

**ANÁLISE DA EVOLUÇÃO ESPAÇO-TEMPORAL DE CONDIÇÕES PARA
FIXAÇÃO E PROLIFERAÇÃO DE VETORES DE *ARBOVÍRUS*, COM
RECURSO A DADOS OBTIDOS POR DETEÇÃO REMOTA**

Carlos Filipe Guerreiro Álvaro

**Dissertação de Mestrado em Gestão do Território, Área de
Especialização em Detecção Remota e Sistemas de Informação
Geográfica**

Janeiro, 2021

Dissertação apresentada para cumprimento dos requisitos necessários à obtenção do grau de Mestre em Gestão do Território, Área de Especialização em Sistemas de Informação Geográfica e Deteção Remota, realizada sob a orientação científica do Professor Doutor José António Pereira Tenedório, do Professor Doutor Fernando Jorge Pedro da Silva Pinto da Rocha e da Doutora Engenheira Teresa Alexandra Gonçalves dos Santos Silva

Esta dissertação de mestrado está enquadrada nos projetos TRIAD (*health Risk and social vulnerability to Arboviral Diseases in mainland Portugal*) | Vulnerabilidade Social e Risco para a Saúde devido às doenças arbovirais em Portugal continental (PTDC/GES-OUT/30210/2017)



*Do not go gentle into that good night,
Old age should burn and rave at close of day;
Rage, rage against the dying of the light.*

*Though wise men at their end know dark is right,
Because their words had forked no lightning, they
Do not go gentle into that good night.*

*Good men, the last wave by, crying how bright
Their frail deeds might have danced in a green bay,
Rage, rage against the dying of the light.*

*Wild men who caught and sang the sun in flight,
And learn, too late, they grieved it on its way,
Do not go gentle into that good night.*

- Dylan Thomas

*Em memória de João Álvaro
e aos que, como ele, não entram gentilmente nessa noite acolhedora*

AGRADECIMENTOS

Ao Professor Doutor José António Tenedório por ter aceite orientar esta dissertação, por tudo o que aprendi e por todo o tempo despendido para a concretização desta dissertação bem como à Doutora Teresa Santos e ao Professor Doutor Jorge Rocha por terem aceitado a coorientação desta dissertação e por todos os conselhos científicos e todos os seus contributos.

Ao Professor Doutor César Capinha pelo conselho científico sobre o *Aedes albopictus*.

Aos meus pais e irmãos por todo o apoio que me deram durante a elaboração desta dissertação. À Ana por todo o apoio que me deu durante o percurso académico desde a licenciatura ao mestrado.

À Sarah, por ter estado sempre presente e me apoiar em tudo o que foi preciso e me incentivar a ir em frente.

Ao Humberto Sousa, um irmão que encontrei na faculdade que nunca me deixou ficar mal, pelas aventuras e pelos serões de estudo e trabalhos académicos.

Ao Henrique Cerqueira, pela sua ajuda, amizade e a sua cara feia quando a conversa não lhe agrada, a quem a dívida no pato será cobrada.

Ao Ricardo Inês e ao Filipe Silveira pela grande amizade, serões na residência e aventuras por Lisboa. Ao Cristiano Martins pela boa disposição e pela sua disponibilidade para ajudar. Ao Rafa, Henrique, Jesus pela amizade, diversão e pelos momentos e festas organizadas no 131 e ao pessoal do bairro do primeiro Sul pelas aventuras e os bons tempos na residência Alfredo de Sousa. A todos os outros que me acompanharam mesmo que brevemente no meu percurso académico.

Aos SAS Nova, em especial à Dr.ª Iva pela disponibilidade e compreensão e ao *Staff* da residência por nos tratarem como família.

Ao Pato Real, por ser palco das discussões de projetos fictícios, intoxicações desnecessárias, pelo bom serviço e pela cerveja barata.

**ANÁLISE DA EVOLUÇÃO ESPAÇO-TEMPORAL DE CONDIÇÕES PARA FIXAÇÃO E
PROLIFERAÇÃO DE VETORES DE ARBOVIRUS, COM RECURSO A DADOS OBTIDOS POR
DETEÇÃO REMOTA**

CARLOS FILIPE GUERREIRO ÁLVARO

RESUMO

As alterações climáticas e a crescente artificialização de espaços naturais têm produzido impactos que se traduzem, designadamente, na alteração de habitats naturais. O impacto destas mudanças nos *arbovírus*, ou seja, nos vírus que são transmitidos por animais artrópodes, com particular destaque para os mosquitos como transmissores destes, têm representado um grande desafio à saúde pública, devido às alterações climáticas e ambientais que favorecem a amplificação da transmissão de vírus e a transposição da barreira entre espécies. As mudanças no uso e ocupação do solo são um fator de potencialização da proliferação das espécies de mosquitos que transmitem doenças como a Dengue, o Zika e Chicungunha. A par disto, a variabilidade espaço-temporal dos índices de vegetação (*Vegetation indices*) tem sido positivamente correlacionada com a taxa de incidência destas doenças. A utilização de tecnologias de observação da terra (OT), como meio para sinalização destes habitats, constitui uma ferramenta de combate permitindo antecipar, prevenir e controlar as novas doenças infecciosas emergentes, em especial destaque, *arbovírus*, para evitar grandes riscos sociais e económicos.

PALAVRAS-CHAVE: Vetor, Uso do Solo, NDVI, Habitats, Arbovírus, Espaço, Tempo, Detecção Remota

SPATIOTEMPORAL ANALYSIS ON THE EVOLUTION OF IDEAL CONDITIONS FOR FIXING AND PROLIFERATING OF ARBOVIRUS VECTORS WITH REMOTE SENSING DATA

CARLOS FILIPE GUERREIRO ÁLVARO

ABSTRACT

Climate change and the growing artificialization of natural spaces have produced impacts that translate, inter alia, to the change of natural habitats. The impact of these changes on arboviruses, that is, on viruses that are transmitted by insect bite, with particular emphasis on mosquitoes as transmitters of these viruses, have represented a major challenge to public health, due to climate and environmental changes that favor amplification and transmission of viruses and the transposition of the barrier between species. Land use and occupation changes are a factor in enhancing the proliferation of mosquito species that transmit diseases such as Dengue, Zika and Chicungunha. Alongside this, the spatiotemporal variability in vegetation indices has been positively correlated with the incidence rate of these diseases. The use of remote sensing, as a means for signaling these habitats, provides a tool that enables anticipating, preventing, and controlling new emerging infectious diseases, especially arboviruses, to avoid major social and economic risks.

KEYWORDS: Vector, Land Use, NDVI, Habitats, Arbovirus, Space, Time, Remote Sensing

ÍNDICE

Introdução.....	1
Capítulo I: O vetor <i>Aedes albopictus</i> e a Detecção Remota	12
I. 1. O vetor <i>Aedes albopictus</i>	13
I. 2. Historia e evolução da dispersão do vetor no mundo e na Europa	14
I. 2.1. História da dispersão da espécie pelo mundo.....	15
I. 2.2. Evolução da dispersão do vetor na Europa	19
I. 3. A Detecção Remota ao serviço da monitorização do vetor.....	21
I. 4. Os métodos e o âmbito geográfico	26
Capítulo II: A preparação de dados <i>Sentinel-2</i> MSI	30
II. 1. Os dados.....	31
II. 2. O pré processamento dos dados	34
II. 2.1. Correções atmosféricas	34
II. 2.2. Operações geométricas sobre a imagem	40
II. 2.3. Estatísticas da imagem e qualidade dos dados	40
Capítulo III: Análise multitemporal do coberto vegetal	44
III. 1. Os índices radiométricos e a sua importância.....	45
III. 2. <i>O Normalized Difference Vegetation Index</i>	47
III. 3. O cálculo do índice e mapeamento multitemporal da vegetação..	49
III. 4. A relação espacial entre os níveis de atividade clorofilina e ocorrência de <i>Aedes albopictus</i>	51
Capítulo IV: Análise da ocupação do solo por <i>Object Based Image Analysis</i> .	54
IV. 1. <i>A Object Based Image Analysis</i>	55
IV. 2. A segmentação de imagem.....	59
IV. 3. A recolha e validação de amostras	65

IV. 4. A extração das estatísticas dos objetos	67
IV. 5. O treino dos classificadores	68
IV. 6. A classificação e nomeação de objetos	76
IV. 7. A validação temática	77
IV. 8. A matriz de admissibilidade e mitigação de imprecisão temática .	83
Capítulo V: A Análise exploratória dos dados	90
V. 1. O conceito de análise exploratória de dados	91
V. 2. Os corredores de propagação potencial	92
Capítulo V: Análise e discussão de resultados	100
Considerações finais e investigação futura.....	114
Referências Bibliográficas	120
Lista de figuras	138
Lista de tabelas	141
Lista de fluxogramas	142

LISTA DE ABREVIATURAS

- AOT** – Aerosol Optical Thickness
- APDA** – Atmospheric Pre-corrected Differential Absorption
- BOA** – Bottom of the Atmosphere Reflectances
- BRDF** – Bidirectional Reflectance Distribution Function
- CDZ** – Chikungunha, Dengue e Zika
- DR** – Detecção Remota
- ECDC** – European centre for disease control
- EE** – Elemento Estruturante
- ESA** – European Space Agency
- ESDA** – Exporatory Spatial Data Analysis
- GBIF** – Global Biodiversity Information facility
- LUT** – Look Up Tables
- MAIT** – Matriz de Admissibilidade de Imprecisão Temática
- MSI** – MultiSpectral Instrument
- OBIA** – Object Based Image Analysis
- OT** – Observação da Terra
- OTB** – OrfeoToolBox
- ROI** – Region of Interest
- RTE** – Radiative Transfer Equation
- SciHub** – Copernicus Open Access Hub
- SC** – Scene Classification
- SIG** – Sistemas de informação Geográfica
- SNAP** – Sentinel Application Platform
- SVM** – Support vector machines
- SVC** – Support Vector Classifier
- TOA** – Top of the Atmosphere Reflectances
- TRIAD** – health Risk and social vulnerability to Arboviral Diseases in mainland Portugal
- TIRS** – Thermal Infrared sensor
- VGI** – Voluntary Geographic Information

INTRODUÇÃO

Um título mais rigoroso do ponto de vista da análise que foi feita nesta dissertação teria sido “análise da evolução espaço-temporal de condições para o estabelecimento e dispersão de vetores de arbovírus, com recurso a dados obtidos por deteção remota”, porém por só se ter detetado esta possível alteração após a submissão, optou por não se proceder à sua alteração.

Entre locais distantes existe uma redução na similaridade entre espécies, isto é, dos padrões mais visíveis na natureza. Esta redução é resultado da limitação na dispersão das espécies limitada a locais próximos ou acessíveis, motivada pela proximidade ou pela mudança dos fatores ecológicos e físicos do meio ambiente que atuam como barreiras. A sua dispersão natural tende a ser espacialmente contígua, porém o comércio, as viagens feitas pelo ser humano têm vindo a quebrar estas barreiras naturais à dispersão, criando condições para que cada vez mais espécies se estabeleçam em locais distantes da sua área de distribuição nativa (Capinha, et al., 2015).

A preocupação crescente causada pelas alterações climáticas e os potenciais efeitos na transmissão de doenças através da picada de mosquitos é uma preocupação a nível mundial. «Os mosquitos transmitem algumas das doenças infecciosas mais importantes do ser humano incluindo a malária, que hoje mata entre de 0,6 e 1,2 milhões de pessoas por ano, a sua maioria crianças em países subdesenvolvidos ou em desenvolvimento» (Charles & Godfray, 2013, p.15). Os mosquitos *Aedes* não transmitem malária, mas outras doenças igualmente perigosas para a saúde pública. «Através do surto de Zika em declarado 2016, a picada de um mosquito infetado *Aedes Aegypti* e *Aedes Albopictus* tornou-se ainda mais perigosa. Estes mesmos mosquitos *Aedes* que transmitem Zika também transmitem dengue, chikungunya e febre amarela, doenças que afetam todos os segmentos da sociedade, mas com particular impacto nos mais pobres e vulneráveis. As doenças causadas pelos mosquitos *Aedes* resultam em centenas de milhares de mortes por ano – mortes que são em grande parte evitáveis através da eliminação de locais de reprodução de mosquitos e da interrupção do contacto entre humanos e mosquitos.» (Claxton & Macias, 2016).

De facto, mesmo em países onde este tipo de fenómenos não era frequente começam a verificar-se cada vez mais ocorrências potenciadas pelas alterações climáticas, mudanças no ambiente local ou o transporte accidental de espécies para partes do mundo onde estas não são nativas. A «investigação futura deverá adotar novas tecnologias, incluindo desenvolvimentos em técnicas de Detecção Remota e de modelização de dinâmicas de sistema, além de permitir uma melhor compreensão e mitigação das doenças transmitidas por mosquitos num mundo em mudança.» (Franklinos *et al.*, 2019, p.302)

A premissa central desta dissertação consta na tarefa número 4 do projeto *health Risk and social vulnerability to Arboviral Diseases in mainland Portugal (TRIAD)*, financiado pela FCT que tem como objetivo estudar a potencialidade de certos vetores (*Aedes albopictus* e *Aedes aegypti*) se estabelecerem em Portugal Continental, procurando avaliar a dispersão de doenças através da utilização de técnicas de análise espacial suportadas em Sistemas de Informação Geográfica (SIG) e Detecção Remota (DR). Este estudo pode contribuir para a avaliação da influência das alterações climáticas na saúde a nível europeu, fazendo uso de soluções inovadoras de conhecimento para a saúde.

