

CARLOS RENATO DOS SANTOS

**CLASSIFICAÇÃO E SEGMENTAÇÃO DE IMAGENS DE
TOMOGRAFIA DE SOLOS VIA PLANOS DE ENTROPIA**

**RECIFE-PE
DEZEMBRO/2017**



UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOMETRIA E ESTATÍSTICA APLICADA

**CLASSIFICAÇÃO E SEGMENTAÇÃO DE IMAGENS DE
TOMOGRAFIA DE SOLOS VIA PLANOS DE ENTROPIA**

Tese apresentada ao Programa de Pós-Graduação em Biometria e Estatística Aplicada como exigência parcial à obtenção do título de Doutor.

Área de Concentração: Estatística aplicada

Orientador: Prof. Dr. Rômulo Simões Cezar Menezes

Coorientador: Prof. Dr. Borko Stosic

RECIFE-PE – DEZEMBRO/2017

Dados Internacionais de Catalogação na Publicação (CIP)
Sistema Integrado de Bibliotecas da UFRPE
Biblioteca Central, Recife-PE, Brasil

S237c Santos, Carlos Renato dos.
Classificação e segmentação de imagens de tomografia de
solos via planos de entropia / Carlos Renato dos Santos. – 2017.
81 f. : il.

Orientador: Rômulo Simões Cezar Menezes.

Coorientador: Borko Stosic.

Tese (Doutorado) – Universidade Federal Rural de
Pernambuco, Programa de Pós-Graduação em Biometria e
Estatística Aplicada, Recife, BR-PE, 2017.

Inclui referências e apêndice(s).

1. Segmentação 2. Classificação de imagens 3. Planos de
Entropia 4. Tomografia de solos I. Menezes, Rômulo Simões Cezar,
orient. II. Stosic, Borko, coorient. III. Título

CDD 310

UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOMETRIA E ESTATÍSTICA APLICADA

**CLASSIFICAÇÃO E SEGMENTAÇÃO DE IMAGENS DE
TOMOGRAFIA DE SOLOS VIA PLANOS DE ENTROPIA**

CARLOS RENATO DOS SANTOS

Tese julgada adequada para obtenção do título de Doutor em Biometria e Estatística Aplicada, defendida e aprovada com distinção em 13/12/2017 pela Banca Examinadora.

Orientador:

Prof. Dr. Rômulo Simões Cezar Menezes
UFPE – Departamento de Energia Nuclear

Banca Examinadora:

Prof. Dr. Borko Stosic
UFRPE – Departamento de Estatística e Informática

Prof. Dr. Richard J. Heck
University of Guelph – School of Environmental Sciences / Canadá

Prof. Dr. Antonio Celso Dantas Antonino
UFPE – CTG – Departamento de Energia Nuclear

Profa. Dra. Maria do Socorro Bezerra de Araújo
UFPE – CFCH – Departamento de Ciências Geográficas

Dedicatória

Esta tese é dedicada a minha Vó
Nair Gomes Pereira, Dona Nair (in memoriam)

Agradecimentos

Primeiramente agradeço a Deus por todos os desafios vencidos até aqui.

A minha família como um todo, em especial a minha mãe Maria Aparecida, a minha esposa Alane de Aquino e ao meu tio Chico pelo apoio e vários momentos vividos durante essa jornada.

Aos meus orientadores Rômulo Menezes e Borko Stosic, os quais somaram sua sabedoria e tempo para me fortalecer nesta proposta de trabalho. Sempre muito disponíveis e ativos em minhas variadas dificuldades e acertos.

Ao Laboratório de Tomografia do Departamento de Energia Nuclear da UFPE pela estrutura, ensinamentos e apoio através dos professores Antonino e Richard Heck, dos alunos Larissa Fernandes, Wellington, Cássia e do físico e responsável técnico Daniel.

Ao amigo Irmão Manoel Valdivino, de Chã de Alegria – PE, pela forte parceria na coleta das amostras.

Aos amigos contemporâneos do PPGBEA que contribuíram diretamente nesta proposta discutindo sobre planos de entropia e programação no R, Ikaro Barreto e o atualmente docente do DEINFO Antonio Samuel.

Aos funcionários Marco Santos e Eduardo Chaves pelo apoio frente a secretaria do PPGBEA.

Ao Grupo de Pesquisa em Energia da Biomassa do Departamento de Energia Nuclear da UFPE, principalmente através das pessoas do pesquisador Dario Primo e alunos de doutorado Eduardo Lucena e Rafael Silva.

A todos os professores do PPGBEA, em especial aqueles que tive contato nas disciplinas e mais ativamente dentro do programa: Guilherme Moreira, Tatijana Stosic e o atual coordenador Moacyr Cunha.

A todos os parceiros e amigos mestrando e doutorando com os quais tive o prazer de compartilhar almoços no nosso restaurante universitário, a sala de estudos, algumas disciplinas e por vezes alguns momentos fora da UFRPE.

A Usinas São José e Petribú pelo acesso e grande apoio técnico em suas propriedades.

A todos que incentivaram, torceram e ajudaram na concretização de um dos dias mais felizes da minha vida !!!

Resumo

Apesar de se tratar de uma das primeiras preocupações no estudo da caracterização de imagens nos mais diversos contextos, o processo segmentar, ou seja, de separar objetos de interesse do restante da imagem nem sempre é um procedimento óbvio ou trivial. Um dos materiais em geociências que mais sofre investigação, inclusive na área de segmentação de imagens, são os solos. Estes desempenham papéis diversos e críticos na biosfera, regulando ciclos biogeoquímicos, fornecendo um *habitat* para microorganismos e um meio para raízes de plantas, armazenando nutrientes e minerais e conduzindo fluidos (como água e gases). A identificação da estrutura dos poros de um determinado segmento do solo é importante, pois fornece informações sobre a mecânica do solo e sobre a forma como pode cumprir sua função. Neste sentido, esta tese propõe um método para uma segmentação adequada, identificando corretamente as fases puras de vazios e sólidos, como também de *voxels* mistos, ou seja, aqueles que de imediato não se enquadram em vazios ou de sólidos, sendo esta separação um dos grandes desafios na segmentação de imagens de tomografia de solos, como também muitas outras imagens digitais. A construção de planos de entropia refletivos pela penetração das imagens na direção da gravidade usando entropia amostral e informação de Shannon permitiram a partir da correlação destes planos construir uma curva de multisegmentação que identifica fases puras a partir de um vale e um pico, sugerindo o método *Entropic Plans Correlations* (EPC). Esta curva foi confrontada com o histograma do método *Pure Voxels Extraction* (PVE), que também realiza busca por regiões puras e adiante calibrada para alocar corretamente os *voxels* mistos para as fases de vazios ou não vazios por meio da teoria de detecção de bordas. O método proposto de EPC, quando comparado com PVE, atendeu aos propósitos de uma boa segmentação de imagens, visto que obteve uma elevada correspondência espacial e diminuiu fortemente a subjetividade encontrada em algumas etapas deste último. Os planos de entropia forneceram ainda possibilidade de uma consistente classificação destas imagens, evidenciando regiões de baixa densidade, ou seja, poros e matéria orgânica e regiões de alta densidade, ou seja, rochas e minerais mais duros, como duas regiões relevantes na separação de solos de mata nativa e do cultivo de cana-de-açúcar. Estes resultados mostraram que confrontar planos de entropia obtidos das imagens de solos, seja por meio da correlação ou da análise multivariada, pode trazer relevantes achados no âmbito da segmentação e classificação destas imagens.

Palavras-chave: segmentação, classificação de imagens, planos de entropia, tomografia de solos.

Abstract

Although it is one of the first concerns in the study of the characterization of images in the most diverse contexts, the segmental process, that is, of separating objects of interest from the rest of the image is not always an obvious or trivial procedure. One of the materials in geosciences that undergoes more investigation, including in the area of segmentation of images, are the soils. They play diverse and critical roles in the biosphere, regulating biogeochemical cycles, providing a habitat for microorganisms and a medium for plant roots, storing nutrients and minerals and conducting fluids (such as water and gases). The identification of the pore structure of a given soil segment is important as it provides information on soil mechanics and how it can perform its function. In this sense, this thesis proposes a method for adequate segmentation, correctly identifying the pure phases of voids and solids, as well as mixed voxels, that is, those that do not immediately fit into voids or solids, this separation being one of the challenges in the segmentation of soil tomography images, as well as many other digital images. The construction of reflective entropic planes by the penetration of images in the direction of gravity using sample entropy and Shannon information allowed from the correlation of these planes to construct a multisegmentation curve that identifies pure phases from a valley and a peak, suggesting the method *Entropic Plans Correlations* (EPC). This curve was compared with the *Pure Voxels Extraction* (PVE) histogram, which also performs pure region searches and is then calibrated to correctly allocate the mixed voxels to the empty or non-void phases through edge detection theory. The proposed method of EPC, when compared with PVE, served the purposes of good image segmentation, since it obtained a high spatial correspondence and strongly diminished the subjectivity found in some stages of the latter. The entropic plans also provided a possibility of a consistent classification of these images, evidencing low density regions, that is, pores and organic matter and regions of high density, that is, harder rocks and minerals, as two relevant regions in the separation of soils of native forest and of the cultivation of sugar cane. These results showed that comparing entropic plans obtained from soil images, either through correlation or multivariate analysis, can bring relevant findings within the segmentation and classification of these images.

Keywords: segmentation, image classification, entropic plans, soil tomography.

LISTA DE FIGURAS

Figura 1. Variação de valores de pixels em parte de um poro na matriz de um solo.....	2
Figura 2. Foto aérea e polígono aproximado do fragmento da Mata de Piedade no município de Igarassu - PE, terras da Usina São José.	7
Figura 3. Estrada que separa a Mata de Piedade e o cultivo de cana-de-açúcar na Usina São José no município de Igarassu - PE	8
Figura 4. Foto aérea e polígono aproximado do fragmento da Mata do Cumbe no município de Araçoiaba – PE, terras da Usina São José.	9
Figura 5. Coleta física das amostras de solo por meio da escavação com facas e armazenamento da estrutura original do solo em cilindro de PVC para posterior tomografia ..	9
Figura 6. Amostra de solo no interior do tomógrafo	11
Figura 7. Escala de Hounsfield.....	12
Figura 8. Imagem com artefato de anel	12
Figura 9. Imagem corrigida	12
Figura 10. Processo geral de uma Análise de componentes principais	16
Figura 11. Exemplo de dendograma com classificação de imagens	18
Figura 12. Método de agrupamento aglomerativo (a) e partitivo (b).....	19
Figura 13. Ilustração da obtenção de um plano entrópico pela penetração de uma imagem na direção da gravidade por medidas de entropia	20
Figura 14. Planos de entropia amostral aplicados na escala de cinza e de Shannon aplicados em voxels de baixa densidade em imagens de tomografia de solos de mata e cana	21
Figura 15. Curvas de multissegmentação obtidas pela correlação de planos de entropia sob variados <i>thresholds</i> acumulados em solos de cana.....	23
Figura 16. Curvas de multissegmentação obtidas pela correlação de planos de entropia sob variados <i>thresholds</i> acumulados em solos de mata	24
Figura 17. Classificação de <i>voxels</i> por meio da curva de multissegmentação proposta pelo método de correlação de planos de entropia (EPC), onde <i>pv</i> são <i>voxels</i> puros vazios, <i>m</i> são voxels mistos e <i>ps</i> são <i>voxels</i> puros sólidos	25
Figura 18. Percentuais de <i>voxels</i> puros vazios (a), mistos (b) e puros sólidos (c) m imagens de tomografia de solos de mata e cana	26

Figura 19. Classificação dos <i>voxels</i> de acordo com a curva de multisegmentação por EPC para a imagem XB300, os <i>pixels</i> em vermelho representam a seleção de cada categoria sob a imagem original em 16 <i>bits</i> (<i>pv</i> = puros vazios, <i>m</i> = mistos e <i>ps</i> = puros sólidos).....	26
Figura 20. Comparativo de dispersão e variabilidade do <i>Vale</i> da curva de multisegmentação em EPC aplicados em imagens de tomografia de solos de mata e cana.....	27
Figura 21. Obtenção de <i>Zepc₁</i> e <i>Zepc₂</i> para XA200.....	29
Figura 22. Correspondência espacial entre EPC corrigido por <i>Zepc₂</i> e PVESeg fatia por fatia	29
Figura 23. Percentuais de <i>voxels</i> puros e mistos de XB300 a partir dos limiares propostos pela curva de correlação de planos de entropia	31
Figura 24. Percentuais de <i>voxels</i> em imagens de solos de mata e cana.....	31
Figura 25. Desvio padrão das imagens e variabilidade de <i>Zepc₁</i> e <i>Zepc₂</i>	32
Figura 26. Curvas de multisegmentação obtidas por correlação de planos de entropia (EPC) apresentado <i>Vale</i> e <i>Pico</i> para pré-segmentação de imagens de tomografia de solos	35
Figura 27. Evolução da fase de puros vazios para vazios totais.....	37
Figura 28. Variação de porosidade por EPCSeg	37
Figura 29. Porosidade média via EPCSeg - Teste de Tukey.....	37
Figura 30. Representação geral de uma curva de explicabilidade total obtida pela dispersão entre a variação de HU e variabilidade explicada pela primeira componente de cada ACP....	38
Figura 31. Curva de explicabilidade total a partir de uma ACP da média dos planos no decorrer da variação de HU	39
Figura 32. Análise de componentes principais utilizando limiar proposto em <i>P_{cet1}</i>	40
Figura 33. Análise de componentes principais utilizando limiar proposto em <i>P_{cet2}</i>	40
Figura 34. CET a partir de todos os <i>pixels</i> dos planos	41
Figura 35. Análise de componentes principais utilizando limiar proposto em <i>P_{cet1}</i>	41
Figura 36. Análise de componentes principais utilizando limiar proposto em <i>P_{cet2}</i>	42
Figura 37. Curva de explicabilidade total a partir de ACP para classificar os <i>pixels</i> dos planos	42
Figura 38. ACP utilizando limiar proposto por <i>P_{cet1}</i>	43
Figura 39. ACP utilizando limiar proposto por <i>P_{cet2}</i>	43
Figura 40. Seleções de <i>pixels/voxels</i> para imagem B4001 baseada CET em ACP ₁	44
Figura 41. Presença de <i>voxels</i> até <i>P_{cet1}</i> em ACP ₁	45
Figura 42. Presença de <i>voxels</i> maiores que <i>P_{cet2}</i> em ACP ₁	45
Figura 43. Entropia de Shannon média em diferentes usos e profundidades para <i>P_{cet1}</i>	46

Figura 44. Entropia de Shannon média em diferentes usos e profundidades para P_{cet2}	46
Figura 45. Entropia amostral média em diferentes usos e profundidades	46
Figura 46. Dendograma da AAg ₁ utilizando P_{cet1} em ACP ₁	47
Figura 47. Clusters da AAg ₁ utilizando P_{cet1} em ACP ₁	48
Figura 48. Dendograma da AAg ₁ utilizando P_{cet2} em ACP ₁	49
Figura 49. Clusters da AAg ₁ utilizando P_{cet2} em ACP ₁	49
Figura 50. Dendograma da AAg ₂ utilizando P_{cet1} em ACP ₂	50
Figura 51. Clusters da AAg ₂ utilizando P_{cet1} em ACP ₂	50
Figura 52. Dendograma da AAg ₂ utilizando P_{cet2} em ACP ₂	51
Figura 53. Clusters da AAg ₂ utilizando P_{cet2} em ACP ₂	51

LISTA DE TABELAS

Tabela 1. Caracterização das amostras coletadas	10
Tabela 2. Medidas de dissimilaridade	17
Tabela 3. <i>Vale</i> e <i>Pico</i> por EPC	25
Tabela 4. <i>P1</i> e <i>P2</i> por PVE.....	25
Tabela 5. Valores de ajuste de EPC buscando similaridade com PVE	30
Tabela 6. Configurações de EPCSeg avaliadas	33
Tabela 7. Valores de porosidade de PVESeg e configurações de EPCSeg	34
Tabela 8. Correspondência espacial com PVESeg	34
Tabela 9. Porosidade e parametrização da segmentação final.....	36
Tabela 10. Valores de <i>threshold</i> com maiores e menores taxas de explicabilidade.....	39
Tabela 11. Valores de <i>threshold</i> com maiores e menores taxas de explicabilidade.....	41
Tabela 12. Valores de <i>threshold</i> com maiores e menores taxas de explicabilidade.....	43
Tabela 13. Correlação cofenética utilizando P_{cet1} em ACP ₁	47
Tabela 14. Correlação cofenética utilizando P_{cet2} em ACP ₁	48
Tabela 15. Correlação cofenética utilizando P_{cet1} em ACP ₂	50
Tabela 16. Correlação cofenética utilizando P_{cet1} em ACP ₂	51
Tabela 17. Grupos em AAg ₁ utilizando P_{cet1} em ACP ₁	52
Tabela 18. Grupos em AAg ₁ utilizando P_{cet2} em ACP ₁	52
Tabela 19. Grupos em AAg ₂ utilizando P_{cet1} em ACP ₂	53
Tabela 20. Grupos em AAg ₂ utilizando P_{cet2} em ACP ₂	53

LISTA DE ABREVIATURAS E SIGLAS

AAG	Análise de agrupamentos
ACP	Análise de componentes principais
ANOVA	Análise de variância
CET	Curva de explicabilidade total
CT	Tomografia computadorizada
EPC	Correlação de planos de entropia
EPCSeg	Segmentação através da correlação de planos de entropia
HU	Unidades de Hounsfield
KL	Critério de seleção de grupos de Krzanowski e Lai
PVE	Extração de <i>voxels</i> puros
PVESeg	Segmentação através da extração de <i>voxels</i> puros
RBMA	Reserva da Biosfera da Mata Atlântica
SE	Entropia amostral
SHE	Entropia de Shannon
UNESCO	Organização das Nações Unidas para Educação, Ciência e Cultura
VAD	<i>Voxels</i> de alta densidade
VBD	<i>Voxels</i> de baixa densidade

SUMÁRIO

1. INTRODUÇÃO	1
1.1 Segmentação de imagens de solos	2
1.2 <i>Pure Voxels Extraction</i> (PVE).....	3
1.3 A mudança de mata nativa para cana-de-acúcar	4
1.4 Objetivos.....	5
1.5 Organização da tese	6
2. MATERIAIS E MÉTODOS	7
2.1 Local, coleta e preparo das amostras	7
2.3 Escaneamento das amostras via CT.....	10
2.4 Informação de Shannon (SHE).....	13
2.5 Entropia amostral (SE)	14
2.6 Coeficiente de correlação de Pearson.....	15
2.7 Análise de componentes principais (ACP).....	15
2.8 Análise de agrupamentos (AAg)	17
3. ENTROPIC PLANS CORRELATIONS (EPC)	20
3.1 Planos de entropia.....	20
3.2 Curva de multisegmentação por EPC e histograma por PVE	22
3.3 Análise do comportamento de curva de multisegmentação	25
3.4 Calibração de EPC por PVESeg.....	28
3.5 Análise de alocação de <i>voxels</i> mistos	30
3.6 Deslocamentos de desvio padrão $Zepc_1$ e $Zepc_2$	32
3.7 Porosidade e correspondência espacial entre EPCSeg e PVESeg.....	33
3.8 Segmentação por EPCSeg	35
4. ANÁLISE MULTIVARIADA DE PLANOS DE ENTROPIA	38
4.1 ACP ₁ : Classificando imagens por meio de médias de planos de entropia	39
4.2 ACP ₂ : Classificando imagens por meio de <i>pixels</i> de planos de entropia	40
4.3 ACP ₃ : Classificando <i>pixels</i> de planos de entropia.....	42
4.4 Discussão sobre ACP e planos de entropia	43
4.5 AAg ₁ : Clusters a partir da sugestão em ACP ₁	47

4.6 AAg ₂ : Clusters a partir da sugestão em ACP ₂	49
4.7 Discussão sobre AAg e planos de entropia	52
5. CONCLUSÕES.....	54
5.1 Sugestões e contribuições para trabalhos futuros	55
REFERÊNCIAS BIBIOGRÁFICAS	56
APÊNDICE	66
Apêndice 1 – Coordenadas geográficas e classificação de solos dos pontos amostrados em Piedade.....	66
Apêndice 2 – Coordenadas geográficas e classificação dos solos dos pontos amostrados em Cumbe.....	67

1. INTRODUÇÃO

Apesar de se tratar de uma das primeiras preocupações no estudo da caracterização de imagens nos mais diversos contextos, o processo de separar objetos de interesse do restante da imagem nem sempre é um procedimento óbvio ou trivial. Tal processo é conhecido por segmentação de imagens, que corresponde a divisão de uma imagem em regiões ou categorias, que correspondem a diferentes objetos ou partes de objetos. Cada *pixel/voxel* em uma imagem é alocado para uma das várias categorias. Em geral, uma boa segmentação é tipicamente uma em que *pixels/voxels* na mesma categoria têm escala de valores multivariados semelhantes e formam uma região conectada, onde *pixels/voxels* vizinhos que estão em diferentes categorias têm valores diferentes.

Os métodos de segmentação de imagens aplicados em geociências podem ser amplamente classificados como métodos globais que geram um único limite aplicado a toda a imagem e métodos locais que adaptam o valor do limiar dependendo de características locais da imagem. No entanto, métodos de segmentação global tem sido frequentemente inferiores à métodos de segmentação local ¹.

Um dos materiais em geociências que mais sofre investigação, inclusive na área de segmentação de imagens, são os solos. Estes desempenham papéis diversos e críticos na biosfera, regulando ciclos biogeoquímicos, fornecendo um *habitat* para microorganismos e um meio para raízes de plantas, armazenando nutrientes e minerais e conduzindo fluidos (como água e gases). Este último é especialmente relevante para a produção agrícola e para a proteção ambiental das emissões de gases. Assim, a identificação da estrutura dos poros de um determinado segmento do solo é importante, pois fornece informações sobre a mecânica do solo e sobre a forma como pode cumprir sua função ².

Neste sentido, uma das técnicas de avaliação de parâmetros dos solos por imagens que têm se tornado bastante comum é a tomografia computadorizada de raios-x (CT), que pode fornecer detalhes da estrutura interna e intacta de blocos de solos, gerando imagens 3D com resolução micrométrica ³. Do ponto de vista estrutural, o solo pode ser definido como uma mistura de componentes sólidos e vazios ⁴. A mistura destes componentes associada a variabilidade natural da atenuação dos raios-x que podem evidenciar um mesmo elemento, fornece um desafio ainda maior na segmentação. Para exemplificar esta questão, considere uma fatia de uma imagem e nesta é selecionada uma região visualmente representativa da fase

de vazios (*pixels* mais escuros). A seguir, apresenta-se um histograma da variação dos valores dos *pixels* observados dentro desta área (Figura 1).

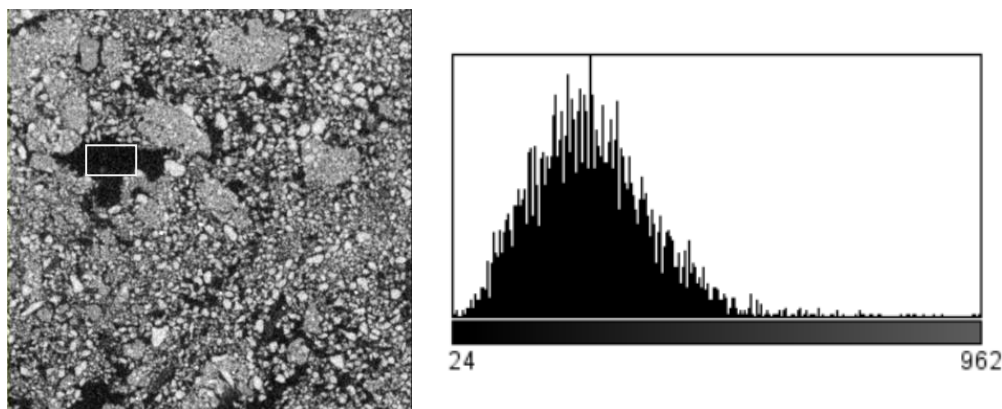


Figura 1. Variação de valores de pixels em parte de um poro na matriz de um solo

Diante desta necessidade enorme de identificar de maneira mais fiel possível as fases do solo em imagens, muitos autores vêm propondo diversas abordagens, que a seguir, apresentam-se em uma breve descrição dos recentes estudos em segmentação de imagens de solos.

1.1 Segmentação de imagens de solos

Muitos métodos têm sido sugeridos ou apenas aplicados na literatura, já que a maioria dos trabalhos necessita em algum momento da segmentação da imagem para investigar os *pixels* ou *voxels* de interesse, sendo uma inesgotável área de investigação e debate. A variação das propostas e aplicações vão desde a seleção manual, com base no *expertise* e satisfação de especialistas, até mesmo métodos totalmente automáticos como por exemplo, o método de Otsu ⁵. Somente no trabalho de Surový et al. ⁶, cujo objetivo era realizar a segmentação de raízes, foram comparados 17 métodos que já encontram-se previamente implementados no *software* ImageJ ⁷. Muitos trabalhos sugerem seus processos de segmentação através de técnicas de agrupamento ^{2; 8}, redes neurais ⁹, *single mean threshold* ¹⁰, utilizando medidas fractais ¹¹, aprimoramentos do método Otsu ¹², correlação de Pearson ¹³ e entropias ^{14; 15; 16; 17; 18; 19}.

Uma técnica comumente utilizada se dá pela avaliação do histograma da escala de cinza das imagens. Esta abordagem assume que as imagens são compostas por regiões com diferentes escalas de níveis de cinza, onde os picos correspondem ou identificam uma determinada região, existindo um vale entre estes picos ²⁰. No caso de imagens de tomografia

de solos, estes picos fornecem informação relevante quanto aos valores representativos das fases de vazios e sólidos, principalmente quando o histograma dispõe de um comportamento pelo menos bimodal, ou seja, apresentam ao menos dois picos, sendo o primeiro representante da fase de vazios e o segundo representante da fase de sólidos ^{21; 22}. Estes picos são em seguida considerados como o ponto de máximo de duas distribuições Gaussianas mistas e decisões sobre a segmentação final levam em conta a variabilidade encontrada internamente nestas ^{1; 23; 24}. Existem, porém, algumas dificuldades encontradas mesmo diante destas importantes referências trazidas pelo histograma, dentre elas, duas são muito comuns:

- Imagens de tomografia de solos, como também muitas outras imagens digitais, dispõem em geral de uma quantidade razoável de *voxels* mistos, ou seja, aqueles que de imediato não se enquadram como menores ou iguais ao primeiro pico ou maiores ou iguais ao segundo pico ²².
- Nem sempre as imagens são pelo menos bimodais, no caso de solos, aqueles que dispõem de pouco volume poroso, em geral apresentam apenas um único pico representante dos sólidos ²⁵.

Um dos métodos que têm buscado atenuar estas dificuldades e tem sido utilizado com muito sucesso em vários estudos é o *Pure Voxels Extraction* ^{25; 26; 27}. A seguir, apresenta-se mais detalhes deste método, que será adotado como referência para o método proposto nesta tese.

1.2 *Pure Voxels Extraction* (PVE)

O PVE é uma ferramenta que tem como objetivo localizar os *voxels* relativamente puros, em geral, dentro de uma vizinhança de 124 vizinhos a partir de cubos de cinco *voxels* de lado, buscando dentro desta vizinhança agrupar aqueles que são similares a um coeficiente de variação proposto pelo pesquisador. Desde modo enquadrando-se como método de segmentação local.

As imagens dos *voxels* puros geradas por essa ferramenta, em geral, possuem histogramas bimodais ou até mesmo multimodais com picos mais evidentes se comparados com histogramas da imagem original, onde cada região próxima aos picos é modelada por uma curva gaussiana representando as fases em que esses *voxels* estão inseridos. Em outras palavras, o PVE tem como principal objetivo tornar mais claros os picos das fases, os quais nem sempre podem ser obtidos diretamente sem esta consideração ^{21; 25; 28}.

