



Ministério da
**Ciência, Tecnologia
e Inovação**



**Observação da influência do uso de séries temporais no mapeamento
de formações campestres nativas e pastagens cultivadas no Cerrado
brasileiro**

Wanderson Santos Costa

Monografia do curso de Introdução ao
Geoprocessamento (SER-300-4), lecionada
pelo Dr. Antônio Miguel Vieira Monteiro.

INPE
São José dos Campos
2014

LISTA DE FIGURAS

	<u>Pág.</u>
Figura 2.1. Fitofisionomias do bioma Cerrado.....	4
Figura 2.2. Regiões de Campo Nativo (Campo Limpo, Sujo e Rupestre).	4
Figura 2.3 Distribuição das classes de uso da terra no bioma Cerrado referente ao ano de 2002. Áreas de Pastagem Cultivada ilustradas em verde.	5
Fonte: Sano <i>et al.</i> (2008)	5
Figura 3.1. Exemplo de série temporal do índice de vegetação no pixel (i,j)	7
Fonte: Adaptada de Eklund e Jönsson (2012).	7
Figura 3.2. Etapas da descoberta de conhecimento em banco de dados. A mineração de dados corresponde a uma das etapas do processo da transformação de dados em conhecimento.....	9
Figura 3.3. Exemplos de retas capazes de separar as classes (verde e laranja). Há um número infinito de retas que separam as duas classes.	11
Figura 3.4. Representação do hiperplano ótimo e dos vetores de suporte.....	12
Figura 3.5. Exemplo de mapeamento para um espaço de maior dimensão, fornecendo uma separação linear.	13
Figura 4.1 Área de estudo do trabalho, localizada no sudoeste do estado de Minas Gerais. Recorte de uma imagem <i>Landsat</i> TM-5 (R5B4G3) da região.....	15
Figura 4.2 OMT-G do trabalho proposto.....	16
Figura 4.3 Sequência de imagens para compor as séries temporais de EVI para 2006 e 2009.	19
Figura 4.4 Abordagem de pixels puros.....	20
Figura 4.5 Trecho de código elaborado em TerraLib 5 que verifica para um determinado pixel se alguns dos seus pixels vizinhos estão contidos no mesmo polígono.	21
Figura 4.6 Para um polígono, o pixel da figura à esquerda (em azul) é considerado um pixel puro, enquanto o da segunda figura não obedece à relação espacial definida no algoritmo e, portanto, não é levado em consideração.....	21
Figura 5.1 Mapa temático gerado a partir da classificação do Experimento 1 (à esquerda) e mapa temático com conjunto verdade (à direita). Os planos de informação temáticos estão sobrepostos em uma imagem <i>Landsat</i> TM-5 (R5B4G3).	24
Figura 5.2 Mapa temático gerado a partir da classificação do Experimento 2 (à esquerda) e mapa temático com conjunto verdade (à direita). Os planos de informação temáticos estão sobrepostos em uma imagem <i>Landsat</i> TM-5 (R5B4G3).	25
Figura 5.3 Mapa temático gerado a partir da classificação do Experimento 3 (à esquerda) e mapa temático com conjunto verdade (à direita). Os planos de informação temáticos estão sobrepostos em uma imagem <i>Landsat</i> TM-5 (R5B4G3).	26

LISTA DE TABELAS

	<u>Pág.</u>
Tabela 5.1 – Resumo dos conjuntos de dados utilizados em cada experimento.	23
Tabela 5.2 – Matriz de confusão do Experimento 1	24
Tabela 5.3 – Matriz de confusão do Experimento 2.....	25
Tabela 5.4 – Matriz de confusão do Experimento 3.....	25

SUMÁRIO

	<u>Pág.</u>
1. INTRODUÇÃO.....	1
2. CERRADO: CAMPOS NATIVOS E PASTAGENS CULTIVADAS.....	3
2.1.1. Campos Nativos.....	4
2.2. Pastagens Cultivadas.....	5
3. SÉRIES TEMPORAIS, MINERAÇÃO DE DADOS E SVM EM SENSORIAMENTO REMOTO.....	7
3.1. Análise de padrões em séries temporais de imagem de sensoriamento.....	7
3.2. Mineração de dados.....	8
3.3. Support Vector Machine (SVM).....	11
4. MATERIAIS E MÉTODOS.....	15
4.1. Área de estudo.....	15
4.2. Geo-objetos de referência.....	16
4.3. Imagens MODIS.....	18
4.4. Extração de atributos, treinamento, avaliação e geração de plano temático.....	19
5. RESULTADOS E DISCUSSÕES.....	23
6. CONCLUSÕES E TRABALHOS FUTUROS.....	27
REFERÊNCIAS BIBLIOGRÁFICAS.....	29

1. INTRODUÇÃO

Mais da metade da área do Cerrado brasileiro tem sido transformada, principalmente, em áreas de pastagem e agricultura, perdendo cerca de 1 milhão de km² de sua vegetação original (MACHADO *et al.*, 2004). As áreas de agricultura cobrem mais de 100.000 km² e as de pastagem cultivada superam 500.000 km², enquanto que as áreas preservadas envolvem apenas cerca de 33.000 km² (KLINK; MACHADO, 2005). Esta situação representa um alto custo ambiental, ocasionando perda de biodiversidade, erosão do solo, degradação das formações vegetais do bioma, poluição da água, mudanças nos eventos de fogo típicos do bioma, instabilidade do ciclo do carbono e prováveis modificações climáticas regionais (KLINK; MACHADO, 2005).

Segundo Walter (2006), dentre as formações vegetais existentes no bioma, encontram-se as formações campestres, que se referem a regiões com predominância de espécies herbáceas e algumas arbustivas, sem ou com pouca ocorrência de árvores na paisagem (RIBEIRO; WALTER, 2008; IBGE, 2012). Considerando que estas fitofisionomias são formações naturais, elas são passíveis de recuperação caso sejam degradadas. Neste caso, o monitoramento deste tipo de vegetação, por meio de imagens de satélite, pode auxiliar na recuperação da sua integridade física, química e biológica assim como na capacidade produtiva da região degradada (RODRIGUES; GANDOLFI, 2001). Com efeito, mapas de uso e cobertura do Cerrado, obtidos a partir imagens de satélites e técnicas de sensoriamento remoto, podem ser usados para monitorar estas áreas.

Existem alguns estudos, como o de Ferreira *et al.* (2013b), que usam técnicas de análise multitemporal baseada em índices de vegetação para classificar áreas de pastagem cultivada no Cerrado. Entretanto, as áreas transformadas em pastagem, de forma similar às formações campestres supracitadas, podem variar de formações com predominância de gramíneas a locais que apresentam dominância de espécies arbustivas, por exemplo (EMBRAPA; INPE, 2011). Desta forma, o mapeamento de áreas de Pastagem e de formações campestres nativas no Cerrado, por meio de imagens de satélites, pode se

tornar uma tarefa difícil quando usada apenas a informação espectral dos alvos (SANO *et al.*, 2008).

Dentro deste contexto, este trabalho tem como objetivo classificar e criar mapas temáticos de áreas de formações campestres nativas e regiões de pastagens cultivadas no Cerrado brasileiro por meio de técnicas de geoprocessamento e da implementação de algoritmos, utilizando métodos de análise de padrões e mineração de dados em dados de séries temporais de imagens, a partir do classificador SVM.

O trabalho foi organizado da seguinte forma: na Seção 2 é apresentada uma breve descrição do Cerrado brasileiro, com ênfase nas formações vegetais e áreas antrópicas de interesse. Na Seção 3 é apresentada uma visão geral sobre a mineração de dados e a análise de padrões em dados de sensoriamento remoto. Os dados e a metodologia descritos no trabalho estão detalhados na Seção 4. Na Seção 5 são apresentados os experimentos realizados e uma discussão sobre os progressos obtidos neste estudo. Finalmente, na Seção 6 são apresentadas as considerações finais e trabalhos futuros.

2. CERRADO: CAMPOS NATIVOS E PASTAGENS CULTIVADAS

Segundo Ferreira *et al.* (2003), o Cerrado brasileiro corresponde a uma das savanas mais úmidas do planeta, apresentando um clima sazonal, com estações seca e chuvosa bem definidas. O termo Cerrado é comumente empregado para indicar o conjunto de ecossistemas (savanas, campos e matas) que ocupa a região central do Brasil (EITEN, 1977).

De forma geral, o Cerrado pode ser entendido como uma formação composta por campos gramíneos coexistentes com árvores e arbustos esparsos. Contudo, várias definições do termo Cerrado e propostas de classificação das formações vegetais do bioma são apresentadas na literatura (WALTER, 2006). Ribeiro e Walter (2008), por exemplo, descrevem o Cerrado como uma região composta de formações florestais, savânicas e campestres. As formações florestais representam áreas com predominância de espécies arbóreas com formação de dossel. As formações savânicas designam regiões com árvores e arbustos esparsos sobre um campo de gramíneas, sem a formação de dossel contínuo. Por outro lado, as formações campestres referem-se a regiões com predominância de espécies herbáceas e algumas arbustivas, sem a ocorrência de árvores na paisagem.

Estas três formações vegetais, representadas na Figura 2.1, são classificadas em diferentes fitofisionomias, baseadas primeiramente na forma (fisionomia), que é definida pela estrutura, pelas formas de crescimento dominantes e por possíveis mudanças estacionais. Posteriormente, consideram-se aspectos do ambiente e da composição florística.