Nesta dissertação foram aplicadas técnicas de deteção remota para o estudo e biogeográfico e ecológico da espécie *Aedes albopictus*. A biogeografia é um ramo da geografia que estuda a distribuição das espécies e ecossistemas no espaço geográfico e através do tempo, procurando responder à questão do porquê as espécies se distribuírem da forma que se distribuem (Huggett, 2004). O estudo biogeográfico de espécies que transmitem doenças, como é o caso de *Aedes albopictus* pode ter um contributo importante para a implementação de medidas de combate a estas. O estudo biogeográfico aplicado a espécies de mosquitos tem sido utilizado em contextos similares ao desta dissertação, um exemplo disso em Medeiros-Sousa et al. 2017, utiliza teorias biogeográficas para o teste de modelos preditivos para análise da composição e riqueza das espécies de mosquitos, bem como o estudo da relação entre a perda de espécies e o potencial benefício para as espécies vetores de doenças na cidade de São Paulo, Brasil. Outros trabalhos onde o ramo da biogeografia estuda espécies de

mosquitos incluem os trabalhos de Kennedy & Lucks, 1999, Foley et al., 2007; Morgan et al., 2009.

A biogeografia como ramo da geografia que estuda a distribuição das espécies e ecossistemas no espaço e no tempo beneficia da introdução das tecnologias de observação da terra e dos sistemas de informação geográfica para responder à pergunta do porquê das espécies se distribuírem da forma como se distribuem. A aplicação de detecção remota e sistemas de informação geográfica o estudo da distribuição de habitats de espécies encontra-se bem explicado no livro «GIS and Remote Sensing Applications in Biogeography and Ecology» de Millington, Walsh & Osborne, 2001 que reúne um conjunto representativo de exemplos das variadas técnicas de análise espacial e abordagens analíticas que são usadas por geógrafos, ecologistas e biogeógrafos para estudar a distribuição de plantas e animais bem como a avaliação dos processos que afetam os padrões observados à escala espacial e temporal. Estas abordagens recorrem à detecção remota e aos sistemas de informação geográfica para a extração das variáveis ecológicas e ambientais necessárias a estudos biogeográficos.

A aplicação da utilização dos sistemas de informação geográfica e da detecção remota para avaliar os parâmetros ecológicos que influenciam a capacidade de estabelecimento de espécies invasoras em determinadas áreas tem as suas fundamentações na preservação dos ecossistemas naturais e artificiais (Joshi, Leeuw & Duren, 2004). «Um dos passos cruciais para avaliar o impacto das espécies invasoras é mapear as mudanças na sua distribuição real e potencial e na abundância relativa em toda uma vasta região durante um período adequado. Embora as abordagens de detecção remota direta e indireta tenham sido usadas há muito tempo para avaliar a invasão de espécies vegetais, a distribuição de animais invasores baseia-se principalmente em métodos indiretos que dependem de *proxies* ambientais de condições adequadas à colonização por uma determinada espécie.» (Rocchini et al., 2015; p.284).

O estudo feito nesta dissertação enquadra-se então no estudo indireto do vetor *Aedes albopictus* através da utilização dos *proxies* ambientais de condições adequadas ao estabelecimento de *Aedes albopictus* já que a «distribuição de mosquitos depende

em grande parte da distribuição espacial dos seus locais de reprodução larvar, da sua gama de voo e da distribuição espacial dos seus hospedeiros preferidos. Estes são todos heterogêneos no espaço e no tempo e o SIG tem, portanto, muitas aplicações potenciais para o estudo dos mosquitos e das doenças que transmitem» (Gimnig, Hightower & Hawley, 2005; p. 27).

A OT e os SIG tornaram-se numa ferramenta potente na ligação entre a biogeografia, ecologia e a geografia médica na qual a observação da terra se encontra frequentemente incorporada. Este campo tornou-se extremamente útil na compreensão do panorama geral da saúde pública e tem sido utilizado para mapear a informação epidemiológica de certas doenças. Este pacote de informação incorpora dados sobre a transmissão epidémica da doença, a distribuição espacial e os determinantes dos estados ou eventos relacionados com a saúde num determinado território com uma determinada população num determinado tempo ou período temporal. As ferramentas SIG são utilizadas em geografia médica para armazenar, visualizar, analisar e interpretar dados geográficos (Kyalo, 2020).

Dado que à data do início desta dissertação o vetor ainda não estava instalado em nenhuma parte de Portugal continental, não existindo dados de ocorrências, a área de foco experimental foi a cidade de Barcelona e áreas adjacentes, pois existem dados de ocorrências de *Aedes albopictus* suficientes para poderem ser efetuados testes de relação espacial com o uso/ocupação do solo e vegetação. O objetivo principal passa assim pela construção de uma metodologia de extração de índices biofísicos e classificação de uso/ocupação do solo sobre Barcelona para posterior utilização no estudo dos fatores de estado de superfície na previsão de potenciais habitats em Portugal continental.

Em maio de 2020 o mosquito é dado como estabelecido pela (ECDC¹) na região do Algarve e na Comunidade Intermunicipal de Tâmega e Sousa o que torna estudos deste género de suma importância para a saúde pública em Portugal.

¹ O Centro Europeu de Prevenção e Controlo das Doenças (ECDC) é o organismo Europeu encarregue de identificar as ameaças para a saúde humana tanto atuais como em vias de aparecimento, avaliando-as e fornecendo informação sobre estas. Em <https://www.ecdc.europa.eu/en> .

Pode-se então enquadrar a OT como uma importante ferramenta para o estudo dos vetores que transmitem doenças, uma vez que tem a capacidade de recolher informação sobre os sistemas físicos, químicos e biológicos da Terra. Esta é feita por meio de tecnologias de DR que permitem estudar as diferentes características biofísicas que potenciam/limitam a distribuição de vetores. A capacidade de gerar informação geográfica multiescalar é uma das grandes vantagens destes métodos, enquanto permitem estudar simultaneamente toda uma parte da terra em tempo real para retirar variáveis meteorológicas e climáticas também permite o estudo de variáveis à escala local como a extração da malha urbana.

Muitos dos estudos remetem para a modelação espacial de distribuição de mosquitos, dado que «até de 700 milhões de pessoas são infetadas e mais de um milhão morrem por ano devido a doenças transmitidas por mosquitos. Embora a grande maioria dos casos ocorram em regiões tropicais e subtropicais endémicas, os padrões internacionais de viagem e migração aumentaram a sua prevalência» (Caraballo, & King, 2014, p.1) em outras regiões. Estes trabalhos de modelação visam prever e mitigar o impacto que os vetores e as doenças que transmitem têm nas comunidades humanas, e comum a todos os estudos encontra-se o uso de dados obtidos por meios de Observação da Terra (Marechal, *et al.*, 2008; Parselia, *et al.*, 2019), em particular a malária (Kogan, 2020; Ferrão *et al.*, 2018).

As alterações climáticas e o transporte accidental de espécies invasoras de mosquitos ou espécies autóctones com capacidade de transporte de determinadas doenças erradicadas anteriormente que podem voltar a transmitir doenças caso estas sejam introduzidas nos seus habitats já motivaram vários estudos. Recorrendo à OT é possível modelar as condições ideais, em áreas onde existe vetor mas não existe doença (Marechal, *et al.*, 2008; Gomes, 2010, Gomes, Rocha & Capinha, 2016), ou em áreas nas quais uma determinada espécie se poderia instalar caso fosse introduzida acidentalmente, ou que já tenha sido introduzida, como é o caso dos vetores das doenças CDZ *Aedes albopictus* e *Aedes aegypti* em Portugal e na Europa.

As escalas para a extração de variáveis biofísicas são muito diversas: enquanto os satélites geoestacionários conseguem extrair dados climáticos para uma porção substancial da superfície da terra, estes trazem uma menor resolução espacial, o que os

torna ideais para estudos à escala do globo ou do continente enquanto que outros satélites são ideais para estudo regionais, com uma resolução média e uma cobertura de imagem na ordem dos milhares de km² e por fim as imagens de alta e muito alta resolução permitem extrair os dados biofísicos a escalas muito maiores, por exemplo ao nível da copa de uma árvore.

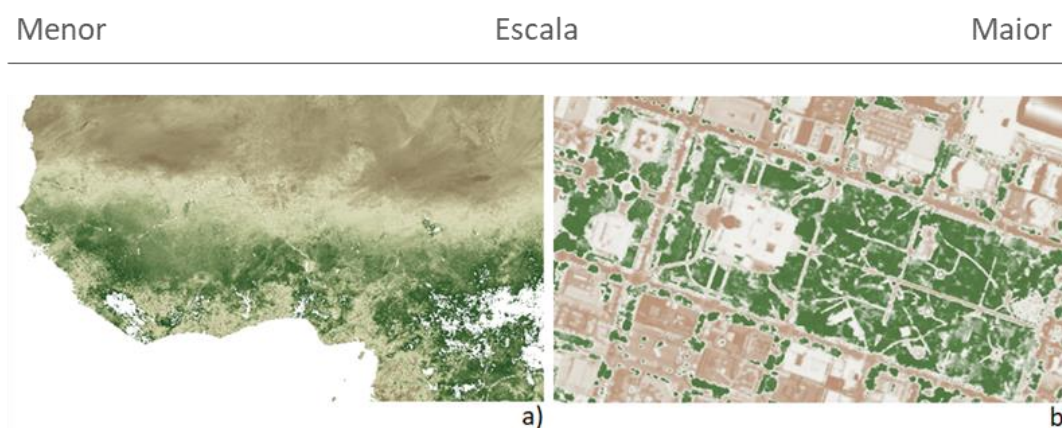


Figura 1 - NDVI à escala regional extraído do sensor Metop AVRRR (LSA SAF, 2019) para a África subsaariana (a), e NDVI extraído à escala urbana extraído obtida por imagem aérea de muito alta resolução (Hänninen, 2019) para o parque do capitólio em Sacramento (b).

Os satélites são, porventura, a fonte de informação mais exaustiva e global. A recente democratização destes dados, informação e conhecimento é uma grande mais valia para o desenvolvimento e para a justiça territorial. São também provavelmente a única fonte de informação global para muitos parâmetros dependendo da aplicação e resolução: grandes conjuntos de dados podem ser derivados de satélites em prazos relativamente curtos, em tempo semi-real (satélites geoestacionários) ou intervalos de alguns dias (satélites de órbita polar), conforme necessário e conforme a tecnologia permita. Estes dados fornecem uma rápida atualização da informação do Indicador, bem como a possibilidade de compilar séries de longa duração e que permite aos governos provas únicas para acompanhar processos, incluindo o estabelecimento de linhas de base para a determinação das tendências futuras, para previsões melhoradas, e para a gestão e mitigação. A OT também conta com a coerência e comparabilidade dos dados obtidos por satélite. (*Group on Earth Observation, 2017*)

Sendo possível obter as características do território, é possível, em teoria, extrair habitats prováveis de mosquitos. Os estudos ao nível da biologia e de dispersão geográfica servirão de base na determinação dos limites geográficos das espécies

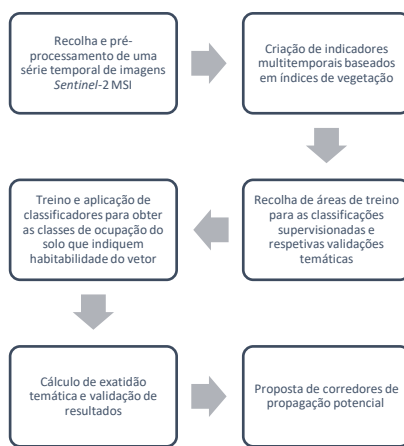
vetores. Os fatores limitantes/potenciadores dos mosquitos compreendem: o clima, a vegetação, a água, o uso e ocupação do solo, altitude e declive, para além dos fatores demográficos e sociais. Estudos anteriores demonstraram uma forte correlação entre a incidência do vetor e altos níveis de presença de vegetação (Lourenço, *et al.*, 2011) e que os habitats preferidos para a reprodução dos mosquitos são áreas alagadas pouco profundas com boa exposição solar, ao mesmo tempo associados a alguma vegetação que forneça sombra. Com as mudanças que se observam no uso e ocupação do solo derivadas pela crescente necessidade de ocupação humana dos espaços, a dispersão destes habitats tem sofrido alterações (com construções de barragens, agricultura intensiva, áreas em construção, etc.).

Considerando os resultados de trabalhos anteriores, cujos objetivos se centravam na modelação remota de suscetibilidade, risco, adequabilidade de habitats, é possível prever a incidência de vetores de *arbovírus* através da adequação dos diferentes habitats. Para modelar habitats prováveis em ambiente SIG é necessário obter informações sobre estes e, então, através do conhecimento biológico do vetor e da doença e das condicionantes biológicas é possível criar padrões de distribuição espacial (Gomes, 2010; Franke *et al.*, 2015), ou determinantes de alcance de determinadas espécies (Capinha, Rocha & Sousa, 2014) ou mesmo mapas de risco (Schaffner *et al.*, 2019).

As atuais condições ambientais podem representar um aumento na suscetibilidade de territórios a doenças transmitidas por mosquitos, como o trio Chicungunha, Dengue e Zika (CDZ), que pode ser exacerbado pelo atual e futuros cenários climáticos. As transformações que se verificam na forma como o espaço é ocupado, conjugadas com o atual panorama de alterações climáticas, podem convergir para a criação de condições mais apropriadas para o surgimento de doenças que tem o mosquito como vetor. O aquecimento da superfície terrestre não é fator suficiente para a instalação da endemicidade do vetor, porém, a forma como a superfície do planeta é transformada, tem potencial para criar condições para a reprodução e atividade dos mosquitos. Por exemplo, a criação de lagos artificiais, ou a plantação de arroz, são os usos do solo mais propícios para a instalação do vetor.

