

Aplicação da análise de componentes principais em dados extraídos automaticamente de imagens de satélite landsat 5 TM

Carla Eva Prichoa¹
Selma Regina Aranha Ribeiro²

¹Universidad Autónoma de Madrid
Ciudad Universitaria de Cantoblanco
C/ Francisco Tomás y Valiente, 1
28049 Madrid-España
cprichoa@yahoo.com.br

²Universidade Estadual de Ponta Grossa
Av. General Carlos Cavalcanti, 4748 – Cep: 84030-900
Ponta Grossa/Paraná – Brasil
selmar.aranha@gmail.com

Abstract. This paper is about a technique which generally reduce the dimensionality of data generated automatically, enabling analyzes with more consistent results and optimizing the validation in the field. A Principal Component Analysis was used in this study to separate and reduce dimensionality of the reflectance values of the targets or descriptors of spectral, spatial and textural extracted from Landsat 5 TM image. In this study, are used images of two landscape units located the hydrographic basin of Pitangui River in the city of Ponta Grossa / PR in order to check the variables with the highest coefficients correlation, since those having higher percentage denote greater similarities with the study units can be made a recognition characteristics as physical forms and locals. The descriptors were coming from Landsat 5 TM in the compositions and spectral R5G4B3 R5G4B3 with insertion MDT equivalent of band 4. The dimensionality of the data was reduced from 39 (R5G4B3 composition) and 35 (composition R5G4B3 and band 4 MDT) for 17 variables or descriptors, which quantitatively characterize the landscape units in this study. Of the 17 variables final, 6 are mainly characterized by compact shapes, solid, convex among others, 9 are represented by digital values and 2 are represented by means of textures.

Key-words: remote sensing, automatic atraction, reducing dimensionality, correlation, landscape units, sensoriamento remoto, extração automática, redução de dimensionalidade, correlação, unidades de paisagem.

1. Introdução

O objetivo deste trabalho consistiu na aplicação da Análise de Componentes Principais em dados espaciais, espectrais e de textura extraídos de imagem Landsat 5TM de duas unidades de paisagem mediante Análise Orientada a Objetos, com o intuito de reduzir a dimensionalidade, filtrando aqueles descritores ou variáveis de maior correlação com as unidades de estudo.

A Análise de Componentes Principais (ACP) é uma técnica matemática da análise multivariada, que possibilita buscas com um grande número de dados disponíveis. (BOUROCHE; SAPORTA, 1982). Permite a identificação das medidas responsáveis pelas maiores variações entre os resultados, sem perdas significativas de informações. Além disso, transforma um conjunto original de variáveis em outro conjunto: os Componentes Principais (CP) de dimensões equivalentes. Essa transformação, em outro conjunto de variáveis, ocorre com a menor perda de informação possível, sendo que esta também busca eliminar algumas variáveis originais que possua pouca informação. Essa redução de variáveis só será possível se as p variáveis iniciais não forem independentes e possuírem coeficientes de correlação não-nulos.

A ACP pode ser resumida em três passos RICHARDS (1993): (a) derivação da matriz de correlação ou de variância covariância, (b) cálculo dos autovetores e autovalores, e (c) transformação linear do conjunto de dados. Esta transformação tem como base a rotação do

espaço de atributos na direção dos autovetores com o intuito de tornar ortogonal o conjunto de dados, promovendo a união das informações de maior correlação nas primeiras componentes (MATHER, 1999). A ACP reduz a dimensionalidade dos dados, concentrando as informações significativas nas primeiras componentes geradas pela transformação (SCHOWENGERDT, 1997). O objetivo da Análise de Componentes Principais é abordar aspectos como a geração, a seleção e a interpretação das componentes investigadas. É possível também determinar as variáveis de maior influência na formação de cada componente, que serão utilizadas para estudos futuros.

