das amostras de treinamento de cada classe são estimados os parâmetros de posição e de variabilidade para cada uma das classes, no espaço de feições. Com base nesses parâmetros (vetor média e matriz de covariâncias) além da suposição de os dados serem regidos pela distribuição gaussiana, é possível estimar a função distribuição de probabilidade subjacente ou populacional de cada classe da selecionada.

A partir da função distribuição de probabilidade observada na amostra para cada classe, é perfeitamente possível estabelecer uma série de fronteiras de decisões entre cada classe, privilegiando a classificação de cada novo pixel naquela classe onde a função de verossimilhança for maximizada.

Quanto maior for a quantidade de atributos (número de variáveis, no caso de imagens digitais, as bandas) medidos e contidos em cada padrão, maior será a dimensionalidade do problema e, por conseguinte o número de parâmetros a serem

estimados. Os parâmetros a serem estimados na matriz de covariâncias se tornam o maior problema quando trabalhamos com problemas de alta dimensionalidade, sendo necessário então uma grande quantidade de amostras de treinamento.

Esse se torna um dos maiores problemas no uso de técnicas estatísticas paramétricas tanto em reconhecimento de padrões como em qualquer outra área de conhecimento em que se faça necessária a utilização da estatística inferencial, pois em muitos casos existe uma dificuldade muito grande de dispomos de um tamanho de amostra suficiente grande e representativa para obtermos estimativas confiáveis para um grande número de parâmetros envolvidos no trabalho, sendo necessária à busca de uma simplificação do conjunto de dados a partir de uma diminuição na dimensionalidade do problema.

É verdadeiro dizer que a alta dimensionalidade dos dados em muitos casos se torna simultaneamente tanto em benefício quanto em problema. Essa matéria tem sido objeto de muitas pesquisas por parte da comunidade científica internacional, com o intuito de minimizar os problemas relacionados e decorrentes do grande número de parâmetros a serem estimados quando dispomos de dados de alta dimensionalidade.

Cabe ressaltar que o aumento da dimensionalidade dos dados resulta em melhoria no entendimento do problema e esperamos, por conseguinte, uma melhoria nos indicadores de acurácia do classificador, porém isto não ocorre de forma linear e crescente, sempre existe um determinado ponto, onde o aumento da dimensionalidade já não traz mais “ganho” de informação e sim um confundimento devido ao crescente aumento de parâmetros a serem estimados a medida que aumenta o número de variáveis associadas em

um estudo, resultando numa diminuição na acurácia do classificador. Esse efeito é conhecido como fenômeno de Hughes ou maldição da dimensionalidade.

A diminuição da dimensionalidade se torna em recurso único muitas vezes. É necessário sabermos que sempre que necessitamos reduzir a dimensionalidade de um problema, estamos por certo perdendo informação, o que devemos nos preocupar é que essa "perda" seja mínima.

Ao selecionamos apenas um conjunto de atributos de uma totalidade, devemos optar sempre por aqueles que expliquem e mantenham quase toda a variabilidade do conjunto de dados original, ou seja, é desejável que os dados resultantes de uma simplificação devido a algum procedimento de diminuição da dimensionalidade sejam representativos e fiéis ao conjunto de dados original.

Uma técnica estatística muito utilizada com o intuito de reduzir a dimensionalidade de problemas é a ACP (Análise de Componentes Principais) que é reconhecida e utilizada nas mais diversas áreas do conhecimento, esta técnica trabalha com a ortogonalização do conjunto de dados, criando componentes que preservam muito da variabilidade do conjunto de dados original já nas primeiras componentes, obtendo-se um ganho extremamente precioso com a redução da dimensionalidade sem que haja perda significativa de informação com essa redução.

Mais adiante, será melhor explicada a técnica de Análise de Componentes Principais, seus fundamentos e os principais problemas verificados na utilização desta em reconhecimento de padrões em imagens digitais.

Os objetivos do nosso trabalho se referem ao problema da redução da dimensionalidade e busca evidenciar novas formas de redução. As imagens do sensor AVIRIS apresentam 224 bandas, e neste caso é inevitável buscarmos uma forma de redução na dimensionalidade do problema, devido à dificuldade em estimar os dois primeiros momentos (vetor média e matriz de variâncias e covariâncias), principalmente pelo fato do segundo momento apresentar um número muito grande de parâmetros associados.

Com isso surge a necessidade de buscarmos métodos capazes de reduzir a quantidade de parâmetros a serem estimados. Nosso trabalho pretende sugerir uma técnica envolvendo modelos teóricos de semivariograma que descreva as características correlacionais da série de bandas das determinadas classes existentes na imagem. A partir da função correlacional será possível reduzir o número de variáveis e ao mesmo tempo o número de parâmetros a serem estimados nos dois primeiros momentos a partir do conjunto de dados.

Podemos dizer que objetivo principal do trabalho é averiguar se tal técnica proposta é capaz de trazer resultados satisfatórios na classificação de padrões, quando trabalhamos com dados do sensor AVIRIS e selecionamos classes distintas na imagem, mas com comportamento espectrais semelhantes. Mais adiante detalharemos a metodologia proposta e o conjunto de diferentes técnicas e áreas do conhecimento envolvidas neste trabalho.

2.3 Análise de Componentes Principais como Redutor de Dimensionalidade

A análise de componentes principais PCA (do inglês Principal Component Analysis), é um método capaz de decompor uma matriz de dados X de posto r como uma

soma de matrizes de posto igual a um, onde o posto é o número que expressa a dimensão de uma matriz.

O procedimento consiste na ortogonalização da matriz de dados original, tendo partida na matriz de variâncias e covariâncias ou na matriz de correlação. Quando as variáveis do conjunto de dados não são mensuradas sob uma mesma unidade de medida é recomendável utilizarmos somente a matriz de correlação para extração das componentes principais.

Com a ortogonalização da matriz de dados originais, estamos criando novas variáveis que são chamadas de “componentes” que não apresentam correlação entre si. Teoricamente o número de Componentes é sempre igual ao número de variáveis originais, Entretanto um número mais restrito de componentes é responsável por grande parte da explicação total.

Os componentes são extraídos na ordem do mais explicativo para o menos explicativo, sendo que o primeiro componente explica o máximo possível da variação do conjunto de dados original, e cada sucessivo componente explica o máximo possível da variação restante não explicada pelos componentes anteriores.

Os componentes principais são entendidos como combinações lineares das variáveis do conjunto de dados original. As projeções da amostra na direção da componente principal são conhecidas como “scores”.

Os autovetores são os coeficientes de cada um dos componentes principais (funções lineares) e os autovalores representam a quantidade da variância que é explicada pelas respectivas componentes principais.

A análise de componentes principais pode ser resumida na seguinte equação matricial:

$$L = V' \Sigma V$$

Onde Σ é a matriz de covariâncias também poderia ser a matriz de correlação, V é a matriz dos autovetores (eigenvectors) e L a matriz dos autovalores (eigenvalues).

Os autovalores são as raízes não-imaginárias da equação polinomial definida por $|\Sigma - \lambda I| = 0$, onde λ é o autovalor e I é uma matriz identidade.

Seja o conjunto de dados $X' = [X_1, X_2, \dots, X_p]$ com uma matriz de variâncias e covariâncias Σ , Sendo $\lambda_1, \lambda_2, \dots, \lambda_p$ os autovalores dessa matriz, com $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$.

Como os componentes principais nada mais são do que combinações lineares das variáveis do conjunto de dados original, podemos representa-las da seguinte forma, segundo a notação utilizada por Johnson and Wichern (1982).

$$Y_1 = l'_1 X = l_{11}X_1 + l_{12}X_2 + \dots + l_{1p}X_p$$

$$Y_2 = l'_2 X = l_{21}X_1 + l_{22}X_2 + \dots + l_{2p}X_p$$

.....

$$Y_p = l'_p X = l_{p1}X_1 + l_{p2}X_2 + \dots + l_{pp}X_p$$

Sendo assim,

$$Var(Y_i) = l'_i \Sigma l_i, \text{ para } i=1,2,\dots,p$$

$$Cov(Y_i, Y_k) = l'_i \Sigma l_k, \text{ para } i \neq k$$

Como é desejável maximizar $Var(Y_1)$, chegamos a uma indeterminação, pois quanto maior for l_1 maior será $Var(Y_1)$, portanto é necessário impor a restrição $l'_1 l_1=1$, fazendo com que os autovetores sejam únicos e normalizados.

Então surge que:

$$Var(Y_1) = l'_1 \Sigma l_1, \text{ sendo que } l'_1 l_1=1$$

$$Var(Y_2) = l'_2 \Sigma l_2, \text{ sendo que } l'_2 l_2=1$$

$$Cov(Y_1, Y_2) = Cov(l'_1 X, l'_2 X) = 0$$

Os autovalores possuem as seguintes propriedades:

$$tr(\Sigma) = \sum_{i=1}^p \lambda_i, \text{ para } i=1,2,\dots,p$$

$$|\Sigma| = \prod_{i=1}^p \lambda_i = \lambda_1 \lambda_2 \dots \lambda_p$$

Ou seja, o traço da matriz de covariâncias, que é a soma da diagonal principal é igual a soma dos autovalores, assim como o determinante da matriz é igual ao produto dos autovalores.

Para encontrarmos os autovetores correspondentes a cada autovalor, será necessária a resolução da seguinte equação matricial:

$$\Sigma a = \lambda a, \text{ onde } a \text{ é o autovetor normalizado}$$

Seja $a_1 = [a_{11}, a_{12}, \dots, a_{1p}]$ o autovetor associado ao 1º autovalor, então a 1ª componente principal será dada pela seguinte combinação linear:

$$Y_1 = a_{11}X_1 + a_{12}X_2 + \dots + a_{1p}X_p$$

O resultado da combinação linear Y_1 para o conjunto total da amostra é conhecido como os scores da 1ª componente principal, o resultado da combinação Y_2 , que utiliza como coeficientes o autovetor associado ao 2º autovalor é conhecido como scores da 2ª componente principal.

A proporção da variabilidade explicada por cada componente principal em relação ao conjunto original de dados, é dado pela seguinte forma:

$$\left(\begin{array}{l} \text{Variabilidade explicada} \\ \text{pela 1ª componente} \end{array} \right) = \frac{\lambda_1}{\sum_{i=1}^n \lambda_i}$$

$$\left(\begin{array}{l} \text{Variabilidade explicada até a} \\ \text{k - esima componente} \end{array} \right) = \frac{\sum_{i=1}^k \lambda_i}{\sum_{i=1}^n \lambda_i}$$

Como dito anteriormente, a Análise de Componentes Principais tem o objetivo de diminuir a dimensionalidade dos problemas sem que haja perda significativa de informação em relação ao conjunto de dados original.

A Análise de Componentes Principais não envolve suposições a cerca da distribuição dos dados, ou seja, não se baseia em premissas acerca do conjunto de dados, mas as propriedades de normalidade multivariada e linearidade podem contribuir para um melhor resultado nesse tipo de análise. Quaisquer fatores que afetem negativamente coeficientes de correlação como outliers e tamanho pequeno de amostra são indesejáveis, uma vez que a ACP se baseia em muitas vezes na matriz de correlação.

Quando possuímos número pequeno de amostras é possível ainda assim utilizarmos uma técnica estatística de simulação denominada de “bootstrap”. A idéia básica desta simulação é considerar a amostra de tamanho “n” disponível como sendo toda a população existente. A partir disso seleciona-se K amostras de tamanho n (tamanho original da amostra), porém utilizando amostragem com reposição. Para um número K consideravelmente grande de simulações realizadas, será possível efetuar testes de hipóteses e até mesmo construir intervalos de confiança, mesmo contando com um número reduzido de amostras.

A análise de componentes principais é uma técnica consagrada dentro da estatística multivariada, em muitos casos ela traz resultados muito satisfatórios com o uso de poucas

componentes. No entanto sabe-se que a ACP deve ser realizada sob todo o conjunto de dados. Em reconhecimento de padrões de imagens digitais isso não é desejável, e nem sempre trará resultados satisfatórios, por dois motivos:

- 1) O custo computacional é muito alto, a ortogonalização do conjunto de dados considerando o grande número de pixels existentes em uma imagem inteira pode ser problemática.
- 2) A técnica sendo aplicada a todos os pixels de uma imagem, necessariamente não leva em conta a informação sobre a verdadeira classe que pertence os pixels. Diferenças entre classes observáveis no conjunto de dados original poderiam ser pulverizadas em diferentes componentes da ACP, dificultando a correta classificação.

Em um capítulo posterior citaremos outros problemas que podem surgir quando utilizamos a técnica de análise de componentes principais em dados de imagens digitais, quando temos o intuito de diminuir dimensionalidade para realizarmos posteriormente uma classificação através da abordagem estatística em reconhecimento de padrões.

Como citamos anteriormente, ressaltamos que o principal objetivo desse trabalho é propor uma metodologia para a redução de dimensionalidade nos dados do sensor AVIRIS, de uma forma alternativa a Análise de Componentes Principais, que é a técnica mais utilizada quando se necessita reduzir dimensionalidade. Para tanto utilizaremos técnicas de geoestatística, mais precisamente dos modelos teóricos de semivariograma, com o intuito de aprimorar uma técnica de extração de parâmetros através dos modelos teóricos de semivariograma e desejamos que os parâmetros obtidos sejam eficazes como discriminador de classes distintas quando posteriormente utilizarmos estes parâmetros para classificação,

utilizando classificadores tradicionais já consagrados em reconhecimento estatístico de padrões.

2.4 Classificadores na Abordagem Estatística

Na abordagem estatística, a classificação é feita através de partições do espaço de feições, com a criação de volumes de espaços conhecidos como regiões de decisão.

Todos os vetores de características situados no interior de uma mesma região de decisão são classificados e atribuídos a uma mesma categoria.

Na maioria dos casos essas regiões de decisão são conexas, ou seja, contíguas, mas dependendo das distribuições subjacentes de cada categoria, essas regiões podem ser caracterizadas por soma de duas regiões não conexas.

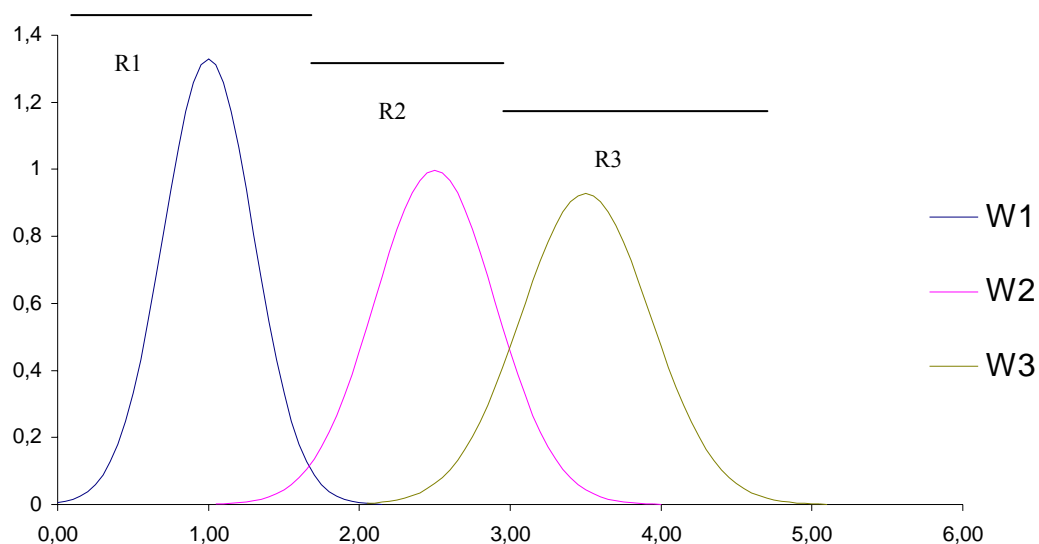


Figura 5 – Distribuição subjacente de 3 classes distintas – Exemplo unidimensional de regiões de decisão conexas.

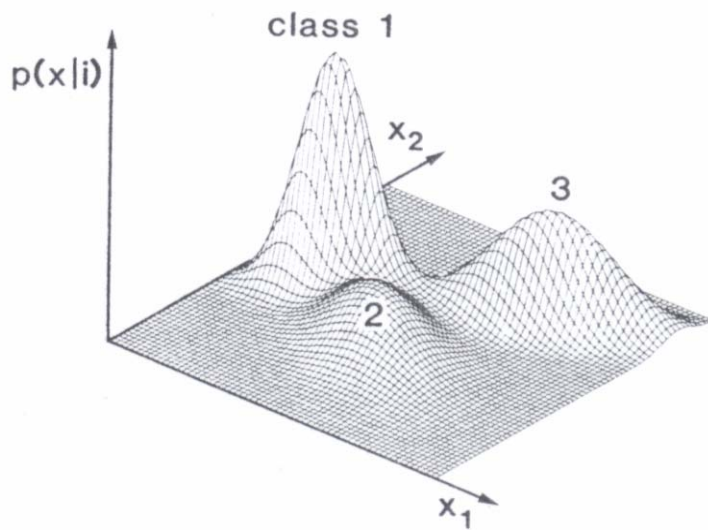


Figura 6 – Exemplo simulado de 3 classes normalmente distribuídas.

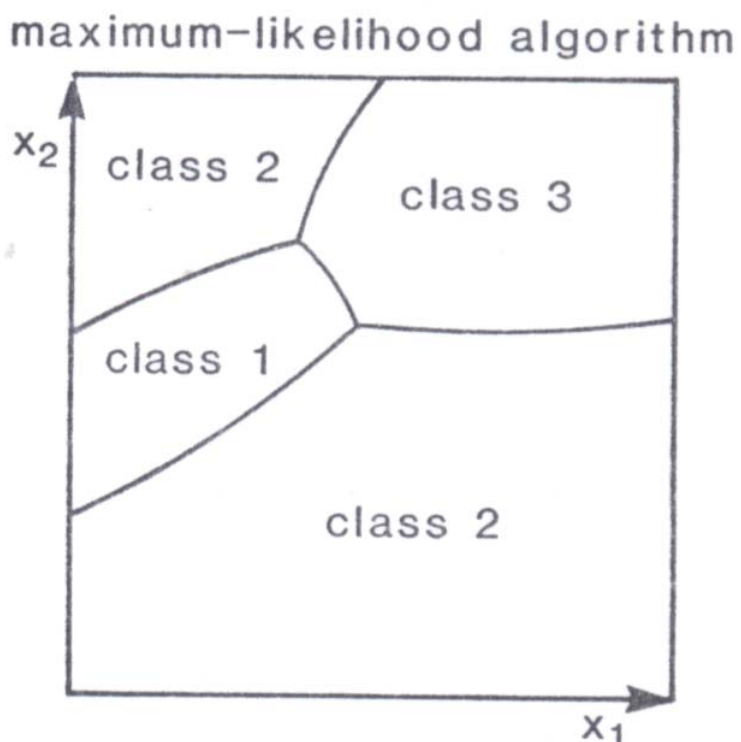


Figura 7 – Regiões de decisão para o exemplo das 3 classes normalmente distribuídas

A qualidade de um classificador pode ser medida através de sua acurácia, que é interpretada como sendo a probabilidade de um padrão qualquer ser classificado

corretamente, isso quando conhecemos a verdade de cada padrão. Devemos sempre perseguir a maximização da probabilidade de acerto, e por consequência a minimização da probabilidade de erro.

Normalmente não se utiliza o total de pixels conhecidos de cada classe na amostra de treinamento, é necessário designar metade deles para campo de teste. Estes pixels serão utilizados para verificar a acurácia do classificador, após serem conhecidos os parâmetros de cada classe obtidos através da amostra de treinamento. O percentual de acerto total verificado no campo de teste é uma estimativa para acurácia que se atingirá no conjunto total de dados de uma imagem.

Quanto mais próximas forem as classes distintas de um determinado estudo quanto às características estudadas, mais difícil será tarefa de atingir um nível alto de classificação correta. Podemos dizer que a similaridade das classes é o principal fator no “confundimento” ou “erro” de classificação, é ele que estabelece o patamar da probabilidade de erro na classificação.

Com o intuito de otimizar o conjunto de feições para minimizar a probabilidade de erro de um classificador qualquer, é desejável maximizar a distância entre os padrões de diferentes classes no espaço de feições. É possível com base em amostras de treinamento de várias classes, estimar a distância entre estas distintas classes. Existem diferentes tipos e formas de medir as distâncias entre classes, dentre elas podemos destacar: Distância de Mahalanobis; Distância de Bhattacharyya; Distância entre vizinhos mais próximos; Distância entre os centros.