Como objetivo central desta dissertação, pretende-se identificar potenciais habitats de vetores de *arbovirus* a partir de dados *Sentinel2 MSI*, com recurso à análise orientada por objetos e cálculo multitemporal de índices de vegetação.

A utilização destes meios de investigação tem como objetivo testar técnicas de Observação da terra para localizar e mapear as áreas mais adequadas para a reprodução e propagação do vetor *Aedes albopictus*. A análise e a representação da dinâmica, no espaço e no tempo, das condições de propagação do vetor, seguem os passos que se identificam no fluxograma 1.



Fluxograma 1 - Passos para análise da dinâmica no espaço e no tempo das condições de propagação do vetor.

A influência que as mudanças do uso/ocupação do solo tem nos padrões de distribuição de mosquitos encontra-se já bem estudada bem como a consequentemente a transmissão de doenças das quais estes são os meios de contágio. A construção de reservatórios, barragens, sistemas de irrigação e a desflorestação têm vindo a criar condições ideais para a fixação de mosquitos vetores de doenças, como a malária e a dengue. No entanto, situações similares não se aplicam apenas a estes exemplos, nem ocorrem somente em países em desenvolvimento, considerando que a expansão urbana desenfreada também tem contribuído significativamente para a criação de habitats que favorecem “condições ótimas”, como a criação de paisagens fragmentadas com abundancia de alimento e abrigo, para algumas populações de vetores (Norris, 2004).

A utilização de dados obtidos por DR para avaliar as condições ambientais que influenciam a proliferação de espécies de vetores, endógenos e invasores também não é nova (Monteiro *et al.*, 2017). Por exemplo, a variabilidade espaço-temporal dos índices

de vegetação (*Vegetation indexes*) tem sido positivamente correlacionada com a taxa de incidência de *arbovirus* (*Atrhopod Borne Viruses*), ou seja, vírus que são transmitidos por mosquitos e outros artrópodes, que conseqüentemente podem causar surtos epidémicos (Lourenço *et al.*, 2011).

Considerando que a vegetação constitui apenas uma das variáveis que afeta a proliferação do mosquito, vários autores combinam-na com dados de temperatura e precipitação como variáveis ambientais dinâmicas para estimar a probabilidade de ocorrência e a abundância em locais específicos (Cleckner, Allen & Bellows, 2011).

Alguns autores também usaram a Deteção Remota para mapear prováveis habitats de larvas de mosquito (Zou, Miller & Schmidtman, 2006; El-Naggar *et al.*, 2017; El-Zeiny, El-Hefni & Sowilem, 2017) com a finalidade de fornecer ferramentas de vigilância de doenças às autoridades com responsabilidade em matéria de saúde pública, permitindo que se adotem medidas preventivas e, assim, se reduza as perdas humanas associadas.

A monitorização das mudanças de uso/ocupação do solo é um processo muito demorado em termos de logística e recursos humanos. Porém, os dados da cobertura do solo são muito importantes em vários aspetos da sociedade. No caso deste estudo, serão utilizados para ajudar a prever e mapear habitats prováveis de *Aedes albopictus*, um mosquito invasor que já se instalou em várias partes da Europa. É então importante compreender este fenómeno do ponto de vista espacial e ambiental, de forma a mitigar ou evitar cenários de surto futuros.

A forma como as transformações na ocupação do solo são suscetíveis de ter um impacto mais direto no risco de emergência de CDZ é pelo facto que potencial nas áreas de reprodução de mosquitos (por exemplo, zonas húmidas) e a taxa de contacto entre pessoas e mosquitos. Com a crescente pressão sobre habitats, não só se criam mais condições para a instalação e proliferação de vetores, como se aproximam as pessoas destes locais.

Mapear as áreas de reprodução e potenciais habitats de mosquito torna-se uma tarefa importante que poderá servir para que se adotem medidas de mitigação dos efeitos das alterações no uso/ocupação do solo, na criação de áreas suscetíveis à

instalação do vetor. Em consequência desse mapeamento pode ocorrer uma diminuição do risco da exposição humana. Este tipo de estudos, de carácter multi-escala e multitemporal podem ser importantes para e os organismos competentes das nações e organizações supranacionais antecipar, prevenir e controlar as novas doenças infecciosas emergentes, em especial destaque, os *arbovirus*, para evitar crises sociais e económicas por via desta eventual nova realidade. Vários estudos referem a importância de utilizar métodos multiescalares para fazer face a complexidade dos fatores que potenciam doenças infecciosas (Garabed *et al.*, 2020; Garira, 2020, 2018; Childs *et al.*, 2019).

As técnicas e métodos utilizados neste estudo fazem parte de um *workflow* de OBIA (*Object Based Image Analysis*) combinado com a extração multitemporal de vegetação. As características de *software*, na sua maioria em acesso aberto e gratuito serão descritos na ótica de os dar a conhecer e algumas expressões e dados técnicos destes *softwares* poderão ser alvo de uma tradução livre sempre que possível para melhor integração no corpo do texto.

A estrutura desta dissertação organiza-se em cinco capítulos que podem ser esquematizados em três fases (figura 2). Serão abordados os temas relativos aos padrões comportamentais e geográficos do vetor *Aedes albopictus*, a teoria, as técnicas e os dados numa primeira instância, seguidos das componentes práticas de correção de imagem e a posterior transformação em informação geográfica relevante.

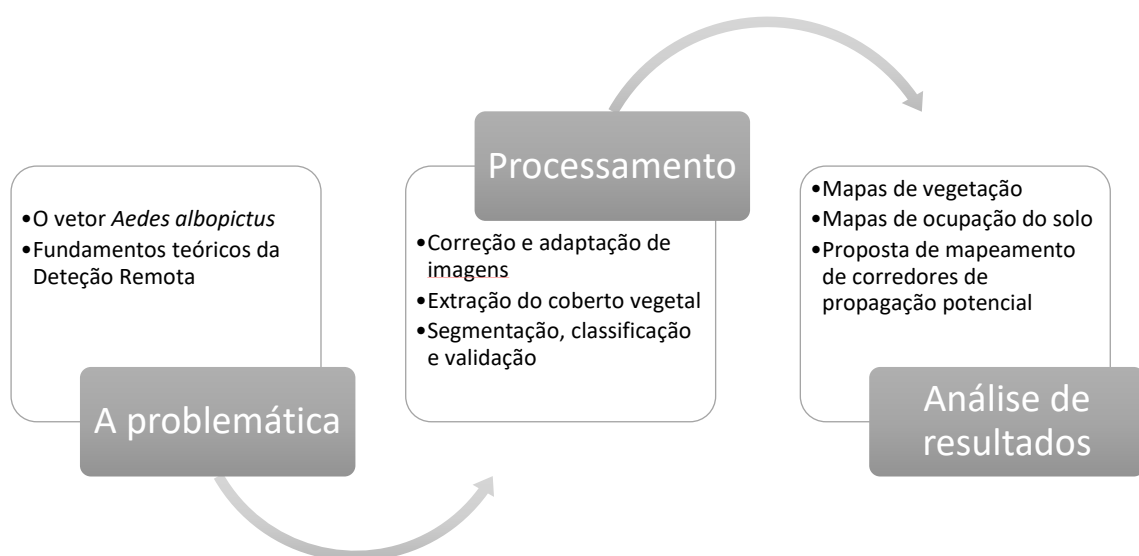


Figura 2 – Esquema da estrutura da dissertação.

Esta estrutura baseia-se na identificação e descrição do caso de estudo - o vetor *Aedes albopictus* - os seus padrões de distribuição e potenciais habitats, recorrendo a bibliografia. Este passo é essencial para perceber o que procurar e que informação extrair das imagens de satélite.

No primeiro passo de processamento foi extraído o coberto vegetal em multitemporal, sobre a qual se fez a análise da relação entre o vetor *Aedes albopictus* e a presença de vegetação, e também porque o índice de vegetação serviu como complemento aos dados de entrada para a classificação de objetos de forma a aumentar a separabilidade das classes. O segundo passo do processamento foi a criação dos mapas de ocupação do solo, seguindo um workflow de OBIA, desde a segmentação até ao produto final, o qual foi validado e melhorado. Por fim, foi feita a análise e proposta de corredores de propagação potencial, aliando os resultados obtidos anteriormente com dados vetoriais pertinentes que poderão servir de previsor de áreas suscetíveis à introdução e instalação de *Aedes albopictus*.

CAPÍTULO I

O vetor *Aedes albopictus* e Detecção Remota

Este capítulo dedica-se ao aprofundamento do conhecimento sobre o Aedes albopictus bem como a apresentação dos conceitos e dos fundamentos necessários à compreensão dos dados, das técnicas e dos métodos utilizados para chegar à resultados obtidos no capítulo VI.

CAPÍTULO I – O vetor *Aedes albopictus* e Detecção Remota

1. Vetor *Aedes albopictus*

O mosquito *Aedes albopictus* é originário das regiões tropicais do sudeste asiático. Esta espécie é conhecida como Mosquito Tigre Asiático pelas suas características (pretos com faixas brancas). O mosquito tem quatro estágios no seu ciclo de vida, ovos, larva, pupa e adulto. O risco de transmissão de doenças pelo *Aedes albopictus* é impulsionado pelas condições climáticas, sendo a temperatura ideal para transmissão situada entre 19,9-29,4°C. Este intervalo de temperatura significa que a sua atividade em regiões temperadas tenha um carácter sazonal (Ryan *et al.*, 2019), mas também em regiões tropicais e subtropicais quando o limiar superior do intervalo de temperatura é superado.

Este mosquito agressivo é ativo durante o dia, alimenta-se de uma vasta gama de hospedeiros incluindo gado, anfíbios, répteis, pássaros e humanos. Isto é especialmente problemático pela capacidade desta espécie em ser vetor de muitas doenças que afetam o ser humano como a febre amarela, a dengue, o Zika e chicungunha.

«Esta espécie mostra uma grande variabilidade genética, variabilidade fisiológica e capacidades de adaptação ecológica. Estas características explicam bem a sua rápida colonização de novos habitats no Hemisfério Norte a partir das suas regiões originais, através da transmissão mundial das suas fases imaturas em contentores artificial, especialmente em pneus usados» (Roiz *et al.*, 2008, citado de Hawley 1988)

Os seus habitats são buracos de árvores, e os seus locais de reprodução na natureza são pequenos e restritos, como corpos de água cercados por vegetação que forneça sombra, portanto, geralmente habita áreas rurais vegetadas. No entanto, a sua flexibilidade ecológica permite colonizar áreas urbanas e muitos tipos de sítios artificiais. Pode reproduzir-se em vasos de flores de cemitério ou ornamentais, poças permanentes ou pequenos lagos ornamentais, latas de refrigerante e recipientes abandonados. Os pneus são particularmente úteis para a reprodução do mosquito porque são

armazenados frequentemente ao ar livre e coletam e retêm a água da chuva por muito tempo. A adição de folhas em decomposição das árvores vizinhas produz condições químicas semelhantes aos buracos de árvores, o que fornece excelentes condições para a sua reprodução. (GISD², 2020).

A sua distância de voo é apenas 200m, pelo que a sua dispersão é geralmente feita através de transporte não intencional em viagens ou transporte de mercadorias. A sua abundância depende da altura do ano e de fatores ambientais como temperatura, luz e disponibilidade de água. Nas áreas de onde são originários, as populações adultas permanecem ativas durante todo o ano, porém, nos climas temperados, os mosquitos demonstram capacidade de “hibernar” no inverno sob forma de ovos e com isso se adaptar e instalar-se em latitudes temperadas. Contudo, é quando as temperaturas são elevadas que o desenvolvimento larvar se dá mais eficientemente, o que resulta num maior número de adultos. (DGS³, 2020)

2. História e evolução da dispersão do vetor no mundo e na Europa

O mosquito *Aedes albopictus* passou por uma dramática expansão global facilitada por atividades humanas, em particular o movimento de pneus usados e outros objetos que possam conter ovos. Aliado a isto, o trânsito passivo através de transportes públicos e privados resultou numa distribuição global generalizada. *Aedes albopictus* está agora listado como uma das 100 principais espécies invasoras pelo Grupo de Especialistas em Espécies Invasoras. (GISD, 2021).

² Estabelecido em 1994, o Grupo especialista em espécies invasoras (ISSG) é uma rede global de especialistas científicos e políticos sobre espécies invasoras que produzem o *Global Invasive Species Database* (GISD). Em <http://www.iucngisd.org/gisd/>.

³ Direção geral de Saúde, Informação geral sobre o mosquito *Aedes albopictus*, em <https://www.dgs.pt/paginas-de-sistema/saude-de-a-a-z/aedes-albopictus-mosquito.aspx>.

2.1. História da dispersão da espécie pelo mundo

Aedes albopictus é originalmente uma espécie zoofílica de florestas da Ásia, que se difundiu para ilhas nos oceanos Índico e Pacífico. Durante a década de 1980 expandiu rapidamente a sua distribuição para a Europa, Estados Unidos e Brasil (Kraemer *et al.*, 2015, citado de Delatte *et al.*, 2009 e Medlock *et al.*, 2012). As ocorrências de *Aedes albopictus* estão agora difundidas por vários continentes, resultantes de «uma expansão de alto alcance em curso alimentada pelo aumento do comércio global e viagens» (Kraemer, *et al.* 2015, p.1) bem como a alta adaptabilidade desta espécie.

