

Universidade de Lisboa
Instituto de Geografia e Ordenamento do Território



**Individualização de Copas de Árvores através de Detecção
Remota**

Sara João Correia Lobo Soares

Relatório de estágio orientado pelo Professor Doutor Fernando Jorge Pedro da Silva
Pinto da Rocha e pelo Dr. Daniel Filipe Carvalho Miranda Pires

Mestrado em Sistemas de Informação Geográfica e Modelação Territorial Aplicados ao
Ordenamento

2019

Universidade de Lisboa
Instituto de Geografia e Ordenamento do Território



**Individualização de Copas de Árvores através de Detecção
Remota**

Sara João Correia Lobo Soares

Relatório de estágio orientado pelo Professor Doutor Fernando Jorge Pedro da Silva
Pinto da Rocha e pelo Dr. Daniel Filipe Carvalho Miranda Pires

Júri:

Presidente: Professor Doutor António Manuel Saraiva Lopes do Instituto de Geografia e
Ordenamento do Território da Universidade de Lisboa;

Vogais:

- Investigadora Teresa Alexandra Gonçalves dos Santos Silva do CICS.NOVA Centro
Interdisciplinar de Ciências Sociais da Universidade Nova de Lisboa.

- Professor Doutor Fernando Jorge Pedro da Silva Pinto da Rocha do Instituto de
Geografia e Ordenamento do Território da Universidade de Lisboa.

Resumo

As florestas desempenham um papel de extrema importância na sociedade e, conseqüentemente, o conhecimento sobre ela é essencial para a sua gestão e usufruto sustentável. A recolha de informação florestal é um processo complexo, tradicionalmente alcançado através de inventários florestais.

A Deteção Remota tem sofrido um desenvolvimento crescente que tem vindo a possibilitar soluções inovadoras e eficazes em diversos campos do conhecimento, nomeadamente no que diz respeito à gestão florestal. A Individualização de Copas de Árvores (ICA) a partir de fotografia aérea é uma operação que poderá facilitar a recolha de informação florestal, permitindo a obtenção indireta de parâmetros físicos das árvores, medidas estas de extrema importância para a gestão florestal.

Assim, o presente relatório de estágio tem como principal objetivo explorar diversos algoritmos que permitem a ICA de forma mais automatizada, num sistema florestal característico de Portugal: o Montado. Trata-se de um sistema florestal único no contexto nacional, no qual a monitorização e correta gestão florestal pode ser melhorada e desenvolvida a partir de novas metodologias.

Neste seguimento foi utilizada como base de trabalho uma fotografia aérea de alta resolução correspondente a uma zona de montado e uma metodologia de trabalho dividida em três fases distintas: a construção de objetos representados na imagem, uma classificação por limiares para seleção dos objetos correspondentes a copas de árvores e a validação da cartografia de Copas de Árvores obtida. Foram aplicados cinco algoritmos que permitiram a construção dos objetos representados na imagem, assentes em bases técnicas distintas. Desde os mais usualmente utilizados nesta tarefa, baseados na segmentação de imagens - *Watershed*, Crescimento de Regiões e Multiresolução – até a técnicas mais avançadas, como os algoritmos de aprendizagem automática (*machine learning*) - *Random Forests* e Máxima Entropia. Na fase seguinte foram selecionados os objetos que correspondiam a copas de árvores através de um processo de classificação por limiares, fazendo uso de um conjunto de variáveis espectrais e morfológicas de interesse.

Foram produzidos mapas de Copas de Árvores para cada algoritmo, verificando-se para os algoritmos de segmentação uma exatidão global entre 0,67 - 0,73 e coeficiente de concordância *Kappa* abaixo de 0,46 e para os algoritmos de *machine learning* uma exatidão global entre 0,79 - 0,81 e coeficiente de concordância *Kappa* acima de 0,55. Estes últimos demonstraram-se assim promissores na tarefa de ICA, em detrimento dos já convencionalmente aplicados.

Palavras-chave: Cartografia de Copas de Árvores, *machine learning*, Segmentação de imagens, Montado.

Abstract

Forests play an important role on world's society and, as a result, the knowledge about them is essential for their sustainable management and use. Forest data collection is a complex task, traditionally achieved by forest inventory.

Remote Sensing has undergone a growing development, which has enabled new and effective solutions in several fields of knowledge, especially regarding to forest management. Tree Crown Individualization (TCI) from aerial photography can make forest data collection an easier task, allowing indirect data collection of tree measures, which are extremely important for forest management.

Therefore, the main purpose of this work is to explore several algorithms that allow TCI from remote sensing data, in a more automated way, in a characteristic Portuguese forest: the cork oak forest. It is a national unique forest, in which monitoring and correct forest management can be of improved and developed through new methodologies.

In order to achieve the presented purpose, was developed a methodology divided into three phases, to apply to a high-resolution aerial photography: object construction phase, threshold classification phase for selection of tree crown objects and the Tree Crown Cartography validation phase.

For the first phase, five algorithms were applied to build image objects. From the most commonly used, based on image segmentation – Watershed, Region Growing and Multiresolution – to more advanced techniques, such as machine learning algorithms – Random Forests and Maximum Entropy. Later, objects corresponding to tree crowns were selected by a threshold classification, making use of a variable set of interest.

Tree Crown maps were produced for each algorithm used, with a global accuracy between 0,67-0,73 and a Kappa coefficient of agreement below 0,46 for image segmentation algorithms, and a global accuracy between 0,79-0,81 and a Kappa coefficient of agreement above 0,55 for machine learning algorithms. Thus, the later ones have shown promising results in the TCI task, in opposition to those conventionally applied.

Keywords: Tree Crown Cartography, Image Segmentation, Machine Learning, Cork Oak Forest.

Índice

Resumo	i
Abstract.....	iii
1. Introdução.....	1
1.1. Enquadramento Geral	1
1.2. Objetivos.....	5
1.3. BIOTA – Entidade de Acolhimento	5
1.4. Organização do relatório.....	6
2. Detecção remota de âmbito florestal	7
2.1. Individualização de copas de árvores	10
2.2. Métodos Baseados na Segmentação	14
2.2.1. Watershed.....	17
2.2.2. Valley-Following	19
2.2.3. Crescimento de Regiões	20
2.2.4. Multiresolução.....	21
2.3. Outros métodos	24
2.3.1. Métodos de deteção.....	24
2.3.2. Métodos de Machine Learning.....	25
3. Metodologia.....	29
3.1. Área de estudo	29
3.2. Procedimentos metodológicos e aquisição de informação	29
3.2.1. Fotografia aérea digital	32
3.2.2. Índice Normalizado de Vegetação	32
3.3. Individualização das Copas	33
3.3.1. Construção de objetos e obtenção de variáveis.....	33
3.3.2. Classificação por limiares	44

3.3.3. Avaliação da exatidão	45
4. Apresentação e discussão dos resultados	51
4.1. Cartografia de Copas de Árvores	51
4.2. Medidas de Validação.....	61
4.3. Comparação das abordagens.....	67
5. Considerações finais e perspectivas futuras	69
Referências bibliográficas	73
Anexos.....	81

Índice de Figuras

Figura 2.1. Esquematização do algoritmo <i>Watershed</i> (adaptado de https://www.mathworks.com/)	18
Figura 2.2. Esquematização da Segmentação por Crescimento de Regiões.	20
Figura 2.3. Fluxograma dos parâmetros da segmentação Multiresolução (adaptado de Trimble, 2012).	22
Figura 3.1. Área de estudo.	29
Figura 3.2. Esquema metodológico.	31
Figura 3.3. Esquema explicativo das medidas complementares de avaliação da exatidão (* Relativamente à copa individualizada ou à copa de referência)	48
Figura 3.4. Pontos utilizados na amostra de validação.	49
Figura 4.1. Copas de Árvores obtidas através da abordagem que utiliza o algoritmo MaxEnt.	55

Índice de Tabelas

Tabela 2.1. Síntese dos algoritmos de segmentação de imagens mais utilizados para a ICA	17
Tabela 2.2. Síntese de outros algoritmos.	24
Tabela 3.1. <i>Softwares</i> de operacionalização.	32
Tabela 3.2. Características dos ortofotomapas utilizados (adaptado de DGT, 2018). ...	32
Tabela 3.3. Tentativas selecionadas para o algoritmo <i>Watershed</i>	36
Tabela 3.4. Tentativas selecionadas para o algoritmo de Crescimento de Regiões.	37
Tabela 3.5. Tentativas selecionadas para o algoritmo de Multiresolução.	38
Tabela 3.6. Código de implementação dos algoritmos de <i>Machine Learning</i>	39
Tabela 3.7. Categorização da AAC (adaptado de Hosmer & Lemeshow, 2000).	42
Tabela 3.8. Tentativas selecionadas para os algoritmos de <i>Machine Learning</i>	43
Tabela 3.9. Variáveis e limiares aplicados para a identificação dos objetos representativos de copas de árvores, por abordagem.	45
Tabela 3.10. Categorização dos valores do coeficiente de concordância <i>Kappa</i> (adaptado de Landis & Koch, 1977).	47
Tabela 4.1. Tentativas finais selecionadas para cada abordagem.	51
Tabela 4.2. Exatidão do produtor e utilizador, erros de omissão e comissão e coeficiente de concordância <i>Kappa</i> para as abordagens comparadas.	62
Tabela 4.3. Medidas complementares de Avaliação de Exatidão, tendo em conta a área (m ²) das Copas individualizadas e de referência.	65
Tabela 4.4. Área de sobreposição de Copas.	67

Lista de siglas e acrónimos

DF	Dimensão Fractal
DGT	Direção Geral do Território
DR	Deteção Remota
EC	Erro de Comissão
ENF	Estratégia Nacional Florestal
EO	Erro de Omissão
EP	Exatidão do Produtor
EU	Exatidão do Utilizador
GLCM	<i>Grey Level Co-occurrence Matrix</i>
ICA	Individualização de Copas de Árvores
IF	Índice de Forma
IFN	Inventário Florestal Nacional
IVP	Infravermelho Próximo
LMF	<i>Local Maxima Filtering</i>
MaxEnt	Máxima Entropia
NDVI	<i>Normalized Difference Vegetation Index</i>
OBIA	<i>Object Based Image Analysis</i>
RF	<i>Random Forests</i>
SIG	Sistemas de Informação Geográfica
TM	<i>Template Matching</i>
VANT	Veículo Aéreo Não Tripulado
ViGra	<i>Vision with generic algorithms</i>
VLate	<i>Vector-based Landscape Analysis Tool</i>

Agradecimentos

Deixo aqui os meus calorosos agradecimentos a todos aqueles que me apoiaram e contribuíram para tornar este trabalho possível.

Primeiramente agradeço à BIOTA, em especial à Patrícia Rodrigues e à Sónia Malveiro por me terem acolhido tão bem e proporcionado a oportunidade de trabalhar com vocês e num contexto novo para mim. Um agradecimento especial ao Daniel, pelo tempo que me dedicou (até nas viagens!), pelo apoio e motivação, sempre tão necessário a quem está no início.

Ao meu orientador, Doutor Jorge Rocha, quero agradecer o tempo facultado, os contactos e apoio no decorrer do trabalho, mesmo quando me deixou com ansiedades (ou desespero!).

Um agradecimento muito especial ao César Capinha pela oportunidade que me deu de trabalhar e aprender com ele, e por todo o dispêndio de tempo no ensino de novas ferramentas, sem as quais este trabalho não seria possível. Foram sem dúvida boas experiências que me despertaram novos olhares e continuam a inspirar ideias.

Aos meus colegas de trabalho, em especial ao Eng.º Paulo Marques tenho a agradecer por todo o apoio e enorme flexibilidade proporcionados para terminar esta, que foi para mim, uma difícil jornada. Ao Mário fico grata por toda a motivação e todas as opiniões.

Um obrigado à minha família sem a qual não teria conseguido. Um obrigado especial à minha mãe, por sempre ter acreditado em mim, pela paciência, opiniões sensatas e discussões disparatadas. Ao Ricardo por suportar todas as minhas longas ausências, pela paciência e apoio constantes.

Por fim, mas não menos importante, aos meus amigos que direta ou indiretamente me acompanharam neste percurso. Um especial agradecimento à Beatriz, pelo “ombro amigo” e toda a ajuda, motivação e visitas de controlo de stress ao *Pasto*.

1. Introdução

1.1. Enquadramento Geral

Os sistemas florestais cobrem cerca de 30% da superfície emersa do planeta e constituem os ecossistemas terrestres mais diversificados e ecologicamente complexos, constituindo um recurso natural importante, quer do ponto de vista da preservação do equilíbrio ecológico, quer do ponto de vista do aproveitamento económico (Ribeiro, 2007). Além da imprescindível função fotossintética e consequente captação de carbono, estes sistemas desempenham funções de grande relevância, quer a nível ecológico, quer económico, quer mesmo social. Entre os inúmeros serviços dos ecossistemas florestais, destacam-se: a proteção dos solos contra a erosão; a conservação dos recursos hídricos, da paisagem e da biodiversidade; e as atividades recreativas. Acresce que são fonte de diversos bens como combustíveis, alimentos, madeira e outras matérias-primas (tais como resina, celulose, cortiça) e concentram a maior parte da biodiversidade terrestre, nomeadamente, de espécies vegetais e animais (Portal do Ambiente e do Cidadão, 2018). Mundialmente mais de 2 mil milhões de pessoas dependem direta ou indiretamente dos seus recursos e estas providenciam mais de 13 milhões de postos de trabalho (WWF, 2018).

A gestão adequada dos diversos sistemas florestais, quer a nível económico como a nível de conservação, é fundamental para garantir que esta dinâmica continue a existir de forma sustentável. Esta gestão requer a recolha de um conjunto de dados confiáveis, que permitam um bom planeamento florestal, a curto e a longo prazo (Fransson *et al.*, 2000). Habitualmente, essa informação centra-se em características físicas dos povoamentos florestais, tais como a altura das árvores, a densidade do povoamento, medidas relativas às dimensões das copas e troncos, entre outras (Zandoná *et al.*, 2008). O conhecimento destes parâmetros pode posteriormente ser utilizado num inúmero conjunto de operações, nomeadamente na monitorização e avaliação do crescimento do sistema florestal, na construção de modelos de previsão da produção, na monitorização da biodiversidade, no planeamento florestal e na realização de estudos de conservação das espécies, em monitorizações da desflorestação e da fitossanidade e ainda na modelação climática (Fransson *et al.*, 2000; Liu, 2013; Zandoná *et al.*, 2008). Em muitos países europeus estes

parâmetros físicos, são obtidos através de um inventário florestal, realizado a cada 10 anos (Koch *et al.*, 2006). Em Portugal, o Inventário Florestal Nacional (INF) foi definido, no sentido lato pela Direção Geral das Florestas (DGF) – Relatório Final da 3ª Revisão do IFN de 2001 – como sendo um conjunto de operações que visam avaliar e monitorizar a extensão e condição dos recursos florestais nacionais (Miranda, 2004).

Tradicionalmente estes parâmetros são obtidos através de métodos diretos, ou seja, que implicam a medição direta dos elementos, referentes neste caso às árvores. A utilização destes métodos convencionais produz informação sobre a floresta que é considerada precisa. No entanto nem sempre existe facilidade neste procedimento, uma vez que são métodos que implicam, na sua maioria, um grande custo e refletem também um trabalho intensivo, sendo necessário a deslocação aos locais em estudo para levantamentos de campo (muitas vezes de difícil acesso) (Liu, 2013). Alguns parâmetros são de complexa medição e possuem um erro relevante associado, como é o caso da altura das árvores ou do diâmetro das copas. De facto, é de sublinhar a impraticabilidade da obtenção de um mapa do tamanho das copas através de medições diretas, uma vez que se trata de um procedimento altamente laborioso, com um erro significativo associado à dificuldade na medição direta das dimensões das copas (Song e Dickinson, 2008). Todos os referidos constrangimentos associados aos métodos convencionais de extração de informação florestal incentivam a optar por outros mais inovadores e a apostar no desenvolvimento de ferramentas que permitam a superação dos primeiros, nomeadamente a interpretação, extração e estimação de informação florestal com recurso a métodos indiretos (Fa Gougeon e Leckie, 2003).

A Deteção Remota (DR), em complementaridade com os Sistemas de Informação Geográfica (SIG), tem-se revelado uma ferramenta de elevado valor para o mapeamento e monitorização de sistemas florestais. Estas tecnologias evoluíram rapidamente desde 1950, existindo atualmente uma vasta gama de sensores de satélites e de veículos aéreos disponíveis para adquirir imagens digitais de diversas resoluções, com eficácia e a custos reduzidos (Ke e Quackenbush, 2011c). As imagens produzidas através destas tecnologias (fotografias aéreas, imagens de satélite, etc.) tornaram possível, através da sua interpretação e análise, a recolha indireta de informação florestal, a um custo reduzido (Ribeiro, 2007).

Como tal, a DR permite não só uma melhor compreensão das características do sistema florestal como um todo, devido à elevada cobertura de análise que proporciona, como também a interpretação das zonas florestais a níveis rigorosos (como é o nível da árvore individualizada) de uma forma mais expedita e com menor custo associado, relativamente aos métodos convencionais utilizados para aquisição da mesma informação (Ke e Quackenbush, 2011a). Acresce que o avanço destas tecnologias ainda proporciona novas oportunidades para a interpretação semiautomática ou mesmo automática dos sistemas florestais e da consequente extração de informação - considerando o sistema florestal como um todo ou mesmo ao nível individual.

A Individualização de Copas de Árvores (ICA) através de DR, principalmente a partir de imagens de alta resolução, é um procedimento estruturante na aquisição indireta de informação física, ao nível da árvore (Liu, 2013). Uma ICA correta proporciona a extração de informação florestal fiável, passível de ser utilizada em vários níveis de gestão florestal (Saliola, 2014). Note-se que o seu delineamento eficaz é essencial para a aquisição exata de variados parâmetros florestais (e.g. diâmetro da copa), a partir dos quais se poderá efetuar a estimação de outros que deles derivem (e.g. diâmetro à altura do peito). Assim, será através de um delineamento bem-sucedido que a informação florestal se aproximará com maior rigor da realidade observada.

A floresta portuguesa é diversificada, sendo constituída essencialmente por formações florestais de eucaliptos, carvalhos, pinheiros, sobreiros e azinheiras. A Estratégia Nacional Florestal (ENF) estima que a floresta portuguesa possua um valor económico de 982 milhões de euros, sendo que 58% deste resulta de produtos e serviços como a cortiça, resina, mel, frutos, cogumelos, plantas, caça e recreação e 30% é resultante da produção de madeira. Apenas 12% deriva de serviços ambientais essenciais como a proteção dos solos e da água (DGRF, 2015). O sector florestal é ainda responsável pela criação de cerca de 92 mil (91.583) postos de trabalho, dos quais quase 25 mil (24.678) correspondem a empregos no sector primário e indústrias transformadoras de base florestal (ICNF, 2018).

O sobreiro e a azinheira representam, respetivamente, 23% e 11% da área florestal, tendo-se vindo a registar uma tendência decrescente das áreas ocupadas por estas espécies desde 1995, e particularmente da azinheira desde meados do século XX (Inventário Florestal,

2010, ICNF). Ainda assim, os montados de sobro e de azinho destacam-se de outros sistemas florestais por serem típicos da floresta portuguesa e apresentarem maior diversidade de funções. Possuem um elevado valor em termos de diversidade biológica, qualidade paisagística e, concretamente nos montados de sobro, de significado económico através da produção de cortiça (Belo *et al.*, 2009; Soares *et al.*, 2015). A cortiça representa de facto o principal valor económico dos montados de sobro. Estes sistemas florestais refletem um enorme potencial, não só devido à cortiça, mas também pelas componentes cinegética, recreativa, silvestre e de pecuária que proporcionam.

O grau de multifuncionalidade dos montados de sobro é elevado face à maioria dos outros sistemas florestais existentes nas zonas temperadas e boreais. Tal facto prende-se com as seguintes realidades: (1) a existência de outros produtos florestais passíveis de serem explorados, adicionalmente à cortiça (e que já tiveram a sua importância no passado, como é o caso da madeira); (2) a existência de uma componente agrícola e pecuária (e não apenas da componente florestal), contrariamente ao que se verifica noutros sistemas florestais; e (3) as componentes cinegética e recreativa terem ou poderem ter um peso absoluto e relativo maior, comparativamente a outros sistemas florestais. Este elevado grau de multifuncionalidade dos montados tem uma consequência económica importante - cria uma flexibilidade que permite a sua fácil adaptação a alterações estruturais na procura dos seus produtos (Carvalho, 2002).

Dadas as características particulares do sistema florestal que é o montado e, em particular o montado de sobro, é pertinente a avaliação da viabilidade de utilização de DR para a ICA, no contexto da gestão florestal em Portugal, não só pelas potencialidades destas tecnologias, mas também pela necessidade de extração de informação florestal de forma mais expedita, e simultaneamente fidedigna, num sistema de reconhecida importância no país (Ribeiro, 2007). O processamento de imagens numa perspectiva da ICA apresenta-se assim como a etapa inicial e fundamental para a aquisição de informação florestal de forma indireta e confiável, uma opção a ser explorada e avaliada para complementar ou mesmo substituir os métodos tradicionais utilizados.

1.2. Objetivos

Procura-se efetuar uma reflexão sobre as perspetivas da utilização de DR, nomeadamente de fotografias aéreas de alta resolução, na ICA. Os objetivos deste relatório de estágio são, portanto:

- Avaliar o estado da arte nesta temática;
- Avaliar o potencial de utilização de fotografias aéreas para a individualização de copas de árvores em montado;
- Explorar as abordagens de processamento de imagens na individualização de copas de sobreiros e azinheiras (espécies arbóreas principais do sistema florestal em análise – o montado).
- Testar, a título experimental, a aplicabilidade dos algoritmos *Random Forests* e de Máxima Entropia para a ICA.
- Criação de uma base de conhecimento/trabalho para futura extração de informação florestal, ao nível da árvore.

1.3. BIOTA – Entidade de Acolhimento

O presente estágio decorreu na empresa BIOTA, por meio de um protocolo com duração de 4 meses.

Fundada por Patrícia Rodrigues e Sónia Malveiro, a BIOTA é uma empresa portuguesa que surgiu em 2007, prestadora de consultoria nas áreas da biodiversidade, ecossistemas, qualidade ecológica da água e caudais ecológicos. Estas componentes são promovidas em âmbitos diversificados nomeadamente, no desenvolvimento de Estudos Ambientais de Base, Programas de Monitorização, Minimização e Compensação, Formação e Divulgação. Mais recentemente, vem desenvolvendo novos projetos na área das Tecnologias de Informação, com especial enfoque nos SIG.

O presente relatório engloba as atividades desenvolvidas no estágio, numa temática sugerida pela própria empresa por se revelar uma necessidade no seu dia-a-dia de trabalho.

1.4. Organização do relatório

O relatório de estágio encontra-se estruturado em cinco capítulos. No primeiro procedeu-se a um breve enquadramento da temática em estudo bem como da entidade de acolhimento. Adicionalmente são apresentados os objetivos do estágio e a organização do relatório. No segundo capítulo foi elaborada uma revisão bibliográfica que visa a apresentação do estado-da-arte da temática da utilização de DR em âmbito florestal e da temática da ICA, mediante a utilização de imagens de alta resolução espacial.

O terceiro capítulo é relativo à área de estudo (enquadramento e caracterização), às metodologias e ao processamento dos dados efetuado. No quarto capítulo são apresentados os resultados obtidos e a discussão dos mesmos, em função dos objetivos inicialmente definidos. Por fim, o quinto e último capítulo refere-se às principais conclusões resultantes do trabalho desenvolvido.

2. Detecção remota de âmbito florestal

Lillesand e Kiefer (2000) definem a DR como a ciência de obter informação de um objeto, área ou fenómeno através da análise de dados adquiridos a partir de um dispositivo que não está em contacto com os mesmos. As imagens produzidas por sensores de DR fornecem uma medida física das propriedades de refletância das superfícies. A informação radiométrica espectral, após um processamento adequado em que se atende aos vários fenómenos físicos que lhe deram origem, pode ser utilizada para inferir características das superfícies observadas e assim gerar mapas e parâmetros, para uma determinada região e momento no tempo (Gabriel, 2013).

As imagens óticas provenientes de DR têm sido amplamente utilizadas em âmbito florestal desde meados de 1970, quando começaram a desenvolver-se as primeiras técnicas para distinguir áreas florestadas de não florestadas (Gomes e Maillard, 2016). Desde então, estas imagens foram utilizadas como complemento em diversos estudos florestais, essencialmente para servir de apoio na produção de cartografia de cobertura vegetal e ocupação do solo. Foi amplamente avaliada a utilização de imagens de satélite de baixa e média resolução (mais de 5 metros de resolução espacial) como ferramenta para monitorizar os recursos florestais (Miranda, 2004). Estas têm igualmente sido bastante úteis no mapeamento de risco de incêndio (Dong *et al.*, 2005), na monitorização da desflorestação (Grinand *et al.*, 2013) e em estudos de biodiversidade (Nagendra *et al.*, 2010). São uma vantagem por serem uma fonte de dados de fácil obtenção e, na maior parte das vezes, gratuita.

