



UNIVERSIDADE DE BRASÍLIA  
INSTITUTO DE CIÊNCIAS HUMANAS  
DEPARTAMENTO DE GEOGRAFIA  
LABORATÓRIO DE SISTEMAS DE INFORMAÇÕES ESPACIAIS

**AVALIAÇÃO DE CLASSIFICADORES NÃO-SUPERVISIONADOS,  
ISODATA E K-MEANS, PARA O USO DA TERRA NA BACIA DO RIO  
DAS ONDAS**

RICARDO VIDA E SILVA

ORIENTADOR: PROF. DR. OSMAR ABÍLIO DE CARVALHO JÚNIOR

BRASÍLIA

JUNHO/2009

RICARDO VIDA E SILVA

**AVALIAÇÃO DE CLASSIFICADORES NÃO-SUPERVISIONADOS,  
ISODATA E K-MEANS, PARA O USO DA TERRA NA BACIA DO RIO  
DAS ONDAS**

Monografia de Graduação submetida ao  
Departamento de Geografia da Universidade  
de Brasília, como parte dos requisitos  
necessários à obtenção do Grau de Bacharel  
em Geografia.

ORIENTADOR: PROF. DR. OSMAR ABÍLIO DE CARVALHO JÚNIOR

BRASÍLIA  
JUNHO/2009

UNIVERSIDADE DE BRASÍLIA  
INSTITUTO DE CIÊNCIAS HUMANAS  
DEPARTAMENTO DE GEOGRAFIA  
LABORATÓRIO DE SISTEMAS DE INFORMAÇÕES ESPACIAIS

**AVALIAÇÃO DE CLASSIFICADORES NÃO-SUPERVISIONADOS, ISODATA E K-MEANS,  
PARA O USO DA TERRA NA BACIA DO RIO DAS ONDAS**

RICARDO VIDA E SILVA

MONOGRAFIA DE GRADUAÇÃO APROVADA PELA BANCA EXAMINADORA,  
CONSTITUÍDA POR:

---

PROF. DR. OSMAR ABÍLIO DE CARVALHO JÚNIOR - UNIVERSIDADE DE BRASÍLIA  
(ORIENTADOR)

---

PROF. DR. ROBERTO ARNALDO TRANCOSO GOMES – UNIVERSIDADE DE BRASÍLIA  
(EXAMINADOR INTERNO)

---

PROF. M.Sc. MARCUS FÁBIO RIBEIRO FARIAS – UNIVERSIDADE DE BRASÍLIA  
(EXAMINADOR INTERNO)

BRASÍLIA, 25 DE JUNHO DE 2009.

## FICHA CATALOGRÁFICA

SILVA, RICARDO VIDA e

Avaliação de Classificadores não-supervisionados, ISODATA e K-MEANS, para o uso da terra na Bacia do Rio das Ondas, 37 p., 297 mm, (UnB-IH-GEA-LSIE, Graduação, 2009).

Monografia de Graduação – Universidade de Brasília. Departamento de Geografia.

1. Classificador automático

2. Uso da terra

3. Matriz de Confusão

4. Sensoriamento Remoto

I. UnB-IH-GEA-LSIE

II. Título (série)

## REFERÊNCIA BIBLIOGRÁFICA

SILVA, Ricardo Vida e. Avaliação de Classificadores não-supervisionados, ISODATA e K-MEANS, para o uso da terra na Bacia do Rio das Ondas. (Monografia de Graduação), Curso de Graduação em Geografia, Universidade de Brasília, 2009. 37 p.

## CESSÃO DE DIREITOS

NOME DO AUTOR: Ricardo Vida e Silva.

TÍTULO DA MONOGRAFIA: “Avaliação de Classificadores não-supervisionados, ISODATA e K-MEANS, para o uso da terra na Bacia do Rio das Ondas”.

GRAU/ANO: Bacharel/2009.

É concedida à Universidade de Brasília permissão para reproduzir cópias desta monografia e emprestar ou vender tais cópias somente para propósitos acadêmicos e científicos. O autor reserva outros direitos de publicação e nenhuma parte desta monografia de graduação pode ser reproduzida sem a autorização por escrito do autor.

---

Ricardo Vida e Silva

## **AGRADECIMENTOS**

Agradeço inicialmente minha mãe, Maria de Fátima da Silva, que de longe sempre me deu carinho e preciosos ensinamentos. Ao Donizeti Vida que sempre acreditou em seus filhos e fez, e ainda faz, de **tudo** para que os nossos sonhos se concretizem. Seguindo os agradecimentos familiares, agradeço aos meus irmãos Luciano Vida, Fernando Vida, Barbara Vida e Mariane Vida que sempre agüentaram (sem reclamar) minhas amolações. Agradeço a Vó Elza que com sabedoria criou essa grande família que tenho orgulho de pertencer.

Ao professor Osmar Abílio pela paciência e dedicação como orientador.

Aos colegas dos LSIE.

Aos meus amigos.



## LISTA DE FIGURAS





Informações sobre o uso e cobertura da terra assumem um relevante papel em políticas de planejamento. As técnicas de sensoriamento remoto auxiliam na aquisição desses dados de forma rápida, repetitiva e confiável e permite um monitoramento constante desse aspecto da paisagem. Nesse sentido, o objetivo desse trabalho é avaliar qual classificador, ISODATA ou K-MEANS, possui maior acurácia tendo como referencia a classificação produzida a partir de Interpretação Visual do uso e cobertura da terra. Com a intenção de produzir uma base de dados mais próxima da realidade, a Interpretação Visual foi feita sobre uma imagem de alta resolução espectral do sensor PRISM/ALOS. Dito isso, a metodologia adotada para analisar os dois classificadores foi comparar a Matriz de Confusão, o índice *Kappa* e área de pixels que não coincidiram, entre o classificador e a Interpretação Visual. Dessa maneira, o processo de avaliação contou com técnicas de estatísticas a fim de comprovar, de forma matemática, qual classificador apresenta melhor resultado para esse tipo de estudo. O classificador K-MEANS em todas as análises demonstrou melhor concordância com a Interpretação Visual do que o classificador ISODATA.

**Palavras chave:** Uso e cobertura da terra, Classificador automático e Matriz de Confusão.

## 1 – INTRODUÇÃO

A crescente interferência do homem no meio natural com a conseqüente transformação do mesmo torna evidente a importância de estudos referentes à paisagem. Esses estudos devem servir de subsídios na elaboração de planejamentos ligados à relação entre homem e natureza, para que desta forma, seja minimizada a degradação ambiental que é proveniente dos processos de desenvolvimento e ocupação do espaço pelas atividades humanas (Christofolletti, 1993).

Entre os estudos referentes à paisagem, o levantamento de uso da terra assume um relevante papel nos projetos e ações voltado ao planejamento, sendo destacado por inúmeros estudiosos que se utilizam de diversos métodos para avaliar as alterações provocadas pelo uso e, assim, fornecer subsídios para o manejo eficiente dos ambientes naturais. O estudo do uso da terra consiste em conhecer como o espaço está sendo ocupado pelo homem, além de caracterizar os tipos de vegetação que reveste o solo. Atualmente se torna cada vez mais necessário aos legisladores e planejadores ter um conhecimento atualizado do uso da terra, permitindo com isso uma melhor elaboração de política de uso e ocupação do solo.

Nesse sentido, as técnicas de sensoriamento remoto (Rosa, 2005) permitem a aquisição de dados de forma global, confiável, rápida e repetitiva, sendo esses dados de grande importância para o levantamento, mapeamento e utilização das informações de uso da terra. O sensoriamento remoto, segundo Curran (1986), é a técnica que permite adquirir informações sobre a superfície terrestre a partir da detecção e registro da energia resultante da interação entre a radiação eletromagnética e a matéria em estudo.

No entanto, Bernardes *et al.* (2007) alerta que devido à geração de grande volume diário de dados sobre a superfície terrestre e sua cobertura vegetal, se torna necessário definir procedimentos confiáveis de mapeamento e interpretação que acompanhe com rapidez suficiente a velocidade em que esses dados são gerados. Apesar do significativo ganho de tempo através de algoritmos de classificação automática, o desempenho dessa técnica ainda depende de características da área de

estudo como o relevo, diversidade e tamanho de fragmentos da cobertura vegetal já que a análise digital considera, quase exclusivamente, a intensidade radiométrica de cada pixel. Entretanto, a classificação visual já incorpora na interpretação da imagem, outros elementos além do valor digital do pixel, isto é, as tradicionais técnicas da fotointerpretação visual.

