

## Aplicação de diferentes métodos de classificadores em imagens Landsat para mapeamento temático do uso da terra

Vanessa Silva Pugliero<sup>1</sup>  
Victor Danilo Manabe<sup>1</sup>  
Jansle Vieira Rocha<sup>1</sup>

<sup>1</sup> Universidade Estadual de Campinas - UNICAMP/FEAGRI  
Av. Candido Rondon, 501 – 13083-875 - Campinas - SP, Brasil  
vpugliero@gmail.com; {victor.manabe; jansle.rocha}@feagri.unicamp.br

**Abstract:** The knowledge of the land use and occupation has been essential for the agricultural and environmental planning. In this context the images of remote sensing have been solving problems. The aim of this study was to identify the best classification to different types of land use in images Landsat-5/TM of 1999 and Landsat-7/ETM+ of 2011, the both with path/row is 220/75, with 12 years of difference time. The study area is located between Pirassununga-SP and Leme-SP. The image classification methods used were: Binary ecode, Parallelepiped, Maximum likelihood, Minimum distance, Mahalanobis distance and Suport vector machine. Images of CBERS-HRC were used to recognize the urban surface. The most confusion among the classes was among the vegetation: sugarcane, citrus and grass, it is related with similarity in their spectral response. Each classification was particularly satisfactory to a kind of land use in these two images although the Maximum likelihood was more useful in Landsat-5/TM as well as the Suport vector machine was very used in Landsat-7/ETM+. The classification accuracy was evaluated through an error matrix, with kappa coefficient of 0.68 and overall accuracy of 76% to image of 2011 and kappa coefficient of 0.83 and overall accuracy of 87% to image of 1999.

**Palavras-chave:** satellite images classification, land use maps, Landsat images, classificação de imagens de satélite, mapas de uso do solo, imagens Landsat.

### 1. Introdução

Um dos objetivos do sensoriamento remoto é a classificação da imagem, sendo feitas as distinções entre alvos registrados, sendo possível porque os objetos apresentam respostas diferentes ao longo do espectro eletromagnéticos (CRÓSTA, 1992). Os dados de sensoriamento remoto são armazenados em matrizes onde um pixel guarda um valor de energia eletromagnética associado a uma linha e coluna. A mudança de uma banda para outra consiste no valor de brilho, também conhecido por número digital (JENSEN, 2009). Segundo Ponzoni e Shimabukuro (2009) sobre as imagens realizam-se os processamentos digitais de imagens tais como as técnicas de correções atmosféricas e de eliminação de ruídos. Os números digitais devem ser convertidos em parâmetros físicos de radiância ou reflectância aparente. Assim, as bandas de um mesmo sensor podem ser comparadas entre si.

Os sensores de satélite imageadores geralmente, operam com resolução radiométrica de 6 ou 8 bits que definem o intervalo de cinza, entre o preto e o branco que as imagens possuem. A imagem passa pelo aumento de contraste através da expansão no histograma. O contraste de uma imagem é uma medida do espalhamento dos níveis de cinza (CRÓSTA, 1992).

Primeiramente, sobre a imagem de satélite adquirida recomenda-se fazer calibração radiométrica Chander et al (2009) propuseram índices para esta calibração de imagens de satélites, inclusive TM e ETM+, levando em conta a radiância espectral e posteriormente a reflectância no topo da atmosfera para as diferentes bandas dos sensores.

As classificações podem ser supervisionadas ou não supervisionadas. Os métodos supervisionados utilizam algoritmos para determinar pixels característicos de determinada classe. O método do *Parallelepiped* considera uma área em forma de um retângulo, definindo os níveis de cinza máximo e mínimo do conjunto de treinamento. A classificação *Maximum likelihood* é baseada no princípio de que a classificação errada de um pixel particular não tem

mais significado do que a classificação errada de qualquer outro pixel na imagem. Os modelos de *Minimum Distance* são caracterizados pela simetria espectral, as classes podem ser definidas com base em dados estatísticos em cada banda espectral. O classificador da distância de *Mahalanobis* é similar ao da distância mínima e usa estatísticas para cada classe, porém, ele supõe que a covariância das amostras são iguais assim se torna mais rápido que o *Minimum Distance*. O método *Support Vector Machine* usa o ângulo entre as amostras de treinamento no espaço de n-dimensões para determinar os pixels para uma determinada classe (ENVI, 2010).