Para a determinação das componentes principais, é necessário calcular a matriz de variância-covariância (Σ), ou a matriz de correlação (R), encontrar os autovalores e os autovetores e, por fim, escrever as combinações lineares, que serão as novas variáveis, denominadas de componentes principais, sendo que cada componente principal é uma combinação linear de todas as variáveis originais, independentes entre si e estimadas com o propósito de reter, em ordem de estimação e em termos da variação total, contida nos dados iniciais, (REGAZZI, 2001).

Neste sentido, as unidades de paisagem delimitadas para esta pesquisa abrangem parte do Primeiro Planalto Paranaense (PPP), Escarpa Devoniana (ED) e Segundo Planalto Paranaense (SSP), e estão inseridas na Bacia Hidrográfica do Rio Pitangui, a qual compreende uma rede de drenagem que segue em direção ao interior do estado (oeste). A Bacia do Rio Pitangui, possui uma área de 927,3 Km², situando-se na porção centro-leste do Estado do Paraná, entre as latitudes 25°46'40" e 25°49'06"S, e as longitudes 49°46'40" e 50°17'38"W. O Rio Pitangui é o principal rio da bacia e provém do PPP, atravessando a ED por meio de vale encaixado e em seguida sendo barrado artificialmente, formando o reservatório de Alagados, utilizado para produção de energia hidrelétrica e abastecimento de água para o município de Ponta Grossa.

O rio Pitangui caracteriza-se por ser um rio geologicamente antigo, isto é, antecede a formação do relevo atual. Suas nascentes localizam-se em trechos meandantes e encaixados (GEALH, MELO, MORO, 2010) no PPP, na Serra das Areias no distrito de Abapã, município de Castro, penetrando no SPP por meio ED e percorrendo os municípios de Carambeí e Ponta Grossa em direção à sua foz. (Figura 1).

A região pertencente às unidades de estudo é composta por rochas sedimentares expostas pelo soerguimento e erosão do terreno (MELO et al., 2010), representa a transição entre o Primeiro e o Segundo Planalto Paranaense. Caracteriza-se por uma região ecotonal, onde a vegetação do PPP, pertencente à zona fitoecológica Floresta Ombrófila Mista Montana forma um mosaico climático-edáfico com a Estepe Gramíneo-Lenhosa do SPP (VELOSO, RANGEL FILHO; LIMA, 1991).

Imperam Cambissolos Háplicos e Neossolos Litólicos formados sobre o arenito da Formação Furnas, o qual origina solos rasos, arenosos e bastante suscetíveis a processos erosivos (EMBRAPA, 2002). O clima na região da bacia do rio Pitangui é do tipo Cfb da classificação de Koeppen (MAACK, 2002), isto é, zona temperada sempre úmida, com mais de cinco geadas anuais. A temperatura média anual está entre 16°C e 22°C e os ventos têm predominância NE no verão e NW no inverno (CRUZ, 2007). A precipitação média anual situa-se entre 1.400mm e 1.800mm, sendo o período mais seco de junho a agosto e período mais chuvoso de dezembro a fevereiro.