No entanto apesar das imagens de satélite de baixa e média resolução representarem uma mais-valia para a análise florestal e permitirem complementar variados estudos, verifica-se que a precisão fornecida por elas não é suficiente para determinados objetivos, sendo um fator limitativo para estudos mais detalhados (Gomes e Maillard, 2016). As imagens de satélite de baixa e média resolução podem ser utilizadas para estudos a grandes escalas, mas para objetivos mais ambiciosos e numa escala mais fina, como é o caso da individualização de determinadas espécies ou da obtenção de características físicas de povoamentos florestais, não são as mais adequadas.

Com o aparecimento das imagens de satélite de alta resolução (inferior a 1 metro) e das fotografias aéreas, estudos mais detalhados tornaram-se possíveis. Tal avanço tecnológico permitiu um aumento na resolução espacial, alterou por completo o processamento de imagens no âmbito florestal e gerou uma linha de desenvolvimento de novas formas de extração de informação (M. Gomes e Maillard, 2016). As fotografias aéreas, no entanto, mostram-se mais vantajosas relativamente às imagens de satélite de alta resolução. Possuem uma resolução espacial equivalente ou por vezes até superior, existe flexibilidade na sua recolha (podendo evitar-se condições meteorológicas desfavoráveis) e é possível obtê-las rapidamente e a baixo custo (Chenari *et al.*, 2017; López, 2012). A fotografia aérea tornou ainda possível retratar as zonas mais inacessíveis com rapidez e precisão.

A maioria das decisões de gestão florestal são tomadas com base em conhecimento adquirido à escala do talhão, onde um talhão representa tipicamente uma área discreta da floresta com espécies e composição estrutural similares, sendo as fotografias aéreas muitas vezes utilizadas para o delineamento manual dos seus limites (Culvenor, 2002). Neste contexto, a interpretação visual de fotografias aéreas afirmou-se uma ferramenta de auxílio a levantamentos de campo em estudos florestais. Contudo esta análise visual é algo trabalhosa (Ke e Quackenbush, 2011c).

A crescente procura por recursos florestais juntamente com a necessidade de os gerir de forma sustentável promove uma tendência de gestão florestal e de recolha de informação a escalas mais finas (Culvenor, 2002). Por conseguinte, a utilização de fotografias aéreas de alta resolução no âmbito florestal ganhou um novo rumo, que vem possibilitar não só recolha e análise de informação a grande escala mas também a escalas mais finas, tornando possível nomeadamente a obtenção de medições de características físicas dos povoamentos florestais (e.g., a altura total, diâmetro de copa, densidade de copas, etc.) (Miranda, 2004). Surge então, para além da perspetiva já existente de análise florestal ao nível do talhão ou da floresta como um todo, outra perspetiva de análise das florestas ao nível da árvore individual e das informações contidas nelas, passíveis de serem obtidas indiretamente (Carneiro, 2015). Torna-se assim possível obter, de forma indireta, a descrição de uma floresta na qual as árvores são representadas por estimativas como a posição, a espécie, a altura e o diâmetro do caule (Korpela, 2005).

A evolução crescente dos sistemas computacionais tem proporcionado o desenvolvimento de novas técnicas de análise destas imagens, de forma automática ou semiautomática, levando a uma transição dos métodos convencionais de análise de imagens para métodos computacionais (Gougeon e Leckie, 2003). Consequentemente, surgiram novas possibilidades no reconhecimento automático de características específicas dos sistemas florestais (Culvenor, 2002) e a possibilidade de criar soluções mais eficazes para a extração de informação relativamente à interpretação manual das fotografias aéreas. Desenvolveram-se técnicas que operam no domínio de imagens individuais (2D) e outras que conjugam informação de diversas imagens (3D), existindo hoje uma larga gama de possibilidades de processamento de fotografias aéreas (Korpela, 2005). Esta evolução, aliada à maior facilidade e baixo custo de aquisição das fotografias aéreas de alta resolução espacial, possibilita fontes viáveis de dados e novas oportunidades para na interpretação florestal (Ke e Quackenbush, 2011c).

Apesar das imagens óticas serem a fonte de informação privilegiada de DR para propósitos florestais (Ke e Quackenbush, 2011c), existe outro tipo de dados de alta resolução, como são os dados de *light detection and ranging* (LiDAR) (Liu, 2013). Os dados provenientes de LiDAR representam uma nuvem de pontos 3D que pode ser utilizada para construir Modelos Digitais do Terreno, Modelos Digitais de Superfície e Modelos de Altura da Cobertura de alta resolução, posteriormente utilizáveis para medição direta da altura de objetos (Ribas, 2011).

A capacidade de medir diretamente a altura e derivar outras medidas a partir desta é uma grande vantagem da tecnologia LiDAR, em relação a outras fontes de DR, abrindo outro mundo a ser explorado no que diz respeito à obtenção de informação florestal indireta (Edson e Wing, 2011; Ribas, 2011). Os dados provenientes de LiDAR têm sido testados para utilização em âmbito florestal, demonstrando resultados promissores na medição da estrutura vertical e horizontal de sistemas florestais (Barilotti e Sepic, 2010; Edson e Wing, 2011; Popescu *et al.*, 2003; Ribas, 2011). Mas a sua obtenção implica ainda custos muito avultados, sendo uma desvantagem bastante considerável, que ainda limita a sua ampla utilização (Gatziolis *et al.*, 2015). Presentemente trona-se mais simples e económico obter coberturas aerofotográficas, considerando as novas possibilidades de utilização de veículos aéreos não tripulados, no atual cenário tecnológico.

2.1. Individualização de copas de árvores

Para a extração de informações florestais individuais são necessárias metodologias de processamento de imagens que permitam proceder à etapa da ICA (Jing *et al.*, 2012). Como foi anteriormente referido, este processo é fundamental pois só através do correto delineamento dos limites das copas de cada árvore é possível a obtenção de informação florestal aproximada à real (Culvenor, 2002).

Técnicas e métodos para separar zonas florestadas de não florestadas, delinear copas de árvores e identificar diferentes espécies têm sido foco de atenção nos últimos anos (Fransson *et al.*, 2000). Muitos dos métodos desenvolvidos inicialmente para a ICA incluem delineamentos manuais através de dados de campo ou da interpretação visual de fotografias aéreas (Koch *et al.*, 2006). No entanto, os que permitam a ICA de forma automática ou pelo menos semiautomática tornam-se cada vez mais cruciais, devido à sua maior eficiência e rapidez.

Têm sido desenvolvidos diversos métodos que permitem a ICA de forma automática ou semiautomática, tendo como base fotografias aéreas ou imagens de alta resolução (Ke e Quackenbush, 2011a). Cabe sublinhar que todos os algoritmos desenvolvidos para este objetivo são baseados neste tipo de imagens (Culvenor, 2002).

A ICA pode ser agrupada em duas categorias de técnicas, consoante o objetivo: as técnicas de deteção das árvores e as técnicas de delineamento das respetivas copas (Ke e Quackenbush, 2008). A deteção das árvores refere-se ao processo de identificação dos seus topos ou da sua localização, enquanto o delineamento das copas se relaciona com o desenho automático dos seus limites. No entanto, apesar de conceptualmente distintos, estes dois processos estão muitas vezes interligados, uma vez que pode ser considerado que a deteção de uma árvore é sempre um processo inerente ao delineamento da sua respetiva copa (Ke e Quackenbush, 2011a). Muitos autores partilham desse pensamento, sendo que combinam os dois processos partindo do pressuposto que a deteção da árvore é anterior ao delineamento da copa (e.g., Culvenor, 2002; Wang *et al.*, 2004; Zhang *et al.*, 2010). Outros autores consideram que o processo de delineamento é equivalente ao processo de deteção das árvores, ou seja, que as árvores são detetadas em consequência do delineamento das suas copas (Gomes e Maillard, 2013; Jing *et al.*, 2012; Ke e

Quackenbush, 2007; Larsen *et al.*, 2011). Apenas um número reduzido de estudos aborda exclusivamente o processo de detecção das árvores (e.g., Pouliot *et al.*, 2005; Vauhkonen *et al.*, 2012).

Entende-se, pois, que a divisão dos métodos para a ICA nestes dois processos não é consensual e depende fortemente do objetivo principal da análise. Como tal, no presente relatório adotou-se generalizadamente a designação de métodos de ICA, mas tendo sempre em consideração a existência de *métodos de detecção* (que executam exclusivamente a detecção da árvore) e de *métodos de detecção-delineamento* (que executam o delineamento das copas com a consequente detecção da árvore).

Existem diversos algoritmos desenvolvidos para a ICA. Os primeiros métodos de ICA, desenvolvidos por volta de 1980 (Bayer e Bucher, 2008; Ke e Quackenbush, 2011c), tentavam identificar as árvores através da detecção dos denominados máximos de intensidade local (Bayer e Bucher, 2008). Os máximos de intensidade local têm por base um padrão de refletância que se verifica numa floresta representada através de uma imagem de alta resolução espacial entre os 30 e os 100 cm: os pixéis de maior intensidade correspondem a copas de árvores enquanto os pixéis envolventes, com menor intensidade, correspondem a áreas de sombra entre as copas. Por outras palavras, a refletância diminui à medida que nos aproximamos dos limites das copas (Ke e Quackenbush, 2008). Este padrão de refletância é a base da maioria dos algoritmos de ICA aplicados atualmente: em que o centro de uma copa aparece radiometricamente mais brilhante que os seus limites (Culvenor, 2002).

A maioria dos métodos de ICA foram aplicados e desenvolvidos para casos de florestas de coníferas, de estrutura madura e com densidades moderadas ou altas. Para árvores com uma estrutura cónica, como é o caso das coníferas, o padrão de refletância referido verifica-se claramente, sendo que os picos de refletância correspondem aos topos das árvores (Ke e Quackenbush, 2008). No entanto, em florestas com outras características, o padrão de refletância nem sempre é consistente, podendo surgir diversos picos de refletância numa mesma árvore (Ke e Quackenbush, 2011a). Note-se ainda que, em florestas que possuam outros tipos de espécies dominantes, como por exemplo espécies decíduas, não existe um pico de refletância característico, uma vez que as árvores não apresentam uma estrutura cónica e podem existir lacunas derivadas de ramos longos nas

copas (Ke e Quackenbush, 2011a). Contudo, a utilização de imagens com resolução espacial inferior a 30 cm potencia também este fenómeno de inconsistência no padrão de refletância. Considerando todas estas condicionantes, a aplicação destes algoritmos de ICA em estruturas florestais com outras características das convencionalmente testadas poderá afetar o desempenho do algoritmo e conseqüentemente produzir um resultado menos eficaz (Ke e Quackenbush, 2011a).

Outros métodos surgiram de modo a dar resposta às condicionantes dos algoritmos que se baseiam neste padrão de refletância, surgindo como uma opção para solucionar alguns problemas. Uns passam pela utilização de características morfológicas para detetar as árvores em detrimento das características espectrais, outros pela utilização de soluções com base em algoritmos computacionalmente mais complexos. Podem ser aplicados em casos onde o padrão de refletância não é consistente e não existam picos de refletância que coincidam com os topos das árvores.

Os métodos mais populares para a ICA baseiam-se em técnicas de processamento orientado a objetos, nomeadamente na segmentação de imagens (Gomes e Maillard, 2016). Podendo entender-se a ICA como um tipo de reconhecimento de objeto, métodos que utilizam segmentação de imagens são bastante adequados para a tarefa, uma vez que permitem a separação do objeto de interesse dos restantes (Yang *et al.*, 2014). A segmentação da imagem atua não só como uma forma de deteção do objeto (neste caso a árvore) mas, ao mesmo tempo, de delineamento dos seus limites. Neste contexto, pode-se considerar que o delineamento da copa é equivalente à deteção da árvore, embora alguns métodos de segmentação apliquem uma fase de deteção de árvores para executar o posterior delineamento das copas.

A segmentação *Watershed*, o *Valley-following* e a segmentação por Crescimento de Regiões são os algoritmos com maior uso na ICA. De modo a ter em conta os diferentes tamanhos das árvores e tentando assim minimizar erros associados, diversos autores utilizam também métodos de segmentação por hierarquias (e.g., Chubey *et al.*, 2006; Gomes e Maillard, 2013). O método mais utilizado para a segmentação hierárquica denomina-se segmentação Multiresolução. Esta permite a criação de vários níveis de segmentação e, portanto, de vários níveis de objetos de diferentes tamanhos, que estão relacionados entre si (Trimble, 2012).

Os algoritmos de segmentação da imagem são geralmente utilizados como método único para a ICA, uma vez que são aplicados a florestas moderadamente/altamente densas, onde não existe na imagem mais elementos que sejam individualizados. Muitas vezes para efeitos de extração de informação florestal, a classificação da segmentação resultante é relevante, pois permite reconhecer a espécie de árvore e conseqüentemente aumentar o conhecimento sobre a floresta (Erikson, 2004). Assim, a classificação dos segmentos resultantes é um procedimento também frequente (e.g., Gougeon, 1997; Gougeon e Leckie, 2003; Ke e Quackenbush, 2007).

Para além dos métodos que utilizam algoritmos de segmentação de imagens têm vindo a surgir mais recentemente outras abordagens que envolvem algoritmos mais elaborados e baseados em aprendizagem automática, como é o caso das Redes Neurais Artificiais (Z. Li *et al.*, 2009). Apostar em metodologias mais evoluídas pode ser uma opção de grande interesse para o melhoramento do processo de ICA e a superação das problemáticas verificadas nos métodos já utilizados neste contexto.

Larsen *et al.* (2011) testaram vários métodos de ICA por tipo de sistema florestal, utilizando imagens óticas de alta resolução, e concluiu que todos eles obtiveram melhores resultados ou equivalentes à interpretação humana. No entanto, chegou também à conclusão que a performance dos diversos algoritmos depende fortemente do tipo de imagem utilizado e do sistema florestal em causa, sendo que não existe um método ótimo para todos os cenários. A abordagem ótima para uma determinada situação depende de várias características da imagem e da estrutura da floresta (densidade, mistura de espécies, tamanho das árvores, etc.). As árvores variam na forma, tamanho, propriedades espectrais, altura, tipo de folhagem e densidade. Para além disso o seu contexto espacial varia com a iluminação, tipo de solo e inclinação. Como tal, a tarefa de ICA não é trivial e pode tornar-se altamente complexa dependendo do número de parâmetros envolvidos (Gomes e Maillard, 2016).

Os principais erros que surgem no processo de ICA relacionam-se ou com o incorreto delineamento da copa ou com copas não delineadas. Considerando uma imagem com uma escala definida, copas pequenas relativamente a essa escala podem não ser detetadas devido à sua fraca resposta espectral e espacial na imagem. O erro associado a copas ou

partes de copas não delineadas denomina-se erro de omissão. Por outro lado, copas de grandes dimensões podem levar a um delineamento incorreto, uma vez que ramos mais longos podem ser individualizados como sendo copas únicas. O erro associado ao incorreto delineamento das copas designa-se erro de comissão (Pouliot *et al.*, 2005).

Por forma a minimizar estes erros inerentes à ICA, é importante não só a utilização de uma imagem adequada como também a escolha do método apropriado, de modo a obter os melhores resultados possíveis. No entanto, como já foi referido, a maioria dos métodos desenvolvidos foram aplicados a florestas com maior incidência de espécies coníferas (Gomes e Maillard, 2013). Assim, para outras características florestais não existe ainda um método definido que seja o adequado, sendo muitas vezes necessário testar diversas metodologias de forma a perceber qual a apropriada à situação em causa.

2.2. Métodos Baseados na Segmentação

A segmentação da imagem consiste na divisão de uma imagem em regiões não sobrepostas, que representam objetos reais da superfície terrestre visualizados através dessa mesma imagem. Os algoritmos de segmentação permitem trabalhar com regiões significativas na imagem ao invés dos pixéis individuais (López, 2012; Yang *et al.*, 2014). Como tal são abordagens valiosas pois permitem ter em consideração a forma, cor e texturas dos pixéis, minimizando o erro associado à definição dos limites dos objetos na imagem (Sousa *et al.*, 2011). Trata-se de uma técnica de análise de imagens orientada a objetos (*Object Based Image Analysis* – OBIA) cuja tarefa básica consiste na junção de elementos, com base em parâmetros de homogeneidade ou na respetiva diferenciação de regiões vizinhas (heterogeneidade) (Schiewe, 2002). Existe uma grande variedade de algoritmos de segmentação, mas os mais utilizados baseiam-se numa de duas propriedades básicas de intensidade/cor: a descontinuidade e a similaridade (Ali, 2006).

Os algoritmos de segmentação de imagens podem ser categorizados essencialmente em: (1) algoritmos de segmentação baseada em limites (*Edge-based*), (2) algoritmos de segmentação baseada em regiões (*Region-based*), (3) algoritmos de segmentação por limiares (*Point-based*) e (4) algoritmos de segmentação combinada (Ali, 2006; López, 2012; Schiewe, 2002).

A segmentação baseada em limites (1) é uma das técnicas mais antigas e continua com uma representação expressiva no processamento de imagens. Estas técnicas dependem dos limites encontrados numa imagem, que marcam a localização de descontinuidades, quer em termos de cor quer de textura (Ali, 2006). Existem diversos algoritmos de segmentação baseada em limites, sendo que alguns deles requerem informação *a priori* que vai permitir a construção de resultados mais confiáveis.

As técnicas de segmentação baseada em regiões (2) centram-se em determinar diretamente regiões, ou seja, os elementos disponíveis são testados no que diz respeito à similaridade entre eles, construindo assim regiões (Schiewe, 2002). Estas técnicas dividem-se em dois grandes grupos: Crescimento de Regiões e Divisão/Junção de Regiões. O primeiro cria regiões homogêneas a partir de pontos iniciais. Por outro lado, no segundo a imagem é inicialmente dividida em regiões arbitrárias, sendo estas posteriormente unidas ou divididas de forma a cumprirem determinadas condições predefinidas, não fazendo uso de pontos iniciais (Ali, 2006).

A segmentação por limiares (3) é uma técnica relativamente simples que pode ser utilizada em situações em que os objetos de interesse podem ser identificados como pixéis que tem valores semelhantes e, ao mesmo tempo, os restantes pixéis da imagem tem valores consideravelmente distintos (Ali, 2006).

Apesar da distinção entre estas principais técnicas de segmentação, certos métodos não deixam de ser uma combinação entre os vários algoritmos (4) (Schiewe, 2002).

No âmbito da ICA, que pode por si só também ser considerado um processo de análise de imagens orientado a objetos, a utilização de algoritmos de segmentação de imagens faz todo o sentido. Pensando neste contexto, a base para uma correta individualização de copas de árvores está relacionada com uma correta segmentação da imagem. Uma segmentação adequada implica a delimitação o mais aproximada possível dos objetos existentes na superfície terrestre e representados na imagem. A obtenção de uma imagem segmentada corretamente pode ser considerada sinónimo de uma cartografia de Copas de Árvores de qualidade, com as copas individualizadas a corresponderem o mais aproximado possível às copas reais existentes.

Os algoritmos de segmentação mais populares para a ICA incluem a segmentação por Crescimento de Regiões (Fan *et al.*, 2001; Erikson, 2003), a segmentação *Watershed* (Yang *et al.*, 2014; Niccolai *et al.*, 2010) e o *Valley-following* (Gougeon e Leckie, 2003). Dentro das diferentes implementações disponíveis nos diversos *softwares* SIG do algoritmo de segmentação por Crescimento de Regiões, a segmentação Multiresolução tem vindo a ganhar popularidade. Isto porque, se encontra implementado num *software* especializado na análise de imagens orientadas a objetos, o que permite a realização de uma gama mais variada de operações automatizadas que visam facilitar a individualização de objetos na imagem (Gomes e Maillard, 2013).

Apresentando diferentes graus de complexidade, os diferentes algoritmos utilizados para a segmentação de imagens, e em específico no âmbito da ICA, pretendem responder de forma eficiente a esta tarefa. A Tabela 2.1 apresenta sumariamente os princípios subjacentes aos algoritmos mais populares no contexto da ICA, que serão analisados em maior detalhe nas subsecções seguintes.

Tabela 2.1. Síntese dos algoritmos de segmentação de imagens mais utilizados para a ICA

Algoritmo	Tipo de segmentação	Princípio Base	Principais características na ICA
<i>Watershed</i>	Combinado (Region-based + edge-based)	Construção de objetos através de gradiente da imagem e de analogia com uma superfície topográfica.	Mais adequado para espécies coníferas; Ocorre muitas vezes segmentação excessiva da imagem; Problemas associados a copas de diferentes tamanhos; Tendência para separar a mesma copa em diversos segmentos ou incluir diversas copas um só segmento.
<i>Valley-Following</i>	<i>Edge-detection</i>	Identifica as partes entre as copas de árvores, derivando de uma analogia com uma superfície topográfica	Apropriado para árvores com estruturas cónicas que criam áreas de sombra entre os indivíduos; Maior sucesso em árvores de idade idêntica; Performance reduzida quando as copas têm formas assimétricas; Tendência a juntar árvores pequenas e a criar mais do que um objeto para a mesma copa.
Crescimento de Regiões	<i>Region-based</i>	Agrupa pixéis em regiões baseando-se em critérios predefinidos de forma a separar objetos na imagem.	Copas com estruturas mais complexas são melhor individualizadas; Tendência em criar mais de um objeto para uma mesma copa (com ramos largos) e de agrupar copas diferentes que se encontrem espacialmente próximas.
Multiresolução	<i>Region-based</i>	Individualiza as regiões de acordo com um processo de crescimento de regiões, criando diversos níveis de objetos de diferentes escalas.	Boa performance aquando a existência de árvores de diferentes tamanhos; Aplicação em imagens multiespectrais;

2.2.1. Watershed

Sendo um dos mais frequentes na segmentação de imagens, este algoritmo é utilizado em diversos cenários e temáticas. Considera uma imagem, geralmente na escala dos cinzentos, como uma superfície topográfica onde os valores representam altitude, ou seja, onde o valor digital para cada pixel pode ser considerado como a elevação nesse mesmo ponto (Ke e Quackenbush, 2008, 2011c). Os segmentos (linhas divisórias) correspondem a zonas de valores de intensidade superiores (máximos locais) que isolam as várias

regiões correspondentes a valores de intensidade mínimos (mínimos locais), construindo assim polígonos isolados (as ‘bacias’) (Niccolai *et al.*, 2010). Este algoritmo é aplicado aos gradientes da imagem em vez de ao espaço espectral original, de modo a assegurar que as linhas divisórias correspondam aos limites reais dos objetos na imagem. Este espaço de gradientes é geralmente produzido através de uma imagem na escala dos cinzentos ou de uma única banda espectral (Yang *et al.*, 2014).

Imaginando que as áreas mais claras correspondem a valores mais elevados e as áreas escuras a valores mais baixos, facilmente se deteta a semelhança entre uma superfície topográfica (Figura 2.1). Ora em termos de superfície é natural pensar em bacias hidrográficas e nas respetivas linhas divisórias de água. O objetivo da segmentação *Watershed* resume-se à procura de regiões de gradientes de intensidade elevados que dividem os mínimos locais vizinhos (as ‘bacias’) (Salman e Liu, 2003). A chave por trás da segmentação *Watershed* é a utilização de uma imagem em que as ‘bacias’ correspondam aos objetos que se querem identificar.

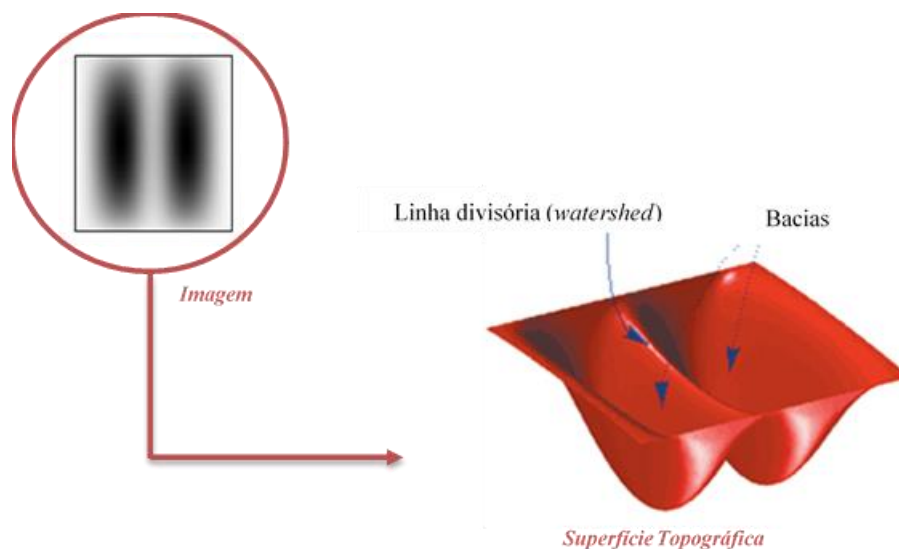


Figura 2.1. Esquemática do algoritmo *Watershed* (adaptado de <https://www.mathworks.com/>)

Uma das maiores desvantagens apontadas aquando da utilização deste algoritmo é o facto de ocorrer uma segmentação excessiva causada pela elevada variação de intensidade em cada região delineada (Ke e Quackenbush, 2011a). Por outro lado, algumas vantagens podem ser apontadas: os limites resultantes formam regiões fechadas e ligadas e esses

mesmos limites das regiões correspondem aos contornos que visualmente aparentam na imagem como contornos óbvios dos objetos (Amoda e Kulkarni, 2013).