Superada a etapa de identificação dos picos, o valor destes, juntamente com a variabilidade interna de cada uma das gaussianas são fornecidos para adotar uma decisão quanto aos *voxels* mistos, ou seja, aqueles que estão entre os dois picos. A alocação destes é feita através da teoria de teoria da detecção de bordas ²⁹, com fins de detectar o *threshold* ótimo que evidencie a mudança da fase vazios para sólidos, tal método está implementado no *plugin CT Segmentation* no software ImageJ. Definimos então PVESeg como sendo a segmentação final, ou seja, superando as etapas de pré-segmentação por PVE e segmentação por *CT Segmentation*. Maiores detalhes podem ser consultados no trabalho de Elliot e Heck ²⁸.

Embora o PVESeg seja um método de segmentação local muito bem-sucedido, sendo utilizado por relevantes trabalhos, existem algumas dificuldades quanto a subjetividade de alguns parâmetros, em ambas etapas, algo que não compromete a qualidade dos resultados, mas consiste de um desafio para o usuário desta metodologia, sendo muitas vezes obtidos parâmetros por tentativa e erro ou a partir da *expertise* de outros pesquisadores em suas particulares segmentações.

A metodologia PVESeg, apresenta muitos pontos em comum com o método proposto nesta tese, e adotaremos ela como referência para sugerir e comparar a proposta de um novo método de segmentação de imagens tomográficas de solos, também com duas etapas, uma pré-segmentação baseada na correlação de planos entrópicos, o *Entropic Plans Correlations* (EPC) e a segmentação também baseada na teoria de detecção de bordas implementada no *plugin CT Segmentation*, porém nesta última etapa com entradas não subjetivas. Definimos a segmentação final por EPC seguido da teoria de detecção de bordas por EPCSeg.

Além de utilizar como referência o PVESeg, este trabalho utiliza amostras de solos em dois usos considerados extremos pela literatura. São avaliadas imagens de solos de mata nativa e sob a produção de cana-de-açúcar. A seção seguinte resume os impactos desta mudança de uso e a importância de avaliar diferentes usos na validação das métricas e métodos usados e propostos neste trabalho.

1.3 A mudança de mata nativa para cana-de-açúcar

Com o aumento da população e dos investimentos governamentais e privados, surge a demanda cada vez mais crescente pela conversão de áreas naturais para urbanização, agricultura e pecuária, como também para o extrativismo, ameaçando fortemente as florestas

tropicais, de modo que nas duas últimas décadas a taxa de desmatamento aumentou de 5,6 para 9,1 milhões de hectares por ano ³⁰.

Esta degradação tem efeitos imediatos e negativos na biodiversidade, causando fragmentação, ou seja, a divisão do meio natural em pequenos blocos de terra, cada vez menores e mais isolados, sendo contornados justamente pelos usos modificados pelo homem. Dentre os biomas brasileiros mais afetados por esta questão, principalmente devido a expansão dos plantios de cana-de-açúcar, pode-se citar a mata atlântica, que hoje encontra-se largamente desmatada, restando pequenos fragmentos, em sua maioria com menos de 1000 hectares ³¹.

As atuais técnicas de manejo da cultura da cana-de-açúcar utilizam um vigoroso revolvimento do solo por ocasião do plantio, com o uso de arados, grades pesadas e subsoladores que acabam refletindo em alterações na estrutura do solo, principalmente na camada superficial. As modificações que ocorrem na estrutura do solo são evidenciadas por alterações nos valores de densidade do solo, resistência mecânica à penetração, porosidade total, porosidade de aeração, armazenagem e disponibilidade de água às plantas, dinâmica de água na superfície e no seu perfil, assim como a consistência e a compactação do solo ³². Muitos trabalhos investigam estes impactos negativos sobre muitas propriedades, dentre elas pode-se citar mudanças nos atributos físicos, químicos, perda de matéria orgânica e biodiversidade em macrofauna ^{33; 34; 35; 36; 37}. Sob estes impactos, espera-se que os métodos aqui propostos sejam capazes de classificar e segmentar bem imagens, mesmo que em diferentes e extremos usos, trata-se de uma das contribuições desta tese que serão melhor resumidas na seção de objetivos, a seguir.

1.4 Objetivos

O principal objetivo deste estudo propor um método de pré-segmentação de imagens baseado na correlação ótima do plano de entropia amostral com planos de informação de Shannon sob variados limiares. A partir deste objetivo principal, outros objetivos específicos também são propostos com fins de validar a técnica:

- Avaliar a viabilidade do uso de planos de entropia para a segmentação de imagens 3D de solos.
- Classificar imagens por meio de planos de entropia.
- Comparar o método proposto, EPCSeg com método similar e já bem estabelecido de PVESeg.

1.5 Organização da tese

Esta tese está organizada em cinco capítulos incluindo esta introdução, onde foram definidos o problema de segmentação de imagens no contexto de solos, os impactos de mudança de uso e os objetivos do estudo, no Capítulo 2 apresenta-se a metodologia de coleta de amostras e reconstrução de imagens, como também aspectos teóricos da teoria da informação, coeficiente de correlação e técnicas multivariadas utilizadas, no Capítulo 3 apresenta-se o método de segmentação proposto, juntamente com resultados correlatos, o Capítulo 4 apresenta-se a classificação de imagens utilizando planos de entropia e análise multivariada e no Capítulo 5 as conclusões do estudo e sugestões para trabalhos futuros.

2. MATERIAIS E MÉTODOS

Este capítulo descreve o local de estudo, a coleta das amostras praticamente intactas, a tomografia, as medidas de entropia e análise multivariada usadas na metodologia proposta para classificar e segmentar as imagens.

2.1 Local, coleta e preparo das amostras

Dois fragmentos de mata nativa contornados pelo cultivo de cana-de-açúcar foram visitados e amostrados. As matas de Piedade e do Cumbe, localizadas no nordeste brasileiro, em Pernambuco, respectivamente nos municípios de Igarassu-PE e Araçoiaba-PE, ambos inseridos no território de uma indústria sucroalcooleira, a Usina São José.

O fragmento de mata atlântica do Engenho Piedade (Figura 2), pertencente a Usina São José, estando localizado no município de Igarassu-PE ($7^{\circ}49'16''$ a $7^{\circ}50'54''$ S; $34^{\circ}49'26''$ a $34^{\circ}00'35''$ O), distante aproximadamente 50 Km da capital Recife, é reconhecido pela UNESCO como uma Zona Núcleo da Reserva da Biosfera da Mata Atlântica (RBMA) brasileira ³⁸ e protegido como Reserva Ecológica da Região Metropolitana do Recife pela Lei Estadual 9.989/87 ³⁹, sendo e em seguida reclassificado como Refúgio da Vida Silvestre através da Lei Estadual 14.324/11 ⁴⁰. Dispõe de uma área de aproximadamente 299 hectares segundo o Cadastro Nacional de Unidades Conservação ⁴¹.



Figura 2. Foto aérea e polígono aproximado do fragmento da Mata de Piedade no município de Igarassu - PE, terras da Usina São José.

De acordo com trabalhos passados realizados nesse fragmento, boa parte dos fragmentos surgiram durante intensa destruição da mata atlântica entre os anos de 1970 e 1980, quando ocorreu uma crise no setor petrolífero e o Proálcool (Programa Nacional do Álcool) foi criado como uma iniciativa do governo brasileiro para enfrentar esta crise e incentivar a produção de álcool combustível^{30; 42}. Sugere-se que a Mata de Piedade tenha sido formada bem antes desse período, sendo estimada, embora sem muita precisão, uma idade maior que 30 anos de uso do solo sob as condições naturais. Dada a razoável conservação neste fragmento, muitos estudos foram e estão sendo realizados nesta localidade. Dentre as principais temáticas estão, estudos florísticos^{43; 44; 45; 46; 47}, efeitos da fragmentação na dinâmica de produção de serapilheira⁴⁸, conservação e distribuição de espécies de pássaros⁴⁹.



Figura 3. Estrada que separa a Mata de Piedade e o cultivo de cana-de-açúcar na Usina São José no município de Igarassu - PE

O outro local fica próximo do distrito de Três Ladeiras na cidade Igarassu-PE, trata-se do fragmento de mata do Engenho Cumbe (7°45'51'' a 7°46'48''S; 35°02'13'' a 35°02'50''O), ou simplesmente Cumbe, também pertencente a Usina São José. Embora o distrito vizinho ao fragmento faça parte do município de Igarassu-PE, boa parte do território do fragmento do Cumbe encontra-se politicamente no município de Araçoiaba-PE. Não existem muitos estudos neste fragmento, o qual apresenta um menor grau de conservação se comparado com o fragmento de Piedade. Apenas um estudo nas proximidades, mas em outro fragmento próximo chamado Cumbe de Baixo com idade média estimada de 12 anos foi

encontrado ⁵⁰. O polígono aproximado do fragmento do Cumbe é apresentado na Figura 4 a seguir.



Figura 4. Foto aérea e polígono aproximado do fragmento da Mata do Cumbe no município de Araçoiaba – PE, terras da Usina São José.

Nestes fragmentos e nos canaviais vizinhos aos mesmos, os pontos escolhidos foram escavados com facas, assegurando ao máximo a estrutura original do ambiente no interior de cilindros de PVC de 7,5 cm de altura por 7,5 cm de diâmetro acoplados a uma lâmina (Figura 5).



Figura 5. Coleta física das amostras de solo por meio da escavação com facas e armazenamento da estrutura original do solo em cilindro de PVC para posterior tomografia

Ao todo, 36 cilindros nas profundidades de 0-10 cm e 10-20 cm, foram coletados no período de 14/03 a 27/03/2016 em 21 pontos, 10 de mata nativa e 11 de cultivo de cana-de-açúcar. As coordenadas geográficas e a classificação dos solos de cada um destes pontos encontram-se nos apêndices desta tese e uma descrição com a nomenclatura para cada amostra é apresentada na Tabela 1, a seguir. A partir dessa nomenclatura é possível saber em que uso e profundidade foram coletadas, pois imagens de cana apresentam A ou C e imagens de mata apresentam B ou D em seus nomes. Nomes de imagens terminados por 0 ou 1 foram coletados na camada de 0-10 cm e os terminados por 2 foram coletados na camada de 10-20 cm.

Tabela 1. Caracterização das amostras coletadas

Uso	Profundidade (cm)	Imagens (Quantidade)
Cana	0 - 10	A1001, A2001, A3001, A4001, A5001, A6001, A7001, XA200, C1001, C3001 (10)
	10 - 20	A1102, A2102, A3102, A4102, A5102, A6102, A7102, C1102, C2102 (9)
Mata	0 - 10	B1001, B2001, B3001, B4001, B5001, B6001, XB300, XB400, D1001, D2001 (10)
	10 - 20	B1102, B2102, B3102, B4102, B5102, B6102, D2102 (7)
Total		36

Após devidamente coletadas as amostras foram secas em uma estufa a 40° C para retirada do efeito da água no escaneamento ou tomografia das amostras, a qual é melhor descrita a seguir.

2.3 Escaneamento das amostras via CT

A análise tomográfica fornece uma poderosa forma de visualizar componentes da estrutura dos solos de forma próxima as encontradas no ambiente natural ou original, através de uma escala de cinza especialmente desenvolvida para estudos com tal técnica, permitindo avaliação global da estrutura do solo e do espaço poroso³. A aquisição preliminar dos dados foi realizada através de um microtomógrafo de raios-x de terceira geração modelo Nikon XT H 225 ST (Figura 6), utilizando uma tensão de 150kV, corrente 180 µA e tempo de exposição de 500 ms, com uma resolução de 45 µm para os *voxels*. Utilizou-se ainda, um filtro de cobre

com espessura de 0,5 mm para minimizar os fótons de baixa intensidade, causadores dos artefatos do tipo feixe de endurecimento (*beam hardening*)⁵¹ e maximizar o contraste entre diferentes fases do solo (sólidos e ar).

Após o escaneamento do volume total do cilindro, na aquisição preliminar, um sub-volume de interesse foi definido e reconstruído por meio do *software* CTPro 3D XT 3.0.3 (Nikon Metrology NV), dando ênfase à parte mais central do cilindro, evitando assim alguma influência das bordas. As projeções axiais 2D reconstruídas mantiveram a mesma resolução espacial da aquisição, ou seja, de 45 μm , sendo salvas a uma resolução radiométrica de 16 bits e o volume final ficou com 790 projeções (*stacks*) de 790 x 790 *pixels*, totalizando assim um volume final de 790³ *voxels*.



Figura 6. Amostra de solo no interior do tomógrafo

Reconstruídas as imagens, alguns tratamentos são necessários com a finalidade de garantir uma padronização e qualidade das informações. A padronização visa tornar interpretável os valores assumidos pelos *voxels*, definindo seus valores em relação a radiodensidade real de materiais proposta por Hounsfield⁵² (Figura 7), adotando-se assim os valores dos *voxels* em unidades de Hounsfield (HU). Visualmente trata-se de uma escala de cinza especialmente adotada em estudos de tomografia, variando entre preto (ar), tons de cinza e branco (sólidos de elevada radiodensidade). A quantidade de tons de cinza depende da resolução radiométrica, nas imagens aqui reconstruídas existe a possibilidade de apresentar até 2¹⁶ cores. Utilizou-se o *software* VGStudio MAX 2.2 (Volumegraphics, Heidelberg, Alemanha), para padronização em HU e aplicação o filtro de Gauss para a minimização de ruídos. Os valores assumidos foram em seguida adicionados em 1000 HU, ou seja, o ar passou a ser representado pelo valor de 0 HU e a água valor igual a 1000 HU.

Além dos artefatos do tipo feixe de endurecimento, citados anteriormente, a qualidade dos dados tomográficos pode ser afetada por outros artefatos, dentre os quais vale destacar os artefatos de anéis que geralmente são originados pelo mau funcionamento dos sensores dos detectores ou uma calibração incorreta dos mesmos ⁵³. Algumas das imagens obtidas apresentaram estes defeitos, os quais podem comprometer fortemente as análises e resultados. Para correção destes anéis utilizou o procedimento sugerido por Munch ⁵⁴ e Zhouping ⁵⁵. A Figura 8 mostra uma fatia defeituosa e a Figura 9 a mesma fatia corrigida. O procedimento de redução destes anéis pode ser encontrado livremente no *software* ImageJ ⁷ através de uma combinação de uso dos plug-ins *Polar Transformer* e *XStripes Filter*.

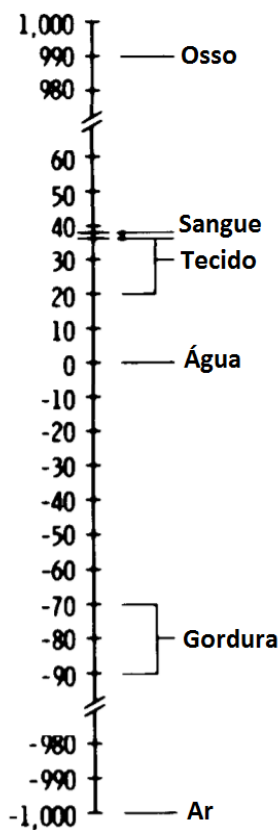


Figura 7. Escala de Hounsfield

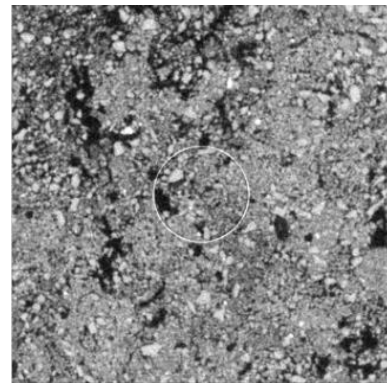


Figura 8. Imagem com artefato de anel

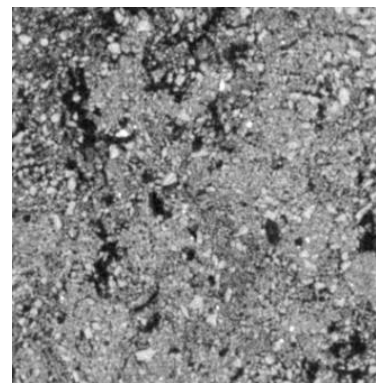


Figura 9. Imagem corrigida

A seguir são apresentados aspectos da teoria da informação, que é o ramo da matemática que descreve como a incerteza deve ser quantificada, manipulada e representada, tendo como quantidade fundamental a entropia ^{56; 57}. Nesta tese, duas medidas de entropia são utilizadas, as entropias de Shannon ⁵⁶ e amostral ⁵⁸.

2.4 Informação de Shannon (SHE)

Considere um conjunto consistindo de n eventos $\{x_1, x_2, \dots, x_n\}$. A incerteza acontece a partir do momento que não podemos ter certeza sobre qual desses eventos irá ocorrer. Porém, baseado em algum conhecimento prévio sobre estes eventos esta incerteza pode ser maior ou menor, por exemplo, se soubermos de duas informações, uma delas, o número total de eventos e a outra, quais destes eventos têm probabilidade de ocorrência diferente de zero. Ou seja, caso conhecida a distribuição de probabilidade de $\{x_1, x_2, \dots, x_n\}$, esta fornece uma quantidade de informação que pode reduzir a incerteza quanto a ocorrência destes eventos. A entropia é então medida levando-se em consideração todas as informações disponíveis, sugerindo assim que a mesma é condicionada a informações sobre possíveis resultados do experimento ^{57; 59}.

Trata-se então a entropia como uma medida de incerteza associada a uma variável aleatória. Para uma variável aleatória X com $p(x) = Pr(X = x)$, $x \in X$, a entropia de Shannon é definida como:

$$H(X) = - \sum_{x \in X} p(x) * \log_2 p(x)$$

onde H é uma medida da informação (mais informação resulta em entropia inferior e vice-versa), $p(x)$ é a função de densidade de probabilidade da variável x . Neste caso os valores de entropia são dados em *bits* porque a base do logaritmo foi assumida como sendo igual a 2 e a probabilidade $p(x)$ é baseada na frequência empírica da variável x que possui n resultados possíveis ⁶⁰.

O uso da informação fornecida pelas entropias para estudo das transformações ocorridas nos solos tem longa data ⁶¹ e tem sido muito intensa até o presente momento. Por se tratar de um meio com fases porosa e sólidas extremamente complexas, porém, muito bem definidas, além de utilizados sob diversos contextos, os solos permitem assim muitas aplicações buscando entender o fluxo ou armazenamento de líquidos ^{18; 62; 63; 64}, análise de riscos em engenharia ⁶⁵, degradação do meio ambiente ^{66; 67}, técnicas de *threshold* visando o tratamento e quantificação em imagens tomográficas ^{4; 16}, a estrutura do solo propriamente dita ^{19; 68}.

2.5 Entropia amostral (SE)

Outra relevante medida de informação é a entropia amostral. É definida como sendo o logaritmo natural negativo da probabilidade condicional de que duas subsequências similares (dentro de um nível de tolerância r), para m pontos, permanecem similares para $m + 1$ pontos. O algoritmo para cálculo da entropia amostral introduzido por Richman e Moorman ⁵⁸ é mostrado a seguir.

Seja $u(1), u(2), \dots, u(N)$ uma sequência de valores ou série temporal de tamanho N . O cálculo da SE (m, r, N) segue as seguintes etapas:

- (1) são construídos $N - m + 1$ vetores de tamanho m , $x_m(i): x_m(1), x_m(2), \dots, x_m(N - m + 1)$, onde $1 \leq i \leq N - m + 1$;
- (2) calcula-se a máxima distância escalar como a distância absoluta máxima entre os elementos de $x(i)$ e $x(j)$:

$$d[x(i), x(j)] = \text{máx}(|u(i + k - 1) - u(j + k - 1)|), \text{ onde } k = 1, \dots, m,$$

- (3) definido um grau de tolerância para similaridade r , como um número real positivo. Para um dado $x_m(i)$, conta-se o número de vetores j ($j = 1 \sim N - m, j \neq i$) os quais $d[x(i), x(j)] < r$ é denotado por $B_i^m(r)$.

$$\text{Logo, para } i = 1 \sim N - m, B^m = \frac{\sum_{i=1}^{N-m} B_i^m(r)}{(N-m)},$$

- (4) similarmente, dado $x_{m+1}(i)$, a contagem do número de vetores j ($j = 1 \sim N - m, j \neq i$) os quais $d[x(i), x(j)] < r$ é denotado por $A_i^m(r)$.

$$\text{Logo, para } i = 1 \sim N - m, A^m = \frac{\sum_{i=1}^{N-m} A_i^m(r)}{(N-m)},$$

- (5) teoricamente, a entropia amostral vetores de tamanho m a um nível de tolerância r é definida como,

$$SE(m, r) = -\ln \left[\frac{A^m(r)}{B^m(r)} \right]$$

A escolha do parâmetro de tolerância r é apontado e utilizado em alguns estudos com variações entre 10 e 20% do desvio padrão das sequências analisadas ⁶⁹. A seguir apresenta-se o coeficiente de correlação de Pearson.

2.6 Coeficiente de correlação de Pearson

Karl Pearson descreveu os fundamentos do coeficiente de correlação em 1896 ⁷⁰, porém, apesar de sua idade secular, esta medida continua utilizada fortemente até hoje ⁷¹. Denotado por r , é uma medida que mede a força da relação linear entre duas variáveis, podendo assumir qualquer valor entre -1 e 1. Existe portanto, a possibilidade do mesmo ser positivo ou negativo e a força ou peso desta correlação será avaliada pela proximidade destes extremos, ou seja: $r = 0$ indica que não há correlação linear, por outro lado, $r = 1$ ou $r = -1$ indicaram respectivamente correlação linear perfeita positiva e negativa.

O cálculo do coeficiente de correlação entre duas variáveis X e Y é obtido da seguinte forma:

- (1) Seja zX_i e zY_i os valores de X e Y padronizados com média 0 e desvio padrão 1. Então cada valor zX_i e zY_i passam pela transformação,

$$zX_i = [X_i - \bar{X}] / \sqrt{\text{var}(X)}$$

$$zY_i = [Y_i - \bar{Y}] / \sqrt{\text{var}(Y)}$$

- (2) O coeficiente de correlação é definido como a soma dos produtos dos pares (zX_i, zY_i) dividido por $n - 1$, ou seja,

$$r = \frac{\sum_{i=1}^n zX_i * zY_i}{n - 1}$$

onde n é o tamanho da amostra.

A seguir são apresentadas as técnicas multivariadas de análise de componentes principais e análise de agrupamentos.

2.7 Análise de componentes principais (ACP)

A ACP é uma técnica multivariada amplamente utilizada nos mais diversos contextos de pesquisa e na ciência do solo pode-se destacar, a contaminação em solos ^{72; 73}, riscos agrícolas ⁷⁴, armazenamento de água e densidade do solo ⁷⁵ e análise espacial de propriedades dos solos ⁷⁶. Trata-se de uma técnica muito popular na busca por padrões em dados com muitas dimensões, transformando um conjunto de variáveis originais em outro conjunto de variáveis de mesma dimensão denominadas de componentes principais, as quais consistem de

uma combinação linear de todas as variáveis originais, reduzindo a dimensão dos dados com menor perda possível de informação, agrupando os indivíduos com base na variância conjunta de suas características. Um roteiro desta metodologia é apresentado a seguir:

- (1) o ponto de partida da ACP é o conjunto de n exemplos ou observações medidos em p variáveis, organizadas matricialmente, as quais em geral são inter-correlacionadas. Nenhuma variável é designada como dependente, e nenhuma observação dispõe de um grupo assumido a priori. (Figura 10.1),
- (2) estes dados são em seguida padronizados com variância 1 em uma nova matriz Z (Figura 10.2),
- (3) a partir de Z é obtida a matriz de autocorrelações R (Figura 10.3),
- (4) a partir de R são obtidos os autovalores ou raízes λ da equação característica $\det[R - \lambda I]$ e seus respectivos autovetores, sendo, por exemplo, \tilde{a}_i o autovetor correspondente ao autovalor λ_i (Figura 10.4),
- (5) em seguida as informações são expressas através de componentes principais, as quais expressam combinações lineares capazes de explicar cada uma algum percentual da variabilidade dos dados (Figura 10.5) ⁷⁷.
- (6) De forma complementar, é possível visualizar a dispersão dos tratamentos em função dos componentes principais (Figura 10.6).

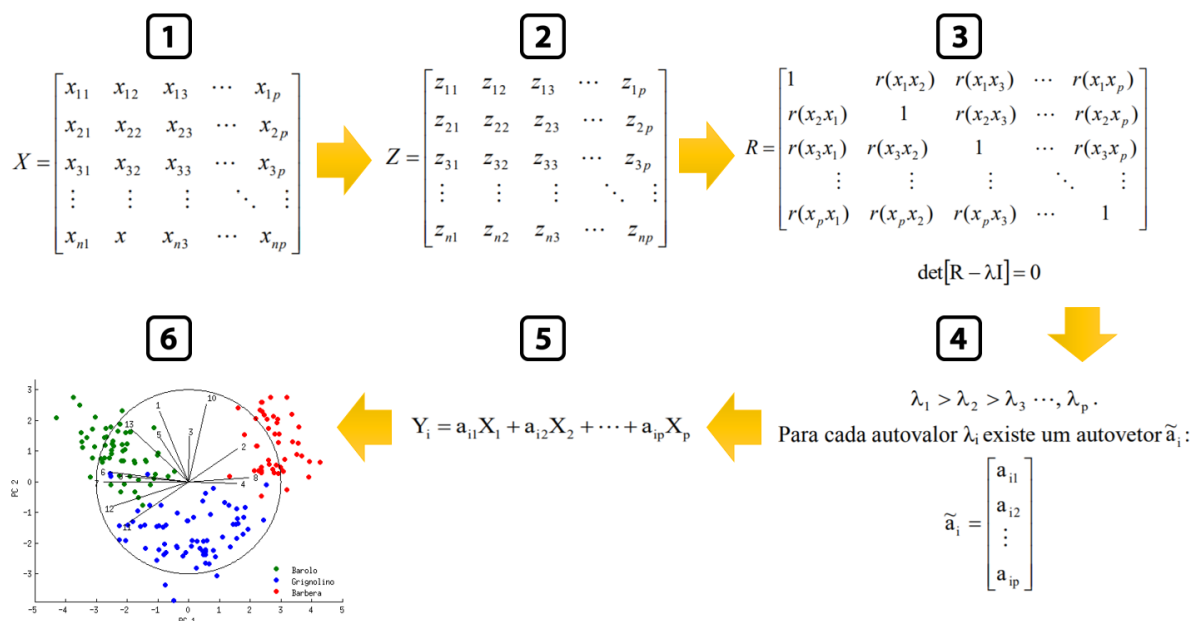


Figura 10. Processo geral de uma Análise de componentes principais

Após concluído o processo acima é possível obter a contribuição de cada componente principal Y_i , a qual é medida em percentual da variabilidade explicada pela referida componente. A contribuição C_i da componente principal Y_i será então dada por:

$$C_i = \frac{Var(Y_i)}{\sum_{i=1}^p Var(Y_i)} * 100 = \frac{\lambda_i}{\lambda_1 + \lambda_2 + \dots + \lambda_p} * 100$$

Maiores informações no âmbito de aspectos matemáticos e de algumas propriedades das componentes principais podem ser encontradas na obra de Johnson e Wichern ⁷⁸.