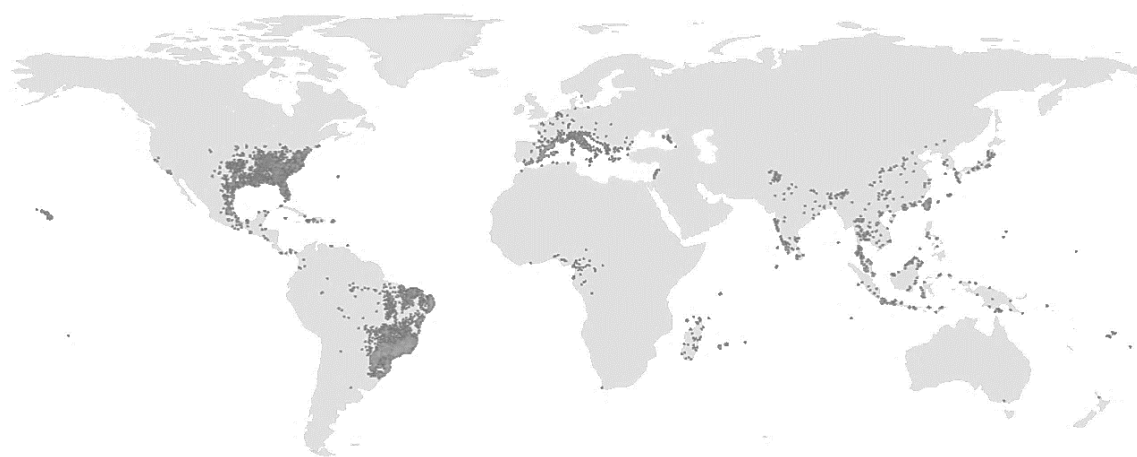


Figura 3 – Distribuição global de ocorrências de *Aedes albopictus* 1990 - 2020 (adaptado de GBIF⁴, 2020)

O *Aedes albopictus* pode ser encontrado em todos os tipos de ocupação, urbano, suburbano, rural bem como terras agrícolas ou florestas profundas (Manorenjitha & Zairi, 2015, citado de Hawley, 1988 e Rao, 1967). A capacidade de prosperar em qualquer condição causa dificuldade no seu controlo por parte das entidades competentes. Assim, as abordagens para o seu controlo devem ser revistas para reduzir eficazmente a população de vetores (Manorenjitha & Zairi, 2015).

⁴ O Sistema Global de Informação sobre Biodiversidade (GBIF) é uma organização internacional dedicada à disponibilização de dados científicos de biodiversidade por intermédio da Internet utilizando web services. A sua missão é facilitar o acesso livre e livre a dados de biodiversidade mundiais, para promover o desenvolvimento sustentável. Em <https://www.gbif.org/pt/>.

É esta adaptabilidade que torna o *Aedes albopictus* numa espécie invasora ameaçadora para vários locais do planeta, como pode ser observado pelo mapa da figura 4 (Kamal *et al.* 2018), que utilizou um algoritmo de máxima entropia para prever a distribuição potencial de *Aedes albopictus* sobre fatores bioclimáticos presentes, para tal foram utilizados modelos de ocorrência baseados em registos e variáveis ambientais. É possível notar de acordo com o resultado que esta espécie tem potencial para se instalar numa grande diversidade de climas.

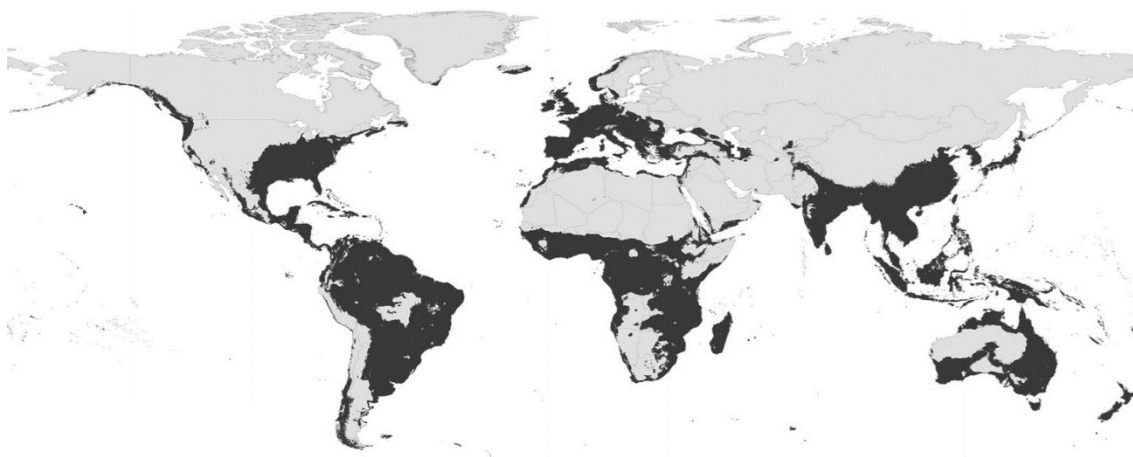


Figura 4 – Distribuição potencial baseada em fatores biofísicos para o ano de 2018 (adaptado de Kamal *et al.* 2018)

Estudos prospetivos da rede científica multidisciplinar EVITAR sobre as características climáticas para a fixação do mosquito tigre selecionaram «Especificamente, áreas que receberam uma precipitação superior a 500 mm, distribuídos por mais de 60 dias por ano e com temperaturas médias acima de 11 °C» (Roiz *et al.*, 2007, pp. 523-524 citado de Eritja *et al.*, 2005). Devido à forte adaptabilidade do *Aedes albopictus* «sempre foi levado em consideração que a espécie pode ser estabelecida localmente em áreas com menor precipitação, especialmente em áreas periurbanas onde ocorre um abastecimento suplementar de água a partir de atividades humanas.» (Roiz *et al.*, 2007, p. 524).

As alterações climáticas e o aumento da temperatura média da Terra terão certamente o seu efeito na distribuição de *Aedes albopictus*. «A distribuição global de *Ae. albopictus* mudou significativamente ao longo das últimas décadas e as indicações

atuais sugerem que é provável que isso continue em cenários climáticos futuros» (Tibatá, Salaman & Graham, 2013, p.238)

Este rearranjo pode levar a que as espécies passem a estabelecer-se em novas áreas, especialmente na direção dos polos ou em altitude, como consequência desta deslocação potenciada pelo comércio mundial e o transporte acidental de indivíduos, bem como, e muito importante, as alterações nos usos do solo, invasões de espécies não nativas serão mais frequentes (Guo, Lenoir & Bonebrake, 2018; Sirami *et al.* 2016; Putten, Macel & Visser, 2010)

Como espécie alvo neste estudo, *Aedes albopictus* é um excelente exemplo de redistribuição geográfica de espécies e invasão de novas áreas geográficas situadas em zonas climáticas distintas. Dado que esta espécie se pode estabelecer em intervalos de temperatura que variam dos 10 aos 25° (Fischer *et al.* 2011), pode-se esperar a sua expansão geográfica devido ao aquecimento global, especialmente para as regiões temperadas. Com as alterações climáticas espera-se um risco mais sazonal em todas as regiões sendo que os maiores aumentos de risco de transmissão de *albopictus* ocorram em cenários intermédios das alterações climáticas (Ryan *et al.*, 2019). Por isto é necessário, preparar medidas de monitorização e controlo de vetores, em áreas suscetíveis, e em fases do ano em que o vetor está ativo.

Apesar de geralmente ser uma espécie de exterior, novos estudos descobriram sinais de que se poderá estar a adaptar a uma vida de interior, procurando alimento e descanso dentro de habitações (Delatte *et al.*, 2010), ao passo que outros estudos demonstraram também sinais da sua capacidade de se reproduzir em áreas interiores, o que pode ter implicações significativas no seu comportamento, especialmente nas áreas urbanas (Dieng *et al.*, 2010), onde isto poderá tornar os esforços de contenção mais difíceis em áreas afetadas.

Segundo Fisher *et al.*, 2011, no que diz respeito ao envelope bioclimático, as ocorrências concentram-se sobretudo em áreas com 500 mm de precipitação média anual e temperaturas médias anuais superiores a 10 °C para populações nativas e invasoras. As populações invasoras distribuíram-se em territórios com médias anuais de precipitação na sua maioria entre 10 e 25 °C e com uma variação anual de precipitação

entre 500 mm e 2000 mm. Enquanto que as populações nativas se caracterizam por ocupar territórios com temperaturas e precipitações ligeiramente mais elevadas quando comparados com as populações invasoras. A Europa destaca-se pelos valores mais baixos de temperatura e precipitação anual. A figura 5 mostra que as que o mosquito tem um envelope bioclimático com temperaturas médias anuais mais elevadas nas áreas onde é nativo com uma temperatura média de 23,7C, a área de invasão excluindo a Europa tem uma temperatura média menor, 21,7C, e a temperatura média mínima onde a espécie se instalou é na europa com 13,8C. Quanto à precipitação a tendência é a mesma com 2028mm de precipitação média anual para os nativos, 1392mm para os invasores excluindo a europa e de apenas 831mm de precipitação anual para a Europa.

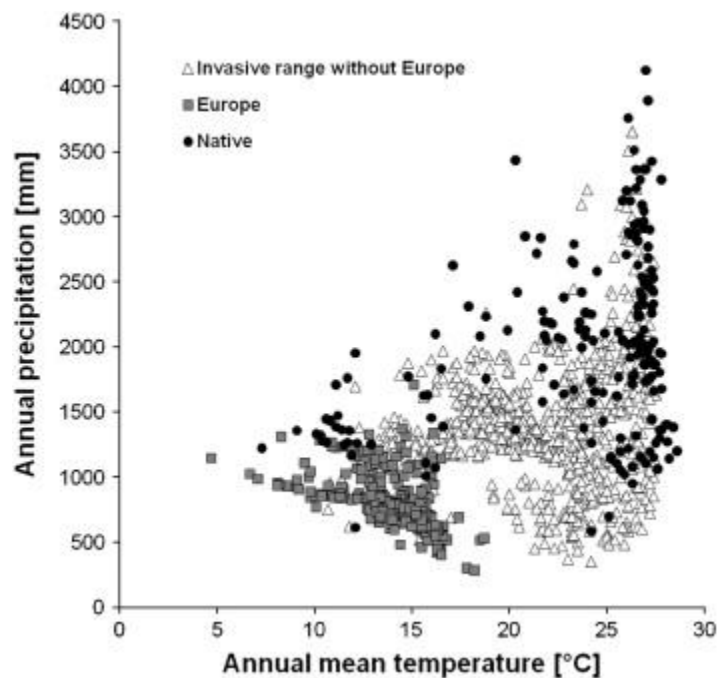


Figura 5 - Envelope bioclimático para *Aedes albopictus* (gama nativa e invasiva), derivado por sobreposição geograficamente explícita de registos de presença com variáveis bioclimáticas anuais (Fisher *et al.* 2011).

Explicações para este comportamento podem ser ligadas ao comportamento de *Aedes albopictus* em relação à temperatura e precipitação. Resultados em laboratório demonstram que em temperaturas mais elevadas, a precipitação tem maior efeito na sobrevivência dos ovos, isto é, quanto mais elevada a temperatura, maior é a necessidade de os contentores de água onde estão os ovos não secarem para se produzirem adultos, enquanto que em ambientes mais frios, a precipitação tem um

contributo menos expressivo para a variação de produção de mosquitos adultos (Alto & Juliano, 2001).

2.2. Evolução da dispersão do vetor na Europa

A Europa tem diversos tipos de clima, sendo estes na sua maior extensão temperados, o que significa que o risco de aumento de atividade de *Aedes albopictus* com as alterações climáticas é real. A espécie já se encontra instalada em vários países da Europa, especialmente nas áreas de clima temperado mediterrânico onde a invasão de *Aedes albopictus* tem sido mais bem sucedida. Diversos estudos têm vindo a ser conduzidos de forma a estimar o impacto e a evolução da sua expansão (Aranda, Eritja & Roiz, 2006; Neteler *et al.*, 2011; Chen, Wang & Bjørnstad 1996; Edes *et al.*, 2007; Caminade *et al.*, 2012).

O Centro Europeu para controlo e prevenção de doenças (ECDC) produz, em conjunto com a Autoridade Europeia de Segurança Alimentar (EFSA)⁵, mapas de distribuição do mosquito na Europa. Estes permitem observar a distribuição, bem como o estado da espécie para todo o continente europeu.

Os dois mapas seguintes (figuras 6 e 7), representam a distribuição de *Aedes albopictus* no continente europeu, a sua legenda tem seis itens, estabelecido, introduzido, ausente, sem dados e desconhecido. Deste mapa interessam as duas primeiras classes da legenda, introduzido e estabelecido, que significam no primeiro caso que o mosquito encontra caminho para estas áreas e há risco de se estabelecer, que significa que já tem uma população estável nessa área. Quando a invasão não sucede em criar uma população estável, e a espécie é extinta nessa área, considera-se que está ausente (Takken & Berg, 2019).

⁵ A Autoridade Europeia para a Segurança Alimentar (EFSA) é um organismo Europeu criado para dar à Comissão Europeia e ao público pareceres científicos independentes sobre a segurança alimentar e os riscos possíveis na cadeia alimentar. Em <https://www.efsa.europa.eu/>.