2.5 Análise de Dados Espaciais (*Estatística Espacial*)

Muitos dos recentes desenvolvimentos metodológicos da estatística, e conseqüentemente do ensino da Estatística tem sido influenciados pelos avanços na área computacional e pelo uso de pacotes computacionais interativos, os quais permitem, além de análises de dados rotineiras, a utilização de gráficos criativos, simulações e transformação de dados de forma fácil e rápida. Ambientes computacionais, gráficos, dinâmicos e avanços nas áreas de mapeamento digital, cartografia computacional e Sistema de Informações Geográficas (*Geographical Information System-GIS*), potencializam desenvolvimentos paralelos de ambientes computacionais, permitindo visualizar dados espaciais, explorar padrões espaciais e relações entre fenômenos de forma interativa.

Análise de Dados Espaciais ou Estatística Espacial pode ser entendido como simplesmente uma aplicação dentro das muitas análises que podem ser desenvolvidas, em relação ao espaço, e que tem por objetivo a visualização e exploração através de mapas mostrando fenômenos no espaço. A análise de Dados Espaciais tem também como objetivo a detecção e reconhecimento de padrões de uma forma mais sofisticada a modelagem dos dados, também em relação ao espaço e em certos casos em relação ao tempo.

Segundo a definição dada por Bailey e Gatrell (1995), “Análise de Dados Espaciais trata das análises onde dados observáveis são obtidos a partir de algum processo operando no espaço e para os quais utilizam-se métodos para descrever ou explicar o comportamento deste processo e sua possível relação com outros fenômenos espaciais. Desta forma o objetivo da Análise de Dados Espaciais é aumentar a compreensão básica do processo, buscar evidências em relação às hipóteses estabelecidas ou ainda predizer valores em áreas onde as observações não foram feitas”.

Existem diferentes classes de problemas espaciais e tipos de dados espaciais. A Análise de Dados Espaciais é dividida em quatro grandes subconjuntos: *Análise de Padrões de Pontos*, *Análise de Dados no Espaço Contínuo*, *Análise de Dados de Área* e *Análise de Dados de Interação Espacial*. Explicaremos brevemente as diferenças entre estes tópicos e os tipos de problemas a que estão relacionados.

Análise de Padrões de Pontos refere-se a pontos discretos no espaço, onde foram observados a ocorrência de eventos e o objetivo é, em geral, ver se existe uma interação espaço-tempo entre os fenômenos ou se eles ocorrem de forma uniforme e aleatória em relação ao espaço. A classe de exemplos aqui trabalhada envolve, por exemplo, a ocorrência de uma determinada doença rara em determinada região. Neste exemplo, verificamos se os casos da doença ocorreram de forma aleatória em relação ao espaço e tempo ou se há uma dependência nos eventos conforme a proximidade de locais no espaço ou no tempo.

Análise de Dados no Espaço Contínuo refere-se a pontos fixos como no caso de padrões de ponto, porém o problema é entender como o processo varia no espaço através dos pontos estudados e fazer previsões para os demais pontos não estudados, esta parte da estatística espacial é também conhecida como Geoestatística. Um exemplo de fácil compreensão seria relacionado ao meio ambiente, como o estudo de poluição do ar. Sabemos que a cada ponto no espaço contínuo do ar, há uma taxa de poluição associada, no caso monitoramos alguns pontos e os resultados seriam generalizados para os demais pontos onde não foram coletadas informações. Este tópico da análise de dados espaciais concentra a maior parcela de estudos na análise de dados espaciais, além disso é a abordagem mais antiga e portanto a mais conhecida, os estudos utilizando técnicas nessa

abordagem são inúmeros, onde podemos destacar, sem dúvida a importância das áreas da Geologia e Agronomia.

Análise de Dados de Área tem como unidade de pesquisa áreas fixas, normalmente divididas segundo limites geográficos e políticos, tais como países, estados e municípios, ou até mesmo, regiões regulares como quadrados chamados de “*lattice*” ou “*pixels*”, utilizados em processamento de imagens de satélite e sensoriamento remoto. Os objetivos principais são examinar padrões nos atributos estudados em relação ao espaço, assim como verificação de relações entre as variáveis estudadas.

Problemas envolvendo *Dados de Interação Espacial* tratam de dados sobre fluxos que ligam um conjunto de localizações que podem ser pontos ou áreas e o objetivo é compreender o padrão de fluxos, construir modelos de tais fluxos e ainda, se possível, usar esta informação para fazer previsões sobre como fluxos podem mudar sob certos cenários. Exemplos destes tipos de dados são os seguintes: fluxos de pessoas, bens ou serviços entre “origens” e “destinos”. Nestes exemplos o objetivo é modelar o padrão de fluxos em termos da acessibilidade geográfica dos “destinos” em relação às “origens” e em termos da demanda das “origens” e da atratividade dos “destinos”. Tais análises são relevantes, em estudos de serviços de saúde, planejamento de centros comerciais, planejamento de transporte, entre outros.

2.5.1 *Geoestatística (Análise de Dados no Espaço Contínuo)*

A variabilidade espacial de características do solo vem sendo uma das preocupações de pesquisadores praticamente desde o início do século. Smith (1910) estudou a disposição de parcelas no campo em experimentos de rendimento de variedades

de milho. Montgomery (1913), preocupado com o efeito do nitrogênio no rendimento do trigo, fez um experimento em mais de 200 parcelas, medindo o rendimento dos grãos.

Os procedimentos usados nestes trabalhos baseavam-se simplesmente na estatística clássica e utilizavam grandes quantidades de dados amostrais, visando caracterizar a distribuição espacial das determinadas características em estudo, utilizando parâmetros como média e desvio padrão para representar um fenômeno e se baseando na hipótese de que as variações de um determinado local para outro são aleatórias.

Krige (1951), desenvolveu um estudo com dados de concentração de ouro e concluiu que somente a informação dada pela variância seria insuficiente para explicar o fenômeno em estudo. Surgiu então a necessidade de levar em consideração a distância entre as observações realizadas no estudo. A partir daí surge o conceito da geoestatística, que leva em consideração a localização geográfica e a dependência espacial.

Um conceito muito importante desenvolvido na geoestatística, sem dúvida é o conceito de variáveis regionalizadas que foram estudadas e desenvolvidas por Matheron a partir de observações de Krige.

Segundo Blais e Carlier (1968), uma variável regionalizada é uma função numérica com distribuição espacial aparentemente contínua, mas cujas variações não podem ser representadas por uma função matemática simples.

A teoria das variáveis regionalizadas pressupõe que a variação de uma característica medida ao longo de uma região é expressa de forma aditiva por três componentes:

- 1) componente estrutural: associada a um valor médio constante ou a uma tendência constante ao longo da região.
- 2) Componente aleatória espacial: Os valores de determinada característica estão correlacionados com a posição espacial na região.
- 3) Erro residual: componente aleatória residual, que não pode ser explicada pelo modelo.

Se \mathbf{x} representa uma posição em uma, duas ou três dimensões, então o valor da variável Z , em \mathbf{x} , é dada por (Burrough, 1987):

$$Z(\mathbf{x}) = m(\mathbf{x}) + \varepsilon'(\mathbf{x}) + \varepsilon''$$

onde:

$m(\mathbf{x})$ é uma função determinística que descreve a componente estrutural de Z em \mathbf{x} ;

$\varepsilon'(\mathbf{x})$ é um termo estocástico, que varia localmente e depende espacialmente de $m(\mathbf{x})$;

ε'' é um ruído aleatório não correlacionado, com distribuição normal de média zero e variância σ^2

2.5.2 Krigagem

Podemos definir krigagem como uma técnica de estimação por interpolação que permite estimar valores em locais onde observações não foram extraídas, a estimação é realizada com base no conjunto de pontos onde os atributos foram medidos.

Diferente dos métodos convencionais de estimação, a krigagem está fundamentada na teoria das variáveis regionalizadas. O primeiro passo na krigagem é definir uma função apropriada para a componente determinística $m(\mathbf{x})$. Para tanto, algumas hipóteses são

necessárias, dentre elas podemos destacar a chamada “estacionaridade de 2ª ordem” que se baseia em duas condições (Burrough, 1987 e David, 1977):

- A componente determinística $m(\mathbf{x})$ é constante, não há tendência na região.
- A variância das diferenças entre duas amostras depende somente da distância h entre elas, ou seja :

$$\text{Var}[Z(\mathbf{x}) - Z(\mathbf{x}+\mathbf{h})] = E\{[Z(\mathbf{x})-Z(\mathbf{x}+\mathbf{h})]^2\} = 2\gamma(\mathbf{h})$$

Onde:

E é o operador esperança matemática;

Var é o operador variância;

γ é a semi-variância.

O termo krigeagem é derivado do nome Daniel G. Krige, que foi o pioneiro a introduzir o uso de médias móveis para evitar a superestimação sistemática de reservas de mineração (Delfiner e Delhomme, 1975).

Inicialmente, o método de krigeagem foi desenvolvido para solucionar problemas de mapeamentos geológicos, mas seu uso expandiu-se com sucesso no mapeamento de solos (Burgess e Webster, 1980), mapeamento hidrológico (Kitanidis e Vomvoris, 1983), mapeamento atmosférico (Lajaunie, 1984) e outros campos correlatos.

A diferença entre a krigeagem e os outros métodos de interpolação são as maneiras como os pesos são atribuídos às diferentes amostras. No caso de interpolação linear simples, por exemplo, os pesos são todos iguais a $1/N$; na interpolação baseada no inverso do quadrado das distâncias, os pesos são definidos como o inverso do quadrado da

distância que separa o valor interpolado dos valores observados. Na krigagem, o procedimento é semelhante ao de interpolação por média móvel ponderada, exceto que aqui os pesos são determinados a partir de uma análise espacial, baseada no semivariograma experimental que leva em conta a distância dos pontos observados, do ponto a ser interpolado. Além disso, a krigagem fornece, em média, estimativas não tendenciosas e com variância mínima.

2.5.3 Semivariograma e Variograma – Ferramentas da Krigagem

O semivariograma ($\gamma(h)$) é uma função utilizada para determinar os pesos que cada amostra coletada na região terá na estimação dos demais pontos onde não foram coletadas informações. De forma intuitiva podemos imaginar que pontos que se encontrem mais próximos dos pontos a serem estimados possuam um peso maior do que amostras mais espaçadas do ponto que se deseja estimar.

Variograma ($2\gamma(h)$) é uma função equivalente ao semivariograma, e portanto também é uma ferramenta básica de suporte as técnicas de Krigagem, que permite representar quantitativamente a variação de um fenômeno regionalizado no espaço (Huijbregts, 1975).

A função do variograma é medir o nível de dependência da variável atributo a partir de mensurações realizadas em pontos num determinado espaço.

Um exemplo de fácil entendimento do uso das técnicas de krigagem seria medir o teor de zinco no solo em alguns pontos espalhados na região de interesse. Com a utilização do variograma seria possível medir a variabilidade do teor de zinco na região de estudo, e a

partir disso construir um modelo capaz de estimar o valor esperado do teor de zinco em qualquer ponto da região de estudo onde não foram coletadas informações.

O variograma ($2\gamma(h)$) pode ser definido como a esperança matemática do quadrado da diferença entre os valores de pontos no espaço separados pelo vetor h .

$$2\gamma(h) = E \{Z(x) - Z(x+h)\}^2$$

Através de uma amostra x_1, x_2, \dots, x_n , o variograma pode ser estimado da seguinte forma:

$$2\hat{\gamma}(h) = \frac{1}{N(h)} \sum_{i=1}^{N(h)} \{Z(x_i) - Z(x_i+h)\}^2$$

Onde:

$N(h)$ = nº de pares de amostras separados por um vetor de distância h .

$Z(x_i)$ e $Z(x_i+h)$ são os valores observados da variável atributo medidos nos pontos x_i e x_i+h , separados pelo vetor h .

A medida de dependência espacial mais conhecida e utilizada é o semivariograma ($\gamma(h)$), que é equivalente ao variograma e pode ser estimado por:

$$\hat{\gamma}(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} \{Z(x_i) - Z(x_i+h)\}^2$$

A fórmula acima possibilita o cálculo do semivariograma experimental para um conjunto de diferentes vetores h . A distância h que separa espacialmente as observações também é conhecida como lag.

Calculando os semivariogramas para diferentes direções, será possível avaliar a estrutura do comportamento dos dados em estudo, verificando a forma de comportamento da variabilidade que pode ser estrutura isotrópica ou anisotrópica.

Isotropia → A dependência espacial na variável atributo é considerada igual para todas as direções.

Anisotropia → A dependência espacial na variável atributo é diferente para as diversas direções espaciais em que se calcula o semivariograma. Ou seja, a variável medida em determinados pontos apresenta uma estrutura de variabilidade diferente conforme a direção que é pré-determinada.

Exemplo: Considere duas variáveis regionalizadas X e Y em uma Região R, sendo X leste, e Y norte.

Norte	03	13	23	33	43	53
	02	12	22	32	42	52
	01	11	21	31	41	51
	00	10	20	30	40	50
		Leste		R		

Figura 8 – Representação espacial das coordenadas “x” e “y”.

Os valores indicam as coordenadas na região dos pontos onde foram coletadas informações, sobre uma determinada variável atributo.

Considerando o semivariograma para a direção de 90°, os pares utilizados para o cálculo serão:

Lag1 → [(00,10);(10,20);(20,30);(30,40);(40,50);(01,11) ;(11,21);...;(43,53)]

Lag2 → [(00,20);(10,30);(20,40);(30,50);(01,21);(11,31);...;(33,53)]

Lag3 → [(00,30);(10,40);(20,50);(01,31);(11,41);...;(23,53)]

Lag4 → [(00,40);(10,50);(01,41);(11,51);...;(13,53)]

Lag5 → [(00,50);(01,51);...;(03,53)]

Para a direção de 0°, os pares serão:

Lag1 → [(00,01);(01,02);(02,03);(10,11);(11,12);...;(52,53)]

Lag2 → [(00,02);(01,03);(10,12);(11,13);...;(51,53)]

Lag3 → [(00,03);(10,13);...;(50,53)]

2.5.4 Principais Modelos Teóricos e Forma Geral do Semivariograma

O padrão normal de um semivariograma, se apresenta de forma intuitiva ao que esperamos de dados de campo, isto é, normalmente as observações mais próximas geograficamente, apresentam comportamento mais semelhantes entre si do que observações separadas por unidades de distâncias maiores.

Portanto esperamos valores pequenos para $\gamma(h)$, quando h é pequeno, e um acréscimo em $\gamma(h)$ a medida que h cresce.

Existem diferentes modelos de semivariogramas que podem ser utilizados convenientemente, conforme o melhor ajuste aos valores observados no semivariograma experimental. Os principais modelos são o exponencial, o linear, o gaussiano e o esférico;

a forma desses modelos assim como os parâmetros envolvidos em cada um deles poderão ser observados na figura 9.

A interpretação prática dos parâmetros envolvidos no semivariograma pode ser feita da seguinte forma:

Range - Alcance (a): Distância dentro da qual as amostras apresentam-se correlacionadas espacialmente.

Nugget – Efeito Pepita (C_0): Por definição $\gamma(0) = 0$. Entretanto a medida que h tende a zero, $\gamma(h)$ se aproxima de um valor positivo, esse valor revela a descontinuidade do semivariograma para distâncias menores do que a menor distância entre as amostras.

Structural Variance - Contribuição (C): é a diferença entre o patamar e o efeito pepita.

Sill – Patamar (C_0+C): Valor do semivariograma correspondente ao seu alcance

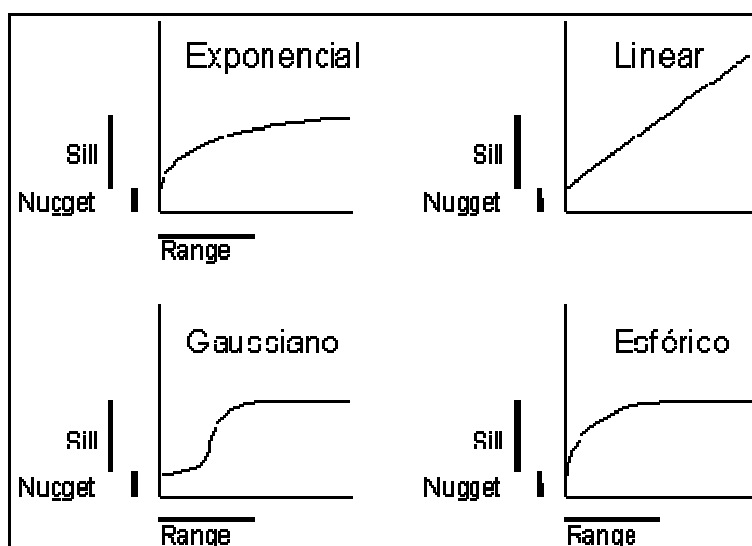


Figura 9- Principais modelos de semivariograma, com respectivos parâmetros.

1. Modelo Exponencial

$$\gamma(h) = C_0 + C[1 - \exp(-h/A)]$$

2. Modelo Esférico

$$\gamma(h) = C_0 + C[1.5(h/a) - 0.5(h/a)^3] \text{ para } h < a$$

$$\gamma(h) = C_0 + C \text{ para } h \geq a$$

3. Modelo Gaussiano

$$\gamma(h) = C_0 + C[1 - \exp(-h^2/A^2)]$$

4. Modelo Linear

$$\gamma(h) = C_0 + [h(C/a)]$$

Existem outros modelos de semivariogramas menos utilizados, assim como outros mais complexos, que normalmente são combinações de dois modelos simples, nosso trabalho não pretende se estender a esses modelos mais sofisticados.

3. METODOLOGIA PROPOSTA

Nesta altura do trabalho podemos evidenciar a quantidade de diferentes áreas do conhecimento que estão desencadeadas neste trabalho. Nesta seção temos a intenção de evidenciar o trabalho como um todo, assim como citar as diferentes técnicas e metodologias envolvidas conjuntamente nesse trabalho.

3.1 Posicionamento do trabalho quanto às diversas áreas do conhecimento

Em sensoriamento remoto, que se baseia na obtenção de informação sobre alvos sem que seja necessário o contato físico, nosso trabalho se encontra dentro de uma área chamada reconhecimento de padrões.

A aplicação de reconhecimento de padrões não é unicamente desenvolvida para sensoriamento remoto, e sim também para outros tipos de problemas onde não estão envolvidas imagens de satélite. Existem diferentes tipos de abordagem em reconhecimento de padrões, o nosso trabalho está localizado simplesmente dentro da abordagem estatística.

Dentro da abordagem estatística em reconhecimento de padrões existem os classificadores supervisionados e os não supervisionados, além dessa subdivisão nos classificadores, ainda podemos dividi-los nos que utilizam métodos paramétricos e nos que utilizam métodos não-paramétricos. Nosso trabalho pretende sugerir um novo processo

para redução da dimensionalidade dos dados, para posterior utilização de classificadores supervisionados tradicionais e consagrados.

A abordagem estatística em reconhecimento de padrões utiliza muitos preceitos da análise estatística multivariada, que concentra uma grande sorte de técnicas diferentes que podem ser utilizadas apropriadamente conforme o tipo de problema de pesquisa. A análise estatística multivariada, como o próprio nome sugere, foi desenvolvida para ser utilizada em problemas onde possuímos mais de duas variáveis. Em reconhecimento de padrões, existem gamas de trabalhos que apontam dificuldades de se conseguir uma classificação satisfatória devido a dificuldade de se estimar os componentes da matriz de covariâncias, principalmente por falta de amostra. Este problema resulta na necessidade de se buscar uma forma de reduzir a dimensionalidade do problema. É muito provável que nenhum estudo até hoje tenha sido efetuado utilizando técnicas de análise de dados espaciais como a geoestatística e mais precisamente com modelos de semivariograma para reduzir a dimensionalidade de dados, sendo assim consideramos que o nosso trabalho surge como uma alternativa inovadora na busca de uma nova forma de se reduzir dimensionalidades.

3.2 Problemas na Utilização de Dados com Alta Dimensionalidade

A alta dimensionalidade de dados é um problema muito comum em muitas áreas do conhecimento que utilizam a estatística multivariada. É verdadeiro dizer que quanto maior for o número de variáveis (maior dimensionalidade), maior terá que ser a amostra para que se possa estimar os parâmetros envolvidos com determinada confiança nas estimativas.

Quando não se tem amostra suficiente para estimar tantos parâmetros de forma confiável, surge a necessidade de diminuirmos a dimensionalidade do problema. A Análise de Componentes Principais é uma técnica capaz de diminuir a dimensionalidade do

problema, preservando muito da variabilidade do conjunto de dados, sem que haja perda significativa na variabilidade do conjunto de dados.