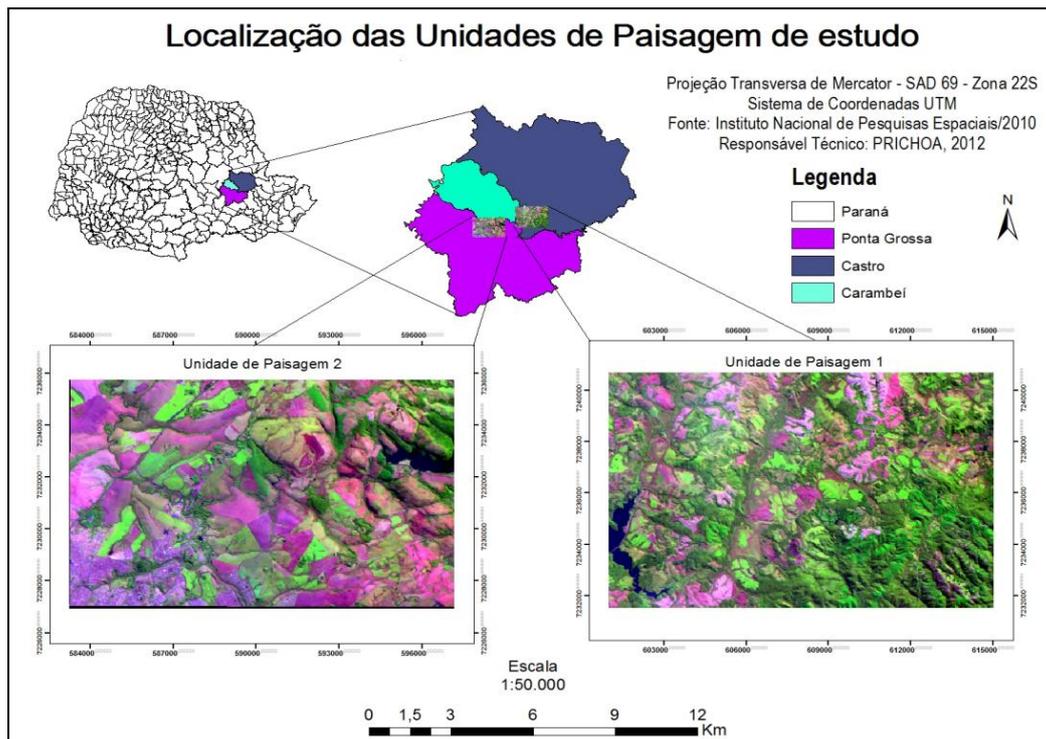


Figura 1: Localização das unidades de paisagem de estudo.

2. Metodologia de trabalho

Os materiais utilizados para a elaboração deste estudo consistem em: Carta topográfica de Ponta Grossa – PR, SG 22-X-C-II/2 (2005) na escala 1:50.000 no formato digital com extensão JPG, imagem Landsat 5 TM (2010), com resolução espacial de 30 metros e 6 bandas espectrais, disponibilizada gratuitamente pelo site do INPE¹, curvas de Nível com equidistância de 20m cedidas pelo Paranacidade no formato DWG, ortoimagem oriunda do Sensor SPOT 5 (2005), com resolução espacial de 5 metros cedida pelo Paranacidade (SEDU-PR).

Para os processamentos foram empregados os seguintes *softwares*: ARCGIS 9.3; ENVI 4.4; ENVI EX e PAST.

A Figura 2 descreve as etapas que originaram os descritores ou variáveis espaciais, espectrais e de textura de maior correlação.

Primeiramente realizou-se a correção geométrica da imagem Landsat 5 TM pela transformação geométrica com 20 pontos de controle e notáveis, com o apoio da ortoimagem e da carta topográfica, no software ENVI 4.4. As curvas de nível foram utilizadas para o processamento do Modelo Digital do Terreno (MDT) usando o interpolador *Inverse Distance Weizhted* (IDW) – Inverso do Quadrado da Distância, no ArcGis 9.3. Nas imagens Landsat e MDT efetivou-se a etapa de Análise Orientada a Objeto, mediante software ENVI EX, a qual originou tabelas com descritores.

Os descritores advindos da Análise Orientada a Objetos² possuem diferentes escalas e dimensões, portanto fez-se necessário a normalização destes dados no software PAST,

¹ Instituto Nacional de Pesquisas espaciais - <http://www.inpe.br/>

² Para maiores informações sobre a técnica consultar PRICHOA, C. E., 2012 Estudo em duas unidades de paisagem da bacia hidrográfica do rio Pitangui/PR mediante estatística multivariada e análise orientada a objetos, Bicen/ UEPG.

tornando representados num intervalo de 0 a 1 e, portanto obedecendo uma distribuição igualitária. Em seguida realizaram-se testes mediante ACP – Análise de Componentes Principais – podendo-se visualizar a correlação por meio de gráficos de dispersão e de tabelas até obter-se o número de variáveis com correlação coerente entre elas.