O mapa da figura 6 mostra uma distribuição concentrada maioritariamente na bacia do Mediterrâneo, tendo a Itália todas as suas regiões com populações estabelecidas de *Aedes albopictus*, bem como regiões no sul da França o litoral mediterrânico espanhol, os países dos Balcãs, no leste a costa da Rússia no Mar Negro, a Geórgia e a Turquia, bem como no médio oriente, Israel, Palestina, Síria e Líbano e Jordão. A espécie tinha sido, na primeira data, introduzida em regiões da Espanha mais afastadas do litoral Mediterrânico e em áreas mais a norte de França, na Suíça, Alemanha, Bélgica, Países Baixos, Áustria, República Checa, Eslováquia. Continua ainda a estabelecer-se em novas áreas nos Balcãs, Hungria, Sérvia, Roménia, Bulgária e Grécia.

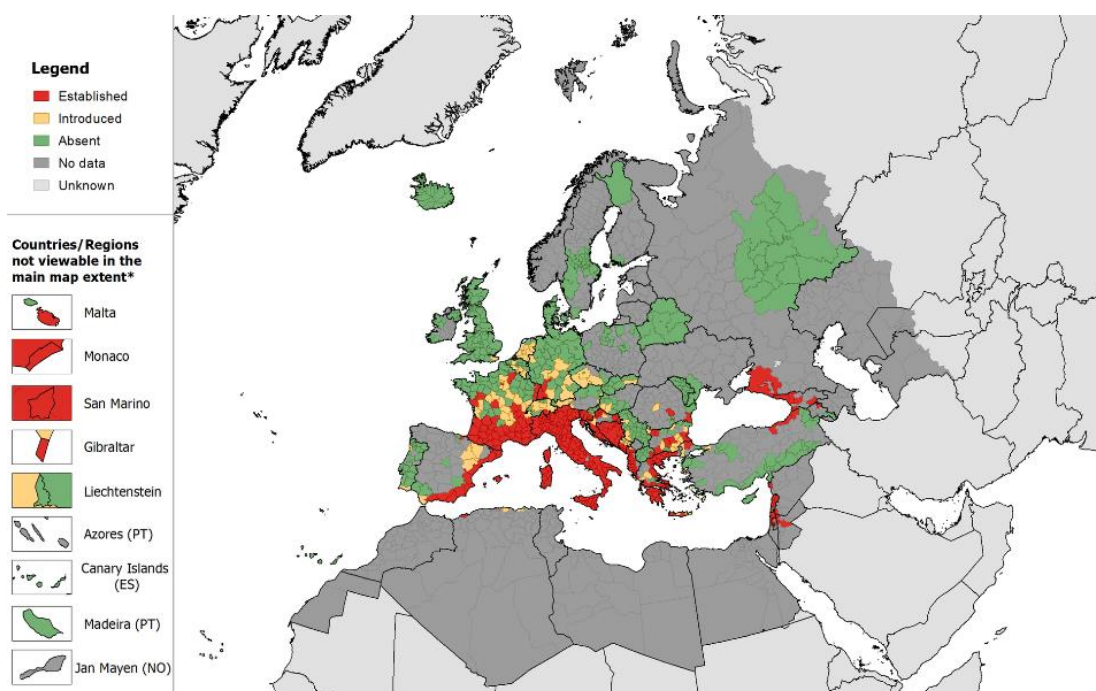


Figura 6 - Mapa da distribuição conhecida de *Aedes albopictus* na Europa a nível regional, agosto de 2019 (ECDC & EFSA, 2019).

O mapa da figura 7 permite confirmar o avanço da invasão da espécie, com apenas 9 meses de intervalo: o estabelecimento de populações em mais áreas na França, e no centro da Europa, como Alemanha e Suíça, e a Eslovénia e a Albânia passaram a ter a totalidade das suas regiões com populações estabelecidas. Nos territórios insulares não se observaram alterações, o segundo mapa contempla informação de que a espécie tenha sido introduzida na Tunísia, porém encontra-se ausente. O *Aedes albopictus* foi observado em praticamente todas as regiões biogeográficas da Europa, incluindo observações na Islândia (Ártico), porém estas

introduções não deram origem a populações estabelecidas, encontrando-se ausente. Portugal aparece no segundo mapa com a região do Algarve e a comunidade intermunicipal de Tâmega e Sousa com populações estabelecidas de *Aedes albopictus*.

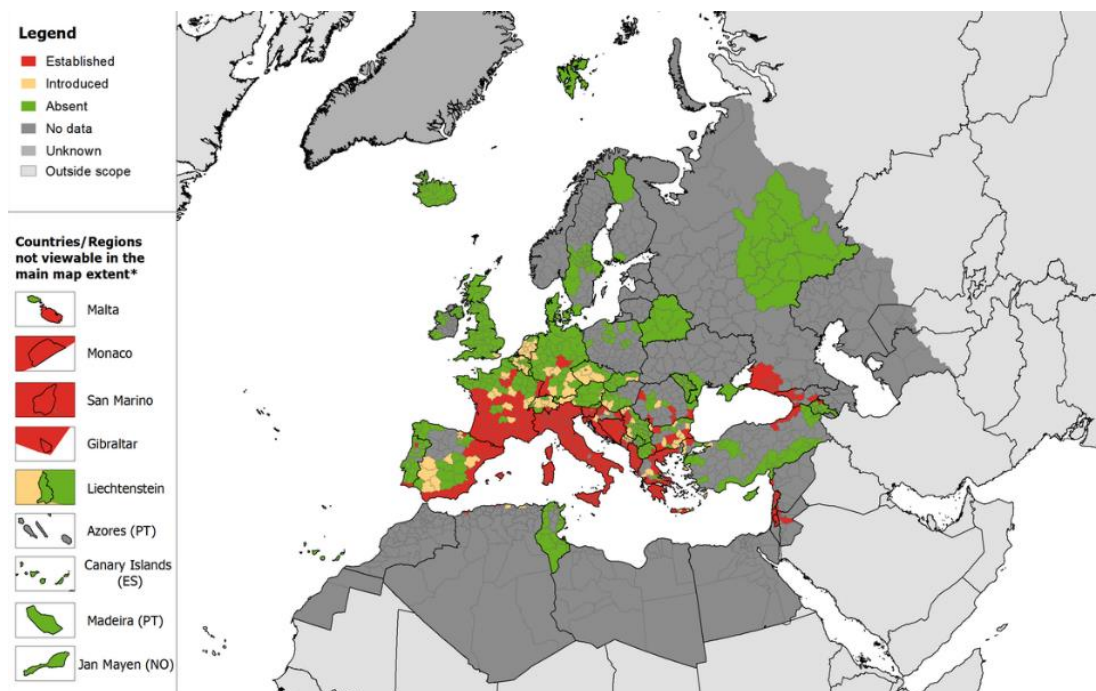


Figura 7 - Mapa da distribuição conhecida de *Aedes albopictus* na Europa a nível regional, maio de 2020 (ECDC & EFSA, 2020).

Apesar de normalmente se reproduzir em áreas de transição de corpos de água ou em buracos nas árvores, a ausência destes não tem sido um impasse para o estabelecimento em novos lugares, dado que é possível colonizarem qualquer recipiente que possa conter água (Caminade *et al.*, 2012). Esta característica torna o mosquito particularmente resiliente do ponto de vista de habitat, aliada à aclimatização que já se verifica nestes mosquitos ao clima europeu.

3. A Detecção Remota ao serviço da monitorização do vetor

O uso/ocupação do solo é uma variável central na modelação de habitats potenciais para *Aedes albopictus*, pela facilidade que o vetor tem em encontrar habitats em áreas artificializadas, mesmo onde a precipitação é escassa. Com base nestes aspetos, e para modelar os habitats, recorreu-se a técnicas baseadas em dados obtidos

por Detecção Remota enquadradas neste tópico, para obter o uso e a ocupação do solo, bem como as variáveis Vegetação e Água.

Os algoritmos de processamento digital de imagens têm a capacidade de extrair largas quantidades de informação relativamente a aspetos biofísicos, e têm por isso sido bastante utilizados para monitorizar e prever condições ideais de reprodução e atividade de vetores. «O mapeamento geoespacial, utilizando a Detecção Remota, oferece o potencial de identificar habitats de larvas numa grande área a um grau que seja difícil ou impossível usando o levantamento convencional do solo» (Zou, Miller & Schmidtman, 2006, p.1034, citado de Hayes *et al.*, 1985; Washino & Wood, 1994; Dale *et al.*, 1998; Hay *et al.*, 1998).

A Detecção Remota consiste numa forma de obtenção de dados geográficos através do processamento de imagens captadas por sensores, sem a presença física ou contacto com o terreno. Os sistemas de Detecção Remota passivos registam a parte refletida do espectro eletromagnético que atinge a superfície da Terra; os sistemas ativos são os dispõem da sua própria fonte de energia.

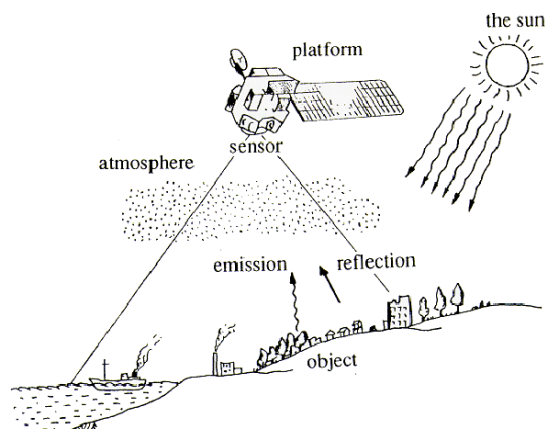


Figura 8 - Processo de Detecção Remota passiva (Muralli, 2015, p. 220).

A radiação eletromagnética é uma forma de energia com propriedades de onda em que a sua maior fonte é o sol. O sol emite energia que viaja na forma de ondas à velocidade da luz, também conhecido como espectro eletromagnético, este propaga-se através do tempo e do espaço numa forma parecida a ondas de água, estas têm também oscilação de forma perpendicular à sua direção de viagem para todas as direções. As ondas do espectro eletromagnético podem ser caracterizadas pela sua frequência e pelo comprimento. O comprimento de onda λ é a distância entre dois picos

da onda e a frequência μ é o número de oscilações completadas por segundo, estas duas são relacionados pela equação $C = \lambda \times \mu$. (Weng, 2010) .

Os sensores captam parte deste espectro, no caso da Detecção Remota passiva a energia do sol que é refletida pela superfície. Esta é captada por um sensor na forma de $\text{watts/m}^2/\text{str}/\mu\text{m}$, isto é, capta a energia refletida em watts por metro quadrado para um determinado comprimento de onda, que confere ao pixel o seu nível radiométrico. A intensidade energética captada pelo sensor, explica-se pelo primeiro princípio da termodinâmica (conservação de energia) em que $E_i(\lambda)$ (energia incidente) é igual à Energia refletida $E_R(\lambda)$ + energia transmitida $E_T(\lambda)$, mais energia absorvida $E_A(\lambda)$.

$$E_i(\lambda) = E_R(\lambda) + E_T(\lambda) + E_A(\lambda) = 1 \quad 1$$

Para obtenção de dados em Detecção Remota os sensores captam a energia refletida pelos objetos associando a cada objeto o seu nível de refletância espectral, que se dá pelo resultado da seguinte equação.

$$P = E_R(\lambda)/E_i(\lambda) \quad 2$$

A reflectância espectral está na origem do nível radiométrico de um pixel, que é convertida de watts para código digital, resultando no nível digital ND, este nível digital nada mais é que o nível de reflectância espectral para um dado pixel em watts que foi codificado em bits. A codificação em bits baseia se na transformação em código binário que representa então a energia captada pelo sensor em $\text{watts/m}^2/\text{str}/\mu\text{m}$, as imagens podem ser codificadas em 2^n bits como por exemplo 2^8 bits assim «se um sensor usa 8 bits para registar informação, estarão disponíveis $2^8 = 256$ níveis digitais, que variam de 0 a 255. Contudo, se apenas são usados 4 bits, então estarão disponíveis só $2^4 = 16$ níveis, que variam de 0 a 15, sendo a resolução radiométrica muito menor.» (Sousa & Silva, 2011, p. 22). Para as imagens *Sentinel 2* MSI, esta resolução é de 2^{12} bits, o que equivale a 4 095 níveis digitais.

As bandas espectrais são assim, camadas de informação codificadas em bits com uma estrutura matricial (linhas e colunas) cuja unidade mínima é o pixel. Cada banda espectral corresponde a uma faixa do espectro eletromagnético registada pelo

sensor, o número de bandas que uma imagem tem varia com a capacidade de cada sensor em captar a energia eletromagnética.

Define-se então uma imagem digital em DR como como fruto de um conjunto de técnicas e ferramentas que nos dão uma matriz de dados, que pode ser armazenada, processada e interpretada em computador, constituída por colunas (Y) e por linhas (X). Esta matriz tem, no caso das bandas *Sentinel 2* MSL L1c e L2A, o seu início no canto superior esquerdo. Este conjunto de linhas e colunas é chamado de matricial ou Raster tem como unidade mínima o pixel, cada um destes é estatisticamente tratado separadamente dos restantes, podendo qualquer pixel assumir uma diferente escala de brilho independentemente da sua localização na matriz.

Como diferentes materiais tem uma refletância diferente nas diferentes bandas do espectro electromagnético, o conhecimento desta teoria permite avançar para a seleção de técnicas de processamento digital de imagem e posterior extração de elementos presentes na superfície da terra nomeadamente a ocupação do solo e coberto vegetal que servirão para modelar habitats prováveis de *Aedes albopictus*.