2.8 Análise de agrupamentos (AAg)

A AAg consiste de uma metodologia multivariada estatística que tem como propósito de avaliar a existência de grupos internamente homogêneos, porém, relativamente heterogêneos entre si, ou seja, dada uma determinada amostra de n indivíduos ou objetos, cada um com medições em p variáveis, a AAg poderá a partir de algum critério de (dis)similaridade ou distância, combinada a um método de ligação ou agrupamento buscar uma alocação destes indivíduos em g grupos ⁷⁸. Esta análise é geralmente organizada em quatro etapas:

- (1) Cálculo de uma matriz de dissimilaridade: Nesta tese, cinco medidas de dissimilaridade são usadas: Euclidiana (*Euclidean*), *Camberra*, *Minkowski*, *Manhattan* e Chebyshev (*Maximum*). Antes do cálculos destas matrizes de distâncias recomenda-se a padronização dos dados, subtraindo cada valor pela média de sua variável e dividindo pelo desvio padrão da variável, de modo a deixar todas as dimensões em uma mesma unidade de medida ⁷⁹. Um resumo dessas medidas de distâncias é apresentado na Tabela 2 a seguir:

Tabela 2. Medidas de dissimilaridade

Distância	Descrição	Fórmula
Euclidiana (<i>Euclidean</i>)	É a distância geométrica entre duas imagens em um espaço multidimensional. A distância entre duas observações (i e j) corresponde à raiz quadrada da soma dos quadrados das diferenças entre os pares de observações ou imagens de i e j para todas as p variáveis ⁷⁸ .	$d_{ij} = \ x_i - x_j\ $ $= \sqrt{(x_i - x_j)^t (x_i - x_j)}$ $= \sqrt{\sum_{k=1}^p (x_{ik} - x_{jk})^2}$

Continuação da Tabela 2.

Canberra	Calcula a soma das diferenças fracionárias entre as coordenadas de pares de objetos .	$d_{ij} = \sum_{k=1}^p \frac{ x_{ik} - x_{jk} }{(x_{ik} + x_{jk})}$
Minkowski	Generalização das demais distâncias, pois as distâncias são basicamente normas de vetores. $\lambda = 1$, <i>Manhattan</i> ; $\lambda = 2$, <i>Euclidean</i> ; $\lambda \rightarrow \infty$, Máxima distância (<i>Maximum</i> ou <i>Chebychev</i>).	$d_{ij} = \left(\sum_{k=1}^p x_{ik} - x_{jk} ^\lambda \right)^{1/\lambda}$

(2) Criação do dendograma com base em métodos de agrupamento: Obtidas as distâncias, sete métodos de agrupamento hierárquico são realizados sobre as mesmas: *Ward*^{80: 81}, *Average*, *Complete*, *Centroid*, *Median* e *Mcquitty*⁸², os quais são também são conjuntamente detalhados no trabalho de Murtagh⁸³. Deste modo, a partir de cada uma das cinco possíveis matrizes de distância, são construídos sete diferentes agrupamentos.

Nos métodos hierárquicos, uma vez que for construído um agrupamento, as observações ou conjunto de observações iterativamente designadas a cada grupo não poderão mais participar ou serem designadas a outro grupo, além de não ser possível definir previamente o número de grupos desejado, sendo necessária para definição do número de grupos uma avaliação adicional, geralmente a partir do dendograma (Figura 11), que é uma estrutura gráfica, na forma de árvore capaz de mostrar o processo de construção.

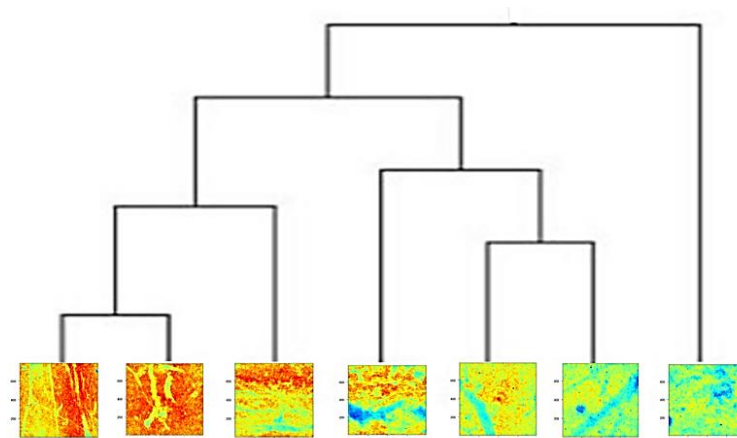


Figura 11. Exemplo de dendograma com classificação de imagens

Os métodos hierárquicos ainda podem ser divididos em aglomerativos, indicando que cada objeto é inicialmente indicado como um *cluster* e após iterativas comparações os grupos são formados (Figura 12.a), ou métodos partitivos, os quais iniciam com todos os objetos em um único grande cluster e após iterativas comparações os grupos são formados (Figura 12.b).

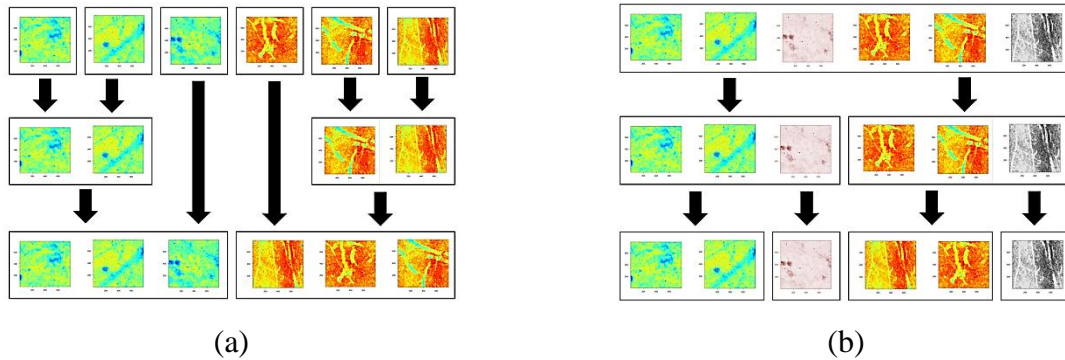


Figura 12. Método de agrupamento aglomerativo (a) e partitivo (b)

Já nos métodos não hierárquicos, existe a possibilidade de uma determinada observação ou um conjunto delas serem designadas a mais de um grupo, além de ser permitida a definição prévia do número de grupos desejados.

- (3) Seleção da melhor combinação distância-agrupamento: Ou seja, da melhor combinação entre matriz de dissimilaridade e método de agrupamento é definida pela maior correlação cofenética introduzida por Sokal e Rohlf ^{84; 85}, a qual tem o poder de avaliar o quanto o dendograma preserva de similaridade com as distâncias observadas na matriz de dissimilaridade original.
- (4) Definição da partição ótima de grupos: Uma vez definido o dendograma com a melhor combinação distância-agrupamento, um dos grandes desafios na AAg é a definição é o recorte adequado do mesmo, que será aqui obtido pelo critério de Krzanowski e Lai ⁸⁶.

Os Capítulos 3 e 4 apresentam a proposta de planos de entropia na segmentação e classificação de imagens de tomografia de solos.

3. ENTROPIC PLANS CORRELATIONS (EPC)

Embora o uso de entropias na segmentação de imagens como um todo não seja novidade, a metodologia aqui proposta para este fim ainda não foi utilizada por outros estudos e devido à similaridade com PVESeg, este servirá como comparativo e até mesmo calibração para o EPCSeg. Este capítulo dispõe do detalhamento da proposta e sua validação, que podem também ser interpretados como resultados preliminares desta tese. A seguir explica-se a idéia e consistência da construção de planos de entropia por meio de imagens tomográficas 3D de solos.

3.1 Planos de entropia

A ideia de construir planos de entropia é o de resumir o grau de desordem de uma imagem 3D em um plano 2D formado por entropias. Seja uma imagem 3D com r fatias (na direção da gravidade) de dimensões $r \times r$. Seja v_{xyz} o *voxel* localizado na posição horizontal x ($1 \leq x \leq r$) e vertical y ($1 \leq y \leq r$) na fatia z ($1 \leq z \leq r$). Ou seja, o primeiro elemento do canto superior esquerdo da primeira fatia é o v_{111} e o último elemento do canto inferior esquerdo da última fatia é o v_{rrr} .

Construindo uma sequência de valores na direção da gravidade, sempre utilizando as mesmas posições de cada uma das fatias e aplicando sobre estas sequências alguma medida de entropia, pode-se refletir o valor desta medida em um plano, o qual definiremos aqui como plano de entropias ou plano entrópico (Figura 13).

Definimos E_{xy} como sendo uma medida de entropia refletida para a posição horizontal x e vertical y do plano de entropias. Ou seja, E_{11} é medida de entropia calculada sobre a sequência de valores $\{v_{111}, v_{112}, v_{113}, \dots, v_{11r}\}$ formada por todos os primeiros *voxels* de cada fatia. Duas medidas de entropia foram utilizadas na criação dos planos. A entropia amostral, que fornece um único plano obtido sobre a escala de cinza original da imagem e a informação

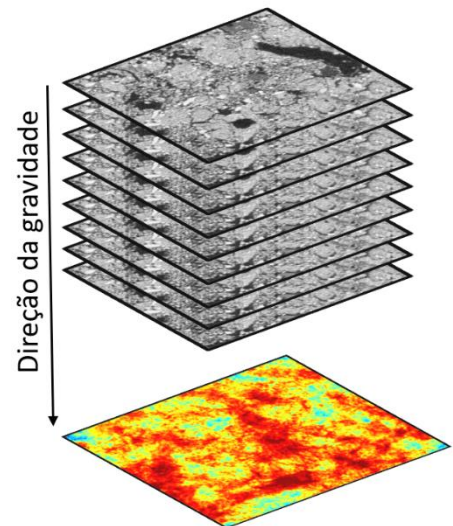


Figura 13. Ilustração da obtenção de um plano entrópico pela penetração de uma imagem na direção da gravidade por medidas de entropia

de Shannon que a partir da seleção de limiares menores ou iguais a 250 HU até limiares menores ou iguais 2600 HU forneceu 2350 planos para cada imagem.

Utilizando inicialmente as primeiras 1000 HU, as quais serão aqui definidos como *voxels* de baixa densidade (VBD), como referência para os planos de informação de Shannon, ou seja, valores representativos entre o ar e água, para imagens de mata e cana, ficou evidente que neste limiar, solos de mata fornecem mais desordem, ou seja, dispõem de uma maior quantidade de *voxels* representativos nesta seleção. Por outro lado, se consideramos toda escala de cinza presente nas imagens, os solos de cana são mais desordenados. Na Figura 14 a seguir, apresentam-se estes planos para duas imagens de cana (XA200 E A7001) e duas imagens de mata (B5001 e B4001).

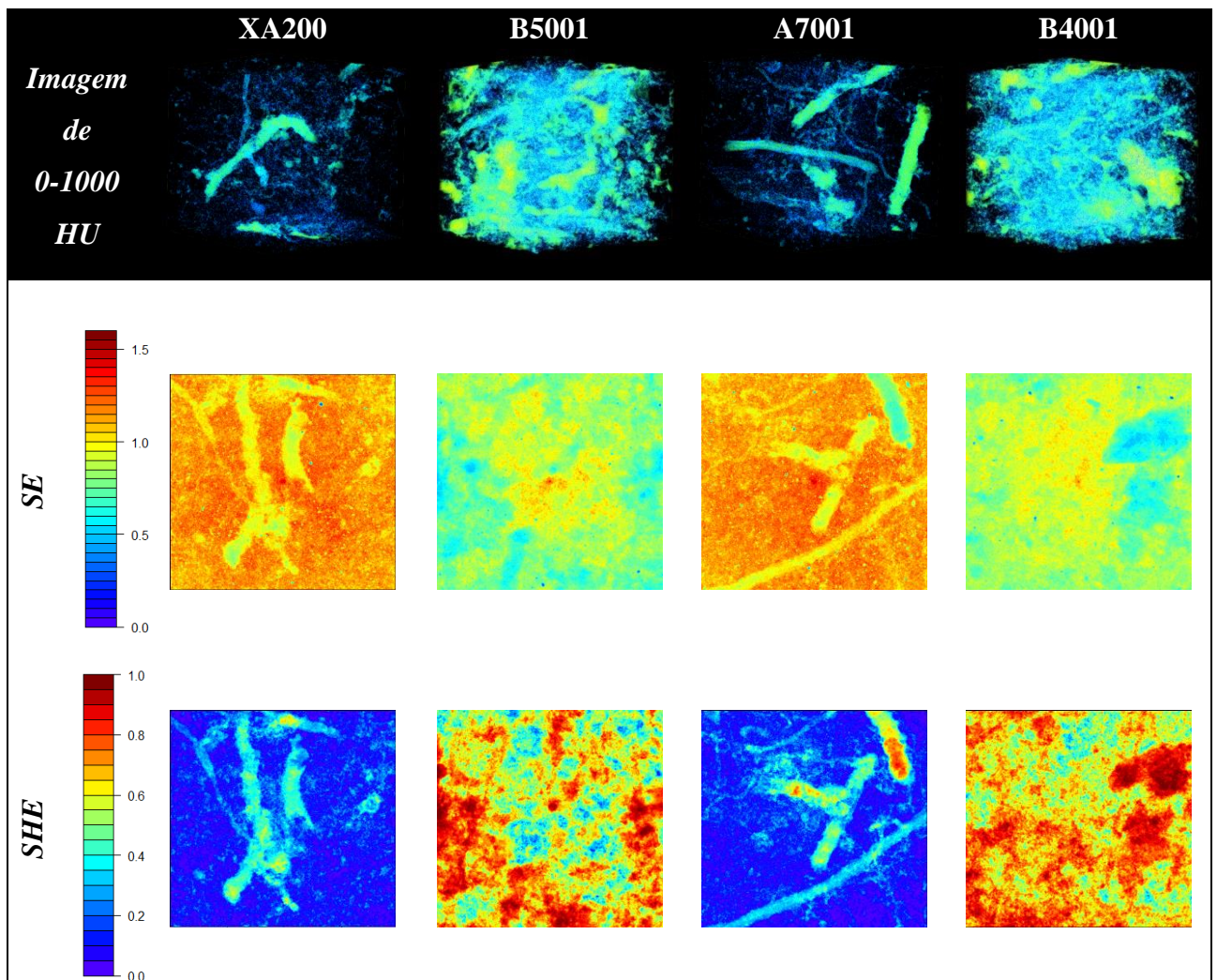


Figura 14. Planos de entropia amostral aplicados na escala de cinza e de Shannon aplicados em voxels de baixa densidade em imagens de tomografia de solos de mata e cana

Ficou evidente que os planos SHE utilizando o *threshold* acumulado até VBD parecem estar correlacionados negativamente com os planos SE, uma vez que áreas com maior desordem em planos SHE parecem coincidir com áreas menos desordenadas em planos SE. Deste modo, pode-se suspeitar que a fase de vazios contribui para uma maior ou menor desordem global, ou seja, solos de mata com maior quantidade de VBD são menos desordenados que solos de cana.

Um estudo detalhado da correlação entre planos SHE para *thresholds* acumulados de 250 a 2600 HU e o plano SE para imagens de solos é apresentado a seguir, buscando avaliar a evolução da correlação entre estes planos no decorrer do *threshold* acumulado. Adiante apresenta-se também a utilidade destas correlações entre planos para obtenção da curva de multissegmentação, que tem importante papel na identificação de *voxels* puros, similarmente aos histogramas gerados em PVE.

3.2 Curva de multissegmentação por EPC e histograma por PVE

Das 36 imagens anteriormente apresentadas na Tabela 1, seis imagens de cada uso foram aleatoriamente designadas para geração de todas as correlações entre planos SHE e o plano SE no decorrer do *threshold* acumulado. Em resumo, a construção da curva de multissegmentação seguiu os passos:

- Seja t um determinado *threshold* assumido na segmentação. Definidos um limite inferior de *threshold* L_i e um limite superior de *threshold* L_s , as imagens foram segmentadas com $t \leq L_i$ até $t \leq L_s$, com acréscimos unitários. Em cada uma destas $L_s - L_i + 1$ operações as imagens são penetradas por entropias SE com dois estados possíveis.
- O valor da correlação linear de Pearson entre cada um dos $L_s - L_i + 1$ planos SE e o plano SE é obtido, conservando as respectivas posições assumidas na matriz de dados destes planos.
- Uma curva de multissegmentação deve ser plotada, de preferência com o valor assumido pela correlação entre planos no eixo das abcissas e o valor do *threshold* nas ordenadas.

O comportamento geral da curva de multissegmentação construída pela correlação de planos de entropia (EPC), fornece um vale e um pico muito evidentes. Uma investigação detalhada da vizinhança destes *thresholds* representantes de *Vale* e *Pico*, permitiu constatar

que da mesma forma que o histograma em PVE fornece pontos de corte para fases puras de vazios e sólidos, informação similar pode ser obtida através da curva de multissegmentação por EPC. Porém neste caso, a vantagem de EPC sobre PVE se dá pela não necessidade de forte interação do usuário com “chutes iniciais” na tentativa de tornar viáveis os picos no histograma pela escolha subjetiva de coeficientes de variação. Em todas as imagens, *Vale* e *Pico* foram obtidos com êxito sem qualquer transformação ou procedimentos de tentativa e erro. A seguir, na Figura 16, apresenta-se a curva de multissegmentação via EPC para seis solos de cana (A1001, A4001, A5001, A5102, A6001, XA200) e na Figura 16, seis solos de mata (B1001, B3001, B4102, B5001, B6001, XB300):

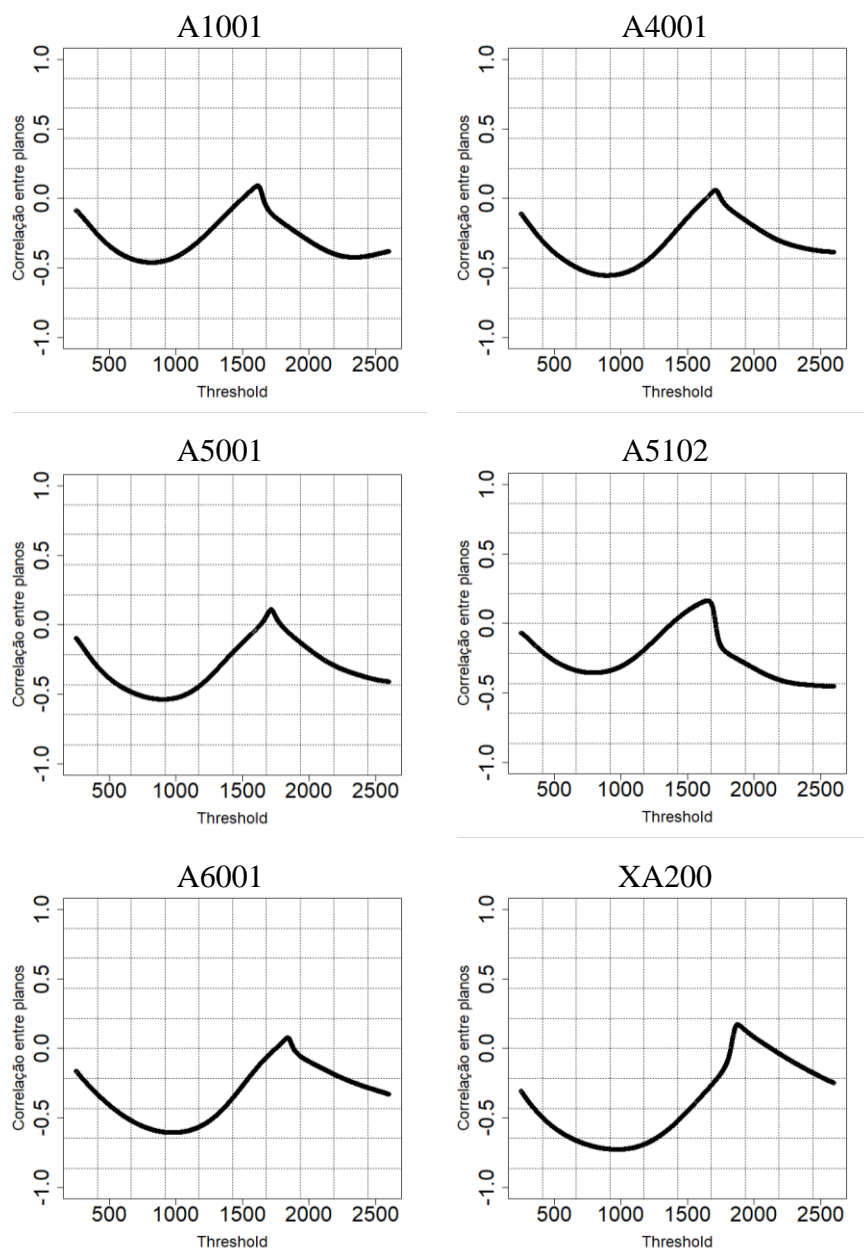


Figura 15. Curvas de multissegmentação obtidas pela correlação de planos de entropia sob variados *thresholds* acumulados em solos de cana

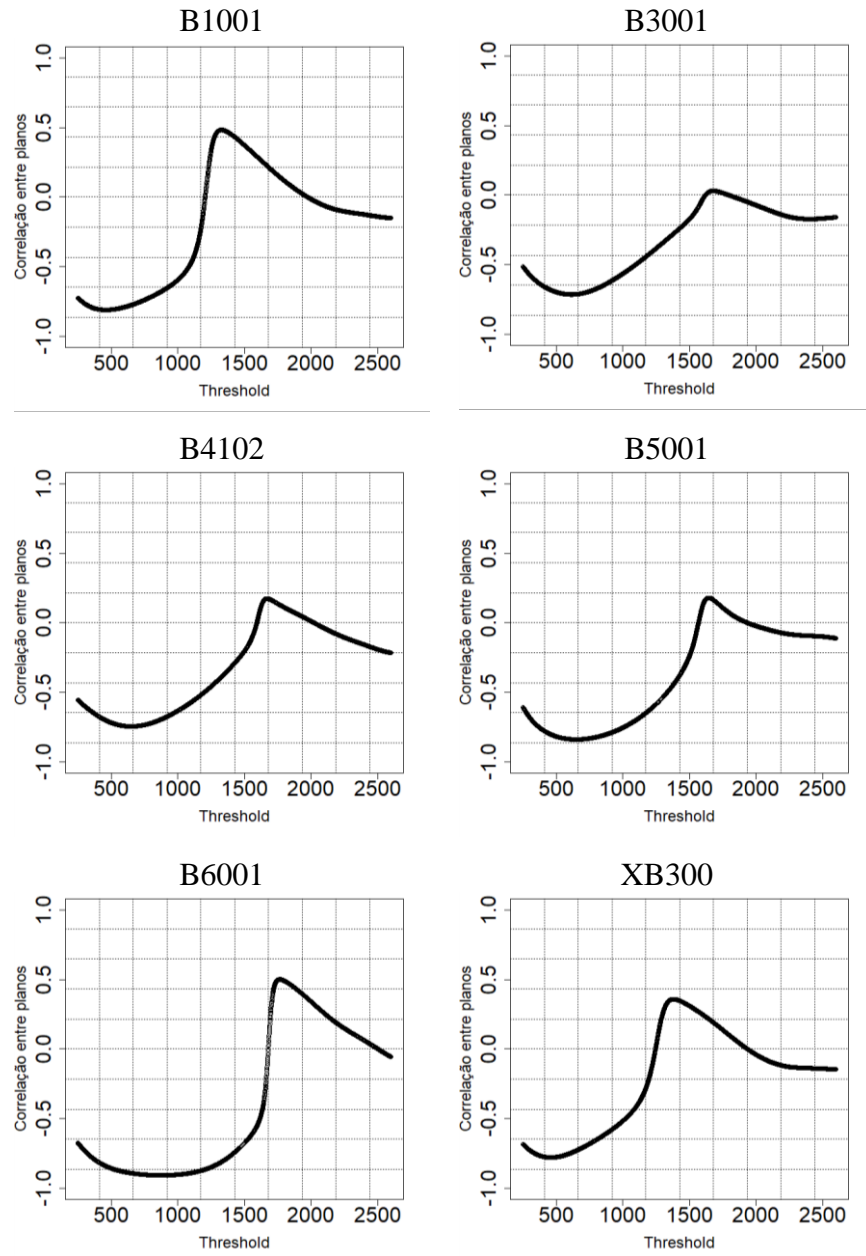


Figura 16. Curvas de multissegmentação obtidas pela correlação de planos de entropia sob variados *thresholds* acumulados em solos de mata

Estas mesmas 12 imagens foram submetidas a extração de *voxels* puros (PVE) e os valores dos picos do histograma *P1* (delimitante da fase de vazios) e *P2* (delimitante da fase de sólidos) foram obtidos. O próximo passo é a comparação entre delimitadores das fases puras de vazios (*Vale* por EPC e *P1* por PVE) e de sólidos (*Pico* por EPC e *P2* por PVE). A Tabela 3 mostra os valores de *Vale* e *Pico* por EPC e a Tabela 4 mostram os valores de *P1* e *P2* por PVE.

Tabela 3. Vale e Pico por EPC

<i>Imagem</i>	<i>Vale</i>	<i>Pico</i>
<i>A1001</i>	807	1613
<i>A4001</i>	896	1710
<i>A5001</i>	903	1714
<i>A5102</i>	794	1649
<i>A6001</i>	973	1839
<i>XA200</i>	968	1876
<i>B1001</i>	456	1326
<i>B3001</i>	610	1678
<i>B4102</i>	648	1669
<i>B5001</i>	650	1644
<i>B6001</i>	875	1763
<i>XB300</i>	457	1381

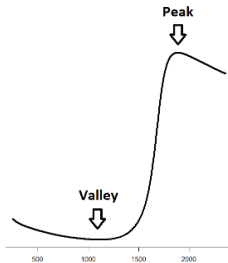
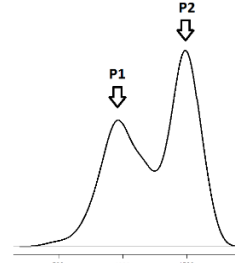


Tabela 4. P1 e P2 por PVE

<i>Imagem</i>	<i>P1</i>	<i>P2</i>
<i>A1001</i>	878	1116
<i>A4001</i>	1001	1221
<i>A5001</i>	1040	1040
<i>A5102</i>	989	1194
<i>A6001</i>	1209	1411
<i>XA200</i>	1100	1395
<i>B1001</i>	455	724
<i>B3001</i>	655	1220
<i>B4102</i>	780	1202
<i>B5001</i>	753	898
<i>B6001</i>	1089	1359
<i>XB300</i>	324	739



3.3 Análise do comportamento de curva de multissegmentação

Como visto anteriormente, EPC pode também delimitar similarmente a PVE as regiões de *voxels* puros das imagens. Consideraremos deste ponto em diante *voxels* menores ou iguais ao *Vale* como puros vazios (*pv*) e os maiores ou iguais ao *Pico* como puros sólidos (*ps*) e os inseridos na região de transição, ou seja, aqueles entre *Vale* e *Pico*, como mistos (*m*) (Figura 17).