Sano e Ferreira (2005), indicam os dados de sensoriamento remoto para monitoramento de mudanças de uso do solo ao longo de um período num mesmo local, inserem-se a adoção de metodologias específicas para objetos em estudo. O mapeamento de culturas agrícolas através de sensoriamento remoto possui características particulares correspondentes àquelas das culturas de interesse dentro da região de estudo (EBERHARDT et al., 2011).

Sano e Ferreira (2005), recomendam que áreas de grande heterogeneidade espectral como pivôs centrais que são facilmente visualizadas nas imagens devido à sua forma circular e áreas urbanas que apresentam padrão espectral e geométrico único pela alta reflectância dos telhados de edifícios e residências sejam delimitadas através de vetorização de polígonos.

Fan et. al. (2007) utilizou pontos aleatórios para validar a classificação das imagens Landsat5-TM de 1998 e de 2003 assim, a acurácia dos mapas e de cada tematização foi conferida através do índice Kappa. As imagens de satélites são vantajosas para estudos na área de agricultura, mas cada imagem tem sua vantagem assim como suas limitações (OZDOGAN et al., 2010). Como produto final deste processo será gerada um mapa de uso e cobertura do solo para os anos de 1999 e 2011, que será feito através de métodos de classificação supervisionada de imagens do Landsat 5-TM e Landsat 7-ETM+.

## 2. Metodologia de Trabalho

A região escolhida para este trabalho situa-se entre Leme-SP e Pirassununga-SP como mostrada na figura 1. No local, existe base de produção bem diversificada, tanto em indústria quanto em agricultura e pecuária. A grande maioria da área é ocupada com agricultura com maior predomínio da cultura da cana de açúcar. Possui plantel de bovinos, suínos e aves e ainda produção de algodão, milho, laranja e tomate, plantio de braquiária e eucalipto.

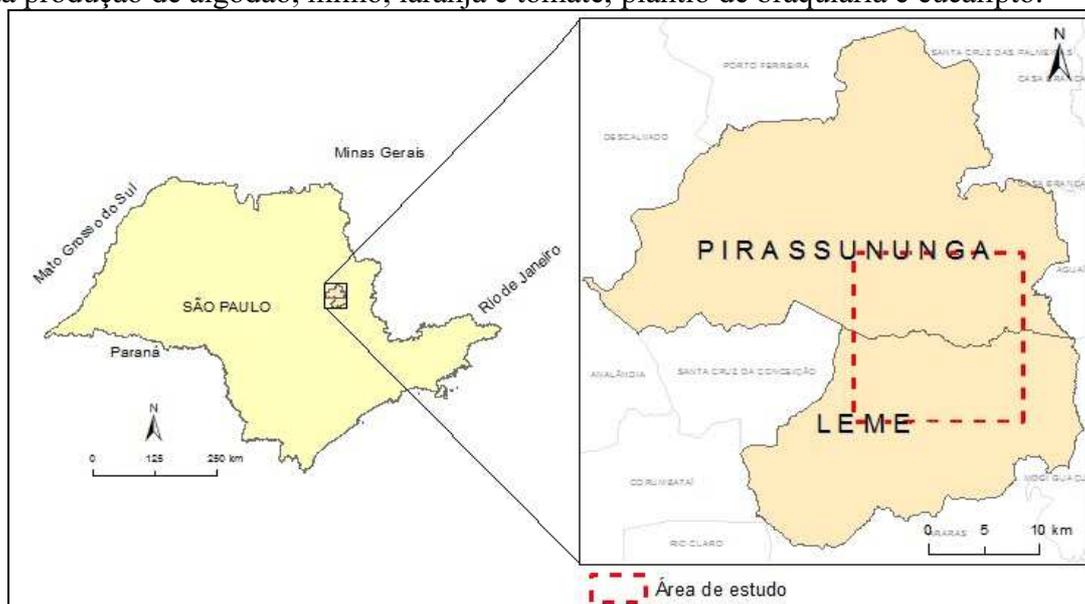


Figura 1. Localização da área de estudo entre os municípios de Pirassununga-SP e Leme-SP.