O processamento de imagem consiste num conjunto de procedimentos, incluindo formatação e correção dos dados, melhoria de imagem, ou classificação automática ou semiautomática. Resumidamente, o processamento digital de imagem tem um conceito bastante simples, os dados originais são inseridos no *software* que aplica uma, ou uma série de equações de forma a produzir um resultado que depois é gravado. Este resultado consiste numa nova imagem que pode ser visualizada ou gravada para processamentos posteriores (Pinho, 2009, citado de Lillesand *et al.*, 2003).

PDI pode ser dividido em três grupos categóricos (Pinho, 2009, citado de Sabins, 1997):

- Restauração de imagem que inclui as operações de pré-processamento como as correções geométricas e radiométricas;
- Melhoria da qualidade visual ou melhoria de imagem, tem como objetivo apenas melhorar a aparência das imagens para ajudar na interpretação e análise visual.

- Extração de informação, que compreende operações de transformação e operações de classificação e análise.

As funções de pré-processamento envolvem as operações que são normalmente necessárias antes da análise e extração de dados principais e são geralmente agrupadas como correções radiométricas ou geométricas.

As correções radiométricas incluem a correção dos dados para irregularidades sensoriais e ruído sonoro ou atmosférico indesejado e a conversão dos dados de modo a que representem com precisão a radiação refletida ou emitida medida pelo sensor.

As correções geométricas incluem a correção de distorções geométricas devido a variações de geometria sensor-Terra, e a conversão dos dados para coordenadas do mundo real na superfície da Terra (Natural Resources Canada, 2020).

As transformações de imagem são operações que envolvem o processamento combinado de dados de várias bandas espectrais. A realização de operações aritméticas permite combinar e transformar as bandas originais em imagens "novas" que melhor exibem ou realçam certas características da superfície (Natural Resources Canada, 2020). Nestas transformações de imagem podem ser incluídos os índices de estado de superfície como o NDVI que já por si permitem extrair informação de uma imagem realçando o diferencial entre a fraca refletância na banda do vermelho da clorofila e a sua elevada refletância na banda do infravermelho próximo.

Por fim a classificação de dados em Deteção Remota e processamento digital de imagem consiste na transformação de dados processados (imagens) em produtos (imagens ou vetores) que permitem a sua representação através de mapas temáticos, atribuindo classes a determinados objetos (Semântica). Pode-se definir esta etapa como a tarefa de extrair classes de informação de uma imagem em que os resultados da classificação de imagens podem ser utilizados para a geração de mapas temáticos. Esta classificação pode ser supervisionada (Semiautomática) ou não supervisionada (Automática) (ESRI, 2020).

mais frio, com uma temperatura média de 9,2 °C, e agosto o mais quente, com 24 °C). A precipitação atinge aproximadamente 588 mm médios por ano (AEMET, 2020).

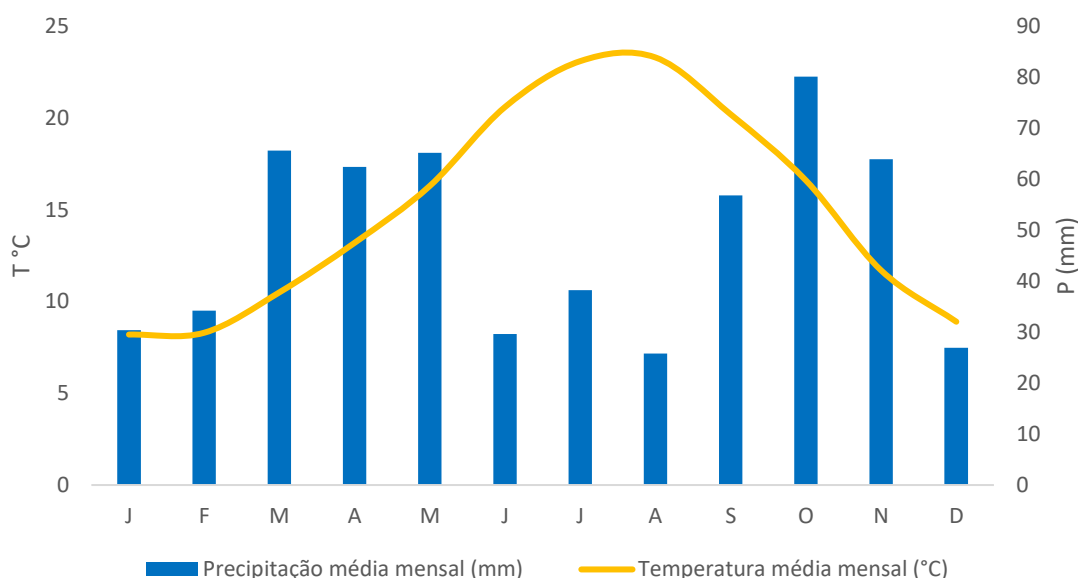


Figura 10 – Gráfico Termopluviométrico para as normais climatológicas do observatório Fabra (período de referência 2007-2016) – (Servei meteorològic de Catalunya⁶, 2017)

A ocupação do solo predominante é o urbano contínuo, seguido do urbano descontínuo e áreas industriais/comerciais; as áreas portuárias também assumem algum destaque. São de notar alguns parques urbanos de dimensões consideráveis situados dentro da cidade. Nos municípios mais afastados da urbe observam-se ainda áreas naturais, como florestas de coníferas e vegetação esclerófila.

A escolha da área de Barcelona serve de teste para encontrar, em função de estudos anteriores feitos sobre a distribuição do vetor, as áreas de habitat prováveis, podendo confrontar os resultados com as ocorrências de *Aedes albopictus*.

Os dados das ocorrências foram recolhidos utilizando ciência voluntária, que consiste em utilizar programas voluntários para recolher grandes quantidades de dados de distribuição de espécies que permitam colmatar as lacunas de informação existentes. Estes projetos podem recolher dados de forma semelhante às sondagens científicas, se

⁶ Normais climatológicas recentes do Serviço meteorològic de Catalunya. *Servei Meteorològic de Catalunya* em <https://www.meteo.cat/wpweb/climatologia/serveis-i-dades-climatiques/normals-climatiques-recents/>.

os protocolos forem suficientemente rigorosos, a principal diferença entre estes dois tipos de recolha de informação, especialmente no que diz respeito à informação geográfica (VGI – *Voluntary Geographic Information*), diz respeito à falta de experiência geral dos cidadãos voluntários que podem levar a erros de identificação ou erros espaciais, que podem reduzir a precisão dos dados (Tiago, Pereira & Capinha, 2017).

A plataforma de onde foram extraídos os dados foi o portal do GBIF. Estes dados foram então georreferenciados, produzindo uma nuvem de pontos com a qual se analisaram possíveis relações com as diferentes variáveis. Na figura 11 encontram-se representados os pontos das ocorrências onde se pode ver a distribuição espacial na área de estudo, sobre uma composição colorida de cor verdadeira de uma imagem Sentinel-2-MSI, sobre a qual se encontram sobrepostos os pontos de ocorrências de *Aedes albopictus* divididas em dois grupos, confirmados e prováveis, separados pelos especialistas que analisaram e validaram as ocorrências enviadas pela população.

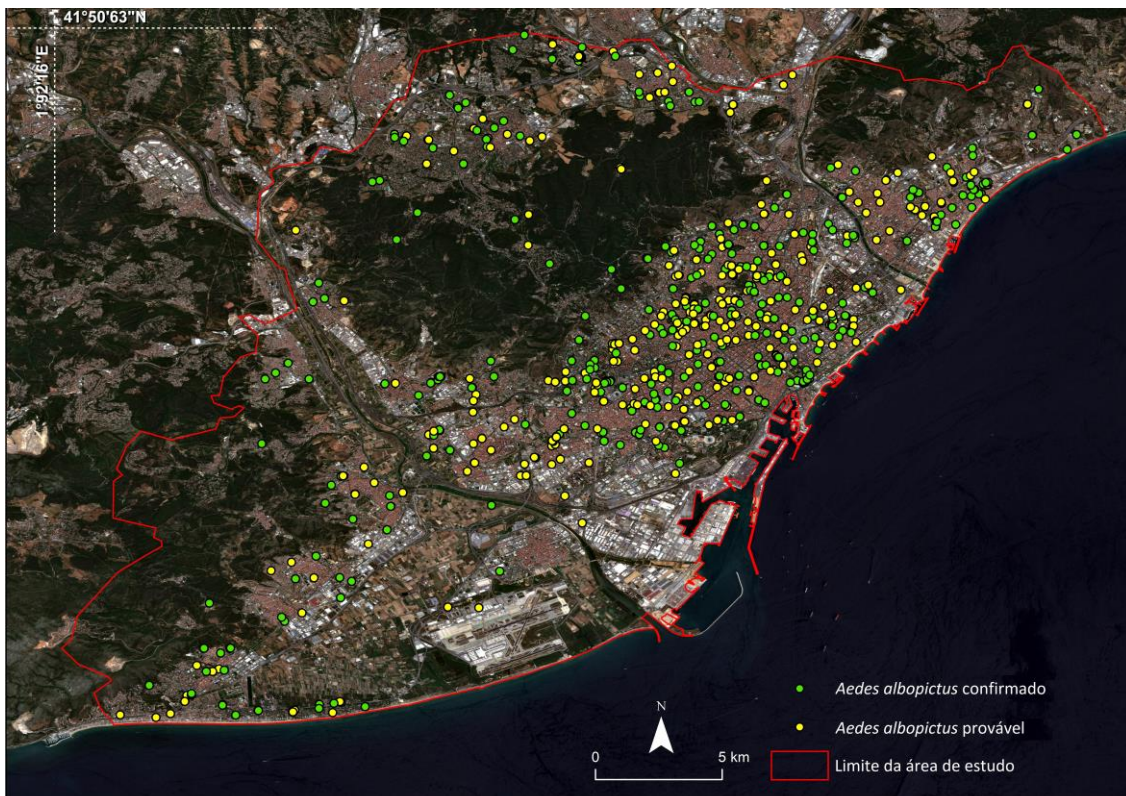


Figura 11 – Contextualização das ocorrências dentro da área de estudo, sobre composição colorida (cor verdadeira) Sentinel-2 MSI.

Neste contexto, estes dados são relevantes para compreender a distribuição das ocorrências e a sua relação com o espaço. Estes registos também trazem a vantagem de cobrir grandes manchas do território e ter um grande número de

observações realizadas e amplamente distribuídas no espaço. Para reduzir o erro associado à identificação errada, os especialistas analisam as fotos tiradas pelos cidadãos para confirmar se, de facto, o seu *Aedes albopictus*, neste estudo foram separados os dados entre ocorrências confirmadas por especialistas e ocorrências prováveis, ambos foram utilizados nas estatísticas espaciais para avaliar a sua relação com a vegetação.

No entanto, a precisão espacial nunca é concedida, pois existe sempre a possibilidade de identificação num local e o carregamento ser feito a partir de outro. Estes registos serão utilizados para medir as áreas onde foram observadas ocorrências de *Aedes albopictus* na zona de Barcelona e arredores e compará-las estatisticamente com a cobertura e vigor da vegetação.

CAPÍTULO II

A preparação de dados Sentinel-2

MSI

Este capítulo dedica-se a uma descrição técnica dos dados Sentinel-2 MSI e das respetivas operações realizadas para as correções atmosféricas para uso em multitemporal bem como as operações geométricas de âmbito geográfico e de uniformização de resoluções entre bandas.

Capítulo II – A preparação de dados *Sentinel-2* MSI

1. Os dados

Os dados foram retirados do portal de acesso aberto *Copernicus*⁷. Nesta primeira começou o pré processamento, tendo logo de partida em atenção alguns aspetos importantes tais como a obstrução da área de estudo por presença de nuvens e o nível de processamento que as imagens descarregadas já têm associado.

Os sensores MSI (*Multispectral Instrument*) a bordo das missões *Sentinel-2 A* e *Sentinel-2 B*, lançados respetivamente a 23 de junho de 2015 e a 7 de março de 2017, serão a fonte primária de dados para este estudo. A plataforma para o descarregamento dos dados foi o *scihub (Copernicus Open Access hub)*, por esta ser uma plataforma dedicada aos dados *Sentinel-2* MSI e oferecer produtos com diferentes níveis de processamento. «Com as suas 13 bandas espectrais, largura de faixa de 290 km e alta frequência de revisita, instrumento MSI do *Sentinel-2* apoia uma ampla gama de estudos de superfície e programas, e reduz o tempo necessário para criar um arquivo de imagens sem nuvens para a Europa. As bandas espectrais do *Sentinel-2* fornecem dados para classificação/mudanças da ocupação do solo, correções atmosféricas e separação nuvens/ neve.» (SUHET, 2015, p.16).

A disponibilidade de produtos com bom desempenho na qualidade dos dados (tanto em termos de precisão quanto em radiometria e geometria) tem uma importância primordial para muitas aplicações. Este é, de fato, um fator facilitador essencial para uma exploração mais fácil de séries temporais, comparações de medições de diferentes sensores ou deteção de mudanças na paisagem (Gascon, *et al.*, 2017).

⁷ Centro de acesso aberto (*Open Access Hub*) de dados *Copernicus*. Este serviço oferece acesso completo, gratuito e aberto aos produtos de utilizador das missões *Sentinel* em <https://scihub.copernicus.eu/dhus/#/home>.

Na tabela 1, estão sumarizadas as características espectrais de cada uma das 13 bandas em ambas as plataformas *Sentinel-2* e as respectivas resoluções espaciais.

Tabela 1 - Bandas espectrais do sensor MSI a bordo da missão *Sentinel-2* (S2A & S2B) (ESA, 2015).