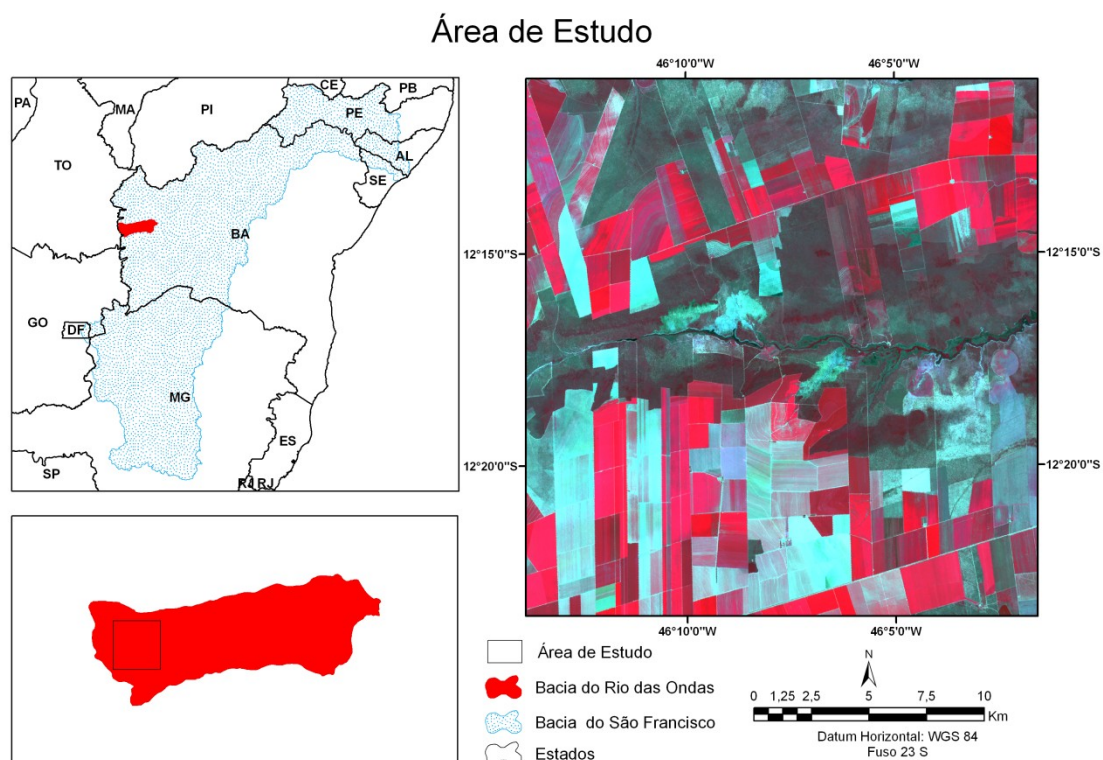
O presente trabalho tem o objetivo de comparar métodos de classificação não supervisionados, ISODATA e K-MEANS, com Interpretação Visual. Com esse trabalho espera-se avaliar qual método de classificação não-supervisionado apresenta melhor resultado ao mapear o uso da terra, utilizando imagens de satélite.

A área utilizada para fazer os testes de classificação insere-se no contexto da bacia hidrográfica do Rio das Ondas. A Lei 9.433/97 afirma a importância da bacia hidrográfica como unidade de análise já que a defini como unidade territorial para implementação da Política Nacional de Recursos Hídricos.

## 2 - ÁREA DE ESTUDO

### 2.1 – Localização

A área de estudo é um quadrante do alto curso da bacia hidrográfica do Rio das Ondas (Figura 1) que se localiza na região oeste da Bahia, entre as coordenadas geográficas  $46^{\circ}1'10''$  -  $46^{\circ}14'15''$  de longitude oeste e  $12^{\circ}12'27''$  -  $12^{\circ}23'59''$  de latitude sul. A bacia do Rio das Ondas é uma sub-bacia do Rio Grande e faz divisa com as sub-bacias do Rio de Janeiro (ao norte) e do Rio das Fêmeas (ao sul). E pertence às sub-bacias da margem esquerda do médio São Francisco. A área drenada pela bacia do Rio das Ondas é de 5.175 Km<sup>2</sup> e envolve parte dos municípios baianos de Barreiras e Luiz Eduardo Magalhães. São afluentes do Rio das Ondas: o rio Borá, o rio das Pedras e o rio Vereda das Lages.



**Figura : Localização da Área de Estudo.**

## 2.2 – Clima

O clima da área da bacia do Rio das Ondas é o Aw tropical (classificação de Koppen), com inverno seco e verão chuvoso. A temperatura média é de 24 °C, com máxima de 32 °C e mínima de 18 °C. A precipitação pluvial anual varia de 900 mm a 1700 mm, na fronteira com os Estados de Goiás e Tocantins. Dessa precipitação, aproximadamente, 95% concentra-se nos meses de outubro a abril (Soares Neto, 2005).

## 2.3 – Geologia

O contexto geológico da bacia do Rio das Ondas compreende o Grupo Urucuia, o Grupo Bambuí, Coberturas Detríticas e sedimentos aluvionares. Soares Neto (2005) caracteriza essas litologias em:

- **Grupo Urucuia:** é o mais expressivo, ocupando 97,76% da área da bacia hidrográfica e compreende toda parte de chapada. A composição litológica desse grupo é constituída por arenitos finos a médios, róseos, impuros, com alguns níveis conglomeráticos. Há uma tendência desses arenitos se tornarem mais argilosos na base;
- **Grupo Bambuí:** detém 1,58% da área da bacia do Rio das Ondas. O Grupo Bambuí compreende litologicamente calcários, calcários dolomíticos, dolomíticos, margas, siltitos, argilitos, folhelhos, arcóseos e os termos intermediários entre essas litologias;
- **Coberturas Detríticas:** ocupam 0,19% e tem ocorrência próxima a foz do Rio das Ondas, as Coberturas Detríticas são compostas por elúvios, ou seja, materiais provenientes do intemperismo das rochas subjacentes e não transportado, e por colúvios, material pouco transportado, acumulado nos sopés das encostas dos platôs;

- **Sedimentos Aluvionares:** esses sedimentos preenchem a calhas e a planície de inundação das principais drenagens. São compostos por areias, cascalhos, silte e argilas (Radambrasil, 1983).

## 2.4 – Geomorfologia

A bacia do Rio das Ondas possui três unidades geomorfológicas: Chapadão do Rio Grande, Patamares do Chapadão e Depressão do Rio Grande. O Chapadão do Rio Grande se caracteriza por superfícies aparentemente homogênea, plana, com inclinação leve orientada para o leste, e por subunidades de planos de topo (partes mais elevadas), rampas (inclinações suaves ao longo das drenagens) e veredas (vales com amplas várzeas). Já, a unidade geomorfológica Patamar do Chapadão se limita com a primeira unidade por ressaltos e por acentuadas escarpas. Essa unidade forma um degrau entre o Chapadão do Rio Grande e Depressão do Rio Grande e encontra-se entre os níveis altimétricos de 500 a 700 metros (Soares Neto, 2005).

A última unidade, a Depressão do Rio Grande inclui, principalmente, rampas suavemente inclinadas com depressões fechadas (Soares Neto, 2005 *apud* Bahia, 1993).

## 2.5 – Solos

Sobre a bacia do Rio das Ondas desenvolvem-se seis tipos de solos que foram identificados como: Latossolo Vermelho-Amarelo, Neossolo Quartzarênico, Gleissolo, Argissolo Vermelho-Amarelo, Neossolo Flúvico e Neossolo Litólico. A classe com maior ocorrência é o Latossolo Vermelho-Amarelo. Esse tipo de solo é normalmente muito profundo e os horizontes A, B e C têm pouca diferenciação, sendo as transições entre eles difusas e graduais. Em geral, são solos fortemente ácidos e com baixa saturação por bases. A segunda classe com maior área de ocorrência é o Neossolo Quartzarênico. Esse solo deriva dos arenitos do Grupo Urucua e aparecem em áreas planas ligeiramente rebaixadas da chapada e geralmente próximo as drenagens. Os Gleissolos ocupam os fundos dos vales e próximo aos canais de drenagem. Uma unidade pouco

representativa é o Argissolo Vermelho-Amarelo que é encontrada em trechos das depressões. Já os Neossolos Flúvicos são solos poucos desenvolvidos e sua constituição é de sedimentos recentes não consolidados. Por último, o Neossolo Litólico freqüentemente associado a afloramentos rochosos. Esse solo é pouco desenvolvido e raso (Soares Neto, 2005).