Neste estudo, duas cenas com aproximadamente 0% de cobertura de nuvem foram obtidas. Ambas do ponto órbita 220/75 sendo uma Landsat TM de 01/07/2011 e outra Landsat ETM+ para o dia 01/07/1999 disponibilizadas pelo INPE.

Primeiramente, foram feitos processamentos dos valores digitais em radiância espectral máxima e mínima aparente do sensor e depois o processamento dos valores de radiância em reflectância aparente no topo da atmosfera para todas as bandas segundo Chandlers et al. (2009).

As bandas (excluindo-se a banda termal) destas cenas foram agrupadas em uma imagem de 6 bandas. A imagem ETM+ teve o registro baseado em uma imagem do mesmo sensor de 01/07/1999 disponibilizada pela USGS, a nova imagem corrigida foi base para o registro na imagem TM, o erro máximo permitido foi menor que 0,5pixel.

Para auxílio na classificação foram adquiridas seis imagens do sensor HRC (resolução espacial de 2,5m) para cobrir a área estudada de 07/07/2008, 15/07/2009 e 01/11/2009. Ainda uma imagem de *Soil Adjust Vegetation Index* (SAVI) (equação 1) foi calculada para diferenciar o solo exposto de culturas rasteiras, com L de 0,5 conforme Huete et.al. (1988).

$$SAVI = [(B4 - B3) / (B4 + B3 + L)] \times (1 + L) \quad (1)$$

A classificação das imagens foram realizadas para os diferentes tipos de uso do solo encontrados nesta área. Para cada tipo de cobertura foram marcados pixels (amostras de treinamento) característicos, e estes serviram como informações de referência para os diversos testes realizados com os classificadores supervisionados utilizados. As legendas foram divididas em vegetação natural, cana de açúcar, citros, palha, água, cultura anual, pastagem, área urbana e solo exposto.

Para a análise de correção dos dados foram gerados 200 pontos aleatórios na área estudada, classificados conforme as máscaras geradas e conferidas para o uso real do local. Utilizou-se o Índice Kappa que é uma medida de concordância usada em escalas nominais que indica o quão legítimas são as interpretações. Também foram trabalhados os erros de omissão e inclusão para cada classe presente nas classificações.

A classificação da área ocorreu em diversas etapas, onde em cada uma delas era definido o melhor classificador para um determinado uso do solo. Os classificadores utilizados para teste estão disponíveis no software ENVI 4.5. Após a identificação da melhor máscara para aquele caso, a área era retirada para a classificação do próximo tipo de cobertura do solo. Este procedimento foi repetido para a obtenção das máscaras de água, palha, solo exposto, mata, cana de açúcar e citros. A máscara de área urbana foi a primeira realizada, porém de forma manual, e retirada sua área daquela classificada a seguir. Para a pastagem foi definido que sua máscara seria a de fundo, ou seja, após a obtenção das máscaras de todos os usos do solo as áreas faltantes de classificação foram consideradas de pastagem.

### 3. Resultados e Discussão

#### 3.1 Landsat 5, sensor TM

No processo de obtenção das máscaras foi necessária a mudança dos histogramas, classificadores e filtros para a realização das séries de testes como apresentado na tabela 1 para cada uso do solo presente na área estudada.

Tabela 1. Caracterização dos classificadores utilizados na imagem TM.