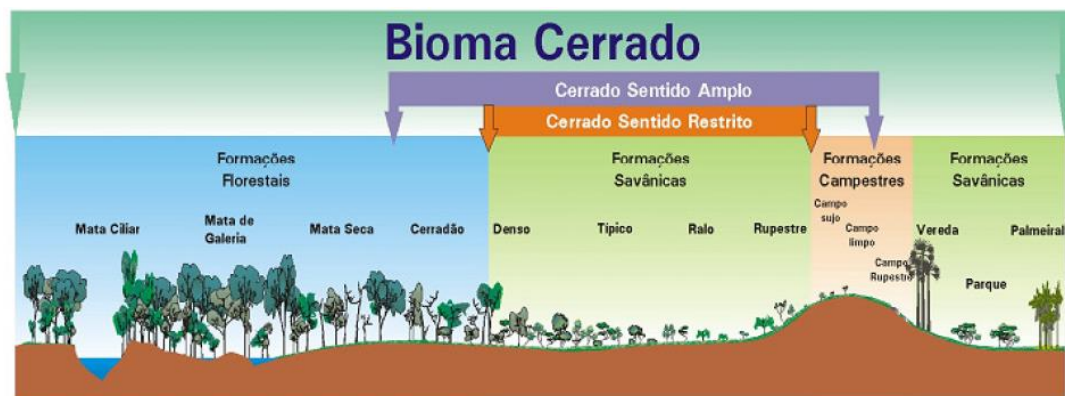


Figura 2.1. Fitofisionomias do bioma Cerrado.

Fonte: Ribeiro e Walter (2008).

2.1.1. Campos Nativos

Dentre as formações vegetais do bioma, este trabalho tem como foco as formações campestres nativas no Cerrado, subdivididas em Campo Limpo, Campo Sujo e Campo Rupestre, exibidas na Figura 2.2.



Figura 2.2. Regiões de Campo Nativo (Campo Limpo, Sujo e Rupestre).

Fonte: IBGE (2012) e COURA (2006).

O Campo Limpo é uma fitofisionomia com predominância herbácea e contém poucos arbustos, com a inexistência de árvores na região. É localizado com frequência nas chapadas, olhos d'água e encostas, com locais que podem apresentar diversas posições topográficas e variações no grau de profundidade e fertilidade do solo (RIBEIRO; WALTER, 2008).

Áreas de Campo Sujo contêm, em seu estrato arbustivo-herbáceo, espécies de arbustos e subarbustos muitas vezes compostas por plantas menos desenvolvidas das espécies arbóreas do Cerrado Sentido Restrito. Podem ser encontradas em solos rasos, com

presença eventual de pequenos afloramentos de rocha de pouca extensão ou mesmo em terrenos profundos e de pouca fertilidade (RIBEIRO; WALTER, 2008).

A fitofisionomia de Campo Rupestre agrupa paisagens em microrrelevos com espécies típicas, geralmente ocupando trechos de afloramento rochosos em altitudes superiores a 900 metros, ocasionalmente a partir de 700 metros, com a presença eventual de arvoretas raquíticas. São, em geral, regiões com solos ácidos e pobres em nutrientes, onde existem altas variações de temperatura e ventos constantes (RIBEIRO; WALTER, 2008).

2.2. Pastagens Cultivadas

De acordo com Machado *et al.* (2004), mais da metade da vegetação original do Cerrado foi transformada em áreas de pastagem, agricultura e outros usos. Além disso, estudos mostram que as mudanças do uso do solo no Cerrado ocorrem com maior intensidade do que na região Amazônica (SANO *et al.*, 2001). Das regiões transformadas no Cerrado, as áreas de Pastagem Cultivada no Cerrado correspondem a cerca de 500.000 km² (MACHADO *et al.*, 2004), conforme ilustrado na Figura 2.3.

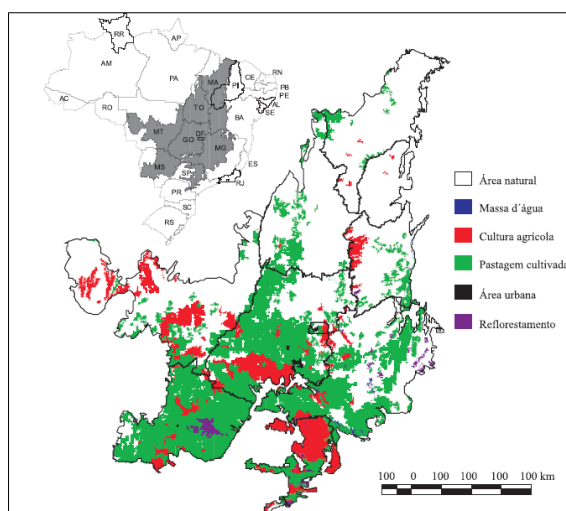


Figura 2.3 Distribuição das classes de uso da terra no bioma Cerrado referente ao ano de 2002. Áreas de Pastagem Cultivada ilustradas em verde.

Fonte: Sano *et al.* (2008)

De acordo com Ferreira *et al.* (2013b), no mínimo, 50% das áreas de pastagem plantada no Cerrado já se encontram severamente degradadas, ocasionando o aumento da erosão e a perda da fertilidade do solo, a redução da biomassa e a predominância de espécies estrangeiras, principalmente as espécies de gramíneas africanas do gênero *Brachiaria*. Com isso, a recuperação destas áreas degradadas pode auxiliar no aumento da renda para os produtores e pode-se reduzir o impacto ambiental no Cerrado por meio da diminuição da erosão, da emissão de dióxido de carbono e da abertura de novas áreas para pastagem (CHAVES *et al.*, (2001)).

A identificação das regiões de pastagem é difícil porque a degradação dos pastos pode, por exemplo, influenciar na porcentagem da cobertura vegetal e na resposta dos índices de vegetação. Confusão e incertezas podem ocorrer quando as pastagens são manejadas de forma inadequada, uma vez que podem aparecer espécies invasoras ou mesmo o renascimento de espécies de arbustos e árvores nativas nestas regiões (FERREIRA *et al.*, 2013b).

Alguns estudos procuram realizar o monitoramento das pastagens cultivadas no bioma. Ferreira *et al.* (2013b), por exemplo, utilizam dados biofísicos das estações secas e úmidas das áreas de pastagem, dados radiométricos (índices de vegetação) e climáticos. Segundo os autores, as estações seca e úmida são bem definidas na região e, portanto, podem auxiliar na identificação das características biofísicas e radiométricas das áreas de pastagem.

Em decorrência das grandes extensões de áreas, da periodicidade marcante da vegetação natural e da confusão espectral entre classes, a necessidade de identificar e discriminar áreas de pastagens cultivadas e pastagens naturais (formações campestres) é um grande desafio (SANO *et al.*, 2008). Consequentemente, para melhorar a discriminação destes alvos é necessário usar dados temporais e de campo, explorando a vantagens de o bioma apresentar estações bem definidas (FERREIRA *et al.*, 2013a).

3. SÉRIES TEMPORAIS, MINERAÇÃO DE DADOS E SVM EM SENSORIAMENTO REMOTO

3.1. Análise de padrões em séries temporais de imagem de sensoriamento

A disponibilidade de uma grande quantidade de dados de sensores remotos ópticos tem tornado cada vez mais acessível e de forma mais detalhada e continuada a observação da Terra. Por exemplo, o sensor MODIS (*Moderate Resolution Imaging Spectroradiometer*) produz dados com alta frequência temporal da superfície terrestre, dos oceanos e da atmosfera. Seus produtos foram projetados para fornecer, de forma consistente, informações espaciais e temporais das condições da vegetação global, que podem ser utilizadas para fazer o monitoramento da cobertura da vegetação (HUETE *et al.*, 2002; ZHANG *et al.*, 2003).

A combinação de informações temporais, espaciais e espectrais, quando adequadamente explorada, permite detectar padrões complexos e importantes para o monitoramento e análise da dinâmica da cobertura da terra (BRUZZONE *et al.*, 2003). Em estudos de análise e caracterização da cobertura vegetal, utilizam-se índices de vegetação para o monitoramento sazonal e interanual dos parâmetros biofísicos, fenológicos e estruturais da vegetação (HUETE *et al.*, 2002). A Figura 3.1 ilustra a criação de uma série temporal de um dado índice de vegetação ($d_{i,j}$). Para cada pixel em questão, pode-se observar uma série temporal que representa a variação da vegetação no decorrer do tempo.

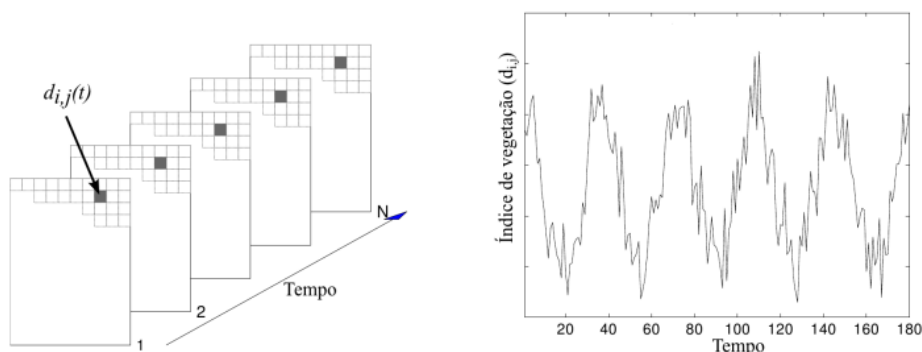


Figura 3.1. Exemplo de série temporal do índice de vegetação no pixel (i,j).

Fonte: Adaptada de Eklund e Jönsson (2012).

Os índices de vegetação produzidos através dos dados MODIS representam medidas aprimoradas das condições da superfície da vegetação e a análise de padrões destes índices pode ser uma ferramenta útil para a observação do uso e cobertura do solo (TUCKER *et al.*, 2005). Produzidos geralmente com resoluções espaciais entre 250 m e 1 km, os índices de vegetação do MODIS mais utilizados são o NDVI e EVI (JUSTICE *et al.*, 2002). O índice EVI (*Enhanced Vegetation Index*) apresenta sensibilidade em áreas com maior quantidade de biomassa verde, além de minimizar os efeitos das influências atmosféricas e do solo (JIANG *et al.*, 2008). Utilizando três comprimentos de onda, o cálculo do EVI é definido por:

$$EVI = G \frac{\rho_{NIR} - \rho_{red}}{\rho_{NIR} + C_1 \times \rho_{red} - C_2 \times \rho_{blue} + L} , \quad (3.3)$$

onde ρ_{NIR} e ρ_{red} são as reflectâncias nas bandas do infravermelho próximo e do vermelho, G equivale a um fator de ganho, L corresponde ao fator de ajuste do solo, C_1 e C_2 são os coeficientes da resistência de aerossóis, que usam a reflectância na banda do azul (ρ_{blue}) para corrigir a influência dos aerossóis na banda do vermelho. Para o sensor MODIS, os coeficientes geralmente utilizados para o cálculo do EVI são: $G = 2,5$, $L = 1$, $C_1 = 6$ e $C_2 = 7,5$ (JIANG *et al.*, 2008).