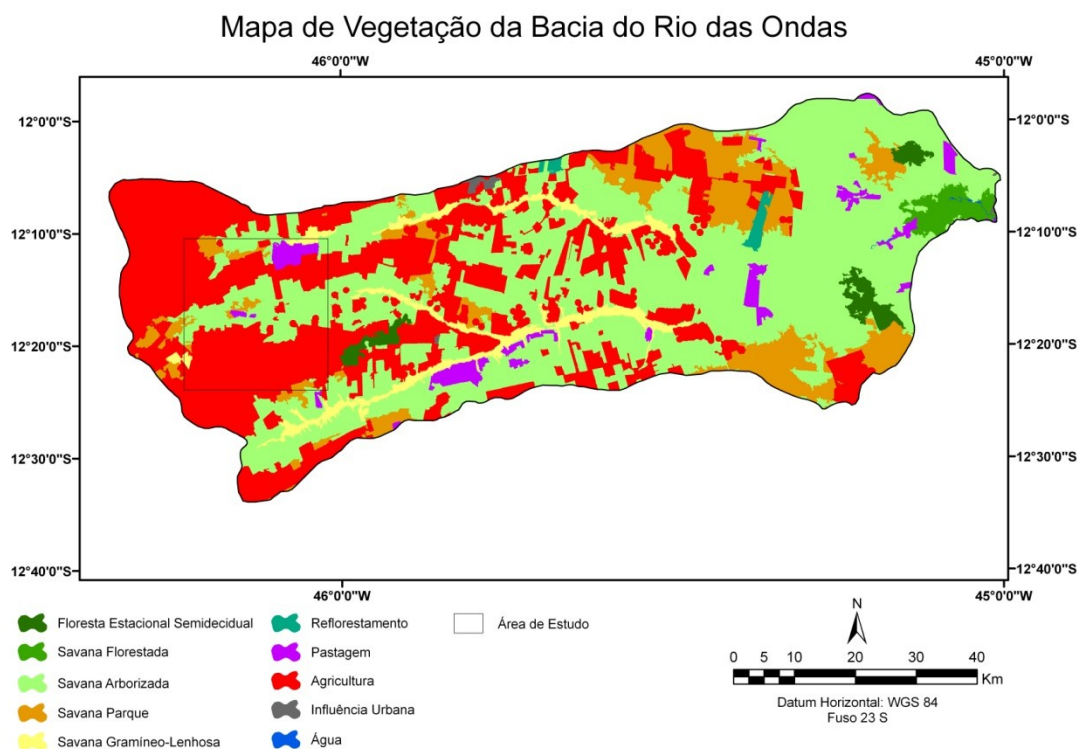
## 2.6 – Vegetação

A bacia do Rio das Ondas se insere no bioma Cerrado, classificado como uma formação do tipo savana tropical e composto por formações florestais, savânicas e campestres.

Entre as fitofisionomias existente no Cerrado foram identificadas as seguintes classes na bacia do Rio das Ondas (PROBIO, 2004; IBGE, 1992), de acordo com a figura 2:

- **Floresta Estacional Semidecidual:** essa classe possui dupla estacionalidade climática, uma tropical com época de intensas chuvas de verão e outra subtropical sem período seco, mas com seca fisiológica provocada pelo frio do inverno;
- **Savana Florestada:** conhecido também como “Cerradão”, esse subgrupo apresenta fisionomia típica e característica restrita das áreas areníticas lixiviadas com solos profundos. Ocorre em clima tropical eminentemente estacional;
- **Savana Arborizada:** formada naturalmente e/ou por ação antrópica e se caracteriza por apresentar uma fisionomia nanofanerofítica rala e outra hemicriptofítica graminóide, contínua, sujeita ao fogo anual;
- **Savana Parque:** é constituída por um estrato graminóide, entremeado por nanofanerófitos isolados;
- **Savana Gramíneo-Lenhosa:** quando natural, prevalecem nessa fisionomia os gramados entremeados por plantas lenhosas raquíticas;

Há também ocorrência na bacia do Rio das Ondas de reflorestamentos, pastagens, culturas agrícolas e áreas de vegetação com influência urbana.



**Figura : Mapa de Vegetação da Bacia do Rio das Ondas (PROBIO, 2004).**

## **2.7 - Aspectos Socioeconômicos**

A bacia do Rio das Ondas envolve partes de dois municípios baianos: Barreiras e Luiz Eduardo Magalhães.

Barreiras é o município com maior área drenada pela bacia do Rio das Ondas. A população é de 129.501 pessoas, na contagem de 2007, e a área territorial é de 7.895 Km<sup>2</sup>, com um PIB per capita de 8.599 reais, no ano de 2006 (IBGE, 2009). O município é descrito como a “Capital de Serviços da Região Oeste da Bahia” e sua economia começou a se fortalecer a partir dos anos 80, com a chegada de agricultores da região sul do país. Esses agricultores introduziram a cultura da soja nas áreas de sequeiro, com os plantios desenvolvidos no período das chuvas (Barreiras, 2009).



Como dito anteriormente, o outro município drenado pela bacia do Rio das Ondas é o Luis Eduardo Magalhães. A população desse município na contagem de 2007 era de 44.265 habitantes. Com área territorial de 4.019 km<sup>2</sup> e PIB per capita de 37.106 reais, em 2006 (IBGE, 2009). Com esse valor, Luis Eduardo Magalhães está entre os municípios baianos com maior desenvolvimento econômico.

De modo geral, a região oeste do Estado da Bahia, de acordo com o site Monitoramento da Expansão das Áreas Irrigadas na região oeste da Bahia, desenvolve tradicionalmente a pecuária extensiva e recentemente, cerca de 10 anos, passou a praticar de modo extraordinário a atividade agrícola. São produzidos na região grãos e fibras (soja, milho, arroz e algodão, principalmente), café, fruticultura irrigada e pecuária (bovina, suína e avícola). Logo, a base econômica da região é praticamente primária. Vale ressaltar que a soja é o produto agrícola que tem maior peso, tanto na área cultivada como na produção (Soares Neto, 2005).

### 3 - MATERIAIS E MÉTODOS

#### 3.1 - Imagens dos sensores ALOS e ASTER

A base de dados do uso da terra foi elaborada a partir de imagens de alta resolução do sensor PRISM/ALOS e a classificação não-supervisionada numa imagem do sensor ASTER.

##### 3.1.1 - Sensor ALOS

O IBGE é o instituto responsável pela distribuição das imagens ALOS para os órgãos do Governo Federal, instituições de pesquisa e demais usuários não comerciais do Brasil. O satélite japonês *Advanced Land Observing Satellite* (ALOS) foi lançado em 2006 pela Agência Espacial Japonesa (JAXA) com a missão de observar e obter imagens de todo o planeta. Essas imagens teriam a finalidade de monitorar desastres ambientais, levantar informações de recursos naturais e, principalmente, dar suporte à cartografia. No entanto, o satélite ALOS entra em fase de operação só no dia 20 de outubro de 2006, após nove meses de calibração (IBGE, 2009).

