



## ANALISE DE CLASSIFICADORES DE IMAGEM NO SOFTWARE MULTSPEC

ANALYSIS OF IMAGE CLASSIFIERS IN MULTSPEC SOFTWARE

Vinícius Fernandes Alves<sup>1</sup>

### RESUMO

Para uma melhor análise do uso e ocupação do solo pode ser utilizado de técnicas de classificação de imagens de satélite, neste trabalho serão apresentadas e discutidas algumas. No intuito de eleger um melhor modelo foram comparados diferentes índices de classificação e objetivos para que possível atingir os resultados esperados. Foi analisada uma imagem do município de São José dos Campos- SP, escolhido por possuir, mancha urbana, vegetação, agricultura e corpos hídricos bem definidos. Neste trabalho foi escolhido o classificado por máxima verossimilhança, que apresentou uma melhor discrepância entre os parâmetros desejados, podendo assim produzir uma imagem mais nítida da região.

**PALAVRAS-CHAVE:** Classificação de Imagem; MultSpec; Máxima Verossimilhança.

### ABSTRACT

For a better analysis of the use and occupation of the land, techniques of classification of satellite images can be used, in this work some will be presented and discussed. In order to choose a better model, different classification indices and objectives were compared in order to achieve the expected results. An image of the municipality of São José dos Campos-SP was analyzed, chosen for having an urban area, vegetation, agriculture and well-defined water bodies. In this work, the one classified by maximum likelihood was chosen, which presented a better discrepancy between the desired parameters, thus being able to produce a clearer image of the region.

---

<sup>1</sup> Licenciado em Geografia na Universidade Federal de Uberlândia (UFU). Bolsista do Programa de Educação Tutorial (PET/MEC) do Instituto de Geografia da UFU. E-mail: [feralvesvini@gmail.com](mailto:feralvesvini@gmail.com)

**KEYWORDS:** Image Classification; MultiSpec; Maximum Likelihood.

## INTRODUÇÃO

O trabalho a seguir trata sobre o tema da seguinte maneira: a classificação de uma imagem em quatro classificadores diferentes, são eles: Máxima verossimilhança, Mínima distância, Paralelepípedo e Distância de Mahalanobis, cada classificador tem suas características e especificidades que serão expostas a seguir.

O classificador por máxima verossimilhança considera a ponderação das distâncias entre as médias dos valores dos pixels das classes, utilizando parâmetros estatísticos, os pixels nessa classificação são delimitados no espaço de uma elipse. Assume que todas as bandas têm distribuição normal e calcula a probabilidade de um dado pixel pertencer a uma classe específica. Ou seja, nesta classificação de máxima verossimilhança cada pixel é mandado à classe que tem a mais alta probabilidade de pertencer. Isso significa que para um pixel na posição  $x$  do espaço multiespectral, um conjunto de probabilidades são computadas e que dão as possibilidades relativas do pixel pertencer a cada classe disponível.

Já a segunda classificação utilizada, a Mínima distância, que seu modelo de trabalho para a obtenção de imagens é diferente da anterior, ela consiste em calcular a medida de distância Euclidiana. Cada pixel será incorporado a um agrupamento através da análise da medida de similaridade de distância Euclidiana. A vantagem dessa classificação é que todos os pixels encontrarão uma média à qual estarão espectralmente mais próximos, e assim, não existirão pixels não-classificados.

A terceira utilizada foi o Paralelepípedo o método que tem como vantagem de ser rápido e simples, uma vez que os valores dos pixels são comparados aos limites que permanecem constantes para cada banda. É frequentemente usado como uma primeira exploração de classificação. Como é um método que não depende de distribuições normais, uma de suas desvantagens é que os paralelepípedos possuem "cantos" e, sendo assim, um pixel pode ser classificado mesmo estando distante espectralmente da média da classe.

O último classificador utilizado foi o de Distância de Mahalanobis na qual a medida da distância do pixel na posição  $x$  do espaço multidimensional ao centro da classe, dividida pelo comprimento do elipsóide na direção de  $x$ . Isso tem a propriedade de minimizar a distância do ponto ao centro de média.

## DESENVOLVIMENTO

### **Objetivos:**

O objetivo geral desse trabalho é realizar o processo de classificação de imagens em um determinado software (ex. multiespec), utilizando quatro classificadores diferentes, e analisar a qualidade dos mapas temáticos produzido

Os objetivos específicos são: Compreender o funcionamento dos quatro classificadores supervisionados; Treinar o processo de seleção de amostras representativas e adequadas em imagens digitais; Interpretar os resultados obtidos com os diferentes classificadores, usando diferentes índices (Kappa, Global, omissão e comissão, presentes na matriz confusão).