O procedimento na análise de componentes principais se baseia na ortogonalização da matriz de dados originais a partir da matriz de covariâncias ou de correlação, assim como a criação de componentes (novas variáveis) em número menor do que as variáveis originais. A análise de componentes principais permite também quantificar o percentual de perda na informação com a diminuição de dimensionalidade adotada.

A Análise de Componentes Principais traz resultados muito satisfatórios em muitos casos, é muito comum conseguir uma grande diminuição na dimensionalidade e mesmo assim preservar quase a totalidade da variabilidade do conjunto de dados original.

Em reconhecimento de padrões na abordagem estatística se utiliza a técnica de análise de componentes principais para a diminuição da dimensionalidade, existem até softwares comerciais com esta técnica já implementada. O principal problema observado nessa técnica quando se utiliza em reconhecimento de padrões é que a ACP tem que ser realizada sobre o total dos dados, não sendo possível ela ser realizada para cada classe separadamente. Além disso, a ACP tem o intuito de criar componentes através dos autovalores no sentido da maior explicação aos dados, isso nem sempre será razoável para discriminar classes distintas em reconhecimento de padrões. Esta situação pode ser facilmente observada na figura 10 que apresenta dados que resultam em uma primeira componente que não é eficaz para gerar escores suficientemente distintos para as duas classes, por outro lado a segunda componente apresenta escores bem distintos para cada classe.

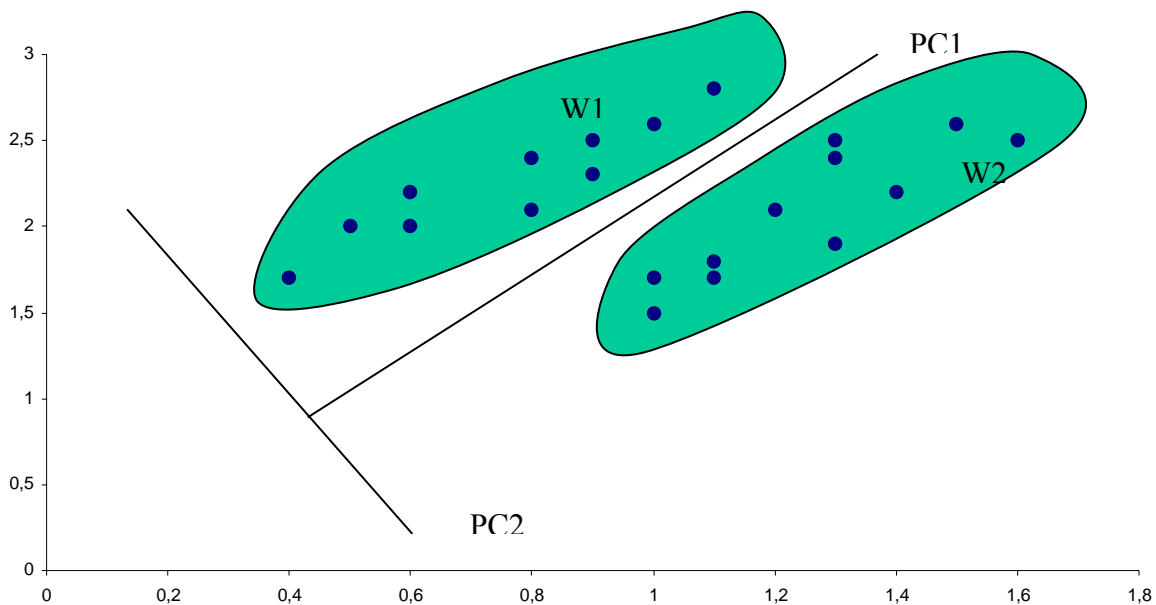


Figura 10- Simulação de um problema da Análise de Componentes Principais na discriminação de classes.

Em nosso trabalho pretendemos utilizar imagens obtidas pelo sensor AVIRIS que possui 224 bandas, ou seja, serão mais de 180 variáveis excluindo aquelas bandas de absorção da água. Será necessário reduzir a dimensionalidade desse problema, e com isso pretendemos desenvolver uma técnica que seja capaz de reduzir, mas que possa fazer isso para cada classe de forma separada.

3.3. Utilizando ferramentas da estatística espacial para reduzir dimensionalidade

Na estatística mais recente existe um grupo de técnicas que foram desenvolvidas para utilização em dados espaciais. A análise de dados espaciais apesar de recente apresenta muitas obras de grande relevância dentro da estatística. A análise de dados espaciais como dito anteriormente pode ser dividida em quatro tipos de abordagens. Em nosso trabalho iremos utilizar os parâmetros do semivariograma, que é uma ferramenta das técnicas de krigagem que por sua vez se encontra dentro de uma das abordagens da análise de dados espaciais, a geoestatística.

Basicamente o semivariograma funcionará como um redutor de dimensionalidade. Selecionaremos amostras de treinamento das diferentes classes e estimaremos o semivariograma experimental para a amostra de cada classe subdividindo o espaço das mais de 200 bandas em partições de bandas consecutivas, quanto maior for o número de partições, maior será a informação sobre a variação espectral de cada classe, porém maior será o número de parâmetros a serem estimados, portanto não podemos propor um número muito grande de partições porque voltariamos ao problema original de alta dimensionalidade.

Calculados os semivariogramas experimentais para a totalidade das classes existentes em uma imagem, buscaremos evidenciar qual modelo teórico se ajusta melhor aos dados observados no semivariograma experimental de cada classe.

Um cenário desejável seria que diferentes classes existentes em uma imagem se ajustassem a diferentes modelos teóricos de semivariograma. Nesse caso a simples identificação do modelo que melhor se ajusta a um pixel seria suficiente para determinação da classe a que pertence o determinado pixel. Por exemplo: considere que em uma imagem se apresentem três classes diferentes (w_1 , w_2 e w_3), extraímos amostras de treinamento de tamanho 200 pixels para cada classe e calculamos o semivariograma experimental médio para cada classe. Ao observarmos os valores do semivariograma experimental para w_1 , chegamos a conclusão de que o modelo linear se ajusta melhor aos dados, para w_2 o melhor ajuste se dá com o modelo exponencial e para w_3 o ajuste melhor é obtido com o modelo gaussiano. Sendo assim, a posteriori poderíamos selecionar um pixel qualquer e calcular o semivariograma experimental para ele, observando qual modelo se ajusta melhor aos dados poderíamos classificar o pixel em uma das três classes. É claro que a classificação somente será satisfatória se cada classe realmente apresentar melhor ajuste a

um determinado modelo do que os demais de forma significativa, no caso de existir um “confundimento”, ou um modelo se ajustar aos dados, somente um pouco melhor do que os demais, possivelmente a acurácia não será consideravelmente alta.

Porém, mesmo que duas ou mais classes distintas tenham melhor ajuste a um mesmo modelo de semivariograma, isso não significa que essas duas classes sejam de difícil discriminação, podemos identificar a classe através dos parâmetros do modelo. Por exemplo, os semivariogramas para as classes w_1 e w_2 se ajustam melhor ao modelo linear que possui dois parâmetros: pepita e patamar, porém a classe w_1 apresenta valores de parâmetros bem distintos dos valores obtidos para a classe w_2 . Portanto os parâmetros do modelo linear serão eficientes para tentar discriminar os pixels pertencentes à classe w_1 e w_2 .

O sistema sensor AVIRIS devido a grande quantidade de bandas existentes é útil principalmente para discriminar classes que apresentam comportamento espectral muito semelhante. Em nosso estudo iremos utilizar classes que apresentam comportamento espectral semelhantes, para estudos onde o comportamento espectral das classes é bem distinto, não há necessidade de utilizar dados provenientes de um sensor hiperspectral com mais de 200 bandas, nesses casos o uso de dados do sensor LANDSAT-TM que contém sete bandas, possivelmente serão suficientes para realizar uma boa classificação.

Quanto aos resultados a serem obtidos, seria desejável que pixels oriundos de uma mesma classe possuam uma grande homogeneidade interna quanto aos parâmetros do semivariograma calculados, ou seja, apresentem pequena variabilidade, ao mesmo tempo em que possua heterogeneidade com os valores dos parâmetros do semivariograma calculados para as demais classes existentes na imagem.

Como dito anteriormente o semivariograma é calculado ao longo de uma direção espacial, em nosso trabalho o semivariograma será calculado ao longo das das bandas espectrais do sensor AVIRIS.

O semivariograma como utilizado na geoestatística apresenta a variabilidade espacial de determinado fenômeno, através das diferentes variáveis que podem ser medidas no espaço. Em nosso trabalho o semivariograma representará a variação espectral das diferentes classes através das radiâncias observadas em níveis de contadores digitais em cada uma das bandas do sistema AVIRIS.

Desejamos que os parâmetros obtidos através dos modelos teóricos que melhor se ajustem aos resultados sejam eficientes para discriminar os pixels das diferentes classes existentes.

A idéia inicial do experimento é selecionar pelo menos dois campos de treinamento não contíguos para cada classe que estará representada no estudo. Preferimos selecionar amostras de diferentes locais para uma mesma classe para evitar que o fator espacial de determinada região influencie demasiadamente nos resultados de uma respectiva classe, ao mesmo tempo em que selecionamos amostras de locais diferentes de uma mesma classe podemos avaliar a homogeneidade dos valores médios dos parâmetros do semivariograma para cada uma delas.

O semivariograma experimental será calculado para cada pixel de cada classe, temos o intuito de subdividir o transecto de 224 (número esse ainda, sem desconsiderar as bandas que necessitam ser descartadas, pois contém comprimentos de onda que são de

absorção da água e, portanto não devem ser utilizadas) de forma consecutiva, em no mínimo quatro e no máximo dez partes. A idéia de particionar o espaço de bandas em determinados intervalos para o cálculo do semivariograma surge por duas premissas:

- 1) Se calcularmos o semivariograma experimental para um pixel e logo após analisar o modelo teórico de semivariograma que melhor se ajusta aos dados, dependendo do modelo, teríamos apenas dois ou três parâmetros que representaria o pixel. Como temos o intuito de trabalhar com classes que apresentem comportamento espectral muito semelhante, uma redução da dimensionalidade da ordem de 200 variáveis para apenas três, sem dúvida não traria resultados satisfatórios para a discriminação dessas classes com comportamento parecido.
- 2) A reflectância de um determinado alvo é determinada pela faixa de comprimento de onda que se refere à banda, é comum, dependendo do alvo que uma pequena variação no comprimento de onda da banda faça com que a reflectância cresça ou diminua significativamente. Como o semivariograma representa a variabilidade ao longo de um transecto, se este for muito grande, existirá muitos componentes no cálculo de cada lag, principalmente dos primeiros. Os resultados obtidos serão na verdade uma média grosseira da variabilidade espectral observada nas bandas. Ao passo que ao particionarmos o transecto de bandas, estaremos observando a variação espectral de cada pixel em faixas restritas de comprimento onda, ou seja, teremos um maior refinamento das informações de cada pixel nas referentes subdivisões do espectro eletromagnético, além de aumentarmos o número de parâmetros que representarão cada pixel.

Para facilitar o entendimento da técnica proposta neste trabalho e os procedimentos a serem executados, realizaremos um exemplo básico de forma clara utilizando um número menor de classes, para que se possa entender a finalidade do estudo.

3.4 Exemplo de aplicação da metodologia proposta

Possuímos uma imagem proveniente do sistema AVIRIS cobrindo uma determinada região, ou seja, temos a informação espectral ao longo das 224 bandas, pixel a pixel ao longo de toda a imagem. Em uma outra imagem temos a informação da realidade terrestre através de uma imagem temática, esta não necessariamente precisa ter dados sobre toda a região, pode existir algumas regiões ou pontos, onde não se tenha informação, pois pode ocorrer que a região não tenha sido totalmente varrida para verificação da realidade terrestre.

A partir da imagem temática que nos informa a realidade terrestre da região, podemos selecionar duas amostras de cada uma das duas classes que estudaremos, por exemplo, milho e trigo. Então, a partir da localização das amostras observadas na imagem temática, selecionaremos os dados da radiância, expressos em níveis de contadores digitais exatamente para os determinados pixels que pertencerão às amostras de treinamento.

A extração das informações da imagem raster pode ser feita através de rotinas já desenvolvidas no Departamento de Engenharia da Universidade PURDUE, USA denominadas de **loadimage** e **loadrow**, que através do MATLAB, que captura as informações dos pixels que são informados ao algoritmo e para as bandas desejadas, também informadas ao algoritmo. A única restrição do se dá pelo fato de somente ser

possível de se capturar informações em uma única vez de pixels que formem uma região retangular.

Suponha que selecionamos duas amostras de 400 pixels cada (20X20) de milho, e duas amostras de mesmo tamanho de trigo e buscamos as informações desses 1600 pixels em todas as 224 bandas existentes nas imagens provenientes do sensor AVIRIS. A estrutura de dados deste exemplo é mostrada na tabela 1.

Tabela 1- Matriz de dados original para 1600 pixels nas 224 bandas do sensor AVIRIS.

		Pixel	Bandas Espectrais (Transecto para cálculo do semivariograma)																
			1	2	3	4	221	222	223	224			
Milho	Campo 1	1																	
		2																	
		..																	
		..																	
		400																	
	Campo 2	401																	
		402																	
		..																	
		..																	
		800																	
Trigo	Campo 1	801																	
		802																	
		..																	
		..																	
		1200																	
	Campo 2	1201																	
		1202																	
		..																	
		..																	
		1600																	

Com a estrutura observada na tabela 1, é possível calcular o semivariograma experimental para os 1600 pixels, a idéia inicial é calcular a semivariância de lag 1 até o 20. O número 20 foi escolhido de forma arbitrária, por entendermos que até esse lag seria suficiente para podermos evidenciar qual modelo melhor se ajustaria ao semivariograma

experimental. A seguinte fórmula nos possibilita o cálculo do semivariograma pixel a pixel:

$$\hat{\gamma}(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} \{Z(x_i) - Z(x_i + h)\}^2$$

Onde h varia de 1 a 20;

$Z(x_i)$ e $Z(x_i + h)$ são os valores da radiância observados nas bandas x_i e $x_i + h$;

$N(h)$ é o número de pares de bandas separados por uma distância h .

Se fossemos calcular um semivariograma único para cada pixel, o número de pares seria o seguinte: $N(1)=223$; $N(2)=222$ $N(3)=221$;...; $N(20)=204$.

Considere que em um primeiro momento subdividíssemos o transecto de 224 bandas em quatro faixas:

Faixa 1 – Banda 1 até 56;

Faixa 2 – Banda 57 até 112;

Faixa 3 – Banda 113 até 168;

Faixa 4 – Banda 169 até 224;

Neste caso para cada faixa, $N(1)=55$; $N(2)=54$; $N(20)=36$. E após calcularmos o semivariograma até lag 20 para as quatro faixas e 1600 pixels, teríamos a estrutura de dados conforme a tabela 2.

Depois de calcularmos a semivariância até lag 20 em cada faixa para os 1600 pixels da amostra, devemos verificar qual modelo teórico de semivariograma se ajusta melhor aos dados observados no semivariograma experimental para cada faixa.

Tabela 2- Matriz de dados contendo as semivariâncias até lag 20 para 1600 pixels em uma subdivisão de 4 faixas de bandas do sensor AVIRIS.

		Faixa 1 (B 1-56)					Faixa 2 (B 57-112)					Faixa 3 (B 113-168)					Faixa 4 (B 169-224)				
		γ_1	γ_2	..	γ_{19}	γ_{20}	γ_1	γ_2	..	γ_{19}	γ_{20}	γ_1	γ_2	..	γ_{19}	γ_{20}	γ_1	γ_2	..	γ_{19}	γ_{20}
Milho	Campo 1	Pixel																			
		1																			
		2																			
		..																			
		400																			
	Campo 2	401																			
		402																			
		..																			
		800																			
		Trigo	Campo 1	801																	
802																					
..																					
1200																					
Campo 2	1201																				
	1202																				
	..																				
	1600																				

Em estudos preliminares já executados em momento anterior para confecção deste trabalho, foram utilizados cinco classes, algumas com comportamento espectral semelhante e outras com comportamento bem diferenciado, em todas as classes e faixas o modelo linear foi o que apresentou melhor ajuste aos dados, é verdade que em algumas faixas para determinadas classes, outros modelos também tiveram bom ajuste, mesmo assim um pouco abaixo do modelo linear. Ao considerarmos que o modelo linear é o mais

parcimonioso dos modelos teóricos de semivariograma, uma vez que apresenta dois parâmetros, enquanto os demais apresentam três, e também foi o que apresentou melhor ajuste em estudos preliminares, optamos por modelar os dados do semivariograma experimental diretamente com o modelo de semivariograma linear, mesmo assim sistematicamente estaremos verificando a qualidade dos ajustes a cada classe e número de partições adotadas.

Considerando que o melhor ajuste se dê com o modelo linear que é formado por dois únicos parâmetros, nossa estrutura matricial estaria representada na forma da tabela 3.

Tabela 3- Matriz de dados contendo os parâmetros do modelo de semivariograma linear em cada uma das quatro faixas de bandas para os 1600 pixels da amostra de treinamento.

		Pixel	Faixa 1 (B 1-56)		Faixa 2 (B 57-112)		Faixa 3 (B 113-168)		Faixa 4 (B 169-224)	
			C_0	C	C_0	C	C_0	C	C_0	C
Milho	Campo 1	1								
		2								
		⋮								
		⋮								
	400									
	Campo 2	401								
		402								
		⋮								
800										
Trigo	Campo 1	801								
		802								
		⋮								
		⋮								
		1200								
	Campo 2	1201								
		1202								
		⋮								
		⋮								
		1600								

Neste estágio, temos oito parâmetros calculados para cada pixel substituindo as 224 variáveis existentes originalmente. É desejável que esses oito parâmetros formem um espaço de dimensão oito com boa separabilidade entre os pixels de milho e trigo, pois assim, ao utilizarmos um classificador já consagrado, como por exemplo, o classificador pela máxima verossimilhança gaussiana, provavelmente alcançará uma excelente acurácia, desde que a distribuição desses parâmetros se aproxime da distribuição normal.

Com a redução dessa ordem (de 224 para 8), seria possível utilizar o classificador pela máxima verossimilhança gaussiana para classificar os demais pixels da imagem, caso fosse confirmada a normalidade dos parâmetros do semivariograma, pois facilmente teríamos tamanho de amostra suficiente para estimar a matriz de covariância para cada classe, o que não seria fácil com a dimensionalidade da ordem anterior.

Para representar a posição e a dispersão dos parâmetros em cada uma das classes, estimaríamos o vetor média e a matriz de covariâncias para cada classe, o que será utilizado no classificador máxima verossimilhança gaussiana para representar a distribuição dos dados em cada classe e para servir de critério na classificação dos outros pixels (campo de teste). Sendo representado da seguinte forma:

$$\text{Vetor Média para milho} \rightarrow \mu_M = [\bar{C}_{01}M, \bar{C}_1M, \bar{C}_{02}M, \bar{C}_2M, \bar{C}_{03}M, \bar{C}_3M, \bar{C}_{04}M, \bar{C}_4M],$$

$$\text{Vetor Média para trigo} \rightarrow \mu_T = [\bar{C}_{01}T, \bar{C}_1T, \bar{C}_{02}T, \bar{C}_2T, \bar{C}_{03}T, \bar{C}_3T, \bar{C}_{04}T, \bar{C}_4T],$$

Matriz de covariâncias para Milho:

$$\Sigma_M = \begin{bmatrix} \hat{\sigma}_{C_{01}M} & \text{Cov}(C_{01}M, C_1M) & \text{Cov}(C_{01}M, C_{02}M) & \dots & \text{Cov}(C_{01}M, C_4M) \\ \text{Cov}(C_{01}M, C_1M) & \hat{\sigma}_{C_1M} & \text{Cov}(C_1M, C_{02}M) & \dots & \text{Cov}(C_1M, C_4M) \\ \text{Cov}(C_{01}M, C_{02}M) & \text{Cov}(C_1M, C_{02}M) & \hat{\sigma}_{C_{02}M} & \dots & \text{Cov}(C_{02}M, C_4M) \\ \dots & \dots & \dots & \dots & \dots \\ \text{Cov}(C_{01}M, C_4M) & \text{Cov}(C_1M, C_4M) & \text{Cov}(C_{02}M, C_4M) & \dots & \hat{\sigma}_{C_4M} \end{bmatrix}$$

Matriz de covariâncias para Trigo:

$$\Sigma_T = \begin{bmatrix} \hat{\sigma}_{C_{01}T} & \text{Cov}(C_{01}T, C_1T) & \text{Cov}(C_{01}T, C_{02}T) & \dots & \text{Cov}(C_{01}T, C_4T) \\ \text{Cov}(C_{01}T, C_1T) & \hat{\sigma}_{C_1T} & \text{Cov}(C_1T, C_{02}T) & \dots & \text{Cov}(C_1T, C_4T) \\ \text{Cov}(C_{01}T, C_{02}T) & \text{Cov}(C_1T, C_{02}T) & \hat{\sigma}_{C_{02}T} & \dots & \text{Cov}(C_{02}T, C_4T) \\ \dots & \dots & \dots & \dots & \dots \\ \text{Cov}(C_{01}T, C_4T) & \text{Cov}(C_1T, C_4T) & \text{Cov}(C_{02}T, C_4T) & \dots & \hat{\sigma}_{C_4T} \end{bmatrix}$$

A distribuição normal multivariada que contém dois parâmetros (vetor média e matriz de covariâncias) é representada da seguinte forma:

$$p(\mathbf{x} / w_i) = \frac{1}{(2\pi)^{p/2} |\Sigma_i|^{1/2}} \exp\left\{ -\frac{1}{2} (\mathbf{x} - \mu_i)' \Sigma_i^{-1} (\mathbf{x} - \mu_i) \right\} \text{ para } i=1,2,\dots,k$$

Onde μ_i e Σ_i são respectivamente os parâmetros vetor média e matriz de covariância associados a classe w_i .