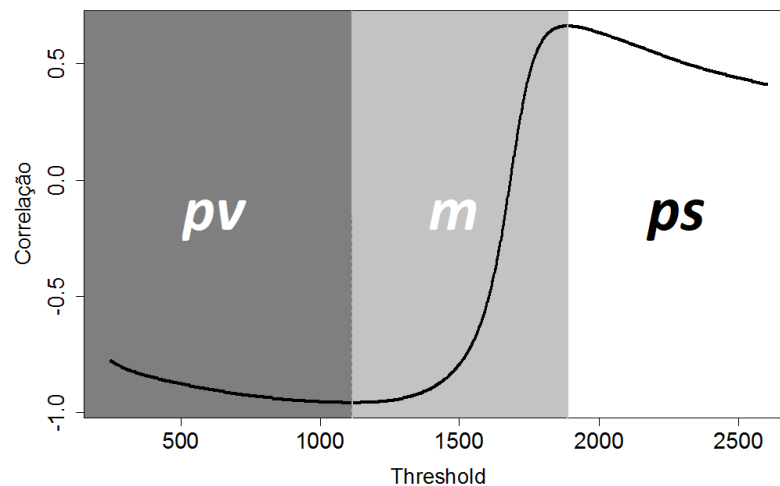
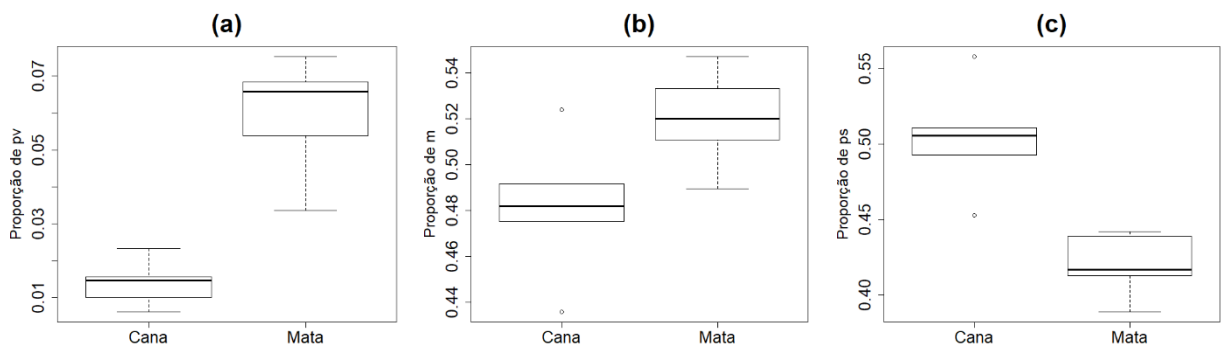


Figura 17. Classificação de *voxels* por meio da curva de multissegmentação proposta pelo método de correlação de planos de entropia (EPC), onde *pv* são *voxels* puros vazios, *m* são *voxels* mistos e *ps* são *voxels* puros sólidos

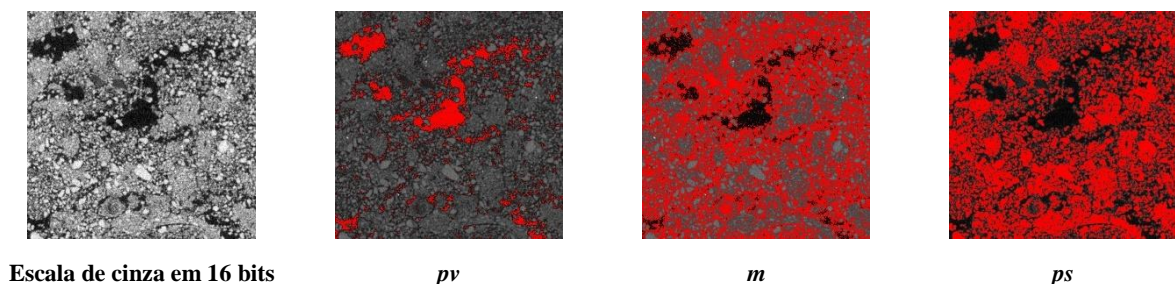
Avaliando a proporção de presença de *voxels pv*, evidenciou-se uma superioridade em solos de mata, essa superioridade em mata também foi observada em *voxels m*, porém para *voxels ps* os solos de cana foram superiores. Estas diferenças foram confirmadas pelo teste de Mann-Whitney ao nível de 5% de significância. Solos de mata e cana apresentaram respectivamente em média 6% e 1,39% de *voxels pv*, justificando uma maior dificuldade de encontrar *P1* em PVE, ou seja, o pico representante da fase pura de vazios em solos de cana, corroborando os achados e afirmativas do trabalho de Jefferies²⁵. Já em termos de *voxels ps*, embora a cana tenha maior presença, a mata também apresenta uma elevada proporção, justificando assim a facilidade em encontrar ao menos *P2* por PVE em imagens de diferentes usos (Figura 18).



a. *Voxels* puros vazios (*pv*), b. *Voxels* mistos (*m*), c. *Voxels* puros sólidos (*ps*)

Figura 18. Percentuais de *voxels* puros vazios (a), mistos (b) e puros sólidos (c) m imagens de tomografia de solos de mata e cana

Avaliando esta classificação, por exemplo, para a imagem XB300, de mata, *pv* correspondem a 7,53%, *m* a 51,02% e *ps* 41,45% (Figura 19).



Escala de cinza em 16 bits

pv

m

ps

Figura 19. Classificação dos *voxels* de acordo com a curva de multissegmentação por EPC para a imagem XB300, os *pixels* em vermelho representam a seleção de cada categoria sob a imagem original em 16 *bits* (*pv* = puros vazios, *m* = mistos e *ps* = puros sólidos)

Além de avaliar a consistência da representação das fases, é importante avaliar se os valores assumidos por *Vale* e *Pico* podem estar relacionados a uma maior ou menor presença de *pv*, *m* ou *ps*. Uma análise do coeficiente de correlação de Pearson entre valores de *Vale* e *pv* (-0,744, *p*-valor = 0,005), *Vale* e *m* (-0,227, *p*-valor = 0,477), *Vale* e *ps* (0,519, *p*-valor = 0,083), *Pico* e *pv* (-0,57, *p*-valor = 0,052), *Pico* e *m* (-0,025, *p*-valor = 0,936) e *Pico* e *ps* (0,279, *p*-valor = 0,379) mostra que apenas o *Vale* está fortemente correlacionado de forma negativa com *pv*. Desta forma, espera-se que valores de *Vale* sejam menores em solos de mata (Figura 20.a e Figura 20.c). Outra correlação muito forte encontrada foi entre valores de *Vale* e *Pico* (0,886, *p*-valor = 0,000) (Figura 20.b), indicando que imagens de valores de *Vale* maiores, tendem a assumir valores de *Pico* maiores, independentemente do uso. Além disso, o teste de Mann-Whitney também apontou uma diferença significativa a 5% de significância entre valores de *Vale* em diferentes usos (Figura 20.c). Este mesmo teste mostrou que o valor de *Pico* não é influenciado por diferentes usos.

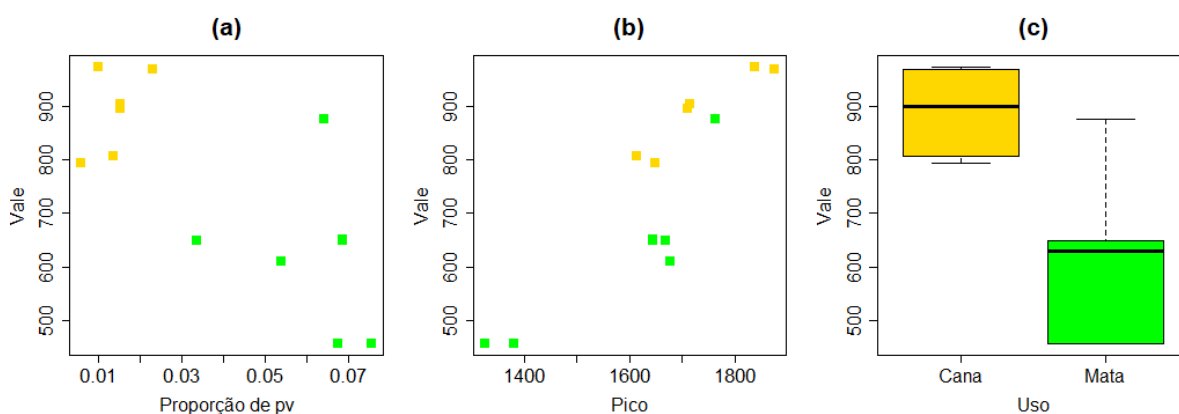


Figura 20. Comparativo de dispersão e variabilidade do *Vale* da curva de multissegmentação em EPC aplicados em imagens de tomografia de solos de mata e cana

A análise de comportamento da curva de multissegmentação demonstra uma coerência em suas estatísticas. Desta forma, a partir de imagens previamente segmentadas por PVESeg busca-se calibrar uma determinada quantidade de desvios padrões que deve deslocar o *Vale* para a direita, ou o *Pico* para esquerda, com fins de tomar uma decisão quanto aos *voxels m*, que na segmentação final deverão fazer parte dos vazios ou não vazios. A esta análise chamaremos de calibração do EPC, apresentada a seguir.

3.4 Calibração de EPC por PVESeg

Um dos desafios na teoria de detecção de bordas em *CT Segmentation* é a definição de uma constante Z_{pve_i} que multiplicada ao desvio padrão da gaussiana possa auxiliar na tomada de decisão quanto a quais *voxels* m serão considerados vazios e não vazios. Ou seja, adotando PVE como método de pré-segmentação, uma vez que se conhece $P1$ e $P2$, um Z_{pve_i} de ótima segmentação é buscado através da teoria de detecção de bordas ²⁹, onde o cálculo otimizado será o seguinte:

$$T_{vazios} = P1 + Z_{pve_1} * DP_{G1}$$

$$T_{sólidos} = P2 - Z_{pve_2} * DP_{G2}$$

onde T_{vazios} e $T_{sólidos}$ são respectivamente o ponto de partida para uma segmentação final, DP_{G1} e DP_{G2} são respectivamente os desvios padrão das gaussianas relativas aos vazios e aos sólidos. O trabalho de Jefferies ²⁵ faz algumas sugestões quanto Z_{epc_1} e Z_{epc_2} .

As 12 imagens foram em boa parte segmentadas por PVESeg a partir de $P2$ uma vez que houve dificuldades de obtenção de $P1$, principalmente em imagens de cana, sendo assim, partindo de $P2$ a maioria das imagens foi bem segmentada com Z_{pve_1} de 1,5 e o Z_{pve_2} de 0,5, sendo a segmentação final uma união destes processos.

Uma vez adequadamente segmentadas por PVESeg, investigou-se quais seriam os valores ótimos de Z_{epc_1} e Z_{epc_2} , ou seja, versões dos Z_{pve_i} de PVE agora para EPC, partindo agora dos valores de *Vale* e *Pico* assumidos em EPC, que podem trazer uma máxima correspondência espacial com as imagens devidamente segmentadas por PVESeg, ou seja, o objetivo é otimizar o seguinte problema:

$$T_{epc_1} = Vale + Z_{epc_1} * DP \cong T_{vazios}$$

$$T_{epc_2} = Pico - Z_{epc_2} * DP \cong T_{sólidos}$$

onde DP neste caso está sendo considerado o desvio padrão da imagem. Em outras palavras, uma vez que se conhece os valores de fronteiras puras por EPC e os desvios padrões de cada imagem, quais seriam os valores de Z_{epc_1} e Z_{epc_2} que tonariam respectivamente T_{epc_1} e T_{epc_2} tão próximos de T_{vazios} e $T_{sólidos}$? A Figura 21 mostra a obtenção destes valores para XA200, onde $Z_{epc_1} = 1,2$ e $Z_{epc_2} = 1,35$.

Segmentada por PVESeg, XA200 apresentou um porosidade de 11,02% contra 11,38% em EPC corrigindo por Z_{epc_2} . Uma comparação fatia a fatia da imagem foi

conduzida mostrando uma concordância média de 99,6% dos *voxels* (Figura 22). Este mesmo procedimento foi realizado para as outras 11 imagens, as quais também apresentaram uma taxa de correspondência espacial acima de 99%.

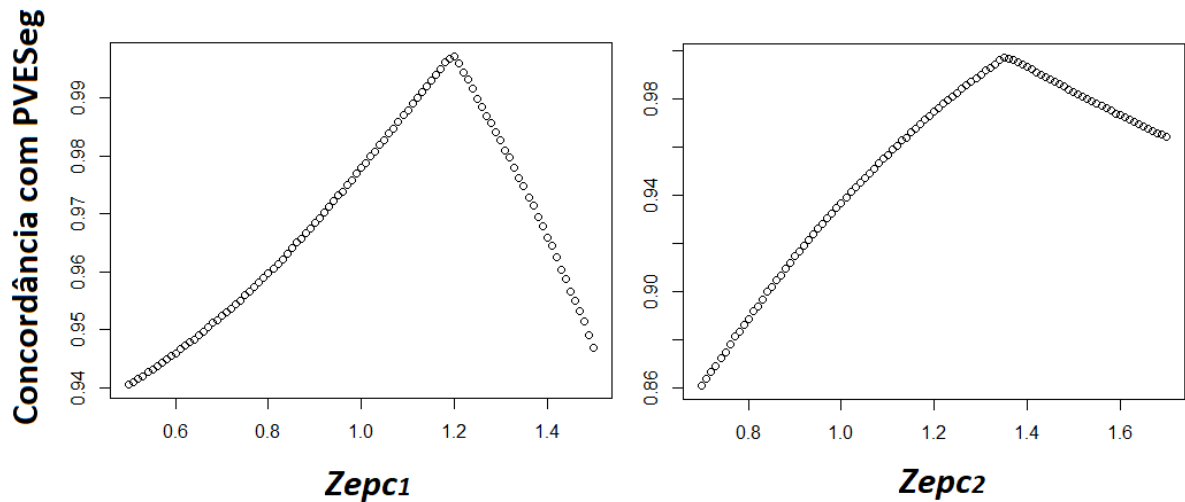


Figura 21. Obtenção de $Zepc_1$ e $Zepc_2$ para XA200

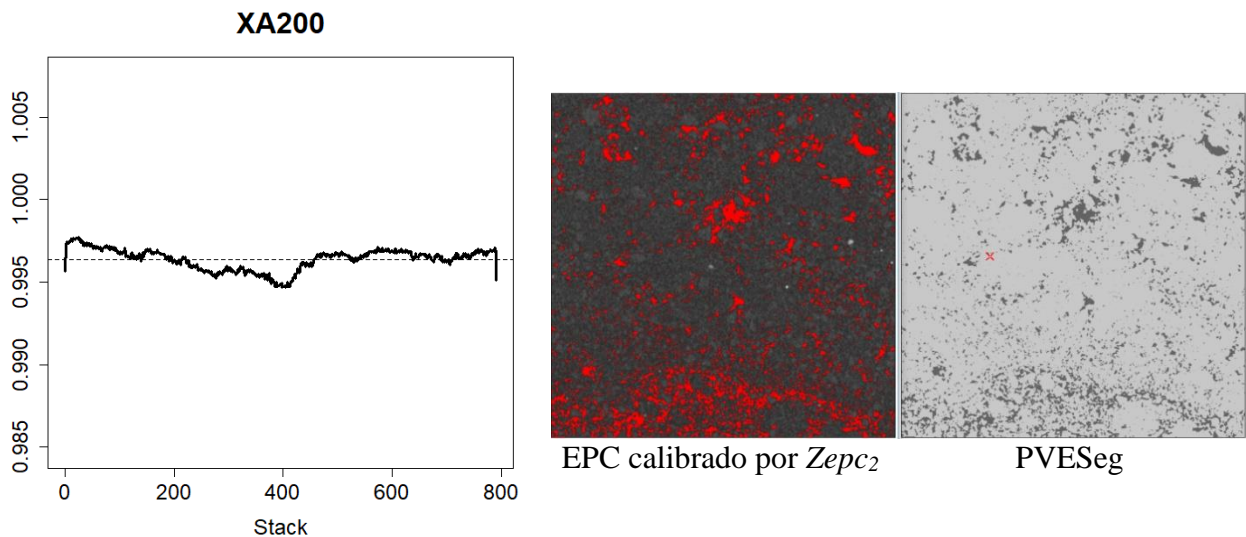


Figura 22. Correspondência espacial entre EPC corrigido por $Zepc_2$ e PVESeg fatia por fatia

Os resultados para XA200 e as demais imagens estão apresentados na Tabela 5, indicando que $Zepc_1$ apresentou uma média de 1,09 para cana e 0,86 para mata, já $Zepc_2$ apresentou uma média de 1,38 para cana e 1,17 para mata. Valores de porosidade também foram mantidos bastante parecidos e estes valores médios podem ser usados como sugestões para segmentação das demais imagens.

Tabela 5. Valores de ajuste de EPC buscando similaridade com PVE

Uso	<i>imagem</i>	<i>Vale</i>	<i>Zepc₁</i>	<i>Pico</i>	<i>Zepc₂</i>	Desvio padrão	<i>Threshold</i>	Porosidade PVESeg	Porosidade EPC + calibração
<i>Cana</i>	<i>A1001</i>	807	0,86	1613	1,39	359	1113,99	0,0725	0,0738
<i>Cana</i>	<i>A4001</i>	896	0,95	1710	1,43	342	1220,94	0,0726	0,0739
<i>Cana</i>	<i>A5001</i>	903	1,01	1714	1,41	335	1241,65	0,073	0,0742
<i>Cana</i>	<i>A5102</i>	794	1,18	1649	1,35	337	1194,05	0,0608	0,062
<i>Cana</i>	<i>A6001</i>	973	1,38	1839	1,36	316	1409,24	0,0807	0,0814
<i>Cana</i>	<i>XA200</i>	968	1,2	1876	1,35	356	1395,4	0,1102	0,1138
<i>Mata</i>	<i>B1001</i>	456	0,52	1326	1,19	509	720,29	0,157	0,1664
<i>Mata</i>	<i>B3001</i>	610	1,23	1678	0,93	495	1217,65	0,237	0,2573
<i>Mata</i>	<i>B4102</i>	648	1,3	1669	1,11	424	1198,36	0,1644	0,1739
<i>Mata</i>	<i>B5001</i>	650	0,5	1644	1,52	490	899,2	0,1347	0,1373
<i>Mata</i>	<i>B6001</i>	875	1,1	1763	1,06	411	1327,34	0,1622	0,1611
<i>Mata</i>	<i>XB300</i>	457	0,53	1381	1,22	529	735,62	0,1849	0,204

Uma vez que EPC foi calibrado por PVESeg, pode-se avaliar quais *voxels m* foram alocados para a fase de vazios e não vazios.

3.5 Análise de alocação de *voxels* mistos

Definiremos m_1 e m_2 como sendo os *voxels m* alocados respectivamente para a fase de vazios e m_2 como sendo os *voxels m* alocados para a fase de vazios e não vazios. Da mesma forma, os vazios totais (V) e sólidos totais (S):

$$V = pv + m_1$$

$$S = ps + m_2$$

Por exemplo, para a imagem de solo de mata XB300, sabemos diante mão, pelo estudo realizado sobre o comportamento da curva de multissegmentação que os percentuais de pv e ps foram respectivamente de 7,53% e 41,45%. Pela tabela anterior a porosidade total obtida através de EPC calibrado por PVESeg foi de 20,4% e que restam 51,02% de *voxels* mistos, desta forma, pode-se concluir que $m_1 = 12,87\%$ e $m_2 = 38,15\%$ (Figura 23).

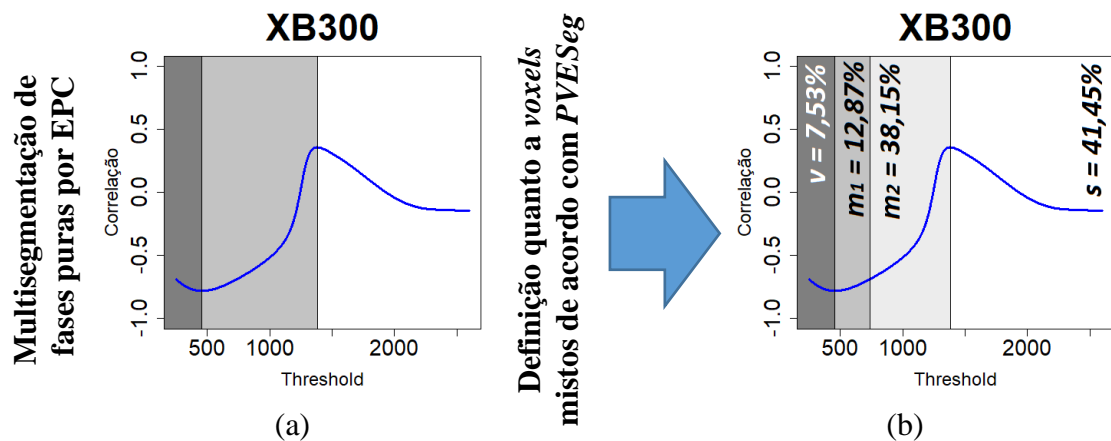
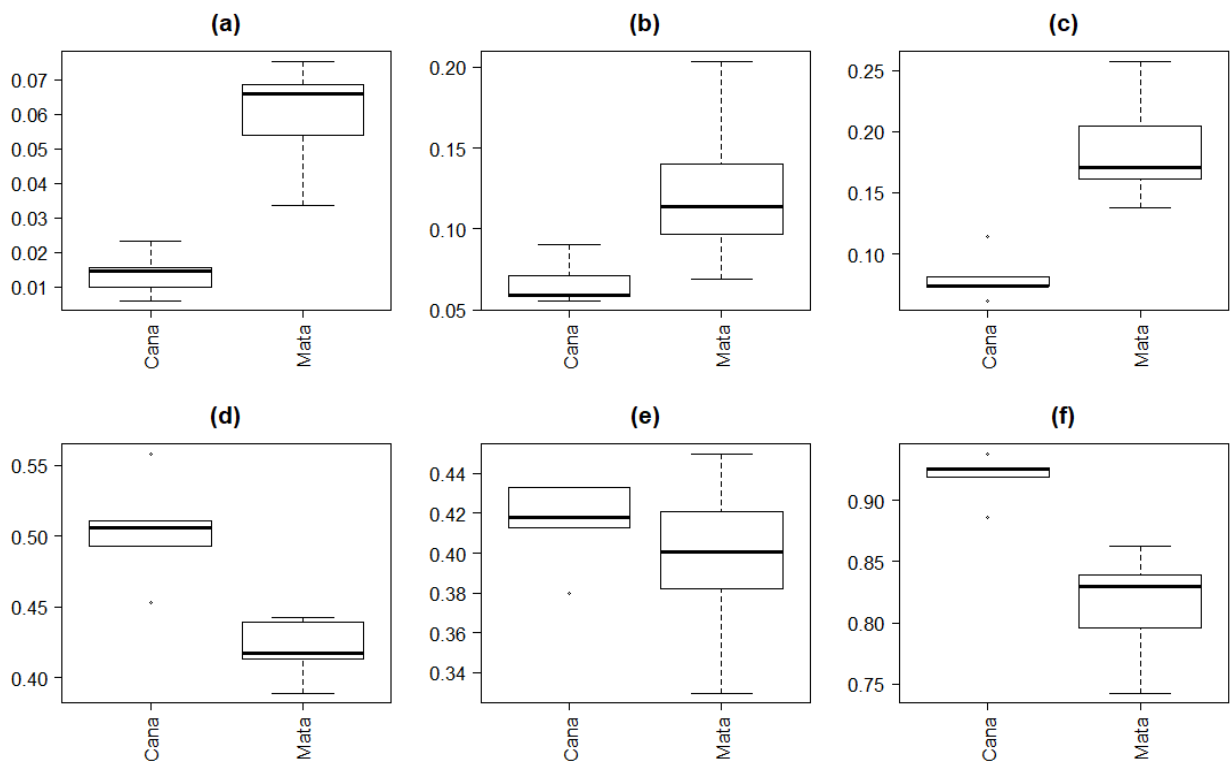


Figura 23. Percentuais de *voxels* puros e mistos de XB300 a partir dos limiares propostos pela curva de correlação de planos de entropia

Foi discutido o comportamento e diferença da proporção de pv e ps entre solos de mata e cana (Figura 24, *a* e *d*). Em relação a alocação de *voxels* m , verificou-se pelo teste de Mann-Whitney a 5% de significância que imagens de mata alocam mais m_1 e que não há diferença na alocação de m_2 entre os usos (Figura 24, *b* e *d*). Ficou também de forma consistente mostrada a superioridade de V em mata e S em cana (Figura 24, *c* e *e*).



a. *Voxels* puros vazios (pv), b. *Voxels* mistos classificados como vazios (m_1), c. Porosidade total ($pv + m_1$), d. *Voxels* puros sólidos (ps), e. *Voxels* mistos classificados como sólidos (m_2), f. Sólidos totais ($ps + m_2$)

Figura 24. Percentuais de *voxels* em imagens de solos de mata e cana

As análises acima demonstram uma consistência na alocação de *voxels m*. Validando a correção de PVE por EPCSeg e sugerindo valores de deslocamento de desvio padrão $Zepc_1$ e $Zepc_2$. A próxima seção realiza uma avaliação minuciosa e sugestões de generalização destes deslocamentos com fins de evitar entradas subjetivas na busca de uma segmentação final.

3.6 Deslocamentos de desvio padrão $Zepc_1$ e $Zepc_2$

Uma análise dos deslocamentos de desvios sugeridos pela calibração é extremamente relevante para buscar formas de sugerir deslocamentos para novas imagens, ou seja, outras imagens que serão diretamente segmentadas por EPCSeg.

Em uma análise conjunta de deslocamentos e desvio padrão, já que estes caminham juntos no processo de segmentação final, solos de mata apresentaram em geral maior variabilidade e menor $Zepc_2$, por outro lado, solos de cana apresentaram menor variabilidade e maior $Zepc_2$. Fato que converge para o que foi visto anteriormente sobre a proporção de m_2 alocada em diferentes usos, pois se em mata temos maior variação com menor deslocamento desta e ao contrário, em cana temos menor variação com maior deslocamento desta, essas proporções de m_2 tendem a se equiparar (Figura 25, *a* e *c*).

Com relação a $Zepc_1$, este não apresentou diferença significativa entre os usos, o que também converge para os achados anteriores sobre a proporção de alocação de m_1 . Pois em mata temos uma maior variabilidade, logo é esperada uma maior alocação de m_1 nestas imagens, pois em cana estaremos fazendo similar deslocamento, porém em uma variação menor (Figura 25, *a* e *b*). Estes achados são também confirmados pelo teste de Mann-Whitney a 5% de significância.

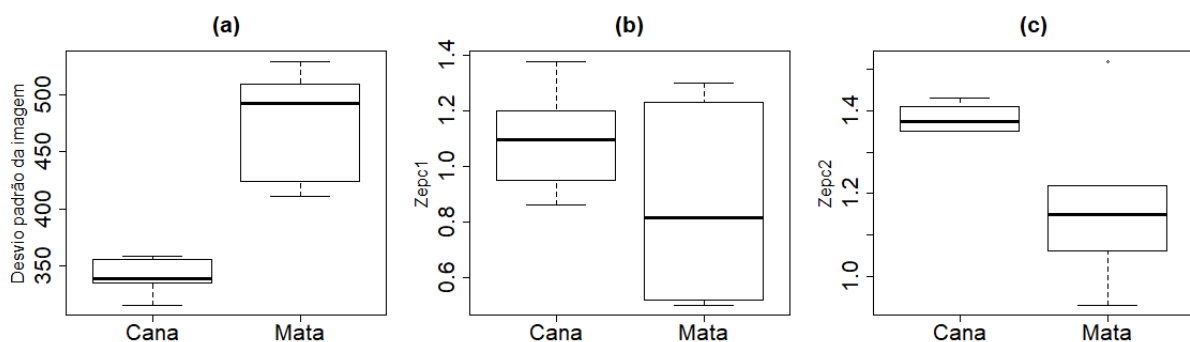


Figura 25. Desvio padrão das imagens e variabilidade de $Zepc_1$ e $Zepc_2$

Uma análise da correlação linear de Pearson entre $Zepc_1$ e *Vale* (0,628, *p*-valor = 0,02), $Zepc_1$ e *Pico* (0,768, *p*-valor = 0,003), $Zepc_2$ e *Vale* (0,403, *p*-valor = 0,193) e $Zepc_2$ e

Pico (0,161, *p*-valor = 0,615), mostra que *Zepc₂* não é influenciado pelos valores assumidos por *Vale* ou *Pico*. Por outro lado, *Zepc₁* tem correlação moderada positiva com *Vale* e forte positiva com *Pico*. A união destes achados sugere que podemos sugerir *Zepc₁* por uma regressão linear simples com *Vale* ou *Pico* como independente e sugerir *Zepc₂* pela média dos diferentes usos.