### **Metodologia:**

Para a execução do trabalho foi selecionada uma imagem do satélite CBERS 4 com a resolução PAN 10, satélite cujo foi predeterminado pelo professor, a região que a imagem representava foi de livre escolha deste que cumprisse todos os requisitos, que eram: Ter água, zona urbana, vegetação, solo e cultivo na imagem, deste modo, a imagem escolhida para a realização do trabalho foi da cidade de São José dos Campos-SP, localizada no Vale do Paraíba paulista margeada pelo Rio Paraíba do Sul.

Foi utilizado o aplicativo MultSpec para a realização das etapas de classificação do trabalho, juntamente com as orientações do professor. A partir do recorte da imagem foram selecionados campos para a sua classificação. Foram selecionados os seguintes alguns campos dentro da imagem sendo alguns como testes e outros como treinamentos, essa divisão foram feita para que se obtivesse um resultado mais preciso e real da classificação cada campo, teste ou treinamento forma devidos em: Água, Área Urbana, Vegetação, Solo, Cultivo e Área Verde.

Imagem 1: Recorte da cidade de São José dos Campos



Fonte: Multspec 2018

Uma legenda foi criada de modo que facilite a compreensão das imagens que serão expostas a seguir;

	Água
	Vegetação
	Área Urbana
	solo
	cultivo
	área verde

Para que pudesse descobrir qual classificação obteve um melhor imagem resultado, foi utilizado de dois diferentes métodos, o método visual, que consiste em observar a imagem e retirar as conclusões e o método matemático que é dado por porcentagens de precisão e pelo índice Kappa, que segundo Baltar e Okano “é uma medida de concordância Interobservador e mede o grau de concordância além do que seria esperado tão somente pelo acaso. Esta medida de concordância tem como valor máximo o 1, onde este valor 1 representa total concordância e os valores próximos e até abaixo de 0, indicam nenhuma concordância, ou a concordância foi exatamente a esperada pelo acaso”.

A tabela a seguir foi executada por Landa e Koch.

Valores de Kappa	Interpretação
<0	No agreement
0-0.19	Poor agreement
0.20-0.39	Fair agreement
0.40-0.59	Moderate agreement
0.60-0.79	Substantial agreement
0.80-1.00	Almost perfect agreement

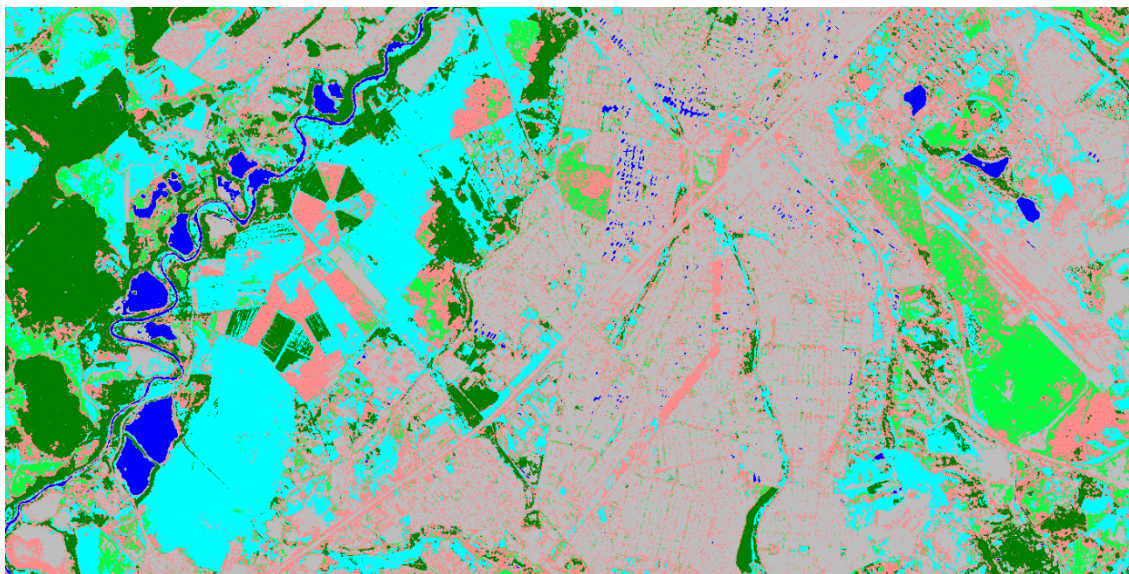
for categorical data. *Biometrics* 1977; 33: 159-174 orial de geoprocessamento SPRING. 2008.