Etapa	Uso do Solo	Classificador	Imagens	Histograma	Filtro
1	Urbano	Manual	-	-	-
2	Água	<i>Binary Ecode</i>	Todas	R: 0,034-0,128 B: 0,110-0,124	-
3	Palha	<i>Parallelepiped</i>	432	B:0,159-0,318	-
4	Solo Exposto	<i>Maximum likelihood</i>	SAVI53	G: 0,058-0,227 B: 0,237-0,328	-
5	Mata	<i>Maximum likelihood</i>	543	R: 0,302-0,393	<i>Median [3x3]</i>
6	Cana	<i>Maximum likelihood</i>	453	G: 0,137-0,381 B: 0,092-0,165	-
7	Citros	<i>Maximum likelihood</i>	453	R: 0,207-0,341 B: 0,092-0,142	-
8	Pasto	Fundo	-	-	-
9	Cultura anual	Manual	-	-	-

Visualmente, o método de *Binary Ecode* utilizado para separar a água das demais classes foi eficiente principalmente para diferenciá-la de solos muito úmidos. Águas com muitos sedimentos puderam ser incorporados na classe de água com base na imagem fusionada.

A mata na composição de imagem RGB543 visualmente se tornou esverdeada e de elevada textura, o filtro fez com o conjunto de pixel se tornassem mais homogêneos e mais fáceis de ser inclusos numa mesma categoria. A modificação do histograma também facilitou a identificação de palha. A composição RGB453 foi útil para classificar citros que adquiriram tonalidade marrom, nesta composição muitos alvos agrícolas foram identificados. Em dois casos houve vetorização da classe: cultura anual (dentro de pivôs agrícolas) e manchas urbanas baseadas na fusão das bandas TM com a banda HRC do CBERS.

A classificação apresentou acurácia de 76% e o índice Kappa de 0,68 conforme tabela 2.

Tabela 2. Erros e índice Kappa.

	Solo	Palha	Cana	Mata	Água	Citros	Pasto	Urbano	Cultura	Total	E. O.	Kappa
Solo E.	53	0	1	0	0	1	1	0	0	56	0,05	
Palha	0	4	0	0	0	0	0	0	0	4	0,00	
Cana	8	0	57	0	0	3	4	0	0	72	0,21	
Mata	0	0	1	9	0	1	3	0	0	14	0,36	
Água	0	0	1	1	1	0	0	0	0	3	0,67	
Citros	1	0	2	4	0	23	0	0	0	30	0,23	
Pasto	0	0	9	0	0	7	2	0	0	18	0,89	
Urbano	0	0	0	0	0	0	0	1	0	1	0,00	
Cultura	0	0	0	0	0	0	0	0	2	2	0,00	
Total	62	4	71	14	1	35	10	1	2	200		0,68
E. I.	0,15	0,00	0,20	0,36	0,00	0,34	0,80	0,00	0,00			

A maioria dos erros de omissões ocorreu para a classe de pasto, água, mata, citros e cana. Os maiores erros de introdução foi para pasto, mata, citros, cana e solo.

Alguns pontos de cana de açúcar foram classificados como solo exposto, mata, água, citros e pasto. Na imagem, há plantios de cana de diversas qualidades em vários estágios. As confusões ocorridas com solo exposto, citros e pasto devem-se à semelhança da resposta espectral de canas com baixas estaturas e da influencia de solos. Apenas um ponto de cana de açúcar foi classificado como mata, este ponto está na transição de uma cultura de cana com uma mata ciliar. O uso do índice de vegetação SAVI, possibilitou menor erro de omissão de

solo exposto, o erro de introdução foi maior devido ao fato destas estarem rentes ao solo desnudo.

O erro de omissão da classe de água está relacionado com cana e mata por conter elevados teores de sedimentos. Estas áreas se apresentam muito úmidas para serem solos e com reflectância baixa no infravermelho próximo para serem vegetações.

A classe de citros teve maior erro de introdução pelas culturas de pasto e cana. Provavelmente, o estágio de desenvolvimento, condições hídricas e presença de solo se assemelharam entre estas classes.

O elevado valor do erro de inclusão e de erro de omissão para pastagem ocorreu por ter sido a máscara de fundo na classificação.

O mapa gerado para as nove classes a partir da imagem TM desta etapa está representado na figura 2.