O satélite ALOS descreve uma órbita circular heliossíncrona (órbita que o satélite fica permanentemente exposto aos raios solares), a 692 km de altitude. O período de revisita do satélite ALOS a um mesmo ponto é de 46 dias (ciclo). Os dados captados são redirecionados para o satélite geoestacionário DRTS que os transmitem a estação de recepção no Japão, a uma taxa de 240 Mbps. O satélite ALOS é constituído por três sensores: *Panchromatic Remote Sensing Instruments for Stereo Mapping* (PRISM), *Advanced Visible and Near Infrared Radiometer type 2* (AVNIR-2) e *Phased Array type L-band Synthetic Aperture Radar* (PALSAR). Como dito anteriormente, foram utilizadas imagens do sensor PRISM/ALOS que opera na faixa da luz visível, com uma (1) banda pancromática de 8 bits e resolução espacial de 2,5 m. A configuração de imageamento é chamada de Triplet, ou seja, é um conjunto de 3 sistemas de imageamento que permite obter simultaneamente cenas com visadas nadir, inclinada para frente e inclinada para trás. Essa configuração torna possível adquirir imagens

estereoscópicas ao longo da trajetória. A tabela a seguir resume as informações do sensor PRISM (IBGE, 2009):

**Tabela : Sensor PRISM/ALOS (IBGE, 2009).**

<b>Características do sensor PRISM</b>	
No. de Bandas	1 (Pancromática)
Comprimento de Onda	0,52 ~ 0,77 microns
No. de Sensores	3 (nadir/para frente/para trás)
Relação Base/Altura (B/H)	1,0 (entre as visadas para frente e para trás)
Resolução Espacial	2,5 m
Largura da Faixa	35 km (modo triplet) 70 km (apenas nadir)
Relação Sinal/Ruído	>70
Função de Transferência de Modulação (MTF)	>0.2
No. de Detectores	28.000/banda (faixa de 70 km) 14.000/banda (faixa de 35 km)
Inclinação da Visada	de -1,5 a +1,5 graus (transversal à trajetória no modo Triplet)
Resolução Radiométrica	8 bits

As imagens utilizadas pertencem ao produto georreferenciado 2B que são submetidas à calibração radiométrica e geométrica, com os pixels alinhados com a grade da projeção UTM. O erro médio quadrático da precisão geométrica absoluta para a visão nadir é de 8m (cross track) e 9m (along track), enquanto que a precisão relativa é de 4m (cross track) e 3m (along track) (JAXA, 2009).

### 3.1.2 - Sensor ASTER

O sensor ASTER é proveniente de um esforço cooperativo entre a NASA – Earth Observing System (EOS), o Japan's Ministry of Economy, Trade and Industry (METI) e o Earth Remote Sensing Data Analysis Center (ERSDAC). Esse sensor encontra-se a bordo satélite TERRA (ou EOS-AM) que opera numa órbita polar e síncrona com o Sol, com inclinação de aproximadamente 98,2 graus e altitude média de 705 km (Rosa, 2005).

Os dados desse sensor vêm sendo utilizados em análises de temperatura, emissividade, refletância e elevação da superfície do terreno. Este sensor consiste de três subsistemas: (a) visível e infravermelho próximo (VNIR – 0,5 $\mu$ m-0,9 $\mu$ m), constituído por 3 bandas espectrais com resolução de 15 metros, (b) infravermelho ondas curtas (SWIR – 1,6 $\mu$ m-2,5  $\mu$ m), com 6 bandas espectrais de resolução espacial de 30 metros e (c) infravermelho termal (TIR), com 5 bandas espectrais de resolução espacial de 90 metros (Abrams, 2000; Fujisada et al., 1998; Yamaguchi et al., 1998). O sensor ASTER é um instrumento de resolução espacial mais alta do satélite EOS-AM e o único que não coleta dados de forma contínua (Rosa, 2005).

A tabela abaixo resume as características do sensor ASTER:

**Tabela : Sensor ASTER.**

<b>Características do Sensor ASTER</b>		
<b>Bandas Espectrais</b>	<b>Resolução Espectral</b>	<b>Resolução Espacial</b>
VNIR	3 bandas: (0,5 - 0,9 $\mu$ m)	15 m
SWIR	6 bandas: (1,6 - 2,5 $\mu$ m)	30 m
TIR	5 bandas: (8 - 12 $\mu$ m)	90 m

As imagens utilizadas (VNIR e SWIR) foram adquiridas já corrigidas do efeito atmosférico. Esse produto do sensor ASTER corresponde à especificação AST07.

### 3.2. – Interpretação Visual do Uso da Terra

A fotointerpretação da imagem pancromática de resolução espacial de 2,5 m do sensor PRISM/ALOS e, posteriormente, a interpretação visual do uso da terra considerou características como tonalidade/cor, textura, forma, tamanho e a sombra da imagem. Esses elementos de fotointerpretação são definidos por Rosa (2005):

- **Tonalidade/cor:** relaciona-se com a intensidade da radiação eletromagnética refletida ou emitida pelos alvos. A tonalidade é as diferentes graduações de cinza e é um elemento essencial na interpretação de fotografias aéreas e de imagens de satélite. Porém, um mesmo alvo pode aparecer com tonalidades diferentes porque variáveis como a hora, local e época do ano interferem nas graduações de cinza desse alvo.
- **Textura:** é o padrão de arranjo espacial dos elementos texturais, sendo o elemento textural a menor feição contínua e homogênea que se distingue na fotografia aérea ou na imagem de satélite.
- **Forma:** de modo geral, as feições naturais apresentam formas irregulares e as feições trabalhadas pelo homem possuem formas geométricas.
- **Tamanho:** dependendo da escala, o tamanho pode ser utilizado para identificar feições individuais.
- **Sombra:** são fenômenos comuns nas imagens de satélite registradas no inverno. Elas acontecem devido à iluminação oblíqua do Sol nesse período, ou da ausência do retorno do sinal, nesse caso, nos sensores ativos.
- **Padrão:** em imagens de satélite a extração de informações constitui basicamente no exame e na identificação de diferentes padrões tonais e texturais.

### 3.3. - Classificação Não-Supervisionada

Com o propósito de aprimorar e aperfeiçoar os procedimentos de classificação foram realizados testes com métodos classificação automatizados sobre a imagem do sensor ASTER. Desta forma, foram avaliados métodos de classificação não supervisionada que estão no âmbito da análise de grupos. A utilização da análise de grupos vem sendo cada vez mais empregada como uma forte ferramenta na investigação científica, contribuindo significativamente para uma análise exploratória dos sistemas existentes. Desta forma, este método estatístico favorece a formulação de hipóteses sobre a estrutura da distribuição interna dos dados. A análise de grupos tem como propósito particionar um conjunto de  $N$  entidades em subconjuntos distintos e não vazios, que sejam tão homogêneos quanto possível (Sampaio, 1990). A classificação dos grupos é definida por intermédio de uma medida de similaridade ou dissimilaridade entre os parâmetros, como por exemplo, o da distância Euclidiana. No final do processo são definidos grupos ou clusters que sejam pertinentes e decorrentes naturais da própria estruturação dos dados. Em outras palavras, na classificação não-supervisionada é o próprio computador que decide, com base em regras estatísticas, quais as classes que vão ser separadas e quais os “pixels” que vão compor cada classe. Nos classificadores não-supervisionados não há qualquer conhecimento prévio sobre os atributos das classes pertinentes a cena. Desta forma, esses classificadores permitem uma análise exploratória dos dados sem a intervenção do usuário. Neste trabalho foram considerados os seguintes métodos: ISODATA e K-MEANS.

O ISODATA e o K-MEANS são métodos não hierárquicos, que primeiramente selecionam centros para os agrupamentos denominados sementes e a partir de uma distância pré-determinada agrupa os elementos em sua volta. Segundo Galvão *et al* (2005), o algoritmo ISODATA baseia-se na análise de agrupamentos em que são identificadas, no espaço de atributos, as nuvens (clusters) formadas por pixels com características similares. De acordo com Tou *et al.* (1974), a vantagem do classificador ISODATA é que ele não exige um conhecimento prévio da área de estudo. No entanto, o usuário tem pouco controle sobre a separação entre as classes ou da determinação precisa do número delas, já que atribui um número mínimo e máximo de classes. Já o

classificador K-MEANS calcula inicialmente as classes distribuindo em uma classe uniformemente no espaço e posteriormente aglomera classe por classe em um processo iterativo. Esse método não-supervisionado utiliza a técnica de distância mínima e a classificação será melhor quanto melhor for agrupada a nuvem de pixels. Nesse caso, a quantidade de classes é atribuída pelo usuário.

### **3.4 – Análise das Classificações**

Com o objetivo de avaliar a qualidade das diferentes classificações ISODATA e K-MEANS, aplicadas na área de estudo, utilizou a interpretação visual como referência, conseguindo, desse modo, avaliar o quanto essas classificações se diferenciavam da base de análise, ou seja, a quantidade de pixels classificados de maneira incorreta. O processo de avaliação baseou-se na Matriz de Confusão e o seu subproduto, o índice *kappa*.