Outro índice utilizado será o índice global, que representa a porcentagem dos pixels classificados corretamente, pelo número total de pixels da imagem, ainda baseado no número de pixels presentes na imagem, é possível extrair mais duas informações, que são; a omissão e a comissão de cada classe. A omissão é o valor em porcentagem de pixels classificados erroneamente, ou seja, nas classes em que eles não pertencem. Já a comissão é o oposto, ou seja, a porcentagem de pixels que foram classificados corretamente em suas respectivas classes. Representado na última linha da matriz de confusão.

### **Desenvolvimento:**

A primeira classificação utilizada foi a Máxima verossimilhança que visualmente apresenta o melhor resultado, pois nele podemos perceber melhor a divisão entre as classes, além de delimitar muito bem a parte de imagem que corre o rio e a área urbana do Município, além de conseguir separar área verde de cultivo e de vegetação, e ser possível observar o solo na área urbana como pode ser visto na imagem a seguir:

Imagem 2: Classificação Máxima verossimilhança.



Neste momento será analisado os resultados matemáticos obtidos por essa classificação. Primeiramente o índice Kappa dessa classificação foi de 82%, que resulta na classificação de amostra perfeita. Além de ter tido seu índice Global de 86,2%. A tabela a seguir mostra a matriz de confusão dessa classificação que contém a porcentagem de precisão de cada classe selecionada e o número de amostras de cada uma encontradas em suas respectivas classes e nas demais que foram utilizadas.

Nome	Número	(%)	Samples	água	Vegetação	Área Urbana	Solo	cultivo	área verde
Água	1	100.0	66	66	0	0	0	0	0
Vegetação	2	99.9	1480	0	1479	0	1	0	0
Área Urbana	3	86.0	333	3	0	2863	405	5	54
Solo	4	92.7	947	0	0	69	878	0	0
Cultivo	5	75.3	1002	0	244	3	0	755	0
Área verde	6	77.8	1868	0	5	30	379	0	1454
TOTAL			8693	69	1728	2965	1663	760	1508
Comissão(%)*			95.7	85.6	96.6	52.8	99.3	96.4	

Tabela 1: Matriz de confusão máxima verossimilhança

A partir da análise da tabela é perceptível que a porcentagem de precisão das amostras é alta, visto que o menor valor encontrado, foi o do Cultivo com 75,3%. É necessário ressaltar a precisão da água com 100% e da vegetação com 99,9%. Devido a

presença da omissão, porcentagem da amostra que fica de fora da classe, que alguns dos valores não alcançaram os 100%, o cultivo como citado anteriormente teve a maior omissão registrada com 14,7%, seguido pela Área Verde com 22,2% e depois a área Urbana com 14%. Os com menores omissões são a Água com nenhuma omissão a vegetação com 0,1% de omissão e o solo com 7,3% de omissão.

Esta tabela também é interessante, não só por trazer a porcentagem de precisão, como também por trazer o valor de confusão da imagem. Visivelmente nota-se que a presença de 405 mostras de solo no campo de área urbana é a maior confusão dessa classificação, mas vale relevar também a grande presença de amostras de vegetação na área de cultivo.