O ajuste de uma regressão linear simples para estimar *Zepc₁* através de *Vale* demonstrou que o intercepto não era significativo, este foi removido, e o novo modelo foi considerado adequado pela ANOVA para testar a inclinação ($F = 194,05$, *p*-valor = 0,000). A mesma abordagem foi usada para estimar *Zepc₂* através de *Pico* ($F = 499,32$, *p*-valor = 0,000), chegando as equações:

$$\hat{Z}_{epc1} = 0,001289 * Vale$$

$$\hat{Z}_{epc2} = 0,0007662 * Pico$$

Anteriormente foi mostrado que *Zepc₂* não está correlacionado com *Pico* e mesmo assim este modelo sem intercepto mostrou-se explicativo, deste modo, não foi descartada esta sugestão, de modo que quatro combinações apresentadas na Tabela 6 são comparadas em termos de porosidade e correspondência espacial com PVESeg:

Tabela 6. Configurações de EPCSeg avaliadas

Método	<i>Zepc₁</i>	<i>Zepc₂</i>
EPCSeg ₁	Regressão (<i>Zepc_{1r}</i>)	Média (<i>Zepc_{2m}</i>)
EPCSeg ₂	Regressão (<i>Zepc_{1r}</i>)	Regressão (<i>Zepc_{2r}</i>)
EPCSeg ₃	Média (<i>Zepc_{1m}</i>)	Regressão (<i>Zepc_{2r}</i>)
EPCSeg ₄	Média (<i>Zepc_{1m}</i>)	Média (<i>Zepc_{2r}</i>)

3.7 Porosidade e correspondência espacial entre EPCSeg e PVESeg

As imagens foram submetidas as quatro configurações de EPCSeg, descritas anteriormente, sobre as quais foram observadas a porosidade e a correspondência espacial, entre elas e em relação a PVESeg. O teste de Shapiro-Wilk confirmou a normalidade de PVESeg ($W=0,906$, *p*-valor = -0,193), EPCSeg₁ ($W = 0,938$, *p*-valor = 0,473), EPCSeg₂ ($W = 0,94$, *p*-valor = 0,507), EPCSeg₃ ($W = 0,867$, *p*-valor = 0,06) e EPCSeg₄ ($W = 0,922$, *p*-valor = 0,307). Tanto uma análise de variância ($F = 0,199$, *p*-valor = 0,937) quanto o teste de

Friedman ($F = 1,575$, $p\text{-valor} = 0,813$) confirmam os valores de porosidades apresentados em PVESeg e nas quatro configurações de EPCSeg são iguais ao nível de 5% de significância. A Tabela 7, mostra os referidos valores.

Tabela 7. Valores de porosidade de PVESeg e configurações de EPCSeg

<i>imagem</i>	<i>Zepc_{1m}</i>	<i>Zepc_{2m}</i>	<i>Zepc_{1r}</i>	<i>Zepc_{2r}</i>	<i>PVESeg</i>	<i>EPCSeg₁</i>	<i>EPCSeg₂</i>	<i>EPCSeg₃</i>	<i>EPCSeg₄</i>
A1001	1,10	1,38	1,04	1,24	0,0725	0,1	0,101	0,111	0,11
A4001	1,10	1,38	1,15	1,31	0,0726	0,1	0,1	0,0933	0,093
A5001	1,10	1,38	1,16	1,31	0,073	0,094	0,094	0,0872	0,085
A6001	1,10	1,38	1,25	1,41	0,0807	0,078	0,075	0,074	0,077
XA200	1,10	1,38	1,25	1,44	0,1102	0,1203	0,1202	0,1	0,108
A5102	1,10	1,38	1,02	1,26	0,0608	0,058	0,071	0,072	0,058
B1001	0,86	1,17	0,59	1,02	0,157	0,183	0,209	0,258	0,256
B3001	0,86	1,17	0,79	1,29	0,237	0,201	0,179	0,179	0,202
B4102	0,86	1,17	0,84	1,28	0,1644	0,1589	0,141	0,141	0,159
B5001	0,86	1,17	0,84	1,26	0,134	0,191	0,187	0,191	0,191
B6001	0,86	1,17	1,13	1,35	0,1622	0,165	0,162	0,122	0,14
XB300	0,86	1,17	0,59	1,06	0,1849	0,2205	0,245	0,291	0,289

Além de apresentarem valores de porosidade próximos, todas as configurações apresentaram também uma correspondência espacial muito próxima com PVESeg, com EPCSeg₁ atingindo em média 97,6% neste quesito. Esta configuração realiza a sugestão de Zepc₁ por meio de regressão e Zepc₂ pela média de cada uso, algo mais coerente com os achados anteriores na calibração dos mesmos (Tabela 8).

Tabela 8. Correspondência espacial com PVESeg

Imagem	EPCSeg₁	EPCSeg₂	EPCSeg₃	EPCSeg₄
A1001	97,1	97,1	96,1	96,1
A4001	97,1	97,1	97,9	97,9
A5001	97,9	97,8	98,5	98,7
A5102	99,6	98,8	98,8	99,6
A6001	99,6	99,3	99,3	99,6
XA200	98,9	99	98,7	99,4
B1001	97,4	94,8	89,9	90
B3001	95,3	93,4	93,3	95,2
B4102	98,5	97,2	97,2	98,5
B5001	94,3	94,7	94,3	94,3
B6001	99,1	99,4	95,9	97,7
XB300	96,4	93,9	89,3	89,5
Média	97,60	96,88	95,77	96,38

Com base nestas informações adotaremos EPCSeg₁ como configuração para segmentar as demais imagens.

3.8 Segmentação por EPCSeg

Para as 24 imagens restantes, inicialmente foram obtidas suas respectivas curvas de multissegmentação pela correlação de planos de entropia (Figura 26).

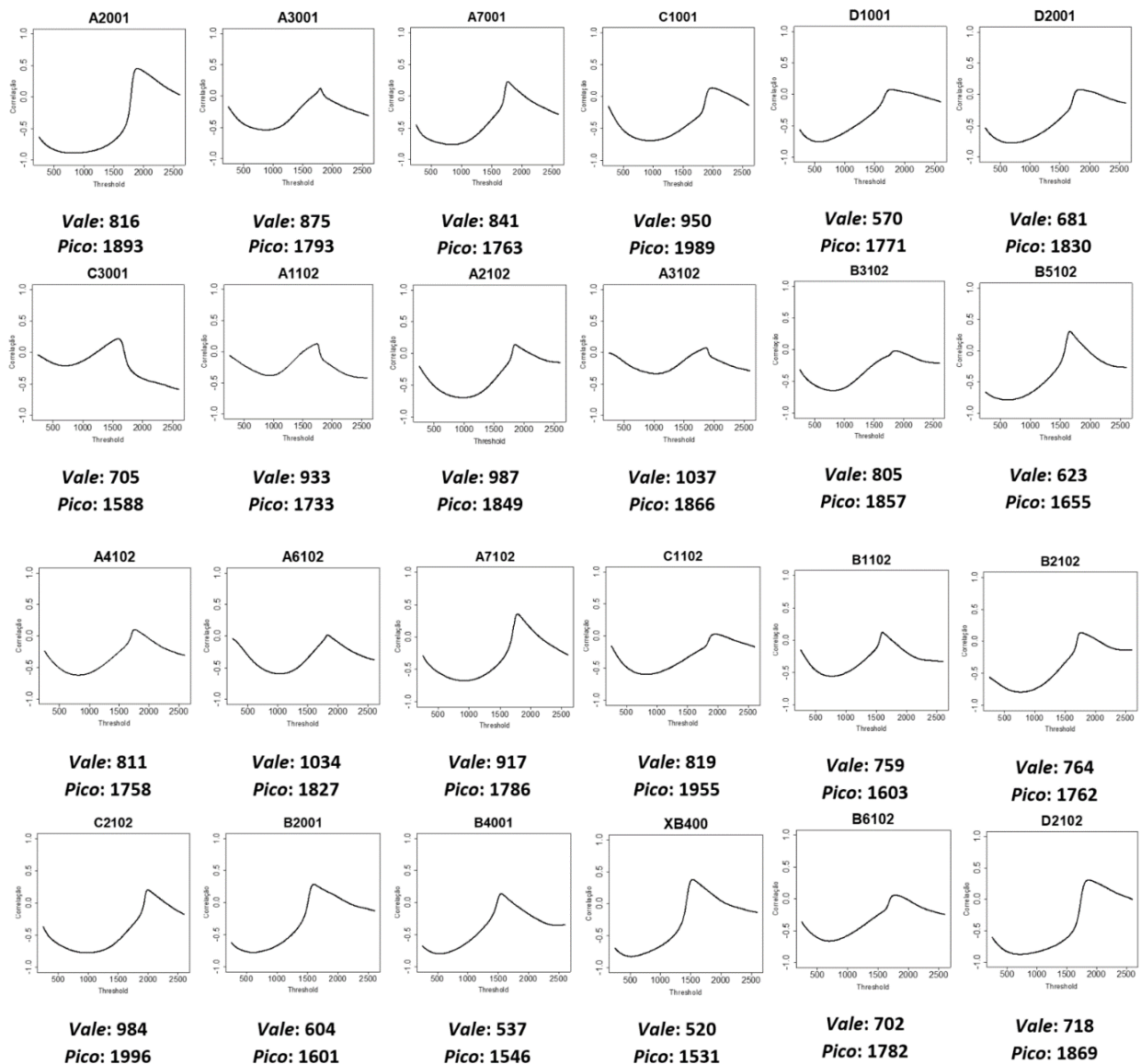


Figura 26. Curvas de multissegmentação obtidas por correlação de planos de entropia (EPC) apresentado Vale e Pico para pré-segmentação de imagens de tomografia de solos

Os deslocamentos de variabilidade foram definidos por $Zepc_{1r}$ e $Zepc_{2m}$. A porosidade e os parâmetros de segmentação final estão apresentados na Tabela 9, a seguir.

Tabela 9. Porosidade e parametrização da segmentação final

<i>Profundidade (cm)</i>	<i>Uso</i>	<i>Imagem</i>	<i>Desvio</i>	<i>Zepc_{1r}</i>	<i>Zepc_{2m}</i>	<i>Porosidade EPCSeg</i>
0 - 10	<i>Cana</i>	A2001	485	1,05	1,38	0,168
0 - 10	<i>Cana</i>	A3001	353	1,13	1,38	0,069
0 - 10	<i>Cana</i>	A7001	353	1,08	1,38	0,094
0 - 10	<i>Cana</i>	C1001	425	1,22	1,38	0,191
0 - 10	<i>Cana</i>	C3001	452	0,91	1,38	0,134
10 - 20	<i>Cana</i>	A1102	308	1,20	1,38	0,058
10 - 20	<i>Cana</i>	A2102	367	1,27	1,38	0,154
10 - 20	<i>Cana</i>	A3102	323	1,34	1,38	0,079
10 - 20	<i>Cana</i>	A4102	387	1,05	1,38	0,119
10 - 20	<i>Cana</i>	A6102	343	1,33	1,38	0,167
10 - 20	<i>Cana</i>	A7102	348	1,18	1,38	0,107
10 - 20	<i>Cana</i>	C1102	454	1,06	1,38	0,167
10 - 20	<i>Cana</i>	C2102	387	1,27	1,38	0,106
0 - 10	<i>Mata</i>	B2001	493	0,78	1,17	0,19
0 - 10	<i>Mata</i>	B4001	491	0,69	1,17	0,173
0 - 10	<i>Mata</i>	XB400	527	0,67	1,17	0,22
0 - 10	<i>Mata</i>	D1001	526	0,73	1,17	0,217
0 - 10	<i>Mata</i>	D2001	495	0,88	1,17	0,214
10 - 20	<i>Mata</i>	B1102	386	0,98	1,17	0,132
10 - 20	<i>Mata</i>	B2102	393	0,98	1,17	0,143
10 - 20	<i>Mata</i>	B3102	390	1,04	1,17	0,167
10 - 20	<i>Mata</i>	B5102	452	0,80	1,17	0,153
10 - 20	<i>Mata</i>	B6102	388	0,90	1,17	0,176
10 - 20	<i>Mata</i>	D2102	526	0,93	1,17	0,218

Na segmentação final, são adicionados *voxels* m_l em regiões antes puras (apenas com p_v) ou em regiões isoladas da imagem (Figura 27). Avaliando as médias de porosidade em relação aos perfis e profundidades por meio de uma ANOVA foram constatadas diferenças significativas ($F = 5,812$, $p\text{-valor} = 0,005$). Em seguida, o pós-teste de Tukey detalhou estas diferenças, trazendo informações que sugerem a consistência do método EPCSeg, ou seja, solos de mata na profundidade de 0-10 dispõem da maior porosidade, enquanto solos de cana na profundidade de 10-20 com dispõem da menor porosidade (Figura 28 e Figura 29).

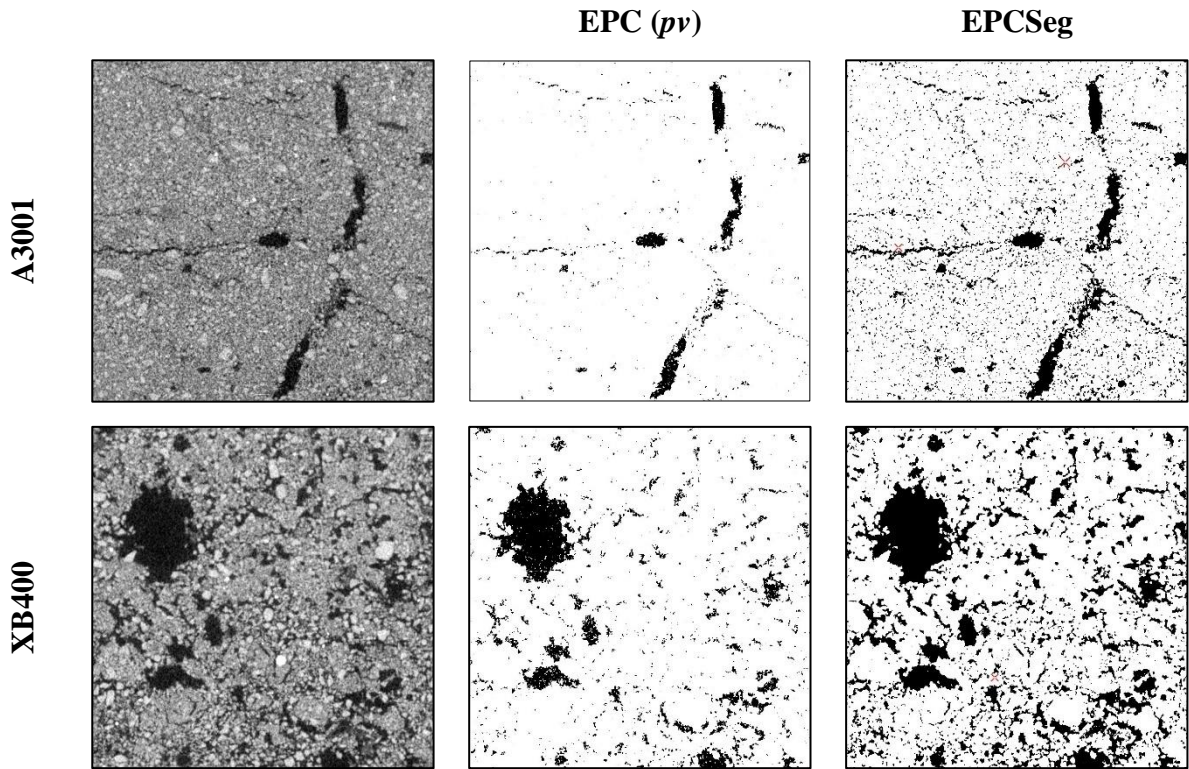
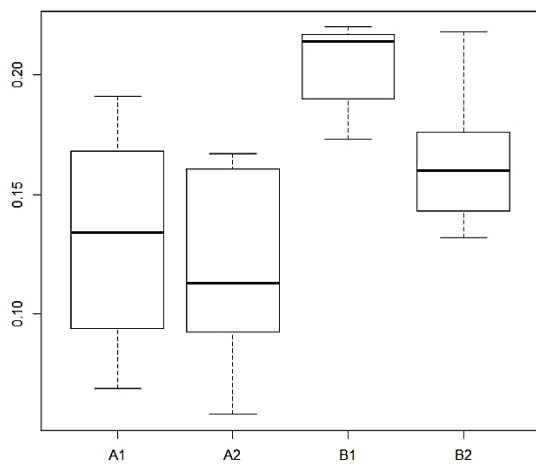
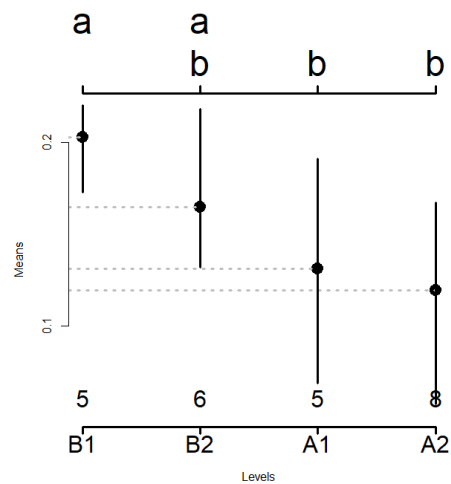


Figura 27. Evolução da fase de poros vazios para vazios totais



A1: Cana de 0-10 cm, A2: Cana de 10-20 cm,
 B1: Mata de 0-10 cm, B2: Mata de 10-20 cm

Figura 28. Variação de porosidade por EPCSeg



A1: Cana de 0-10 cm, A2: Cana de 10-20 cm,
 B1: Mata de 0-10 cm, B2: Mata de 10-20 cm

Figura 29. Porosidade média via EPCSeg - Teste de Tukey

4. ANÁLISE MULTIVARIADA DE PLANOS DE ENTROPIA

Os resultados do capítulo anterior motivam uma investigação mais profunda a respeito de como planos de entropia podem ajudar a classificar imagens de tomografia de solos em diferentes usos. O objetivo deste capítulo é justamente unificar a abordagem de planos de entropia com análise multivariada, de modo a obter uma possível classificação destas imagens. Utilizando-se análise de componentes principais (ACP), buscou-se entender qual(is) *thresholds* otimizam a distinção entre estas imagens. A presente análise seguirá passos similares ao capítulo anterior, porém, ao invés de colecionar as correlações entre o plano amostral e planos de informação de Shannon com variados *thresholds*, são colecionadas a máxima explicabilidade da primeira componente de cada ACP no decorrer da variação de HU.

Foram realizadas 2350 ACP's, ou seja, utilizando como variáveis de entrada o plano amostral e planos de Shannon variando de 250 a 2600 HU. Avaliando a dispersão entre a variação de HU utilizada para construir planos de Shannon e a explicabilidade da primeira componente, foi possível a construção e investigação da curva de explicabilidade total (CET). A CET foi obtida a partir da variação explicada pela primeira componente em cada uma das 2350 ACP's, mostrando assim a evolução da variação explicada no decorrer da variação do *threshold* de 250 a 2600 HU. A CET apresenta dois picos (P_{cet1} e P_{cet2}) e um vale (V_{cet}) muito evidentes, Figura 30.

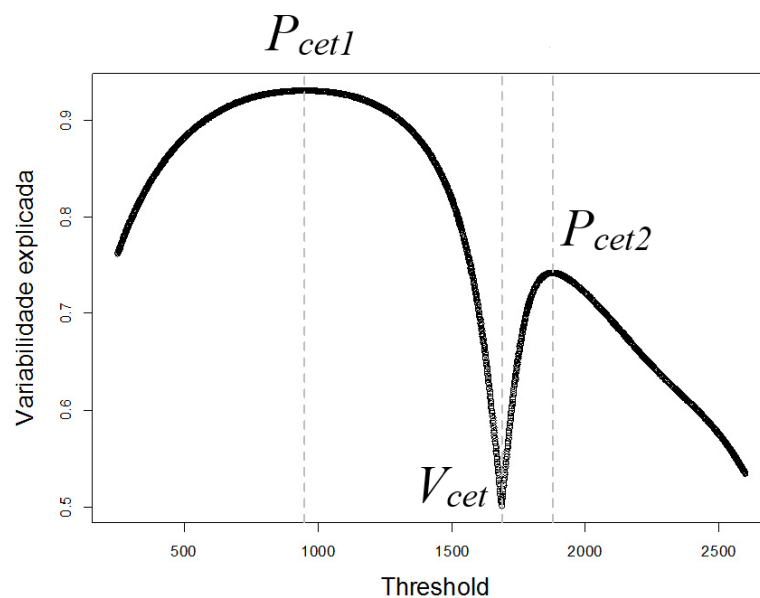


Figura 30. Representação geral de uma curva de explicabilidade total obtida pela dispersão entre a variação de HU e variabilidade explicada pela primeira componente de cada ACP

Por representarem um maior explicabilidade da variação, os picos representam limiares que podem classificar melhor as imagens e são em momento posterior melhor investigados por meio de uma análise de agrupamentos (AAg) e da visualização de sua representação nas imagens originais. As ACP's e conseqüentemente a CET foram obtidas de três formas com dois objetivos, classificar imagens: utilizando as médias dos planos (ACP₁) e utilizando os valores de *pixels* dos planos (ACP₂) e classificar pixels: utilizando os pixels dos planos de entropia e rotulando os mesmos como provenientes de mata ou cana (ACP₃). Os resultados de cada análise são apresentados a seguir.

4.1 ACP₁: Classificando imagens por meio de médias de planos de entropia

Para cada limiar é obtido um banco de dados contendo a média dos planos de informação de Shannon que combinados a média dos planos de entropia amostral fornecem as duas variáveis da análise. Desta forma, cada ACP será conduzida em uma base de 36 linhas, uma para cada imagem e duas variáveis, uma para cada média de planos. As imagens são identificadas como sendo oriundas de mata ou cana. A curva de explicabilidade, com picos e vale, é mostrada na Figura 31 e estes detalhados na Tabela 10, a seguir:

Tabela 10. Valores de *threshold* com maiores e menores taxas de explicabilidade

Referência	Threshold	Variação explicada
P_{cet1}	1116	97,8%
V_{cet}	1682	50,1%
P_{cet2}	1886	83,1%

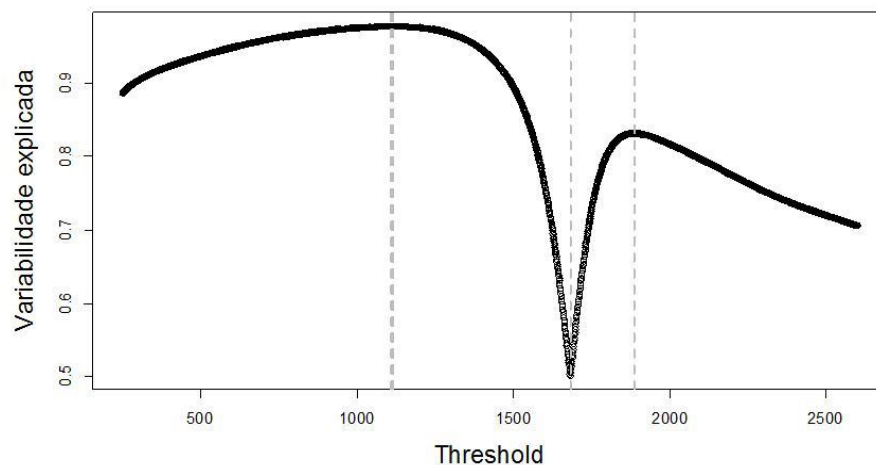


Figura 31. Curva de explicabilidade total a partir de uma ACP da média dos planos no decorrer da variação de HU

Duas ACP's presentes nos dois picos (P_{cet1} e P_{cet2}) são melhor representadas a seguir (Figura 32 e Figura 33).