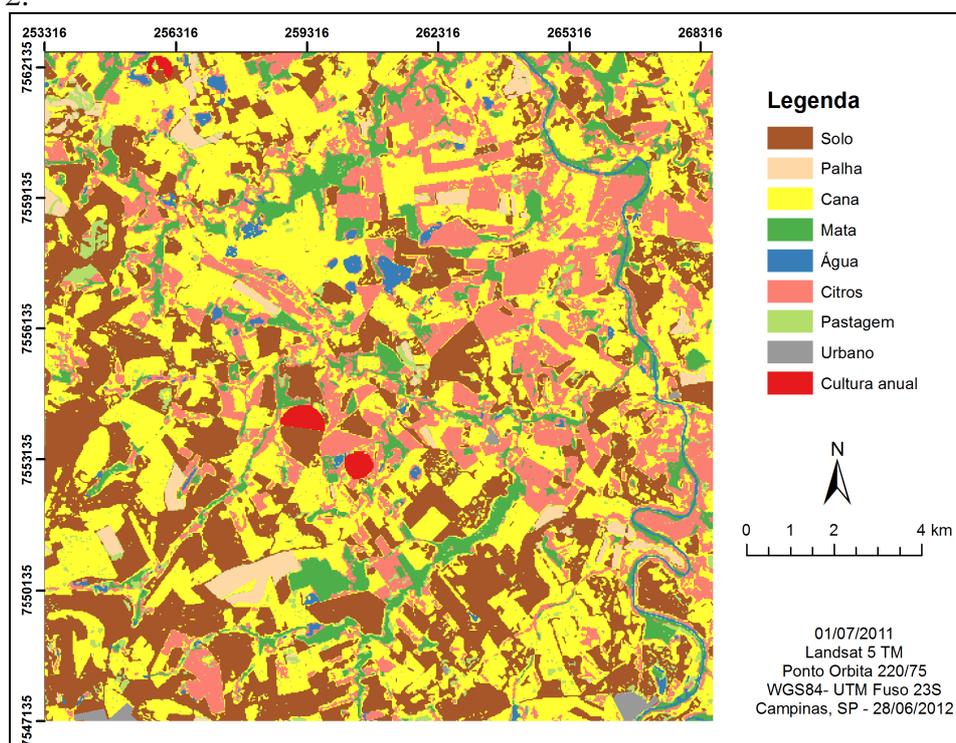


Figura 2. Mapa de uso e ocupação do solo para o ano de 2011.

### 3.2. Landsat 7, sensor ETM+

No processo de obtenção das máscaras não foi necessária a mudança dos histogramas para a realização dos testes, somente foram modificados os classificadores e as bandas utilizadas. A seguir na tabela 3, serão apresentados os classificadores e as bandas para cada uso do solo presente na área estudada, na ordem em que foram trabalhados. A figura 3 mostra o mapa de uso do solo para o ano de 1999, gerado a partir das classificações.

Tabela 3. Caracterização dos classificadores utilizados na imagem ETM+

Etapa	Uso do Solo	Classificador	Imagens
1	Urbano	Manual	-
2	Água	<i>Parallelepiped</i>	456
3	Palha	<i>Support Vector Machine</i>	Todas
4	Solo Exposto	<i>Support Vector Machine</i>	456
5	Mata	<i>Mahalanobis Distance</i>	Todas
6	Cana	<i>Support Vector Machine</i>	345
7	Citros	<i>Minimum Distance</i>	Todas
8	Pasto	Fundo	-