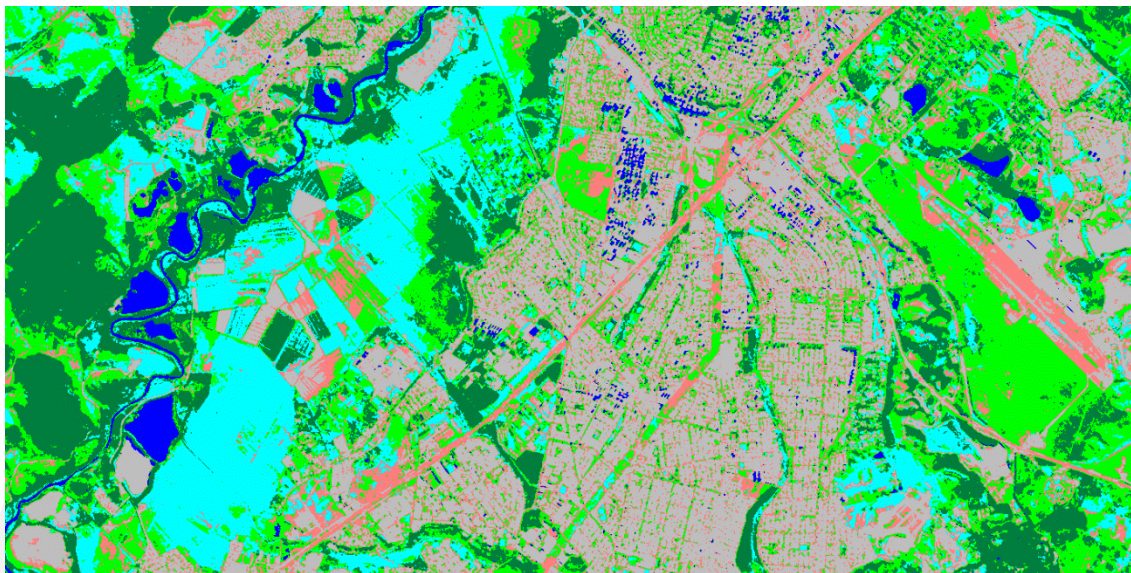
Class	Samples	Porcentagem	Area (Hectares)
1 água	19.275	1.9	192,750
2 Vegetação	123.749	12.4	1.237,490
3 Área Urbana	367.303	36.7	3.673,030
4 Solo	221.803	22.1	2.218,030
5 Cultivo	196.621	19.6	1.966,210
6 Área Verde	73.066	7.3	730,660
Total	1.001.817	100.0	10.018,170

Tabela 2: Distribuição das classes por área máxima verossimilhança

Esta outra interessante tabela mostra a distribuição das classes por área na imagem, ou seja, o número de hectares que cada classe ocupa, sua porcentagem perante a imagem total e o número de amostras que cada classe ocupa. Na imagem selecionada pode perceber que área urbana se sobressai com mais de 3.500 hectares e representando cerca de 36,7 da imagem, outro destaque é região m que tem a presença de água que significa menos de 2% da imagem totalizando 192,75 hectares.

A segunda classificação utilizada foi a do classificador denominado Mínima distância, a imagem produzida por esse classificador também tem uma boa resolução visual porém um pouco pior que a anterior, principalmente na área urbana, onde não é mais possível ver a diferenciação das ruas com nitidez, e a presença do solo diminui além da menor distinção entre áreas verdes e vegetação.

Imagem 3: Classificação Mínima Distância



A matriz de confusão dessa imagem, é a seguinte:

Nome	Número	(%)	Samples	água	Vegetação	Área Urbana	Solo	cultivo	área verde
Água	1	100.0	66	66	0	0	0	0	0
Vegetação	2	100.0	1480	0	1480	0	0	0	0
Área Urbana	3	75.1	3330	9	17	2501	506	1	296
Solo	4	40.5	947	0	0	562	384	0	1
Cultivo	5	81.3	1002	0	187	0	0	815	0
Área verde	6	91.2	1868	0	165	0	0	0	1703
TOTAL			8693	75	1849	3063	890	816	2000
Comissão (%)*				88.0	80.0	81.7	43.1	99.9	85.2

Tabela 3: Matriz de confusão Mínima distância

A partir da análise dessa tabela podemos ver a mudança da precisão em relação ao classificador anterior, principalmente nas classes de Área Urbana e a de solo, na qual ambas entraram em confusão entre si, provando essa confusão cerca de 60% da classe solo estava classificada como Área Urbana, essas duas também tem as maiores omissões visíveis na classe, a de Solo com 59,5% e a da Área Urbana com 24,9%. A água manteve sua boa precisão juntamente com a Vegetação, as duas com omissão de 0%. Cultivo e Área Verde tiveram omissões medianas de 18,7% e 8,85 respectivamente

O índice Kappa dessa imagem é de 73,5%, que a qualifica como uma imagem muito boa, seu índice Global foi de 79,9%, sendo assim seus índices inferiores a



classificação anterior mas não a impedindo de ser uma imagem de boa qualidade visual, que foi comprovada matematicamente a partir desses índices e da tabela.