3.2. Mineração de dados

Segundo Lambin e Lindman (2006), a riqueza de dados de sensoriamento remoto disponíveis atualmente permite explorar questões sobre o sistema terrestre em escalas anteriormente ineficazes. Os autores relatam que a maioria das pesquisas em mudanças na cobertura do solo vinha sendo conduzidas pela disponibilidade de dados. Porém, com o aumento do número de sensores e, conseqüentemente, do volume de informações, as pesquisas passaram a ser guiadas pela necessidade de processamento e análise dos dados, considerando que um melhor entendimento dos conjuntos de dados é essencial para a análise das mudanças da cobertura de solo e dos impactos dos processos do sistema terrestre. Portanto, o crescente montante de dados de sensoriamento remoto

torna-se um caso apropriado para a aplicação de técnicas de mineração de dados (RUSHING et al., 2005).

A mineração de dados representa um dos passos intermediários da descoberta de conhecimento em um determinado banco de dados, ilustrada na Figura 3.2, e envolve a descoberta de modelos apropriados para um conjunto de dados e a determinação de padrões a partir destes (FAYYAD *et al.*, 1996).



Figura 3.2. Etapas da descoberta de conhecimento em banco de dados. A mineração de dados corresponde a uma das etapas do processo da transformação de dados em conhecimento.

Fonte: Adaptada de Fayyad *et al.* (1996).

A classificação faz a predição e o mapeamento dos dados de um conjunto em uma ou mais classes categóricas (FAYYAD *et al.*, 1996; MILLER; HAN, 2009). A geração de modelos de classificação de dados é um processo composto de duas fases: aprendizado e teste. Na primeira etapa, aplica-se um algoritmo de classificação sobre um conjunto de dados de treinamento e, como resultado, tem-se a construção de um modelo de classificação, o classificador propriamente dito. O conjunto de treinamento, geralmente, equivale a um conjunto de amostras escolhidas aleatoriamente a partir da base de dados que se deseja analisar. Cada instância do conjunto de treinamento apresenta dois tipos de atributos: o atributo classe, o qual indica a classe que a instância pertence; e os atributos preditivos, cujos valores são avaliados a fim de descobrir o modo como eles relacionam-se com o atributo classe.

A fase de teste inicia-se após o classificador ser construído e seu objetivo é estimar e analisar a acurácia usando um conjunto de dados de teste. As amostras deste conjunto também são selecionadas de maneira aleatória, a partir da base de dados, e não podem ser utilizadas na construção do classificador. Ou seja, as amostras de teste devem ser

diferentes das amostras que foram escolhidas para gerar o conjunto de treinamento. A acurácia de um classificador em um conjunto de teste caracteriza a porcentagem de amostras que são corretamente classificadas por ele. Se a acurácia é considerada aceitável, pode-se utilizar o classificador para classificar novas amostras em que o rótulo de classe não é conhecido (HAN; KAMBER, 2006).

Antes de realizar a classificação e predição, é importante que os dados sejam pré-processados através, por exemplo, da redução de ruídos e tratamento de valores ausentes para um determinado atributo. Segundo Lu *et al.* (2004), é importante que algumas condições devem ser satisfeitas antes de se implementar alguma técnica de detecção de mudanças. Considerando as imagens de sensoriamento, é necessário que o registro e as calibrações atmosférica e radiométrica sejam precisas, além de, sempre que possível, fazer a seleção de imagens com a mesma resolução espacial e espectral. Além disso, é fundamental que fatores que ocasionem ruídos, como falhas nos dados e cobertura de nuvens, também sejam tratados.

Tais procedimentos podem auxiliar na redução da confusão durante a fase de aprendizagem. Além disso, quando métodos que envolvem medições de distâncias são utilizados na fase de aprendizado, os dados podem ser transformados através da normalização. Geralmente, este processo envolve ajustar para um conjunto de atributos todos os valores em uma escala, de forma que estes se encontrem em um intervalo específico, como de -1,0 a 1,0, podendo evitar que alguns atributos, por apresentarem uma escala de valores maior que outros, possam influenciar o classificador de forma tendenciosa.

Algoritmos de mineração de dados são bastante utilizados em aplicações de sensoriamento remoto. Por exemplo, Noma *et al.* (2013) comparam diversos métodos de classificação, como árvores de decisão, vizinhos mais próximos, Máquina de Vetores de Suporte (SVM) e Floresta de Caminhos Ótimos (OPF, *Optimum Path Forest*) para mapear áreas de vegetação. Os autores utilizam dados temporais correspondentes aos índices EVI em uma região do Mato Grosso com áreas de agricultura, floresta e

pastagem. Neste caso, o método SVM apresentou os melhores resultados para o estudo em questão.

3.3. Support Vector Machine (SVM)

Embasado na Teoria da Aprendizagem Estatística (VAPNIK, 1995), o algoritmo de máquinas de vetores de suporte (*Support Vector Machines*, SVMs). (BOSER *et al.*, 1992; CORTES; VAPNIK, 1995) tem recebido atenção emergente da academia. O método tem sido eficiente em classificações com dados de alta dimensionalidade, muitas vezes com resultados superiores aos de outros algoritmos de aprendizado, como as redes neurais (SUNG; MUKKAMALA, 2003; DING; DUBCHAK, 2001).

Apesar de o tempo da fase de treinamento dos algoritmos de SVM seja considerado moderado, a precisão do método é alta devido a sua capacidade de modelar limites de decisão não lineares complexos, apresentando uma boa generalização. Como é um classificador não-paramétrico, o SVM tem como característica a independência de modelos de distribuição estatística (MELGANI; BRUZZONE, 2004).

O algoritmo SVM utiliza, de forma geral, um mapeamento não linear para transformar os dados do conjunto de treinamento original em uma nova dimensão. Dentro desta nova dimensão, o algoritmo faz uma busca pelo limite de decisão linear ótimo para separar duas classes: o hiperplano cuja margem de separação seja máxima (HAN; KAMBER, 2006). O caso mais simples que o algoritmo trata é o de duas classes, em que os dados são linearmente separáveis, como ilustrado na Figura 3.3.

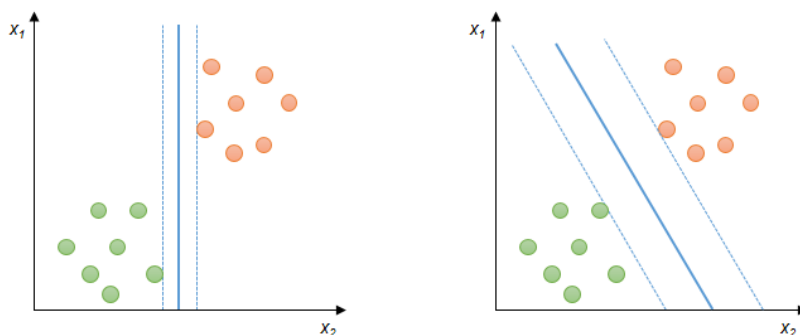


Figura 3.3. Exemplos de retas capazes de separar as classes (verde e laranja). Há um número infinito de retas que separam as duas classes.

Como existem infinitas retas que podem separar essas classes, o problema, então, reside em encontrar aquela que mais se adequa no processo de classificação. Generalizando, procura-se encontrar a superfície de decisão entre as classes baseada no comportamento geométrico das amostras no espaço de atributos, ou seja, busca-se o hiperplano que minimize o erro de classificação. Para isso, o SVM pesquisa o hiperplano de máxima margem de separação, onde a margem corresponde à distância mais curta do hiperplano para as amostras de treinamento mais próximas de cada classe (HAN; KAMBER, 2006). As amostras localizadas nos limites da margem de separação são as responsáveis por determinar o hiperplano, e são chamadas de vetores de suporte. A Figura 3.4 ilustra a escolha de um hiperplano de seleção ótimo, com os vetores de suportes contornados com uma borda mais escura.

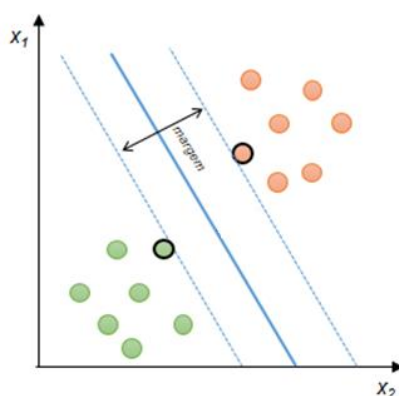


Figura 3.4. Representação do hiperplano ótimo e dos vetores de suporte.

Em suma, a busca do hiperplano de separação ótima é modelada como um problema de otimização e pode ser solucionada com a introdução de uma função langrangiana (THEODORIDIS; KOUTROMBAS, 2008). Uma vez que os vetores de suportes e o hiperplano de margem máxima de separação são encontrados, constrói-se o classificador: a máquina de vetores de suporte, que pode ser utilizada para classificar dados linearmente separáveis.

No entanto, aplicações em situações reais cujos dados sejam linearmente separáveis são bastante difíceis de ser encontradas. Para contornar este problema, o algoritmo SVM linear pode ser estendido para classificação de dados não linearmente separáveis, sendo

capaz de descobrir limites de decisão não lineares no conjunto de amostras de treinamento. Para isso, faz-se a transformação dos dados de treinamento originais para um novo espaço dimensional através do mapeamento não linear. Com os dados transformados, o algoritmo procura por um hiperplano linear no novo espaço de atributos (HAN; KAMBER, 2006). A Figura 3.5 ilustra o processo de transformação dos dados de entrada em uma nova dimensão.