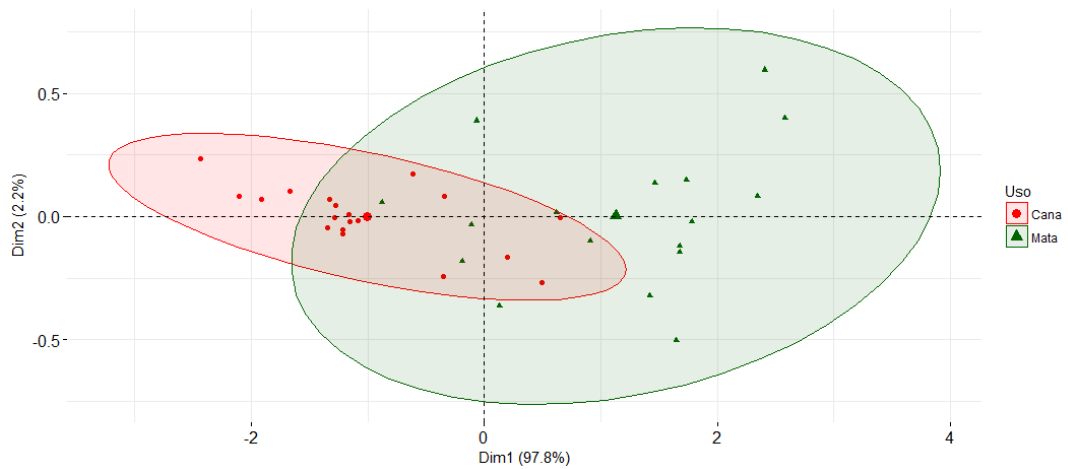


Figura 32. Análise de componentes principais utilizando limiar proposto em P_{cet1}

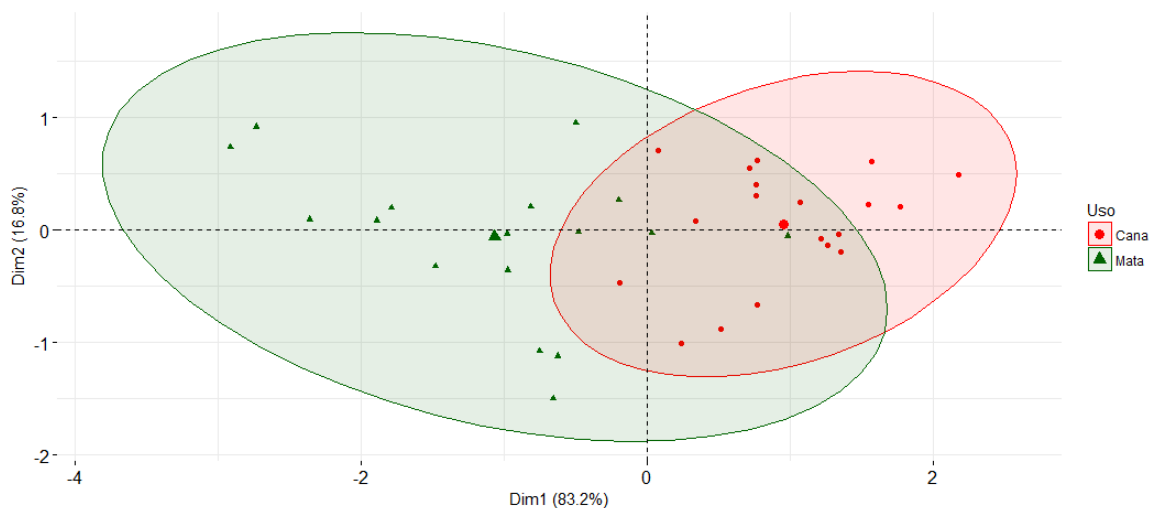


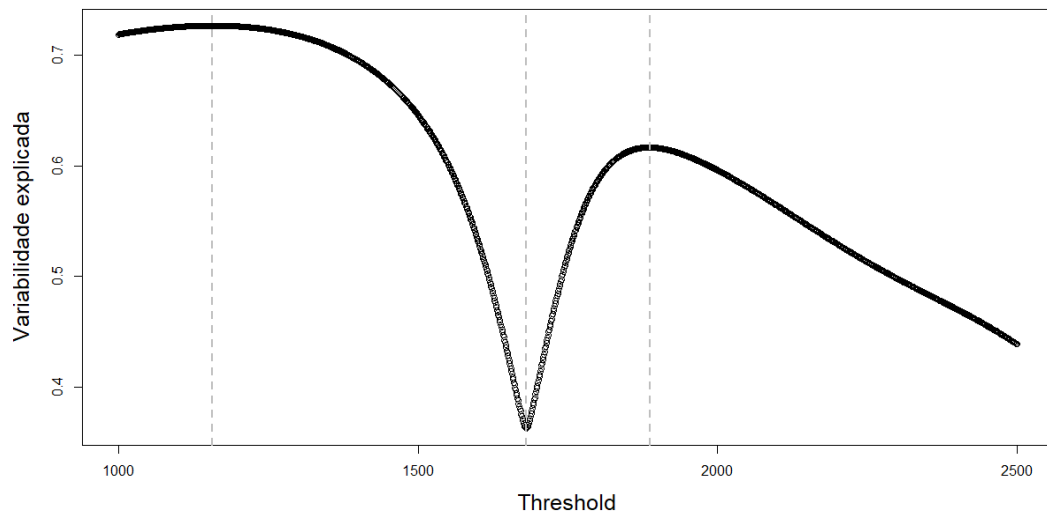
Figura 33. Análise de componentes principais utilizando limiar proposto em P_{cet2}

4.2 ACP₂: Classificando imagens por meio de *pixels* de planos de entropia

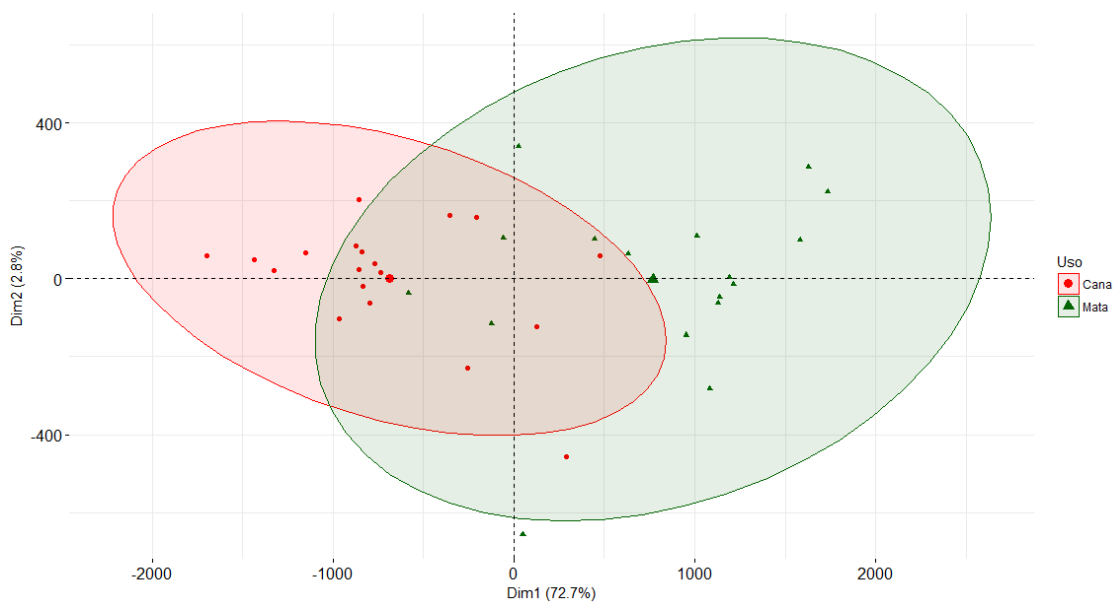
Nesta abordagem cada plano de informação de Shannon e cada plano de entropia amostral são organizados em vetores linha e em seguida unificados para cada imagem. Desta forma foi construído em cada limiar, um banco de dados de 36 linhas (correspondentes a cada imagem) e 1248200 variáveis (correspondentes a cada *pixel*). Os picos de interesse e a explicabilidade nas 2350 ACP's estão resumidas a seguir (Tabela 11; Figura 34):

Tabela 11. Valores de *threshold* com maiores e menores taxas de explicabilidade

Referência	Threshold	Variação explicada
P_{cet1}	1161	72,6%
V_{cet}	1681	36,2%
P_{cet2}	1887	61,6%

**Figura 34.** CET a partir de todos os *pixels* dos planos

Duas ACP's presentes nos dois picos (P_{cet1} e P_{cet2}) são melhor representadas a seguir (Figura 35 e Figura 36):

**Figura 35.** Análise de componentes principais utilizando limiar proposto em P_{cet1}

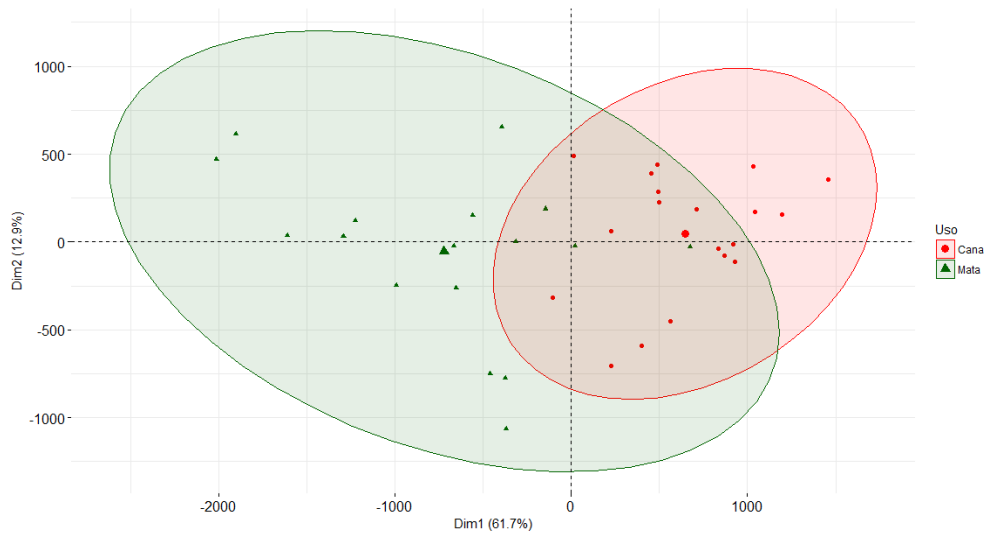


Figura 36. Análise de componentes principais utilizando limiar proposto em P_{cet2}

4.3 ACP3: Classificando *pixels* de planos de entropia

Outra forma de buscar a classificação das imagens pode ser através de classificação da menor unidade observada nos planos, os *pixels*. Uma vez que para mata temos 17 planos e para cana 19, de cada entropia, cada um fornecendo 790^2 *pixels*. Cada ACP foi conduzida em um banco de dados com 10609700 *pixels* de mata e 11857900 de *pixels* de cana, totalizando 22467600 *pixels* a serem classificados a partir de duas variáveis, a informação de Shannon em cada limiar e a entropia amostral. Foram obtidos os seguintes *thresholds* para picos e vale (Figura 37; Tabela 12):

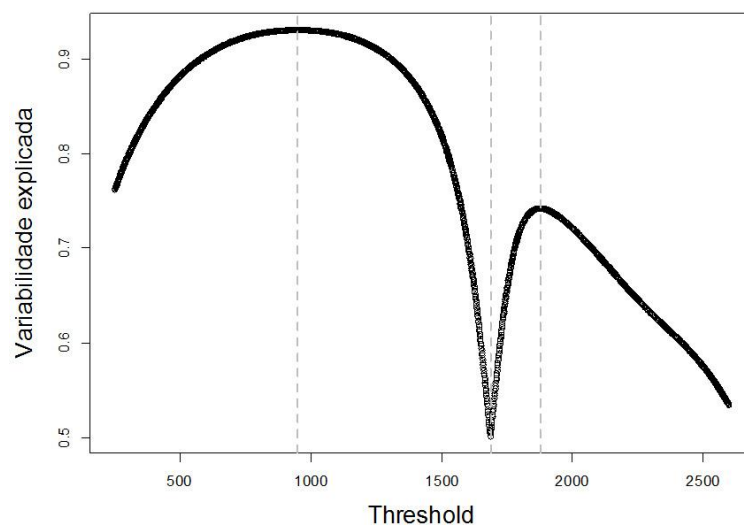


Figura 37. Curva de explicabilidade total a partir de ACP para classificar os *pixels* dos planos

Tabela 12. Valores de *threshold* com maiores e menores taxas de explicabilidade

Referência	Threshold	Variação explicada
P_{cet1}	954	93,1%
V_{cet}	1687	50,1%
P_{cet2}	1877	74,2%

Identificados estes limiares, apresenta-se a representação gráfica das duas componentes principais para valores sugeridos por P_{cet1} e P_{cet2} respectivamente (Figura 38 e Figura 39):

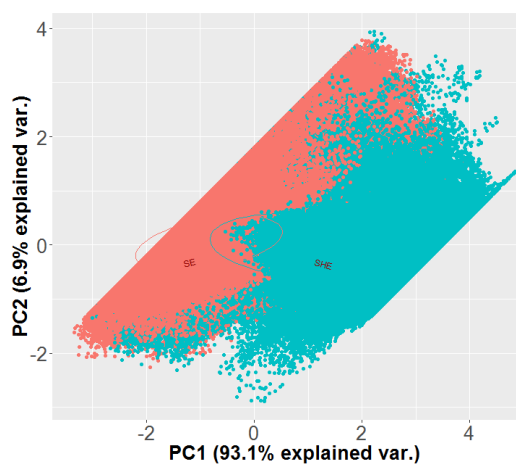


Figura 38. ACP utilizando limiar proposto por P_{cet1}

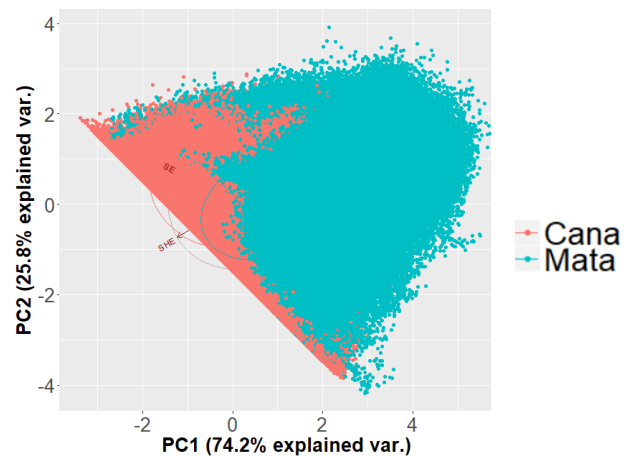


Figura 39. ACP utilizando limiar proposto por P_{cet2}

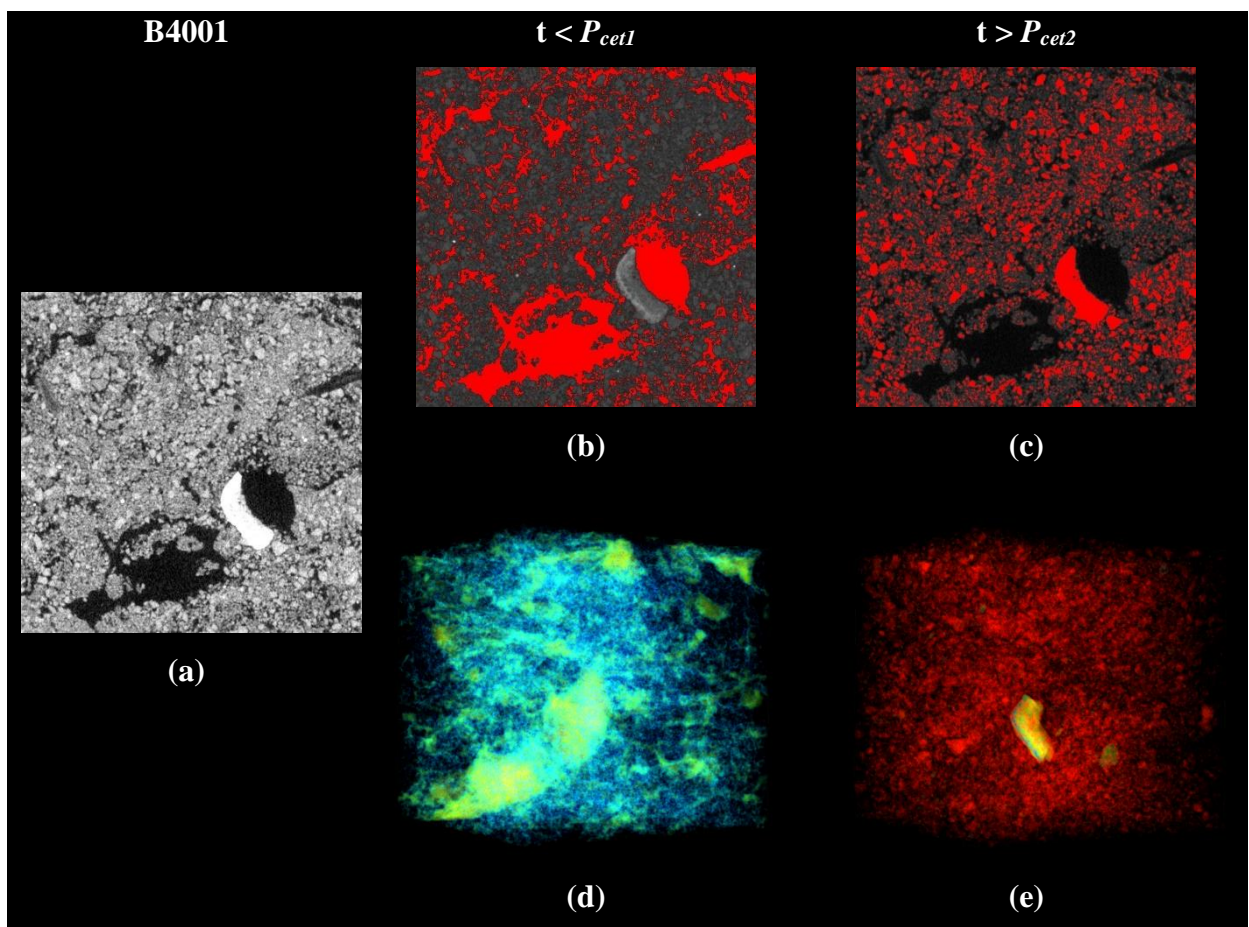
4.4 Discussão sobre ACP e planos de entropia

As três CET's construídas nas análises de componentes principais anteriores, além confirmarem a existência de dois picos e um vale muito evidentes, onde tais picos apontam duas sugestões de limiares que melhor explicam a variabilidade dos planos e/ou *pixels*, onde o contrário pode ser afirmado para o vale, demonstram ainda que duas vizinhanças de *threshold* precisam ser investigadas, uma nas proximidades de 1000 HU e outra nas proximidades de 1800 HU.

A Figura 40 (b e c), apresenta destacado em vermelho duas seleções de *pixels* para uma determinada fatia e (d e e) mostra a representação 3D da imagem B4001 baseadas na

sugestão de ACP_1 . As seleções demonstram que até P_{cet1} a maioria dos *pixels/voxels* selecionados correspondem a fase vazios, ou seja, com tonalidade mais escuras. Por outro lado, os maiores ou iguais a P_{cet2} correspondem a *pixels* mais claros, ou seja, sólidos dos solos.

Na Seção 3.1 sobre planos de entropia e solos, foram considerados como *voxels* de baixa densidade (VBD), aqueles com radiodensidade menor ou igual a 1000 HU, ou seja, em vizinhança similar a proposta nas três ACP's, onde foi observado um maior grau de desordem global e menor grau de informação de Shannon em solos de cana (com menor presença de VBD), sendo o contrário visualizado em solos de mata (com maior presença de VBD). As ACP's além de confirmarem estes achados sugerem ainda outra classe de *voxels* em vizinhança maiores que 1800 HU como pertinentes a uma diferenciação entre estas imagens, os quais chamaremos aqui de VAD ou *voxels* de alta densidade.



(a) Escala de cinza, (b) $Pixels\ p < P_{cet1}$, (c) $Pixels\ p > P_{cet2}$, (d) $Voxels\ v < P_{cet1}$, (e) $Voxels\ v > P_{cet2}$

Figura 40. Seleções de *pixels/voxels* para imagem B4001 baseada CET em ACP_1

Uma análise do percentual da presença na vizinhança de VBD e VAD nas imagens mostra que há uma inversão na presença destes quando comparados os dois usos (Figura 41 e Figura 42). Ou seja, maior presença de VBD em mata e maior presença de VAD em cana. Este comportamento é acompanhado pelas médias dos planos de informação de Shannon (Figura 43 e Figura 44), ou seja, na vizinhança dos VBD a informação de Shannon é mais presente em mata, porém, na vizinhança da VAD esta mesma informação é mais presente em cana, neste último caso, concordando com a desordem global mostrada pela entropia amostral na escala de cinza (Figura 45). Sugere-se assim uma correlação negativa em VBD e positiva em VAD entre essas duas entropias.

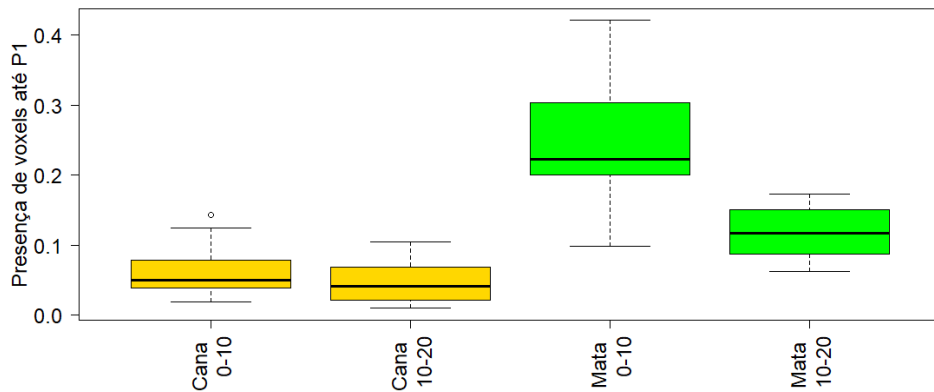


Figura 41. Presença de *voxels* até P_{cet1} em ACP_1

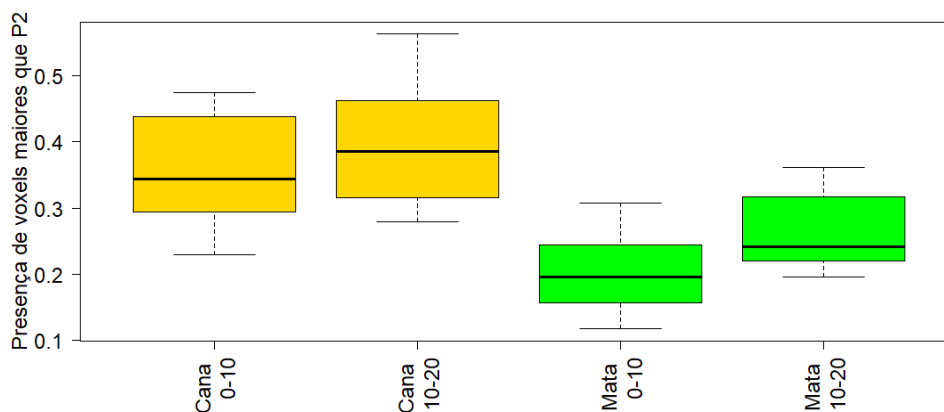


Figura 42. Presença de *voxels* maiores que P_{cet2} em ACP_1

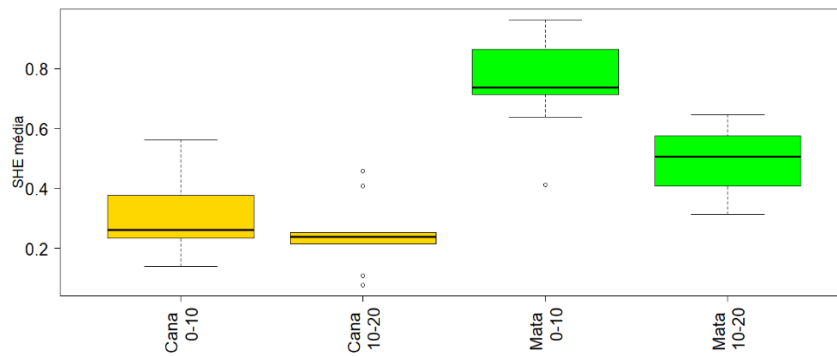


Figura 43. Entropia de Shannon média em diferentes usos e profundidades para P_{cet1}

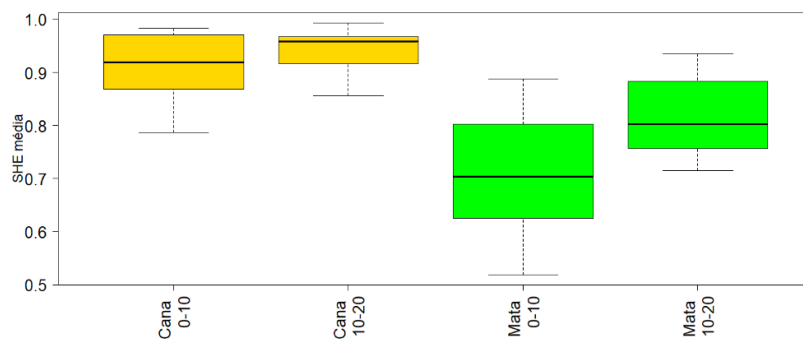


Figura 44. Entropia de Shannon média em diferentes usos e profundidades para P_{cet2}

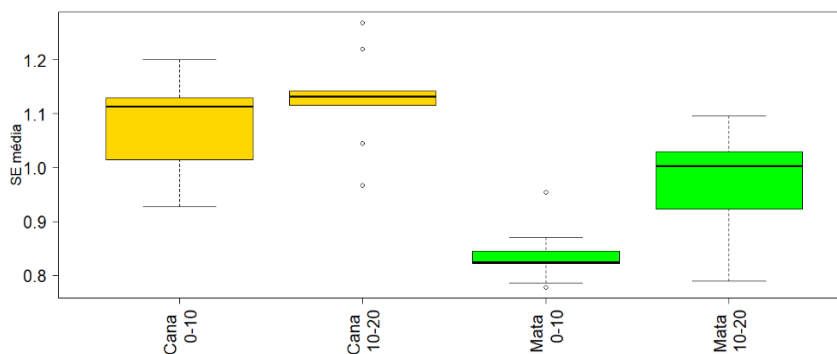


Figura 45. Entropia amostral média em diferentes usos e profundidades

A partir de sugestões das análises de componentes principais (ACP_1 , ACP_2 e ACP_3) aplica-se o método multivariado de análise de agrupamentos ou clusters (AAG). Com fins de utilizar um dendograma mais adequado foram combinadas cinco medidas de dissimilaridade (*Euclidean*, *Maximum*, *Manhattan*, *Canberra* e *Minkowski*) e sete métodos de ligação ou formas de construir o dendograma (*Ward*, *Average*, *Complete*, *Single*, *Centroid*, *Median* e

Mcquitty). As 35 combinações possíveis foram avaliadas a partir da correlação cofenética e a maior delas adotada^{84; 85; 87}. Uma vez definido o dendograma, a partição ótima do mesmo foi obtida a partir do critério KL⁸⁶. Os resultados da AAg são mostrados e discutidos nas subseções 4.5 a 4.7 a seguir.

4.5 AAg₁: Clusters a partir da sugestão em ACP₁

A ACP₁ sugere dois thresholds ($P_{cet1} = 1116$ e $P_{cet2} = 1886$) como ótimos na explicação da variabilidade da média de planos de entropia. Para P_{cet1} o dendograma que apresentou melhor correlação cofenética foi contruído com matriz de similaridade *Maximum* e método de agrupamento *Average* (Tabela 13), já a partição ótima pelo critério KL apresentou três grupos de imagens (Figura 46 e Figura 47).

Tabela 13. Correlação cofenética utilizando P_{cet1} em ACP₁

Agrupamento	Matriz de dissimilaridade				
	<i>Euclidean</i>	<i>Maximum</i>	<i>Manhattan</i>	<i>Canberra</i>	<i>Minkowski</i>
<i>Ward</i>	0,773	0,787	0,773	0,059	0,773
<i>Average</i>	0,784	0,792*	0,784	0,058	0,784
<i>Complete</i>	0,767	0,718	0,719	0,059	0,767
<i>Single</i>	0,701	0,732	0,704	0,044	0,701
<i>Centroid</i>	0,785	0,791	0,783	0,058	0,785
<i>Median</i>	0,669	0,58	0,682	0,058	0,669
<i>Mcquitty</i>	0,642	0,788	0,781	0,058	0,642

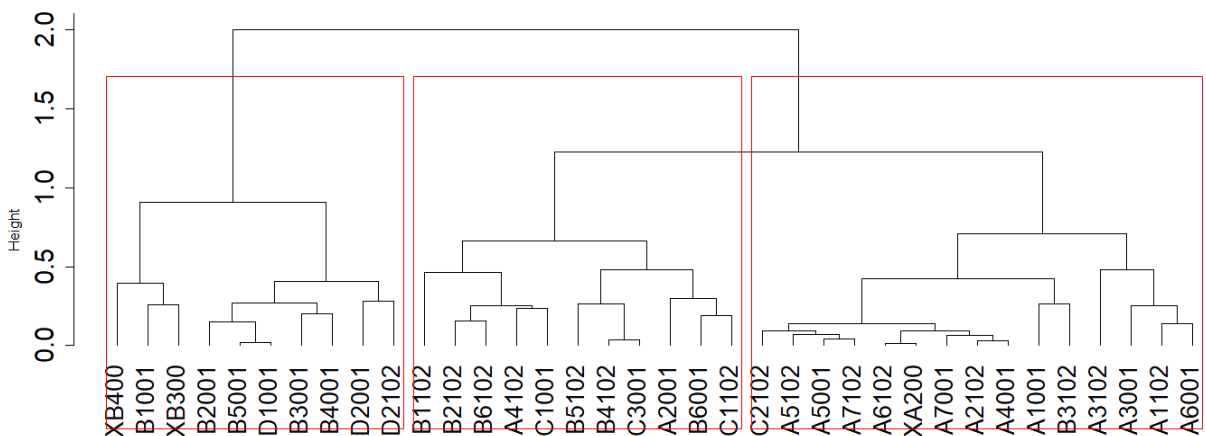


Figura 46. Dendograma da AAg₁ utilizando P_{cet1} em ACP₁

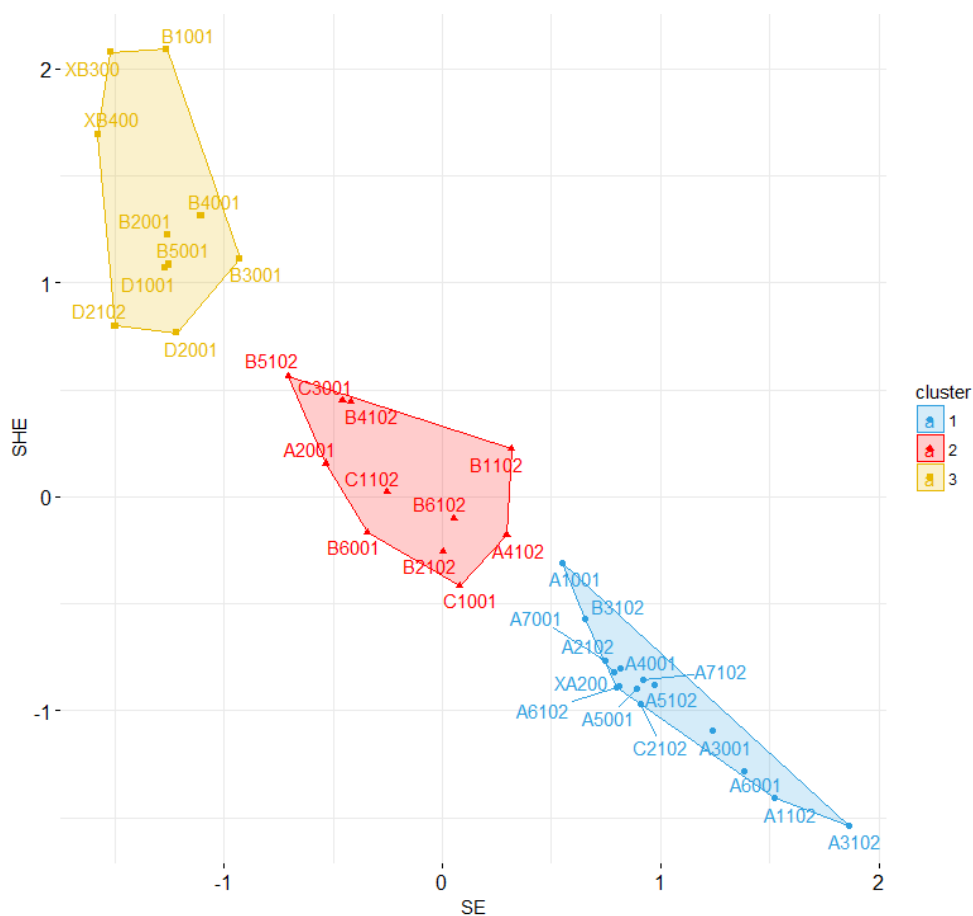


Figura 47. Clusters da AAg₁ utilizando P_{cet1} em ACP₁

Para P_{cet2} o dendograma que apresentou melhor correlação cofenética foi contruído com matriz de similaridade *Manhattan* e método de agrupamento *Average* (Tabela 14) e a partição ótima pelo critério KL apresentou três grupos de imagens (Figura 48 e Figura 49).