Para estabelecer a regra de decisão Bayesiana seria necessário multiplicar a função distribuição de cada classe com a probabilidade da referente classe (priori), surgindo assim a distribuição à posteriori, da seguinte forma:

$$p(w_i / x) = p(w_i) \cdot \frac{1}{(2\pi)^{p/2} |\Sigma_i|^{1/2}} \exp\left\{\frac{-1}{2} (\mathbf{x} - \mu_i)' \Sigma_i^{-1} (\mathbf{x} - \mu_i)\right\}$$

O termo $(2\pi)^{p/2}$ é constante para todas as classes e portanto pode ser desconsiderado, para simplificação dos cálculos sem que haja prejuízo na regra de classificação.

$$p(w_i / x) = p(w_i) \cdot \frac{1}{|\Sigma_i|^{1/2}} \exp\left\{\frac{-1}{2} (\mathbf{x} - \mu_i)' \Sigma_i^{-1} (\mathbf{x} - \mu_i)\right\}$$

No sentido de tornarmos a função de decisão mais simplificada para a discriminação da máxima verossimilhança gaussiana é possível extrairmos o logaritmo natural.

$$\ln[p(w_i / x)] = \ln[p(w_i)] - \frac{1}{2} \ln|\Sigma_i| - \frac{1}{2} (\mathbf{x} - \mu_i)' \Sigma_i^{-1} (\mathbf{x} - \mu_i)$$

Para simplificar a notação, será considerado $G_i(x) = \ln[p(w_i / x)]$. Assim, a expressão toma a seguinte forma:

$$G_i(x) = \ln[p(w_i)] - \frac{1}{2} \ln|\Sigma_i| - \frac{1}{2} (\mathbf{x} - \mu_i)' \Sigma_i^{-1} (\mathbf{x} - \mu_i)$$

A partir das funções $G_i(x)$ é possível estabelecer as regras de decisões e portanto as regiões de decisão para cada classe. A regra de classificação para um padrão \mathbf{x} qualquer será a seguinte:

$$\mathbf{x} \in w_i \text{ se } G_i(x) > G_j(x) \forall j \neq i$$

Em muitos é possível que não tenhamos informações à priori sobre as probabilidades de cada classe em uma imagem. Mesmo assim podemos utilizar o classificador Bayesiano optando pela ‘priori não-informativa’ considerando que todas as classes tenham a mesma probabilidade, ou seja, considerando $P(w_i) = 1/k$, para todas as k classes. Nesses casos as funções de decisão se tornam mais simples e assume a seguinte forma:

$$G_i(x) = -\frac{1}{2} \ln |\Sigma_i| - \frac{1}{2} (\mathbf{x} - \mu_i)' \Sigma_i^{-1} (\mathbf{x} - \mu_i) \text{ , sendo possível a simplificação pelo}$$

termo $1/2$.

$$G_i(x) = -\ln |\Sigma_i| - (\mathbf{x} - \mu_i)' \Sigma_i^{-1} (\mathbf{x} - \mu_i)$$

Em casos onde for comprovada a homogeneidade de variâncias, ou seja, onde podemos assumir que a matriz de covariâncias sejam iguais para todas as classes w_i as funções de decisão tomam a forma da conhecida distância de Mahalanobis.

$$G_i(x) = -(\mathbf{x} - \mu_i)' \Sigma^{-1} (\mathbf{x} - \mu_i)$$

Porém é preciso ter muito cuidado com este tipo de suposição, é necessário ter evidências muito fortes sobre a verdadeira matriz de covariâncias para assumir que elas sejam idênticas para todas as classes. A suposição feita de forma errônea pode trazer uma série de problemas no momento da classificação.

No exemplo que estamos trabalhando não temos evidências sobre a igualdade na matriz de covariâncias para os parâmetros do semivariograma para as classes milho e trigo, ao mesmo tempo em que não temos informação a priori sobre as probabilidades de cada uma dessas classes. Neste caso é mais apropriado utilizarmos a priori não informativa que assume probabilidade $\frac{1}{2}$ para cada classe. Assim sendo, a regra de decisão para discriminação das classes milho e trigo para os pixels dos campos de teste será a seguinte:

$$G_M(x) = -\frac{1}{2} \ln |\Sigma_M| - \frac{1}{2} (\mathbf{x} - \mu_M)' \Sigma_M^{-1} (\mathbf{x} - \mu_M)$$

$$G_T(x) = -\frac{1}{2} \ln |\Sigma_T| - \frac{1}{2} (\mathbf{x} - \mu_T)' \Sigma_T^{-1} (\mathbf{x} - \mu_T)$$

$\mathbf{x} \in \text{milho}$ se $G_M(x) > G_T(x)$; e $\mathbf{x} \in \text{trigo}$ se $G_M(x) < G_T(x)$

Sendo $\mu_M, \mu_T, \Sigma_M, \Sigma_T$ as respectivas estimativas para os parâmetros verdadeiros, vetor média e matriz de covariâncias para as classes milho e trigo, onde as componentes já foram detalhadas anteriormente.

Como citamos anteriormente desejamos utilizar a técnica proposta para tentar classificar/discriminar classes com comportamentos espectrais muito semelhantes, neste trabalho utilizaremos seis classes distintas, três da cultura de milho e três da cultura de soja, onde se distinguem essencialmente pela forma do plantio e, portanto realmente apresentam um comportamento espectral muito semelhante. A imagem coletada pelo sensor AVIRIS, que contém os dados que serão estudados, possui um grande número de classes, é uma imagem de 1992 de uma área que cobre uma região agrícola do estado de Indiana nos Estados Unidos e pode ser observada na figura 11.

A cena mostrada na figura 11 se refere à imagem do sensor AVIRIS da região de interesse, juntamente com a imagem temática contendo a verdade terrestre. A imagem foi obtida no início da época de crescimento, ocorrendo por isto grande quantidade de solo exposto, o que trará maior dificuldade ainda na discriminação de classes com comportamentos tão semelhantes.

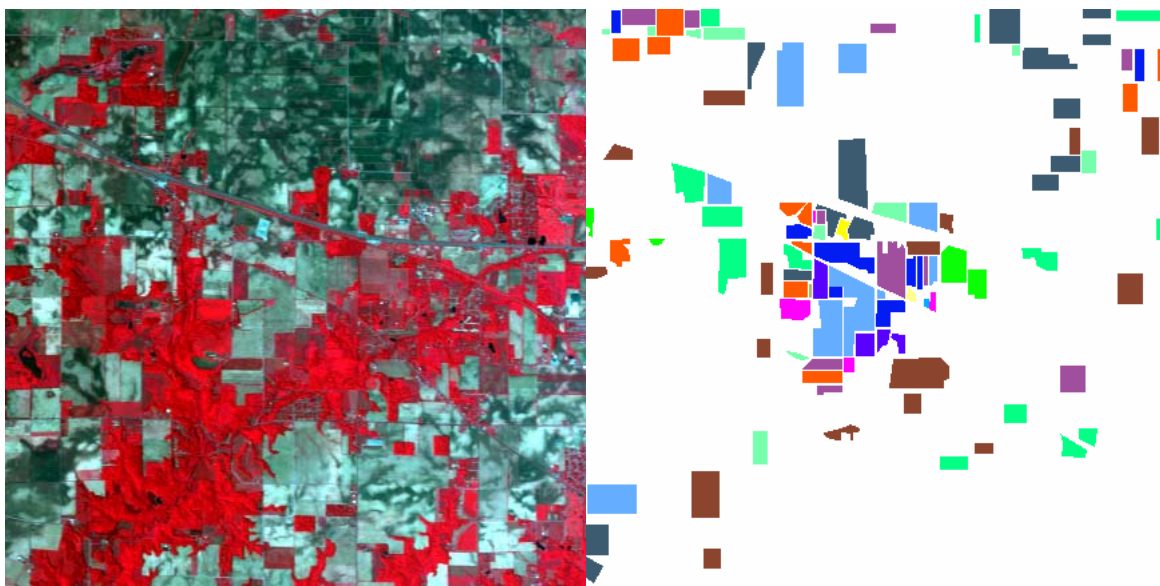


Figura 11 – (a) Imagem do sensor AVIRIS bandas 17,27 e 50 (b) Imagem temática da verdade terrestre

4. ESTUDO DE CASO - APLICAÇÃO EM DADOS REAIS DO SENSOR AVIRIS

4.1 Seleção de Bandas espectrais

Como etapa inicial para seleção de amostras das classes, é necessário realizar uma análise preliminar na imagem, com o intuito de trabalharmos somente com os dados que tragam informações essencialmente relevantes sobre as classes. Em dados de imagens digitais provenientes de sensores hiperspectrais é comum descartarmos algumas bandas de comprimentos de onda que estejam sendo afetadas pelo efeito de absorção, ocasionadas pelos constituintes como vapor d'água, por exemplo, que causam um nível muito alto de ruído em algumas bandas. Como resultado desta análise preliminar resultou na utilização de 190 bandas, sendo descartadas as demais. Para efeito de cálculo dos semivariogramas serão consideradas igualmente espaçadas as bandas que serão utilizadas no estudo, isso permitirá que não haja deficiências em alguns lags devido a retirada de algumas bandas de absorção. Para facilitação no entendimento, deste ponto em diante, a banda 30, por exemplo, será reconhecida como 30ª banda aceita e não como a banda 30 inicialmente rotulada no sistema AVIRIS.

4.2 Definição do número de partições e critérios dos locais de partições

Serão realizados estudos utilizando uma série de partições diferentes. Alguns critérios foram adotados para estabelecer o número de partições a serem simuladas. Estabeleceu-se que cada partição terá o mesmo número de bandas, isso favorece que os *lags* comuns calculados em cada partição tenham o mesmo número de componentes

amostrais e, portanto, a confiança nas estimativas será aproximadamente a mesma no que se refere ao método, sendo influenciadora na confiança apenas a variabilidade intrínseca dos dados. Adotou-se também que cada lag calculado para qualquer partição seja proveniente de pelo menos de uma amostra de 5 pares de bandas, um número menor que este pode trazer estimativas com erros muito grandes. Por fim, estabeleceu-se que o número de partições seria de no mínimo 4, pois um número menor que este provavelmente seria insatisfatório para discriminar classes tão semelhantes, uma vez que reduziria o número de parâmetros de 190 para 6. As tabelas 4 e 5 mostram com detalhe as partições que serão realizadas no estudo.

Tabela 4- Número de partições adotadas no estudo e características de cada partição.

Nº de partições	Número de bandas em cada partição	Lag máximo a ser atingido em cada partição
4	48	20
6	32	20
8	24	15
10	19	12

Tabela 5- Transecto de bandas por número de partições.

4 Partições	6 Partições	8 Partições	10 Partições
B(1-48)	B(1-32)	B(1-24)	B(1-19)
B(49-96)	B(33-64)	B(25-48)	B(20-38)
B(97-144)	B(65-96)	B(49-72)	B(39-57)
B(145-190)	B(97-128)	B(73-96)	B(58-76)
	B(129-160)	B(97-120)	B(77-95)
	B(161-190)	B(121-144)	B(96-114)
		B(145-168)	B(115-133)
		B(169-190)	B(134-152)
			B(153-171)
			B(172-190)

4.3 Descrição das classes

Foram selecionadas seis classes para fazerem parte do estudo, são classes que apresentam um comportamento espectral muito semelhante. Para as seis classes selecionadas, foram extraídos todos os pixels da imagem AVIRIS através de uma varredura na imagem temática, onde se observa a verdade terrestre do local.

As classes selecionadas foram: corn (milho plantio lavrado tradicional), corn min (milho plantio intermediário), corn notill (milho plantio direto), soy clean (soja plantio lavrado tradicional), soy min (soja plantio intermediário), soy notill (soja plantio direto). Os pixels referentes a cada classe foram divididos em duas partes, pixel a pixel, um pertencendo ao grupo de amostras de treinamento e outro subsequente pertencendo a amostras de campo de teste, processo este de forma sucessiva.

Para ilustrarmos a semelhança no comportamento espectral dessas classes, construímos um gráfico contendo as médias observadas nas amostras de treinamento em cada uma das 190 bandas para cada classe, que pode ser observada na figura 12.

Tabela 6-Quantidade de pixels por classe em amostra de treinamento e teste

Classe	Total de Pixels	Amostra de Treinamento	Amostra de Teste
Corn	3642	1821	1821
Corn Min	2442	1221	1221
Soy Clean	4654	2327	2327
Corn Notill	1720	860	860
Soy Min	5030	2515	2515
Soy Notill	2264	1132	1132

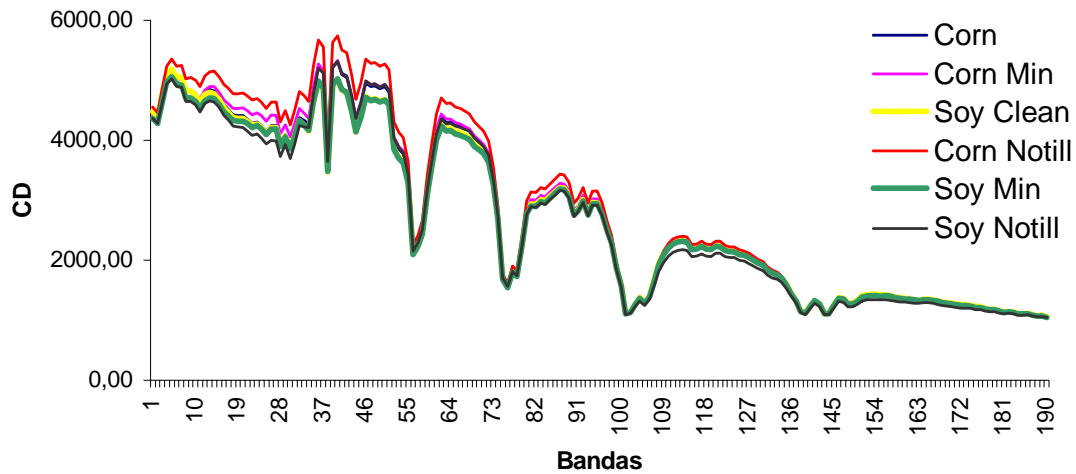


Figura 12- Média amostral espectral para as diferentes classes

4.4 Resultados dos Experimentos e Análise

Para realizarmos os procedimentos necessários para a implementação da técnica proposta neste trabalho, foram desenvolvidas muitas rotinas no software MATLAB. Essas rotinas foram desenvolvidas para poderem particionar o transecto de 190 bandas em quantidades diferentes; calcular os semivariogramas experimentais para cada partição, até o número de lag desejado; calcular os parâmetros do semivariograma ajustado; e por fim a classificação dos pixels de teste segundo regra bayesiana, através do critério da máxima verossimilhança gaussiana. Além disso, foram criadas algumas rotinas para verificar o ajuste dos dados do semivariogramas experimental em relação ao semivariograma linear ajustado conforme os parâmetros calculados, e verificação da normalidade dos parâmetros calculados para cada classe.

Realizaremos diferentes tipos de experimentos. Em um deles será efetuada a classificação utilizando todas as combinações de classes duas a duas, para as partições 4, 6, 8 e 10, em outro será feito um experimento utilizando as três classes da mesma cultura, ou

seja um experimento será realizado com as três classes de soja e o outro com as três classes de milho, utilizando as partições 4,6,8 e 10, por fim, será realizado o experimento utilizando as 6 classes do estudo para as mesmas partições. Para todos estes experimentos ainda iremos realizar estudos comparando os resultados obtidos com a suposição de matriz de covariâncias idênticas, com os resultados assumindo matriz de variâncias heterogêneas.

4.4.1 Experimentos utilizando duas classes

Neste experimento iremos utilizar as seis classes combinando elas duas a duas, e procederemos a classificação a partir dos parâmetros do semivariograma linear calculados para cada partição do transecto de 190 bandas. A partir destes resultados poderá ser possível detectar que par de classes tem a classificação correta mais difícil, além disso, poderemos observar se a acurácia aumenta à medida que aumentamos o número de partições e, portanto aumentamos o número de parâmetros.

A ordem na realização das etapas dos procedimentos da técnica foi a seguinte:

- 1) Particionou-se o transecto de 190 bandas em quatro partes;
- 2) Calculou-se as semivariâncias até lag 20 para cada pixel de todas as amostras de treinamento para as 4 partições;
- 3) A partir de cada semivariância até lag 20 foram calculados os parâmetros do semivariograma linear para cada pixel na 1ª partição; depois se repete para as outras três partições; sistematicamente é verificado o ajuste dos dados ao semivariograma linear;
- 4) Organizou-se para cada pixel os oito parâmetros obtidos (dois para cada partição, coeficiente linear ou efeito pepita e coeficiente angular ou contribuição);

- 5) Observou-se através de histogramas, se os parâmetros podem ser assumidos como sendo regidos por uma distribuição gaussiana para cada classe;
- 6) Estima-se o vetor média e a matriz de covariâncias para cada classe com respeito aos oito parâmetros;
- 7) Conhecidos os dois primeiros momentos estimados, esta estimada a função distribuição de probabilidade de cada classe e, portanto é possível realizar a classificação, conforme o critério da máxima verossimilhança gaussiana;
- 8) Realiza-se a classificação dos pixels de teste utilizando os parâmetros das 2 classes na combinação de estudo, classificou-se o pixel naquela classe onde a função verossimilhança é maximizada, se repete esta etapa para as outras catorze combinações de classes duas a duas;
- 9) Verificou-se a acurácia para cada par de classes analisadas;
- 10) Repete os itens de 1 a 9, para a partição de tamanho 6, depois para 8 e 10, considerando as diferenças nos números devido a cada partição.

4.4.1.1 Resultados utilizando matriz de covariâncias distintas para cada classe

Os resultados obtidos a partir da classificação pela máxima verossimilhança gaussiana utilizando os parâmetros do semivariograma ajustado, para cada combinação de par de classes podem ser observados na tabela 7.

Os valores mostrados na tabela 7 se referem ao percentual de pixels de teste da classe que foram classificados corretamente em uma determinada partição. Por exemplo, na combinação 1 (corn e corn min) na partição 6, 68,20% dos pixels de corn foram classificados corretamente e 91,48% dos pixels de corn min foram classificados corretamente, portanto acurácia geral ponderada para a combinação 1 em 6 partições foi de 77,54% como pode ser observado na tabela 8.