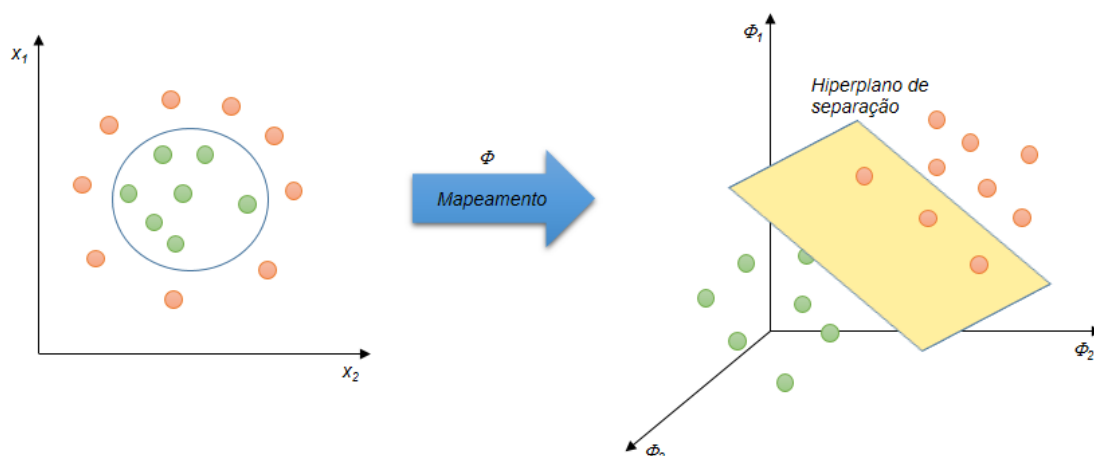


Figura 3.5. Exemplo de mapeamento para um espaço de maior dimensão, fornecendo uma separação linear.

A determinação dos parâmetros do hiperplano de separação na nova dimensão corresponde outra vez a um problema de otimização quadrática langrangiana. No entanto, quanto maior a dimensionalidade, maior será o custo computacional envolvido, fazendo-se necessário calcular o produto interno dos vetores em cada dimensão. Para resolver esta situação, pode-se substituir o produto interno presente no problema de otimização por uma função *Kernel* K , de forma que todos os cálculos são realizados nas amostras de treinamento iniciais, reduzindo potencialmente o custo computacional (THEODORIDIS; KOUTROMBAS, 2008).

Alguns dos *kernels* mais empregados em aplicações gerais são as funções de base radial (*Radial Basis Functions*, RBF) e as polinomiais, descritos a seguir.

- Linear: $K(x_i, x_j) = \langle x_i, x_j \rangle$.

- Polinomial: $K(x_i, x_j) = (\langle x_i, x_j \rangle + 1)^d$.

- RBF: $K(x_i, x_j) = e^{-\frac{\|x_i - x_j\|^2}{2\gamma^2}}$.

Vale ressaltar que $d \in \mathbb{N}^*$ e $\gamma \in \mathbb{R}_+^*$ são parâmetros que precisam de ajustes no processo de classificação. Com a seleção do *kernel*, pode-se em seguida realizar o processo de decisão de forma similar ao processo de predição de dados linearmente separáveis. Neste caso, porém, há a introdução do parâmetro C , que deve ser ajustado e é responsável pela regularização associada à quantidade de amostras classificadas erroneamente. Geralmente, tanto o parâmetro C quanto o parâmetro γ do *kernel* RBF são obtidos empiricamente, através de pesquisa exaustiva por meio de validação cruzada.

Como vantagem, a complexidade do classificador é definida pelo número de vetores de suporte em vez da dimensionalidade dos dados. Além disso, se outras amostras de treinamento forem removidas e a fase de treinamento for repetida, e estas não sejam vetores de suporte, o mesmo hiperplano de separação será encontrado. Desta forma, o número de vetores de suporte pode ser usado para calcular o limite superior da taxa de erro esperado no classificador, independentemente da dimensão dos dados (HAN; KAMBER, 2006).

Embora o SVM seja, originalmente, esquematizado para uma classificação binária, é possível compor várias SVMs para tratar problemas com mais de duas classes. As principais abordagens usam como base a decomposição de um problema multiclasse em subproblemas binários e, em seguida, faz-se a reconstrução do problema por meio da combinação dos resultados destes subproblemas.

4. MATERIAIS E MÉTODOS

4.1. Área de estudo

A área de estudo faz parte do complexo do Cerrado, e apresenta regiões naturais e antrópicas, com áreas de pastagem cultivada. Nesta região, ilustrada na Figura 4.1, está presente o Parque Nacional da Serra da Canastra e regiões vizinhas, onde são encontradas todas as fitofisionomias do grupo das formações campestres nativas (IBGE, 2012). Esta região foi escolhida por apresentar regiões de Campo Nativo e Pastagem Cultivada.

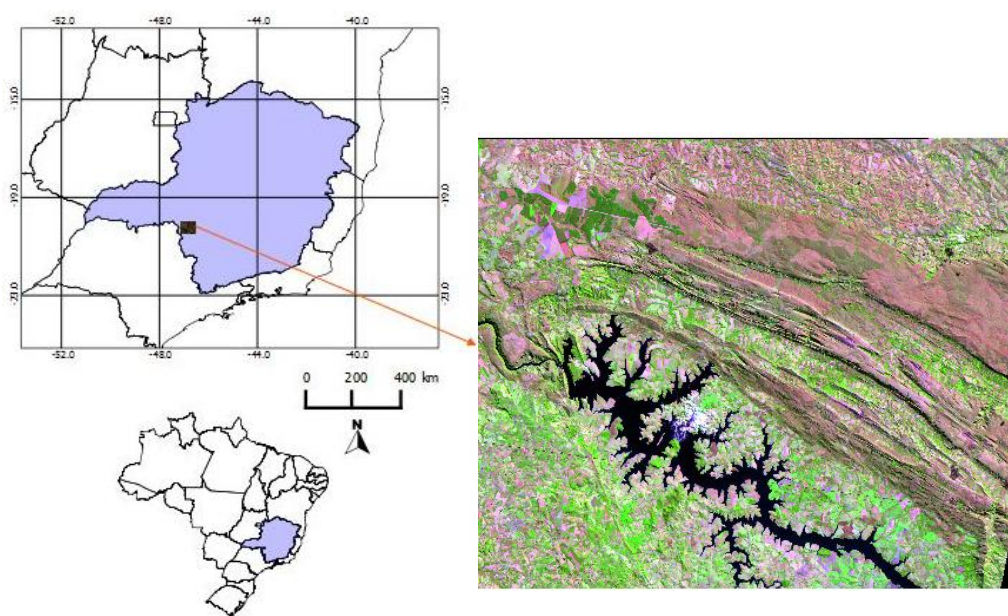


Figura 4.1 Área de estudo do trabalho, localizada no sudoeste do estado de Minas Gerais. Recorte de uma imagem *Landsat* TM-5 (R5B4G3) da região.

Os dados e as técnicas que foram utilizadas para a geração de mapas temáticos como plano de informação resultante da classificação proposta neste trabalho estão esquematizados no modelo OMG-T exposto a seguir (Figura 4.2). O modelo apresenta uma descrição dos tipos de dados de entrada e as etapas de processamento dos dados que foram usados para o desenvolvimento desta pesquisa.

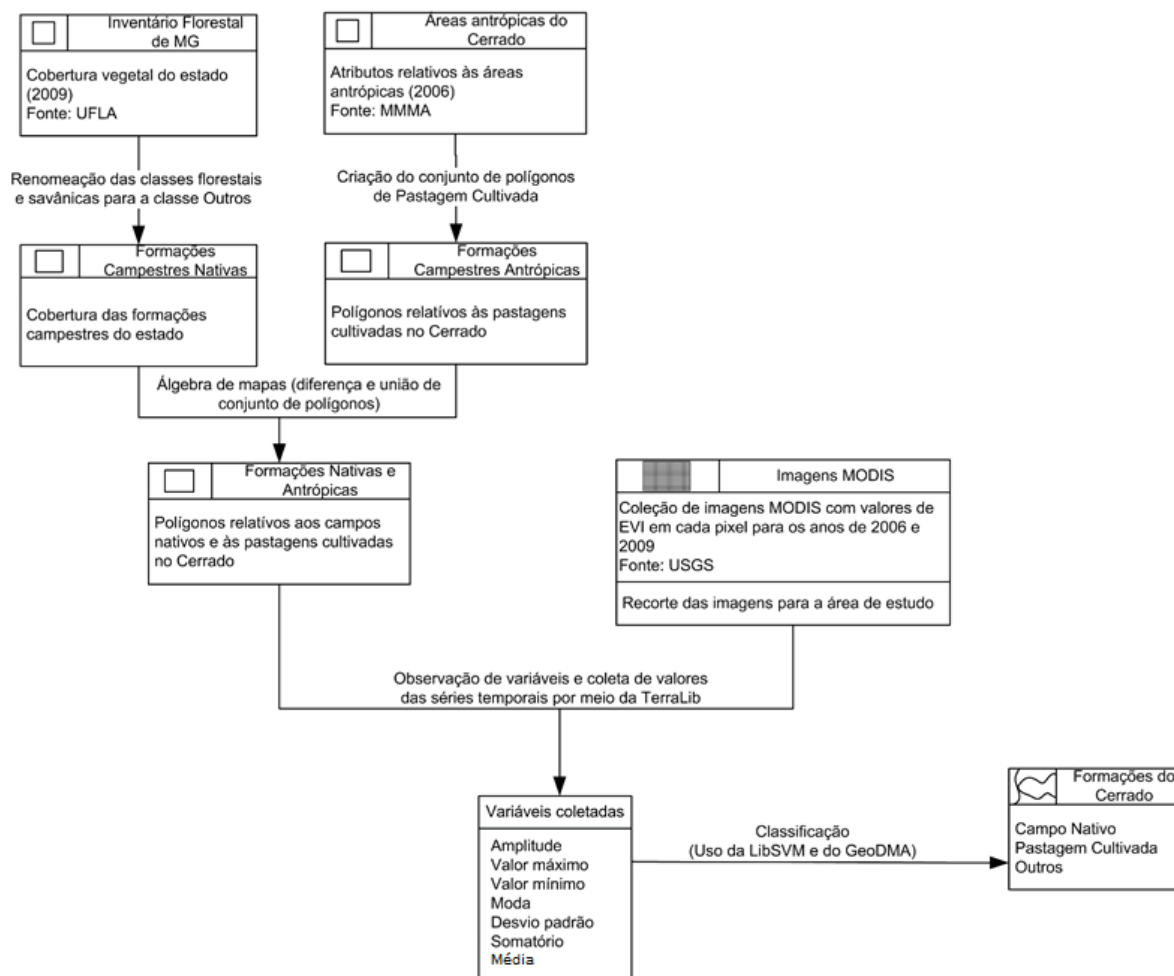


Figura 4.2 OMT-G do trabalho proposto.