Nome da banda	Resolução espacial (m)	<i>Sentinel-2A</i>		<i>Sentinel-2B</i>	
		Comprimento de onda central (nm)	Largura de banda (nm)	Comprimento de onda central (nm)	Largura de banda (nm)
Aerosol	60	442.7	21	442.2	21
Azul	10	492.4	66	492.1	66
Verde	10	559.8	36	559.0	36
Vermelho	10	664.6	31	664.9	31
<i>Red Edge 1</i>	20	704.1	15	703.8	16
<i>Red Edge 2</i>	20	740.5	15	739.1	15
<i>Red Edge 3</i>	20	782.8	20	779.7	20
Próximo Infravermelho	10	832.8	106	832.9	106
<i>Red Edge 4</i>	20	864.7	21	864.0	22
Vapor de água	60	945.1	20	943.2	21
Cirros	60	1373.5	31	1376.9	30
SWIR 1	20	1613.7	91	1610.4	94
SWIR 2	20	2202.4	175	2185.7	185

Um dos constrangimentos, que tem sido apontado em diversos estudos, é relativo à utilização de dados *Sentinel* em OBIA, pois este método é tradicionalmente utilizado em imagens de alta resolução (como IKONOS, QuickBird ou Worldview), ao passo que para imagens de média ou baixa resolução é preferida a classificação baseada em pixels (Chen *et al.*, 2018). Os dados *Sentinel*, apesar de serem - comparativamente a outras plataformas - dados de alta resolução, não o são tanto quando comparados com as plataformas tradicionalmente utilizadas em OBIA. No entanto, estudos anteriores demonstraram que este método quando aplicado a imagens *Sentinel-2* é capaz de produzir bons resultados ao nível da classificação, que ultrapassam os métodos de classificação pixel a pixel no que diz respeito à precisão temática (Belgiu & Csilik, 2018; Kaplan & Avdan, 2017; Tijssen, 2018).

O primeiro aspeto tido em conta, com base no trabalho desenvolvido no âmbito desta dissertação é o nível de processamento das imagens. Existem vários níveis disponíveis para descarregar a partir da plataforma Copernicus Open Access Hub ou pelo *software* SNAP, porém nem todos eles estão disponíveis para os utilizadores.

Aos produtos de nível 0 não é possível o acesso público em geral, este produto contém os dados em bruto que servem de base para gerar todos os outros produtos. O

nível 1A também não disponível ao público, resulta da descompressão dos dados, nesta etapa é desenvolvido o modelo geométrico, permitindo localizar qualquer pixel na imagem. (SUHET, 2015)

A partir do nível 1B, já está disponível o acesso público. Este produto contém a imagem com as devidas correções radiométricas como a correção de sinal escuro, correção de uniformidade de resposta dos pixels, correção de diafonia, identificação e correção de pixels com defeito, restauração de bandas de alta resolução espacial (deconvolução e ruído), armazenamento das bandas espectrais de 60 m). Este produto contém também os valores de radiância para o topo da atmosfera e da geometria captada pelo sensor. Adicionalmente, este produto inclui também geometria refinada, que foi utilizada para gerar os produtos de nível 1C. As coordenadas de cada pixel para os produtos 1A e 1B referem-se ao centro do pixel (SUHET, 2015).

Os produtos de nível 1C, contém imagens de 100 por 100km orto retificadas utilizando um modelo digital de elevação (DEM) para projetar a imagem com as respectivas coordenadas cartográficas. São também fornecidas medições radiométricas para as reflectâncias no topo da atmosfera com todos os parâmetros para as transformar em radiâncias. Os produtos de nível 1C também são reamostrados em resoluções espaciais constantes, dependentes da resolução nativa das diferentes bandas espectrais, de 10, 20 e 60m. As coordenadas dos pixels referem se ao seu canto superior esquerdo. Neste produto ainda vem incluídas máscaras para a água e nuvens, e dados ECMWF para a coluna total de ozono, vapor de água e pressão média no nível do mar (SUHET, 2015).

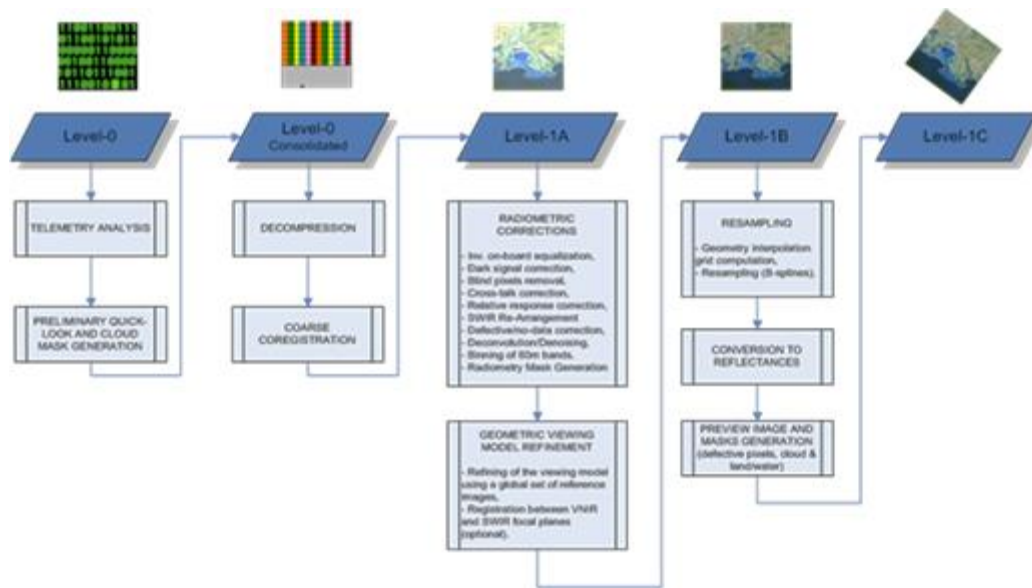


Figura 12 – Níveis de processamento de imagens Sentinel-2 MSI, desde o nível 0 até ao nível 1C (SUHET, 2015)

O último nível de processamento, e aquele que utilizado nesta dissertação para a extração de informação geográfica é o nível 2A, contém as refletâncias na base da atmosfera (BOA) derivadas do produto 1C de entrada (SUHET, 2015), pelo que este produto é em tudo igual ao anterior, exceto que é mais indicado para a extração de informação geográfica ao nível do solo que o anterior, pois a operação que o transforma realiza as necessárias correções atmosféricas descritas em seguida, este é o produto indicado para análises multitemporais, por mitigar o ruído atmosférico.

2. Pré-processamento dos dados

2.1. Correções atmosféricas

A correção atmosférica «de imagens de satélite antes da extração da vegetação é essencial para remover o ruído e aumentar a interpretabilidade dos dados da imagem. Isto é particularmente relevante quando uma série temporal de imagens é usada ou quando uma área é abrangida por muitas imagens, uma vez que é necessário tornar essas imagens compatíveis espacial e espectralmente» (Fatiha *et al.*, 2013, p.670).

Esta operação não foi realizada para a totalidade dos produtos por uma questão de poder de processamento, pelo que serão descarregados os produtos de nível 2A que estiverem disponíveis na plataforma Copernicus opens access hub. No entanto, e pelo facto de a sua produção não ser sistemática, nem todas as imagens a utilizar neste estudo poderão ser descarregadas já nesse formato, o que justifica a necessidade de as processar a partir do nível 1C.

Para a realização dos mapas de vegetação multitemporais, foi necessária a transformação dos produtos de nível 1C em produtos nível 2A utilizando o processador Sen2cor. Este *plug-in* para o *software* SNAP processa, gera e formata produtos *Sentinel-2* nível 2A a partir de produtos 1C, para tal, este realiza a correção atmosférica, de terreno e cirrus, sobre os dados de entrada.

No Sen2Cor para o processamento de dados do *Sentinel-2* para o nível 2A existem duas fases: a classificação de cena e posterior correção atmosférica descritos por Mueller-Wilm (2018).

Na classificação de cenas (CC) gera-se um mapa de classificação de cena e de dois indicadores de qualidade e, com base em algoritmos de deteção, são produzidos também um mapa de probabilidade de nuvens e de neve que se constituem como produtos adicionais e input necessário para a correção atmosférica.

O algoritmo para a classificação de cena deve ser adequado para detetar nuvens, neve e sombras de nuvens bem como gerar um mapa de classificação, que consiste em 4 classes diferentes para nuvens (incluindo cirros), juntamente com seis classificações diferentes para sombras, sombras de nuvens, vegetação, solos / desertos, água e neve. Os dados de entrada para esta parte são o nível Produto 1C, ou seja, reflectância TOA (*Top Of Atmosphere*) medida nas resoluções de 60/20 ou 10metros. (Richter, Louis & Berthelot, 2011). É nesta fase que se aplicam correções sobre a vegetação senescente, que apesar de ter um alto brilho, já tem uma atividade clorofilina reduzida, isto é feito correlacionando as bandas 3 (verde) e 8 (Próximo infravermelho) de forma a normalizar este tipo de vegetação, separando-a da vegetação saudável.

A correção atmosférica utiliza um conjunto de LUT (*look-up tables*) geradas via *libRadtran*. Este pacote de *software* contém ferramentas para cálculos de transferência radiativa na atmosfera terrestre, e através do programa *uvspec*, oferece resoluções para a equação de transmissão radiativa descrita, que permite adaptá-la às propriedades óticas da imagem (Mayer & Kylling, 2005).

Segundo os mesmos autores a equação para a transferência radiativa (*Radiative Transfer Equation* ou RTE) pode ser escrita da seguinte forma:

$$\frac{dL}{\beta ds} = -L + J \quad 3$$

Onde a função de origem J é definida como:

$$J = \frac{\omega}{4\pi} \int p(\Omega, \Omega') L(\Omega') d\Omega' + (1 - \omega) B(T) \quad 4$$

Nesta equação, L está para a radiância na localização (x, y, z), β é o coeficiente de extinção de volume, ω o albedo de dispersão único, $p(\Omega, \Omega')$ dá a função de fase que para a probabilidade de um evento de dispersão redistribuir radiação da direção Ω' para Ω , e B(T) a função de *Planck*. Ao contrário da maioria dos outros modelos de transferência radiativa, o modelo de *uvspec* não se baseia num único método para resolver a equação de transmissão radiativa, mas numa série de diferentes soluções para a mesma. Após a sua resolução, uma série de quantidades radiativas são calculadas. Estes incluem as irradiações $E \downarrow$ diretas e difusas para baixo e a irradiação $E \uparrow$ ascendente (Mayer & Kylling, 2005):

$$E \downarrow = \int_{2\pi} L(\Omega) \cos\theta d\Omega \quad 5$$

$$= \int_0^{2\pi} \int_{\pi/2}^{\pi} L(\theta, \phi) \cos\theta \sin\theta d\theta d\phi \quad 6$$

$$E \uparrow = \int_0^{2\pi} \int_0^{\pi/2} L(\theta, \phi) \cos\theta \sin\theta d\theta d\phi$$

E os fluxos actínicos correspondentes

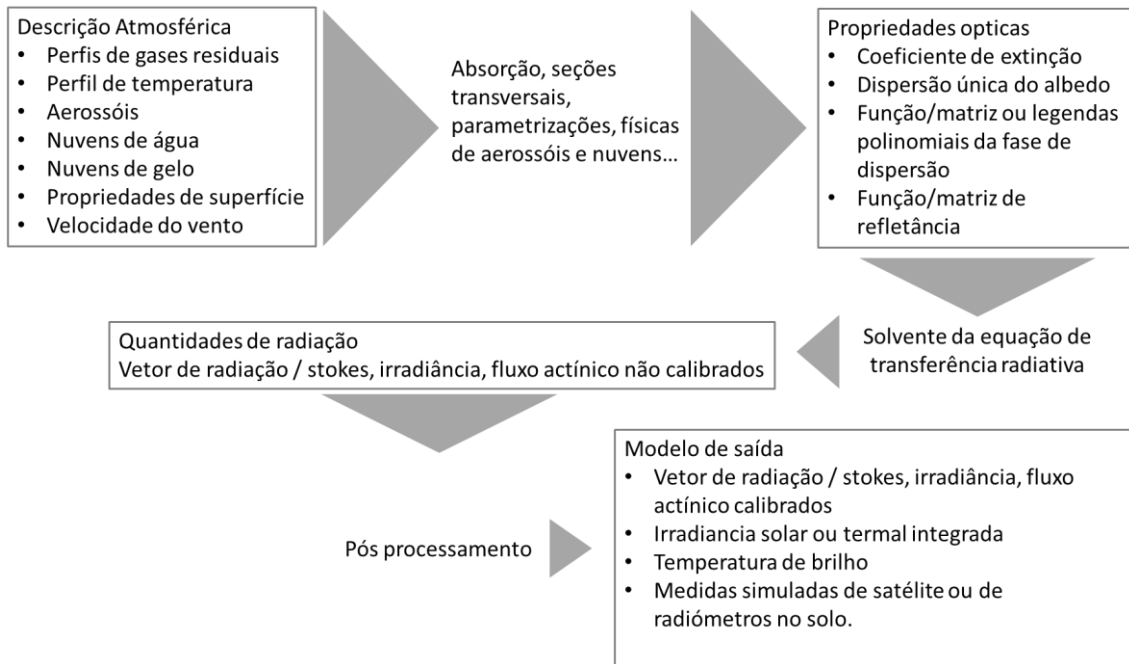
$$F \downarrow = \int_{2\pi} L(\Omega) d\Omega \quad 7$$

$$= \int_0^{2\pi} \int_{\pi/2}^{\pi} L(\theta, \phi) \sin \theta d\theta d\phi \quad 8$$

$$F \uparrow = \int_0^{2\pi} \int_0^{\pi/2} L(\theta, \phi) \sin \theta d\theta d\phi$$

Em que θ diz respeito aos ângulos polares e ϕ ao azimute.