A Matriz de Confusão ou Matriz de Erros tem a finalidade de avaliar o resultado de uma classificação ao comparar a classificação com os dados de uma base de referência, agrupando-os numa tabela. O coeficiente *kappa* ( $k$ ) é um subproduto da matriz de confusão e auxilia na avaliação do classificador. Uma de suas vantagens é que ele incorpora a informação dos pixels mal classificados, e não apenas dos pixels bem classificados, como a exatidão global (Jensen, 1986).

A equação do coeficiente *Kappa* é dada por (Cohen, 1960):

$$\hat{k} = \frac{N \sum_{i=1}^{\gamma} x_{ii} - \sum_{i=1}^{\gamma} x_{i+} x_{+i}}{N^2 - \sum_{i=1}^{\gamma} x_{i+} x_{+i}}$$

onde,

$\Sigma$  representa o somatório em cada linha e coluna;

$\gamma$  é o número de linhas e de colunas;

$N$  é o número total de pontos, o somatório de toda matriz;

Dividindo o numerador e o denominador por  $N^2$

$$\hat{k} = \frac{\theta_1 - \theta_2}{1 - \theta_2}$$

onde

$$\theta_1 = \frac{\sum_{i=1}^{\gamma} x_{ii}}{N}$$

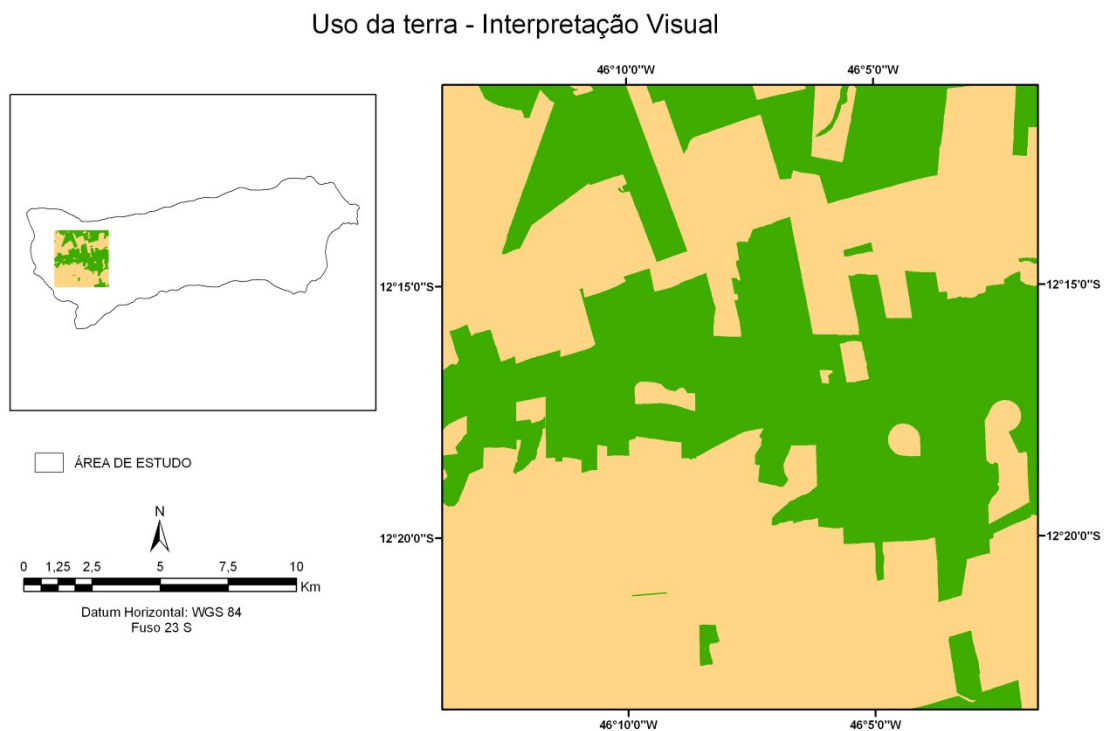
$$\theta_2 = \frac{\sum_{i=1}^{\gamma} x_{i+} x_{+i}}{N^2}$$



## 4 – RESULTADOS

### 4.1 – Interpretação Visual

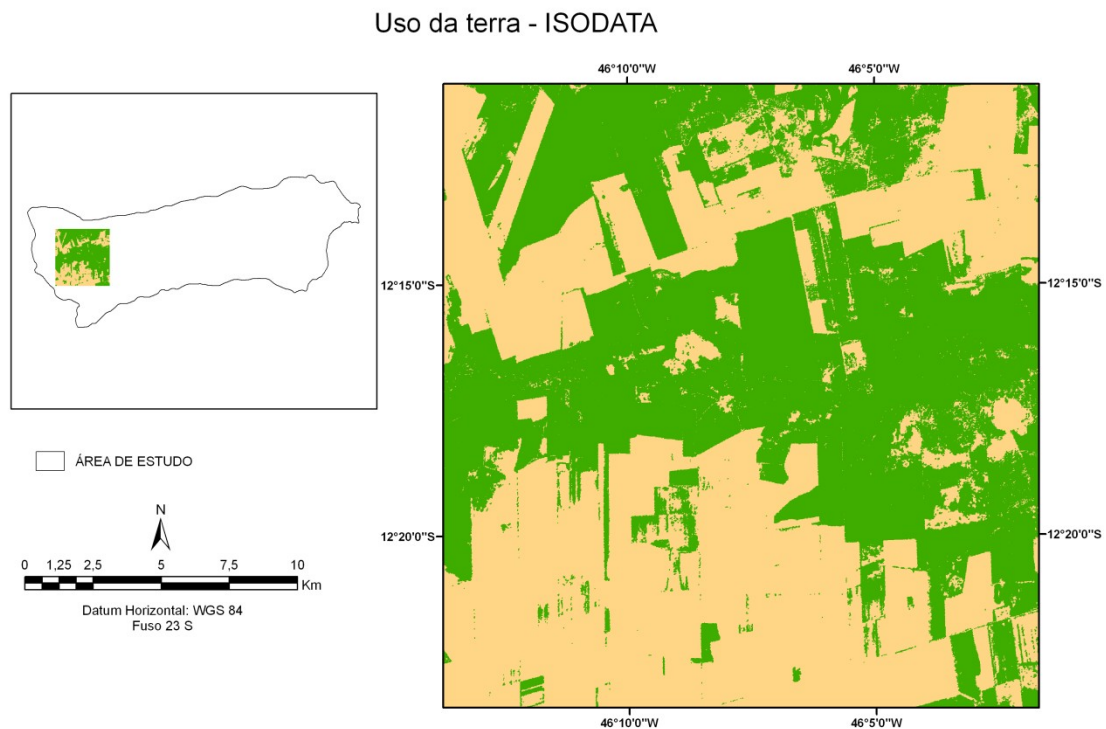
A interpretação visual da área de estudo (Figura 3), a partir da imagem do sensor PRISM/ALOS, estabeleceu duas classes de uso da terra: a área alterada e a área natural. Na classe de área alterada incluiu áreas de agricultura, sede de propriedade, pecuária, pivô e reflorestamento, totalizando, aproximadamente, 359 Km<sup>2</sup>. A outra classe, área natural, abrangeu a vegetação natural e massa d'água, com cerca de 226 Km<sup>2</sup> de área.