Tabela 7- Acurácia obtida na classificação por classe em cada combinação por partição (em%)

Combinação de par de classes	Classes	Número de Partições (Número de Parâmetros)			
		4 (8)	6 (12)	8 (16)	10 (20)
Combinação 1	corn	64,52	68,20	65,02	63,04
	corn min	92,38	91,48	92,55	94,51
Combinação 2	corn min	92,87	92,05	94,18	96,40
	corn notill	50,23	67,21	66,16	56,04
Combinação 3	corn	80,12	80,28	82,70	85,94
	corn notill	85,00	89,88	92,32	86,74
Combinação 4	soy clean	82,34	78,73	81,35	83,15
	soy min	87,27	89,98	87,63	87,93
Combinação 5	soy clean	97,63	98,41	97,80	97,76
	soy notill	84,01	84,98	91,07	90,99
Combinação 6	soy min	96,70	96,62	96,38	95,78
	soy notill	48,58	57,68	58,83	63,52
Combinação 7	corn	63,43	64,80	63,21	64,85
	soy clean	89,17	92,31	94,54	93,77
Combinação 8	corn	49,26	64,42	56,62	58,54
	soy min	94,36	95,43	95,59	96,58
Combinação 9	corn	96,43	95,06	93,41	92,59
	soy notill	62,90	73,85	82,77	84,72
Combinação 10	corn min	92,55	93,53	95,99	96,40
	soy clean	90,29	94,07	90,72	92,39
Combinação 11	corn min	85,50	86,40	88,62	90,00
	soy min	71,97	84,45	74,99	80,07
Combinação 12	corn min	97,62	98,12	97,46	97,62
	soy notill	37,90	49,29	52,47	59,27
Combinação 13	corn notill	91,28	94,42	95,58	92,33
	soy clean	96,82	97,94	97,89	97,81
Combinação 14	corn notill	59,88	73,84	78,14	67,33
	soy min	89,18	91,18	88,99	92,84
Combinação 15	corn notill	96,74	91,86	92,79	81,98
	soy notill	42,40	64,66	68,37	85,07

Tabela 8- Acurácia ponderada observada em cada combinação por partição (em %)

Combinação de par de classes	Número de Partições (Número de Parâmetros)			
	4 (8)	6 (12)	8 (16)	10 (20)
Combinação 1	75,70	77,54	76,07	75,67
Combinação 2	75,25	81,78	82,60	79,72
Combinação 3	81,69	83,36	85,79	86,20
Combinação 4	84,90	84,57	84,61	85,63
Combinação 5	93,17	94,01	95,60	95,54
Combinação 6	81,76	84,53	84,72	85,77
Combinação 7	77,87	80,23	80,79	81,07
Combinação 8	75,42	82,41	79,22	80,60
Combinação 9	83,58	86,93	89,33	89,57
Combinação 10	91,07	93,88	92,53	93,77
Combinação 11	76,39	85,09	79,44	83,32
Combinação 12	68,89	74,63	75,82	79,17
Combinação 13	95,33	96,99	97,27	96,33
Combinação 14	81,71	86,76	86,23	86,34
Combinação 15	65,86	76,40	78,91	83,74
% de acerto	81,32	85,06	84,79	85,68

Tabela 9- Acurácia média observada em cada classe independente da combinação, conforme número de partições (em %)

Classes	Número de Partições (Número de Parâmetros)			
	4 (8)	6 (12)	8 (16)	10 (20)
corn	70,75	74,55	72,19	72,99
corn min	92,18	92,32	93,76	94,99
soy clean	91,25	92,29	92,46	92,98
corn notill	76,63	83,44	85,00	76,88
soy min	87,90	91,53	88,72	90,64
soy notill	55,16	66,09	70,70	76,71

4.4.1.2 Resultados assumindo matriz de covariâncias iguais para todas as classes

Os resultados obtidos a partir da classificação pela máxima verossimilhança gaussiana utilizando os parâmetros do semivariograma ajustado, supondo homogeneidade de variâncias, ou seja, utilizando matriz de covariâncias iguais para cada combinação de par de classes pode ser observada na tabela 10.

Tabela 10- Acurácia obtida na classificação por classe em cada combinação por partição, supondo homogeneidade de variâncias (em %)

Combinação de par de classes	Classes	Número de Partições (Número de Parâmetros)			
		4 (8)	6 (12)	8 (16)	10 (20)
Combinação 1	corn	77,76	81,11	81,82	79,41
	corn min	71,66	74,20	74,44	74,53
Combinação 2	corn min	83,46	85,67	87,71	88,21
	corn notill	75,58	78,84	81,16	81,40
Combinação 3	corn	74,19	77,48	78,36	76,44
	corn notill	85,00	83,14	87,79	85,70
Combinação 4	soy clean	81,82	83,54	83,20	83,03
	soy min	80,44	85,29	80,83	80,04
Combinação 5	soy clean	94,28	94,80	95,02	96,09
	soy notill	91,61	87,20	92,93	92,14
Combinação 6	soy min	85,33	85,77	85,29	85,85
	soy notill	71,38	68,46	71,73	74,82
Combinação 7	corn	70,90	74,74	71,99	73,04
	soy clean	75,16	77,87	73,61	74,82
Combinação 8	corn	70,62	75,34	74,14	78,25
	soy min	82,5	88,51	81,87	84,45
Combinação 9	corn	81,60	86,38	85,06	88,85
	soy notill	80,04	80,48	81,45	84,36
Combinação 10	corn min	81,08	83,13	82,15	83,78
	soy clean	88,23	92,52	91,58	91,41
Combinação 11	corn min	70,92	73,63	75,92	78,71
	soy min	69,98	77,73	74,03	80,68
Combinação 12	corn min	76,90	85,91	85,01	86,49
	soy notill	70,76	76,59	78,89	83,22
Combinação 13	corn notill	93,26	94,07	95,58	95,12
	soy clean	94,97	97,38	97,29	96,86
Combinação 14	corn notill	80,93	82,44	85,23	82,56
	soy min	78,21	88,51	87,71	88,55
Combinação 15	corn notill	77,56	79,3	76,98	79,65
	soy notill	83,30	85,25	85,69	87,54

Tabela 11- Acurácia ponderada observada em cada combinação por partição, supondo homogeneidade de variâncias (em %)

Combinação de par de classes	Número de Partições (Número de Parâmetros)			
	4 (8)	6 (12)	8 (16)	10 (20)
Combinação 1	75,31	78,34	78,86	77,45
Combinação 2	80,20	82,85	85,00	85,40
Combinação 3	77,66	79,30	81,38	79,41
Combinação 4	81,10	84,45	81,97	81,48
Combinação 5	93,41	92,31	94,34	94,80
Combinação 6	81,00	80,40	81,08	82,43
Combinação 7	73,29	76,50	72,90	74,04
Combinação 8	77,51	82,98	78,62	81,85
Combinação 9	81,00	84,12	83,68	87,13
Combinação 10	85,77	89,29	88,33	88,78
Combinação 11	70,29	76,39	74,65	80,04
Combinação 12	73,95	81,43	82,07	84,92
Combinação 13	94,51	96,49	96,83	96,39
Combinação 14	78,90	86,96	87,08	87,02
Combinação 15	80,82	82,68	81,93	84,13
% de acerto	80,28	83,63	82,94	84,06

Tabela 12- Acurácia média observada em cada classe independente da combinação, conforme número de partições, supondo matriz de covariâncias iguais (em %)

Classes	Número de Partições (Número de Parâmetros)			
	4 (8)	6 (12)	8 (16)	10 (20)
corn	75,01	79,01	78,27	79,20
corn min	76,80	80,51	81,05	82,34
soy clean	86,89	89,22	88,14	88,44
corn notill	82,47	83,56	85,35	84,89
soy min	79,29	85,16	81,95	83,91
soy notill	79,42	79,60	82,14	84,42

4.4.1.3 Análise dos resultados de classificação utilizando duas classes

Em primeiro momento, só estava prevista a classificação utilizando matriz de covariâncias diferentes para cada classe. Acontece que foi constatada em algumas combinações, uma grande diferença entre a acurácia observada em uma classe em comparação a outra, quando utilizamos matriz de covariâncias distintas. Esse problema

pode ser visto nas combinações 1,2,6,7,8,12 e 14, e também é possível observar que em todas essas combinações, as classes que apresentam acurácia baixa sempre são as classes *corn*, *corn notill* e *soy notill*. Esse resultado indica que há uma dificuldade em classificar novos pixels nestas classes quando são confrontadas com as classes *corn min* e *soy min*.

Observando a função verossimilhança gaussiana $\frac{1}{|\Sigma_i|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{x} - \mu_i)' \Sigma_i^{-1} (\mathbf{x} - \mu_i)\right\}$,

surgiu a idéia de que o termo $|\Sigma_i|$ (determinante da matriz de covariâncias) poderia estar influenciando na classificação, se algumas classes apresentassem grandes diferenças neste termo. Isto poderia influenciar diretamente nos casos onde as classes apresentassem muita similaridade nos parâmetros de posição (primeiro momento), por isso optamos por também simular a classificação utilizando matriz de covariâncias iguais para as classes de cada combinação, a fim de observar se há uma melhora na acurácia em algumas combinações, uma vez que tornaríamos o termo determinante da matriz de covariâncias igual para cada classe em cada uma das combinações.

Comparando os resultados da tabela 7 classificação utilizando (matriz de covariâncias distintas) com a tabela 10 podemos observar que a acurácia melhorou consideravelmente nas classes que apresentavam uma acurácia reduzida nas combinações 1,2,6,7,8,12 e 14, porém é comum também observar nestas mesmas combinações, uma redução na acurácia da outra classe que apresentava uma acurácia maior, o que é perfeitamente previsível. Comparando ainda estas tabelas fica evidente que os valores de acurácia de cada classe/combinação apresentam menor variabilidade quando utilizamos a classificação considerando matrizes de covariâncias iguais, por exemplo, o valor mais baixo que observamos na tabela de acurácias considerando matriz de covariâncias iguais é 68,46%, valor este da acurácia para os pixels da classe *soy notill* na combinação 6, ou seja, quando é confrontada com a classe *soy min* sendo utilizada a partição 6. Na tabela 7

podemos observar exatamente 30 valores de acurácia menores que 68,46%. Em contrapartida há escassez de valores altos, considerando somente valores acima de 93%, somente são observados para a classe *soy clean* na combinações 6 e 13 e para a classe *corn notill* na combinação 13, enquanto quando utilizamos matriz de covariâncias distintas este percentual é facilmente atingido em várias combinações.

Verificando a acurácia geral para cada combinação para os dois métodos nas tabelas 8 e 11, observamos que houve aumento de acurácia nas combinações 1 e 14 nas partições 6, 8 e 10; nas combinações 2 e 12 em todas as partições e na combinação 8 nas partições 4, 6 e 10; não houve aumento nas combinações 6 e 7 onde se imaginaria que seria possível obter algum aumento. Nas demais combinações, onde não se esperaria melhora alguma, os resultados se confirmaram, somente na combinação 5 para a partição 4 houve um pequeno aumento.

Ao observarmos os valores médios de acurácia em cada partição, nos dois métodos podemos constatar que a classificação foi mais adequada quando não supomos matriz de covariâncias iguais, a melhor acurácia média foi observada na partição 10, onde foi atingido um índice de 85,68%, o pior resultado foi observado na partição 4 quando utilizamos matriz de covariâncias iguais, neste o índice foi de 80,28%.

A classe *corn min* foi a que apresentou maior percentual de acerto independentemente da combinação chegando quase a 95%, ao considerarmos a partição 10 com matriz de covariâncias distintas que foi a partição e método que maximizou a acurácia, já a classe *corn* foi a que apresentou menor acurácia, o percentual para esta classe ficou em 72,99%.

Através destes resultados podemos observar que realmente o fato de a matriz de covariâncias ser distinta para as classes, faz com que as classes que apresentem maior variância (entenda por isso apresentar maior valor para o termo $|\sum_i|$) apresentem uma probabilidade de erro relativamente grande em consideração com aquelas classes que apresentam menor variância. A tabela 13 nos mostra somente o termo $|\sum_i|$ para cada classe observada nas 4 partições trabalhadas.

Tabela 13- Valores observados para o determinante da matriz de covariâncias para cada classe em cada partição

Classes	Número de Partições (Número de Parâmetros)			
	4 (8)	6 (12)	8 (16)	10 (20)
corn	7,05E+64	3,02E+89	2,54E+111	5,33E+132
corn min	5,82E+63	1,16E+88	2,25E+109	7,51E+130
soy clean	2,45E+64	4,30E+88	3,27E+110	7,26E+131
corn notill	4,10E+64	6,20E+88	2,06E+110	4,29E+132
soy min	1,03E+64	1,10E+88	5,11E+109	1,11E+131
soy notill	5,50E+65	3,90E+89	9,55E+110	1,76E+132

Tabela 14-Número de vezes que o termo determinante da matriz de covariâncias para cada classe é maior em relação à classe corn min

Classes	Número de Partições (Número de Parâmetros)			
	4 (8)	6 (12)	8 (16)	10 (20)
corn	12,11	25,93	113,08	70,93
corn min	1,00	1,00	1,00	1,00
soy clean	4,21	3,70	14,53	9,67
corn notill	7,05	5,33	9,17	57,15
soy min	1,78	0,94	2,27	1,47
soy notill	94,47	33,56	42,43	23,40

Se compararmos os resultados observados na tabela 11 com os resultados das tabelas 13 e 14, podemos observar que a é confirmada a suposição inicial de que o termo determinante da matriz de covariâncias poderia estar influenciando demasiadamente na acurácia de cada classe. Note que as classes que apresentam menor percentual de pixels

classificados corretamente foram *corn*, *soy notill* e *corn notill*, e estas são as classes que apresentam maiores valores para o termo determinante da matriz de covariância, termo este, que se encontra no denominador da função de verossimilhança gaussiana, ou seja, altos valores para este termo, fazem com que a função assumam menores valores, dificultando a maximização da função verossimilhança nesta classe em relação a outras classes que apresentam menores valores para o determinante da matriz de covariâncias.

Para facilitar o entendimento de como a variabilidade influencia na classificação, ou seja, no percentual que cada classe atingirá de acertos, iremos simular através de um exemplo univariado. Suponha duas classes que apresentaram medidas centrais parecidas, porém uma delas apresentará uma variância maior. Deseja-se mostrar com este exemplo, que aquelas classes que apresentem maior variabilidade tendem a apresentar um menor percentual de acerto, ou seja, uma maior probabilidade de erro, assim como foi visto nos resultados reais analisados anteriormente.

Suponha que existam duas classes normalmente distribuídas com médias próximas, mas com variâncias consideravelmente diferentes, como pode ser observado na Figura 13.

A classe W_2 apresenta maior variabilidade que a classe W_1 , ao utilizarmos o classificador Bayesiano através do critério da máxima verossimilhança gaussiana, a probabilidade de erro esperada, ou seja, a probabilidade de classificarmos uma unidade incorretamente será maior para a classe W_2 .

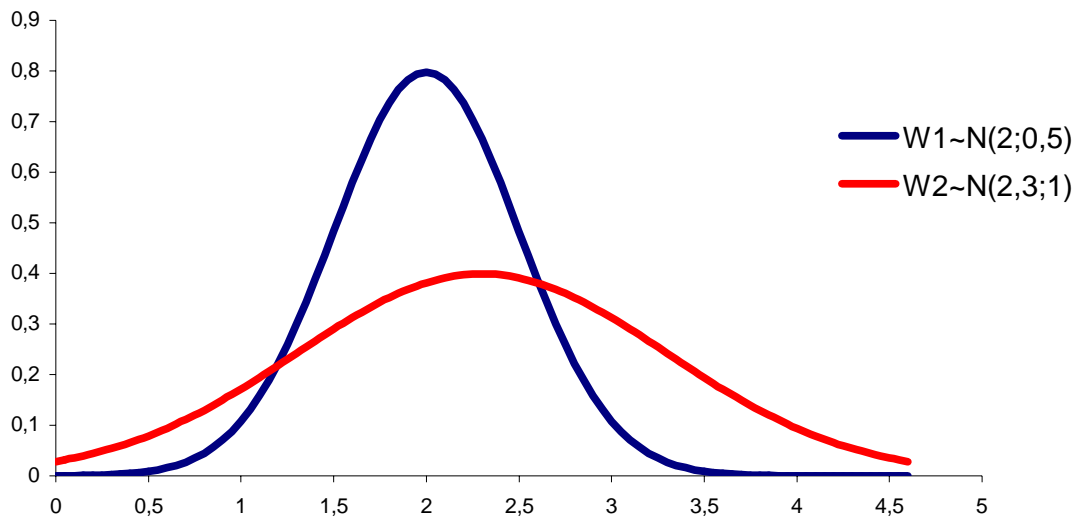


Figura 13- Função distribuição de probabilidade para duas classes

A classe W_1 apresenta média 2 e desvio padrão 0,5; enquanto a classe W_2 apresenta média 2,3 e desvio padrão 1. Podemos observar que as duas distribuições tem seu centro aproximado, sendo a curva da classe W_2 mais alargada, devido a sua maior variabilidade. Considerando os parâmetros de cada distribuição, a classificação seria realizada da seguinte forma:

$$G_1 = \frac{1}{0,5\sqrt{2\pi}} \exp\left\{\frac{-1}{2} \frac{(x_i - 2)^2}{0,5^2}\right\}$$

$$G_2 = \frac{1}{\sqrt{2\pi}} \exp\left\{\frac{-1}{2} \frac{(x_i - 2,3)^2}{1^2}\right\}$$

$$x \in W_1 \text{ se } G_1 > G_2 \Leftrightarrow x \in W_1 \text{ se } 1,19141 < x < 2,60859$$

$$x \in W_2 \text{ se } G_1 < G_2 \Leftrightarrow x \in W_2 \text{ se } x < 1,19141 \text{ ou } x > 2,60859$$

$$P(x \text{ class } W_2 / W_2) = F_2(1,19141) + (1 - F_2(2,60859)) = 0,1338 + 0,3788 = 0,5126$$

$$P(x \text{ class } W_1 / W_1) = F_1(2,60859) - F_1(1,19141) = 0,8353$$

Sendo assim, esperamos classificar corretamente 83,53% dos padrões da classe W_1 , enquanto que da classe W_2 esperamos classificar corretamente apenas 51,26% dos padrões corretamente. Principal razão para esta diferença se dá pelo fato da classe W_2 apresentar uma variabilidade significativamente maior que a classe W_1 .

Outros resultados que merecem destaque são os comportamentos semelhantes da tendência de melhora na acurácia à medida que aumentamos o número de partições nas classificações utilizando matriz de covariâncias distintas e na classificação supondo homogeneidade de variâncias. A média da acurácia na partição 10 foi a melhor nas duas formas, a única exceção se deu devido a partição 6 ter apresentado um resultado levemente superior ao encontrado na partição 8.

Ao observarmos a acurácia média de cada classe quando utilizamos matriz de covariâncias distintas na tabela 9, é possível verificar que a classe *corn notill* apresenta um comportamento bem distinto das demais classes. Esta classe apresenta uma piora relativamente grande na partição 10 em relação a partição 8, note que utilizando 8 partições a classe atingiu uma acurácia de 85% enquanto que quando foi trabalhada com 10 partições, este índice recuou para menos de 77%. Este resultado pode ser explicado por esta classe apresentar o menor tamanho de amostra disponível (860 pixels) entre as classes, e quando passamos de 8 partições para 10, estamos aumentando a dimensionalidade de 16 para 20 e com isso o número de componentes a serem estimados na matriz de covariâncias aumenta de 136 para 210; isto pode ser a causa para a grande diminuição observada na acurácia desta classe.

Através da tabela 7 nas combinações 2,3,13,14 e 15, que são as combinações que contém a classe *corn notill* podemos observar que sempre há um decréscimo na acurácia desta classe quando aumentamos o número de partições de 8 para 10.

4.4.2 Experimentos utilizando três classes

Neste experimento iremos utilizar as seis classes separando-as em dois grupos conforme sua cultura, ou seja, no primeiro grupo farão parte as classes *corn*, *corn min* e *corn notill*, e no segundo, as classes *soy clean*, *soy min* e *soy notill*. Procederemos a classificação a partir dos parâmetros do semivariograma linear calculados para cada partição do transecto de 190 bandas. A partir destes resultados poderá ser possível detectar em qual grupo de cultura a classificação correta é mais difícil, além disso, poderemos observar como se comporta o nível de acurácia à medida que alteramos o número de partições no transecto.

4.4.2.1 Resultados utilizando matriz de covariâncias distintas para cada classe

Os resultados obtidos a partir da classificação pela máxima verossimilhança gaussiana utilizando os parâmetros do semivariograma ajustado para cada conjunto de cultura pode ser observado na tabela 15 e 16.

Os valores mostrados na tabela 15 se referem ao número de pixels de teste de cada classe de soja cuja verdade terrestre é conhecida que foram classificados em cada uma das três classes possíveis de soja. Os valores da diagonal principal são os pixels que foram classificados corretamente, os valores fora da diagonal principal se referem ao número de pixels que foram classificados erroneamente. Por exemplo, na partição 8, 1886 dos 2327 pixels da classe *soy clean* foram classificados corretamente, portanto 441 foram classificados de forma errada, destes 417 foram classificados em *soy min* e 24 em *soy*

notill, perfazendo então uma acurácia de 81,5% para a classe soy clean, somando a estes resultados, os achados nas outras classes, foi possível classificar 78,19% dos pixels corretamente na partição 8. Na partição 10 este percentual se elevou para 79,53.