4.2. Geo-objetos de referência

O mapa de cobertura vegetal do Cerrado fornecido pelo Inventário Florestal de Minas Gerais do ano de 2009 foi utilizado para a obtenção das formações campestres nativas (Campo Limpo, Sujo e Rupestre). O mapa é o resultado do esforço conjunto de diversas instituições, dentre as quais a Universidade Federal de Lavras apresenta papel de destaque na produção dos mapeamentos obtidos (SCOLFORO *et al.*, 2008). O *shapefile* contém polígonos para todo o estado de Minas Gerais categorizados com as seguintes classes: Campo, Campo Cerrado, Campo Rupestre, Cerrado, Cerradão, Eucalipto, Floresta Estacional Decidual Montana, Floresta Estacional Decidual Sub Montana,

Floresta Estacional Semidecidual Montana, Floresta Estacional Semidecidual Sub Montana, Floresta Ombrófila Alto Montana, Floresta Ombrófila Montana, Floresta Ombrófila Sub Montana, Pinus, Urbanização, Vereda, Água.

É importante ressaltar que o mapa de cobertura vegetal citado contém duas classes que representam as formações campestres nativas do Cerrado: a classe Campo, que combina as fitofisionomias de Campo Limpo e Campo Sujo; e a classe Campo Rupestre, definida pela fitofisionomia homônima.

Em relação às pastagens cultivadas do Cerrado, o Ministério do Meio Ambiente (MMA) oferece um *shapefile*, relativo ao ano de 2006, que distingue a cobertura antrópica da natural no bioma e faz a discriminação das áreas antrópicas do Cerrado, como mostra a Figura 5.4. O mapeamento subdivide as áreas antrópicas em quatro classes: Agricultura, Pastagem Cultivada, Florestamento/reflorestamento e Área Degradada por Mineração. Dessa forma, o mapa do MMA foi também usado como referência para obtenção de áreas de pastagem cultivada.

Como um *shapefile* cobre todo o estado de Minas Gerais e outro cobre todo o bioma, foi realizada uma segmentação de regiões para a área de estudo com a ferramenta GeoDMA e, em seguida, a interseção dos dois *shapefiles* com o resultado desta segmentação para que estes apenas contenham polígonos dentro da área de estudo. O sistema GeoDMA (Geographic Data Mining Analyst) (KORTING *et al.*, 2013) é uma plataforma de código aberto que possui implementadas técnicas de segmentação de imagens, extração e seleção de atributos, métodos de análise multitemporal para detecção de mudanças e mineração de dados espaciais.

Ambos os mapas de referência resultantes foram gerados utilizando as ferramentas de álgebra de mapas disponíveis no *software* TerraView (INPE, 2013). Em seguida, todos os polígonos do mapa de cobertura do Inventário de Minas Gerais correspondentes às classes de formações vegetais, exceto Campo e Campo Rupestre, foram reclassificados para a classe Outros. Além disso, as classes Campo e Campo Rupestre foram reclassificadas para a classe Campo Nativo.

Já no mapa do MMA apenas os polígonos atribuídos à classe Pastagem Cultivada foram considerados, criando um *shapefile* com apenas os polígonos categorizados para esta classe. Em seguida, foi realizada a operação de diferença entre os dois conjuntos de polígonos, com o intuito de eliminar polígonos com sobreposição e, conseqüentemente, regiões que apresentassem conflitos de classe.

Com isso, criaram-se dois arquivos *shapefile* sem polígonos sobrepostos: um com polígonos associados a duas classes: Campo Nativo e Outros para o ano de 2009; e outro com polígonos categorizados como Pastagem Cultivada para o ano de 2006. Pode-se ressaltar que estes dois *shapefiles* podem ser unidos, resultando num arquivo com todos os polígonos de interesse. Estes polígonos são usados nas etapas de extração de atributos e treinamento, e também como referência para avaliar os resultados da classificação.

4.3. Imagens MODIS

A presença de duas estações bem definidas no Cerrado pode auxiliar na detecção de mudanças nas coberturas vegetais a partir da observação de séries temporais. Para analisar esta situação, foram utilizadas as imagens do sensor MODIS, com correção atmosférica e resolução espacial de 250 m, para os períodos seco e úmido da região e os dados temporais dos índices de vegetação EVI. Especificamente, foram utilizados perfis anuais como vetores de características para 2006 e 2009. Cada perfil anual foi extraído da composição EVI de 16 dias do MODIS, resultando em 23 instantes para cada perfil. A Figura 5.7 ilustra um resumo da sequência de imagens obtidas dos dados MODIS para os 23 instantes dos anos de 2006 e 2009. Como descrito na seção anterior, para cada imagem, cada pixel corresponde a um valor de índice de vegetação (EVI), de modo que os valores ordenados desta sequência de imagens resulta em uma série temporal.

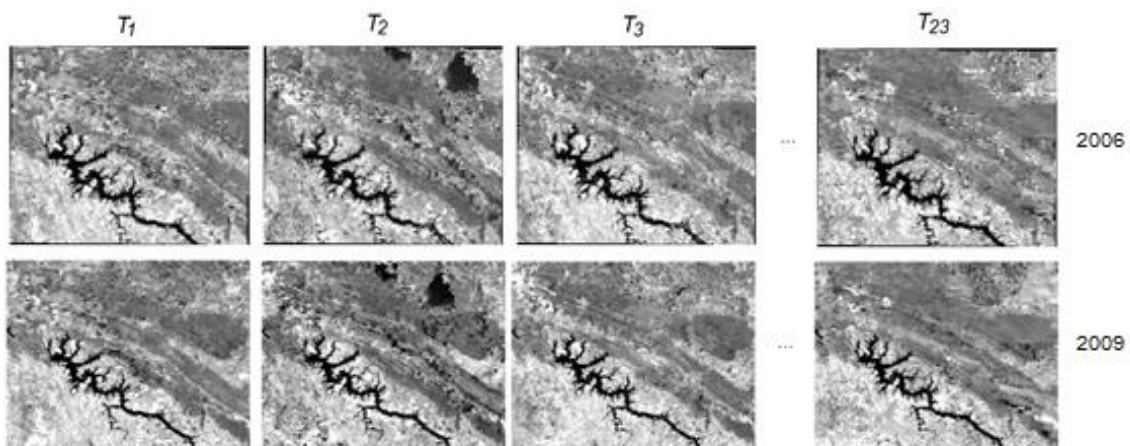


Figura 4.3 Sequência de imagens para compor as séries temporais de EVI para 2006 e 2009.

Vale ressaltar que foi necessário realizar o georreferenciamento das imagens usando como referência a base de dados da NASA (*Global Land Cover Facility - GLCF*). Além disso, foi realizado o recorte das imagens para a área de estudo por meio do ambiente *ENVI*.

4.4. Extração de atributos, treinamento, avaliação e geração de plano temático

Para a extração dos atributos preditivos, cada amostra foi considerada como o conjunto de informações relativas a um polígono do conjunto de regiões dos *shapefiles*, criados a partir dos mapas de referência. Os atributos foram pré-processados, conforme descrito na seção anterior, e foram transformados por meio da normalização.

Em relação às imagens MODIS, para cada uma das 23 imagens dos índices EVI, os atributos relacionados com os índices de vegetação foram extraídos a partir da criação de algoritmos usando a biblioteca TerraLib 5. Foi necessário implementar um algoritmo para associar a média dos valores de todos os pixels (no caso, os índices de vegetação) para cada um dos polígonos. Os atributos dos índices de vegetação correspondem, assim, à média dos valores dos pixels dentro de cada polígono, para cada imagem EVI no instante T_i ($i = 1, 2, \dots, 23$).

Para evitar o problema de pixels de borda, foi utilizada uma abordagem para definir os atributos de índices de vegetação para cada polígono. Usando a biblioteca TerraLib 5, foi implementado um algoritmo que define o valor do atributo EVI para cada polígono em cada uma das imagens T_1 a T_{23} , equivalente à média dos “pixels puros” do polígono em questão, conforme ilustrado na Figura 4.4.

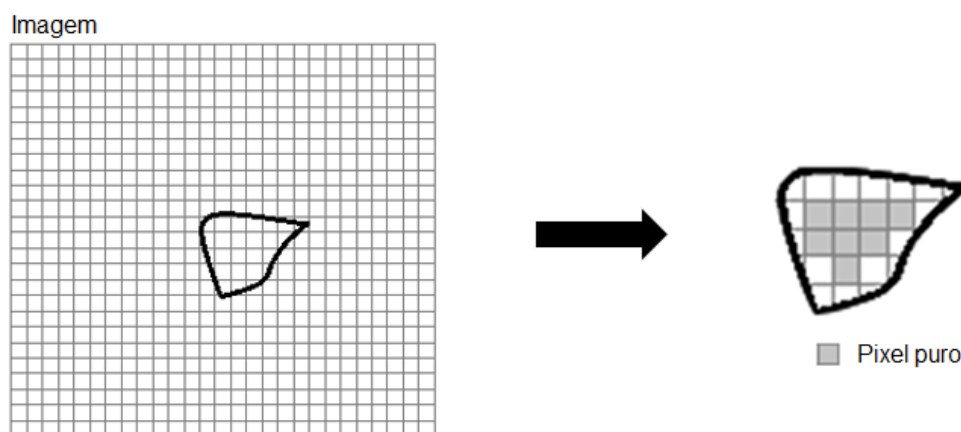


Figura 4.4 Abordagem de pixels puros.