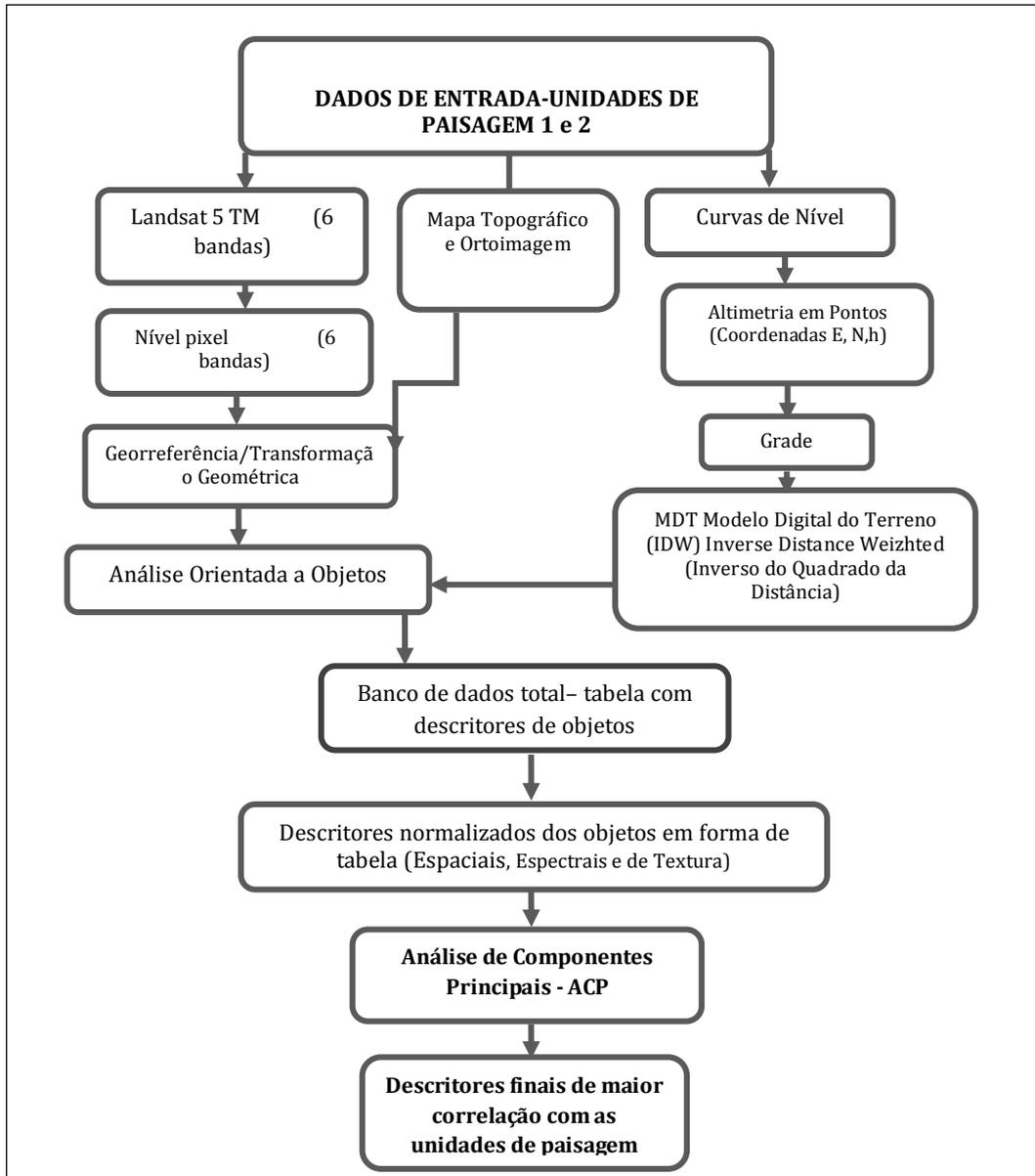


Figura 2: Diagrama de blocos das etapas realizadas para pesquisa.