Tabela 15-Matriz de confusão para as três classes de soja e acurácias por partição.

Número de Partições	Classes verdadeiras	Classificação			Total	Resultados	
		soy clean	soy min	soy notill		Acurácia	Ac. Geral
4	soy clean	1902	395	30	2327	81,74	76,70
	soy min	307	2140	68	2515	85,09	
	soy notill	66	526	540	1132	47,70	
6	soy clean	1817	487	23	2327	78,08	77,35
	soy min	244	2192	79	2515	87,16	
	soy notill	74	446	612	1132	54,06	
8	soy clean	1886	417	24	2327	81,05	78,19
	soy min	302	2132	81	2515	84,77	
	soy notill	53	426	653	1132	57,69	
10	soy clean	1927	372	28	2327	82,81	79,53
	soy min	292	2122	101	2515	84,37	
	soy notill	50	380	702	1132	62,01	

Tabela 16- Matriz de confusão para as três classes de milho e acurácias por partição.

Número de Partições	Classes verdadeiras	Classificação			Total	Resultados	
		corn	corn min	corn notill		Acurácia	Ac. Geral
4	corn	1064	555	202	1821	58,43	64,99
	corn min	71	1077	73	1221	88,21	
	corn notill	71	394	395	860	45,93	
6	corn	1111	513	197	1821	61,01	69,50
	corn min	83	1061	77	1221	86,90	
	corn notill	64	256	540	860	62,79	
8	corn	1072	576	173	1821	58,87	69,61
	corn min	71	1092	58	1221	89,43	
	corn notill	30	278	552	860	64,19	
10	corn	1046	634	141	1821	57,44	67,81
	corn min	52	1135	34	1221	92,96	
	corn notill	50	345	465	860	54,07	

4.4.2.2 Resultados assumindo matriz de covariâncias iguais para todas as classes

Os resultados obtidos a partir da classificação pela máxima verossimilhança gaussiana utilizando os parâmetros do semivariograma ajustado para cada conjunto de

cultura, supondo homogeneidade de variâncias entre os parâmetros dentro de cada cultura pode ser observado na tabela 17 e 18.

Tabela 17- Matriz de confusão para as três classes de soja e acurácias por partição, supondo matriz de covariâncias iguais

Número de Partições	Classes verdadeiras	Classificação			Total	Resultados	
		soy clean	soy min	soy notill		Acurácia	Ac. Geral
4	soy clean	1822	449	56	2327	78,30	71,69
	soy min	528	1782	205	2515	70,85	
	soy notill	51	402	679	1132	59,98	
6	soy clean	1827	443	57	2327	78,51	72,36
	soy min	449	1814	252	2515	72,13	
	soy notill	93	357	682	1132	60,25	
8	soy clean	1828	445	54	2327	78,56	72,31
	soy min	499	1768	248	2515	70,30	
	soy notill	53	355	724	1132	63,96	
10	soy clean	1870	418	39	2327	80,36	72,97
	soy min	532	1762	221	2515	70,06	
	soy notill	62	343	727	1132	64,22	

Tabela 18- Matriz de confusão para as três classes de milho e acurácias por partição, supondo matriz de covariâncias iguais

Número de Partições	Classes verdadeiras	Classificação			Total	Resultados	
		corn	corn min	corn notill		Acurácia	Ac. Geral
4	corn	1182	236	403	1821	64,91	64,58
	corn min	323	760	138	1221	62,24	
	corn notill	85	197	578	860	67,21	
6	corn	1170	267	384	1821	64,25	65,48
	corn min	309	800	112	1221	65,52	
	corn notill	100	175	585	860	68,02	
8	corn	1214	279	328	1821	66,67	67,30
	corn min	303	804	114	1221	65,85	
	corn notill	75	177	608	860	70,70	
10	corn	1152	307	362	1821	63,26	66,04
	corn min	283	814	124	1221	66,67	
	corn notill	84	165	611	860	71,05	

4.4.2.3 Análise dos resultados de classificação utilizando três classes

Ao realizarmos a classificação utilizando as três classes de soja foi observada uma acurácia acima do patamar de 75%, a tabela 17 mostra a matriz de confusão para as 4 diferentes partições utilizadas. A classe que apresentou maior percentual de pixels classificados corretamente foi *soy min* e a que mostrou pior acurácia foi a classes *soy notill*. O percentual de pixels classificados corretamente apresentou relativo aumento a medida em que se aumentava o número de partições, começando em 76,70% para 4 partições, atingindo 79,53% quando trabalhamos com 10 partições.

Para as três classes de milho não foi possível atingir o mesmo percentual de pixels classificados corretamente observados nas classes de soja. A acurácia não ultrapassou a faixa dos 70% em nenhuma das partições. A classe *corn min* foi a que atingiu maior acurácia, as classes *corn* e *corn notill* alternaram o menor aproveitamento, conforme o número de partições utilizadas. Diferentemente do que se esperava o percentual de acerto não aumentou a medida que se aumentava o número de partições e portanto de variáveis aleatórias, a partição 10 mostrou resultados menos satisfatórios do que os resultados encontrados nas partições 6 e 8.

A classe *corn notill* assim como na análise utilizando duas classes, continuou a apresentar queda na acurácia quando utilizada classificação com base em 10 partições, este resultado pode ser observado na tabela 16.

Os resultados de classificação melhores para as três classes de soja podem ter sido obtidos devido ao maior número de pixels de treinamento existentes nestas classes em relação as classes de milho. São 5974 pixels de treinamento para soja contra 3902 pixels de treinamento de milho.

Considerando os resultados utilizando suposição de matriz de covariâncias iguais para cada conjunto de classes de mesma cultura foi possível observar que:

- a) Assim como na análise anterior, utilizando matriz de covariâncias distintas para cada classe, o melhor resultado de classificação foi observado quando se trabalha com as classes de soja;
- b) Da mesma forma com o observado nos experimentos utilizando 2 classes, os resultados em geral são mais satisfatórios quando não supomos matriz de covariâncias iguais, independentemente do número de partições utilizadas no estudo.
- c) Manteve-se a tendência de que classes que apresentam percentuais de acerto maiores diminuam quando se utiliza matriz de covariâncias iguais, da mesma forma, que tende a aumentar a acurácia quando se utiliza matriz de covariâncias iguais naquelas classes que apresentam menores percentuais de classificação correta, isto pode ser observado com maior facilidade nas classes de milho.
- d) A tendência normal de aumento no percentual de pixels classificados corretamente com o aumento do número de partições, não ocorreu em nenhum dos grupos de cultura.
- e) Observando a matriz de confusão para as classes de milho e soja quando utilizamos matrizes de covariâncias distintas, os percentuais de acerto de cada classe apresentam maiores diferenças entre si. Por exemplo, é muito mais

provável classificarmos um pixel de *corn* em *corn min* do que o inverso, isto tende a fazer com que o número de pixels classificados em *corn min* seja superestimado em relação ao número verdadeiro e, por conseguinte o número de pixels classificados em *corn* seja subestimado.

4.4.3 Experimentos utilizando seis classes

Para estes experimentos serão estudadas todas as seis (três de soja e três de milho) classes simultaneamente. Estes experimentos tendem a apresentar uma acurácia menor em relação aos demais experimentos, uma vez que há uma possibilidade de classificar um pixel corretamente, contra cinco de classificar de forma errada.

4.4.3.1 Resultados utilizando matriz de covariâncias distintas para cada classe

Os resultados obtidos a partir da classificação pela máxima verossimilhança gaussiana utilizando os parâmetros do semivariograma ajustado para cada diferente tipo de partição pode ser observado na tabela 19.

Tabela 19- Matriz de confusão para as seis classes estudadas e acurácias por partição

Número de Partições	Classes verdadeiras	Classificação						Total	Resultados	
		corn	corn min	soy clean	corn notill	soy min	soy notill		Acurácia	Ac. Geral
4	corn	403	344	533	82	421	38	1821	22,13	52,08
	corn min	28	967	48	35	127	16	1221	79,20	
	soy clean	91	166	1717	7	325	21	2327	73,79	
	corn notill	29	309	25	307	179	11	860	35,70	
	soy min	18	663	272	110	1415	37	2515	56,26	
	soy notill	29	525	53	76	115	334	1132	29,51	
6	corn	543	338	477	125	306	32	1821	29,82	58,16
	corn min	53	922	49	50	139	8	1221	75,51	
	soy clean	62	96	1698	11	441	19	2327	72,97	
	corn notill	32	196	12	455	148	17	860	52,91	
	soy min	26	341	210	129	1769	40	2515	70,34	
	soy notill	43	313	48	72	299	357	1132	31,54	
8	corn	483	381	444	114	366	33	1821	26,52	56,32
	corn min	37	1006	18	40	100	20	1221	82,39	
	soy clean	34	174	1719	12	367	21	2327	73,87	
	corn notill	12	232	11	499	91	15	860	58,02	
	soy min	12	550	266	169	1480	38	2515	58,85	
	soy notill	15	326	50	97	269	375	1132	33,13	
10	corn	499	430	418	76	359	39	1821	27,40	58,64
	corn min	30	1038	26	18	96	13	1221	85,01	
	soy clean	46	148	1780	13	322	18	2327	76,49	
	corn notill	20	250	16	385	166	23	860	44,77	
	soy min	8	461	268	98	1620	60	2515	64,41	
	soy notill	18	276	43	42	284	469	1132	41,43	

4.4.3.2 Resultados utilizando matriz de covariâncias iguais para todas as classes

Os resultados obtidos a partir da classificação pela máxima verossimilhança gaussiana utilizando os parâmetros do semivariograma ajustado para cada diferente tipo de partição, considerando as seis classes distintas com mesma matriz de covariâncias, pode ser observado na tabela 20.

Tabela 20- Matriz de confusão para as seis classes estudadas e acurácias por partição, supondo matriz de covariâncias iguais

Número de Partições	Classes verdadeiras	Classificação						Total	Resultados	
		corn	corn min	soy clean	corn notill	soy min	soy notill		Acurácia	Ac. Geral
4	corn	637	167	357	302	298	60	1821	34,98	46,79
	corn min	243	546	47	140	182	63	1221	44,72	
	soy clean	408	141	1463	29	261	25	2327	62,87	
	corn notill	47	148	5	538	82	40	860	62,56	
	soy min	172	419	442	322	1089	71	2515	43,30	
	soy notill	58	392	41	109	184	348	1132	30,74	
6	corn	729	267	279	333	191	22	1821	40,03	51,19
	corn min	220	643	62	107	162	27	1221	52,66	
	soy clean	363	60	1506	44	327	27	2327	64,72	
	corn notill	84	128	1	546	69	32	860	63,49	
	soy min	133	383	465	200	1281	53	2515	50,93	
	soy notill	74	246	42	110	309	351	1132	31,01	
8	corn	676	319	293	300	200	33	1821	37,12	48,18
	corn min	219	683	38	111	133	37	1221	55,94	
	soy clean	508	111	1337	37	314	20	2327	57,46	
	corn notill	68	146	2	581	32	31	860	67,56	
	soy min	168	459	534	218	1076	60	2515	42,78	
	soy notill	57	234	41	146	249	405	1132	35,78	
10	corn	732	292	290	295	185	27	1821	40,20	50,13
	corn min	204	682	63	126	126	20	1221	55,86	
	soy clean	544	103	1358	35	268	19	2327	58,36	
	corn notill	64	134	2	573	62	25	860	66,63	
	soy min	161	394	482	228	1164	86	2515	46,28	
	soy notill	48	177	43	128	294	442	1132	39,05	

4.4.3.3 Análise dos resultados de classificação utilizando seis classes

Ao realizarmos o experimento utilizando as seis classes foi observada uma acurácia acima de 50%. Em uma primeira vista este patamar pode ser considerado baixo, mas se avaliarmos que estamos trabalhando com seis classes com comportamento espectral muito semelhante, e que para cada pixel a ser classificado temos uma chance de acertar contra cinco de errar, podemos admitir que este percentual de acerto é bastante razoável.

A partição que apresentou melhores resultados foi a partição 10, nela 58,64% dos pixels foram classificados corretamente. Dentre as classes destaca-se o percentual de acertos atingido nos pixels de *corn min* 85,01%, a classe *corn* foi a que apresentou pior aproveitamento, sua acurácia não ultrapassou os 30%.

Quando utilizamos a suposição de matriz de covariâncias iguais para todas as classes, os resultados encontrados se assemelham aos resultados encontrados quando trabalhávamos com 2 ou 3 classes. A acurácia quando supomos matriz de covariâncias iguais apresenta menor acurácia independentemente do número de partições utilizadas; eleva-se o percentual de acerto nas classes que apresentavam menor acurácia, porém diminui o percentual de acerto nas classes em que se obteve resultados mais satisfatórios; o número de pixels classificados em cada classe se aproxima mais do verdadeiro número de pixels existentes em cada classe. A partição 6 apresentou maior acurácia, *soy clean* foi a classe que apresentou maior percentual de pixels classificados corretamente 64,72%, enquanto *soy notill* foi a que apresentou pior acurácia 40,03%.

5. CONCLUSÕES E RECOMENDAÇÕES

A capacidade de riqueza de informação obtida pelos sensores hiperespectrais está sendo capaz de discriminar classes com comportamentos espectrais muito semelhantes, isto não era possível até então com o uso de imagens multispectrais. Ao mesmo tempo em que se torna mais comum a disponibilização de imagens hiperespectrais, surge o problema na utilização de dados com alta dimensionalidade. Com o aumento da dimensionalidade, é necessário se estimar um número muito grande de parâmetros e para isso é necessário contar com um tamanho de amostra consideravelmente grande para que estas estimativas sejam confiáveis. Em situações reais dificilmente contamos com amostras abundantes, normalmente o tamanho de amostra é insuficiente para estimarmos com precisão principalmente o segundo momento (matriz de covariâncias). Este fato resulta normalmente numa performance insatisfatória dos classificadores, uma vez que os parâmetros utilizados foram estimados com base em amostras insuficientes. Este problema vem sendo fruto de muito estudo entre a comunidade científica internacional.

Em nosso trabalho, estamos sugerindo a utilização de técnicas da geoestatística, mais precisamente da ferramenta semivariograma, com o intuito de reduzir a dimensionalidade do problema sem que haja perda significativa de informação, supondo que os parâmetros calculados para cada partição do transecto de bandas do sensor AVIRIS sejam eficazes para uma boa discriminação de classes que apresentem comportamento espectrais muito semelhantes.

A idéia inicial do uso do semivariograma como redutor de dimensionalidade, se deveu ao fato de que cada banda pode ser autocorrelacionada com suas bandas anteriores, e que esta autocorrelação pode acontecer em níveis diferentes mesmo para classes semelhantes. Se tal fato fosse verdadeiro, poderíamos considerar que os parâmetros do semivariograma traduzem uma idéia de autocorrelação e, portanto, estes parâmetros muito provavelmente seriam eficientes para discriminação de classes. Além disso, com a redução da dimensionalidade poderíamos utilizar um classificador tradicional conforme a distribuição observada nestes parâmetros.

Os resultados encontrados em nosso estudo indicaram que o modelo de semivariograma linear se ajusta melhor aos dados observáveis do semivariograma experimental calculados para as partições de bandas do sensor AVIRIS. O nível de classificação correta atingindo pelo método proposto pode ser considerado satisfatório. Quando utilizamos duas classes obtivemos uma acurácia média levemente superior a 85% com um pico de 97,27% quando trabalhamos com a separação das classes milho plantio direto (corn notill) e soja plantio lavrado (soy clean). Os resultados ao trabalharmos com três classes de mesma cultura se revelaram em patamares diferentes para a cultura de soja e milho. A classificação utilizando três classes de soja chegou próxima a 80%, enquanto que a acurácia para as culturas de milho teve seu pico em 69,50%. Uma possível explicação para esta diferença no nível de acurácia poderia ser o tamanho de amostra mais reduzido nas classes de milho em relação às classes de soja. Se trabalhássemos com tamanho de amostras iguais para cada classe talvez a acurácia nestas duas culturas fossem mais próximas. Ao utilizarmos as seis classes de estudo, foi possível obter uma acurácia de 58,64%; evidenciando que a técnica trouxe acréscimo, mas que carece de mais investigações e estudos, com a possibilidade de se atingir níveis maiores de acurácia. Ressaltamos que as seis classes trabalhadas apresentam um comportamento espectral muito

semelhante o que torna inviável, por exemplo, a utilização de dados de sensores multispectrais como LANDSAT que apresenta apenas 6 bandas.

Os resultados de classificação apontaram para um nível de acurácia mais alto naquelas classes que apresentavam menor variabilidade, também foi possível verificar que a classe com notill apresentava uma tendência de decréscimo na acurácia quando se aumentava o número de partições e, portanto, de parâmetros, por esta contar com um menor tamanho de amostra disponível.

A utilização do classificador pela máxima verossimilhança gaussiano é plausível, uma vez que a simples análise de histograma dos parâmetros calculados para cada classe evidencia um comportamento muito similar a distribuição normal.

A idéia de utilizar matriz de covariâncias iguais para comparar os resultados, apesar de trazer resultados interessantes (aumenta a acurácia de classes com maior variabilidade e diminui a acurácia das classes com menor variabilidade) e podem ser satisfatórios dependendo do tipo de abordagem que se queira dar, pois aproxima os valores de acurácia para cada classe analisada, foi utilizada neste trabalho, mas carece de um embasamento maior, não pretendemos sugerir que esta metodologia traga resultados satisfatórios, uma vez que não esta comprovada sua eficácia nos demais casos.

Como sugestões para trabalhos futuros, podemos citar os seguintes aspectos:

- 1) Estudos adicionais da técnica, como observação do comportamento de acurácia segundo diferentes tipos de distribuição dos dados, através de simulações.

- 2) Buscar a criação de um critério para subdivisão do transecto de bandas do sensor AVIRIS, tal que seja minimizada a probabilidade de erro e, portanto sejam criados intervalos de bandas, não necessariamente com o mesmo tamanho como neste trabalho, e que sejam capazes de trazer melhores resultados.

- 3) Utilização de outros parâmetros que sejam eficazes para discriminar classes com comportamento espectral semelhante, juntando-se aos parâmetros do semivariograma que se mostraram eficientes na classificação.

6. BIBLIOGRAFIA

BAILEY, T.C. and GATRELL, A. C. (1995) *Interactive Spatial Data Analysis*. London: Longman.

BITTENCOURT, H.R. (2001). *Comparação da discriminação logística com o método da máxima verossimilhança gaussiana na classificação de imagens digitais*. CEPSRM,UFRGS, Porto Alegre.

BLAIS, R.A. and CARLIER,P.A. (1967). *Applications of geostatistics in ore evaluation, in Ore Reserve Estimation and Grade Control. Canadian Institute for Mining & Metallurgy Special Publication 9, 41-68.*

BURGESS, T. M. and WEBSTER, R. (1980) *Optimal interpolation and isarithmic mapping of soil properties. The semi-variogram and punctual Kriging*. Journal of Soil Science. p. 315-331.

BURROUGH, P.A.(1987) *Principles of geographical information systems for land resources assessment*. Oxford.Clarendon Press.

CAMPOS, T. E. (2001). *Técnicas de Seleção de Características com Aplicações em Reconhecimento de Faces*.

CRESSIE, N. (1991). *Statistics for Spatial Data*, John Wiley.Chichester.

CURRAN, P. J. (1988).*The Semivariogram in Remote Sensing: An Introduction. Remote Sensing of Environment*, vol. 24, p. 493-507.

DAVID, M (1977). *Geoestatistical ore reserve estimation*. New York, Elsevier Scientific.

DELFINER P. and. Delhomme, J.P (1975) *Optimum interpolation by kriging*. In: *J.D.Davis and M.J.McCullagh (eds), Display and Analysis of Spatial Data*. New York. John Wiley and Sons. p. 96-114.

DIDIER,D. (1997).*Fuzzy Information Engineering*. Wiley Computer Publishing,USA.

DUDA,R.O. and HART,P.E. (1973).*Pattern Classification and Scene Analysis*. Wiley-Interscience, USA.

EFRON,B. and TIBSHIRANI,R. (1993) *An Introduction to the Bootstrap*. Chapman and Hall.

ERBERT, M. (2001). *Uso da Análise discriminante regularizada no reconhecimento de padrões em imagens digitais hiperespectral de sensoriamento remoto*. CEPSRM,UFRGS, Porto Alegre

FUKUNAGA, K. (1990). *Statistical Pattern Reconigition*. 2^a ed. Academic Press.