Foi considerado, no cálculo do valor EVI para um polígono em cada uma das 23 imagens, apenas os “pixels puros”, ou seja, os pixels cujos seus vizinhos também pertencem à mesma classe, evitando, assim, que pixels de borda influenciem no cálculo da média dos valores de EVI de um dado polígono. A partir desta estratégia, consequentemente, polígonos muito pequenos, que estão inclusos em único pixel são desconsiderados do conjunto de amostras. A Figura 4.5 ilustra o trecho de código onde é definido se um determinado pixel em um polígono pode ser definido como ‘puro’, a partir da observação de alguns dos pixels vizinhos. Por meio das operações de álgebra de mapas disponíveis na biblioteca TerraLib 5, é analisado se os pixels vizinhos também estão contidos no polígono em questão. A Figura 4.6 exibe dois exemplos de pixels no processo de decisão do algoritmo. O primeiro pixel seria considerado no cálculo dos atributos, enquanto o segundo não entraria no cálculo dos índices EVI.

```

purePixel = true;

inputRasterGrid->gridToGeo(column, row - 1, x, y);
coordPixelUp = new te::gm::Point(x, y, sridInputRaster, 0);
purePixel = purePixel && polygon->contains(coordPixelUp);

inputRasterGrid->gridToGeo(column, row + 1, x, y);
coordPixelDown = new te::gm::Point(x, y, sridInputRaster, 0);
purePixel = purePixel && polygon->contains(coordPixelDown);

inputRasterGrid->gridToGeo(column - 1, row, x, y);
coordPixelLeft = new te::gm::Point(x, y, sridInputRaster, 0);
purePixel = purePixel && polygon->contains(coordPixelLeft);

inputRasterGrid->gridToGeo(column + 1, row, x, y);
coordPixelRight = new te::gm::Point(x, y, sridInputRaster, 0);
purePixel = purePixel && polygon->contains(coordPixelRight);

```

Figura 4.5 Trecho de código elaborado em TerraLib 5 que verifica para um determinado pixel se alguns dos seus pixels vizinhos estão contidos no mesmo polígono.

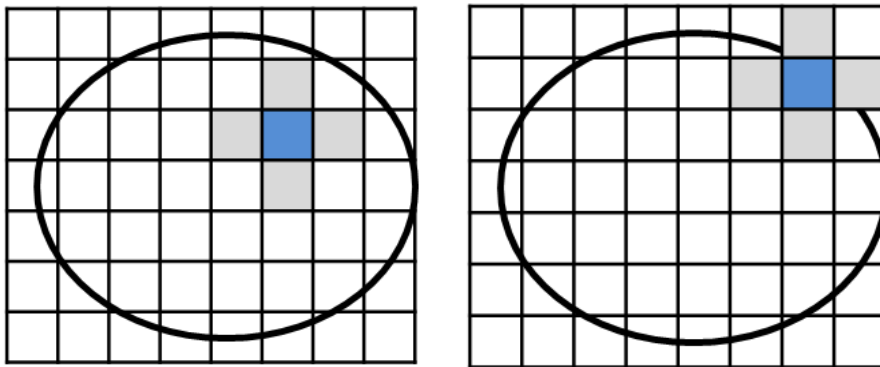


Figura 4.6 Para um polígono, o pixel da figura à esquerda (em azul) é considerado um pixel puro, enquanto o da segunda figura não obedece à relação espacial definida no algoritmo e, portanto, não é levado em consideração.

Com a geração da série temporal para cada um dos polígonos, o algoritmo elaborado neste trabalho também faz o cálculo e armazenamento das variáveis derivadas da série temporal extraída, sendo estas variáveis também utilizadas como atributos nos experimentos realizados. As variáveis extraídas a partir da série temporal são os valores de: mínimo, máximo, amplitude, soma, moda, média e desvio padrão.

Foi utilizada a estratégia de divisão percentual do conjunto de total de amostras, sendo 2/3 das amostras para o conjunto de treinamento e 1/3 para o conjunto de teste. Para o

método SVM, utilizou-se a biblioteca LibSVM (CHANG; LIN, 2011), que disponibiliza ferramentas para criação de algoritmos que particionem um conjunto de dados em treinamento e teste, treinem o classificador e realizem a predição do conjunto de teste, além de estimar os parâmetros C e γ do modelo e *kernel* do classificador.

No processo de avaliação, as matrizes de confusão e os índices Kappa (COHEN, 1960) foram usados. A partir da matriz de confusão, podem-se calcular medidas estatísticas descritivas, como a acurácia dos resultados, para que a confusão entre classes possa ser claramente apresentada e entendida (CONGALTON; GREEN, 2008). No caso da classificação SVM, os cálculos da matriz de confusão e do índice Kappa global foram implementados a partir de algoritmos em linguagem R e C++.

É importante salientar que a presença inerente de ruídos nas imagens, tais como falhas nos dados temporais e cobertura por nuvens, podem restringir a utilização dos indicadores de fenologia vegetal para a classificação da cobertura do solo, podendo contribuir na redução da acurácia dos resultados (LAMBIN; LINDERMAN, 2006).

Após utilizar e avaliar o classificador SVM, são construídos como saída do processo de classificação planos de informação referentes aos mapas temáticos para cada um dos experimentos classificados, sendo possível gerar uma camada de informação a partir da integração e processamento dos dados de entrada.

5. RESULTADOS E DISCUSSÕES

Para a geração e análise dos resultados, três experimentos foram realizados. Cada experimento usa um conjunto de atributos específico no processo de classificação, como mostra a Tabela 5.1.

Tabela 5.1 – Resumo dos conjuntos de dados utilizados em cada experimento.

Experimentos	Atributos preditivos utilizados
1	23 atributos (Série temporal)
2	Atributos extraídos a partir da série temporal
3	Série temporal e atributos extraídos a partir da série temporal

As classes de interesse utilizadas neste estudo foram: Campo Nativo, Pastagem Cultivada e Outros. O método de classificação SVM foi utilizado nos três experimentos. Os parâmetros do método SVM foram ajustados para cada teste, usando a técnica de busca exaustiva com validação cruzada. O índice Kappa foi calculado para todos os resultados de classificação com o intuito de mensurar a concordância entre os resultados dos classificadores e o conjunto de referência.

A partir dos algoritmos implementados em TerraLib 5 utilizando a abordagem de pixels “puros” descrita na seção anterior, foi extraído um conjunto com um total de 508 amostras para realizar os três experimentos, onde cada amostra é relativa a atributos de um polígono. Como a estratégia de divisão percentual de 2/3 das amostras para o conjunto de treinamento e 1/3 para o conjunto de teste foi utilizada, o conjunto de treinamento continha 337 amostras e o de teste, 171 amostras.

No processo de classificação do Experimento 1, avaliou-se o potencial das séries temporais das imagens EVI como único conjunto de atributos na classificação, a fim de analisar a contribuição das séries temporais em si. A classificação resultou em uma acurácia de 84,21% e índice Kappa de 0,74. A matriz de confusão do Experimento 1 é ilustrada na Tabela 5.2.

Tabela 5.2 – Matriz de confusão do Experimento 1

<i>Class./Ref.</i>	Campo Nativo	Outros	Pastagem cultivada
Campo Nativo	47	3	5
Outros	8	18	3
Pastagem Cultivada	3	5	79

A partir da classificação realizada pelo SVM, pôde ser criado o plano de informação correspondente ao mapa temático equivalente aos resultados do Experimento 1. A Figura 5.1 ilustra o mapa temático do Experimento 1 e o mapa temático relativo ao conjunto verdade, respectivamente.

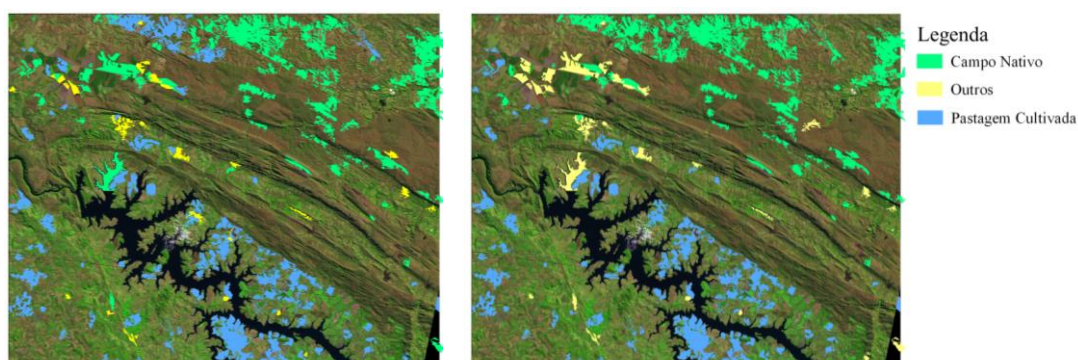


Figura 5.1 Mapa temático gerado a partir da classificação do Experimento 1 (à esquerda) e mapa temático com conjunto verdade (à direita). Os planos de informação temáticos estão sobrepostos em uma imagem *Landsat TM-5 (R5B4G3)*.

No Experimento 2, foi avaliada a contribuição das variáveis extraídas a partir de cada uma das séries temporais. Os atributos preditivos utilizados foram os valores de mínimo, máximo, amplitude, soma, moda, média e desvio padrão, todos também extraídos e implementados em um algoritmo em TerraLib 5. Os resultados utilizando apenas estas informações não implicaram em uma boa classificação, se comparada com o Experimento 1. O valor de acurácia foi de apenas 76,60%, com um índice Kappa de 0,61. A matriz de confusão para o Experimento 2 pode ser observada na Tabela 5.3, em que se pode observar que várias amostras de Campo Nativo foram classificadas como Pastagem Cultivada.