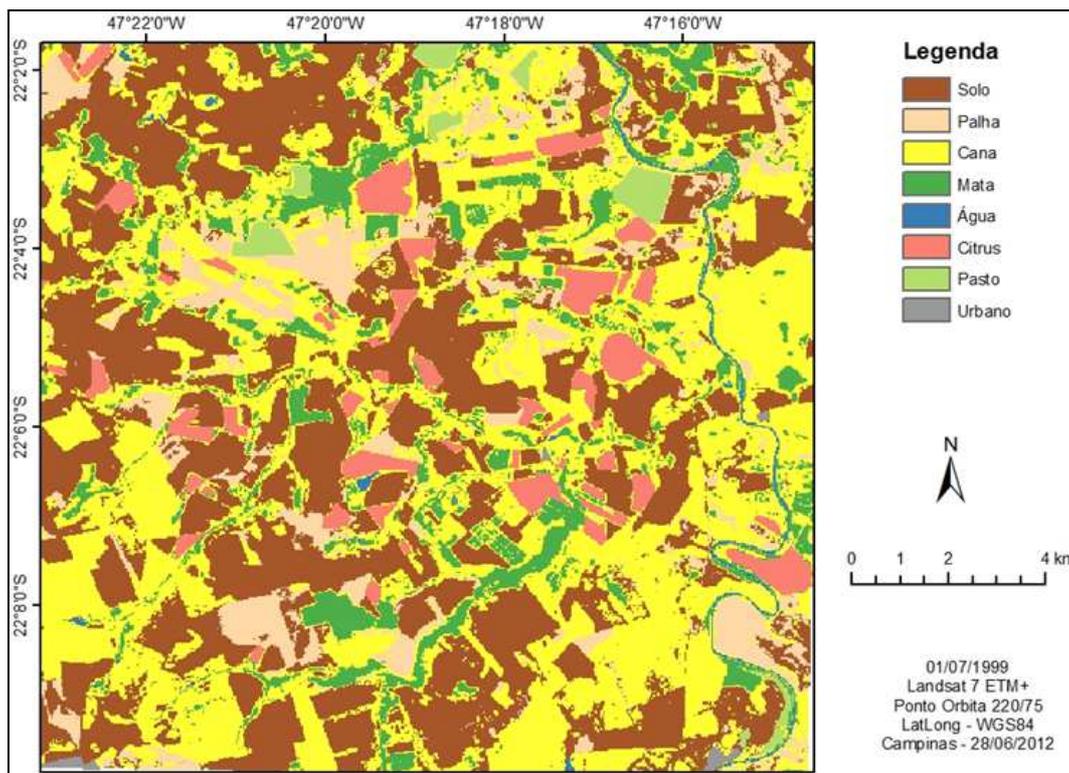


Figura 3. Mapa de uso do solo e ocupação do solo para o ano de 1999.

Para a classificação dos diferentes tipos de uso foi encontrado dificuldade para a geração de máscara naqueles alvos que apresentavam algum tipo de cobertura vegetal. Por isto, foram primeiro classificados e retirados para futura classificação as áreas urbanas, solo exposto e solo coberto de palha. Os alvos citros, mata e cana por apresentarem uma cobertura vegetal necessitaram um maior cuidado na coleta das áreas para a análise e posterior classificação.

Na tabela 4 foi feita a análise dos erros de inclusão e omissão das máscaras e a apresentação do índice Kappa para a classificação da área. A classificação apresentou uma acurácia de 87% e o Kappa de 0,83.

Tabela 4. Erros e índice Kappa.

	Solo	Palha	Cana	Mata	Água	Citros	Pasto	Urbano	Total	E. O.	Kappa
Solo E.	54	1	1	0	0	0	0	0	56	0,04	
Palha	2	18	0	0	0	0	0	0	20	0,10	
Cana	0	0	66	0	0	1	0	0	67	0,01	
Mata	2	0	9	19	0	2	0	0	32	0,41	
Água	0	0	1	0	1	0	0	0	2	0,50	
Citros	0	0	1	2	0	9	0	0	12	0,25	
Pasto	1	0	3	0	0	0	7	0	11	0,36	
Urbano	0	0	0	0	0	0	0	0	0	0,00	
Total	59	19	81	21	1	12	7	0	200		0,83
E. I.	0,08	0,05	0,19	0,10	0,00	0,25	0,00	0,00			