3. Resultados e discussão

A Análise de Componentes Principais é uma técnica estatística que mediante correlação e variância formou na primeira componente a maioria das informações contidas em todas as variáveis. Nesta pesquisa, esta análise foi utilizada para reduzir a dimensionalidade e concomitantemente avaliar a correlação entre as variáveis (descritores dos segmentos ou regiões).

A Figura 03 mostra o gráfico de dispersão contendo todas as variáveis (35) originadas a partir da Análise Orientada a Objeto com as bandas espectrais (bandas 5R4G3B). E a Figura 04 exibe o gráfico contendo o total de variáveis (39) advindas da técnica Orientada a Objeto para as bandas espectrais (bandas 5R4G3B) combinadas com a variável física (banda

4 MDT). Devido ao vasto número de variáveis, nota-se que muitas delas ficaram posicionadas tão próximas umas das outras que impossibilitou a identificação. Entretanto, outras variáveis ficaram posicionadas fora da elipse denotando maior heterogeneidade com as unidades de paisagem. Por isso, para chegar a um número coerente de variáveis, necessitou-se eliminar gradativamente àquelas de menor correlação ou maior heterogeneidade, na medida em que se processavam os gráficos de dispersão.

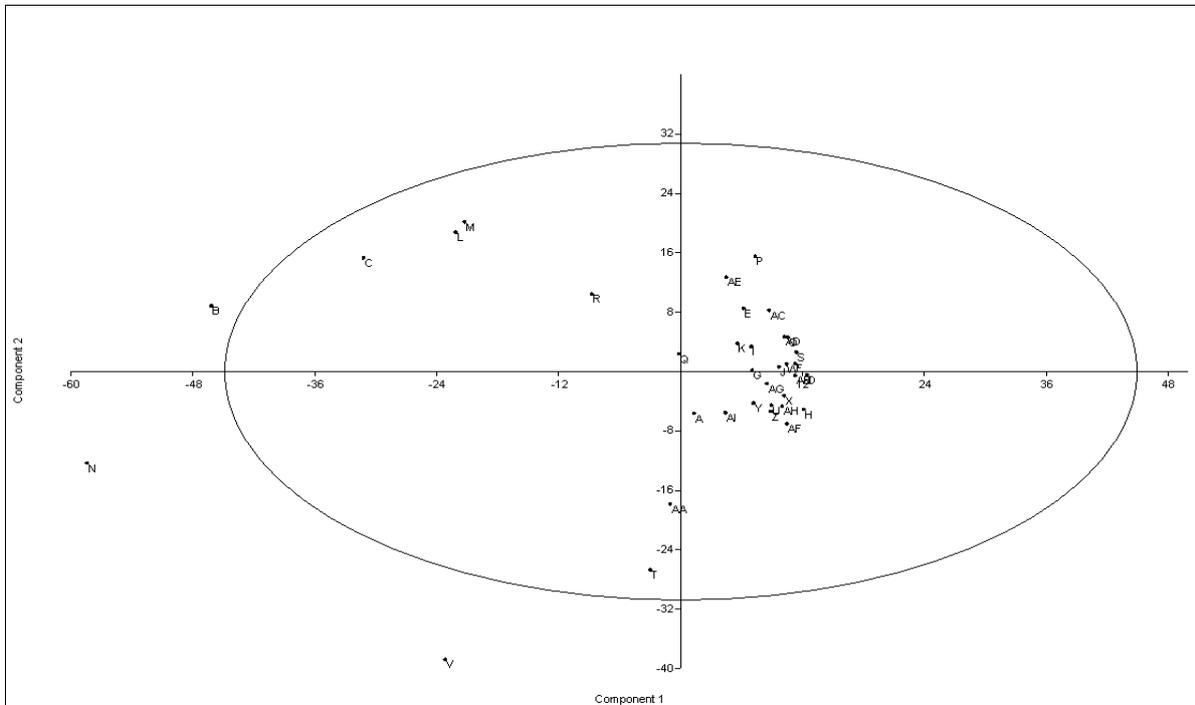


Figura 03 – Gráfico de dispersão contendo o total das variáveis (35) extraídas das imagens Landsat 5 TM (bandas 5R4GB3).