Tabela 5.3 – Matriz de confusão do Experimento 2

<i>Class./Ref.</i>	Campo Nativo	Outros	Pastagem cultivada
Campo Nativo	39	2	14
Outros	9	16	4
Pastagem Cultivada	4	7	76

Em seguida, o plano de informação referente ao mapa temático do Experimento 2 foi gerado, como ilustrado na Figura 5.2.

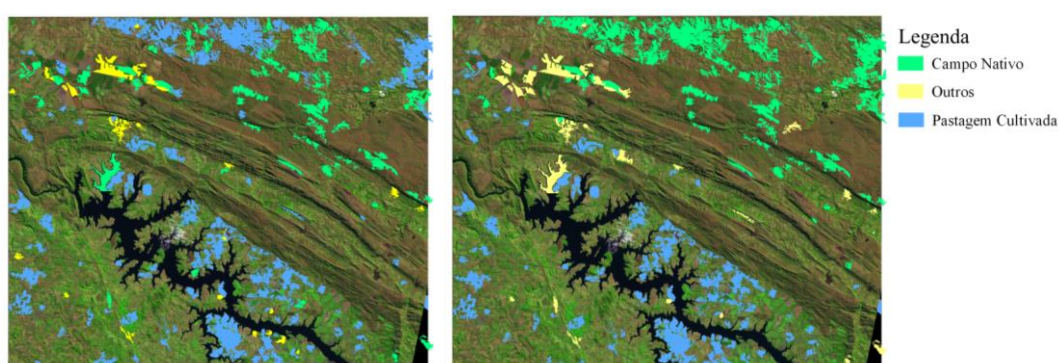


Figura 5.2 Mapa temático gerado a partir da classificação do Experimento 2 (à esquerda) e mapa temático com conjunto verdade (à direita). Os planos de informação temáticos estão sobrepostos em uma imagem *Landsat* TM-5 (R5B4G3).

Por fim, no Experimento 3 foi analisado se a combinação dos atributos do Experimento 1 e do Experimento 2 alcançaria melhores resultados. A partir da avaliação dos resultados, pode-se perceber um pequeno aumento na acurácia para o SVM em relação aos outros testes aplicados. O Experimento 3 alcançou o melhor resultado das classificações realizadas e atingiu acurácia de 85,96% e índice Kappa de 0,77, apenas com um aumento na acurácia de 1,75% em relação ao Experimento 1, sem trazer um acréscimo mais significativo. A Tabela 5.4 ilustra a matriz de confusão para o Experimento 3.

Tabela 5.4 – Matriz de confusão do Experimento 3

<i>Class./Ref.</i>	Campo Nativo	Outros	Pastagem cultivada
Campo Nativo	48	2	5
Outros	9	17	3
Pastagem Cultivada	1	4	82

O mapa temático gerado a partir da classificação para o Experimento 3 pode ser observado na Figura 5.3.

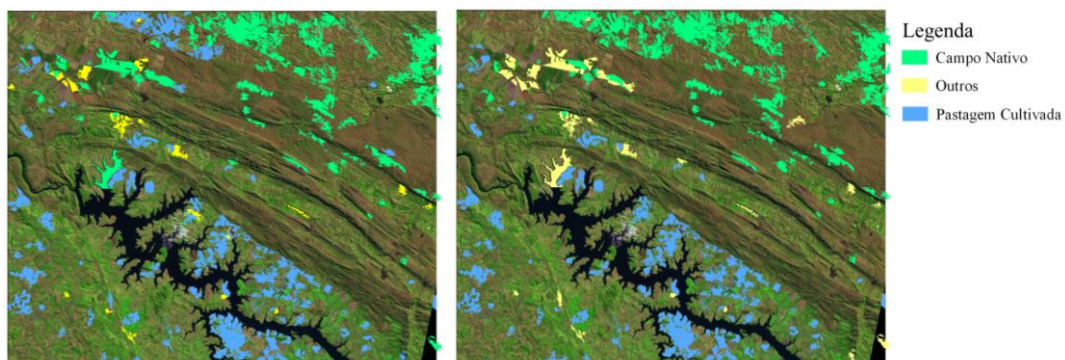


Figura 5.3 Mapa temático gerado a partir da classificação do Experimento 3 (à esquerda) e mapa temático com conjunto verdade (à direita). Os planos de informação temáticos estão sobrepostos em uma imagem *Landsat TM-5 (R5B4G3)*.

Em relação ao custo computacional, o algoritmo implementado em TerraLib 5 para extração dos atributos, observando apenas os pixels “puros” do polígono, apresentou um tempo de cerca de 5 horas para realizar a extração do conjunto de atributos. Ao observar a classificação por SVM, quando utilizado sem o ajuste de parâmetros, o classificador realiza as predições em questão de milissegundos, porém os valores de acurácia obtidos para as classificações não ultrapassaram mais que 60% sem tal ajuste. Para ocorrer uma melhora no processo de classificação, é necessário que seja feita a estimativa dos parâmetros C e γ mais adequados, a qual foi a principal responsável pelo aumento no custo de processamento, acrescentando um tempo médio de 15 segundos para cada um dos experimentos.

Do ponto de vista operacional, a criação de um modelo de classificação por SVM com os atributos do Experimento 3 apresentaram resultados similares ao Experimento 1, já que a melhora na acurácia foi pequena para o mapeamento das três classes. Além disso, pode-se considerar que, a partir da construção de mapas temáticos como saída do processo de classificação, o trabalho pode contribuir na integração de dados de sensoriamento remoto em sistemas de informação geográfica, uma vez que foi possível gerar uma camada de informação a partir dos dados de entrada.

6. CONCLUSÕES E TRABALHOS FUTUROS

O trabalho teve o intuito de classificar e criar mapas temáticos de áreas de formações campestres nativas e regiões de pastagens cultivadas no Cerrado brasileiro a partir da implementação de algoritmos e de técnicas de geoprocessamento. Para isso, uma região do Parque da Serra da Canastra, situada em Minas Gerais, foi utilizada como área de estudo, uma vez que podem ser encontradas áreas correspondentes às classes de interesse nestas localizações.

O processo de análise de dados de sensoriamento remoto foi realizado com auxílio dos ambientes computacionais GeoDMA e da biblioteca LibSVM. Algoritmos foram implementados na biblioteca TerraLib 5, em linguagem R e C++ para a classificação das amostras, avaliação e validação dos resultados, além da possibilidade de considerar apenas os pixels “puros” para o cálculo das amostras. Além da implementação dos algoritmos, várias operações de geoprocessamento foram utilizadas para que os planos de informações temáticos fossem gerados no final do processo de classificação, tais como álgebras de mapas, reclassificação de geo-objetos, recorte e registro de imagens.

Três experimentos foram conduzidos, combinando atributos preditivos relativos e derivados das séries temporais de índices de vegetação para os anos de 2006 e 2009 da área de estudo. Índices de vegetação EVI provenientes do sensor MODIS foram utilizados para a extração dos atributos preditivos, o treinamento e a avaliação do conjunto de amostras.

No tocante ao tempo de processamento, a extração das amostras teve um custo de cerca de cinco horas a partir do algoritmo implementado. O método SVM apresentou um custo computacional mínimo para classificação dos resultados, sendo que o fator predominante para este custo foi a estimação dos parâmetros mais adequados.

O algoritmo SVM apresentou bons resultados na separação entre campo nativo e pastagens cultivadas para as amostras em questão. O melhor resultado obtido teve

acurácia de 85,96% e índice Kappa de 0,77, combinando todos os atributos referentes à série temporal e as variáveis que foram extraídas a partir desta.

É importante ressaltar que o plugin GeoDMA está em processo de transição para a linguagem TerraLib 5 (foi originalmente criado em TerraLib 4). A construção do algoritmo SVM no *plugin* já está em andamento, faltando ajustar alguns fatores relativos à compatibilidade e padronização dos parâmetros com a sintaxe da TerraLib 5 e com os algoritmos de classificação já existentes na ferramenta. O novo sistema ainda está em fase preliminar e, como trabalhos futuros, pretende-se concluir a construção do classificador SVM e também da interface para comunicação com o usuário.

REFERÊNCIAS BIBLIOGRÁFICAS

- BOSER, B. E.; GUYON, I. M.; VAPNIK, V. N. A training algorithm for optimal margin classifiers. In: ANNUAL WORKSHOP ON COMPUTATIONAL LEARNING THEORY, 5., 1992, Pittsburgh, Pennsylvania, USA. **Proceedings...** Pennsylvania: ACM, 1992. p. 144-152.
- BRUZZONE, L.; SMITS, P.; TILTON, J. Foreword special issue on analysis of multitemporal remote sensing images. **IEEE Transactions on Geoscience and Remote Sensing**, v. 41, n. 11, p. 2419-2422, nov. 2003.
- CÂMARA, G.; VINHAS, L.; FERREIRA, K.; QUEIROZ, G.; SOUZA, R.; MONTEIRO, A.; CARVALHO, M.; CASANOVA, M.; FREITAS, U. TerraLib: An open source GIS library for large-scale environmental and socio-economic applications. **Open Source Approaches in Spatial Data Handling**, Berlin, v. 2, p. 247-270, 2008.
- CHANG, C.-C.; LIN, C.-J. Libsvm: a library for support vector machines. **ACM Transaction on Intelligent Systems and Technology**, v. 2, n. 3, p. 1-27, 2011.
- CHAVES, J. M.; MOREIRA, L.; SANO, E. E.; BEZERRA, H. S.; FEITOZA, L. Uso da técnica de segmentação na identificação dos principais tipos de pastagens cultivadas do Cerrado. In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 10. (SBSR), 2001, Foz do Iguaçu. **Anais...** São José dos Campos: INPE, 2001. p. 31-33. CD-ROM, On-line. ISBN 85-17-00016-1. Disponível em:<<http://urlib.net/dpi.inpe.br/lise/2001/09.12.16.36>>. Acesso em: 09 jun. 2014.
- COHEN, J. A. Coeficiente of agreement for nominal scales. **Educational and Psychological Measurement**, n. 20, p. 37-46, 1960.
- CONGALTON, R. G.; GREEN, R. **Assessing the accuracy of remotely sensed data: principles and practices**. 2. ed. Boca Raton: CRC Press, p. 1-183, 2008.
- CORTES, C.; VAPNIK, V. Support-Vector Networks. **Maching Learning**, v. 20, n. 3, p. 273-297, set. 1995.
- COURA, S. M. D. C. **Mapeamento de vegetação do estado de Minas Gerais utilizando dados MODIS**. 2006. 150 p. (INPE-14657-TDI/1213). Dissertação de Mestrado - Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2006. Disponível em:<<http://urlib.net/sid.inpe.br/MTC-m13@80/2006/12.21.13.36>>. Acesso em: 09 jun. 2014.
- DING, C. H. Q.; DUBCHAK, I. Multi-class protein fold recognition using support vector machines and neural networks. **Bioinformatics**, v. 17, n. 4, p. 349-358, 2001.