**Figura : Mapa de Uso da Terra a partir da Interpretação Visual.**

## 4.2 - Classificação Não-Supervisionada

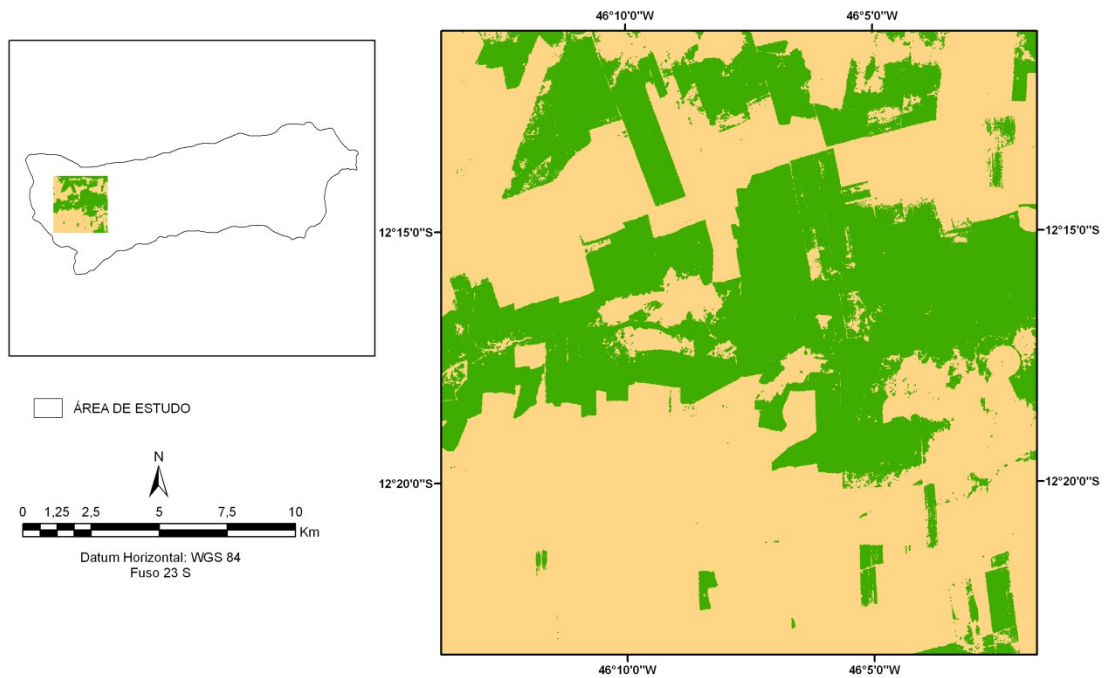
A classificação automatizada pelo método ISODATA estabeleceu as duas classes de uso da terra, após aglutinar as classes que apareceram em excesso à duas classe pré-determinada (Figura 4). A classe de área alterada possui, aproximadamente, 327 Km<sup>2</sup> e a classe de área natural 259 Km<sup>2</sup>.



**Figura : Mapa de Uso da Terra a partir da classificação ISODATA**

O método K-MEANS classificou as duas classes (Figura 5). A área alterada com cerca de 393 Km<sup>2</sup> e área natural 193 Km<sup>2</sup>.

#### Uso da terra - KMEANS



**Figura : Mapa de Uso da Terra a partir da classificação K-MEANS**

#### 4.3 - Comparações entre as classificações

A interpretação visual delimitou de modo satisfatório as duas classes propostas, tanto em relação à forma como em relação ao tamanho das classes. As duas classes assumiram formas geométricas, pois a classe natural, ao fazer limite com a classe transformada, teve que adotar o contorno desta. De forma geral, a classificação pelo método ISODATA e K-MEANS ficaram heterogêneas e os pixels apareceram espalhados. Tal fato aconteceu devido à incapacidade do classificador automático em diferenciar, ou agrupar, os pixels tendo como único parâmetro, o valor do pixel. Além dessa comparação preliminar, foi utilizado a Matriz de Confusão e o índice *Kappa* a fim de avaliar qual das classificações aproximou mais da Interpretação Visual do uso da terra. A Matriz de Confusão e o índice *Kappa* foram gerados pelo software ENVI 4.4.

A tabela 3 tem os dados da Matriz de confusão para o classificador ISODATA.

Tabela : Matriz de Confusão do classificador ISODATA, adaptada de Richards, J. A. (1986).

		Interpretação Visual		Total	Inclusão (%)	Pixels bem classificados (%)
		Área Natural	Área Alterada			
Classificação ISODATA	Área Natural	735177	84477	819654	10,31	89,69
	Área Alterada	164642	482506	647148	25,44	74,56
Total pixels de campo		899819	566983	1466802		Exatidão Global (%)
Omissão (%)		18,30	14,90			83,02

Esse trabalho adotou duas classes: Área Natural e Área Alterada. A Matriz de Confusão se restringe as linhas e as colunas dessas duas classes. A diagonal principal (células cinza) da matriz fornece a quantidade de pixels corretamente classificados para cada classe correspondente. Por exemplo, para Área Natural, na classificação ISODATA foram classificados corretamente 735.177 pixels. No entanto, 84.477 pixels foram classificados incorretamente, pertencendo a Área Alterada.

Portanto, para 819.654 pixels da Área Natural, 735.177 (89,69%) foram bem classificados, enquanto o restante 84.477 (10,31%) foram classificados de maneira incorreta. Este erro de classificação é denominado como **erro de inclusão, commission**, já que inclui pixel em uma classe quando na verdade ele pertence à outra. Esse tipo de análise também pode ser feita para a Área Alterada do classificador ISODATA.

Do ponto de vista da Interpretação Visual, agora para a Área Alterada, temos 482.506 pixels bem classificados e 84.477 classificados incorretamente. Esses pixels classificados incorretamente pertencem ao **erro de omissão, omission**, ou seja, 84.477 (14,90%) pixels foram omitidos da classe correta e atribuídos a outra classe.

Na última coluna da tabela aparecem os valores da exatidão específica de cada classe, ou seja, a percentagem de pixels classificados corretamente. A última célula dessa coluna fornece a **exatidão global, accuracy**, do classificador ISODATA. Esse percentual corresponde a divisão entre os pixels bem classificados (1.217.683) pelo total de pixel das duas classes (1.466.802). Portanto, a percentagem da exatidão global do classificador ISODATA é de 83,02%.

A Matriz de Confusão do classificador K-MEANS (Tabela 4) apresenta os seguintes valores:

**Tabela : Matriz de Confusão do classificador K-MEANS, adaptada de Richards, J. A. (1986).**

		Interpretação Visual		Total	Inclusão (%)	Pixels bem classificados (%)
		Área Natural	Área Alterada			
Classificação K-MEANS	Área Natural	836384	146646	983030	14,92	85,08
	Área Alterada	63435	420337	483772	13,11	86,89
Total pixels de campo		899819	566983	1466802		Exatidão Global (%)
Omissão (%)		7,05	25,86			85,68

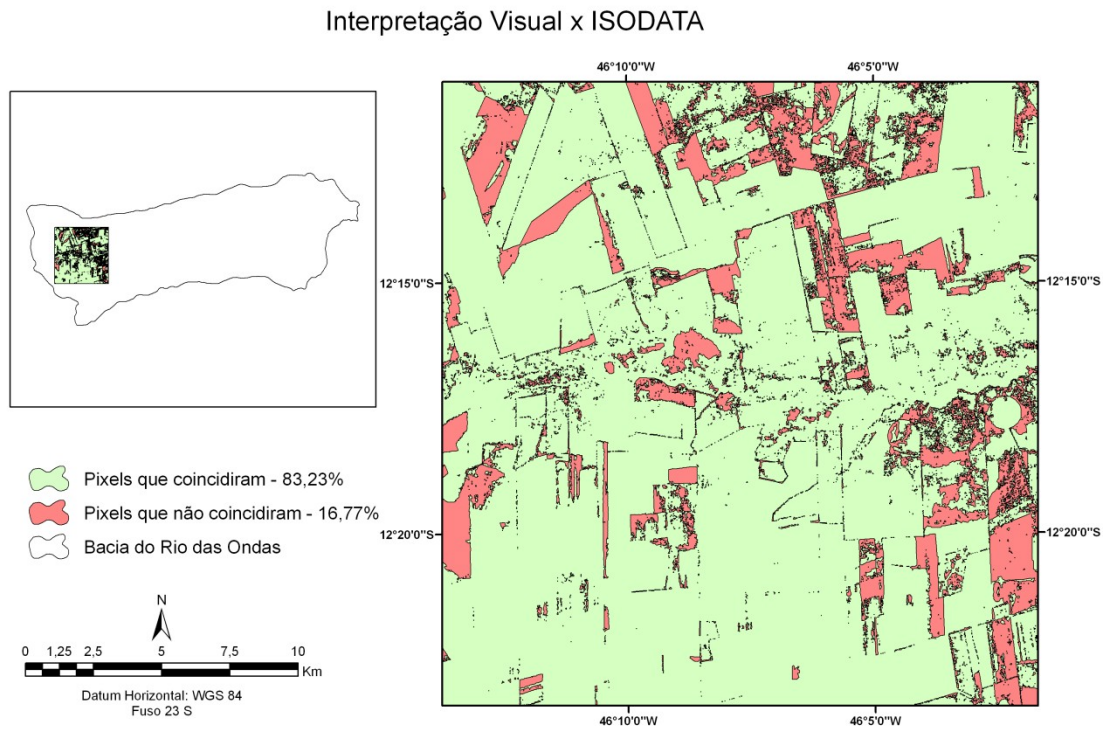
É possível deduzir da tabela acima que a exatidão específica da Área Alterada do classificador K-MEANS (86,89%) é melhor que do classificador ISODATA (74,56%) para a mesma classe. No entanto, o classificador K-MEANS não obteve um melhor resultado para a Área Natural (85,08%) do que o classificador ISODATA (89,69%). Por fim, com base na exatidão global, o classificador K-MEANS apresenta maior semelhança com a Interpretação Visual, com a percentagem de 85,68% contra 83,02% do classificador ISODATA.