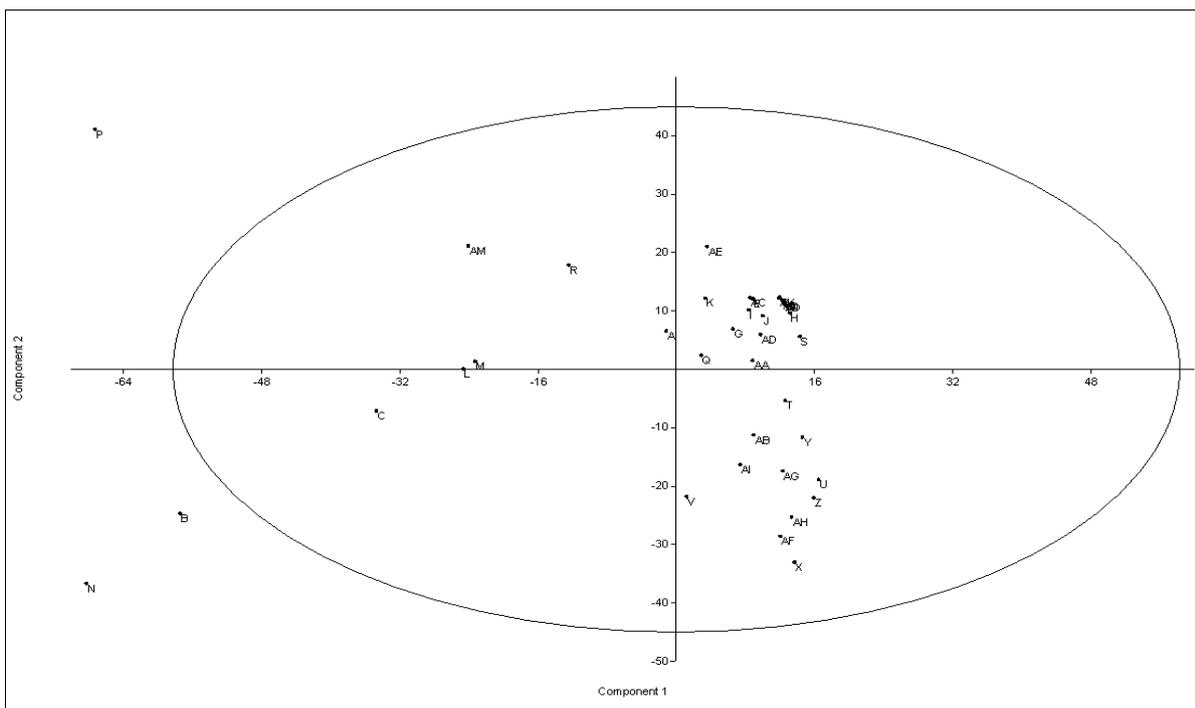


Figura 04 – Gráfico de dispersão contendo o total das variáveis (39) extraídas das imagens Landsat 5 TM (bandas 5RG4B3 e banda 4MDT).