Cana de açúcar apresentou a maior ocorrência de pontos incidentes, desta forma pode-se dizer que é a classe de maior presença na área. Seu erro por omissão foi o mais baixo e o erro por inclusão é o segundo maior, isto indica que a máscara gerada para cana de açúcar superestima a sua área, classificando nele outros usos, porém identifica todas as áreas onde realmente se encontra a cana de açúcar. A segunda classe de maior ocorrência é a de solo exposto. Esta classe apresenta um comportamento radiométrico bem diferenciável das demais

classes. A maior parte dos erros de omissão da classe de mata ocorreu para a classe de cana de açúcar, há dificuldade da diferenciação das classes com cobertura vegetal, pois nesta época da imagem classificada a cana de açúcar que ainda não foi colhida apresenta grandes quantidades de biomassa, apresentando um comportamento espectral muito próxima ao de mata. O mesmo ocorre para as classes de citros e pasto que apresentaram dificuldades para a classificação devido ao comportamento espectrais semelhantes ao de cana de açúcar e mata, e conseqüentemente a isto, um numero maior de erros entre estas quatro classes. O baixo valor do erro de inclusão e valor mais elevado do erro de omissão para pastagem ocorre por ela ser a máscara de fundo na classificação, diminuindo desta sua área.

O solo com cobertura de palha por apresentar uma alta reflectância teve uma diferenciação das demais classes mais facilitada.

As áreas urbanas e água por serem as de menor tamanha em toda área estudada, na amostragem de dados não apresentaram dados suficientes ou não apresentaram dados para a análise.

### 3.3. Evolução da área de estudo em 12 anos (1999 e 2011)

Os classificadores utilizados resultaram em diferentes acuráneas para cada tipo de uso e cobertura da terra para os diferentes anos das imagens. E isto pode variar dependendo da área de estudo.

Consideraram-se *Maximum likelihood* e *Support Vector Machine* para classificar solo exposto, enquanto Silva et.al. (2011), ao utilizar os métodos *Minimum Distance* e *Mahalanobis Distance* obtiveram boa acurácea para solos expostos na imagem trabalhada em outra região. Demarchi et. al. (2011) aponta que o método *Parallelepiped* obteve menor acurácea para a classificação da imagem em geral, porém este mesmo classificador foi eficiente para classes de água na imagem na imagem Landsat7/ETM+ e palha Landsat5/TM.

Através das classificações feitas, aceitando-se as estatísticas de erros e o índice kappa calculados, para a área estudada, foi conferida a variação de cada classe de uso e ocupação do solo. As classes de solo exposto e palha foram somadas à classe de cana de açúcar que no local o solo representa o preparo para o plantio e a palha o resíduo de colheita de um único uso.

Conforme a figura 4 percebem-se o aparecimento da cultura anual em 2011 inexistente em 1999 caracterizado pelos pivôs de irrigação, aumento da área urbana e citricultura e queda nas áreas de mata.

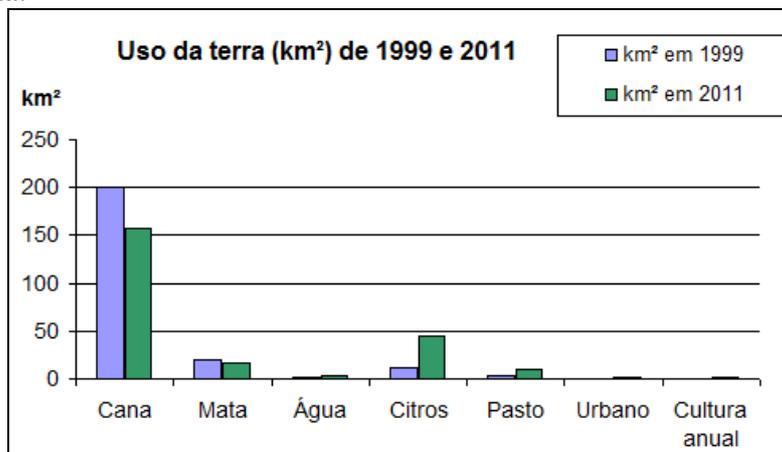


Figura 4: Gráfico de comparação entre áreas dos diversos usos em 1999 e 2011.

Nota-se que a região mantém forte cultivo de cana de açúcar registrado pelos maiores valores de área nos dois anos apresentados e tende a ser produtora agrícola pelo imperceptível acréscimo de urbanização.