No caso específico da ICA, pode-se verificar a separação de uma copa em diferentes objetos quando os ramos estão relativamente afastados, ou a inclusão de várias copas no mesmo quando não existe separação espacial entre as mesmas. Uma variante deste algoritmo poderá permitir colmatar alguns destes problemas, com a utilização de pontos como marcadores das diferentes copas que permitam um melhor controlo do processo de segmentação (Ke e Quackenbush, 2008; Niccolai *et al.*, 2010; Wang *et al.*, 2004).

2.2.2. Valley-Following

Desenvolvido por Gougeon em 1995 especificamente com o objetivo de executar a ICA numa floresta madura de coníferas, o método *Valley-Following* utiliza um princípio semelhante à segmentação *Watershed*, derivando da analogia com uma superfície topográfica. Esta abordagem baseia-se em seguir vales de sombra entre copas de árvores (Tsendbazar, 2011). O fenómeno verificado nas imagens de alta resolução, em que as árvores aparecem geralmente como objetos de maior intensidade rodeados de áreas sombreadas mais escuras, é a base deste algoritmo.

As áreas de intensidades mais baixas ('vales') existentes entre as copas são identificadas e as restantes áreas correspondentes às copas são delineadas com a forma da copa através de um sistema baseado em regras (Gougeon, 1995). O algoritmo encontra os 'vales' através de uma busca pelos mínimos locais de toda a imagem, utilizando uma janela de procura. O pixel central de cada janela de procura é um mínimo local, se todos os outros pixels na janela tiverem um valor superior. Toda a imagem é examinada pela janela de procura até estarem identificados todos os 'vales'.

Contrariamente à segmentação *Watershed*, este algoritmo realiza uma procura não pelos máximos locais de refletância, mas sim pelos mínimos locais, correspondentes aos denominados 'vales'. Assim identifica as zonas de sombra que envolvem as copas das árvores, sendo que os segmentos construídos correspondem aos limites da copa da árvore (Strîmbu e Strîmbu, 2015). Esta abordagem permite obter bons resultados em florestas de coníferas de densidade média a elevada fazendo uso de imagens com resoluções espaciais

entre os 30 e os 60 cm. No entanto, para individualização de árvores decíduas é possível mas geralmente não é bem-sucedida pois a forma das suas copas e a sua distribuição faz com que seja menos comum a presença de áreas sombreadas significativas entre copas (F Gougeon e Leckie, 2006).

2.2.3. Crescimento de Regiões

O algoritmo de Crescimento de Regiões encontra-se amplamente difundido para a execução de tarefas de segmentação de imagens nas mais variadas temáticas e objetivos (e.g., Bins *et al.*, 1996; Pohle e Toennies, 2001; van der Sande *et al.*, 2003). No âmbito da ICA, é igualmente frequente a sua utilização. A segmentação por Crescimento de Regiões inicia-se a partir de um conjunto de pontos, generalizadamente designados por *sementes*. Com base em critérios de homogeneidade espectral, cada *semente* é comparada com os pixéis vizinhos. Dois pixéis vizinhos são agrupados se esses critérios de homogeneidade forem satisfeitos. Este processo dá então origem a uma região de crescimento que herda as propriedades dos pixéis que lhe deram origem (Figura 2.2). Este processo é sequencial até que os critérios de homogeneidade deixem de ser satisfeitos, dando assim origem a regiões espectralmente homogêneas de acordo com os critérios utilizados (Ke e Quackenbush, 2008).

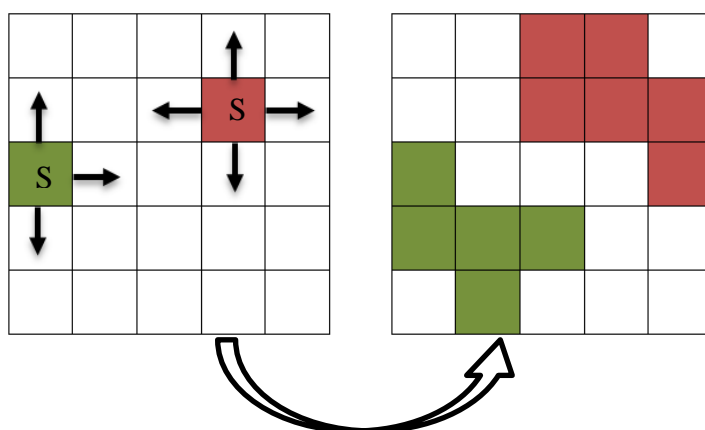


Figura 2.2. Esquematização da Segmentação por Crescimento de Regiões.

Uma das grandes vantagens deste algoritmo é a possibilidade de selecionar as *sementes* mais adequadas para o crescimento das regiões pretendidas, podendo esta seleção ser

manual ou automática (Fan *et al.*, 2001). No caso da ICA a seleção das *sementes* está habitualmente associada à procura pelos topos das árvores, que correspondem efetivamente, em termos espectrais, aos máximos locais na imagem. Os resultados obtidos através deste algoritmo estão altamente dependentes não só da definição adequada dos critérios de homogeneidade para o objetivo em questão, mas também da qualidade das *sementes* utilizadas (Ali, 2006). A utilização de um conjunto de *sementes* que seja representativo da realidade observada é bastante relevante para a construção correta dos objetos da imagem, o mais aproximados possível dos objetos existentes na superfície terrestre (Bechtel *et al.*, 2008).

2.2.4. Multiresolução

A segmentação Multiresolução é uma técnica de Crescimento de Regiões que se encontra desenvolvida num *software* comercial específico, o eCognition®, da *Definiens Imaging*. Este *software* foi especificamente desenvolvido para o processamento de imagens, e mais concretamente direcionado para o processamento de imagens orientado a objetos.

Este algoritmo é fomentado através de um processo heurístico de decisão, baseando-se num critério de fusão, determinado pela heterogeneidade ou homogeneidade das características espectrais e espaciais dos objetos da imagem (Trimble, 2012). A cada passo, um par de objetos é unido num objeto maior, sendo que a decisão dessa junção está relacionada com critérios de homogeneidade, descrevendo a similaridade de objetos adjacentes (Baatz e Schäpe, 2000). No entanto, apesar de ser uma técnica de Crescimento de Regiões, este algoritmo é baseado em parâmetros ajustáveis que definem os critérios de homogeneidade ou de heterogeneidade espacial e espectral.

Desta forma é possível aumentar a aproximação e correspondência entre os objetos da imagem e os da superfície terrestre (Barbosa, 2016). Neste contexto, os critérios de homogeneidade são compostos por quatro parâmetros que definem a homogeneidade total relativa para os objetos de imagem resultantes. Estes parâmetros dividem-se em forma, cor, compacidade e suavidade. A forma e a cor estão inter-relacionadas uma vez que, ao alterar o valor do critério de forma, o de cor é automaticamente modificado. Quanto maior o valor do parâmetro de forma, menor o contributo dos valores espectrais dos objetos da

imagem para o critério de homogeneidade. No caso de o critério de forma apresentar o valor máximo de 1, os objetos resultantes estariam otimizados para a homogeneidade espacial. No entanto não são aconselháveis valores superiores a 0.9 pois caso não exista um mínimo de informação espectral da imagem, os objetos resultantes em nada se irão relacionar com a informação espectral (Trimble, 2012).

Por outro lado, a compacidade e suavidade, também relacionados entre si, tem a função de otimizar os objetos da imagem. O aumento do valor de suavidade, em oposição ao da compacidade, previne a complexidade nas fronteiras dos objetos (Trimble, 2012).

Para além destes, o parâmetro escala também é considerado. Este define a heterogeneidade máxima permitida para os objetos de imagem resultantes, e consequentemente a sua dimensão. Para dados com maior heterogeneidade, os objetos resultantes para um determinado valor deste parâmetro serão de menores dimensões do que para dados com maior homogeneidade (Trimble, 2012). A variação deste parâmetro reflete-se na construção de vários níveis de segmentação, interferindo na dimensão, forma e número de objetos. Quanto maior o seu valor, maiores os objetos resultantes (Barbosa, 2016). A imagem deve ser segmentada numa escala que permita identificar a menor característica de interesse.

A esquematização da organização dos parâmetros possíveis de definir aquando da segmentação de uma imagem com o algoritmo de Multiresolução encontra-se representada na Figura 2.3.

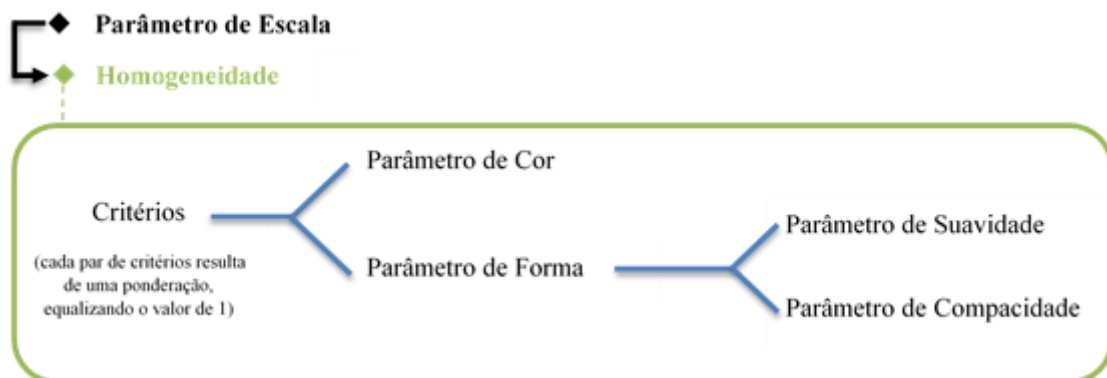


Figura 2.3. Fluxograma dos parâmetros da segmentação Multiresolução (adaptado de Trimble, 2012).

Existe também a possibilidade de ponderar as bandas espectrais da imagem, definindo assim o peso que cada uma terá na implementação do algoritmo. Quanto maior o peso atribuído a uma banda espectral, mais importância será dada à informação dessa mesma banda no processo de segmentação (Trimble, 2012).

Resumidamente, os objetos de imagem obtidos podem ser influenciados através da calibração de parâmetros de escala, critérios de homogeneidade espectral e de forma dos objetos resultantes (Chubey *et al.*, 2006) e da ponderação de bandas. Esta flexibilidade e variedade na calibração do algoritmo é uma grande mais-valia relativamente a outras implementações de segmentação por Crescimento de Regiões.

Outra das grandes vantagens da implementação deste algoritmo é a possibilidade de associar automaticamente a cada objeto, além da informação espectral, informação relativa à forma, tamanho, textura e relações de vizinhança dos objetos (Trimble, 2012). Existe uma vasta gama de métricas que podem ser calculadas para cada objeto de imagem criado, acrescentando assim mais informação passível de ser utilizada posteriormente em processos de classificação (auxiliando na distinção dos objetos de diferentes classes).

Esta técnica oferece ainda a possibilidade de criação de uma segmentação hierárquica, ou seja, a criação de vários níveis de segmentação que representam objetos de diferentes escalas (Trimble, 2012). Os diferentes objetos criados nas diferentes segmentações ficam inter-relacionados, existindo assim uma dependência dos objetos criados em níveis inferiores com os de níveis superiores. Para a ICA esta valência é de extrema importância. Isto porque, sendo os diferentes tamanhos de copas existentes numa floresta uma das principais fontes de erro para a ICA, a implementação deste algoritmo poderá auxiliar na sua redução (Gomes e Maillard, 2013; Ke e Quackenbush, 2007; Saliola, 2014).

Assim, contrariamente ao algoritmo de Crescimento de Regiões implementado no *software* de SIG clássico, este possibilita outro tipo de operações e resultados que podem ser mais adequados para os objetivos específicos da tarefa em análise.

2.3. Outros métodos

Na sequência do que atrás ficou dito (ver secção 2.1) importa sublinhar a existência de outros métodos, que não são especificamente algoritmos de segmentação de imagens. Uns que têm vindo a ser testados e outros que carecem ainda de experimentação, como é o caso de algoritmos de *machine learning*. Na Tabela 2.2 encontram-se resumidos um conjunto de algoritmos, que serão descritos em maior detalhe nas secções seguintes.

Tabela 2.2. Síntese de outros algoritmos.

	Princípio Base	Características
<i>Template Matching (TM)</i>	Calcula a correlação entre a imagem e uma amostra do elemento a identificar, geralmente uma amostra retirada da imagem.	Inclusão de características espectrais, estruturais e texturais; Complicado de utilizar em florestas complexas; Facilidade em detetar copas de maiores dimensões.
<i>Local Maxima Filtering (LMF)</i>	Identificação dos pontos de maior refletância dentro de uma janela de procura definida.	Apropriado para identificação de espécies coníferas; Resultados são afetados por distribuição espacial das árvores, variação dos tamanhos das copas, tamanho da janela de procura.
<i>Random Forests</i>	Classificador binário baseado em <i>machine learning</i> que permite prever o comportamento de uma variável resposta em função de determinadas variáveis explicativas.	Eficiência computacional; Produção de um resultado probabilístico; Classificação binária.
MaxEnt	Classificador binário que permite estimar a distribuição alvo encontrando a distribuição mais uniforme (máxima entropia) sujeita às características derivadas dos dados de treino do modelo.	

2.3.1. Métodos de deteção

Nomeadamente no que se refere à deteção de árvores, são exemplos o método do *Local Maxima Filtering (LMF)* e o método de *Template Matching (TM)*. O método do LMF foi desenvolvido baseando-se no padrão de refletância já descrito (secção 2.1): os topos das árvores são detetados assim que os valores de maior refletância – máximos de intensidade locais - são detetados. Assim, filtrando uma imagem para encontrar os máximos locais detetam-se os topos das árvores (Wang *et al.*, 2004). Esta filtragem é realizada através de uma janela de procura. A janela de procura percorre toda a imagem e o pixel central da janela é identificado como um topo de árvore, se tiver o maior valor dentro dos pixéis da janela. O tamanho da janela é definido pelo utilizador, baseando se na relação entre a

resolução da imagem e o tamanho das copas das árvores. Para florestas com copas de árvores de diferentes tamanhos, janelas grandes causam a não detecção das árvores mais pequenas (maior erro de omissão) enquanto para janelas pequenas as árvores maiores são contabilizadas várias vezes (maior erro de comissão) (Ke e Quackenbush, 2011b). Como tal, a articulação entre o tamanho da janela de procura e a variação no tamanho das copas, é a maior limitação na aplicação deste método. Procedimentos alternativos que tentam utilizar uma janela de procura de tamanho variável foram aplicados de forma a tentar minimizar os erros criados, melhorando significativamente os resultados finais (Wulder *et al.*, 2000).

Já o método TM é uma técnica de processamento de imagem utilizada para o reconhecimento de objetos. Aplica-se através de procura de uma correspondência entre um modelo do objeto de interesse, denominado de *template*, e as várias regiões de uma imagem. São gerados uma série de *templates* que caracterizam o aspeto de um objeto numa imagem, tendo em consideração propriedades geométricas e espectrais. Posteriormente um processo de procura de correlação é implementado, procurando a melhor correspondência entre a imagem e os *templates* gerados. O objeto assume-se como localizado onde a medida de correspondência atinge o valor máximo (Wang *et al.*, 2004).

A forma mais simples de criar os *templates* de um objeto de interesse é obtendo amostras representativas do objeto a partir da imagem. A grande vantagem deste método é o facto de, para além de permitir uma análise das copas de árvore a partir das suas características espectrais, também a permitir quanto às suas características estruturais e texturais. No entanto, o seu uso em florestas mais complexas torna-se muitas vezes impraticável, uma vez que pode não existir um *template* único que represente adequadamente as copas das árvores (Gomes & Maillard, 2016). Assim, os erros de reconhecimento das copas aumentam, não só com a irregularidade das mesmas, mas também com a densidade do sistema florestal em causa

2.3.2. Métodos de Machine Learning

Outros métodos mais recentes e inovadores, nomeadamente algoritmos de aprendizagem automática (*machine learning*), são cada vez mais utilizados no âmbito da ICA (Li *et al.*, 2009; Schroff *et al.*, 2008). As metodologias de classificação binária (*one-class*

classification) são uma das mais promissoras e que têm vindo a ser testadas em diversas operações de análise e processamento de imagens de alta resolução (e.g., Ao *et al.*, 2017; Immitzer *et al.*, 2012; Li e Guo, 2010). Nos métodos de classificação binária, pontos pertencentes a uma classe específica de interesse são denotados como dados de presença, enquanto pontos pertencentes a outras classes são referidos como dados de ausência. Assim, um classificador binário tem como objetivo gerar um modelo de decisão a partir de uma amostra de treino constituída por dados de presença e ausência, para desta forma reconhecer os dados de presença numa amostra de teste (Ao *et al.*, 2017).

Facilmente se reconhece a utilidade deste tipo de algoritmos no âmbito da ICA, sendo esta um processo de individualização de uma única classe de interesse – as copas das árvores. Estes algoritmos de classificação parecem, pois, ajustados ao interesse da ICA. Contudo, apesar da sua utilização frequente em outras operações de processamento de imagem, neste contexto da ICA não foi até á data suficientemente explorado.

Um desses métodos é o algoritmo de *Random Forests* (RF). Trata-se de um algoritmo de Árvores de Decisão, assente em *machine learning*, que tem vindo a ganhar popularidade na comunidade científica, principalmente pela sua aplicabilidade em várias tarefas de classificação supervisionada. Este algoritmo permite, não só a execução de tarefas de classificação, mas também de regressão, permitindo assim a previsão de uma variável resposta (contínua no caso de regressão, categórica no caso de classificação) tendo como base um determinado número de variáveis explicativas.

A ideia base das Árvores de Decisão é que ao nível de cada árvore se tomem decisões acerca da estrutura do nível seguinte, o que se traduz na sucessiva geração de nós descendentes, nos quais a heterogeneidade da variável resposta é sucessivamente atenuada, concebendo-se conseqüentemente previsões progressivamente mais precisas para cada um desses novos nós criados (Santos, 2018). Podem ser apontadas diversas vantagens a este algoritmo, entre as quais a sua eficiência computacional no treino e classificação e o seu resultado probabilístico. Os classificadores de Árvores de Decisão são amplamente utilizados, apesar de sempre terem demonstrado problemas relativamente a falta de generalização e sobreajustamento (*over-fitting*) (Schroff *et al.*, 2008).

Ora a principal ideia das RF é tentar mitigar os problemas encontrados quando se aplicam Árvores de Decisão (Schroff *et al.*, 2008). Inicialmente proposto por Breiman (2001), o algoritmo *Random Forests* consiste na construção de árvores de decisão, usando amostras aleatórias dos dados de treino do modelo e escolhendo aleatoriamente amostras de variáveis para cada árvore em separado. Este algoritmo é distinto do método convencional de Árvores de Decisão, pois para cada árvore é executada uma amostragem aleatória, não só dos dados mas também das variáveis explicativas. Assim, este algoritmo não considera todas as variáveis em cada nó, para determinar o melhor limiar (*threshold*), mas sim uma amostra do conjunto inicial de variáveis.

O número de variáveis por nó é tipicamente a raiz quadrada do número total de variáveis. Estes dois mecanismos, a amostragem aleatória dos dados e o uso de uma amostra de variáveis aleatórias para cada nó, cria árvores muito diferentes entre si e sem correlação alguma (Puissant *et al.*, 2014). Como tal, este algoritmo permitiu inserir aleatoriedade no treino das árvores e combinar o resultado de múltiplas árvores aleatórias num único classificador (Schroff *et al.*, 2008). Cada árvore neste algoritmo é considerada um classificador fraco. No entanto, um grande número de árvores treinadas com dados escolhidos aleatoriamente irá levar a uma única decisão, com base na maioria (Kapás *et al.*, 2016), ou seja, irá definir-se uma classe final para um objeto, baseando-se no voto maioritário de todas as árvores do modelo.

O número de árvores a criar é um parâmetro que deve ser definido pelo utilizador, para que seja suficientemente elevado para capturar toda a variabilidade dos dados de treino do modelo e ao mesmo tempo permitir uma boa precisão na classificação resultante (Puissant *et al.*, 2014). As RF demonstraram um menor erro de teste do que as Árvores de Decisão convencionais, mantendo a eficiência computacional (Schroff *et al.*, 2008).

Outro algoritmo bem conhecido de classificação binária é a abordagem de Máxima Entropia (MaxEnt) proposta por Jaynes (1957). O algoritmo MaxEnt é, à semelhança do RF, uma técnica de *machine learning* que permite a previsão de uma variável resposta binária, em função de um conjunto de variáveis explicativas. Este algoritmo foi desenvolvido para a modelação da distribuição de espécies e para a modelação de nichos ecológicos, utilizando apenas dados de presença de espécies. Permite estimar a distribuição alvo encontrando a distribuição mais uniforme (máxima entropia) sujeita às

características derivadas dos dados de treino do modelo (Ao *et al.*, 2017). Assim, é possível prever a presença e ausência de uma espécie em função de um determinado conjunto de variáveis. Uma vez que esta técnica permite a resolução de um problema de classificação de uma classe, na modelação de nichos ecológicos, também poderá contribuir para a classificação de uma classe no âmbito da DR. Li e Guo (2010) aplicaram este algoritmo a uma imagem de alta resolução (30 cm) para a classificação de uma única classe e comparou os resultados obtidos com um outro método conhecido de classificação binária (*one-class support vector machine*). Os resultados desta comparação revelaram-se promissores, demonstrando a aplicabilidade do MaxEnt na classificação binária de imagens. Ora tendo estes aspetos em consideração, facilmente se transpõe a sua aplicabilidade para a ICA, permitindo classificar a presença e ausência de copas de árvores numa determinada imagem.

A aplicação destes algoritmos no âmbito da ICA, está ainda muito pouco estudada, contudo demonstram um potencial que merece ser explorado devido à grande compatibilidade técnica entre os dois procedimentos. Acresce que poderão solucionar-se algumas das principais limitações, verificadas para os restantes algoritmos utilizadas na ICA, o que seria de grande interesse.

3. Metodologia

3.1. Área de estudo

A região escolhida para a aplicação deste estudo corresponde a uma área de montado de sobre e azinho (distrito de Setúbal, concelho de Santiago do Cacém, freguesia de Abela, 37°59' N / 8°33' W) (Figura 3.1). Trata-se de uma herdade denominada de Monte das Lebres com uma extensão de cerca de 58 hectares. A escolha desta área deve-se ao facto de ser de fácil acesso e acima de tudo de interesse para a BIOTA, pelo facto de representar um sistema florestal dominado por sobreiros (*Quercus suber*) e azinheiras (*Quercus ilex*), espécies florestais de grande interesse para a extração de informação no âmbito do trabalho realizado pela empresa.

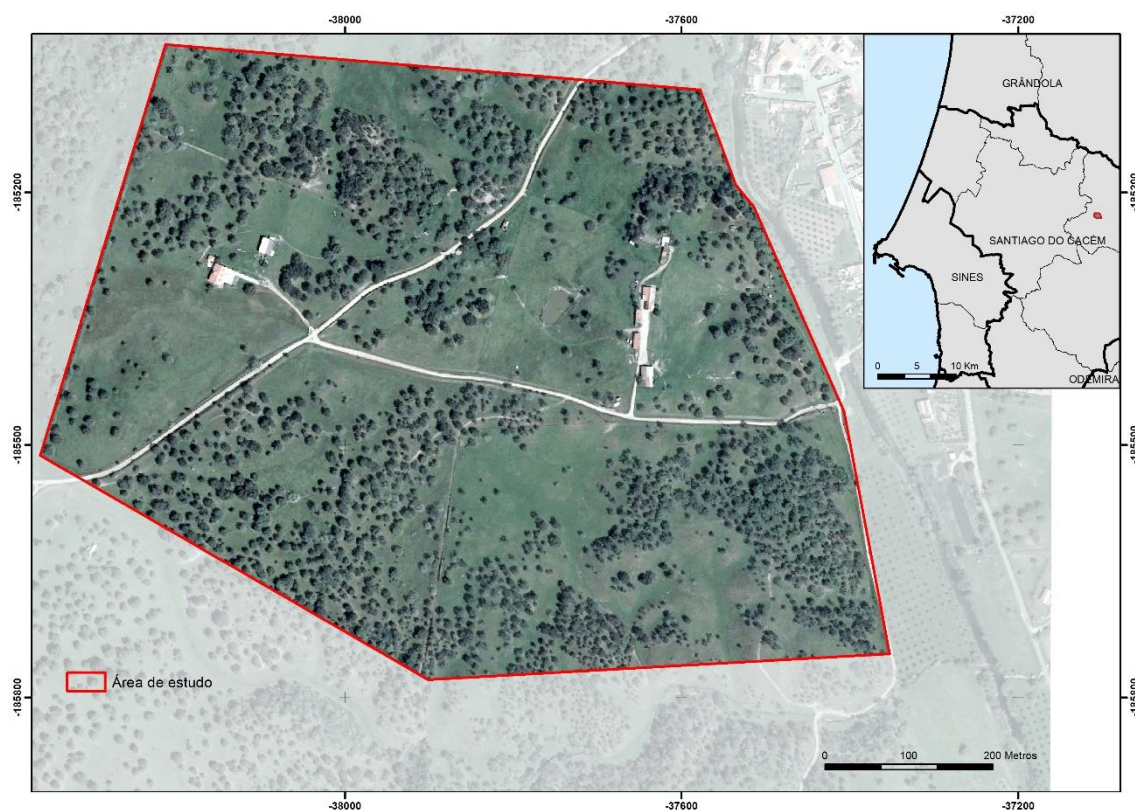


Figura 3.1. Área de estudo.