- HUIJBREGTS, C. (1975). *Regionalized variables and quantitative analysis of spatial data*, in *Display and analysis of spatial data*. Ed. J.C.Davis and M.J.McCullagh, J.Wiley and Sons, London, 38-53.
- ISAAKS, E. H. and SRIVASTAVA, R. M. (1989). *An Introduction to Applied Geostatistics*. Oxford University Press. Oxford.
- JAIN, A. K. and ROBERT, P. W. *Statistical Pattern Recognition: A Review*. IEEE Transactions on Pattern Analysis and machine Intelligence, vol. 22, p. 4-34, 2000.
- JOHNSON, R. A. and WICHERN, D. W. (1982) *Applied Multivariate Statistical Analysis*. Prentice Hall.New Jersey.
- KITANIDIS, P.K. and VOMVORIS, E.G.(1983). *A geoestatistical approach to the inverse problem in groundwater modelling and one dimensional simulations*. Water Resources Research. p. 677-690.
- KRIGE, D.G (1951) A statistical approach to some basic mine evaluation problems on the Witwatersrand. Johannesburg Chemistry Metallurgy Mining Society South African. p.119-139.
- LAJAUNIE,C. (1984) *In: Geostatistics for Natural Resources Characterization*, ed. A.G. Journel.p 877-891.
- LILLESAND, T. and KIEFER, P. (1987). *Remote Sensing and Image Interpretation*. 3^a ed. Wiley. Chinchester.

MARDIA, K. V. and KENT, J. T. (1979). *Multivariate Analysis*. Academic Press. London.

NOVO, E. (1992), *Sensoriamento Remoto, Princípios e aplicações*. 2ª ed. São Paulo: Edgar Blücher.

RICHARDS, J. (1993). *Remote Sensing Digital Image Analysis*. 2ª ed. Springer-Verlag.

RIPLEY, B. D. (1981). *Spatial Statistics*. John Wiley. Chinchester.

STEFFEN, C. A. MORAES, E. C. e GAMA, F. (1996) *Radiometria Óptica espectral*. VIII Simpósio Nacional de Sensoriamento Remoto. Salvador.

THERRIEN, C. (1989). *Decision Estimation and Classification*. John Wiley.

Referências Eletrônicas (Internet):

INPE- Instituto de Pesquisas Espaciais – Manual do SPRING, disponível em:

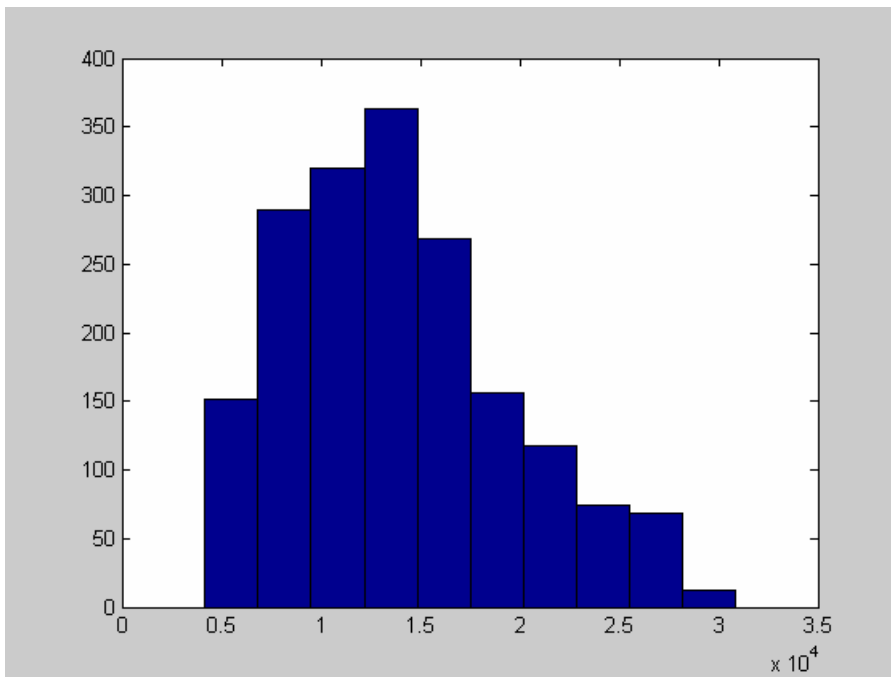
<http://www.dpi.inpe.br/spring/usuario/geoestati.htm>

JPL – Jet Propulsion Laboratory – NASA – What's AVIRIS? Disponível em:

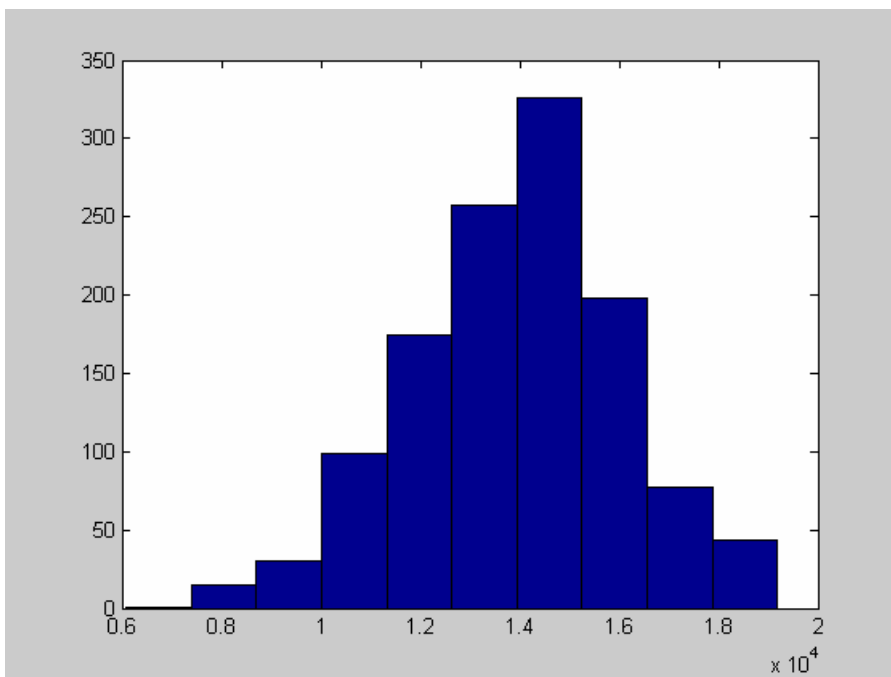
<http://makalu.jpl.nasa.gov>

7. ANEXOS

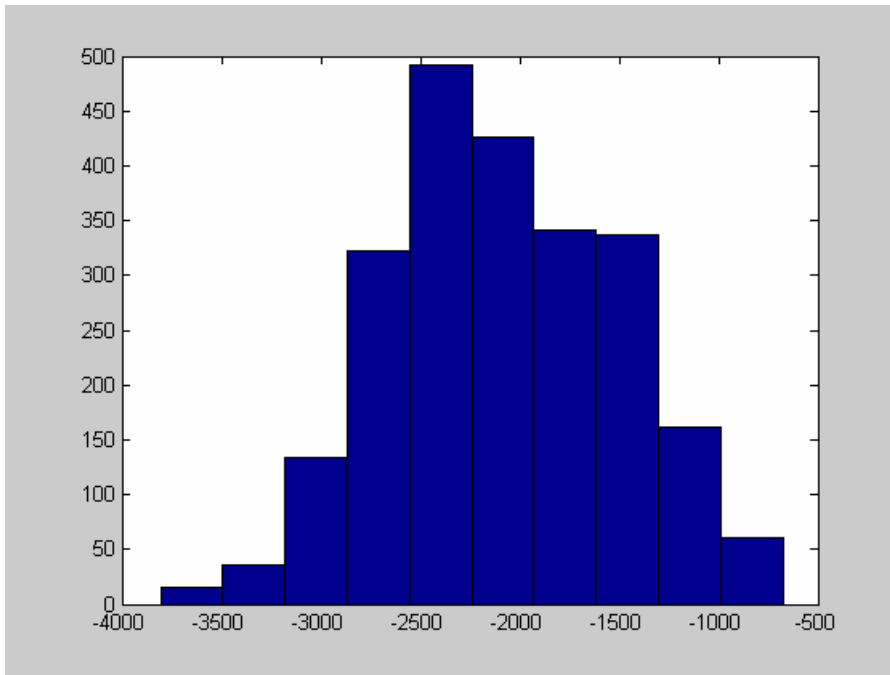
Histograma classe com 4 partições – 2ª partição – parâmetro contribuição



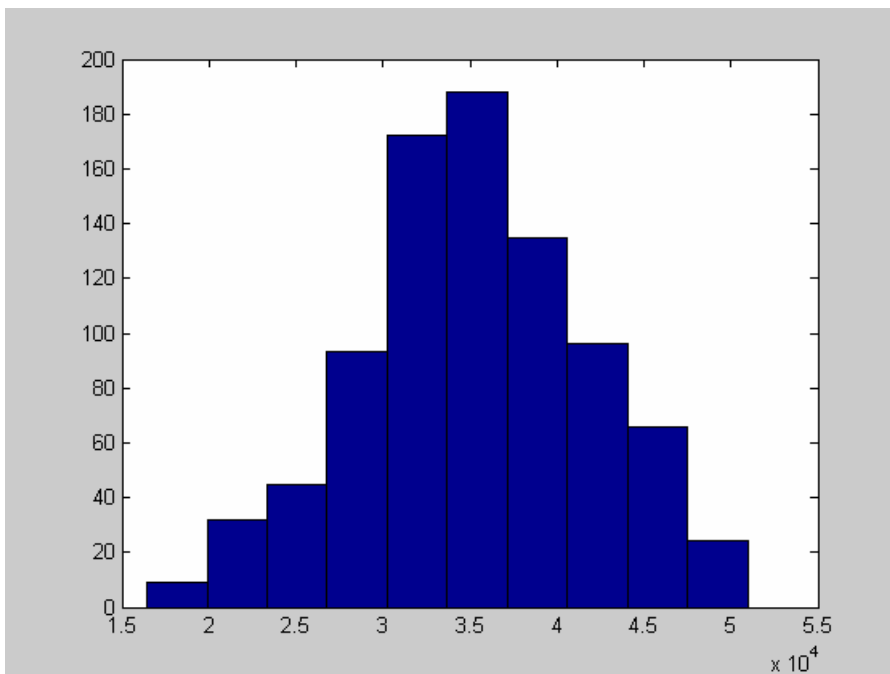
Histograma classe com min 4 partições – 3ª partição – parâmetro contribuição



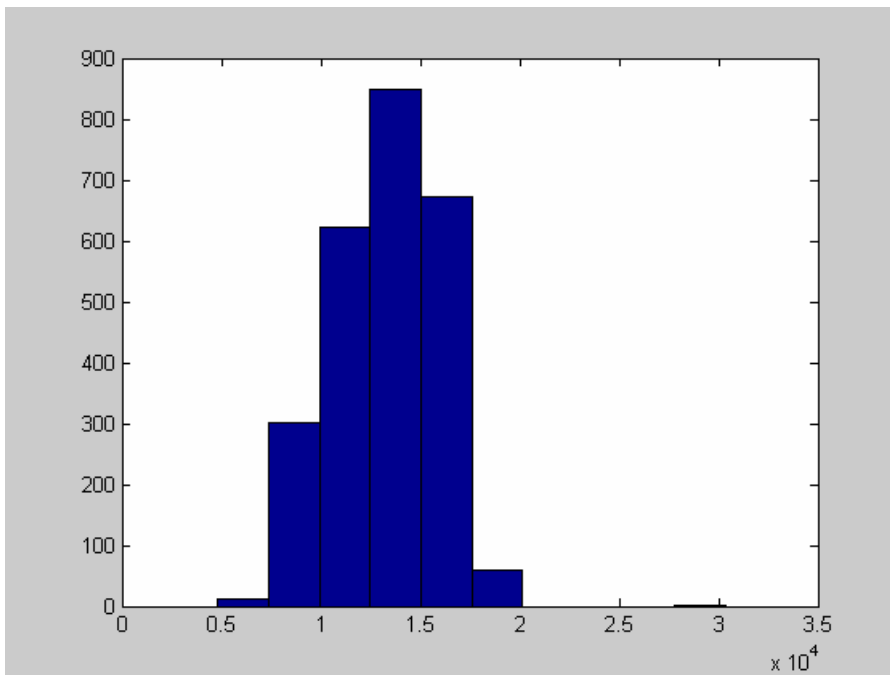
Histograma classe soy clean 4 partições – 4ª partição – parâmetro pepita



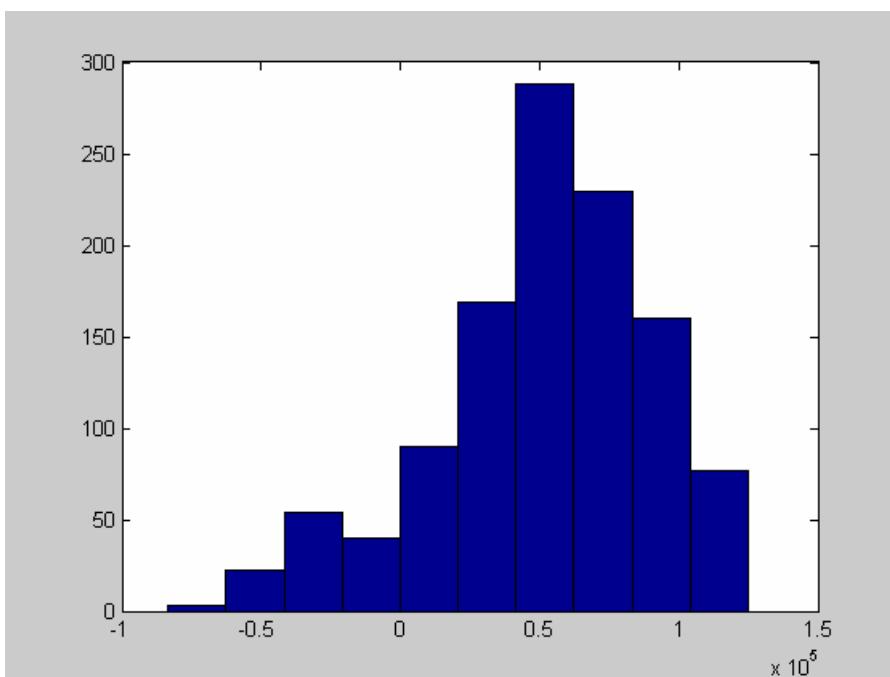
Histograma classe corn notill 4 partições – 3ª partição – parâmetro pepita



Histograma classe soy min 4 partições – 3ª partição – parâmetro contribuição



Histograma classe soy notill 4 partições – 1ª partição – parâmetro pepita



Médias para os parâmetros do semivariograma linear ajustado para cada classe utilizando 4 partições

Classes	Partições							
	Efeito Pepita (Coeficiente linear)				Contribuição (Coeficiente angular)			
	1	2	3	4	1	2	3	4
corn	6,97E+4	5,09E+5	3,05E+4	-1,96E+3	1,21E+4	1,40E+4	1,31E+4	7,87E+2
corn min	7,78E+4	5,05E+5	3,19E+4	-2,16E+3	6,52E+3	1,38E+4	1,39E+4	8,44E+2
soy clean	5,56E+4	4,78E+5	3,11E+4	-2,09E+3	9,76E+3	1,20E+4	1,36E+4	8,66E+2
corn notill	9,52E+4	5,94E+5	3,53E+4	-2,28E+3	9,28E+3	1,66E+4	1,52E+4	8,68E+2
soy min	6,26E+4	4,64E+5	2,94E+4	-2,09E+3	7,87E+3	1,17E+4	1,33E+4	8,14E+2
soy notill	5,15E+4	4,83E+5	2,77E+4	-1,69E+3	2,28E+4	1,55E+4	1,12E+4	6,34E+2

Matriz de covariâncias classe corn – 4 partições

2,89E+09	8,88E+09	4,87E+08	-2,76E+07	-3,36E+08	2,84E+08	1,77E+08	1,15E+07
8,88E+09	2,91E+10	1,59E+09	-8,87E+07	-8,97E+08	9,20E+08	5,80E+08	3,65E+07
4,87E+08	1,59E+09	9,05E+07	-5,17E+06	-5,58E+07	4,92E+07	3,31E+07	2,14E+06
-2,76E+07	-8,87E+07	-5,17E+06	3,48E+05	4,21E+06	-2,54E+06	-2,10E+06	-1,37E+05
-3,36E+08	-8,97E+08	-5,58E+07	4,21E+06	8,20E+07	-2,33E+07	-2,40E+07	-1,69E+06
2,84E+08	9,20E+08	4,92E+07	-2,54E+06	-2,33E+07	3,08E+07	1,70E+07	1,07E+06
1,77E+08	5,80E+08	3,31E+07	-2,10E+06	-2,40E+07	1,70E+07	1,32E+07	8,38E+05
1,15E+07	3,65E+07	2,14E+06	-1,37E+05	-1,69E+06	1,07E+06	8,38E+05	5,63E+04

Matriz de covariâncias classe corn min– 4 partições

1,20E+09	3,20E+09	1,54E+08	-6,90E+06	-1,02E+08	1,15E+08	4,70E+07	3,55E+06
3,20E+09	1,01E+10	5,06E+08	-2,46E+07	-2,22E+08	3,34E+08	1,70E+08	1,16E+07
1,54E+08	5,06E+08	3,04E+07	-1,64E+06	-1,55E+07	1,53E+07	1,06E+07	7,01E+05
-6,90E+06	-2,46E+07	-1,64E+06	1,29E+05	1,13E+06	-5,95E+05	-7,32E+05	-4,84E+04
-1,02E+08	-2,22E+08	-1,55E+07	1,13E+06	2,35E+07	-6,18E+06	-6,14E+06	-4,65E+05
1,15E+08	3,34E+08	1,53E+07	-5,95E+05	-6,18E+06	1,25E+07	4,40E+06	3,25E+05
4,70E+07	1,70E+08	1,06E+07	-7,32E+05	-6,14E+06	4,40E+06	4,54E+06	2,86E+05
3,55E+06	1,16E+07	7,01E+05	-4,84E+04	-4,65E+05	3,25E+05	2,86E+05	2,02E+04

Matriz de covariâncias classe soy clean – 4 partições

2,11E+09	6,93E+09	4,01E+08	-2,08E+07	-2,69E+08	2,09E+08	1,40E+08	8,45E+06
6,93E+09	2,47E+10	1,47E+09	-7,70E+07	-8,53E+08	7,15E+08	5,22E+08	3,15E+07
4,01E+08	1,47E+09	9,57E+07	-5,29E+06	-5,59E+07	4,01E+07	3,43E+07	2,22E+06
-2,08E+07	-7,70E+07	-5,29E+06	3,30E+05	3,31E+06	-1,93E+06	-2,01E+06	-1,37E+05
-2,69E+08	-8,53E+08	-5,59E+07	3,31E+06	4,88E+07	-2,29E+07	-2,04E+07	-1,37E+06
2,09E+08	7,15E+08	4,01E+07	-1,93E+06	-2,29E+07	2,25E+07	1,36E+07	7,85E+05
1,40E+08	5,22E+08	3,43E+07	-2,01E+06	-2,04E+07	1,36E+07	1,29E+07	8,41E+05
8,45E+06	3,15E+07	2,22E+06	-1,37E+05	-1,37E+06	7,85E+05	8,41E+05	5,84E+04

Matriz de covariâncias classe corn notill– 4 partições

1,30E+09	3,48E+09	2,13E+08	-1,49E+07	-2,29E+08	1,00E+08	8,47E+07	6,34E+06
3,48E+09	1,23E+10	6,65E+08	-3,41E+07	-2,32E+08	3,95E+08	2,32E+08	1,46E+07
2,13E+08	6,65E+08	4,31E+07	-2,79E+06	-3,68E+07	1,85E+07	1,62E+07	1,17E+06
-1,49E+07	-3,41E+07	-2,79E+06	2,76E+05	4,87E+06	-6,08E+05	-1,33E+06	-1,12E+05
-2,29E+08	-2,32E+08	-3,68E+07	4,87E+06	1,28E+08	4,98E+06	-2,10E+07	-1,97E+06
1,00E+08	3,95E+08	1,85E+07	-6,08E+05	4,98E+06	1,48E+07	5,42E+06	2,84E+05
8,47E+07	2,32E+08	1,62E+07	-1,33E+06	-2,10E+07	5,42E+06	7,14E+06	5,45E+05
6,34E+06	1,46E+07	1,17E+06	-1,12E+05	-1,97E+06	2,84E+05	5,45E+05	4,62E+04