A tabela a seguir mostra a divisão por área da imagem:

Class	Samples	Porcentagem	Area (Hectares)	
1	água	26.551	2.7	265,510
2	Vegetação	182.832	18.3	1.828,320
3	Área Urbana	246.599	24.6	2.465,990
4	Solo	136.736	13.6	1.367,360
5	cultivo	165.058	16.5	1.650,580
6	área verde	244.041	24.4	2.440,410
	Total	1.001.817	100.0	10.018,170

Tabela 4: Divisão de classes por área, Mínima Distância

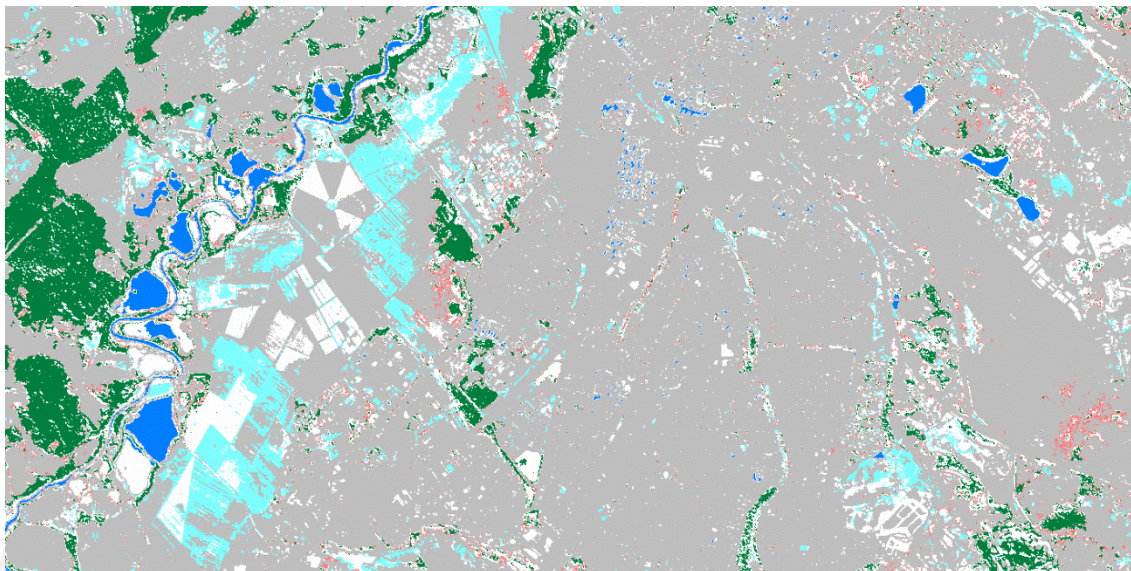
A área da água aumentou em relação à anterior, cerca de 70 ha, ainda assim a área urbana ficou com a maior porcentagem 24,65 de área, porém a vantagem que ela levava sobre as demais na classificação anterior, diminuiu relativamente, isso devido principalmente ao aumento da omissão de sua amostra. O processo inverso ocorreu com a amostras de Área verde, com a diminuição da sua taxa de omissão, ela teve um aumento em área classificada.

A terceira imagem obtida para a realização do trabalho foi dada a partir do classificador Paralelepípedo. A partir de sua análise visual podemos perceber que não foi classificado nenhum pixel da classe Área Verde, isso devida as características dessa classificação, na qual apresenta muitas sobreposições entre as classes além de áreas que não conseguiram ser classificadas ficando em branco na imagem.

Seu índice Kappa obteve o valor de 40,5%, que segundo a tabela de classificação do índice no início do trabalho é considerada uma imagem de moderada qualidade visual, seu índice global foi de 58,2%, deste modo, com valores obtidos menores que os das classificações anteriores, e assim justificando uma imagem de menor qualidade de classificação.

Nota-se que a classe Área Urbana se sobressaiu em relação as demais, principalmente sobre áreas de vegetação e cultivo. Essa anormalidade visível na imagem também foi provada pela matriz de confusão que vem a seguir:

Imagem 4: Classificação Paralelepípedo



Nota-se que a classe Área Urbana se sobressaiu em relação às demais, principalmente sobre áreas de vegetação e cultivo. Essa anormalidade visível na imagem também foi provada pela matriz de confusão que vem a seguir:

Nome	Número	(%)	Samples	água	Vegetação	Área Urbana	Solo	cultivo	área verde	Em branco
Água	1	100.0	66	66	0	0	0	0	0	0
Vegetação	2	83.4	1480	0	1234	0	0	0	0	246
Área Urbana	3	99.2	3330	0	1	3303	10	0	0	16
Solo	4	0.0	947	0	0	947	0	0	0	0
Cultivo	5	45.7	1002	0	0	40	0	458	0	504
Área verde	6	0.0	1868	0	8	1745	88	0	0	27
TOTAL			8693	66	1243	6035	98	458	0	793
Comissão (%)*				100.0	99.3	54.7	0.0	100.0	0.0	

Tabela 5: Matriz de confusão Paralelepípedo

Como dito anteriormente e percebido visualmente, constata-se zerado os valores de classificação da Área Verde, e que seus pixels foram distribuídos entre outras classes. Além do baixíssimo valor do Solo, tendo somente sua maioria de pixels classificados como área urbana e com um valor de omissão de 100% assim como a área verde, logo em seguida com 44,3% de omissão vem a classe cultivo, cujo grande parte de seus pixels não foram classificados ficando em Branco. Com as menores omissões tem as classes, Água, Área Urbana e vegetação, como 0%; 0,8%;16,6% cada uma respectivamente.

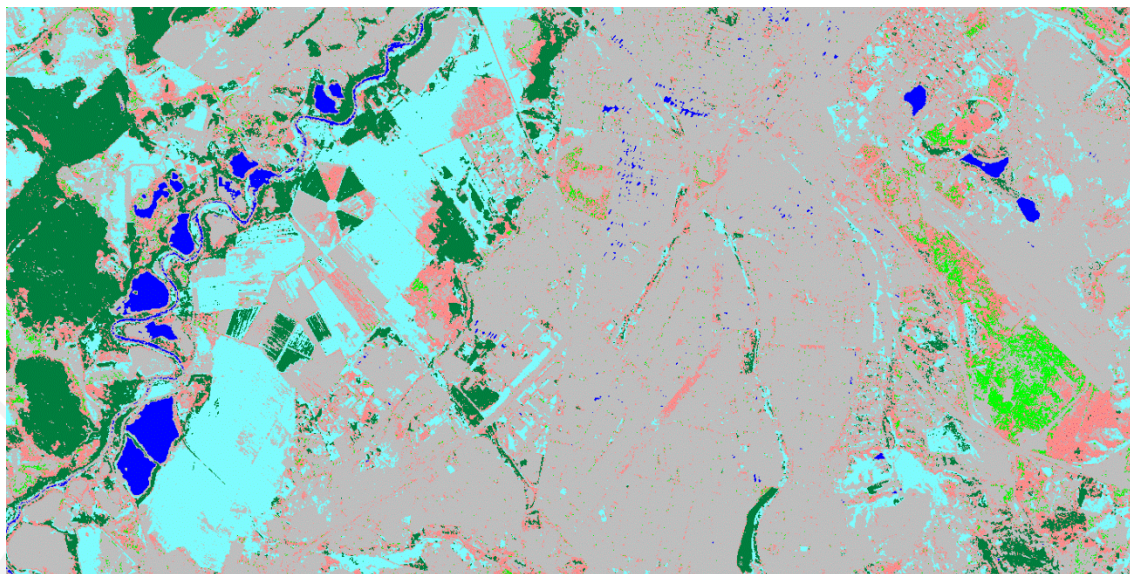
Class	Samples	Percentage	Area (Hectares)
1 água	15.457	1.5	154,570
2 Vegetação	84.244	8.4	842,440
3 Área Urbana	689.224	68.8	6.892,240
4 Solo	19.615	2.0	196,150
5 cultivo	52.668	5.3	526,680
6 área verde	0	0.0	0,000
Threshold	140.609	14.0	1.406,090
Total	1.001.817	100.0	10.018,170

Tabela 6: Classificação por área Paralelepípedo

Na tabela acima podemos comprovar com números um fato visualmente percebido, que é a aparição de pixels em brancos, representando a segunda maior parcela da imagem com 1.406,090 hectares, ou seja, 14%, ficando atrás somente da classe Área Urbana que na imagem é predominantemente visível com 68,8% da área da imagem total, além de comprovar o que foi dito anteriormente sobre a não aparição da classe Área Verde na classificação.