3.2. Procedimentos metodológicos e aquisição de informação

Os procedimentos metodológicos utilizados neste estudo para a ICA podem ser divididos em três etapas principais: (1) a obtenção de objetos e variáveis, (2) classificação por

limiares e (3) a validação dos resultados. As primeiras duas etapas correspondem ao processo de ICA propriamente dito.

A primeira etapa metodológica (1) foi conseguida através da aplicação de diversos algoritmos, subdivididos em dois grupos: os algoritmos de segmentação de imagens e os algoritmos de *machine learning*. Estes representam abordagens distintas para a criação dos objetos na imagem. Esta etapa incluiu também o cálculo das variáveis que foram necessárias na etapa seguinte.

A segunda etapa metodológica (2) representa a fase de individualização dos objetos que correspondem efetivamente a copas de árvores, recorrendo à definição de limiares para as variáveis calculadas na etapa anterior que permitam a sua correta distinção. Apesar de frequentemente a ICA ser um processo único, no caso específico em análise foi necessária a implementação desta etapa metodológica intermédia, posterior à aplicação do algoritmo de ICA. Por se tratar de um sistema florestal de densidade reduzida, relativamente aos sistemas florestais onde é habitualmente aplicada a ICA, existem outros elementos na imagem, para além das copas das árvores, que também são delineados aquando da aplicação do algoritmo. Para tal, de modo a selecionar apenas os objetos de interesse – as copas – é necessária esta etapa que permite filtrar os objetos que representam outros elementos.

Por último, a terceira etapa metodológica (3) representa a validação dos resultados obtidos para cada uma das abordagens, sendo essencial para a definição de conclusões.

No intuito de tornar mais perceptíveis os diversos procedimentos metodológicos realizados, e a forma como se relacionam entre si, é apresentado um modelo conceptual dos processos afetos à ICA (**Figura 3.2**).

Os diversos procedimentos efetuados foram operacionalizados maioritariamente em ambiente SIG (ArcGIS 10.5; Saga GIS; *eCognition Developer*), com o sistema de projeção PT-TM06/ETRS89, sendo que os algoritmos de *machine learning* foram implementados em ambiente de desenvolvimento integrado (*Rstudio*) e com recurso a linguagem de programação *R*, sendo posteriormente realizadas operações em ambiente SIG (ArcGIS 10.5). Todas as operações de validação dos resultados obtidos a partir dos

diversos procedimentos metodológicos decorreram em ambiente SIG (ArcGIS 10.5). Uma breve descrição dos mesmos encontra-se apresentada na Tabela 3.1.

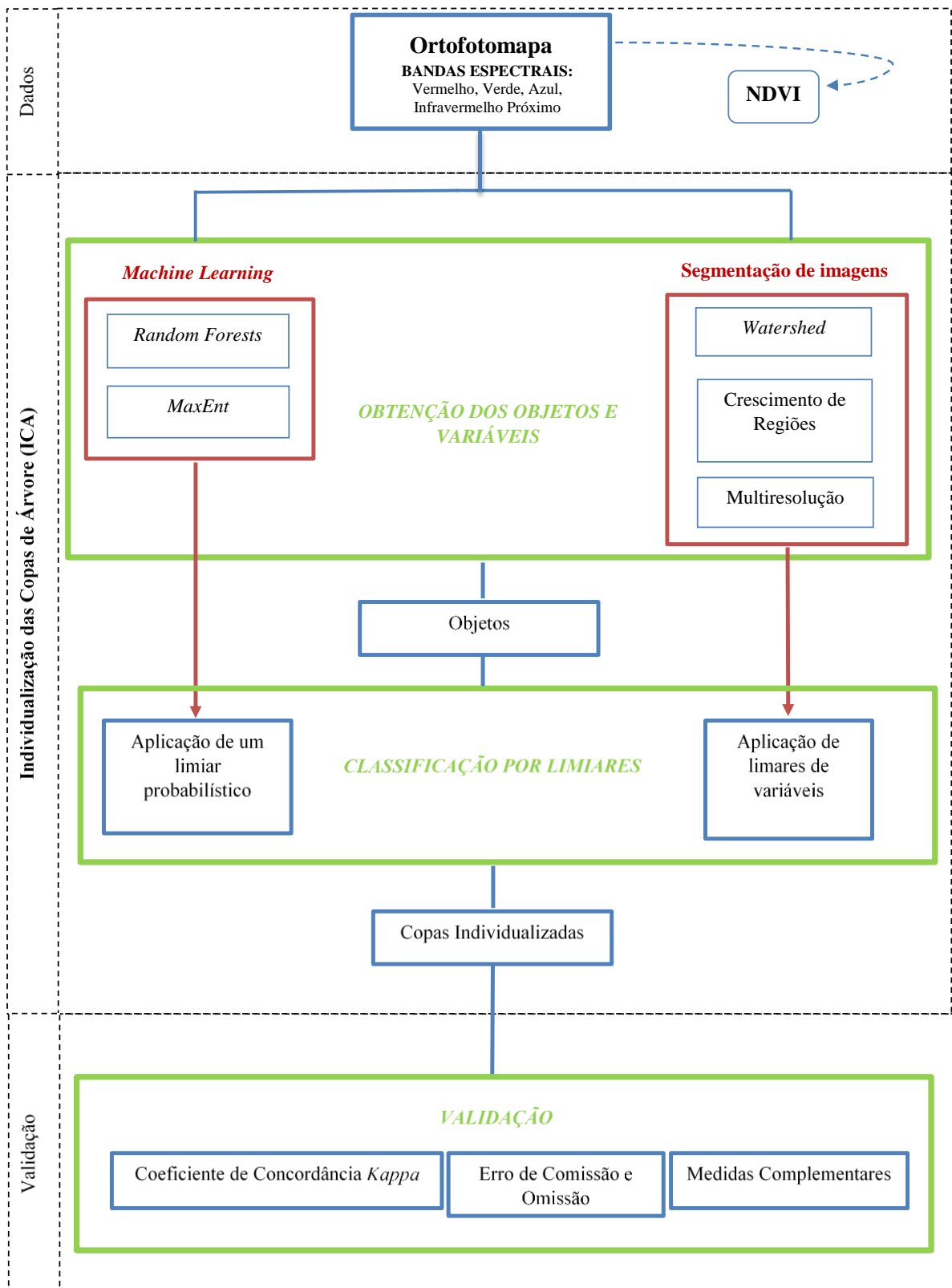






Figura 3.2. Esquema metodológico.

Tabela 3.1. Softwares de operacionalização.

Programa	Versão	Fonte	Descrição
 SAGA <i>System for Automated Geoscientific Analyses</i>	2.3.2	Livre	SIG para trabalhar com informação geográfica.
 ArcGIS ESRI	10.5	ESRI	SIG para trabalhar com informação geográfica.
 eCognition	9	Trimble Navigation	Software de análise de imagens orientada a objetos.
 R Studio	3.5.1	Livre	Ambiente de desenvolvimento integrado com recurso a linguagem R para análise estatística de dados.

3.2.1. Fotografia aérea digital

Foram utilizados os ortofotomapas disponibilizados pela Direção Geral do Território (DGT), produzidos através de fotografia aérea digital, respeitantes ao ano de 2015. Estão disponíveis quatro bandas na imagem: vermelho (610-660 nm), verde (535 – 585 nm), azul (430-490 nm) e o Infravermelho Próximo (760-900 nm). As características dos ortofotomapas encontram-se sumarizadas na Tabela 3.2.

Tabela 3.2. Características dos ortofotomapas utilizados (adaptado de DGT, 2018).

Parâmetros	Valores
Data de recolha	Entre 1 e 13 Abril 2015
Origem	Fotografia digital, obtida com câmara aerofotogramétrica
Câmaras Aerofotogramétricas	UltraCam Xp-wa, Vexcel
Orientação externa	GNSS e IMU
Resolução Radiométrica/Espectral	RGB e Infravermelho Próximo (NIR)
Resolução Espacial	50 cm
Altura de voo aproximada	6000 m
Sistema de referência	PT-TM06/ETRS89

3.2.2. Índice Normalizado de Vegetação

O NDVI (*Normalized Difference Vegetation Index*) é um indicador que avalia a condição geral da vegetação, analisando a atividade fotossintética, e que permite ainda a distinção da vegetação de outros elementos. Para a vegetação, o comprimento de onda de maior

absorção é na faixa dos 650 nm, correspondente à cor vermelha do visível. Já os comprimentos de onda refletidos pela vegetação centram-se entre os 760 nm e os 900 nm, que correspondem ao Infravermelho Próximo (IVP), i.e., *red edge*. Ora o NDVI é obtido através da razão entre a diferença da refletância do IVP e do Vermelho e a soma das mesmas variáveis:

$$NDVI = \frac{IVP + \text{vermelho}}{IVP - \text{vermelho}} \quad (3.1)$$

Este indicador pode variar entre -1 e 1, onde os valores iguais ou inferiores a zero correspondem a ausência de vegetação (Huete *et al.*, 2002). Evidentemente que esta informação é de grande utilidade para a ICA, uma vez que poderá auxiliar na distinção de vegetação relativamente a outros elementos na imagem.

Assim, atendendo à relevância deste índice, procedeu-se ao seu cálculo através das bandas de imagem respetivas, fazendo uso da ferramenta de cálculo de Índices de Vegetação do *software* SAGA GIS.

3.3. Individualização das Copas

3.3.1. Construção de objetos e obtenção de variáveis

A obtenção de objetos a partir de diferentes algoritmos de processamento de imagens constituiu a primeira etapa e a mais importante no processo metodológico da ICA, sendo fundamental para o desenvolvimento de uma cartografia de Copas de Árvores de qualidade. A base para uma ICA bem-sucedida implica a delimitação o mais aproximada possível da realidade dos objetos existentes na imagem. Como tal, o teste de diferentes algoritmos é necessário para, desta forma, obter o conjunto de objetos que melhor representam os elementos na imagem, nomeadamente as copas das árvores. A obtenção de uma correta imagem segmentada facilita a posterior seleção dos objetos de interesse (correspondentes às copas das árvores) através de classificação por limiares, fazendo uso de um leque de variáveis de interesse.

As variáveis utilizadas para a definição destes limiares foram variáveis de natureza espectral, das bandas utilizadas, mas também de natureza morfológica, permitindo a melhor diferenciação entre os diversos objetos. Todas as variáveis utilizadas foram calculadas aquando da obtenção dos objetos. A única exceção ocorreu com as abordagens de algoritmos de *machine learning* onde não existiu o cálculo das variáveis para utilizar na classificação de limiares. Isto, em consequência do resultado probabilístico característico destes algoritmos, distinto dos resultados obtidos pelas restantes abordagens.

As variáveis de cariz morfológico foram obtidas para todas as abordagens, com exceção do algoritmo de Multiresolução, com recurso a uma extensão do ArcMap 10.5 denominada VLate (*Vector-based Analysis Tools*). Esta ferramenta é específica para a análise de paisagem, providenciando o conjunto das métricas mais comuns na análise dos seus padrões. Estas métricas estão agrupadas em grupos (área, forma, limites, proximidade, diversidade, entre outros) que descrevem em geral a forma, configuração e composição da paisagem e cobrem os principais aspetos da estrutura da mesma (ArcGIS, 2018).

A partir desta ferramenta é possível obter diversas métricas relacionadas com a forma de cada um dos objetos. Dentro de todas as métricas a utilizar e possíveis de calcular, consideraram-se mais relevantes o Índice de Forma (IF) e a Dimensão Fractal (DF). O IF dá uma medida do desvio de um objeto relativamente à forma quadrada. Ou seja, mede a complexidade da forma de um objeto comparada a um objeto *standard* (quadrado) do mesmo tamanho. Este índice apresenta valor unitário para objetos quadrados e aumenta (sem limite) à medida que estes se tornam geometricamente mais complexos (Turner *et al.*, 2001). Por sua vez, a DF mede a irregularidade de um objeto no que diz respeito aos seus contornos, através de uma proporção perímetro-área. O seu valor é menor quando o objeto possui um perímetro reduzido em relação à sua área. Se o objeto for mais complexo e fragmentado, o perímetro aumenta e induz um valor de DF superior (Turner *et al.*, 2001).

Dentro das variáveis espectrais utilizadas destacam-se: as médias das bandas Vermelha, Verde, Azul e IVP, do NDVI e do brilho. Todas estas as variáveis foram também calculadas com recurso ao ArcMap 10.5.

A exceção ao cálculo das variáveis utilizadas ocorre na implementação do algoritmo de Multiresolução. Uma vez que a implementação desse algoritmo permite a extração automática de determinadas variáveis para cada um dos objetos, como foi referido anteriormente (ver secção 2.2.4), extraíram-se um conjunto de variáveis de forma distinta da efetuada nas restantes abordagens. O *software* utilizado permite o cálculo automático de um grande número de variáveis, o que poderá ser uma mais-valia na etapa de distinção dos objetos correspondentes às copas. Assim, para além do cálculo das mesmas variáveis obtidas para os restantes algoritmos utilizados, extraiu-se adicionalmente outra variável de forma – *Roundness* – e ainda uma de textura – Homogeneidade GLCM (*Grey Level Co-occurrence Matrix*). A primeira descreve a semelhança entre a forma de um objecto e uma elipse. Quanto maior a semelhança menor o valor de *Roundness* (Trimble, 2012).

A GLCM caracteriza a textura de uma imagem calculando o quão frequentemente pares de píxeis com valores específicos, e numa relação espacial específica, ocorrem numa imagem. A construção destas matrizes permite extrair medidas estatísticas que representam características texturais da imagem, como é o caso da homogeneidade. A homogeneidade refere-se à distribuição dos elementos mais próxima da diagonal da GLCM. Esta informação pode ser computada para diversas distâncias e orientações (Trimble, 2012).

Seguidamente será explicada em detalhe a implementação dos diversos algoritmos de construção dos objetos utilizados.

3.3.1.1. *Watershed*

Este método foi operacionalizado com recurso ao *software* SAGA GIS através da ferramenta *Watershed Segmentation* presente na biblioteca ViGra (*Vision with generic algorithms*). A biblioteca apresenta um conjunto de ferramentas específicas para o processamento e análise de imagens, focando-se principalmente em algoritmos personalizados e estruturas de dados (SAGA GIS, 2018)

A implementação do algoritmo através desta ferramenta pressupõe a calibração de apenas um parâmetro – a largura da banda (*Band width*). Este parâmetro define o raio de procura,

em pixels, dos gradientes máximos de intensidade que vão delimitar as bacias, delineando as respectivas linhas divisórias, como já foi referido (ver secção 2.2.1).

Foram realizadas diversas tentativas de segmentação fazendo uso das várias bandas disponíveis, e com diferentes valores de largura de banda, por forma a otimizar os resultados da segmentação. Muitas vezes este algoritmo vê-se aplicado diretamente numa imagem na escala dos cinzentos (Jing *et al.*, 2012). No entanto, esta opção pode implicar perda de informação para a ICA, uma vez que o valor de cada pixel seria uma única amostra de um espaço de cores que variam entre preto (com menor intensidade) e branco (com maior intensidade). Assim optou-se por utilizar como dados de entrada para o algoritmo, informação que esteja associada à melhor distinção espectral da vegetação. Para além do NDVI, a banda do IVP e a banda do Vermelho são também bastante relevantes na distinção da vegetação, como já foi mencionado (ver secção 3.2.2). Testou-se igualmente a utilização das restantes bandas espectrais isoladamente.

Após uma procura exaustiva, inserida num amplo processo de tentativa-erro, seleccionaram-se as três tentativas correspondentes aos melhores resultados visuais, para prosseguir para as etapas metodológicas seguintes (Tabela 3.3).

Tabela 3.3. Tentativas seleccionadas para o algoritmo *Watershed*.

Tentativas	Dados	Parâmetros de Calibração
		Janela de Procura
1	Vermelho	6
2	Infravermelho Próximo	
3	NDVI	

3.3.1.2. Crescimento de Regiões

A implementação deste algoritmo foi feita com recurso à ferramenta do *software* SAGA GIS denominada *Object based image segmentation*. Esta permite a aplicação de um algoritmo de crescimento de regiões, gerando *sementes* através da definição de uma janela de procura por parte do utilizador. Uma vez que o resultado deste algoritmo de segmentação é altamente dependente, não só de uma medida de semelhança adequada mas também da qualidade das *sementes*, a geração destas é uma parte importante do procedimento de segmentação (Bechtel *et al.*, 2008).

Assim, de forma a definir a melhor janela de procura para as *sementes* recorreu-se à ferramenta de geração de seментes do *software* SAGA GIS para obter os pontos que melhor detetassem as copas das árvores, fazendo uso essencialmente dos dados relativos ao NDVI. A escolha do melhor valor para a janela de procura resultou de um processo de tentativa-erro, dado que não existe um valor ótímo especificado. Através da observação das diversas *sementes* produzidas pelas janelas de procura testadas, obtiveram-se os valores aparentemente mais adequados para esse parâmetro, de acordo com o objetivo em causa.

Após esta análise implementou-se finalmente o algoritmo de segmentação, calibrando a janela de procura com os valores obtidos anteriormente. Para além da definição da janela de procura foi também necessário calibrar o parâmetro de generalização. Este relaciona-se com os limites mais ou menos suavizados e complexos dos objetos a criar. Após diversas tentativas concluiu-se através da análise visual que o melhor parâmetro de generalização seria o correspondente ao valor 5.

À semelhança do método anterior, foram também utilizadas diversas combinações de valores de parâmetros e dados de entrada para o algoritmo. A Tabela 3.4 resume as melhores tentativas efetuadas, visualmente promissoras para prosseguirem para as seguintes etapas metodológicas.

Tabela 3.4. Tentativas seleccionadas para o algoritmo de Crescimento de Regiões.

Tentativas	Dados	Parâmetros de Calibração	
		Janela de Procura	Generalização
1	Vermelho; Verde;	10	5
2	Azul; NDVI	8	
3	Vermelho;	10	
4	Verde; Azul;	9	
5	IVP; NDVI	8	

3.3.1.3. Multiresolução

De acordo com o que foi exposto anteriormente (ver secção 2.2.4), este algoritmo constitui uma técnica de Crescimento de Regiões que se encontra especificamente

implementada no *software* eCognition®, e possui a valência de permitir a calibração dos diversos critérios de homogeneidade, que lhe são inerentes.

À semelhança dos casos anteriores, aplicou-se um processo de tentativa-erro, testando variada informação espectral de entrada e valores distintos para os critérios de homogeneidade (forma, cor, compacidade e suavidade). Para além destes critérios, o parâmetro da escala e a ponderação das bandas utilizadas foram também testados para diferentes combinações de valores.

A Tabela 3.5 resume as melhores tentativas registadas, à semelhança das abordagens anteriores.

Tabela 3.5. Tentativas selecionadas para o algoritmo de Multiresolução.

Tentativas	Dados	Parâmetros de Calibração						
		Escala	Cor	Forma	Suavidade	Compacidade	Peso das bandas da imagem	
1	IVP, Vermelho, Verde	50	0,8	0,2	0,5	0,5	3,1,1	
2		40						
3		60						
4	Vermelho, Verde, Azul	20					1,1,1	
5		30						
6		40						
7		50						
8	60							
9	Vermelho, Verde, Azul, IVP	40						1,1,1,1
10		50						
11	Vermelho, IVP, Verde, Azul, NDVI	40						

3.3.1.4. *Random Forests e MaxEnt*

Para a execução destes dois algoritmos, recorreu-se ao *software Rstudio*, implementando-se um *script* em linguagem R (Tabela 3.6).

Tabela 3.6. Código de implementação dos algoritmos de *Machine Learning*.

Linha	Código
1	<code>library(raster)</code>
2	<code>library(randomForest)</code>
3	<code>library(pROC)</code>
4	<code>library(dismo)</code>
5	<code>#</code>
6	<code>#Bandas</code>
7	<code>setwd("...") #Set working directory</code>
8	<code>#</code>
9	<code>b1 <- raster("RED_CLIP.tif")</code>
10	<code>b2 <- raster("GREEN_CLIP.tif")</code>
11	<code>b3 <- raster("BLUE_CLIP.tif")</code>
12	<code>b4 <- raster("NIR_CLIP.tif")</code>
13	<code>b5 <- raster("NDVI_etrS_Clip.tif")</code>
14	<code>b6 <- raster("ShapeIx2.tif")</code>
15	<code>b7 <- raster("FractDim.tif")</code>
16	<code>#</code>
17	<code>#Criação de um raster stack</code>
18	<code>bands_stack <- stack(b1, b2, b3, b4, b5, b6, b7)</code>
19	<code>#</code>
20	<code>#Pontos de Presença</code>
21	<code>points.presence <- shapefile("Pontos_arv_RF_etrS.shp")</code>
22	<code>#</code>
23	<code>#Função que seleciona aleatoriamente pontos num raster stack</code>
24	<code>sampleStack <- function(r, n) {</code>
25	<code> rowSample <- sample(1:r@nrows, size = n, replace = TRUE)</code>
26	<code> colSample <- sample(1:r@ncols, size = n, replace = TRUE)</code>
27	<code> pairs <- data.frame("rowInd" = rowSample, "colInd" = colSample)</code>
28	<code> out <- as.data.frame(cbind(pairs, as.data.frame(t(apply(pairs, MARGIN = 1, function(x)</code>
29	<code> {return(r[x[1],x[2]])}))))))</code>
30	<code> colnames(out)[3:ncol(out)] <- names(r)</code>
31	<code> return(out)</code>
32	<code>}</code>
33	<code>#</code>
34	<code>#Aplicação da função</code>
35	<code>random.sample <- sampleStack(r=bands_stack, n=1000)</code>
36	<code>random.sample <- random.sample[,3:9] #os numeros das colunas podem variar</code>
37	<code># random.sample <- random.sample[,3:7].</code>
38	<code>#</code>
39	<code>#Vetor de 0's</code>
40	<code>zero.vector <- rep(0, NROW(random.sample))</code>
41	<code>#</code>
42	<code>#Extração dos valores das bandas para cada ponto de presença</code>
43	<code>presence.sample <- extract(bands_stack, points.presence, method='simple')</code>
44	<code>#</code>

```

44 #Vector de 1's
45 ones.vector <- rep(1, NROW(presence.sample))
46 #
47 #Preparação para validação cruzada
48 #assign folds
49 nb.folds = 3
50 #
51 base.data <- rbind(random.sample,presence.sample)#dados para os pontos de copas e pontos
aleatorios
52 #
53 #Números aleatórios para ausências
54 fold <- sample(1:nb.folds, NROW(base.data), replace = TRUE) #atribui a cada ponto um numero
de 1 a 3
55 base.data <- cbind(base.data, fold)
56 #
57 #Criação da variável dependente: presença/ausencia de copa
58 dependent <- c(zero.vector,ones.vector)
59 base.data <- cbind(dependent, base.data)
60 #Vector para AUC
61 AUC.hold.rf <- c()
62 AUC.hold.max <- c()
63 #
64 #Aplicação do processo de validação cruzada
65 for (i in 1:nb.folds) {
66 #Separar dados para calibração e validação
67 #Dados Calibração
68 base.data.calibration <- subset(base.data, fold != i)
69 #base.data.calibration <- base.data.calibration[,1:6]
70 base.data.calibration <- base.data.calibration[,1:8]# numero das colunas podem variar consoante
o numero de bandas a usar
71 rownames(base.data.calibration) <- NULL
72 #Dados Validação
73 base.data.validation <- subset(base.data, fold == i)
74 rownames(base.data.validation) <- NULL
75 #
76 #Inicio da calibração
77 #Treino do modelo RandomForest e MaxEnt
78 dependent <- base.data.calibration[,1]
79 predictors <- base.data.calibration[,-1]
80 #
81 #Modelo RF
82 RF.model <- randomForest(x=predictors, y=as.factor(dependent), ntree=1000)
83 predict.RF <- predict(RF.model, base.data.validation, type="prob")
84 AUC.hold.rf <- rbind(AUC.hold.rf,auc(base.data.validation$dependent,predict.RF[,2]))
85 #
86 #Modelo Maxent
87 MaxEnt.model <- maxent(predictors, as.factor(dependent))
88 predict.max <- predict(MaxEnt.model, base.data.validation)

```

```

89     AUC.hold.max <- rbind(AUC.hold.max, auc(base.data.validation$dependent, predict.max))
90   }
91   #
92   #Modelo Completo RF
93   #base.data.full.model <- base.data[,1:6] #numeros das colunas podem variar consoante o numero
de bandas a analisar
94   base.data.full.model <- base.data[,1:8]
95   dependent <- base.data.full.model[,1]
96   predictors <- base.data.full.model[,-1]
97   RF.model <- randomForest(x=predictors, y=as.factor(dependent), ntree=1000)
98   predict.RF <- predict(bands_stack, RF.model, type="prob", na.action=na.pass)
99   #
100  #Modelo Completo MaxEnt
101  #base.data.full.model <- base.data[,1:6]#numeros das colunas podem variar consoante o numero
de bandas a analisar
102  base.data.full.model <- base.data[,1:8]
103  dependent <- base.data.full.model[,1]
104  predictors <- base.data.full.model[,-1]
105  MaxEnt.model <- maxent(predictors, as.factor(dependent))
106  predict.max <- predict(bands_stack, MaxEnt.model, na.action=na.pass)
107  #
108  #Exportação dos resultados
109  writeRaster((1-predict.RF), filename='prob_map_RF_var.asc', overwrite=TRUE)
110  writeRaster(predict.max, filename='prob_map_MAX_var.asc', overwrite=TRUE)
111  writeRaster(mean((1-predict.RF), predict.max), filename='prob_map_mean_var.asc',
overwrite=TRUE)
112  #
113  #Obtenção da AAC
114  mean(AUC.hold.rf)
115  mean(AUC.hold.max)

```

Estes algoritmos, utilizados como classificadores binários, permitem prever os pixels que correspondem às copas das árvores, com base num conjunto de variáveis explicativas, em pontos de presença que representam copas de árvore e em pontos que representam pseudo-ausências, utilizados para treinar o modelo. Como tal o primeiro passo consistiu na obtenção dos pontos de presença de copas de árvores.