#### 4. Conclusão

Os resultados obtidos das classificações supervisionadas testadas são satisfatórios. Não houve uma classificação eficiente para todas as classes de uso e nenhuma coincidente entre as duas imagens para a mesma classe. Foi positivo considerar o teste de todas as possibilidades de classificação supervisionada fornecida pelo software e escolher a melhor bem como alteração no histograma, composição de canais espectrais, filtros e cálculos de índices de vegetação resultando em acurácia de 87% e índice Kappa de 0,83 para 1999 e acurácia de 76% e o índice Kappa de 0,68 para 2011.

Neste trabalho, os métodos mais usuais foram *Support Vector Machine* e o *Maximum likelihood*. Cada conjunto de soluções é eficiente para uma determinada classe e os resultados apresentados para os diferentes classificadores destoam daqueles encontrados em outros trabalhos.

#### 5. Referências bibliográficas

Chander, G., Markham, B., Helder, D. Summary of current radiometric calibration coefficients for Landsat MSS, TM, ETM+, and EO-1 ALI sensors. **Remote Sensing of Environment**. 2009, 893–903; 10p.

Crosta, A.P. **Processamento digital de imagens de sensoriamento remoto**. Campinas: UNICAMP, 1992. 170p.

Demarchi, J.C.; Sartori, A.A.C.; Zimback, C.R. Métodos de classificação de imagens orbitais para o mapeamento do uso do solo: estudo de caso na Sub-Bacia do Córrego Três Barras. In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 15., 2011, Curitiba. Anais... São José dos Campos: INPE, 2011. p. 2644-2651. Disponível em <<http://www.dsr.inpe.br/sbsr2011/files/p0678.pdf>> Acesso em 26 out. 2012.

Eberhardt, I.D.R; Rizzi, R.; Risso, J.; Fernandes, S.L.; Bernardy, R. Mapeamento da área de cana-de-açúcar em Porto Xavier-RS por meio de imagens Landsat. In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 15., 2011, Curitiba. Anais... São José dos Campos: INPE, 2011. p. 0229-0306. Disponível em:< <http://www.dsr.inpe.br/sbsr2011/files/p1123.pdf>> Acesso em 26 out. 2012.

ENVI. **Tutorial em Português 4.5: Classificação**. 32p.

Fan, F.; Weng, Q.; Wang, Y. Land Use and Land Cover Change in Guangzhou, China, from 1998 to 2003, Based on Landsat TM /ETM+ Imagery. **Sensors** 2007. 2007. 1323-1342; 20p.

Huete, A.R. A soil-adjust vegetation index (SAVI). **Remote sensing of Environment**, 25, p.295-309, 1988.

Jensen, J.R. **Sensoriamento remoto do ambiente: uma perspectiva em recursos terrestres**. São José dos Campos: Parêntese, 2009. 604 p.

Ozdogan, M.; Yang, Y.; Allez, G.; Cervantes, C. Remote Sensing of Irrigated Agriculture: Opportunities and Challenges. **Remote Sensing**. 2010, 2274-2304.

Ponzoni, F.J., Shimabukuro, Y.E. **Sensoriamento remoto no estudo da vegetação**. São José dos Campos, SP: Parêntese, 2009. 127p.

Sano, E. E., Ferreira, L. G. Monitoramento semidetalhado (escala de 1:250.000) de ocupação de solos do Cerrado: considerações e proposta metodológica. In: Simpósio Brasileiro de Sensoriamento Remoto (SBSR), 12., 2005, Goiânia. Anais... São José dos Campos: INPE, 2005. Artigos, p. 3309-3316. Disponível em <[http://www.agencia.cnptia.embrapa.br/recursos/Monitoramento\\_OcupacaoID-8tSbyl8QDN.pdf](http://www.agencia.cnptia.embrapa.br/recursos/Monitoramento_OcupacaoID-8tSbyl8QDN.pdf)> Acesso em 20 jun. 2012.

Silva, A.M., Mello, J.R.B., Boschi, R.S. Avaliação de classificadores para o mapeamento de uso da terra. In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 15., 2011, Curitiba. Anais... São José dos Campos: INPE, 2011. p. 7324-7331. Disponível em <<http://www.dsr.inpe.br/sbsr2011/files/p1090.pdf>> Acesso em 22 out. 2012.