Tabela 14. Correlação cofenética utilizando P_{cet2} em ACP₁

Agrupamento	Matriz de dissimilaridade				
	<i>Euclidean</i>	<i>Maximum</i>	<i>Manhattan</i>	<i>Canberra</i>	<i>Minkowski</i>
<i>Ward</i>	0,617	0,606	0,593	0,181	0,617
<i>Average</i>	0,704	0,742	0,754*	0,185	0,704
<i>Complete</i>	0,743	0,692	0,729	0,179	0,743
<i>Single</i>	0,6	0,699	0,545	0,156	0,6
<i>Centroid</i>	0,745	0,702	0,744	0,185	0,745
<i>Median</i>	0,716	0,6	0,572	0,139	0,716
<i>Mcquitty</i>	0,696	0,669	0,749	0,186	0,696

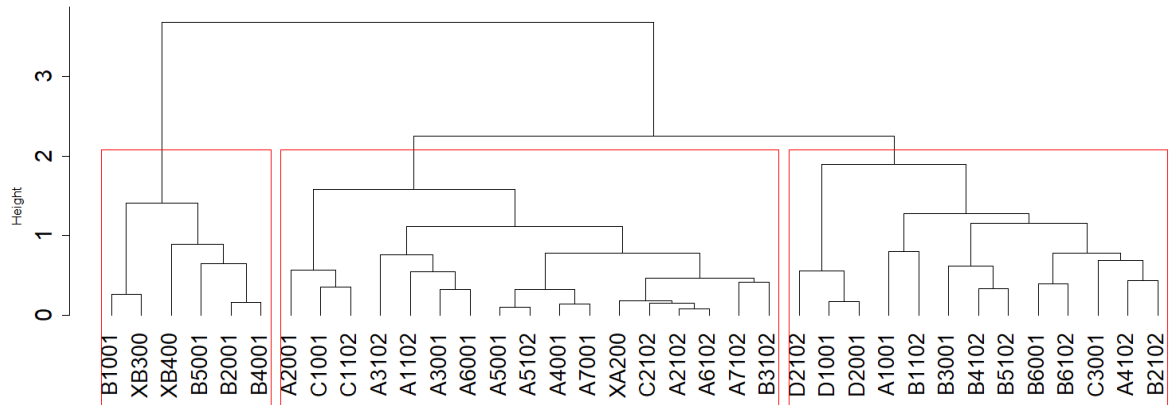


Figura 48. Dendrograma da AAg₁ utilizando P_{cet2} em ACP₁

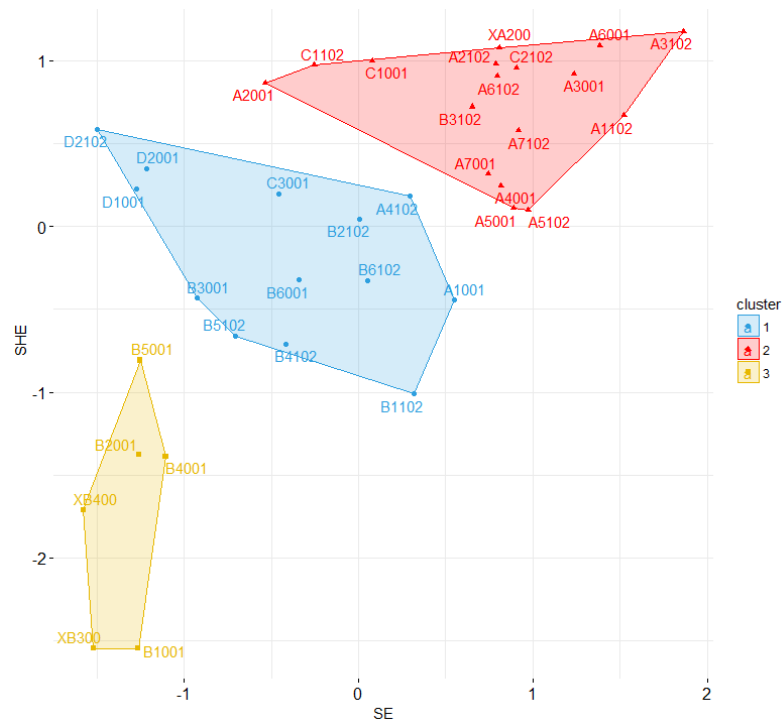


Figura 49. Clusters da AAg₁ utilizando P_{cet2} em ACP₁

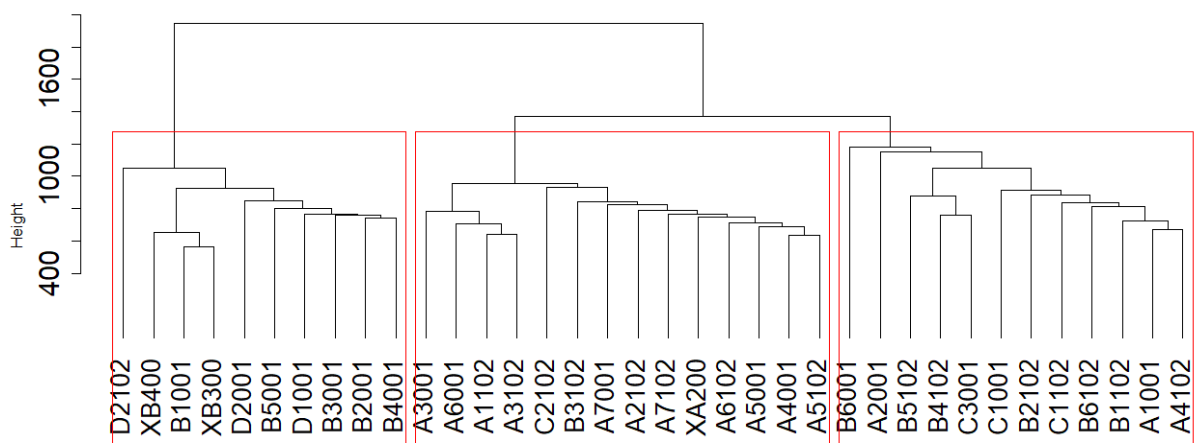
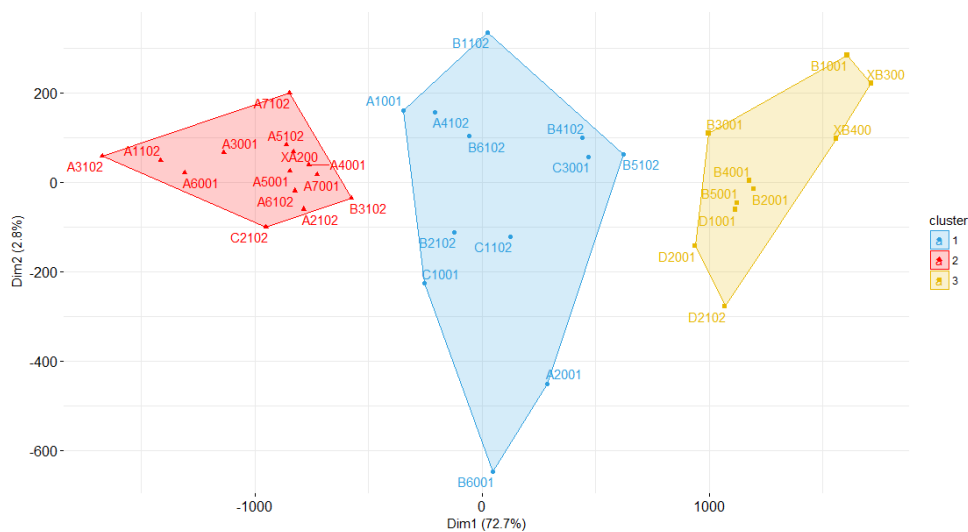
4.6 AAg₂: Clusters a partir da sugestão em ACP₂

A ACP₂ sugere dois *thresholds* ($P_{cet1} = 1161$ e $P_{cet2} = 1887$) como ótimos na explicação da variabilidade dos planos de entropia através de seus *pixels*. Para P_{cet1} o dendrograma que apresentou melhor correlação cofenética foi contruído com matriz de similaridade *Euclidean* e método de agrupamento *Average* (Tabela 15) e a partição ótima pelo critério KL apresentou três grupos de imagens (Figura 50 e Figura 51).

Tabela 15. Correlação cofenética utilizando P_{cet1} em ACP₂

Agrupamento	Matrix de dissimilaridade				
	<i>Euclidean</i>	<i>Maximum</i>	<i>Manhattan</i>	<i>Canberra</i>	<i>Minkowski</i>
<i>Ward</i>	0,74	0,378	0,734	0,285	0,74
<i>Average</i>	0,746**	0,85	0,739	0,297	0,746
<i>Complete</i>	0,66	0,762	0,656	0,279	0,66
<i>Single</i>	0,418	0,832	0,396	0,162	0,418
<i>Centroid</i>	0,718	0,807	0,722	0,297	0,718
<i>Median</i>	0,128	0,783	0,109	0,23	0,128
<i>Mcquitty</i>	0,738	0,861	0,736	0,272	0,738

** Apesar da matriz de dissimilaridade obtida por *Maximum* combinada a maioria das técnicas de agrupamento fornecer correlação cofenética alta, estas combinações não forneceram dendogramas factíveis a seleção ótima de clusters pelo critério KL, sendo assim adotada outra medida de dissimilaridade, *Euclidean*, com a maior correlação cofenética.

Figura 50. Dendrograma da AAg₂ utilizando P_{cet1} em ACP₂Figura 51. Clusters da AAg₂ utilizando P_{cet1} em ACP₂

Para P_{cet2} o dendrograma que apresentou melhor correlação cofenética foi contruído com matriz de similaridade *Manhattan* e método de agrupamento *Average* (Tabela 16) e a partição ótima pelo critério KL apresentou três grupos de imagens (Figura 52 e Figura 53).

Tabela 16. Correlação cofenética utilizando P_{cet1} em ACP_2

Agrupamento	Matrix de dissimilaridade				
	<i>Euclidean</i>	<i>Maximum</i>	<i>Manhattan</i>	<i>Canberra</i>	<i>Minkowski</i>
<i>Ward</i>	0,672	0,365	0,682	0,372	0,672
<i>Average</i>	0,776	0,729	0,769*	0,393	0,776
<i>Complete</i>	0,732	0,594	0,72	0,328	0,732
<i>Single</i>	0,68	0,624	0,635	0,345	0,68
<i>Centroid</i>	0,708	0,593	0,662	0,387	0,708
<i>Median</i>	0,217	0,437	0,151	0,358	0,217
<i>Mcquitty</i>	0,718	0,67	0,726	0,194	0,718

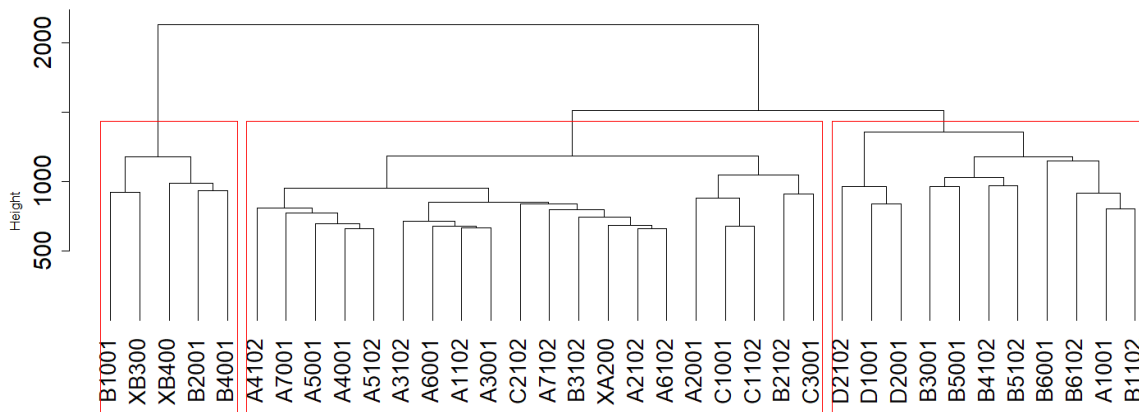


Figura 52. Dendrograma da AAg_2 utilizando P_{cet2} em ACP_2

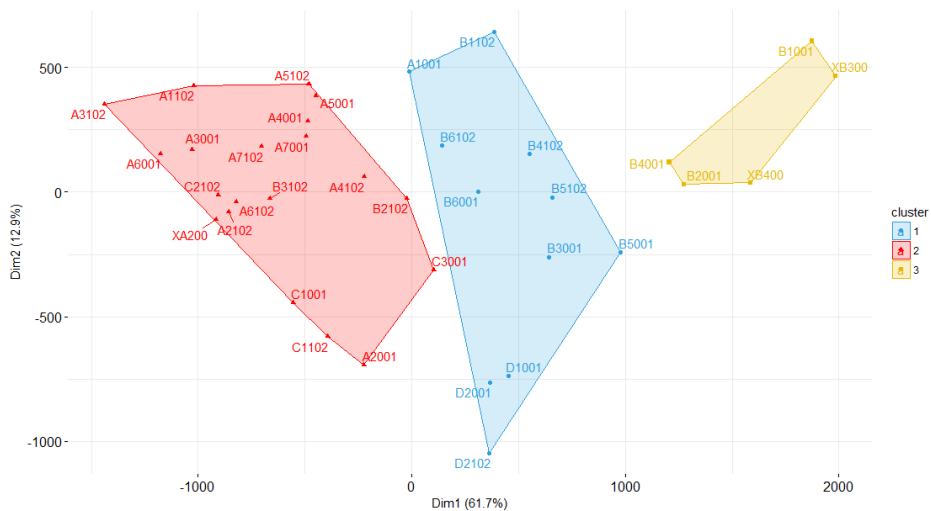


Figura 53. Clusters da AAg_2 utilizando P_{cet2} em ACP_2

4.7 Discussão sobre AAg e planos de entropia

A classificação de imagens de solos por planos de entropia e análise de agrupamentos utilizando os picos da CET evidenciados por ACP₁ e ACP₂ mostrou-se consistente, uma vez que foi possível separar as imagens em grupos que separam os usos e validam os achados e discussão da metodologia de segmentação proposta nesta tese. Em AAg₁ por P_{cet1} em ACP₁ foram separados três grupos, um com 100% de solos de mata, sendo 90% na profundidade de 0-10 cm, um com 93,2% de solos de cana nas duas profundidades e outro solos de mata na profundidade de 10-20 cm e solos de cana (Tabela 17).

Tabela 17. Grupos em AAg₁ utilizando P_{cet1} em ACP₁

Grupo	Imagens	Mata 0-10 cm	Mata 10-20 cm	Cana 0-10 cm	Cana 10-20 cm
1	10	9 (90%)	1 (10%)	0 (0%)	0 (0%)
2	11	1 (9%)	5 (45,5%)	3 (27,3%)	2 (18,1%)
3	15	0 (0%)	1 (6,8%)	7 (46,6%)	7 (46,6%)

Em AAg₁ por P_{cet2} em ACP₁ foram separados três grupos, um com 100% de solos de mata na profundidade de 0-10 cm, outro com 94% de solos de cana nas duas profundidades e outro com 76,8% de solos de mata nas duas profundidades (Tabela 18).

Tabela 18. Grupos em AAg₁ utilizando P_{cet2} em ACP₁

Grupo	Imagens	Mata 0-10 cm	Mata 10-20 cm	Cana 0-10 cm	Cana 10-20 cm
1	6	6 (100%)	0 (0%)	0 (0%)	0 (0%)
2	17	0 (0%)	1 (6%)	8 (47%)	8 (47%)
3	13	4 (30,7%)	6 (46,1%)	2 (15,3%)	1 (7,9%)

Em AAg₂ por P_{cet1} em ACP₂ foram separados três grupos, um com 100% de solos de mata, sendo 90% na profundidade de 0-10 cm, um com 92,8% de solos de cana nas duas profundidades e outro com 41,7% de solos de mata na profundidade de 10-20 cm e 50% de solos de canas duas profundidades (Tabela 19).

Tabela 19. Grupos em AAg₂ utilizando P_{cet1} em ACP₂

Grupo	Imagens	Mata 0-10 cm	Mata 10-20 cm	Cana 0-10 cm	Cana 10-20 cm
1	10	9 (90%)	1 (10%)	0 (0%)	0 (0%)
2	14	0 (0%)	1 (7,2%)	6 (42,8%)	7 (50%)
3	12	1 (8,3%)	5 (41,7%)	4 (33,3%)	2 (16,7%)

Em AAg₂ por P_{cet2} em ACP₂ foram separados três grupos, um com 100% de solos de mata na profundidade de 0-10 cm, outro com 90% de solos de cana nas duas profundidades e outro com 91% de solos de mata nas duas profundidades (Tabela 20).

Tabela 20. Grupos em AAg₂ utilizando P_{cet2} em ACP₂

Grupo	Imagens	Mata 0-10 cm	Mata 10-20 cm	Cana 0-10 cm	Cana 10-20 cm
1	5	5 (100%)	0 (0%)	0 (0%)	0 (0%)
2	20	0 (0%)	2 (10%)	9 (45%)	9 (45%)
3	11	5 (45,5%)	5 (45,5%)	1 (9%)	0 (0%)

A classificação demonstra consistência inclusive quando é adotado como referência P_{cet1} tanto na ACP₁ quanto na ACP₂, uma vez que P_{cet1} pertence a vizinhança de *voxels* de baixa densidade, onde já foi visualizado algum confundimento entre solos de mata de 10 – 20 cm e solos de cana. Quando a referência se dá por P_{cet2} , pertencente a vizinhança de *voxels* de alta densidade, este confundimento é atenuado.

5. CONCLUSÕES

Esta tese demonstrou que confrontar planos de entropia no decorrer da variação de unidades de Hounsfield das imagens de solos, seja por meio da correlação ou da máxima explicabilidade em análise de componentes principais, pode trazer relevantes achados no âmbito da segmentação e classificação destas imagens, respectivamente.

A abordagem por planos de entropia elucidou a contribuição da porosidade na desordem dos solos, seja em nível global, na escala de cinza original, por meio de planos de entropia amostral, ou de forma segmentada, por meio de planos de informação de Shannon, estes últimos foram úteis para evidenciar que solos de mata fornecem mais informação na vizinhança de *voxels* de baixa densidade (VBD) e solos de cana na vizinhança de *voxels* de alta densidade (VAD).

O método de pré-segmentação proposto, o *Entropic Plans Correlations* (EPC), mostrou-se poderoso em evidenciar *voxels* puros da mesma forma que o método *Pure Voxels Extraction* (PVE), até de forma mais conservadora, evitando assim, falsos *voxels* puros, além de retirar a subjetividade das propostas de coeficientes de variação por tentativa e erro com fins de forçar o surgimento do primeiro pico do histograma (P_1), na maioria das vezes muito difícil em solos com pouca porosidade.

Uma calibração das fases puras obtidas por EPC através de imagens segmentadas por PVE seguido da teoria de detecção de bordas (PVESeg), permitiu entender a alocação de *voxels* mistos, além do importante papel que o desvio padrão das imagens fornece nesta alocação, pois foi claro que quanto maior a porosidade, esperam-se desvios padrões maiores. Embora o valor de *Vale* na curva de multisegmentação seja em geral inferior em solos de mata e o tamanho do deslocamento de *Vale* ($Zepc_1$) não apresente diferença significativa entre os usos, este será aplicado ou deslocará um desvio padrão maior em mata, o que faz com que estes solos possam alocar uma maior proporção de *voxels* mistos para os vazios totais. Por outro lado, o tamanho do deslocamento de *Pico* ($Zepc_2$) é mais contundente em solos de cana, mas *Pico* não apresenta diferença significativa entre os usos, ocorre então uma compensação que deixa a alocação de *voxels* mistos para a fase não vazios totais similar entre os usos. Pois, em mata é feito um menor deslocamento de um desvio maior, ao contrário, em cana é feito um maior deslocamento de um desvio menor.

A sugestão de estimar os deslocamentos do desvio padrão ($Zepc_1$ e $Zepc_2$) por meio de regressão linear simples sem intercepto ou pela média dos usos também apresentou bom

desempenho na comparação da porosidade total e na correspondência espacial entre EPC seguindo da teoria de detecção de bordas (EPCSeg) e PVESeg, além de retirar a subjetividade destes deslocamentos. Por fim, a segmentação de outras 24 imagens apresentou resultados consistentes, fortalecendo ainda mais esta proposta.

A utilização de planos de entropia como entrada de classificação de imagens através de análise multivariada permitiu a construção da curva de explicabilidade total (CET) por meio da análise de componentes principais, a qual enfatizou e validou muitos dos achados na construção do método EPCSeg, destacando as vizinhanças de voxels de baixa (VBD) e alta (VAD) densidade, além de agrupar consistentemente imagens nestas vizinhanças por meio de análise de agrupamentos (AAg).

5.1 Sugestões e contribuições para trabalhos futuros

Muito ainda pode ser investigado nesta linha de planos de entropia, deste o uso de outras medidas de entropia, uso de outros materiais e até mesmo alterações na resolução espacial das imagens. A segmentação por EPCSeg pode ainda permitir estudos derivados, tais como análises de lacunaridade, sucularidade, dependência espacial e percolação do espaço poroso dos solos.

REFERÊNCIAS BIBLIOGRÁFICAS

- 1 WANG, W. et al. Comparison of image segmentation methods in simulated 2D and 3D microtomographic images of soil aggregates. **Geoderma**, v. 162, n. 3, p. 231-241, 2011/05/15/ 2011. ISSN 0016-7061. Disponível em: < <http://www.sciencedirect.com/science/article/pii/S0016706111000152> >.
- 2 OJEDA-MAGAÑA, B. et al. Identification of pore spaces in 3D CT soil images using PFCM partitional clustering. **Geoderma**, v. 217–218, p. 90-101, 4// 2014. ISSN 0016-7061. Disponível em: < <http://www.sciencedirect.com/science/article/pii/S0016706113004126> >.
- 3 TAINA, I. A.; HECK, R. J.; ELLIOT, T. R. Application of X-ray computed tomography to soil science: A literature review. **Canadian Journal of Soil Science**, v. 88, n. 1, p. 1-19, 2008/02/01 2008. ISSN 0008-4271. Disponível em: < <https://doi.org/10.4141/CJSS06027> >. Acesso em: 2017/07/16.
- 4 MARTÍN-SOTOCA, J. J. et al. New segmentation method based on fractal properties using singularity maps. **Geoderma**, v. 287, p. 40-53, 2/1/ 2017. ISSN 0016-7061. Disponível em: < <http://www.sciencedirect.com/science/article/pii/S0016706116303925> >.
- 5 OTSU, N. A threshold selection method from gray-level histograms. **IEEE Transactions on Systems, Man and Cybernetics**, v. 9, n. 1, p. 62-66, 1979.
- 6 SUROVÝ, P. et al. Importance of automatic threshold for image segmentation for accurate measurement of fine roots of woody plants. **Forestry Journal**, v. 60, n. 4, p. 244, 2014. ISSN 03231046. Disponível em: < <http://www.degruyter.com/view/j/forj.2014.60.issue-4/forj-2015-0007/forj-2015-0007.xml> >. Acesso em: 2017-02-14t16:26:09.301+01:00.
- 7 RASBAND, W. S. **ImageJ**. Bethesda, Maryland, USA: U. S. National Institutes of Health 1997-2016.
- 8 MANDAVA, A. K.; REGENTOVA, E. E.; BERLI, M. A Semi-automated Segmentation of Soil X-ray Microtomography. **Procedia Computer Science**, v. 46, p. 1468-1475, 2015/01/01 2015. ISSN 1877-0509. Disponível em: < <http://www.sciencedirect.com/science/article/pii/S1877050915001301> >.
- 9 CORTINA-JANUCHS, M. G. et al. Detection of pore space in CT soil images using artificial neural networks. **Biogeosciences**, v. 8, n. 2, p. 279-288, 2011. ISSN 1726-4189. Disponível em: < <http://www.biogeosciences.net/8/279/2011/> >.