A análise entre os classificadores também pode ser feita comparando o subproduto da Matriz de Confusão, o índice *Kappa*. Como mencionado anteriormente, esse índice tem a vantagem de incorporar informações dos pixels classificados incorretamente, e não apenas dos classificados corretamente como a exatidão global.

O índice *Kappa* do classificador ISODATA é 0,6510 e o classificador K-MEANS possui índice *Kappa* de 0,6896. O índice *Kappa* é um valor que vai do intervalo de 0 a 1, isto é, quanto mais próximo de 1 mais semelhante é a classificação. Dito isso, o classificador K-MEANS, também pelo índice *Kappa*, apresentou melhor resultado que o classificador ISODATA.

Outra forma de avaliar as duas classificações consistiu em comparar a área total dos pixels que não coincidiram entre a classificação ISODATA ou a classificação K-MEANS com a Interpretação Visual.

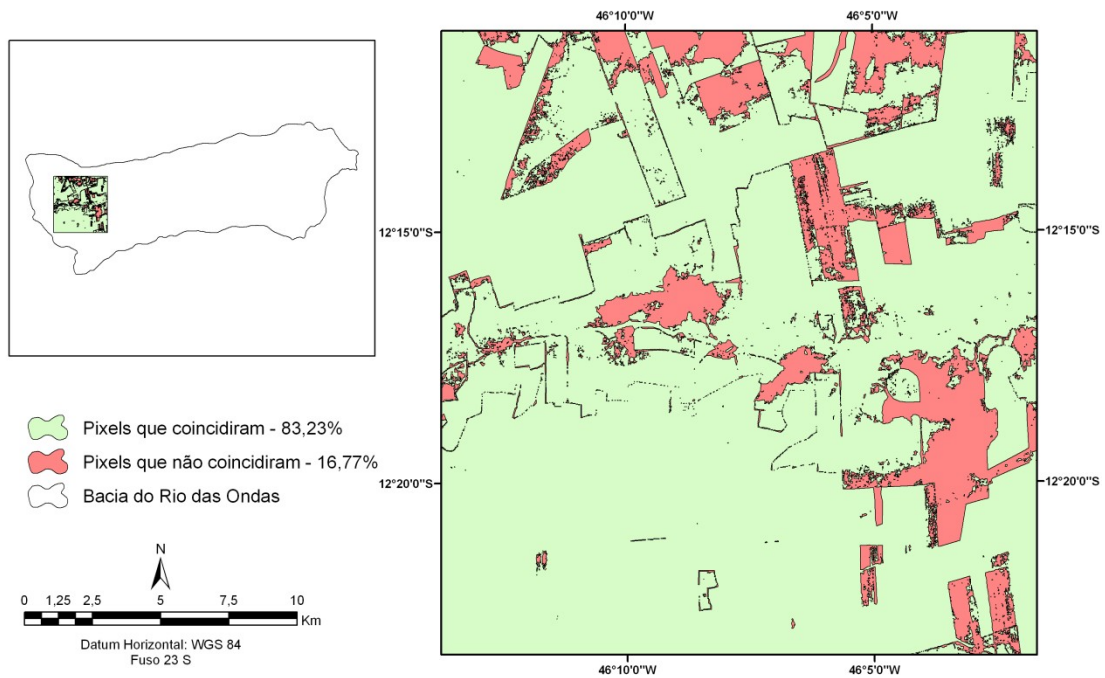
A figura 6 mostra área dos pixels que coincidiram e dos pixels que não coincidiram entre a classificação ISODATA e a Interpretação Visual. A área de pixels que não coincidiram correspondem a aproximadamente 98 Km<sup>2</sup>(16,77%).



**Figura : Comparação entre a Interpretação Visual x classificação ISODATA.**

O classificador K-MEANS demonstrou novamente maior concordância com a Interpretação Visual, pois a área de pixels que não coincidiram correspondem a aproximadamente 83 Km<sup>2</sup> ou 14,23% da área (Figura 7).

### Interpretação Visual x K-MEANS



**Figura : Comparação entre a Interpretação Visual x classificação K-MEANS.**

Os classificadores, de maneira geral, não conseguiram diferenciar com precisão as áreas de vegetação fotossinteticamente ativas assim como as não ativas. Especificamente, no classificador K-MEANS foi possível diferenciar as áreas de solo exposto na Área Natural, enquanto na classe Área Alterada não foi possível fazer a diferenciação de solo exposto e agricultura.

## 5 – CONSIDERAÇÕES FINAIS

A Interpretação Visual da imagem de alta resolução espectral (2,5m) do sensor PRISM/ALOS produziu uma base de dados confiável já que classificou as classes de Área Alterada e Área Natural por meio de conhecimento em fotointerpretação.

A metodologia adotada para avaliar a acurácia dos classificadores automáticos, quando comparados com a Interpretação Visual, atendeu o objetivo proposto. Ao analisar as Matrizes de Confusão, os índices *Kappa* e as áreas de pixels que não coincidiram, foi possível determinar qual classificador tem maior concordância com a base de dados produzidos a partir da Interpretação Visual. Dito isso, conclui-se que o classificador K-MEANS possui melhor acurácia que o classificador ISODATA.

Esse tipo de trabalho é importante porque facilita ao pesquisador escolher, antecipadamente, qual desses métodos se enquadra melhor em pesquisas sobre o uso da terra. Outra vantagem de se utilizar um classificador automático é a rapidez em que são produzidas as informações, mesmo quando a classificação é feita em áreas de grandes dimensões.

O método utilizado para avaliar os classificadores não é a única forma de abordar o tema já que existem outros procedimentos estatísticos e diferentes índices de concordância que podem ser incorporados a esse tipo de estudo.

Esse estudo ficaria mais completo se ao invés de comparar só as duas classes, Área transformada e Área natural, fosse incluído um número maior de classes de uso da terra. Com isso, o uso da terra representaria com mais detalhe a realidade da área de estudo.



## REFERÊNCIA BIBLIOGRÁFICA

ABRAMS, M. 2000. **The Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER): data products for the high spatial resolution imager on NASA's Terra platform.** International Journal of Remote Sensing, 21(5): 847–859.

BAHIA. Secretária de Recursos Hídricos, Saneamento e Habitação. Coordenação de Recursos Hídricos. **Plano Diretor de Recursos Hídricos: bacia do Rio Grande.** Salvador: HIGESA, 1993. 266 p.

BARREIRAS. **Prefeitura Municipal de Barreiras (BA).** Disponível em: <<http://www.barreiras.ba.gov.br/site>>. Acesso em: 25 de abril 2009.

BERNARDES, T.; ALVES, H. M. R.; VIEIRA, T. G. C.; ANDRADE, H. Avaliação da acurácia do mapeamento do uso da terra no complexo Serra Negra, Patrocínio, MG, por interpretação visual e classificação automática de imagens Landsat. In: XIII SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 2007, Florianópolis. **Anais do XIII Simpósio Brasileiro de Sensoriamento Remoto.** Florianópolis, Brasil, 21-26 abril 2007, INPE, p. 5587-5594.