EITEN, G. **Delimitação do conceito de cerrado**. Rio de Janeiro: Arquivos do Jardim Botânico, v. 21, p. 125-134, 1977.

EKLUNDH, L.; JÖNSSON, P. **TIMESAT 3.1 software manual**. Lund, Sweden: Malmö University, 2012. 182 p.

EMPRESA BRASILEIRA DE PESQUISAS AGROPECUÁRIAS (EMBRAPA)/ INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS (INPE). **Levantamento de informações de uso e cobertura da terra na Amazônia** - sumário executivo. Brasília/São José dos Campos:TerraClass. 2011.

FAYYAD, U.; SHAPIRO, G.; SMYTH, P. The KDD process for extracting useful knowledge from volumes of data. **Communications of the ACM**, New York, NY, v. 39, n. 11, p. 27-34, 1996.

FERREIRA, L. G.; FERNANDEZ, L. E.; SANO, E. E.; FIELD, C.; SOUSA, S. B.; ARANTES, A. E.; ARAÚJO, F. M. Biophysical Properties of Cultivated Pastures in the Brazilian Savanna Biome: An Analysis in the Spatial-Temporal Domains Based on Ground and Satellite Data. **Remote Sensing**, v. 5, n. 1, p. 307-326, 2013.

FERREIRA, L. G.; SANO, E. E.; FERNANDEZ, L. E.; ARAÚJO, F. M. Biophysical characteristics and fire occurrence of cultivated pastures in the Brazilian savanna observed by moderate resolution satellite data. **International Journal of Remote Sensing**, v. 34, n. 1, p. 154-167, 2013.

FERREIRA, L. G.; YOSHIOKA, H.; HUETE, A.; SANO, E. E. Seasonal landscape and spectral vegetation index dynamics in the Brazilian Cerrado: An analysis within the Large-Scale Biosphere–Atmosphere Experiment in Amazônia (LBA). **Remote Sensing of Environment**, v. 87, n. 4, p. 534-550, 2003.

HAN, J.; KAMBER, M. **Data mining: concepts and techniques**. 2. ed. San Francisco, CA: Morgan Kaufmann Publishers, 2006, 743 p.

HUETE, A.; DIDAN, K.; MIURA, T.; RODRIGUEZ, E. P.; GAO, X.; FERREIRA, L. G. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. **Remote Sensing of Environment**, v. 83, n. 1, p. 195-213, 2002.

_____. **Manual técnico da vegetação brasileira**. 2. ed. Rio de Janeiro: Manuais Técnicos em Geociências, v. 1, 2012.

INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS (INPE). **TerraView**. São José dos Campos, SP: Divisão de Processamento de Imagens, 2013. Disponível em: <www.dpi.inpe.br/terraview>. Acesso em: 26 Jan. 2014.

JIANG, Z.; HUETE, A. R.; DIDAN, K.; MIURA, T. Development of a two-band enhanced vegetation index without a blue band. **Remote Sensing of Environment**, v. 112, n. 10, p. 3833-3845, 2008.

JUSTICE, C. O.; TOWNSHEND, J. R. G.; VERMOTE, E. F.; MASUOKA, E.; WOLFE, R. E.; SALEOUS, N.; ROY, D. P.; MORISETTE, J. T. An overview of MODIS Land data processing and product status. **Remote Sensing of Environment**, v. 83, n. 1, p. 3-15, 2002.

KLINK, C.; MACHADO, R. Conservation of the Brazilian Cerrado. **Conservation Biology**, v. 19, n. 3, p. 707-713, jun. 2005.

KORTING, T. S.; FONSECA, L. M. G.; CÂMARA, G. GeoDMA - Geographic Data Mining Analyst: a framework for GIScience. **Computers & Geosciences**, v. 57, p. 133-145, 2013. doi: <10.1016/j.cageo.2013.02.007>.

LAMBIN, E.; LINDERMAN, M. Time series of remote sensing data for land change science. **IEEE Transactions on Geoscience and Remote Sensing**, v. 44, n. 7, p. 1926-1928, 2006.

LU, D.; MAUSEL, P.; BRONDÍZIO, E.; MORAN, E. Change detection techniques. **International Journal of Remote Sensing**, v. 25, n. 12, p. 2365-2401, 2004.

MACHADO, R. B.; RAMOS NETO, M.; PEREIRA, P.; CALDAS, E.; GONÇALVES, N.; SANTOS, K.; TABOR, K.; STEININGER, M. **Estimativa de perda da área do Cerrado Brasileiro**. Relatório Técnico. Conservação Internacional, Brasília, DF. 2004.

MALHADO, A. C. M.; PIRES, G. F.; COSTA, M. H. Cerrado Conservation is Essential to Protect the Amazon Rainforest. **Ambio**, v. 39, n. 8, p. 580-584, 2010.

MELGANI, F.; BRUZZONE, L. Classification of hyperspectral remote sensing images with support vector machines. **IEEE Transaction on Geoscience and Remote Sensing**, v. 42, n. 8, p. 1778-1790, 2004.

MILLER, H. J.; HAN, J. **Geographic data mining and knowledge discovery**. Boca Raton, FL: CRC Press, 2009, 445 p.

NOMA, A.; KORTING, T. S.; FONSECA, L. M. G.; PAPA, J. P. Uma comparação entre classificadores usando regiões e perfis EVI para agricultura. In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 16. (SBSR), 2013, Foz do Iguaçu. **Anais...** São José dos Campos: INPE, 2013. p. 2250-2257. DVD, Internet. ISBN 978-85-17-00066-9 (Internet), 978-85-17-00065-2 (DVD). Disponível em:<<http://urlib.net/3ERPFQRTRW34M/3E7GDT8>>. Acesso em: 09 jun. 2014.

RIBEIRO, J. F.; WALTER, B. M. T. As principais fitofisionomias do Bioma Cerrado. In: SANO, S. M.; ALMEIDA, S. P.; RIBEIRO, J. F. **Cerrado: ecologia e flora**. Brasília: EMBRAPA, v. 1, p. 152-212, 2008.

RODRIGUES, R. R.; GANDOLFI, S. Recomposição de florestas nativas: princípios gerais e subsídios para uma definição metodológica. **Revista Brasileira de Horticultura Ornamental**, Campinas, v. 2, n. 1, p. 4-15, 2001.

- RUSHING, J.; RAMACHANDRAN, R.; NAIR, U.; GRAVES, S.; WELCH, R.; LIN, H. ADaM: a data mining toolkit for scientists and engineers. **Computers and Geosciences**, v. 31, n. 5, p. 607-618, jun. 2005.
- SANO, E. E.; BARCELLOS, A. O.; BEZERRA, H. S. Assessing the spatial distribution of cultivated pastures in the Brazilian savanna. **Pasturas Tropicales**, v. 22, n. 3, p. 2-15, 2001.
- SANO, E. E.; ROSA, R.; LUÍS, J.; BRITO, S.; FERREIRA, G. Mapeamento semidetalhado do uso da terra do Bioma Cerrado. **Pesquisa Agropecuária Brasileira**, Brasília, v. 43, n. 1, p. 153-156, 2008.
- SCOLFORO, J. R.; MELLO, J. M.; OLIVEIRA, A. D. **Inventário florestal de Minas Gerais: cerrado - florística, estrutura, diversidade, similaridade, distribuição diamétrica e de altura, volumetria, tendências de crescimento e áreas aptas para manejo florestal**. Lavras: UFLA, 2008.
- SUNG, A. H.; MUKKAMALA, S. **Identifying important features for intrusion detection using support vector machines and neural networks**. In: SYMPOSIUM ON APPLICATIONS AND THE INTERNET, 2003, Washington. **Proceedings...** Washington, DC, USA: IEEE Computer Society. 2003. p. 209-216.
- THEODORIDIS, S.; KOUTROUMBAS, K. **Pattern Recognition**. USA: Academic Press, 2008, 984 p.
- TUCKER, C. J.; PINZON, J. E.; BROWN, M. E.; SLAYBACK, D. A.; PAK, E. W.; MAHONEY, R.; VERMOTE, E. F.; SALEOUS, N. E. An extended AVHRR 8-km NDVI dataset compatible with MODIS and SPOT vegetation NDVI data. **International Journal of Remote Sensing**, v. 26, n. 20, p. 4485-4498, 2005.
- VAPNIK, V. N. **The nature of statistical learning theory**. New York, NY, USA.: Springer-Verlag New York, Inc., 1995, 314 p.
- WALTER, B. M. T. **Fitofisionomias do bioma Cerrado : síntese terminológica e relações florísticas**. 2006. 389p. Tese (Doutorado em Ecologia) - Universidade de Brasília, Brasília. 2006.
- ZHANG, X.; FRIEDL, M. A.; SCHAAF, C. B.; STRAHLER, A. H.; HODGES, J. C.; GAO, F.; REED, B. C.; HUETE, A. Monitoring vegetation phenology using MODIS. **Remote Sensing of Environment**, v. 84, n. 3, p. 471-475, 2003.