A ACP embasou o julgamento e a importância das próprias variáveis originais geradas, ou seja, os descritores advindos da Análise Orientada a Objetos. As variáveis com maior peso ou correlação na combinação linear foram consideradas as mais importantes do ponto de vista estatístico. Esta análise reduziu ao máximo o número de variáveis heterogêneas e, posteriormente, possibilitou a interpretação. O coeficiente de correlação obtido nas Componentes CP 1 com CP 2 aproximou-se de 50%, denotando assim as variáveis de maior homogeneidade (Tabela 1). Na primeira coluna encontram-se as composições espectrais utilizadas para este estudo correspondentes a unidades 1 e 2, representadas pela imagem com três bandas do espectro eletromagnético (bandas R5G4B3) Landsat 5 TM e também as mesmas imagens associadas ao MDT.

Tabela 1 – Percentuais de Correlação para as variáveis finais da Análise de Componentes Principais.

IMAGENS	Percentual de Correlação
Landsat bandas 5R4G3B – unidade de paisagem 1	33,86%
Landsat bandas 5R4G3B – unidade de paisagem 2	29,72%
Landsat bandas 5R4G3B e banda 4 MDT- unidade de paisagem 1	47,34%
Landsat bandas 5R4G3B e banda 4 MDT- unidade de paisagem 2	41,44%

Julga-se que a homogeneidade das variáveis com as unidades de paisagem depende do posicionamento das variáveis no gráfico de dispersão, ou seja, posicionando-as na parte interior da elipse a heterogeneidade com as unidades será menor. Dessa forma, ao atingir os percentuais de correlação (Tabela 1) para as duas unidades em diferentes composições espectrais (bandas R5G4B3 e R5G4B3 e banda 4 MDT) a correlação foi mais coerente entre 17 variáveis, as quais podem ser vistas na Tabela 2.

Destacam-se na primeira coluna as variáveis espaciais, a saber: compacidade, convexidade, solidez, fator de forma, retangularidade e relação área-furos; Na segunda coluna têm-se as variáveis Espectrais: intensidade, valor máximo de pixel da banda 1, média do valor digital da banda 1, valor máximo de pixel da banda 2, valor máximo de pixel da banda 3, média do valor digital da banda 3, valor mínimo de pixel da banda 4 MDT, valor máximo de pixel da banda 4 MDT, média do valor digital da banda 4 MDT; Na terceira e última coluna tem-se as variáveis de Textura: textura média e entropia de textura.

Tabela 2 – Variáveis de maior homogeneidade com as unidades de paisagem.

Variáveis Espaciais	Variáveis Espectrais	Variáveis de Textura
(D) compacidade; (E) convexidade; (F) solidez; (H) fator de forma; (J) retangularidade; (O) relação área – furos.	(S) intensidade; (Y) valor máx. de pixel da banda 1; (Z) média do valor digital da banda 1; (AC) valor máx. de pixel da banda 2; (AG) valor máx. de pixel da banda 3; (AH) média do valor digital da banda 3; (AJ) valor mín. de pixel da banda 4 MDT; (AK) valor máx. de pixel da banda 4 MDT; (AL) média do valor digital da banda 4 MDT.	(U) textura média; (W) entropia de textura.

4. Conclusões

As técnicas estatísticas permitem a redução da dimensionalidade dos descritores espaciais, espectrais e de textura mediante Componentes Principais e também conhecimento prévio daqueles de maior correlação com unidades de paisagem.

A aplicabilidade da Análise de Componentes Principais nesta pesquisa comprovou que é possível extrair e selecionar automaticamente variáveis ou descritores de maior relação com unidades de estudo com a intenção de otimizar análises quantitativas mantendo um nível de coerência e eficácia para cada tipo de apreciação.

A associação das variáveis de maior correlação desta pesquisa aludem características tanto físicas quanto de uso das unidades de paisagem, porém, tais especificidades não foram abordadas nesta síntese. Por se tratar de um tema atual e eficaz, a utilização da ACP contribuiu na sintetização dos descritores ou variáveis, podendo-se notar na Tabela 03 a representação de algumas variáveis que tiveram maior correlação.

Tabela 3: Representação de descritores: espacial, espectral e de textura.