- 10 BECKERS, E. et al. Coupling X-ray microtomography and macroscopic soil measurements: a method to enhance near-saturation functions? **Hydrol. Earth Syst. Sci.**, v. 18, n. 5, p. 1805-1817, 2014. ISSN 1607-7938. Disponível em: < <http://www.hydrol-earth-syst-sci.net/18/1805/2014/> >.
- 11 BERTHE, K. A.; VINOY, K. J. A New Method for Segmentation using Fractal Properties of Images. **International Journal of Eletronics & Comunication Technologies**, v. 2, n. 4, 2011. ISSN 2230-7109.
- 12 HAPCA, S. M. et al. New Local Thresholding Method for Soil Images by Minimizing Grayscale Intra-Class Variance. **Vadose Zone Journal**, v. 12, n. 3, 2013.
- 13 BRINK, A. D. Grey-level thresholding of images using a correlation criterion. **Pattern Recognition Letters**, v. 9, n. 5, p. 335-341, 1989/06/01/ 1989. ISSN 0167-8655. Disponível em: < <http://www.sciencedirect.com/science/article/pii/0167865589900627> >.
- 14 KAPUR, J. N.; SAHOO, P. K.; WONG, A. K. C. A new method for gray-level picture thresholding using the entropy of the histogram. **Computer Vision, Graphics, and Image Processing**, v. 29, n. 3, p. 273-285, 1985/03/01/ 1985. ISSN 0734-189X. Disponível em: < <http://www.sciencedirect.com/science/article/pii/0734189X85901252> >.
- 15 LIN, Q.; OU, C. Tsallis entropy and the long-range correlation in image thresholding. **Signal Processing**, v. 92, n. 12, p. 2931-2939, 2012/12/01/ 2012. ISSN 0165-1684. Disponível em: < <http://www.sciencedirect.com/science/article/pii/S0165168412001703> >.
- 16 NIE, F. et al. A novel generalized entropy and its application in image thresholding. **Signal Processing**, v. 134, p. 23-34, 5// 2017. ISSN 0165-1684. Disponível em: < <http://www.sciencedirect.com/science/article/pii/S016516841630295X> >.
- 17 SAHOO, P.; WILKINS, C.; YEAGER, J. Threshold selection using Renyi's entropy. **Pattern Recognition**, v. 30, n. 1, p. 71-84, 1997/01/01/ 1997. ISSN 0031-3203. Disponível em: < <http://www.sciencedirect.com/science/article/pii/S0031320396000659> >.
- 18 ZHU, Q. Y.; ZHUANG, Y. J.; YU, H. Z. Entropy generation due to three-dimensional double-diffusive convection of power-law fluids in heterogeneous porous media. **International Journal of Heat and Mass Transfer**, v. 106, p. 61-82, 3// 2017. ISSN 0017-9310. Disponível em: < <http://www.sciencedirect.com/science/article/pii/S0017931016320385> >.
- 19 MARTÍN, M. Á.; REYES, M.; TAGUAS, F. J. Estimating soil bulk density with information metrics of soil texture. **Geoderma**, v. 287, p. 66-70, 2/1/ 2017. ISSN

- 0016-7061. Disponível em: <
<http://www.sciencedirect.com/science/article/pii/S0016706116303950>>.
- 20 CHAUBEY, A. K. Comparison of The Local and Global Thresholding Methods in Image Segmentation. **World Journal of Research and Review** v. 2, n. 1, p. 01-04, 2016. ISSN 2455-3956.
- 21 COSTA, L. F. et al. Microtomografia computadorizada de raios-X na caracterização morfométrica dos poros de Neossolo Regolítico Eutrofico. **Journal of Environmental Analysis and Progress**, v. 1, n. 1, 2016. ISSN 2525-815X.
- 22 COSTA, L. et al. Espaço poroso em solos brasileiros usando tomografia computadorizada de raios-X (Porous space in Brazilian soils using X-ray computed tomography). **Revista Brasileira de Geografia Física**, v. 9, n. 3, 2016.
- 23 SCHLÜTER, S.; WELLER, U.; VOGEL, H.-J. Segmentation of X-ray microtomography images of soil using gradient masks. **Computers & Geosciences**, v. 36, n. 10, p. 1246-1251, 10// 2010. ISSN 0098-3004. Disponível em: <
<http://www.sciencedirect.com/science/article/pii/S0098300410001457>>.
- 24 TARQUIS, A. M. et al. Pore network complexity and thresholding of 3D soil images. **Ecological Complexity**, v. 6, n. 3, p. 230-239, 2009/09/01/ 2009. ISSN 1476-945X. Disponível em: <
<http://www.sciencedirect.com/science/article/pii/S1476945X09000580>>.
- 25 JEFFERIES, D. A. et al. Characterizing soil surface structure in a temperate tree-based intercropping system using X-ray computed tomography. **Agroforestry Systems**, v. 88, n. 4, p. 645-656, 2014/08/01 2014. ISSN 1572-9680. Disponível em: <
<http://dx.doi.org/10.1007/s10457-014-9699-0>>.
- 26 PIRES, L. F. et al. Soil structure changes induced by tillage systems. **Soil and Tillage Research**, v. 165, p. 66-79, 1// 2017. ISSN 0167-1987. Disponível em: <
<http://www.sciencedirect.com/science/article/pii/S0167198716301313>>.
- 27 CAPLAN, J. S. et al. Nitrogen-mediated effects of elevated CO2 on intra-aggregate soil pore structure. **Global Change Biology**, v. 23, n. 4, p. 1585-1597, 2017. ISSN 1365-2486. Disponível em: < <http://dx.doi.org/10.1111/gcb.13496>>.
- 28 ELLIOT, T. R.; HECK, R. J. A comparison of 2D vs. 3D thresholding of X-ray CT imagery. **Canadian Journal of Soil Science**, v. 87, n. 4, p. 405-412, 2007/08/01 2007. ISSN 0008-4271. Disponível em: < <https://doi.org/10.4141/CJSS06017>>. Acesso em: 2017/07/02.

- 29 MARR, D.; HILDRETH, E. Theory of Edge Detection. **Proceedings of the Royal Society of London, Series B, Biological Sciences**, v. 207, n. 1167, p. 187-217, 1980.
- 30 MORAES, M. C. P. D.; MELLO, K. D.; TOPPA, R. H. Protected areas and agricultural expansion: Biodiversity conservation versus economic growth in the Southeast of Brazil. **Journal of Environmental Management**, v. 188, p. 73-84, 3/1/ 2017. ISSN 0301-4797. Disponível em: < <http://www.sciencedirect.com/science/article/pii/S0301479716309677> >.
- 31 HADDAD, N. M. et al. Habitat fragmentation and its lasting impact on Earth's ecosystems. **Science Advances**, v. 1, n. 2, 2015. Disponível em: < <http://advances.sciencemag.org/content/1/2/e1500052.abstract> >.
- 32 CENTURION, J. F. et al. Influência do cultivo da cana-de-açúcar e da mineralogia da fração argila nas propriedades físicas de latossolos vermelhos. **Revista Brasileira de Ciência do Solo**, v. 31, p. 199-209, 2007. ISSN 0100-0683. Disponível em: < http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-06832007000200002&nrm=iso >.
- 33 CHERUBIN, M. R. et al. Sugarcane expansion in Brazilian tropical soils—Effects of land use change on soil chemical attributes. **Agriculture, Ecosystems & Environment**, v. 211, p. 173-184, 12/15/ 2015. ISSN 0167-8809. Disponível em: < <http://www.sciencedirect.com/science/article/pii/S0167880915002248> >.
- 34 FRANCO, A. L. C. et al. Soil carbon, nitrogen and phosphorus changes under sugarcane expansion in Brazil. **Science of The Total Environment**, v. 515–516, p. 30-38, 5/15/ 2015. ISSN 0048-9697. Disponível em: < <http://www.sciencedirect.com/science/article/pii/S0048969715001618> >.
- 35 BICALHO, E. S. et al. Spatial variability structure of soil CO₂ emission and soil attributes in a sugarcane area. **Agriculture, Ecosystems & Environment**, v. 189, p. 206-215, 2014/05/01/ 2014. ISSN 0167-8809. Disponível em: < <http://www.sciencedirect.com/science/article/pii/S0167880914001911> >.
- 36 FRANCO, A. L. C. et al. Loss of soil (macro)fauna due to the expansion of Brazilian sugarcane acreage. **Science of The Total Environment**, v. 563–564, p. 160-168, 9/1/ 2016. ISSN 0048-9697. Disponível em: < <http://www.sciencedirect.com/science/article/pii/S0048969716308117> >.
- 37 HAGHIGHI, F.; GORJI, M.; SHORAFI, M. A study of the effects of land use changes on soil physical properties and organic matter. **Land Degradation & Development**, v. 21, n. 5, p. 496-502, 2010. ISSN 1099-145X. Disponível em: < <http://dx.doi.org/10.1002/ldr.999> >.

- 38 UNESCO. **Reserva da Biosfera da Mata Atlântica : revisão e atualização dos limites e zoneamento da Reserva da Biosfera da Mata Atlântica em base cartográfica digitalizada : Fase VI** São Paulo: MAB-UNESCO. 38: 119 p. 2009.
- 39 PERNAMBUCO. Lei nº 9.989, de 13 de janeiro de 1987. Define as reservas ecológicas da Região Metropolitana do Recife. 1987. Disponível em: < <http://legis.alepe.pe.gov.br/arquivoTexto.aspx?tiponorma=1&numero=9989&complemento=0&ano=1987&tipo=&url=> >. Acesso em: 29/12.
- 40 _____. Lei nº 14.324, de 3 de junho de 2011. Categoriza as Reservas Ecológicas da Região Metropolitana do Recife, e dá outras providências. 2011. Disponível em: < <http://legis.alepe.pe.gov.br/arquivoTexto.aspx?tiponorma=1&numero=14324&complemento=0&ano=2011&tipo=> >.
- 41 BRASIL. Ministério do Meio Ambiente - Cadastro Nacional de Unidades de Conservação. Unidade de Conservação: Refúgio de Vida Silvestre da Usina São José. Disponível em: < <http://sistemas.mma.gov.br/cnuc/index.php?ido=relatorioparametrizado.exibeRelatorio&relatorioPadrao=true&idUc=2254> >. Acesso em: 28/12.
- 42 RANTA, P. et al. The fragmented Atlantic rain forest of Brazil: size, shape and distribution of forest fragments. **Biodiversity & Conservation**, v. 7, n. 3, p. 385-403, 1998/03/01 1998. ISSN 1572-9710. Disponível em: < <https://doi.org/10.1023/A:1008885813543> >.
- 43 ARAUJO, D.; ALVES, M. Climbing plants of a fragmented area of lowland Atlantic Forest, Igarassu, Pernambuco (northeastern Brazil). **Phytotaxa**, p. 24, 2014-08-31 2010. ISSN 1179-3163. Disponível em: < <http://biotaxa.org/Phytotaxa/article/view/phytotaxa.8.1.1> >.
- 44 LUNA, N. K. M. D.; EDLLEY PESSOA, M. A. Flora da Usina São José, Igarassu, Pernambuco: Zingiberales. **Rodriguésia - Revista do Jardim Botânico do Rio de Janeiro**, v. 67, n. 1, 2016.
- 45 ROCHA, K. D. et al. Caracterização da vegetação arbórea adulta em um fragmento de floresta atlântica, Igarassu, PE. **Revista Brasileira de Ciências Agrárias**, v. 3, n. 1, p. 35-41, 2008.
- 46 NASCIMENTO, L. M. D. **Sucessão secundária em áreas de mata atlântica de Pernambuco: mudanças florísticas e estruturais**. 2010. (Tese de doutorado). Departamento de Biologia - Pós-graduação em Botânica, Universidade Federal Rural de Pernambuco
- 47 SANTOS, S. O.; ALVES, M. Sinopse taxonômica da família Lauracea na porção norte da Floresta Atlântica brasileira. **Revista Brasileira de Biociências**, 2013.

- 48 SCHESSL, M.; SILVA, W. L. D.; GOTTSBERGER, G. Effects of fragmentation on forest structure and litter dynamics in Atlantic rainforest in Pernambuco, Brazil. **Flora - Morphology, Distribution, Functional Ecology of Plants**, v. 203, n. 3, p. 215-228, 4/20/ 2008. ISSN 0367-2530. Disponível em: < <http://www.sciencedirect.com/science/article/pii/S0367253008000133> >.
- 49 PEREIRA, G. A.; ARAÚJO, H. F. P.; AZEVEDO-JÚNIOR, S. M. Distribution and conservation of three important bird groups of the Atlantic Forest in north-east Brazil. **Brazilian Journal of Biology**, v. 76, p. 1004-1020, 2016. ISSN 1519-6984. Disponível em: < http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1519-69842016000401004&nrm=iso >.
- 50 KIMMEL, T. M. et al. Pollination and seed dispersal modes of woody species of 12-year-old secondary forest in the Atlantic Forest region of Pernambuco, NE Brazil. **Flora - Morphology, Distribution, Functional Ecology of Plants**, v. 205, n. 8, p. 540-547, 2010/01/01/ 2010. ISSN 0367-2530. Disponível em: < <http://www.sciencedirect.com/science/article/pii/S0367253009001558> >.
- 51 SCARFE, W. C.; FARMAN, A. G. What is Cone-Beam CT and How Does it Work? **Dental Clinics of North America**, v. 52, n. 4, p. 707-730, 2008/10/01/ 2008. ISSN 0011-8532. Disponível em: < <http://www.sciencedirect.com/science/article/pii/S001185320800044X> >.
- 52 HOUNSFIELD, G. N. **Godfrey N. Hounsfield - Nobel Lecture: Computed Medical Imaging. Nobel Lectures, Physiology or Medicine 1971-1980**. Singapore: Editor Jan Lindsten, World Scientific Publishing Co. 1992.
- 53 BOAS, F. E.; FLEISCHMANN, D. CT artifacts: causes and reduction techniques. **Imaging in Medicine**, v. 4, n. 2, p. 229–240, 2012. ISSN 1755-5191.
- 54 MÜNCH, B. et al. Stripe and ring artifact removal with combined wavelet — Fourier filtering. **Optics Express**, v. 17, n. 10, p. 8567-8591, 2009/05/11 2009. Disponível em: < <http://www.opticsexpress.org/abstract.cfm?URI=oe-17-10-8567> >.
- 55 ZHOUPING, W.; SHELDON, W.; DEAN, C. Ring artifacts removal from synchrotron CT image slices. **Journal of Instrumentation**, v. 8, n. 06, p. C06006, 2013. ISSN 1748-0221. Disponível em: < <http://stacks.iop.org/1748-0221/8/i=06/a=C06006> >.
- 56 SHANNON, C. E. A Mathematical Theory of Communication. **Bell System Technical Journal**, v. 27, n. 3, p. 379-423, 1948. ISSN 1538-7305. Disponível em: < <http://dx.doi.org/10.1002/j.1538-7305.1948.tb01338.x> >.

- 57 SINGH, V. P. THE USE OF ENTROPY IN HYDROLOGY AND WATER RESOURCES. **Hydrological Processes**, v. 11, n. 6, p. 587-626, 1997. ISSN 1099-1085. Disponível em: < [http://dx.doi.org/10.1002/\(SICI\)1099-1085\(199705\)11:6<587::AID-HYP479>3.0.CO;2-P](http://dx.doi.org/10.1002/(SICI)1099-1085(199705)11:6<587::AID-HYP479>3.0.CO;2-P) >.
- 58 RICHMAN, J. S.; MOORMAN, J. R. Physiological time-series analysis using approximate entropy and sample entropy. **American Journal of Physiology - Heart and Circulatory Physiology**, v. 278, n. 6, p. H2039, 2000. Disponível em: < <http://ajpheart.physiology.org/content/278/6/H2039.abstract> >.
- 59 MARTÍN, M. A.; REY, J.-M. On the role of Shannon's entropy as a measure of heterogeneity. **Geoderma**, v. 98, n. 1, p. 1-3, 2000/11/01/ 2000. ISSN 0016-7061. Disponível em: < <http://www.sciencedirect.com/science/article/pii/S0016706100000495> >.
- 60 RODRIGUES DA SILVA, V. D. P. et al. Shannon information entropy for assessing space–time variability of rainfall and streamflow in semiarid region. **Science of The Total Environment**, v. 544, p. 330-338, 2016/02/15/ 2016. ISSN 0048-9697. Disponível em: < <http://www.sciencedirect.com/science/article/pii/S0048969715310639> >.
- 61 IBÁÑEZ, J. J. et al. Pedodiversity and global soil patterns at coarse scales (with discussion). **Geoderma**, v. 83, n. 3, p. 171-192, 1998/05/01/ 1998. ISSN 0016-7061. Disponível em: < <http://www.sciencedirect.com/science/article/pii/S001670619700147X> >.
- 62 CÁMARA, J.; GÓMEZ-MIGUEL, V.; MARTÍN, M. Á. Lithologic control on soil texture heterogeneity. **Geoderma**, v. 287, p. 157-163, 2/1/ 2017. ISSN 0016-7061. Disponível em: < <http://www.sciencedirect.com/science/article/pii/S0016706116303937> >.
- 63 NAZEM, M. et al. A stable Maximum-Entropy Meshless method for analysis of porous media. **Computers and Geotechnics**, v. 80, p. 248-260, 12// 2016. ISSN 0266-352X. Disponível em: < <http://www.sciencedirect.com/science/article/pii/S0266352X16301896> >.
- 64 GAO, S. et al. Estimating the spatial distribution of soil moisture based on Bayesian maximum entropy method with auxiliary data from remote sensing. **International Journal of Applied Earth Observation and Geoinformation**, v. 32, p. 54-66, 10// 2014. ISSN 0303-2434. Disponível em: < <http://www.sciencedirect.com/science/article/pii/S0303243414000531> >.
- 65 WANG, Y. et al. Risk analysis on swell–shrink capacity of expansive soils with efficacy coefficient method and entropy coefficient method. **Applied Clay Science**, v. 99, p. 275-281, 9// 2014. ISSN 0169-1317. Disponível em: < <http://www.sciencedirect.com/science/article/pii/S0169131714002609> >.

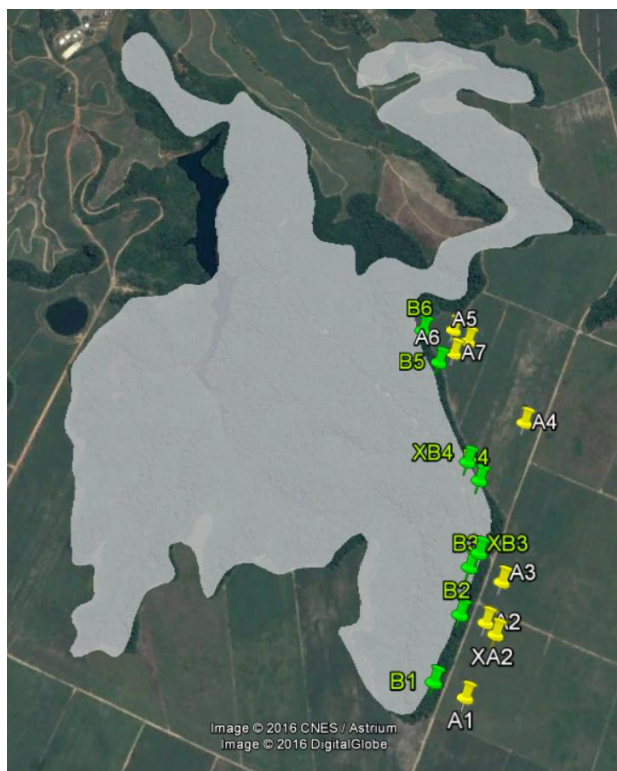
- 66 ZAMBON, I. et al. Land quality, sustainable development and environmental degradation in agricultural districts: A computational approach based on entropy indexes. **Environmental Impact Assessment Review**, v. 64, p. 37-46, 5// 2017. ISSN 0195-9255. Disponível em: < <http://www.sciencedirect.com/science/article/pii/S019592551630364X> >.
- 67 SOHOULANDE DJEBOU, D. C.; SINGH, V. P. Retrieving vegetation growth patterns from soil moisture, precipitation and temperature using maximum entropy. **Ecological Modelling**, v. 309–310, p. 10-21, // 2015. ISSN 0304-3800. Disponível em: < <http://www.sciencedirect.com/science/article/pii/S0304380015001593> >. Acesso em: 2015/8/24/.
- 68 TORRE, I. G.; LOSADA, J. C.; TARQUIS, A. M. Multiscaling properties of soil images. **Biosystems Engineering**, ISSN 1537-5110. Disponível em: < <http://www.sciencedirect.com/science/article/pii/S1537511016305128> >.
- 69 SHUANGCHENG, L. et al. Measurement of climate complexity using sample entropy. **International Journal of Climatology**, v. 26, n. 15, p. 2131-2139, 2006. ISSN 1097-0088. Disponível em: < <http://dx.doi.org/10.1002/joc.1357> >.
- 70 PEARSON, K. Mathematical Contributions to the Theory of Evolution. III. Regression, Heredity, and Panmixia. **Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character**, v. 187, p. 253, 1896. Disponível em: < <http://rsta.royalsocietypublishing.org/content/187/253.abstract> >.
- 71 RATNER, B. The correlation coefficient: Its values range between +1/-1, or do they? **Journal of Targeting, Measurement and Analysis for Marketing**, v. 17, n. 2, p. 139-142, 2009/06/01 2009. ISSN 1479-1862. Disponível em: < <http://dx.doi.org/10.1057/jt.2009.5> >.
- 72 GOLOBOČANIN, D. D.; ŠKRBIĆ, B. D.; MILJEVIĆ, N. R. Principal component analysis for soil contamination with PAHs. **Chemometrics and Intelligent Laboratory Systems**, v. 72, n. 2, p. 219-223, 2004/07/28/ 2004. ISSN 0169-7439. Disponível em: < <http://www.sciencedirect.com/science/article/pii/S0169743904000383> >.
- 73 ŠKRBIĆ, B.; ĐURIŠIĆ-MLADENOVIĆ, N. Principal component analysis for soil contamination with organochlorine compounds. **Chemosphere**, v. 68, n. 11, p. 2144-2152, 2007/08/01/ 2007. ISSN 0045-6535. Disponível em: < <http://www.sciencedirect.com/science/article/pii/S0045653507002020> >.
- 74 ARIAS, O. V. et al. Homogenisation of a soil properties map by principal component analysis to define index agricultural insurance policies. **Geoderma**, 2017/01/23/ 2017.

- ISSN 0016-7061. Disponível em: <
<http://www.sciencedirect.com/science/article/pii/S0016706117300812> >.
- 75 BOLUWADE, A.; MADRAMOOTOO, C. A. Independent principal component analysis for simulation of soil water content and bulk density in a Canadian Watershed. **International Soil and Water Conservation Research**, v. 4, n. 3, p. 151-158, 2016/09/01/ 2016. ISSN 2095-6339. Disponível em: <
<http://www.sciencedirect.com/science/article/pii/S2095633916300065> >.
- 76 LEVI, M. R.; RASMUSSEN, C. Covariate selection with iterative principal component analysis for predicting physical soil properties. **Geoderma**, v. 219, p. 46-57, 2014/05/01/ 2014. ISSN 0016-7061. Disponível em: <
<http://www.sciencedirect.com/science/article/pii/S0016706113004412> >.
- 77 YAO, F.; COQUERY, J.; LÊ CAO, K.-A. Independent Principal Component Analysis for biologically meaningful dimension reduction of large biological data sets. **BMC Bioinformatics**, v. 13, n. 1, p. 24, 2012/02/03 2012. ISSN 1471-2105. Disponível em: <
<http://dx.doi.org/10.1186/1471-2105-13-24> >.
- 78 JOHNSON, R. A.; WICHERN, D. W. **Applied Multivariate Statistical Analysis**. 6th. Pearson, 2007.
- 79 KASSAMBARA, A. **Practical Guide To Cluster Analysis in R: Unsupervised Machine Learning**. 1. STHDA, 2017.
- 80 WARD, J. H. Hierarchical Grouping to Optimize an Objective Function. **Journal of the American Statistical Association**, v. 58, n. 301, p. 236-244, 1963/03/01 1963. ISSN 0162-1459. Disponível em: <
<http://www.tandfonline.com/doi/abs/10.1080/01621459.1963.10500845> >.
- 81 MURTAGH, F.; LEGENDRE, P. Ward's Hierarchical Agglomerative Clustering Method: Which Algorithms Implement Ward's Criterion? **Journal of Classification**, v. 31, n. 3, p. 274-295, 2014/10/01 2014. ISSN 1432-1343. Disponível em: <
<http://dx.doi.org/10.1007/s00357-014-9161-z> >.
- 82 MCQUITTY, L. L. Hierarchical Linkage Analysis for the Isolation of Types. **Educational and Psychological Measurement**, v. 20, n. 1, p. 55-67, 1960/04/01 1960. ISSN 0013-1644. Disponível em: <
<https://doi.org/10.1177/001316446002000106> >. Acesso em: 2017/11/19.
- 83 MURTAGH, F.; CONTRERAS, P. Algorithms for hierarchical clustering: an overview. **Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery**, v. 2, n. 1, p. 86-97, 2012. ISSN 1942-4795. Disponível em: <
<http://https://doi.org/10.1002/widm.53> >.

- 84 R. SOKAL, R.; ROHLF, F. **Sokal RR, Rohlf FJ. The comparison of dendrograms by objective methods. Taxon 11: 33-40.** 1962. 33-40
- 85 SARAÇLI, S.; DOĞAN, N.; DOĞAN, İ. Comparison of hierarchical cluster analysis methods by cophenetic correlation. **Journal of Inequalities and Applications**, v. 2013, n. 1, p. 203, 2013/04/23 2013. ISSN 1029-242X. Disponível em: < <https://doi.org/10.1186/1029-242X-2013-203> >.
- 86 KRZANOWSKI, W. J.; LAI, Y. T. A Criterion for Determining the Number of Groups in a Data Set Using Sum-of-Squares Clustering. **Biometrics**, v. 44, n. 1, p. 23-34, 1988. ISSN 0006341X, 15410420. Disponível em: < <http://www.jstor.org/stable/2531893> >.
- 87 FARRIS, J. **On the Cophenetic Correlation Coefficient.** 1969. 279

APÊNDICE

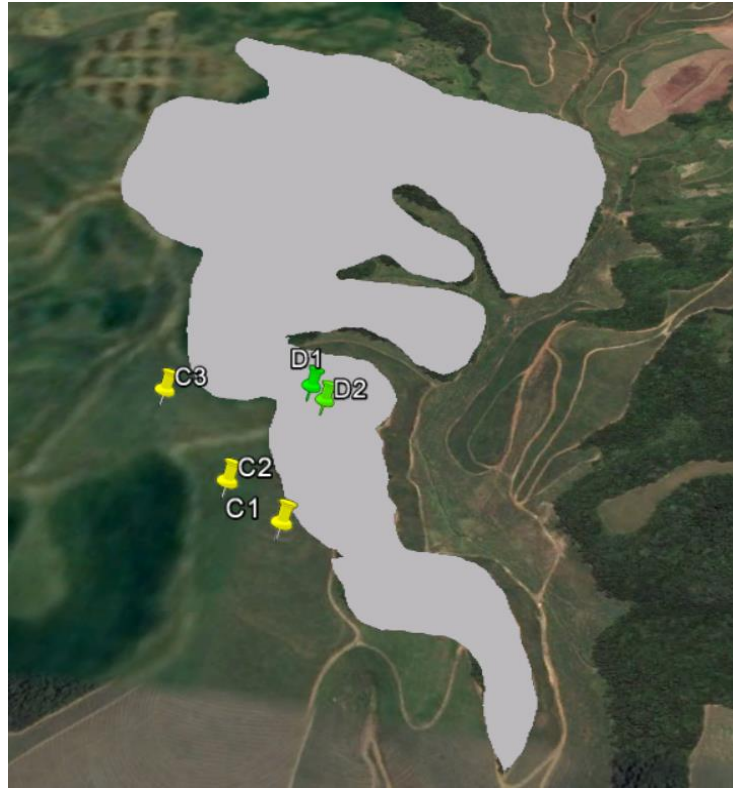
Apêndice 1 – Coordenadas geográficas e classificação de solos dos pontos amostrados em Piedade



Ponto	Latitude	Longitude	Tipo de solo
B1	-7,84791	-34,9966	Espodossolo
B2	-7,84562	-34,99575	Argissolo amarelo / Latossolo amarelo
B3	-7,84396	-34,99542	Espodossolo / Latossolo
XB3	-7,84396	-34,99542	Espodossolo / Latossolo
B4	-7,84084	-34,99511	Latossolo amarelo
XB4	-7,84084	-34,99511	Latossolo amarelo
B5	-7,8365	-34,99666	Latossolo amarelo
B6	-7,83527	-34,99729	Latossolo acinzentado
A1	-7,84836	-34,99551	Espodossolo
A2	-7,84583	-34,99483	Argissolo acinzentado
XA2	-7,84583	-34,99483	Argissolo acinzentado
A3	-7,84441	-34,99428	Espodossolo
A4	-7,8386	-34,99346	Latossolo amarelo
A5	-7,8356	-34,99558	Latossolo amarelo
A6	-7,83519	-34,99612	Latossolo
A7	-7,8361	-34,9961	Latossolo

Coordenadas em graus decimais.

Apêndice 2 – Coordenadas geográficas e classificação dos solos dos pontos amostrados em Cumbe



Ponto	Latitude	Longitude	Tipo de solo
D1	-7,77408	-35,0429	Latossolo vermelho-amarelo
D2	-7,77435	-35,04267	Latossolo vermelho-amarelo
C1	-7,77488	-35,04385	Latossolo vermelho-amarelo
C2	-7,77513	-35,0441	Latossolo vermelho-amarelo
C3	-7,77431	-35,0450	Latossolo vermelho-amarelo

Coordenadas em graus decimais.