BOLFE, E. L.; FONSECA, E. L.; PEREIRA, R. S.; MADRUGA, P. R. A. Verificação da exatidão em classificação digital de povoamentos florestais em imagem orbital mediante três índices. In: XI Simpósio Brasileiro de Sensoriamento Remoto, 2003, Belo Horizonte. **Anais do XI Simpósio Brasileiro de Sensoriamento Remoto.** Belo Horizonte, Brasil, 05-10 abril 2003, INPE, p. 2671-2677.

BRASIL. Lei n. 9.433, de 8 de janeiro de 1997. Institui a Política Nacional de Recursos Hídricos. **Lex:** SIRVINSKAS, L. P. (Org.). Legislação de Direito Ambiental. 3. ed. São Paulo: Rideel, 2008.

BRITES, R. S.; SOARES, V. P.; RIBEIRO, C. A. A. S. Comparação de Desempenho entre Três Índices de Exatidão Aplicados a Classificações de Imagens Orbitais. In: VIII Simpósio Brasileiro de Sensoriamento Remoto, 1996, Salvador. **Anais do VIII Simpósio Brasileiro de Sensoriamento Remoto.** Salvador, Brasil, 14-19 abril 1996, INPE, p. 813-821.

CHRISTOFOLETTI, A. Morfologia do relevo na média Bacia do rio Corumbataí. In: V SIMPÓSIO DE GEOGRAFIA FÍSICA APLICADA, 1993, São Paulo. **Anais do V Simpósio de Geografia Física Aplicada**. São Paulo: USP, 1993. p. 3011-3022.

COHEN, J. A Coefficient of Agreement for Nominal Scales. **Educational and Psychological Measurement**. v.20, n. 1, p. 37-46, 1960.

CURRAN, P. J. **Principles of Remote Sensing**. London: Longman, 1986.

EMBRAPA. **Monitoramento da Expansão das Áreas Irrigadas na região oeste da Bahia**. Disponível em: < <http://www.bndes.cnpm.embrapa.br>>. Acesso em: 19 de abril 2009.

FIGUEIREDO, S. M. M.; CARVALHO, L. M. T. Análise comparativa entre técnicas de classificação digital em imagens Landsat no mapeamento do uso e ocupação do solo em Capixaba, Acre. In: XIII SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 2007, Florianópolis. **Anais do XIII Simpósio Brasileiro de Sensoriamento Remoto**. Florianópolis, Brasil, 21-26 abril 2007, INPE, p. 6729-6736.

FUJISADA, H.; SAKUMA, F.; ONO, A.; KUDOH, M. 1998. **Design and preflight performance of ASTER instrument protoflight model**. IEEE Transactions on Geoscience and Remote Sensing, 36(4): 1152–1160.

GALVÃO, W. S.; MENESES, P. R. Uso de SIG e do classificador ISODATA na geração do mapa de regiões geoambientais homogêneas na bacia do rio São Francisco, como um produto estratégico de apoio às ações de planejamento de redes fluviométricas. In: XII SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 2005, Goiânia. **Anais XII Simpósio Brasileiro de Sensoriamento Remoto**. Goiânia, Brasil, 16-21 abril 2005, INPE, p. 2177-2184.

GASPAR, M. T. P. **Sistema Aquífero Urucuia: caracterização regional e propostas de gestão**. 2006. 158 f. Tese (doutorado) - Universidade de Brasília.

GÓES, C. A.; FILHO, W. L. M.; QUIRINO, W.; CARVALHO, M. Avaliação da acurácia de classificadores, utilizando técnica de fusão de bandas dos sensores ETM+/LANDSAT-7 e CCD/CBERS-1. In: XIII Simpósio Brasileiro de Sensoriamento Remoto, 2007,

Florianópolis. **Anais do XIII Simpósio Brasileiro de Sensoriamento Remoto**. Florianópolis, Brasil, 21-26 abril 2007, INPE, p. 5815-5821.

IBGE. Banco de Dados. **Mapa de Biomas do Brasil**. Escala 1:5.000.000. Disponível em: <<http://mapas.ibge.gov.br/biomas2/viewer.htm>>. Acesso em: 23 de maio 2009.

\_\_\_\_\_. Banco de Dados. **Cidades@**. Disponível em: <<http://www.ibge.gov.br/cidadesat>>. Acesso em: 19 de abril 2009.

\_\_\_\_\_. **Manual Técnico da Vegetação Brasileira**. Rio de Janeiro: IBGE, 1992.

\_\_\_\_\_. Produtos e Serviços. **ALOS**. Disponível em: <<http://www.ibge.gov.br/alos>>. Acesso em: 25 de maio 2009.

JAXA. Satellites & Spacecraft. Disponível em: <[http://www.jaxa.jp/index\\_e.html](http://www.jaxa.jp/index_e.html)>. Acesso em: 27 de maio 2009.

Jensen, J. R. **Introductory Digital Image Processing**, Prentice-Hall, Englewood Cliffs, New Jersey, p. 379, 1986.

MENDONÇA, L. E. R.; SOARES, V. P.; GLERIANI, J. M.; RIBEIRO, G. A.; RIBEIRO, C. A. A. S. Mapeamento do uso da terra baseado em imagem Ikonos II com a utilização de algoritmos classificadores por pixels e por regiões. In: XIII Simpósio Brasileiro de Sensoriamento Remoto, 2007, Florianópolis. **Anais do XIII Simpósio Brasileiro de Sensoriamento Remoto**. Florianópolis, Brasil, 21-26 abril 2007, INPE, p. 603-610.

MENESES, P. R.; MADEIRA NETTO, J. S (Coord). **Sensoriamento remoto: reflectância dos alvos naturais**. 1. ed. Brasília: Editora UnB, 2001.

MORAES, L. S. **Diagnóstico de uso e ocupação da Bacia do Rio de Ondas**. 2003. 150 f. Dissertação (mestrado) – Universidade Católica de Brasília.

OLIVEIRA, S. N de. **Metodologia para delimitação e análise multivariada de atributos morfométricos de bacia de drenagem usando modelo digital de elevação hidrologicamente corrigido**. 2008. 74 f. Dissertação (mestrado) - Universidade de Brasília.

PROBIO. Projeto de conservação e utilização sustentável da diversidade biológica brasileira. **Levantamento da cobertura vegetal e do uso do solo do bioma Caatinga**. 2004. Disponível em: <<http://mapas.mma.gov.br/i3geo>>. Acesso em: 29 de maio 2009.

PROJETO RADAMBRASIL. **Levantamento de Recursos Naturais**. Folha Brasília SD. 23. Rio de Janeiro, 1983.

ROSA, R. **Introdução ao sensoriamento remoto**. 5. ed. Uberlândia: Edufu, 2003. 238 p.

RICHARDS, J.A. **Remote sensing digital image analysis - an introduction**. 2. ed. Springer-Verlag, Berlin, 1993.

SAMPAIO CEM. **Análise de grupo via programação matemática**. 1990. 95 f. Dissertação (mestrado) - Universidade de Brasília.

SANTOS, M. **Por uma geografia nova: da crítica da geografia a uma geografia crítica**. 6. ed. São Paulo: EDUSP, 2004.

SOARES NETO, J. P. **Avaliação geoambiental da Bacia do Rio das Ondas no oeste da Bahia**. 2005. 233 f. Tese (doutorado) - Universidade de Brasília.

SULSOFT. Guia do ENVI em Português. 2007

TOU, J. T.; GONZALEZ, R. C. **Pattern Recognition Principles**, Addison-Wesley Publishing Company, Reading, Massachusetts. 1974.

WEBER, E.; FONTANA, D. C.; DUCATI, J.; SILVA, T. R.; SARAIVA, T. S.; FREITAS, L. A. S. Comparação entre resultados de classificação de imagens Landsat e CBERS para estimativa de área cultivada com soja no planalto do Rio Grande do Sul. In: XII Simpósio Brasileiro de Sensoriamento Remoto, 2005, Goiânia. **Anais XII Simpósio**

**Brasileiro de Sensoriamento Remoto.** Goiânia, Brasil, 16-21 abril 2005, INPE, p. 321-328.

YAMAGUCHI Y; KAHLE AB; TSU H; KAWAKAMI T; PNIEL M. **Overview of Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER).** IEEE Transactions on Geoscience and Remote Sensing, 36(4): 1062–1071, 1998.