Com recurso ao *software* ArcMap 10.5 e através de interpretação visual das imagens, obteve-se uma amostra de 50 pontos associados à existência de copas de árvores (Anexo 1). Para a amostra de pseudo-ausências foram selecionados aleatoriamente 1000 pixels. Esta amostra não pode ser considerada de verdadeiras ausências, uma vez que foi gerada aleatoriamente para toda a imagem. Obteve-se assim um conjunto de pontos representando presenças e pseudo-ausências de copas de árvores.

Para cada um dos pontos de presença e pseudo-ausência de copas de árvores, extraíram-se os valores de cada variável explicativa utilizada, obtendo-se finalmente os dados para a aplicação dos algoritmos. Da totalidade destes dados, uma parte é utilizada como dados de treino do modelo e a restante para testá-lo. Desta forma, o modelo prevê a presença de copas dos dados de teste, com base nos dados de treino.

O recurso a estes algoritmos permite ainda avaliar a qualidade das previsões obtidas. Esta avaliação pode ser conseguida através da validação cruzada. O procedimento de validação cruzada permite que a totalidade dos dados seja utilizada como treino do modelo e igualmente como teste (Elith *et al.*, 2008). Assim, dividiu-se aleatoriamente a amostra de dados em três conjuntos e aplicou-se o algoritmo a cada um deles. Isto é, um dos conjuntos foi utilizado para treinar o modelo (*training set*) e os restantes dois foram utilizados para testar o modelo treinado (*test set*).

Este processo foi efetuado para cada um dos três conjuntos criados, ou seja, para que deste modo todos eles sejam utilizados como treino e como teste. Seguidamente, para determinar a precisão das previsões obtidas, utilizou-se como medida a Área Abaixo da Curva (AAC) ROC (*Receiver operating characteristic*). A curva ROC corresponde a uma representação gráfica bidimensional entre duas métricas estatísticas que medem a performance de um teste binário à sensibilidade e especificidade. Quanto à AAC, esta corresponde à porção de área da unidade quadrada sob a curva, pelo que o seu valor estará sempre compreendido entre 0 e 1 e a sua interpretação permite o diagnóstico da capacidade de discriminação de um modelo (Santos, 2018). Os seus valores podem ser classificados de acordo com a Tabela 3.7.

Tabela 3.7. Categorização da AAC (adaptado de Hosmer & Lemeshow, 2000).

Valor de AAC	Qualidade da Previsão
0,91 – 1,0	Excelente
0,81 - 0,9	Bom
0,71 – 0,8	Razoável
0,5 – 0,7	Fraco

À semelhança das abordagens já executadas, foi aplicado um processo de tentativa-erro fazendo uso de diferentes variáveis explicativas. De igual forma, as mesmas referem-se a variáveis espectrais, mas também a variáveis de forma. A inserção de variáveis na

implementação destes algoritmos requer um formato matricial. As variáveis espectrais podem ser inseridas diretamente através das respectivas bandas da imagem. No entanto, para as variáveis de natureza morfológica é necessário executar uma transformação que permita a sua utilização. Como tal, foi selecionado o melhor resultado obtido nas abordagens de efetuadas anteriormente (segmentação obtida com recurso ao algoritmo de Multiresolução) e criada uma camada matricial com as variáveis morfológicas já calculadas para cada objeto – IF e DF.

Cabe sublinhar, que apesar de serem considerados algoritmos que procedem à construção de objetos na imagem, para este estudo, não o são em rigor, dado o seu resultado probabilístico e conseqüentemente matricial. A transformação para verdadeiros objetos (vetoriais) acontece em simultâneo com a etapa seguinte, como será detalhado em pormenor adiante. As tentativas que produziram os resultados visualmente mais promissores foram selecionadas para utilização nas etapas metodológicas posteriores (Tabela 3.8).

Tabela 3.8. Tentativas selecionadas para os algoritmos de *Machine Learning*.

Tentativas	Algoritmo	Nº presenças	Nº Pseudo-ausências	Dados	AAC
1	<i>Random Forests</i>	50	1000	Vermelho, Verde, Azul, IVP e NDVI.	0,79
2				Vermelho, Verde, Azul, IVP, NDVI, IF, DF.	0,82
1	<i>MaxEnt</i>			Vermelho, Verde, Azul, IVP e NDVI.	0,86
2				Vermelho, Verde, Azul, IVP, NDVI, IF, DF.	0,87

Acontece frequentemente que a combinação de resultados de diferentes algoritmos produz um melhor resultado do que os demonstrados por cada um deles isoladamente (Dzeroski e Zenko, 2002). Razão pela qual se optou por uma tentativa dessa natureza, com o objetivo de alcançar um resultado com mais qualidade do que o produzido por cada um dos modelos. Para tal executaram-se também tentativas para a combinação dos resultados dos dois algoritmos realizados. A combinação obteve-se através da computação da média dos resultados produzidos pelos dois algoritmos.

3.3.2. *Classificação por limiares*

Estando os objetos construídos e as respectivas variáveis obtidas, procedeu-se à segunda etapa metodológica do estudo: a individualização dos objetos que correspondem efetivamente às copas das árvores, objetivo final, para cada um dos resultados.

Para tal, realizou-se uma classificação por limiares, isto é, procedeu-se à seleção visual dos limiares das diversas variáveis calculadas, para os quais os objetos correspondentes a copas possam ser individualizados. A obtenção dos intervalos de valores para as variáveis e a própria escolha das mesmas, para cada abordagem, foi operacionalizada através de uma procura intensiva e de um processo de tentativa-erro. No caso das abordagens que fazem uso de algoritmos de *machine learning* na etapa de construção de objetos, em consequência do seu resultado probabilístico, a variável utilizada foi efetivamente a probabilidade de existência de uma copa de árvore.

Tendo esta abordagem uma base distinta das restantes, com um resultado matricial que representa a probabilidade de determinado píxel pertencer a uma copa de árvore, a aplicação de um limiar de probabilidade faz todo o sentido. A aplicação deste limiar permitiu a obtenção de uma camada matricial binária, onde valores unitários representam áreas de copas e valores nulos as restantes. Tendo-se detetado a existência de diversos pixéis isolados com probabilidade elevada de ser copa de árvore, mas que claramente não representavam áreas de copas, foi aplicado um filtro para minimizar esses pixéis isolados. A aplicação de um limiar à área dos objetos criados foi também ação complementar a esta operação, para eliminar pequenas zonas ainda existentes que não representavam de maneira nenhuma a copas de árvores. A imagem binária foi então convertida numa camada vetorial de polígonos para alcançar os objetos e consequentemente a cartografia final.

Para as restantes abordagens, que produzem resultados diretamente em formato vetorial, aos objetos obtidos após a classificação por limiares foi atribuído o valor unitário, sendo estes os que representam copas de árvores. Aos restantes foi atribuído o valor nulo, sendo que estes correspondem a todos os demais elementos existentes na superfície terrestre.

Na Tabela 3.9 podem observar-se as variáveis e os respetivos intervalos de valores aplicados para a classificação por limiares, nas tentativas de cada abordagem. Note-se que para todas as tentativas realizadas em cada uma das seis, as copas obtidas após a classificação por limiares tiveram como base a aplicação dos mesmos valores de limiares às variáveis, sendo que estes só se alteraram entre abordagens.

Após a classificação por limiares, que permite individualizar os objetos relativos a copas de árvores, dá-se por concluído o processo de ICA, obtendo-se a cartografia final das Copas de Árvores.

Tabela 3.9. Variáveis e limiares aplicados para a identificação dos objetos representativos de copas de árvores, por abordagem.

Algoritmo de construção de objetos	Variáveis	Intervalos de valores
<i>Watershed</i>	IVP	145 - 168
	Área	> 10 m ²
	Vermelho	54 – 69,1
	IF	< 2
	DF	< 5
Crescimento de Regiões	Vermelho	55 - 67
	Verde	67 - 82
	IVP	145 - 170
	IF	1,2 – 2,3
	DF	< 2
	Brilho Médio	60 - 80
	Área	>10 m ²
Multiresolução	Vermelho	53 - 69
	Verde	67 - 95
	IF	< 3,5
	<i>Roundness</i>	< 2
	IVP	146 – 166,2
	Homogeneidade GLMC (Todas as direções)	0,28 – 0,4
<i>Random Forests</i>	Probabilidade	>0,1
	Área	>10 m ²
MaxEnt	Probabilidade	>0,5
	Área	>10 m ²
Combinação (RF + MaxEnt)	Probabilidade	>0,3
	Área	>10 m ²

3.3.3. Avaliação da exatidão

A avaliação da exatidão constituiu a última etapa do procedimento metodológico proposto. Um mapa ou qualquer outro modelo só se torna válido após o teste e avaliação da sua aptidão para representar a informação pretendida, dado que os resultados de

qualquer tipo de análise nem sempre traduzem a realidade com precisão. A avaliação da exatidão de qualquer mapa permite obter informação descritiva importante para avaliar a utilidade do mesmo para determinada aplicação (Stehman, 2001). Com a ICA não é diferente. Logo, é essencial avaliar a qualidade dos mapas obtidos, para dessa forma ser possível retirar conclusões credíveis e consistentes.

A avaliação da exatidão é um processo que consiste na confrontação entre o mapa a avaliar e um conjunto de dados de referência, que se assume representarem a realidade de forma fidedigna. Este confronto é expresso através de um conjunto de medidas de avaliação, apresentadas através da denominada matriz de confusão, e que dão informação sobre a probabilidade de ocorrerem atribuições de classes corretas (Costa, 2008). Nesta matriz são expressos os valores das medidas de avaliação específicas – exatidão do produtor (medida do erro de omissão), exatidão do utilizador (medida do erro de comissão) e o coeficiente de concordância *Kappa* – que permitem tirar conclusões sobre a qualidade e confiabilidade de um resultado (Barbosa, 2016).

A Exatidão do Produtor (EP) representa a probabilidade de um objeto de referência de uma classe estar realmente representado na cartografia por essa mesma classe, sendo o erro de omissão (EO) a probabilidade de um objeto estar mal classificado. A Exatidão do Utilizador (EU) traduz a probabilidade de um objeto classificado com determinada classe pertencer verdadeiramente a essa classe, sendo o erro de comissão (EC) a probabilidade do objeto classificado representar na realidade outra classe.

Proposto por Jacob Cohen em 1960, o coeficiente de concordância *Kappa* é uma medida de concordância usada em escalas nominais, que fornece uma ideia do quanto, casuisticamente, as observações registadas se destacam das esperadas, indicando-nos assim o quão legítimas as interpretações são. Trata-se de um método estatístico para avaliar o nível de concordância ou reprodutibilidade entre dois conjuntos de dados. A utilização deste coeficiente permite avaliar satisfatoriamente toda a imagem classificada, porque considera nos seus cálculos toda a matriz de confusão. Os valores deste coeficiente variam entre -1 e +1, sendo mais forte a concordância entre os conjuntos de dados quanto mais elevado for o seu valor. Os valores podem ser classificados de acordo com a Tabela 3.10.

Tabela 3.10. Categorização dos valores do coeficiente de concordância *Kappa* (adaptado de Landis & Koch, 1977).

Coeficiente de Concordância <i>Kappa</i>	Concordância da Classificação
0,81 – 1,0	Excelente
0,61 - 0,8	Muito Bom
0,41 – 0,6	Boa
0,21 – 0,4	Razoável
0,0 – 0,2	Má
<0,00	Péssima

O processo de validação consistiu na construção de uma amostra de copas de árvores da área em estudo (copas de referência), delineadas manualmente através da interpretação visual da imagem, com a qual cada mapa de Copas de Árvores foi comparado. Foram delineadas um total de 40 copas de árvores com recurso ao *software* ArcMap 10.5 (Anexo 2) e foi amostrado, pela totalidade da área de estudo, um conjunto de observações (pontos) para a amostra de validação. É através desta amostra que será realizada a comparação entre as copas de referência delineadas e as copas obtidas em cada resultado. Ambos foram utilizados para a obtenção de uma matriz de confusão.

No entanto, a matriz de confusão não tem em conta a área ocupada pelas classes avaliadas nem a forma dos objetos. Para obter uma melhor perspectiva da qualidade das copas individualizadas, a forma e a área das mesmas tem importância, uma vez que uma copa esta delineada corretamente se tiver a forma e área equivalente à da copa de referência. Como tal, considerou-se relevante o cálculo de medidas de avaliação complementares às inicialmente descritas, que permitissem a análise tendo em consideração este aspeto, obviamente sempre em comparação com as copas de referência. Estas medidas permitiram avaliar a qualidade da cartografia produzida, pelo facto de transmitirem informação acerca da área corretamente e/ou incorretamente considerada como copa de árvore. Foram computadas em termos percentuais e agrupadas da seguinte forma (Figura 3.3):

- Medidas obtidas tendo por base a área das copas individualizadas pela respetiva abordagem (percentagem de área da copa individualizada que foi identificada como copa, mas não corresponde efetivamente a uma e percentagem da área da copa individualizada que corresponde);
- Medidas obtidas tendo por base a área das copas de referência (percentagem da área da copa de referência que não foi identificada como copa de árvore e percentagem da área da copa de referência que efetivamente o foi).

Através do cálculo destas medidas é possível obter informação sobre a totalidade de área de copas que foi correta ou incorretamente delineada, seja pelo facto de ter sido delineada, mas não corresponder a uma copa, seja pelo facto de não ter sido delineada, mas corresponder realmente a uma.

O procedimento para a realização desta avaliação e da obtenção de todas as métricas de avaliação de exatidão referidas foi executado com recurso ao *software* ArcMap 10.5 e encontra-se explicado em detalhe nas secções seguintes.

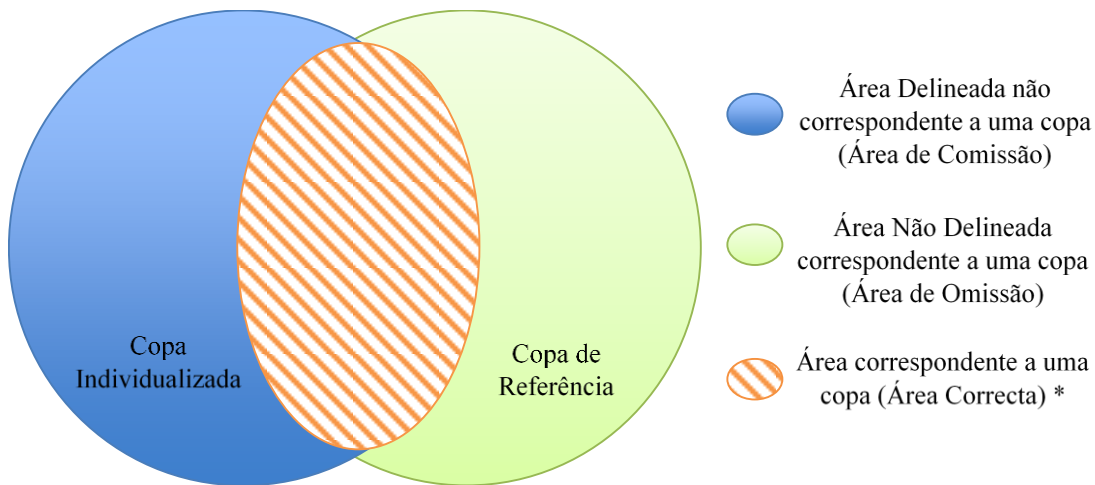


Figura 3.3. Esquema explicativo das medidas complementares de avaliação da exatidão (* Relativamente à copa individualizada ou à copa de referência)

3.3.3.1. Amostragem

O cuidado na escolha dos pontos e a quantidade de dados amostrados estão diretamente relacionados com a qualidade do produto final de uma aplicação sobre o modelo. A dimensão da amostra deve ser suficiente para uma avaliação significativa e coerente. A quantidade de pontos amostrados, bem como o cuidado na sua escolha, é decisiva. De uma forma geral, quanto maior é a amostra, maior é a confiança dos resultados da avaliação da exatidão (Costa, 2008). Contudo, quanto maior a quantidade de pontos representativos da superfície real, maior será o esforço computacional para que estes sejam armazenados, recuperados, processados, até que se alcance o produto final da

aplicação (Barbosa, 2016). Assim, de forma a encontrar um compromisso adequado entre o esforço computacional e a dimensão significativa da amostra de validação, considerou-se a criação de um total de 800 pontos para a amostra de validação (divididos equitativamente entre as copas de árvores de referência e a restante área não considerada copa) (Figura 3.4).

O processo de amostragem foi comum para todas os mapas de ICA obtidos, tendo sido considerada a mesma amostra de validação, em todos os resultados. A utilização dos pontos de amostra de validação com a mesma localização para os diferentes resultados de ICA obtidos foi considerado o mais indicado para avaliar a exatidão, no presente caso de estudo. Esta amostra foi obtida através de uma amostragem aleatória, de forma a evitar efeitos de autocorrelação espacial.

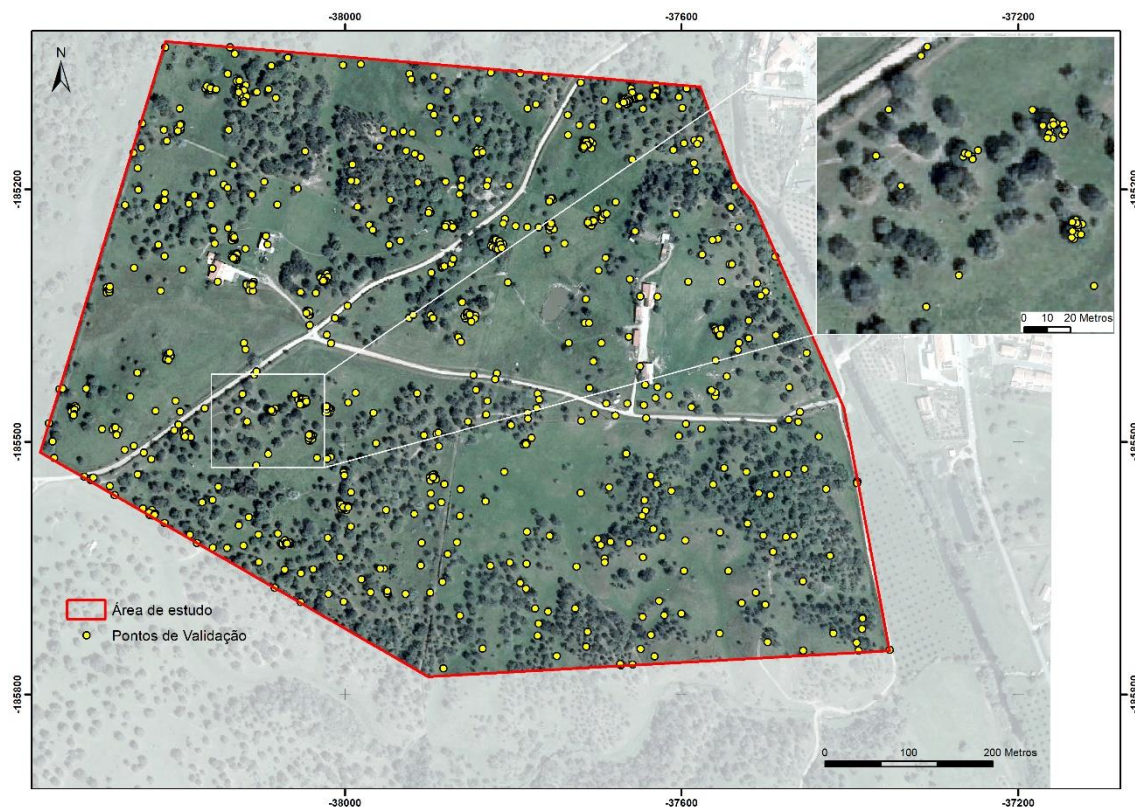


Figura 3.4. Pontos utilizados na amostra de validação.

3.3.3.2. Recolha de dados para a amostra de validação

Após a seleção do conjunto de pontos da amostra de validação, procedeu-se ao seu confronto com as copas de referência e seguidamente com cada um dos resultados

obtidos. Para efetuar este processo de confrontação, é fundamental a recolha de dois tipos de informação para cada ponto da amostra de validação: (1) os dados de referência, que são o rótulo da classe, originário dos dados considerados corretos (as copas de referência) e (2) os dados a avaliar, que são o rótulo da classe, originado a partir do mapa a ser avaliado.

Assim, foi feita uma interpretação ao conjunto de pontos amostrados e verificado quais os pontos da amostra de validação que se intersectam com as copas de referência. A esses pontos foi atribuído o valor unitário, correspondente a uma copa de árvore, sendo que aos restantes foi atribuído o valor nulo, correspondente a elementos que não são copas. Desta forma estão criados os dados de referência. Quanto aos dados a avaliar, para cada um dos resultados, classificaram-se com valor 1 os pontos da amostra de validação que correspondiam a copas individualizadas, que se intersectavam com as copas de referência, e com valor 0 os restantes. Obtiveram-se assim, para cada mapa, os dados a avaliar. Os dados de referência mantêm-se sempre os mesmos, tendo em conta que são obtidos considerando as copas de referência, sendo que os dados a avaliar diferem consoante o mapa a avaliar uma vez que são obtidos a partir do resultado que se quer avaliar.

Após efetuadas estas operações foi assim possível a obtenção da matriz de confusão para cada resultado de ICA e as respetivas medidas de avaliação da qualidade global e específica das classes (C1 – copas – e C0 – ausência de copas) da cartografia produzida.

Esta avaliação da exatidão e o cálculo das respetivas medidas de avaliação foi efetuada para as tentativas identificadas anteriormente para cada abordagem utilizada (Anexo 3), tendo sido posteriormente selecionada a tentativa com melhores resultados em cada uma das abordagens. Obtiveram-se assim seis mapas finais da ICA, discutidos no capítulo seguinte.

4. Apresentação e discussão dos resultados

Dentro das diversas tentativas em cada abordagem, foram finalmente identificadas as que obtiveram melhores resultados, tendo em conta o conjunto das medidas de avaliação de exatidão e a qualidade visual da cartografia produzida (Tabela 4.1). Desta forma foi analisado o melhor resultado obtido em cada uma das abordagens executadas, totalizando seis mapas de Copas de Árvores.

Tabela 4.1. Tentativas finais selecionadas para cada abordagem.

Abordagem (de acordo com o algoritmo de construção de objetos)	Tentativa
<i>Watershed</i>	2
Crescimento de Regiões	3
Multiresolução	7
RF	2
MaxEnt	1
Combinação (RF e MaxEnt)	2

Neste capítulo são apresentados os produtos finais obtidos para as diferentes abordagens executadas para a ICA na secção anterior, e discutidas as suas representações e as respetivas medidas de validação calculadas.

4.1. Cartografia de Copas de Árvores

A partir dos mapas produzidos (da Figura 4.1 à **Erro! A origem da referência não foi encontrada.**) é possível retirar visualmente algumas conclusões relevantes, mesmo antes de proceder à análise das medidas de validação.

Nem todas as copas são detetadas e individualizadas, verificando-se um maior número de copas não individualizadas nas abordagens que envolveram o algoritmo de Crescimento de Regiões e o *Watershed*. A comparação visual do resultado destas duas com as restantes denota claramente um menor número de copas identificadas, e mais especificamente no caso da abordagem de Crescimento de Regiões. Esta individualiza um número de copas muito inferior ao das restantes.