DESCRITOR																															
Espacial:	Forma Compacta 	Forma não Compacta 																													
Espectral:	Valor Digital Máximo = 80 <table border="1"> <tr><td>15</td><td>40</td><td>10</td></tr> <tr><td>50</td><td>15</td><td>20</td></tr> <tr><td>80</td><td>20</td><td>45</td></tr> </table>	15	40	10	50	15	20	80	20	45	Valor Digital Mínimo = 10 <table border="1"> <tr><td>15</td><td>40</td><td>10</td></tr> <tr><td>50</td><td>15</td><td>20</td></tr> <tr><td>80</td><td>20</td><td>45</td></tr> </table>	15	40	10	50	15	20	80	20	45	Média do Valor Digital = $295/9 = \mathbf{32,77}$ <table border="1"> <tr><td>15</td><td>40</td><td>10</td></tr> <tr><td>50</td><td>15</td><td>20</td></tr> <tr><td>80</td><td>20</td><td>45</td></tr> </table>		15	40	10	50	15	20	80	20	45
15	40	10																													
50	15	20																													
80	20	45																													
15	40	10																													
50	15	20																													
80	20	45																													
15	40	10																													
50	15	20																													
80	20	45																													
Textura	Textura Lisa 	Textura Rugosa 																													

Agradecimentos

As autoras agradecem a Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pelo subsídio financeiro o qual possibilitou o desenvolvimento de uma pesquisa mais abrangente (Dissertação de Mestrado), onde se pôde extrair este estudo o qual faz parte.

Referências Bibliográficas

- BOROUCHE, J. M., SAPORTA. G. **Análise de dados**. Zahar Editores. Rio de Janeiro, 1982. 116 p.
- CRUZ, G. C. F. Alguns aspectos do clima dos Campos Gerais. In: MELO, M.S.; MORO, R.S.; GUIMARÃES, G.B. (Eds.). **Patrimônio Natural dos Campos Gerais do Paraná**. Ponta Grossa: Ed. UEPG, 2007. p. 59-72.
- EMBRAPA. **Centro Nacional de Pesquisa de Florestas. Zoneamento Ecológico para Plantios Florestais no Estado do Paraná**. Brasília: EMBRAPA-DDT, 2002.

GEALH, A.M.; MELO, M.S.; MORO, R.S. (Orgs.). **Pitangui, Rio de Contrastes**: seus lugares, seus peixes, sua gente. Ponta Grossa: Ed. UEPG, 2010. 199p.

MATHER, P.M. **Computer processing of remotely sensed images**. New York: John Wiley e Sons, Inc., 1999. 292 p.

MAACK, R. **Geografia física do Estado do Paraná**. Curitiba: Imprensa Oficial, 2002. 440p.

MELO, M. S. et al. Relevo e hidrografia dos Campos Gerais. In: MELO, M. S.; MORO, R. S.; GUIMARÃES, G. B. **Patrimônio Natural dos Campos Gerais do Paraná**. Ponta Grossa: Editora UEPG, 2007. p. 49-58.

REGAZZI, A. J. **INF 766 - Análise Multivariada**. Viçosa: Universidade Federal de Viçosa, Centro de Ciências Exatas e Tecnológicas. Departamento de Informática. 2001. 166p.

RICHARDS, John. A. **Remote Sensing Digital Image Analysis - An Introduction**. 2nd ed. Springer-Verlag, Berlin. 1993. 281p.

SCHOWENGERDT, R.A. **Remote Sensing: Models and Methods for Image Processing**. 3. ed. London: Academic Press. 1997. 560 p.

VELOSO, H. P.; RANGEL FILHO, A. L.; LIMA, J. C. A. **Classificação da Vegetação Brasileira, adaptada a um sistema universal**. Rio de Janeiro: IBGE, 1991. 123p. Disponível em: <http://biblioteca.ibge.gov.br/visualizacao/monografias/GEBIS%20-%20RJ/classificacaovegetal.pdf>. Acesso em: 18 jan. 2012.