Matriz de covariâncias classe soy min – 4 partições

1,39E+09	4,16E+09	2,22E+08	-1,33E+07	-1,70E+08	1,32E+08	8,50E+07	4,99E+06
4,16E+09	1,37E+10	7,05E+08	-4,03E+07	-3,84E+08	4,41E+08	2,67E+08	1,49E+07
2,22E+08	7,05E+08	4,28E+07	-2,55E+06	-3,05E+07	2,12E+07	1,57E+07	1,01E+06
-1,33E+07	-4,03E+07	-2,55E+06	1,92E+05	2,57E+06	-1,06E+06	-1,09E+06	-7,34E+04
-1,70E+08	-3,84E+08	-3,05E+07	2,57E+06	5,48E+07	-6,77E+06	-1,37E+07	-1,02E+06
1,32E+08	4,41E+08	2,12E+07	-1,06E+06	-6,77E+06	1,58E+07	7,40E+06	3,86E+05
8,50E+07	2,67E+08	1,57E+07	-1,09E+06	-1,37E+07	7,40E+06	6,65E+06	4,11E+05
4,99E+06	1,49E+07	1,01E+06	-7,34E+04	-1,02E+06	3,86E+05	4,11E+05	2,96E+04

Matriz de covariâncias classe soy notill – 4 partições

1,47E+09	1,14E+09	1,99E+08	-2,39E+07	-9,12E+08	-1,15E+08	1,28E+08	9,07E+06
1,14E+09	5,13E+09	2,06E+08	-6,14E+06	1,08E+08	1,83E+08	5,88E+07	2,78E+06
1,99E+08	2,06E+08	3,61E+07	-3,92E+06	-1,32E+08	-1,64E+07	2,13E+07	1,50E+06
-2,39E+07	-6,14E+06	-3,92E+06	5,44E+05	1,96E+07	3,17E+06	-2,80E+06	-2,05E+05
-9,12E+08	1,08E+08	-1,32E+08	1,96E+07	7,89E+08	1,38E+08	-1,00E+08	-7,33E+06
-1,15E+08	1,83E+08	-1,64E+07	3,17E+06	1,38E+08	3,14E+07	-1,59E+07	-1,18E+06
1,28E+08	5,88E+07	2,13E+07	-2,80E+06	-1,00E+08	-1,59E+07	1,49E+07	1,06E+06
9,07E+06	2,78E+06	1,50E+06	-2,05E+05	-7,33E+06	-1,18E+06	1,06E+06	7,78E+04

Coefficientes de determinação por pixel ao ajuste do semivariograma linear para 4 partições

Primeiros 40 pixels para cada classe em cada partição – corn e corn min

corn				corn min			
1	2	3	4	1	2	3	4
0,9965	0,8918	0,9925	0,9878	0,9808	0,8823	0,9940	0,9829
0,9959	0,8895	0,9931	0,9874	0,9868	0,8818	0,9940	0,9825
0,9965	0,8879	0,9923	0,9888	0,9728	0,8783	0,9937	0,9833
0,9972	0,8859	0,9935	0,9884	0,9500	0,8797	0,9942	0,9830
0,9972	0,8903	0,9936	0,9878	0,9704	0,8761	0,9939	0,9828
0,9952	0,8938	0,9929	0,9884	0,9772	0,8713	0,9937	0,9830
0,9868	0,8857	0,9930	0,9873	0,9753	0,8734	0,9945	0,9835
0,9965	0,8925	0,9923	0,9879	0,8978	0,8758	0,9938	0,9829
0,9958	0,8900	0,9922	0,9882	0,9209	0,8757	0,9944	0,9834
0,9963	0,8890	0,9927	0,9888	0,9722	0,8726	0,9943	0,9822
0,9965	0,8866	0,9929	0,9866	0,9809	0,8822	0,9940	0,9834
0,9962	0,8833	0,9935	0,9883	0,9835	0,8793	0,9943	0,9821
0,9947	0,8869	0,9932	0,9879	0,9729	0,8777	0,9941	0,9839
0,9910	0,8862	0,9929	0,9873	0,9608	0,8782	0,9941	0,9843
0,9985	0,8737	0,9945	0,9877	0,9678	0,8792	0,9938	0,9832
0,9983	0,8745	0,9948	0,9870	0,9706	0,8733	0,9943	0,9827
0,9939	0,8729	0,9942	0,9881	0,9630	0,8731	0,9941	0,9832
0,9974	0,8755	0,9944	0,9876	0,9485	0,8732	0,9938	0,9834
0,9907	0,8757	0,9937	0,9873	0,9604	0,8747	0,9943	0,9819
0,9869	0,8793	0,9939	0,9866	0,9746	0,8754	0,9948	0,9829
0,9754	0,8759	0,9937	0,9870	0,9843	0,8779	0,9946	0,9831
0,9926	0,8791	0,9932	0,9873	0,9880	0,8806	0,9941	0,9834
0,9983	0,8804	0,9951	0,9883	0,9759	0,8780	0,9939	0,9829
0,9984	0,8734	0,9954	0,9863	0,9640	0,8818	0,9938	0,9833
0,9845	0,8748	0,9936	0,9871	0,9691	0,8770	0,9937	0,9829
0,9424	0,8833	0,9940	0,9873	0,9749	0,8738	0,9938	0,9832
0,9631	0,8822	0,9936	0,9858	0,9589	0,8705	0,9938	0,9847
0,9961	0,8781	0,9944	0,9861	0,9618	0,8714	0,9941	0,9837
0,9832	0,8770	0,9940	0,9855	0,9782	0,8750	0,9943	0,9829
0,9841	0,8808	0,9941	0,9872	0,9876	0,8710	0,9949	0,9819
0,9896	0,8818	0,9938	0,9869	0,9882	0,8817	0,9939	0,9829
0,9943	0,8798	0,9940	0,9872	0,9915	0,8757	0,9941	0,9822
0,9983	0,8785	0,9947	0,9872	0,9751	0,8756	0,9943	0,9827
0,9967	0,8901	0,9920	0,9868	0,9592	0,8798	0,9939	0,9829
0,9958	0,8897	0,9926	0,9877	0,9691	0,8772	0,9939	0,9829
0,9956	0,8877	0,9927	0,9871	0,9631	0,8750	0,9935	0,9835
0,9967	0,8863	0,9925	0,9872	0,9495	0,8711	0,9938	0,9836
0,9958	0,8840	0,9930	0,9873	0,9726	0,8705	0,9942	0,9826
0,9946	0,8863	0,9928	0,9870	0,9930	0,8752	0,9943	0,9836
0,9954	0,8833	0,9936	0,9882	0,9921	0,8739	0,9950	0,9823

Coefficientes de determinação por pixel ao ajuste do semivariograma linear para 4 partições

Primeiros 40 pixels para cada classe em cada partição – soy clean e corn notill

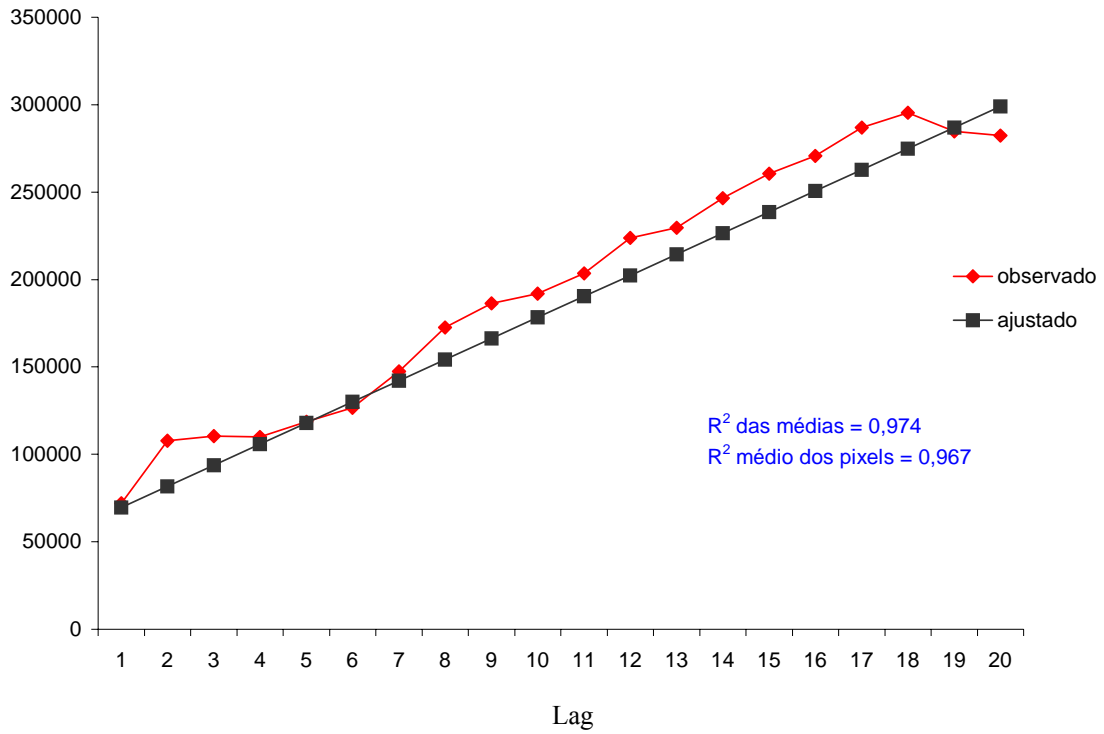
soy clean				corn notill			
1	2	3	4	1	2	3	4
0,9356	0,8680	0,9935	0,9874	0,9688	0,8748	0,9947	0,9822
0,9631	0,8697	0,9938	0,9881	0,9620	0,8736	0,9940	0,9831
0,9973	0,8637	0,9948	0,9871	0,9742	0,8755	0,9944	0,9825
0,9968	0,8647	0,9950	0,9877	0,9767	0,8734	0,9942	0,9817
0,9907	0,8652	0,9943	0,9878	0,9710	0,8760	0,9945	0,9819
0,9962	0,8677	0,9945	0,9881	0,9605	0,8763	0,9943	0,9819
0,9973	0,8651	0,9950	0,9881	0,9853	0,8788	0,9946	0,9811
0,9915	0,8674	0,9941	0,9879	0,9641	0,8747	0,9939	0,9818
0,9842	0,8651	0,9938	0,9886	0,9583	0,8717	0,9946	0,9817
0,9937	0,8680	0,9942	0,9872	0,9798	0,8744	0,9941	0,9826
0,9898	0,8659	0,9944	0,9874	0,9784	0,8724	0,9938	0,9838
0,9979	0,8874	0,9951	0,9860	0,9613	0,8749	0,9947	0,9819
0,9969	0,8719	0,9956	0,9863	0,9642	0,8754	0,9942	0,9827
0,9976	0,8714	0,9956	0,9866	0,9824	0,8770	0,9940	0,9830
0,9963	0,8650	0,9952	0,9858	0,9662	0,8760	0,9945	0,9816
0,9973	0,8652	0,9952	0,9867	0,9697	0,8736	0,9946	0,9821
0,9978	0,8666	0,9947	0,9867	0,9843	0,8764	0,9943	0,9839
0,9973	0,8624	0,9950	0,9871	0,9626	0,8722	0,9943	0,9825
0,9971	0,8656	0,9946	0,9866	0,9641	0,8769	0,9943	0,9821
0,9980	0,8648	0,9951	0,9862	0,9761	0,8779	0,9940	0,9839
0,9979	0,8667	0,9949	0,9878	0,9796	0,8772	0,9946	0,9818
0,9984	0,8687	0,9948	0,9870	0,9571	0,8759	0,9942	0,9823
0,9935	0,8677	0,9945	0,9878	0,9609	0,8765	0,9945	0,9830
0,9957	0,8657	0,9942	0,9865	0,9696	0,8803	0,9944	0,9832
0,9962	0,8696	0,9949	0,9869	0,9556	0,8780	0,9937	0,9824
0,9760	0,8647	0,9947	0,9875	0,9544	0,8778	0,9939	0,9833
0,9598	0,8718	0,9936	0,9874	0,9497	0,8814	0,9937	0,9834
0,9836	0,8725	0,9940	0,9867	0,9428	0,8817	0,9938	0,9827
0,9924	0,8669	0,9947	0,9888	0,9508	0,8765	0,9937	0,9829
0,9842	0,8706	0,9939	0,9868	0,9303	0,8757	0,9938	0,9841
0,9834	0,8716	0,9944	0,9883	0,9174	0,8803	0,9941	0,9829
0,9757	0,8725	0,9942	0,9877	0,9554	0,8771	0,9932	0,9827
0,9894	0,8701	0,9940	0,9874	0,9498	0,8772	0,9940	0,9829
0,9977	0,8831	0,9948	0,9879	0,9399	0,8774	0,9940	0,9831
0,9967	0,8731	0,9951	0,9868	0,9284	0,8825	0,9936	0,9840
0,9960	0,8750	0,9950	0,9872	0,9433	0,8784	0,9940	0,9831
0,9971	0,8689	0,9949	0,9860	0,9611	0,8777	0,9946	0,9833
0,9971	0,8699	0,9947	0,9879	0,9310	0,8808	0,9938	0,9838
0,9979	0,8686	0,9952	0,9882	0,9629	0,8716	0,9935	0,9827
0,9974	0,8679	0,9948	0,9877	0,9624	0,8774	0,9945	0,9831

Coefficientes de determinação por pixel ao ajuste do semivariograma linear para 4 partições

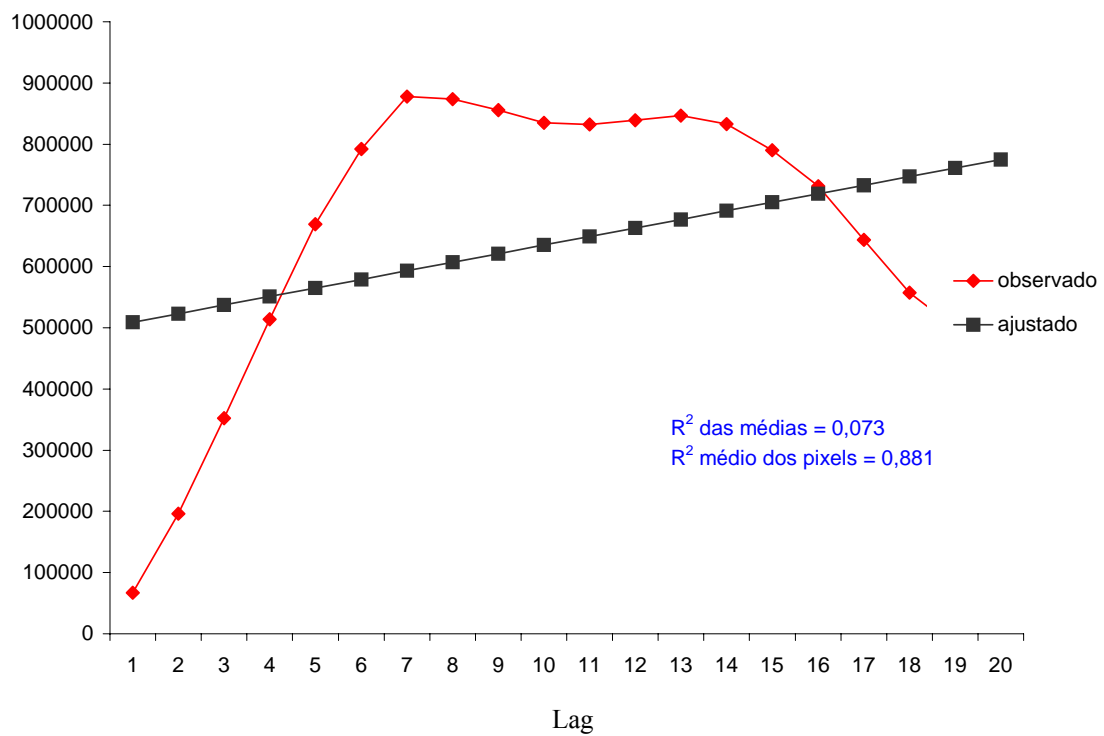
Primeiros 40 pixels para cada classe em cada partição – soy min e soy notill

soy min				soy notill			
1	2	3	4	1	2	3	4
0,9981	0,8682	0,9947	0,9859	0,9854	0,8758	0,9943	0,9828
0,9978	0,8681	0,9948	0,9876	0,9732	0,8728	0,9945	0,9822
0,9973	0,8685	0,9946	0,9863	0,9633	0,8751	0,9940	0,9816
0,9978	0,8712	0,9951	0,9859	0,9641	0,8782	0,9941	0,9822
0,9955	0,8647	0,9950	0,9859	0,9304	0,8773	0,9939	0,9824
0,9798	0,8741	0,9937	0,9871	0,9465	0,8732	0,9939	0,9833
0,9835	0,8661	0,9939	0,9861	0,9506	0,8763	0,9938	0,9820
0,9976	0,8722	0,9952	0,9864	0,9623	0,8756	0,9945	0,9814
0,9981	0,8691	0,9948	0,9869	0,9519	0,8755	0,9944	0,9821
0,9979	0,8666	0,9951	0,9858	0,9701	0,8761	0,9940	0,9836
0,9970	0,8695	0,9950	0,9861	0,9688	0,8767	0,9942	0,9824
0,9979	0,8716	0,9954	0,9850	0,9678	0,8759	0,9939	0,9829
0,9873	0,8654	0,9941	0,9851	0,9861	0,8747	0,9945	0,9836
0,9783	0,8734	0,9936	0,9866	0,9813	0,8756	0,9944	0,9835
0,9871	0,8669	0,9935	0,9859	0,9723	0,8785	0,9947	0,9817
0,9979	0,8688	0,9957	0,9869	0,9749	0,8773	0,9946	0,9825
0,9960	0,8722	0,9945	0,9859	0,9457	0,8754	0,9937	0,9821
0,9980	0,8769	0,9949	0,9867	0,9501	0,8732	0,9945	0,9825
0,9853	0,8665	0,9941	0,9859	0,9558	0,8781	0,9941	0,9817
0,9533	0,8704	0,9939	0,9867	0,9574	0,8720	0,9942	0,9816
0,9685	0,8752	0,9938	0,9854	0,9487	0,8742	0,9942	0,9840
0,9826	0,8784	0,9933	0,9874	0,9574	0,8754	0,9941	0,9829
0,9943	0,8863	0,9938	0,9873	0,9540	0,8772	0,9941	0,9828
0,9972	0,8797	0,9939	0,9877	0,9620	0,8746	0,9943	0,9831
0,9981	0,8782	0,9947	0,9860	0,9553	0,8789	0,9938	0,9833
0,9981	0,8811	0,9943	0,9855	0,9814	0,8735	0,9945	0,9824
0,9981	0,8703	0,9949	0,9871	0,9838	0,8756	0,9945	0,9820
0,9978	0,8662	0,9950	0,9874	0,9870	0,8771	0,9945	0,9817
0,9981	0,8679	0,9951	0,9839	0,9839	0,8752	0,9947	0,9823
0,9981	0,8709	0,9958	0,9861	0,9189	0,8775	0,9936	0,9824
0,9962	0,8640	0,9947	0,9857	0,9906	0,8768	0,9946	0,9819
0,9685	0,8692	0,9935	0,9865	0,9745	0,8785	0,9948	0,9819
0,9734	0,8711	0,9943	0,9857	0,9701	0,8764	0,9946	0,9815
0,9942	0,8718	0,9943	0,9848	0,9678	0,8732	0,9946	0,9827
0,9837	0,8700	0,9938	0,9849	0,9575	0,8760	0,9942	0,9821
0,9977	0,8723	0,9946	0,9860	0,9358	0,8756	0,9940	0,9827
0,9983	0,8760	0,9949	0,9864	0,9388	0,8742	0,9942	0,9823
0,9863	0,8690	0,9940	0,9854	0,9652	0,8705	0,9942	0,9827
0,9899	0,8647	0,9946	0,9857	0,9463	0,8738	0,9944	0,9835
0,9864	0,8693	0,9944	0,9855	0,9637	0,8757	0,9946	0,9829

Comparação do semivariograma experimental médio observado e semivariograma linear médio ajustado - 1ª partição classe corn (1821 pixels)



Comparação semivariograma experimental médio observado e semivariograma linear médio ajustado - 2ª partição classe corn (1821 pixels)



Comparação semivariograma experimental médio observado e semivariograma linear médio ajustado - 3ª partição classe corn (1821 pixels)