Contrariamente, aquelas que se baseiam em algoritmos de *machine learning* realizam a individualização de grande parte das copas, sendo as que mais árvores detetam e

individualizam. No entanto, é de notar que muitas vezes várias copas encontram-se individualizadas num mesmo polígono. Isto é, apesar de a maioria das copas ser detetada e de, aparentemente, parte dos seus limites estar consideravelmente bem delineado, acontece que muitas se encontram juntas no mesmo polígono.

Este é um padrão comum em todos os métodos utilizados: a distinção entre copas contíguas é um processo que não é totalmente bem conseguido. Este padrão afigura-se, contudo, mais ou menos evidente, consoante a abordagem utilizada. Era de esperar que existisse esta dificuldade, como foi referido anteriormente como sendo uma das mais apontadas no processo de ICA (ver Tabela 2.1).

No caso específico do sistema florestal em análise, onde ocorre naturalmente a existência de árvores contíguas e não existe uma organização específica com um compasso definido entre árvores (tal como se encontra em plantações ou culturas), era de esperar que esta dificuldade fosse mais notória. No entanto, apesar de ser um padrão geral que acontece um pouco em toda a cartografia resultante, algumas abordagens lidam melhor com este aspeto que outras.

Observamos que, na abordagem que utiliza o algoritmo MaxEnt (Figura 4.1) este padrão é muito consistente, sendo que existe um grande número de copas que são detetadas e individualizadas, mas que se encontram ligadas umas às outras.

Na abordagem de RF este padrão está presente, mas de uma forma menos acentuada (Figura 4.2). A produção de cartografia da combinação das duas abordagens de *machine learning* permitiu minimizar esta característica, verificando-se um menor número de polígonos que representam diversas copas (Figura 4.). Nas restantes abordagens este padrão acontece com menor expressão. Por outro lado, a abordagem que utiliza o algoritmo de Crescimento de Regiões aparenta produzir melhores resultados, dentro deste contexto. No entanto é importante sublinhar que esta individualiza um número muito menor de copas do que as que recorrem a outros algoritmos de construção de objetos, o que se deve ter em consideração na conclusão relativa à expressividade deste padrão.

Nota-se também alguma dificuldade em separar as zonas de sombra das respetivas copas. É nas abordagens que utilizam algoritmos de *machine learning* e também um pouco na

que utiliza o algoritmo de Multiresolução (Figura 4.4), que se nota mais vezes a existência de zonas de sombra que estão presentes nas áreas de copas individualizadas, como uma espécie de extensão das mesmas. Embora isto se verifique, neste estudo encontra-se relativamente minimizada visto ter sido utilizada uma imagem correspondente ao período do final da manhã do mês de abril, onde a iluminação relativa produz menores extensões de sombra.

Existe ainda dificuldade na distinção de copas relativamente ao solo. Esta é devida essencialmente à semelhança espectral entre os dois elementos da superfície terrestre, que muitas vezes dificulta a distinção dos objetos na imagem. Na abordagem Multiresolução esta dificuldade é mais evidente, sendo, no entanto, uma dificuldade generalizada. Esta confusão espectral está também presente no caso da abordagem que utiliza o algoritmo RF, neste caso entre copa e água, observando-se que a segunda é individualizada como copa. Mais uma vez, a cartografia produzida pela combinação das duas abordagens de *machine learning* permitiu inocular esta confusão espectral e parece adequar-se a outras potencialmente existentes. Dado que o sistema florestal, que é o montado, regista uma reduzida densidade de árvores, ocorre uma maior presença de outros elementos da superfície terrestre, o que vem a gerar a confusão espectral constatada. Assim, poderá eventualmente ser um fator a considerar, as situações em que a densidade do povoamento possa ser mais elevada, contribuindo para a melhoria na cartografia produzida.

Outra característica, exclusiva das abordagens que utilizam algoritmos de segmentação de imagens, é o facto de determinadas copas (em geral as de maiores dimensões) se encontrarem repartidas em mais do que um objeto. Esta partição excessiva verifica-se especialmente na abordagem que utiliza o algoritmo de *Watershed* (Figura 4.). Estando este algoritmo frequentemente associado à segmentação excessiva da imagem (Ke & Quackenbush, 2011a), observando-se no resultado obtido aquando da implementação do algoritmo (Anexo 4), já era de esperar que fosse um erro que sobressaísse com maior evidência no seu resultado final. Na abordagem resultante do Crescimento de Regiões (Figura 4.) também é possível verificar esta característica, com menor impacto, e igualmente observada no resultado obtido após a sua implementação (Anexo 5). Na abordagem resultante da segmentação Multiresolução a sua presença é ainda menos significativa, existindo uma clara diferença nos tamanhos dos objetos criados aquando da

sua implementação (Anexo 6), resultando em objetos mais aproximados aos elementos reais da superfície terrestre representados através da imagem.

Constata-se um aspeto interessante na cartografia produzida por abordagens que utilizaram algoritmos de *machine learning*: as copas individualizadas possuem frequentemente lacunas no seu interior, contrariamente às obtidas por outras abordagens. Acontece que as espécies florestais constituintes do montado (sobreiros e azinheiras), são árvores que apresentam uma copa com uma forma irregular e, em muitos casos, existem ramos de diferentes tamanhos e orientações que fazem com que existam aberturas nas suas copas (Ke & Quackenbush, 2011a). Através das abordagens que utilizam algoritmos de *machine learning*, ao contrário das restantes, identificam-se essas mesmas aberturas e individualizam-se as copas de acordo com a existência das mesmas. Ora este resultado é bastante interessante e único dentro das abordagens testadas. Estas refletem assim uma mais-valia relativamente às abordagens que fazem uso de algoritmos de segmentação de imagens e que se demonstraram incapazes de detetar esta característica.

Em conclusão, através da análise visual é possível perceber que as abordagens que utilizam algoritmos de *machine learning* produzem uma cartografia mais adequada relativamente às que fazem uso de algoritmos de segmentação de imagens, ainda que, dentro dos últimos, a segmentação Multiresolução resulte também visualmente mais razoavelmente adequada.

Apesar dos diversos problemas inerentes ao processo da ICA, muitos deles comuns a qualquer das abordagens, estes algoritmos revelam-se promissores na execução da tarefa. Em detalhe, o algoritmo RF apresenta um resultado bastante aceitável, superior ao do algoritmo MaxEnt, no entanto a combinação dos dois algoritmos apresenta igualmente um resultado satisfatório uma vez que permite a eliminação de alguns erros produzidos por ambos, produzindo um resultado intermédio e que demonstra ser visualmente aceitável.

Seguidamente efetuaram-se os cálculos das medidas de avaliação da exatidão dos resultados de forma a ser possível retirar conclusões mais credíveis, não unicamente baseadas em análise visual.

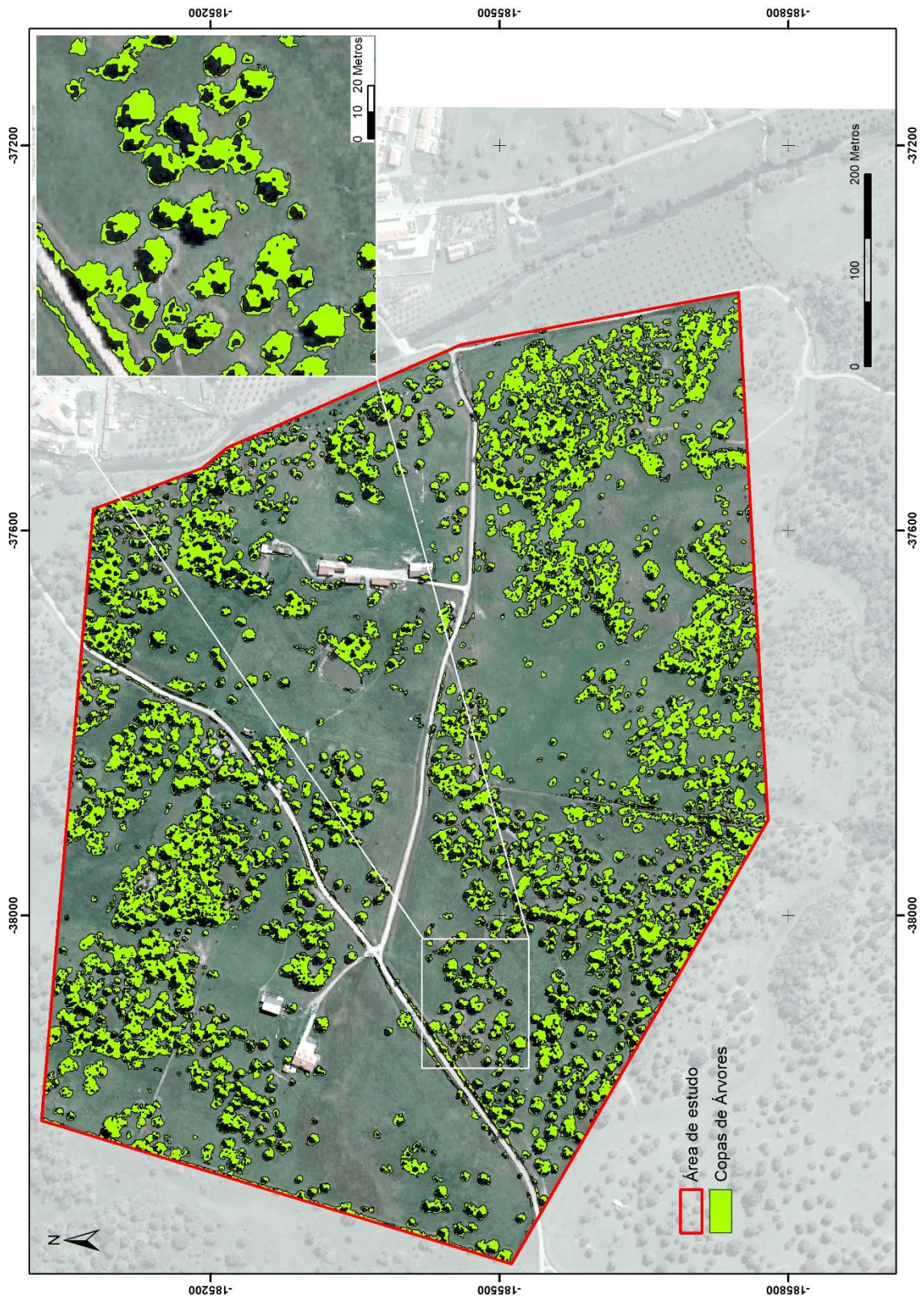


Figura 4.1. Copas de Árvores obtidas através da abordagem que utiliza o algoritmo MaxEnt.

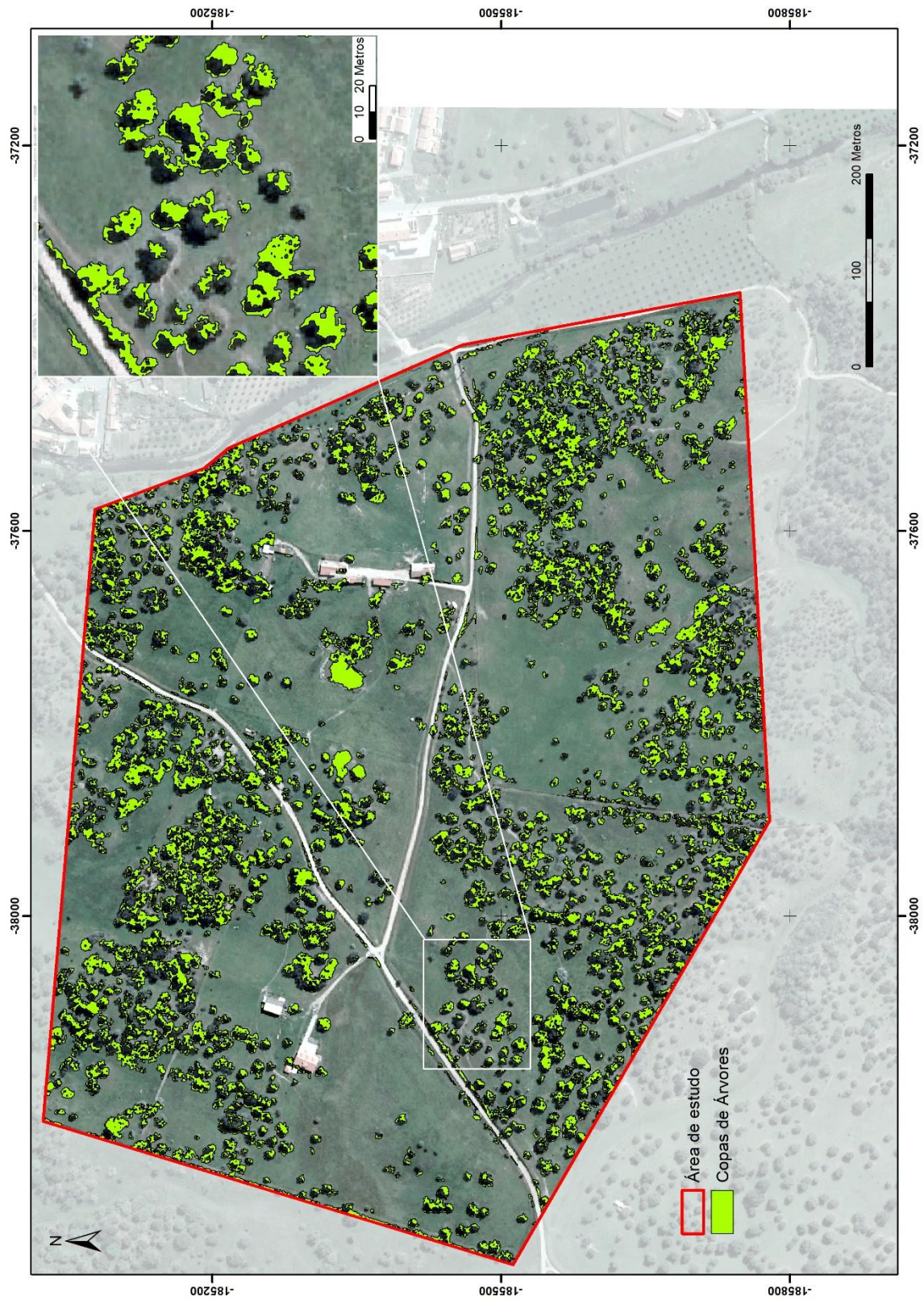


Figura 4.2. Copas de Árvores obtidas através da abordagem que utiliza o algoritmo *Random Forests*.



Figura 4.3. Copas de Árvores obtidas através da combinação das duas abordagens que utilizam algoritmos de *Machine Learning*.



Figura 4.4. Copas de Árvores obtidas através da abordagem que utiliza o algoritmo Multiresolução.



Figura 4.5. Copas de Árvores obtidas através da abordagem que utiliza o algoritmo *Watershed*.



Figura 4.6. Copas de Árvores obtidas através da abordagem que utiliza o algoritmo de Crescimento de Regiões.

4.2. Medidas de Validação

Para a avaliação da cartografia de Copas de Árvores realizada para as seis abordagens em estudo, calculou-se um conjunto de medidas que visam esse objetivo, permitindo assim tirar conclusões confiáveis sobre a qualidade da cartografia. São essas medidas: a exatidão do produtor e do utilizador, erro de Omissão e Comissão e o Coeficiente de Concordância *Kappa*. Estas medidas resultam das matrizes de confusão obtidas para cada abordagem (Anexo 7 a Anexo 12).

Esta avaliação nem sempre compreendeu as duas classes existentes (C1 – copas – e C0 – ausência de copas). Relativamente à EP, que representa a percentagem de pontos de referência de uma classe que estão realmente representados na cartografia por essa mesma classe, foi excluída *a priori* do processo de avaliação a classe C0, na medida em que praticamente não houve observações mal classificadas (EP superior a 98% para toda a cartografia). Isto demonstra que os pontos de referência correspondentes a áreas de ausência de copas estão corretamente classificados (com exceção de uma percentagem pouco significativa de 0,02%). Já no caso da EU, respeitante à percentagem de pontos classificados em determinada classe que pertencem verdadeiramente a essa classe, excluiu-se à partida a classe C1, pelo facto de não existirem observações mal classificadas (EU superior a 97% para toda a cartografia). Isto demonstra que os pontos que foram classificados como copas, em cada uma das abordagens, correspondem efetivamente a copas de árvores. Apenas 0.03% correspondem a áreas de ausência de copas. Não foi por isso necessário apresentar os valores destas classes na Tabela 4.2 uma vez que são valores desprezíveis, no sentido em que não comprometem a qualidade da cartografia.

Procedendo à análise da Tabela 4.2 é possível retirar algumas conclusões. Em primeiro lugar, num panorama mais geral, verifica-se que as abordagens que apresentam os valores de erro mais reduzidos, bem como os EC e EO, são as abordagens RF, MaxEnt e a combinação de ambos. Ou seja, as abordagens que utilizam técnicas de *machine learning* produzem erros menores comparativamente às outras, que utilizam algoritmos de segmentação de imagens. Dentro das técnicas de segmentação de imagens, a que utiliza o algoritmo de *Watershed* apresenta os valores de erro mais elevados e a que utiliza o algoritmo de Multiresolução apresenta os valores de erro mais reduzidos. No entanto,

qualquer abordagem que utilize algoritmos de segmentação apresenta um menor desempenho relativamente às com algoritmos *de machine learning*.

Tabela 4.2. Exatidão do produtor e utilizador, erros de omissão e comissão e coeficiente de concordância *Kappa* para as abordagens comparadas.

Abordagem (de acordo com o algoritmo de construção de objetos)	Exatidão do Produtor	Erro de Omissão	Exatidão do Utilizador	Erro de Comissão	Exatidão Global	Coeficiente de Concordância <i>Kappa</i>
	C1		C0			
<i>Watershed</i>	0,35	0,65	0,60	0,40	0,67	0,34
Crescimento de Regiões	0,31	0,69	0,59	0,41	0,65	0,30
Multiresolução	0,46	0,54	0,65	0,35	0,73	0,46
<i>Random Forests</i>	0,64	0,36	0,73	0,27	0,81	0,62
MaxEnt	0,55	0,45	0,69	0,31	0,77	0,55
Combinação	0,59	0,41	0,71	0,29	0,79	0,57

Analisando mais detalhadamente, a abordagem que utiliza o algoritmo RF apresenta o melhor valor de EP – 0,64 -, o que equivale a um EO de 0,36. Apenas 36% das observações pertencentes à classe das C1 foram classificadas de forma incorreta (como sendo C0). De igual forma esta abordagem obteve o melhor valor de EU – 0,73 -, o que se traduz num EC de 0,27. Apenas 27% das observações classificadas como não sendo C0 são na realidade C1. A abordagem que utiliza o algoritmo MaxEnt, embora não com valores tão elevados, apresenta igualmente valores satisfatórios (EP de 0,55 e EO de 0,45; EU de 0,69 e EC de 0,31). A combinação destes dois algoritmos produz um resultado ‘intermédio’ entre as duas abordagens, como seria de esperar por se tratar de uma média. Este é relevante uma vez que permite atenuar os defeitos produzidos por cada uma das abordagens individualmente e, apesar de não ser o que produz valores mais satisfatórios das medidas de avaliação (EP de 0,59 e EO de 0,41; EU de 0,71 e EC de 0,29), poderá ser o mais adequado no contexto.

Por outro lado, observa-se que a abordagem que faz uso do Crescimento de Regiões apresenta um valor de EP de 0,31, resultante num EO de 0,69, e um valor de EU de 0,59, resultando num EC de 0,41, desempenho este muito semelhante ao do algoritmo *Watershed* (EP de 0,35 e EO de 0,65; EU de 0,60 e EC de 0,40), embora um pouco

inferior. Ora nota-se que a diferença existente entre estas duas e a abordagem de RF é elevada, confirmando a suspeita já existente de que os algoritmos de *machine learning* executam com maior sucesso a tarefa da ICA. A abordagem de Multiresolução apresenta uma EP de 0,46 (EO de 0,54) e um EU de 0,65 (EC de 0,35) o que, apesar de não ser tão expressivo como os valores obtidos para a RF, é relativamente satisfatório no conjunto de abordagens com algoritmos de segmentação de imagens. Dentro deste conjunto, a utilização do algoritmo de Multiresolução produz resultados mais satisfatórios, com maiores valores de exatidão do que os restantes. A possibilidade de calibração dos critérios de homogeneidade proporcionada pelo algoritmo de segmentação permite uma melhor adequação dos objetos construídos, consoante o contexto pretendido. O facto de ser possível também o cálculo de um maior número de variáveis, que podem ser utilizadas na etapa de classificação por limiar, pode igualmente auxiliar a obtenção de melhores resultados. Estas características podem ser decisivas na obtenção de uma cartografia de Copas de Árvores de maior qualidade.

Apesar de para a exatidão global (EG) se verificarem valores considerados bons, e que indicam uma boa qualidade da cartografia produzida, é necessário ter atenção aquando da análise desta medida. Isto porque este valor elevado nem sempre traduz um mapa de qualidade. Trata-se de uma medida que tem em conta apenas os pontos bem classificados em cada classe em relação ao total de pontos amostrados. Pode ser influenciada por uma classe que esteja muito bem classificada e ‘esconder’ outras que não estejam tao bem definidas. Por outras palavras, a existência de uma classe com bons resultados pode tornar o valor de EG superior, apesar de poderem existir classes com um desempenho mais à quem, disfarçando assim a verdadeira qualidade do mapa. No entanto verifica-se a tendência já referida anteriormente, sendo a abordagem que utiliza RF a que possui um maior valor de EG (0,81) e a de Crescimento de Regiões a que possui um menor valor (0,30).

Em relação ao coeficiente de concordância *Kappa*, é de notar que, vindo de encontro aos valores das medidas já referidas, a abordagem de RF apresenta uma qualidade categorizada como ‘muito boa’ (0,62). As restantes abordagens de *machine learning* apresentam valores que categorizam a qualidade como sendo ‘boa’. Verifica-se que não existe nenhuma abordagem em que a qualidade da cartografia possa ser considerada ‘má’. No entanto, à semelhança do que é perceptível a partir dos valores das medidas

mencionadas, as abordagens *Watershed* e Crescimento de Regiões apresentam os valores de *Kappa* menos interessantes, correspondendo a uma qualidade ‘razoável’.

Pode afirmar-se que, em termos de análise da matriz de confusão, a cartografia de Copas de Árvores produzida pelas diferentes abordagens apresenta na sua generalidade valores aceitáveis. No que diz respeito às abordagens de *machine learning* é possível afirmar que apresentam já valores minimamente satisfatórios.

Ao longo deste estudo foi-se clarificando a ideia de que as abordagens que fazem uso dos algoritmos mais populares na ICA não foram efetivamente as que obtiveram os melhores resultados. Os resultados dos parâmetros analisados até aqui vêm a sustentar ainda mais esta ideia. Apesar de verificados resultados satisfatórios na utilização dos primeiros, esses resultados dizem respeito frequentemente à execução da tarefa para sistemas florestais distintos do analisado no presente trabalho (Ke e Quackenbush, 2008) (Ke e Quackenbush, 2011a). Para o sistema florestal do montado, estes algoritmos não demonstraram grande adequação. Sendo o montado um sistema florestal complexo, com especificidades próprias, uma organização distinta e uma composição de espécies florestais única, pressupõe também uma metodologia mais apurada de forma a executar a tarefa com eficácia. A obtenção de resultados mais satisfatórios com algoritmos de *machine learning* vem corroborar este pressuposto, confirmando-se assim a aplicabilidade destes algoritmos no âmbito da ICA.

É importante por outro lado analisar as medidas complementares calculadas, que avaliam a qualidade da cartografia de Copas produzida, mas tendo em consideração a área das copas individualizadas em comparação com as de referência.

Com o auxílio da Tabela 4.3, na área de Copas de Referência, é possível concluir que, em média, o valor máximo de percentagem de área correta foi de 55,49%, através da abordagem RF – o melhor resultado para esta medida. O segundo melhor resultado advém do algoritmo de Multiresolução - 53,75% -, não existindo uma diferença muito significativa entre ambos. Fica evidente aqui que, contrariamente ao verificado com a análise da matriz de confusão, o algoritmo de Multiresolução apresenta um valor de área correta significativo e mais elevado do que as outras abordagens de *machine learning* (MaxEnt e Combinação). Mesmo não tendo obtido resultados muito satisfatórios de

acordo com as medidas da matriz de confusão, demonstra-se aqui que as copas de referência possuem, em média, 53,75% da sua área individualizada por via da Multiresolução, o que é aceitável para esta medida. Adicionalmente, a percentagem de área de comissão é inferior à registada para a abordagem com RF. Em média, 51,25% da área de uma copa foi individualizada sem, no entanto, pertencer a uma copa na realidade, sendo que a restante área individualizada (48,75%) é verdadeiramente coincidente com uma copa de árvore. No caso da abordagem de RF, em média, 55,31% da área de uma copa individualizada não representa uma copa na realidade, sendo que apenas os restantes 44,69% representam, o que não é tão positivo. No entanto a área de omissão observada na abordagem Multiresolução é superior, resultando numa percentagem de área de copa de referência individualizada menor relativamente à abordagem de RF, que por sua vez deve ser encarado como positivo para esta última.