A última imagem solicitada, foi obtida a partir do uso do classificador Distância de Mahalanobis, o índice Kappa dessa classificação ficou em 59.8% sendo também classificada como moderada, o valor de seu índice global foi de 71,4%, desta maneira sendo melhor que a classificação anterior, porém pior que as demais. Em seguida pode perceber a influência desses valores obtidos, a partir da análise visual da imagem a seguir:

Imagem 5: Distância de Mahalanobis



Todas as classes foram representadas na imagem, não deixando dessa maneira, pixels em branco, porém podemos ver novamente a Área Urbana se sobressaindo sobre as demais, assim como a área de cultivos, além da visível diminuição. A Água continua bem delimitada quase não sofrendo distorções ou delimitações.

Nome	Número	(%)	Samples	água	Vegetação	Área Urbana	Solo	cultivo	área verde
Água	1	100.0	66	66	0	0	0	0	0
Vegetação	2	94.2	1480	0	1394	75	11	0	0
Área Urbana	3	98.0	3330	0	0	3263	56	3	8
Solo	4	1.7	947	0	0	931	16	0	0
Cultivo	5	73.7	1002	0	243	21	0	738	0
Área verde	6	39.0	1868	0	0	731	409	0	728
TOTAL			8693	66	1637	5021	492	741	736
Comissão (%)*				100.0	85.2	65.0	3.3	99.6	98.9

Tabela 7: Matriz de confusão Distância de Mahalanobis

A partir da análise da tabela podemos comprovar com valores o que foi visto na imagem, por exemplo, a imagem retratou muito pouco a presença de pixels da classe solo, que foi comprovado pela matriz, representando 98,3% de omissão, outro grande valor de omissão observado é o da Área verde com 61%. Classes como a água com 0%, a vegetação com 5,8% e o cultivo com 26,3% de omissão. Além de notar que a maior confusão mostrada pela tabela é do solo com a Área Urbana.

A tabela a seguir apresenta a relação das classes com a área de cada uma.

Class	Samples	Porcentagem	Area (Hectares)	
1	água	16.408	1.6	164,080
2	Vegetação	103.831	10.4	1.038,310
3	Área Urbana	603.564	60.2	6.035,640
4	Solo	102.983	10.3	1.029,830
5	cultivo	156.416	15.6	1.564,160
6	área verde	18.615	1.9	186,150
	Total	1.001.817	100.0	10.018,170

Tabela 8: Classificação por área Distância de Mahalanobis

A partir da análise dessa tabela, podemos tirar de conclusão que novamente a área urbana da imagem se sobressai com 60,2% da área total, já a classe de solos juntamente com a da água são as que menos aparecem nessa classificação, com menos de 2% cada uma, esse alto valor de Área Urbana pode ser explicado pela classe ter agregado como sua classificação os pixels de solo.

## CONSIDERAÇÕES FINAIS

A partir das aulas teóricas juntamente com o auxílio do professor que serviram como guia e facilitaram a produção deste trabalho prático no qual trouxe para os alunos uma nova experiência na área da Geografia, a classificação de imagens dentro da disciplina de Sensoriamento Remoto.

Baseado na metodologia recomendada os resultados obtidos foram os esperados para cada classificação, o classificador que obteve os melhores resultados, tanto visual, quanto matemático foi o de Máxima verossimilhança, visualmente pois ele representou melhor as classes pedidas, com uma boa resolução, e de modo que pudéssemos distinguir bem as classes, foi a classificação que mais se aproximou do real. Quanta a de pior resultado foi a de Paralelepípedo, também por ser um modelo mais simples e mais rápido deste modo mais vulnerável a possíveis erros de classificação.

Essa parte da Geografia deve ser bastante explorada pelos geógrafos, pois além de ser extremamente interessante e útil para o monitoramento de regiões classificação de áreas entre outros usos, é mais uma oportunidade no mercado de trabalho para o profissional em Geografia e desta maneira ela deve ser aprendida e executada corretamente.

## REFERÊNCIAS

INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS (INPE) - Manuais: tut Landis JR, Koch GG. **The measurement of observer agreement**

for categorical data. **Biometrics orial de geoprocessamento** 1977; 33: 159-174 SPRING. 2008.

BALTAR, Valéria Troncoso; OKANO, Valdir. **Análise de Concordância - Kappa**. Disponível em: <<http://www.lee.dante.br/pesquisa/kappa/>>. Acesso em: 04 jul. 2018.

MENESES, Paulo Roberto; ALMEIDA, Tati de. **INTRODUÇÃO AO PROCESSAMENTO DE IMAGENS DE SENSORIAMENTO REMOTO**. Brasília: Unb, 2012.