Tabela 4.3. Medidas complementares de Avaliação de Exatidão, tendo em conta a área (m²) das Copas individualizadas e de referência.

Medidas complementares de Avaliação de Qualidade				
Abordagem (de acordo com o algoritmo de construção de objetos)	Área de Copas Individualizada		Área de Copas de Referência	
	Área de Comissão (%)	Área Correta (%)	Área de Omissão (%)	Área Correta (%)
Watershed	41,02	57,62	85,97	13,26
Crescimento de Regiões	37,76	62,24	64,97	35,03
Multiresolução	51,25	48,75	46,25	53,75
Random Forests	55,31	44,69	44,51	55,49
MaxEnt	47,07	52,93	55,71	44,29
Combinação	48,61	51,39	50,21	49,79

Quanto às abordagens que registaram um pior desempenho, estas são coincidentes com o verificado nas medidas da matriz de confusão. Em média, apenas 13,26% da área de uma copa de referência foi individualizada como copa na abordagem que utiliza o *Watershed*, sendo que 85,97% da área de uma copa de referência não é individualizada por esta. Para a abordagem de Crescimento de Regiões estas medidas apresentam valores mais positivos, embora não deixem de ser pouco convincentes. Volta-se a verificar uma discrepância no sentido da interpretação dos resultados indicativos destas medidas, relativamente aos das medidas da matriz de confusão. A abordagem que produz uma cartografia de menor qualidade nas medidas complementares calculadas, é a que utiliza o algoritmo de *Watershed*, contrariamente ao verificado através da análise das medidas da matriz de confusão.

Verifica-se que os algoritmos que fazem uso de *machine learning* possuem valores de área de comissão mais elevados relativamente as restantes abordagens. Este facto é devido essencialmente à não separação de copas contíguas, o que produz objetos que representam não apenas uma copa mas diversas copas. Isto faz com que, aquando do cálculo das medidas, a área que seja identificada como fazendo parte da uma determinada copa faça de facto parte de outras, mas que se encontra no mesmo objeto. Este aspeto verifica-se frequentemente na cartografia produzida por estas abordagens sendo um dos principais pontos negativos.

Em suma, apesar das pequenas diferenças observadas, é possível distinguir uma tendência geral: abordagens com algoritmos de *machine learning* produzem uma cartografia de Copas de Árvores de maior qualidade, relativamente a abordagens de segmentação de imagens. Apesar dos problemas encontrados na cartografia produzida pelas primeiras, o seu desempenho não deixa de ser superior às segundas, individualizando um maior número de copas e produzindo copas com uma área e forma que se assemelha mais com a real.

Dentro dos métodos de segmentação, a abordagem de Multiresolução produz um resultado também razoável, dentro deste conjunto de abordagens. Apesar de não serem detetadas tantas copas e de os valores das medidas de avaliação não serem tão satisfatórios, não deixam ainda assim de ser razoáveis. No entanto esta abordagem não permite reprodutibilidade metodológica independente, sendo uma desvantagem considerável relativamente às de *machine learning*.

A abordagem que faz uso do algoritmo de *Watershed* apresenta um desempenho francamente reduzido. Contrariamente a todas as outras, a sua implementação apenas permite introduzir informação de uma única banda, não permitindo combinar a informação multiespectral, de mais do que uma banda espectral. Este fator poderá vir ao encontro dos resultados menos eficazes desta abordagem quando comparadas com outras, que tiveram em conta na sua implementação uma maior quantidade de informação.

Deve-se sublinhar que a avaliação foi realizada através de uma amostra de copas de referência e não de todas as copas existentes na imagem, i.e., podem existir copas mal classificadas e individualizadas, mas que não foram incluídas na amostra de validação.

4.3. Comparação das abordagens

De forma a obter uma medida de comparação das diferentes abordagens constantes deste estudo, remete-se para a análise da Tabela 4.4 e da .

Tabela 4.4. Área de sobreposição de Copas.

Abordagens Coincidentes (Nº)	Área de Copas (%)
1	25,85
2	16,52
3	18,79
4	20,16
5	13,71
6	4,96

Observa-se que as copas das árvores são reconhecidas de formas distintas pelas várias abordagens (1 a 6). Em geral, a zona central das copas das árvores é reconhecida como tal por um número crescente de abordagens. Quer isto dizer que existe uma coincidência nos processos analíticos das diversas abordagens, que conduzem a resultados similares para a definição como copas dessas áreas.

Da área total reconhecida como copa de árvore pelas abordagens, apenas 4,96% é identificada como tal por todas as abordagens, contra 25,85% por uma única. No entanto uma percentagem considerável – 38,83% - é definida como copa de árvore por quatro ou mais, sendo que se pode afirmar com algum grau de confiança, que essa área pertence efetivamente a uma copa de árvore na superfície terrestre.

A definição dos limites das copas é que demonstra ser a parte mais sensível e complexa do processo, diferindo entre as abordagens utilizadas, podendo constituir uma explicação para o elevado valor da área identificada como copa apenas por três ou menos abordagens (61,16%).

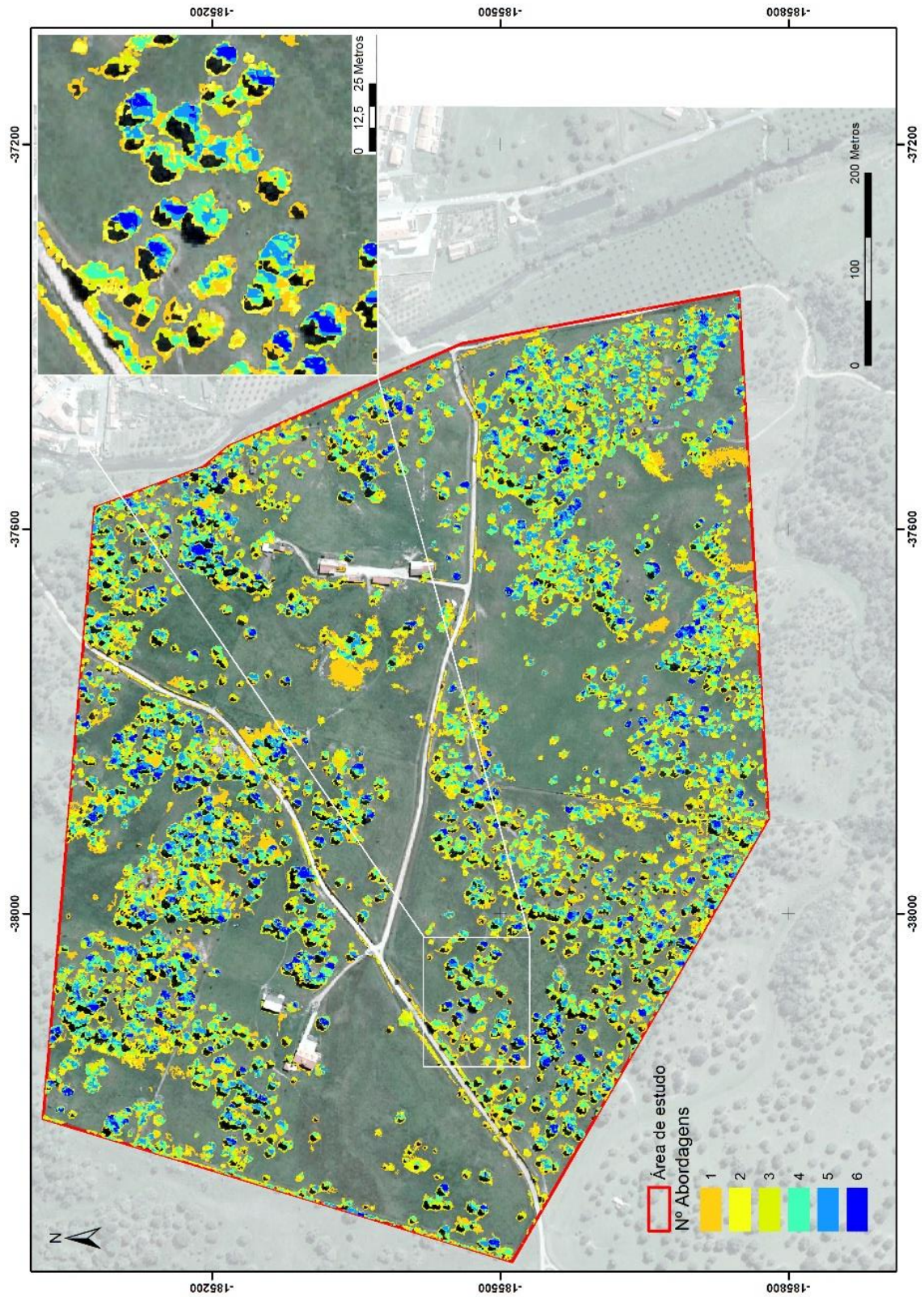


Figura 4.7. Áreas Individualizadas como Copas, coincidentes para as diversas abordagens (1 - 6).

5. Considerações finais e perspectivas futuras

Atendendo à importância do processo de ICA para a obtenção de cartografia de Copas de Árvores e considerando as potencialidades que um conhecimento mais aprofundado sobre esta temática poderá vir a representar no contexto de extração e estimação de diversos parâmetros florestais de uma forma indireta, o presente trabalho teve como objetivos explorar diferentes abordagens e construir, em consequência, uma base de conhecimento que reúna informação passível de ser utilizada na extração de informação florestal. Esta base de conhecimento vem possibilitar um maior sucesso na extração de parâmetros florestais, essenciais para uma gestão florestal de qualidade. Sendo a ICA uma etapa estruturante para a extração de parâmetros florestais de qualidade, torna-se central a exploração de diferentes abordagens e a compilação de informação relativa às mesmas.

O trabalho apresentado seguiu uma linha de desenvolvimento que se considerou adequada dentro dos objetivos definidos, mas qualquer projeto que proceda a operações de processamento e análise de imagens inclui erros que, naturalmente, se querem reduzidos, mas não devem ser ignorados. O caso da ICA não é diferente. Assim, parte das fragilidades dos resultados obtidos poderão estar relacionados com a influência conjunta ou individual de erros de amostragem, erros na classificação por limares e erros na informação de referência. O facto de os dados de referência terem sido obtidos através da análise visual da imagem e do delineamento manual pode ter influenciado os resultados obtidos. A utilização de 40 copas de referência, apesar de ter sido considerado um compromisso aceitável, pode não ter sido suficiente, para a realidade que a área de estudo representa. Assim, a amostra de validação usada foi suficiente para desenvolver as medidas de validação, mas deverá ser melhorada de forma a obterem-se estimativas mais exatas acerca da qualidade da cartografia produzida.

Por outro lado, as variáveis utilizadas e os intervalos respetivos de valores da classificação por limiaries podem não ter sido os mais adequados. A utilização de diferentes variáveis poderá otimizar os resultados e a utilização de um método alternativo para a classificação por limiaries poderá ser uma alternativa que permita a melhoria da cartografia produzida. Estes erros não podem ser totalmente eliminados, contudo a consciência deles pode levar

a que e a sua ocorrência e magnitude sejam minimizadas, reduzindo a sua expressão nas conclusões que os resultados permitem construir.

Apesar de na sua generalidade a tarefa de ICA não ter sido alcançada com resultados excelentes, a metodologia desenvolvida permitiu retirar algumas considerações válidas. É importante sublinhar que apesar das conclusões apresentadas, o trabalho de investigação e desenvolvimento de uma abordagem metodológica eficiente e executável, no âmbito da ICA, não está esgotado. A exploração da utilização de outra informação para a implementação destas abordagens, de outras ferramentas, ou mesmo da exploração de outros algoritmos é uma crescente linha de desenvolvimento.

Os resultados evidenciaram que as abordagens de *machine learning* adotadas são uma linha de desenvolvimento a aprofundar para o objetivo em causa. Não só demonstraram valores de exatidão consideravelmente aceitáveis, e os melhores dentro das abordagens executadas, como visualmente produziram uma cartografia com uma qualidade no geral satisfatória e significativamente melhor do que a produzida por outras abordagens. Acresce sublinhar que são capazes de gerir aspetos específicos do sistema florestal do montado, tais como as lacunas existentes nas copas das espécies florestas constituintes. No entanto, continuam a carecer de aprofundamento na sua aplicação, nomeadamente na tentativa de correção dos principais problemas verificados ao nível da junção de diversas copas numa única e na distinção das zonas de sombra.

Mais se verifica que as abordagens frequentemente utilizadas no âmbito da ICA não demonstraram ser adequadas no caso específico em análise neste trabalho. A única exceção é a abordagem com o algoritmo de Multiresolução. Este, apesar da sua menor performance relativamente as abordagens *machine learning*, produz um resultado satisfatório no âmbito das abordagens com algoritmos de segmentação de imagens. A sua exploração mais aprofundada poderia (atendendo das amplas possibilidades oferecidas pela implementação do algoritmo) ser uma opção a desenvolver, de forma a melhorar a cartografia final. No entanto, o facto de não ser um procedimento metodológico com reprodutibilidade independente, sendo só possível de ser realizada com recurso ao *software* eCognition, torna difícil a difusão da sua utilização. Esta situação, embora à primeira vista possa parecer de pouca relevância, pode em contexto operacional dificultar o processo e consequentemente a obtenção de cartografia de Copas de Árvores, por se

tratar de um *software* comercial a que nem sempre é possível aceder. Assim, esta situação pode comprometer a viabilidade da exploração desta abordagem, por oposição às abordagens de *machine learning*, passíveis de acesso livre.

Outro aspeto que pode ter influenciado os resultados obtidos prende-se com as imagens utilizadas. A investigação de outro tipo de dados para a execução da ICA, nomeadamente a utilização de outro tipo de tecnologias, tais como os Veículos Aéreos Não Tripulados (VANT) ou os dados LiDar, é uma opção relevante neste contexto uma vez que pode permitir otimizar os procedimentos e eventualmente trazer novas oportunidades no âmbito da ICA. Técnicas de fotogrametria e reconstrução 3D têm vindo a demonstrar possibilidades promissoras, possibilitando a estimação de dados 3D a partir de fotografias áreas de alta resolução os quais constituem fontes para os procedimentos de ICA (Birdal *et al.*, 2017; Panagiotidis *et al.*, 2017).

Finalmente, a execução da tarefa de ICA num sistema florestal como o montado pode igualmente ter influenciado os resultados obtidos, pelas suas características distintas dos sistemas florestais onde frequentemente se aplica a ICA (Ke e Quackenbush, 2011a) (Ke e Quackenbush, 2008).

Concluído este trabalho verifica-se existirem possibilidades de melhoria que podem ser testadas neste contexto, de forma a produzir uma cartografia de Copas de Árvores mais precisa. Está criada uma base de conhecimento que abre caminho na extração indireta de informação florestal, na medida em que os objetivos principais foram cumpridos e os resultados obtidos são aceitáveis. Saliente-se a importância e a ajuda preciosa que o teste a diversas abordagens de ICA e a sua comparação poderão ter na definição da abordagem adequada para obter informação florestal de qualidade, logo mais aproximada da realidade.

Numa avaliação geral afigura-se possível a ICA em sistemas florestais mais complexos como é o montado, nomeadamente recorrendo à exploração de algoritmos de *machine learning* que demonstram um enorme potencial. Esta exploração ficará aberta a um desenvolvimento mais aprofundado. No entanto a tarefa da ICA continua ainda fortemente dependente da intervenção e decisão do utilizador.

Referências bibliográficas

- Ali, W. (2006). Tree Detection using Color, and Texture Cues for Autonomous Navigation in Forest Environment. *Image (Rochester, N.Y.)*, (June), 1–91.
- Amoda, N., e Kulkarni, R. K. (2013). Image Segmentation and Detection using Watershed Transform and Region Based Image Retrieval. *International Journal of Emerging Trends e Technology in Computer Science*, 2(2), 89–94.
- Ao, Z., Su, Y., Li, W., Guo, Q., e Zhang, J. (2017). One-class classification of airborne LiDAR data in urban areas using a presence and background learning algorithm. *Remote Sensing*, 9(10).
- ArcGIS (2018). Vector-based Landscape Analysis Tools (Extension for ArcGIS 10) 2.0 beta: V-LATE 2.0 beta. Acedido em 13 de Agosto de 2018, em: <https://www.arcgis.com/>
- Baatz, M., e Schäpe, A. (2000). Multiresolution segmentation: An optimization approach for high quality multi-scale image segmentation. *Angewandte Geographische Informationsverarbeitung XI Beiträge Zum AGIT-Symposium Salzburg*.
- Barbosa, Â. S. (2016). Determinação Semiautomática de limiares para Classificação de conjuntos multitemporais de imagens de satélite. Dissertação de Mestrado em Sistemas de Informação Geográfica e Modelação Territorial Aplicados ao Ordenamento. Lisboa: Instituto de Geografia e Ordenamento do Território, Universidade de Lisboa.
- Barilotti, A., e Sepic, F. (2010). Assessment of forestry parameters at single-tree level by using methods of LIDAR data analysis and processing. *Ambiência Guarapuava*, 6(4), 81–92.
- Bayer, S., e Bucher, T. (2008). Application of Remote Sensing and GIS Methods for the Automatic Extraction of Single Trees Based on Digital Aerial Images and Elevation Models.
- Bechtel, B., Ringeler, A., e Böhner, J. (2008). Segmentation for object extraction of trees using MATLAB and SAGA. *Hamburger Beiträge Zur Physischen Geographie Und Landschaftsökologie*, 19, 1–12.
- Belo, C. C., Pereira, M. S., Moreira, A. C., Coelho, I. S., Onofre, N., e Paulo, A. A. (2009). Montado-capítulo 8 (Vol. Ecossistem, pp. 251–293).

- Bins, L. S., Fonseca, L. M. G., Erthal, G. J., e Ii, F. M. (1996). Satellite Imagery Segmentation: a region growing approach. *III Simposia Brasileiro de Sensoriamento Remoto, Salvador, Brasil, 14-19 Abril 1996, INPE*, 677–680.
- Birdal, A., Avdan, U., e Tarik, T. (2017). Estimating tree heights with images from an unmanned aerial vehicle. *Geomatics, Natural Hazards and Risk*, 5705(November), 1–13.
- Breiman, L. (2001). Random Forests. *Machine Learning*, 45, 5-32.
- Carneiro, D. (2015). Uso de sensoriamento remoto como ferramenta para estimar características dendrométricas de um povoamento de *Pinus taeda*. Universidade Federal de Santa Catarina.
- Carvalho M., A. (2002). A economia do sector da cortiça em Portugal. *Faculdade de Economia e Gestão*, 238.
- Chenari, A., Erfanifard, Y., Dehghani, M., e Pourghasemi, H. R. (2017). Woodland mapping at single-tree levels using object-oriented classification of unmanned aerial vehicle (UAV) images. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, 42(4W4), 43–49.
- Chubey, M. S., Franklin, S. E., e Wulder, M. a. (2006). Object-based analysis of Ikonos-2 imagery for extraction of forest inventory parameters. *Photogrammetric Engineering and Remote Sensing*, 72(4), 383–394.
- Costa, H. (2008). Estratégia Multi-temporal para Produção Automática de Cartografia de Ocupação do Solo com Imagens AWiFS. Dissertação de Mestrado em Ciência e Sistemas de Informação Geográfica. Lisboa: Instituto Superior de Estatística e Gestão de Informação, Universidade Nova de Lisboa.
- Culvenor, D. S. (2002). TIDA: An algorithm for the delineation of tree crowns in high spatial resolution remotely sensed imagery. *Computers and Geosciences*, 28(1), 33–44.
- DGRF (2015). Estratégia Nacional para as Florestas.
- DGT (2018). Fotografia aérea digital. Acedido a 5 de Maio de 2018, em: <http://www.dgterritorio.pt/>
- Dong, X., Li-min, D., Guo-fan, S., Lei, T., e Hui, W. (2005). Forest fire risk zone mapping from satellite images and GIS for Baihe Forestry Bureau, Jilin, China. *Journal of Forestry Research*, 16(3), 169–174.
- Dzeroski, S., e Zenko, B. (2002). Is Combining Classifiers Better than Selecting the Best One? *Machine Learning*, 54(3), 255–273.

- Edson, C., e Wing, M. G. (2011). *Airborne light detection and ranging (LiDAR) for individual tree stem location, height, and biomass measurements. Remote Sensing* (Vol. 3).
- Elith, J., Leathwick, J. R., e Hastie, T. (2008). A working guide to boosted regression trees. *Journal of Animal Ecology*, 77(4), 802–813.
- Erikson, M. (2003). Segmentation of individual tree crowns in colour aerial photographs using region growing supported by fuzzy rules. *Canadian Journal of Forest Research*, 33(8), 1557–1563.
- Fan, J., Yau, D. K. Y., Elmagarmid, a K., e Aref, W. G. (2001). Automatic image segmentation by integrating color-edge extraction and seeded region growing. *IEEE Transactions on Image Processing*, 10(10), 1454–1466.
- Fransson, J. E. S., Walter, F., e Ulander, L. M. H. (2000). Estimation of forest parameters using CARABAS-II VHF SAR data. *IEEE Transactions on Geoscience and Remote Sensing*, 38(2), 720–727.
- Gabriel, C. (2013). Análise Comparada de Segmentação e Classificação orientada por objetos de uma imagem WorldView-2. Dissertação de Mestrado em Gestão do Território. Lisboa: Faculdade de Ciências Sociais e Humanas, Universidade Nova de Lisboa.
- Gatziolis, D., Lienard, J. F., Vogs, A., e Strigul, N. S. (2015). 3D tree dimensionality assessment using photogrammetry and small unmanned aerial vehicles. *PLoS ONE*, 10(9), 1–22.
- Gomes, M. F., e Maillard, P. (2013). Identification of urban tree crown in a tropical environment using WorldView-2 data: problems and perspectives. *Proceedings of SPIE – The International Society for Optical Engineering*.
- Gomes, M., e Maillard, P. (2016). Detection of Tree Corwns in Very High Spatial Resolution Images. In *Environmental Applications of Remote Sensing*. InTech.
- Gougeon, F. (1997). Recognizing the forest from the trees: Individual tree crown delineation classification and regrouping for inventory purposes. *Third International Airborne Remote Sensing Conference and Exhibition*, 2(July), 807–814.
- Gougeon, F. A. (1995). A crown-following approach to the automatic delineation of individual tree crowns in high spatial resolution aerial images. *Canadian Journal of Remote Sensing*, 21(3), 274–284.
- Gougeon, F., e Leckie, D. (2003). *Forest information extraction from high spatial resolution images using an individual tree crown approach. Forestry*.

- Gougeon, F., e Leckie, D. G. (2006). The individual tree crown approach applied to Ikonos images of a coniferous plantation area. *Photogrammetric Engineering and Remote Sensing*, 72, 1287–1297.
- Grinand, C., Rakotomalala, F., Gond, V., Vaudry, R., Bernoux, M., e Vieilledent, G. (2013). Estimating deforestation in tropical humid and dry forests in Madagascar from 2000 to 2010 using multi-date Landsat satellite images and the random forests classifier. *Remote Sensing of Environment*, 139, 68–80.
- Hosmer, D. W., e Lemeshow, S. (2000). *Applied Logistic Regression*.
- Huete, A., Didan, K., Miura, T., Rodriguez, E. P., Gao, X., e Ferreira, L. G. (2002). Overview of the radiometric and biophysical performance of the MODIS vegetation indices, 83, 195–213.
- ICNF (2018). Importância Económica. Acedido a 19 de Novembro de 2018, em: <http://www2.icnf.pt/>
- INF (2010). Inventário Nacional Florestal.
- Immitzer, M., Atzberger, C., e Koukal, T. (2012). Tree species classification with Random forest using very high spatial resolution 8-band worldView-2 satellite data. *Remote Sensing*, 4(9), 2661–2693.
- Jaynes, E. T. (1957). Information Theory and Statistical Mechanics. *Physical Review*, 106 (4), 620-630.
- Jing, L., Hu, B., Noland, T., e Li, J. (2012). An individual tree crown delineation method based on multi-scale segmentation of imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 70, 88–98.
- Kapás, Z., Lefkovits, L., e Szilágyi, L. (2016). Automatic Detection and Segmentation of Brain Tumor Using Random Forest Approach, 301–312.
- Ke, Y., e Quackenbush, L. (2007). Forest Species Classification and Tree Crown Delineation Using Quickbird Imagery. *ASPRS 2007 Annual Conference*, 9.
- Ke, Y., e Quackenbush, L. J. (2008). Comparison of individual tree crown detection and delineation methods. *ASPRS 2008 Annual Conference Proceedings*, 11.
- Ke, Y., e Quackenbush, L. J. (2011a). A comparison of three methods for automatic tree crown detection and delineation from high spatial resolution imagery. *International Journal of Remote Sensing*, 32(13), 3625–3647.
- Ke, Y., e Quackenbush, L. J. (2011b). A review of methods for automatic individual tree-crown detection. *International Journal of Remote Sensing*, 32(17), 4725–4747.
- Ke, Y., e Quackenbush, L. J. (2011c). A review of methods for automatic individual tree-

- crown detection and delineation from passive remote sensing. *International Journal of Remote Sensing*, 32(17), 4725–4747.
- Koch, B., Heyder, U., e Weinacker, H. (2006). Detection of Individual Tree Crowns in Airborne Lidar Data. *Photogrammetric Engineering e Remote Sensing*, 72(4), 357–363.
- Korpela, I. (2005). *Individual Tree Measurements by Means of Digital Aerial Photogrammetry. Silva Fennica Monographs*.
- Landis, J. R., e Koch, G. G. (1977). The measurement of observer agreement for categorical data. *International Biometric Society*, 33(1), 159–174.
- Larsen, M., Eriksson, M., Descombes, X., Perrin, G., Brandtberg, T., e Gougeon, F. A. (2011). Comparison of six individual tree crown detection algorithms evaluated under varying forest conditions. *International Journal of Remote Sensing*, 32(20), 5827–5852.
- Lillesand, T. M. e Kiefer, R. W. (2000). *Remote Sensing and image segmentation*. 4th Edition. New York: Kiefer Wiley and sons.
- Li, W., e Guo, Q. (2010). A maximum entropy approach to one-class classification of remote sensing imagery. *International Journal of Remote Sensing*, 31(8), 2227–2235.
- Li, Z., Hayward, R., Zhang, J., Liu, Y., e Walker, R. (2009). Towards automatic tree crown detection and delineation in spectral feature space using PCNN and morphological reconstruction. *Proceedings - International Conference on Image Processing, ICIP*, (December), 1705–1708.
- Liu, H. (2013). Automated Treetop Detection and Tree Crown Identification.
- López, J. P. A. (2012). Object-based methods for mapping and monitoring of urban trees with multitemporal image analysis. *Faculty of Geo-Information Science and Earth Observation*, 176.
- Miranda, P. (2004). Aplicação da detecção remota em inventário florestal. Dissertação de Douturamento em Ciências Florestais. Vila Real: Universidade de Trás-os-montes e Alto Douro.
- Nagendra, H., Rocchini, D., Ghate, R., Sharma, B., e Pareeth, S. (2010). Assessing plant diversity in a dry tropical forest: Comparing the utility of landsat and ikonos satellite images. *Remote Sensing*, 2(2), 478–496.
- Niccolai, A., Hohl, A., Niccolai, M., e Oliver, C. D. (2010). Integration of varying spatial, spectral and temporal high-resolution optical images for individual tree crown

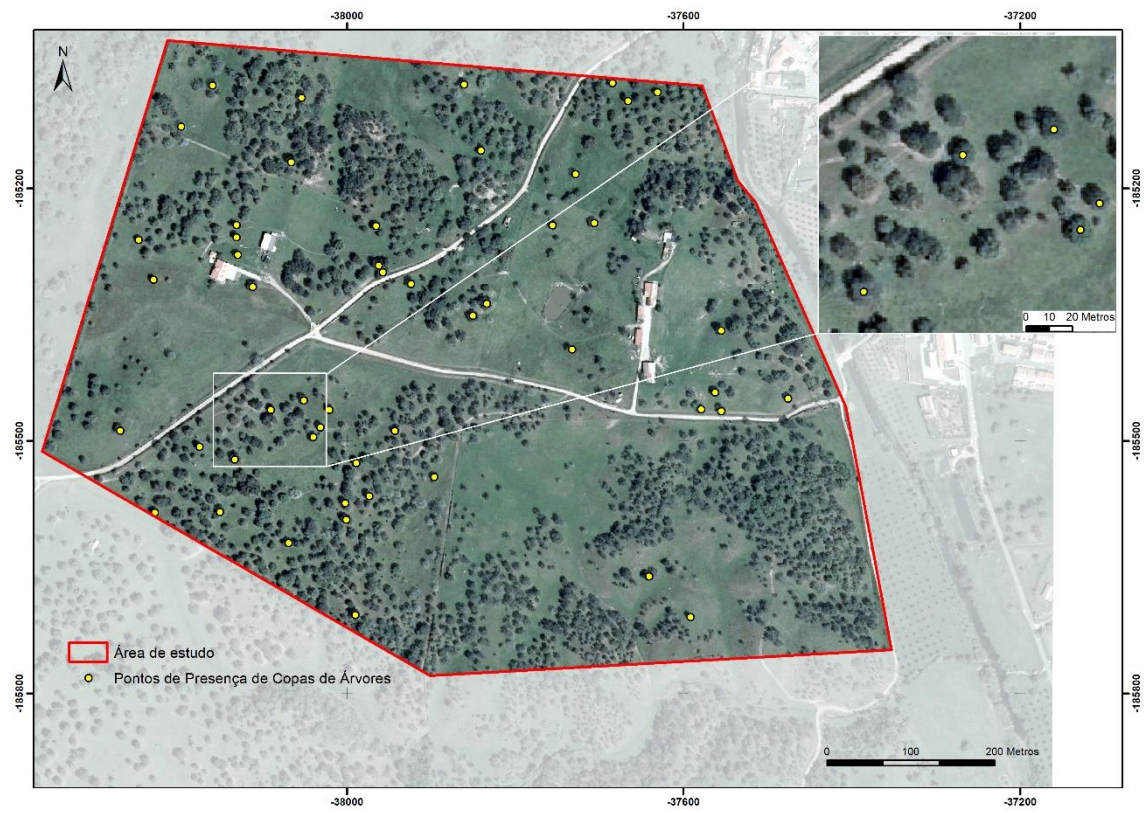
- isolation. *International Journal of Remote Sensing*, 31(19), 5061–5088.
- Panagiotidis, D., Abdollahnejad, A., Surový, P., e Chiteculo, V. (2017). Determining tree height and crown diameter from high-resolution UAV imagery. *International Journal of Remote Sensing*, 38(8–10), 2392–2410.
- Pohle, R., e Toennies, K. D. (2001). Segmentation of medical images using adaptive region growing, 1337–1346.
- Popescu, S. C., Wynne, R. H., e Nelson, R. F. (2003). Measuring individual tree crown diameter with lidar and assessing its influence on estimating forest volume and biomass. *Remote Sensing*, 29(5), 564–577.
- Portal do Ambiente e do Cidadão (2018). *Sobre a importância das florestas*. Acedido a 19 de julho de 2018, em: <http://ambiente.maiadigital.pt/>
- Pouliot, D. A., King, D. J., e Pitt, D. G. (2005). Development and evaluation of an automated tree detection–delineation algorithm for monitoring regenerating coniferous forests. *Canadian Journal of Forest Research*, 35(10), 2332–2345.
- Puissant, A., Rougier, S., e Stumpf, A. (2014). Object-oriented mapping of urban trees using Random Forest classifiers. *International Journal of Applied Earth Observation and Geoinformation*, 26, 235–245.
- Ribas, R. (2011). Individualização de Árvores em ambiente florestal nativo utilizando métodos de segmentação em Modelos Digitais produzidos a partir de tecnologia LIDAR. Dissertação de Mestrado em Análise e Modelagem de Sistemas Ambientais. Universidade Federal de Minas Gerais.
- Ribeiro, J. (2007). A Detecção Remota no Inventário Florestal: Análise das potencialidades da utilização das imagens de satélite. Dissertação de Mestrado em Ciência e Sistemas de Informação Geográfica. Lisboa: Instituto Superior de Estatística e Gestão da Informação, Universidade Nova de Lisboa.
- SAGA GIS (2018). SAGA-GIS Module Library Documentation. Acedido em 10 de Março de 2018, em: <http://www.saga-gis.org/> Março de 2018, em: <http://www.saga-gis.org/>
- Saliola, A. (2014). Object-based Image Analysis for the Delineation of Canopy Gaps and Individual Tree Crowns using Multi-source Data: A Case Study in Haliburton Forest, Ontario.
- Salman, N. H., e Liu, C. Q. (2003). Image segmentation and edge detection based on watershed techniques. *International Journal of Computers and Applications*, 25(4), 258–263.

- Santos, J. M. (2018). Análise e modelação espaço-temporal do mosquito vetor do dengue na ilha da Madeira. Dissertação de Mestrado em Sistemas de Informação Geográfica e Modelação Territorial aplicados ao Ordenamento. Lisboa: Instituto de Geografia e Ordenamento do Território, Universidade de Lisboa.
- Schiewe, J. (2002). Segmentation of high-resolution remotely sensed data - concepts, applications and problems. *Symposium on Geospatial Theory, Processing and Applications*.
- Schroff, F., Criminisi, A., e Zisserman, A. (2008). Object Class Segmentation using Random Forests. *Proceedings of the British Machine Vision Conference 2008*, 54.1-54.10.
- Soares, P., Firmino, P., Tomé, M., Campagnolo, M., Oliveira, J., Oliveira, B., e Araújo, J. (2015). A utilização de Veículos Aéreos Não Tripulados no inventário florestal – o caso do montado de sobro. *Viiiicng.Ordemengenheiros.Pt*, 1–8.
- Song, C., e Dickinson, M. B. (2008). Extracting forest canopy structure from spatial information of high resolution optical imagery: Tree crown size versus leaf area index. *International Journal of Remote Sensing*, 29(19), 5605–5622.
- Sousa, A. M. O., Mesquita, P., Gonçalves, A. C., Marques, J. R., e Silva, L. L. (2011). Detecção de vegetação arbórea através de segmentação e classificação orientada a objecto de imagens multiespectrais de alta resolução (Quickbird). *VI Congresso Ibérico de Agro-Engenharia*, 1–8.
- Stehman, S. (2001). Statistical Rigor and Practical Utility in Thematic Map Accuracy Assessment. *ISPRS Journal of Photogrammetry and Remote Sensing*, 67(6), 727–734.
- Strîmbu, V. F., e Strîmbu, B. M. (2015). A graph-based segmentation algorithm for tree crown extraction using airborne LiDAR data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 104, 30–43.
- Trimble. (2012). eCognition Developer XD 2.0.4 User Guide. Munique: Trimble Documentation.
- Tsendbazar, N. (2011). Object based image analysis of GEO-EYE VHR data to model above ground carbon stock in Himalayan mid-hill forests, Nepal.
- Turner, M. G., Gardner, R. H., e O'Neill, R. V. (2001). *Landscape Ecology in Theory and Practice*. Springer.
- van der Sande, C. J., de Jong, S. M., e de Roo, A. P. J. (2003). A segmentation and classification approach of IKONOS-2 imagery for land cover mapping to assist flood

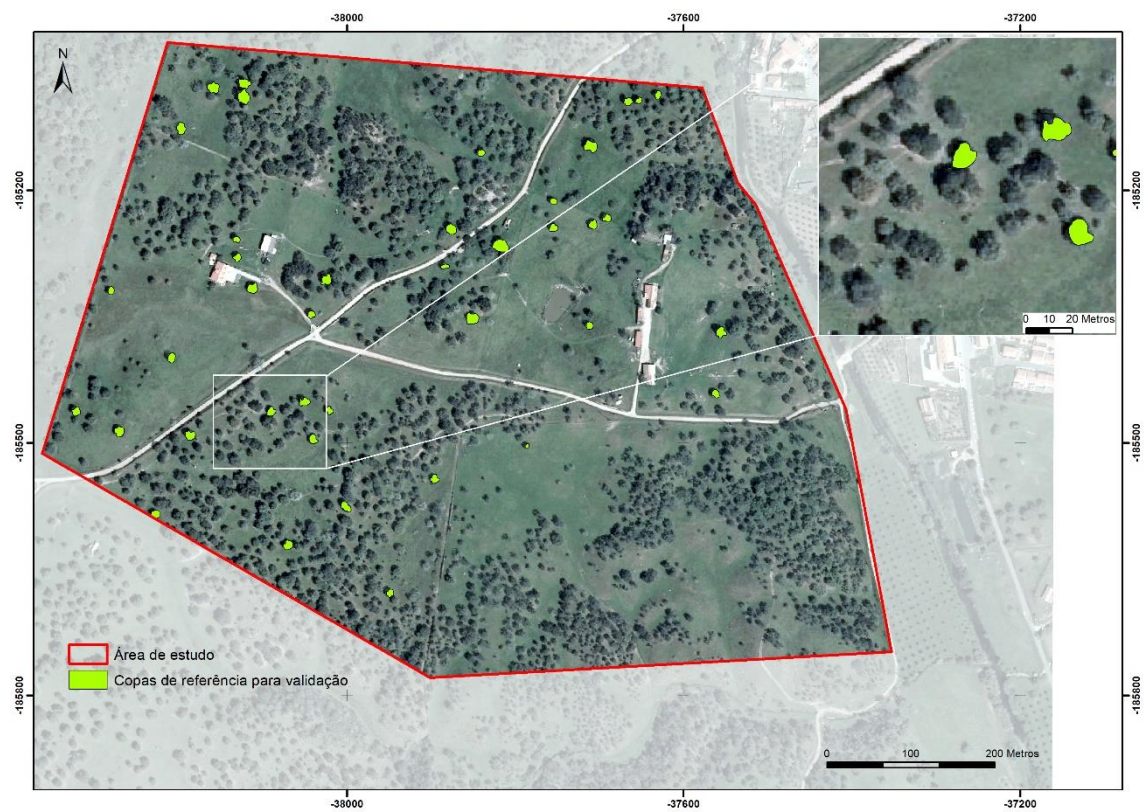
- risk and flood damage assessment. *International Journal of Applied Earth Observation and Geoinformation*, 4(3), 217–229.
- Vauhkonen, J., Ene, L., Gupta, S., Heinzl, J., Holmgren, J., Pitkänen, J., ... Maltamo, M. (2012). Comparative testing of single-tree detection algorithms under different types of forest. *Forestry*, 85(1), 27–40.
- Wang, L., Gong, P., e Biging, G. S. (2004). Individual Tree-Crown Delineation and Treetop Detection in High-Spatial-Resolution Aerial Imagery. *Photogrammetric Engineering e Remote Sensing*, 70(3), 351–357.
- Wulder, M., Niemann, K. O., e Goodenough, D. G. (2000). Local maximum filtering for the extraction of tree locations and basal area from high spatial resolution imagery. *Remote Sensing of Environment*, 73(1), 103–114.
- WWF (2018). *The importance of Forests cannot be underestimated*. Acedido a 19 de julho de 2018, em: <https://wwf.panda.org/>
- Yang, J., He, Y., e Caspersen, J. (2014). A multi-band watershed segmentation method for individual tree crown delineation from high resolution multispectral aerial image. *International Geoscience and Remote Sensing Symposium (IGARSS)*, 1588–1591.
- Zandoná, D. F., Lingnau, C., e Nakajima, N. Y. (2008). Varredura a Laser aerotransportado para estimativa de variáveis dendrométricas Airborne Laser Scanner technology for estimating dendrometric variables. *Scientia Forestalis, Piracicaba*, 36(80), 295–306.
- Zhang, W., Ke, Y., Quackenbush, L. J., e Zhang, L. (2010). Using error-in-variable regression to predict tree diameter and crown width from remotely sensed imagery. *Canadian Journal of Forest Research*, 40(6), 1095–1108.

Anexos

Anexo 1. Pontos de Presença utilizados nos algoritmos de *Machine Learning*.



Anexo 2. Copas de Referência

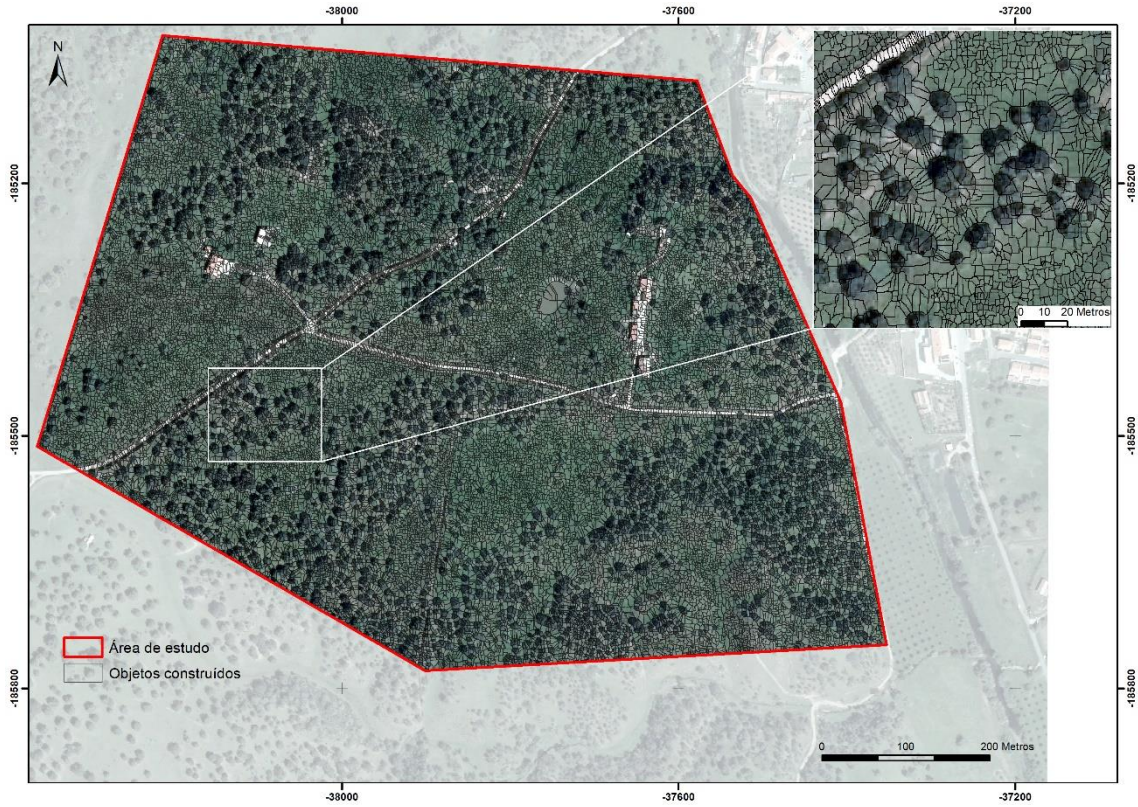


Anexo 3. Medidas de avaliação da qualidade da cartografia produzida para as diversas tentativas em cada abordagem.

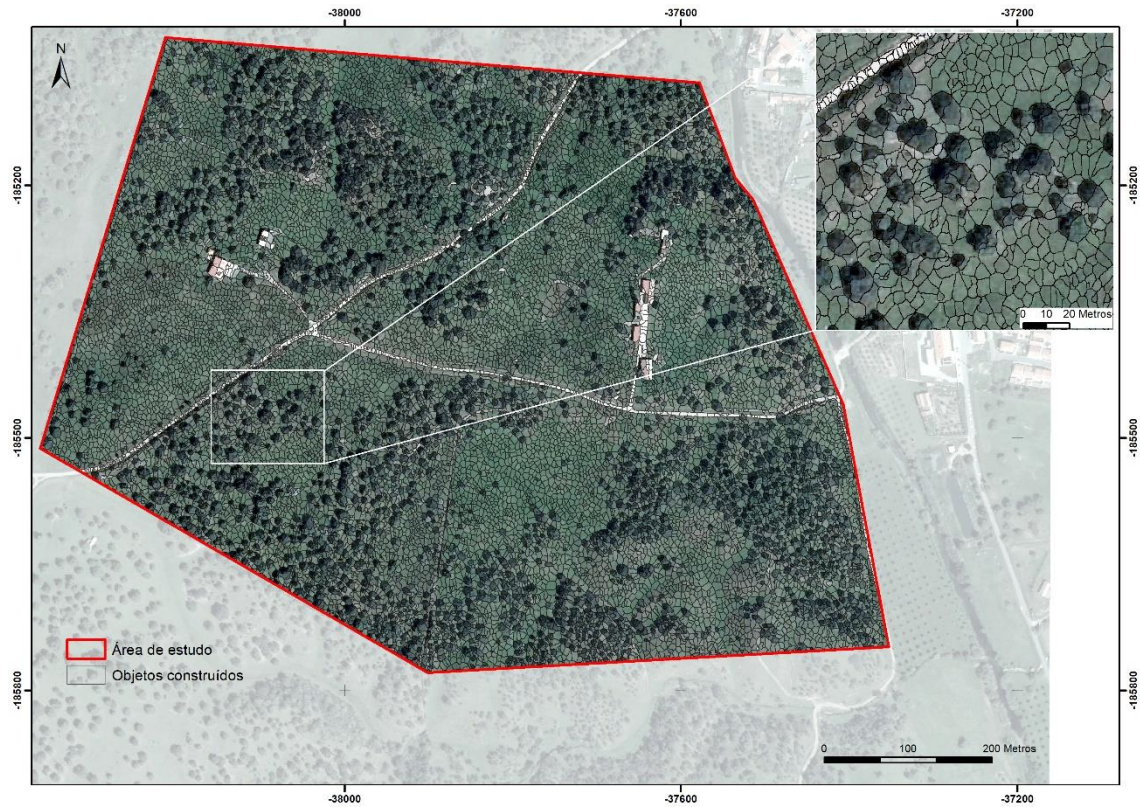
Abordagem (de acordo com o algoritmo de construção de objetos)	Tentativa	Medidas complementares de avaliação				Matriz de Confusão para a Classe de Interesse (Copas)			
		Área de Copas Individualizadas		Área de Copas de Referência		Coeficiente de Concordância <i>a Kappa</i>	EU (C0)	EP (C1)	EG
		Área de Comissão (%)	Área Correta (%)	Área de Omissão (%)	Área Correta (%)				
WS	1	42,38	57,62	86,74	13,26	0,285	0,58	0,29	0,64
	2	41,02	58,98	85,97	14,03	0,34	0,60	0,35	0,67
	3	37,03	62,973	86,59	13,41	0,33	0,6297	0,34	0,67
RG	1	48,18	51,83	74,23	25,77	0,37	0,61	0,38	0,69
	2	40,48	59,52	78,22	21,78	0,24	0,57	0,24	0,62
	3	37,76	62,24	64,97	35,03	0,30	0,59	0,31	0,65
	4	36,96	63,04	70,24	29,76	0,25	0,57	0,25	0,62
	5	36,58	63,42	71,97	28,03	0,32	0,59	0,32	0,66
MR	1	46,11	53,89	52,99	47,01	0,47	0,65	0,47	0,73
	2	46,14	53,86	57,87	42,13	0,39	0,62	0,39	0,69
	3	52,70	47,3	49,81	50,19	0,34	0,60	0,35	0,67
	4	35,32	64,68	76,85	23,15	0,44	0,64	0,44	0,72
	5	41,56	58,45	64,92	35,08	0,48	0,66	0,49	0,74
	6	44,37	55,63	56,68	43,32	0,47	0,65	0,48	0,74
	7	51,25	48,75	46,25	53,75	0,46	0,65	0,46	0,73
	8	63,88	36,12	49,24	50,76	0,38	0,62	0,39	0,69
	9	45,57	54,43	53,08	46,92	0,42	0,63	0,42	0,71
	10	51,98	48,02	48,27	51,73	0,43	0,64	0,44	0,72
	11	45,15	54,85	56,53	43,47	0,41	0,63	0,42	0,71
RF	1	26,86	73,14	84,23	15,77	0,13	0,54	0,13	0,57
	2	55,31	44,69	44,51	55,49	0,62	0,73	0,64	0,81
MaxEnt	1	47,07	52,93	55,71	44,29	0,55	0,69	0,55	0,77
	2	42,07	57,93	57,37	42,63	0,47	0,66	0,48	0,74

Combinação (RF e MaxEnt)	1	41,74	58,26	65,78	34,22	0,48	0,66	0,49	0,74
	2	48,61	51,39	50,21	49,79	0,57	0,71	0,59	0,79

Anexo 4. Objetos construídos na 1ª etapa metodológica, com o algoritmo *Watershed*.



Anexo 5. Objetos construídos na 1ª etapa metodológica, com o algoritmo de Crescimento de Regiões.



Anexo 6. Objetos construídos na 1ª etapa metodológica, com o algoritmo Multiresolução.



Anexo 5. Matriz de Confusão obtida para a abordagem que utiliza o algoritmo *Watershed*.

<i>Watershed</i>			
CLASSE	C0	C1	Total
C0	398	262	660
C1	2	138	140
Total	400	400	800

Anexo 6. Matriz de Confusão obtida para a abordagem que utiliza o algoritmo de Crescimento de Regiões.

Crescimento de Regiões			
CLASSE	C0	C1	Total
C0	398	277	675
C1	2	123	125
Total	400	400	800

Anexo 7. Matriz de Confusão obtida para a abordagem que utiliza o algoritmo de Multiresolução.

Multiresolução			
CLASSE	C0	C1	Total
C0	398	216	614
C1	2	184	186
Total	400	400	800

Anexo 8. Matriz de Confusão obtida para a abordagem que utiliza o algoritmo *Random Forests*.

<i>Random Forests</i>			
CLASSE	C0	C1	Total
C0	392	143	535
C1	8	257	265
Total	400	400	800

Anexo 9. Matriz de Confusão obtida para a abordagem que utiliza o algoritmo MaxEnt.

MaxEnt			
CLASSE	C0	C1	Total
C0	397	179	576
C1	3	221	224
Total	400	400	800

Anexo 10. Matriz de Confusão obtida para a abordagem que combina os algoritmos de *Machine Learning*.

Combinação (RF e MaxEnt)			
Classe	C0	C1	Total
C0	392	163	555
C1	8	237	245
Total	400	400	800