O esquema seguinte retirado da nota técnica para o *software LibRadan*, sobre o qual assenta o processador Sen2cor, mostra o processo pelo qual se aplicam as correções atmosféricas e em que parâmetros assentam a escolha da resolução para a RTE.



Fluxograma 2 – Estrutura do modelo de *uvspec* (adaptado de Mayer & Kylling, 2005).

No Sen2Cor o modelo é executado uma vez para gerar um grande conjunto de LUT (Look-up tables) com funções específicas para o sensor (brilho, transmissões diretas e difusas, fluxos solares diretos e difusos, e albedo esférico), isto explica uma grande variedade de condições atmosféricas, geometrias solares e elevações do solo. Esta base de dados é gerada com uma alta resolução espectral (0,6 nm) e, em seguida, é feita uma reamostragem com respostas espectrais do sensor. Esta LUT é utilizada como um

modelo simplificado para inverter a equação de transferência radiativa e calcular a reflectância na base da atmosfera (BOA) (ESA, 2020).

O processamento de base é do tipo aerossol rural/continental. Outras *look-up tables* (LUT) também podem ser usadas de acordo com a localização geográfica da imagem e climatologia. No Sen2Cor, existem dois modelos atmosféricos (verão de latitude média e inverno de latitude média), dois tipos de aerossóis (rurais e marítimos), seis concentrações de ozono (dependendo do caso de verão ou inverno), e seis ou quatro quantidades diferentes da coluna de vapor de água (dependendo do verão ou do inverno). Os elementos gasosos da atmosfera e aerossóis podem ser derivados pelo próprio algoritmo ou fixadas pelo operador *a priori*. A espessura ótica de aerossóis (Aerosol Optical Thickness, AOT) pode ser derivada a partir da própria imagem, recorrendo a pixels de áreas com comportamento refletivo conhecido, de preferência vegetação densa escura ou corpos de água; se não puderem ser encontradas, limites incrementais podem ser aplicados na banda 12 para identificar áreas de referência com brilhos médios. A visibilidade e AOT correspondentes são derivados da correlação da banda 12 (SWIR) com a banda 2 (Blue) (ESA, 2020; Richter, Louis & Berthelot, 2011).

A detecção do vapor de água sobre terra é feita recorrendo ao algoritmo de absorção Diferencial Atmosférica Pré-corrigida (Atmospheric Pre-corrected Differential Absorption, APDA), que é aplicado as bandas 8a e 9, em que a primeira é a banda de referência para a janela atmosférica e a banda nova é o canal de medida na região de absorção; opcionalmente é possível remover a neblina (Mueller-Wilm *et al.*, 2013).

Adicionalmente, o processador Sen2Cor inclui várias correções opcionais que podem ser ativadas, como correção de cirros, correção do terreno, correção de adjacência e correções empíricas de função de distribuição de reflexão bidirecional (*Bidirectional Reflectance Distribution Function*, BRDF). O produto final contém imagens de diferentes resoluções espaciais: 10 metros para as bandas 2, 3, 4 e 8, 20 metros para as mesmas e para as bandas 5, 6, 7, 8a, 11 e 12, e por fim, 60 metros para todas as anteriores mais as bandas 1 e 9, a banda 10 desaparece no nível 2A. Os valores de da base da atmosfera são codificados em JPEG2000 com o valor de quantificação de 10 000,

um fator de 1/1000 precisa de ser aplicado aos níveis digitais no nível 2A para obter os valores físicos de refletância da superfície (Main-Knorn *et al.*, 2017).

Como resultado obtém-se as imagens de reflexão corrigida para base da atmosfera. Opcionalmente pode-se ainda obter imagens de terreno e cirrus corrigidas. O seu formato de produto de saída é equivalente ao produto de usuário nível 1C: imagens em formato JPEG 2000 com três resoluções diferentes, 60, 20 e 10 m (ESA, 2019).

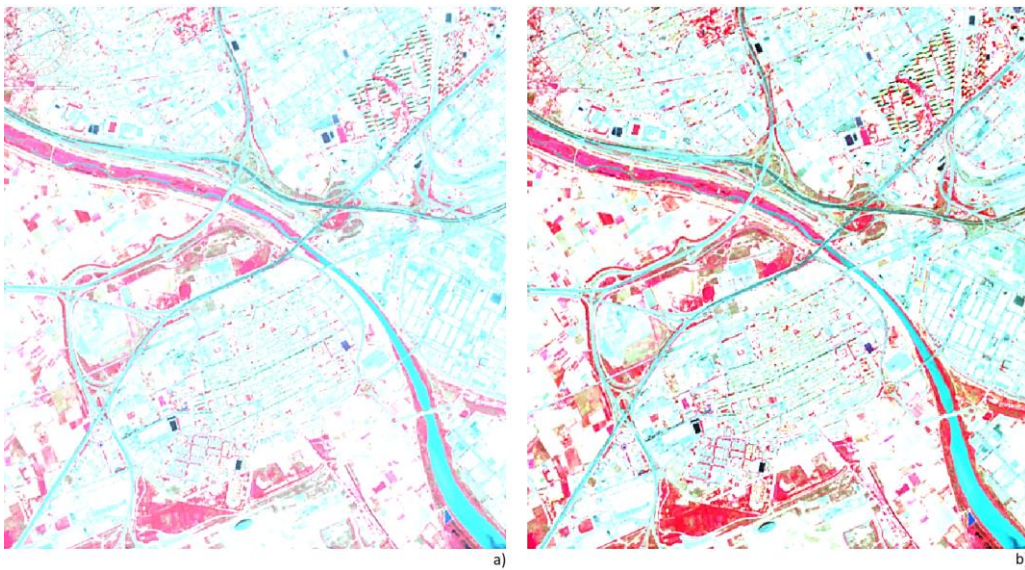


Figura 13 – Exemplo de transformação de uma imagem nível 1C (a) para uma imagem nível 2A (b) da área do estuário do rio El Llobregat. As imagens Nível 1C também podem ser referidas como TOA *image data*, e as imagens de nível 2A como BOA *Image data*.

Numa primeira instância, identificaram-se 99 imagens *Sentinel-2* MSI que poderão ser utilizadas, esta identificação foi feita com base nas amostras disponíveis no website, e o critério foi a presença de nebulosidade sobre a área de estudo. Como a Imagem que contém a área de estudo é na sua maioria oceano, tornou-se necessário que esta verificação fosse feita com recurso a análise visual, procurando as imagens nas quais área de estudo estivesse totalmente desimpedida de nuvens, mesmo quando a imagem não tem um valor de 0% da nebulosidade. Este número de imagens poderá ser reduzido ao iniciar o pré-processamento e as avaliações estatísticas de qualidade.

2.3. Operações geométricas sobre a imagem

No sentido de otimizar os dados para utilização e diminuir o poder de processamento foi necessário proceder a duas operações geométricas distintas sobre as imagens.

O *subset* recorta a imagem de forma a ajustá-la à área de estudo, isto permite otimizar o poder computacional necessário às operações de processamento. A aplicação desta operação permite retirar apenas uma parte da imagem adequando-a às dimensões da área de estudo. Em *software SNAP* apenas é possível diminuir o número de colunas e linhas, pela forma irregular da área de estudo, existem áreas que lhe são externas que foram alvo de processamento e recortadas apenas no pós-processamento.

A segunda operação foi o *resample*, que permite uniformizar as resoluções espaciais de todas as bandas, viabilizando assim mais combinações, que permitiram uma melhor análise visual da área de estudo e dos tipos de coberto de solo que a compõem, algo importante para a seleção de áreas de treino para os algoritmos de classificação utilizados em OBIA. A resolução de saída foi de 10m, pela complexidade da área de estudo que requer a melhor resolução possível para a sua análise, portanto, o *resample* foi aplicado sobre as bandas com resoluções de 20m (*Red Edge 2, 3, 3 e 4, SWIR 1 e 2*) e 60 m (*Aerosol, water vapour e cirrus*). O método selecionado para a interpolação das bandas de menor resolução foi o *Nearest*, em que o valor de cada pixel é definido como o valor de pixel de entrada mais próximo.

2.2. Estatísticas da imagem e qualidade dos dados

A análise das estatísticas das imagens permite que se afira a qualidade das imagens processadas pelo Sen2Cor e descartar aquelas que não atinjam níveis de qualidade aceitáveis.

A imagem é uma matriz de linhas e colunas constituída por pixels, e cada um destes pixels é tratado como uma unidade individual e separada das restantes, tendo cada pixel um nível digital associado para cada uma das bandas da imagem. Isto permite calcular estatísticas digitais com base num universo (n° total de pixéis na imagem) ou numa amostra (*Spatial Subset*) da imagem. Para se avaliar a qualidade de uma imagem é necessário atender a estas estatísticas uni-variados e multivariados calculados a partir dos níveis de brilho ou Radiométricos dos pixéis de cada banda que compõe a imagem. As estatísticas básicas principais que nos permitem ver as características da imagem são o histograma, o mínimo e o máximo, a amplitude, a média, o desvio padrão, e a matriz de correlação.

Para avaliar a qualidades dos dados e efetuar as devidas operações de melhoramento ou, em última instância, descartar imagens, recorreu-se aos indicadores estatísticos acima mencionados. No final processaram se apenas as imagens que se entendeu terem qualidade suficiente para processamento digital.

Em processamento digital de imagem os histogramas revelam-se de primária importância quando se pretende analisar as características da imagem com base na frequência e distribuição dos pixéis pelos níveis digitais. Os histogramas são amplamente utilizados para a avaliação da qualidade dos dados, pois a sua forma descreve a amplitude do contraste de uma imagem e permite analisar entre outros o seu grau de homogeneidade. Isto acontece pois as bandas individuais são tipicamente codificadas em *bits*; no caso do dos dados *Sentinel-2* MSI, estas encontram se codificadas em 2^{12} bits, o que significa que os valores variam entre zero e 4 095. Tabulando a frequência da ocorrência de cada um dos seus valores oferece uma forma de informação estatística que pode ser visualizada sob a forma de um gráfico (Jensen, 2015).

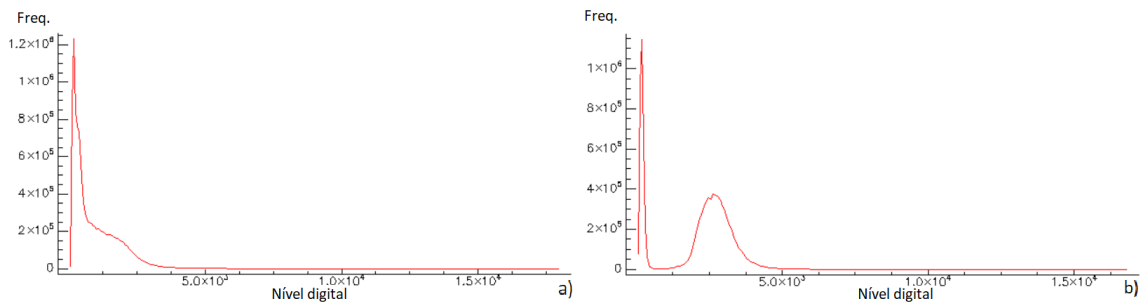


Figura 14 – Comparação entre histogramas de bandas diferentes, histograma a) corresponde à banda 4 (Vermelho) e o histograma b) corresponde a banda 8 (PIV). Os histogramas permitem avaliar a distribuição de níveis digitais na imagem, neste exemplo, o PIV mostra mais frequências em níveis mais altos pela influência da vegetação presente na cena.

A média é um indicador estatístico que divide pelo número total de pixels a soma do valor radiométrico de todos eles. O resultado da média total indica o nível “brilho” global da imagem, em que $\sum x_i$ é a soma total de brilhos registados para todos os pixels e n o total de pixels, a média \bar{X} é dada por:

$$\bar{X} = \frac{\sum x_i}{n} \quad 9$$

Quando mais elevado o valor da média em relação ao nível radiométrico máximo, mais clara a imagem, por outro lado, quanto mais baixa, mais escura como exemplificado na comparação entre as bandas 3 (vermelho) com uma média de 936,8 e a banda 8 (PIV) com uma média de 2038,4.

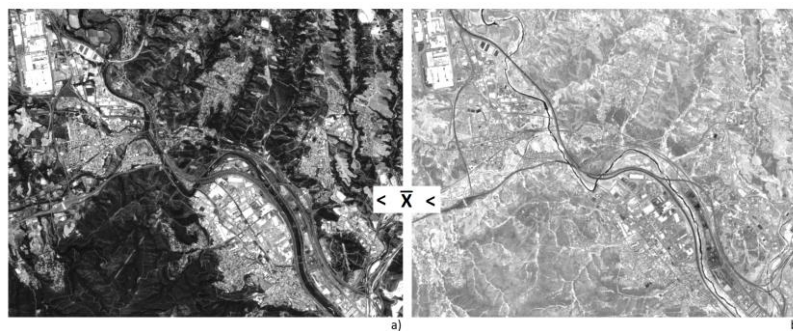


Figura 15 – Comparação entre o brilho de duas bandas com médias diferentes das localidades de Sant Andreu de la Barca, Castellbisbal e Martorell e áreas circundantes, a imagem a) corresponde à banda 4 e a imagem b) à banda 8.

O desvio-padrão indica a dispersão média dos valores dos níveis digitais da imagem em relação ao seu valor médio, o desvio padrão é dado pela raiz quadrada da

variancia que serve de atenuante dos desvios às médias. O desvio padrão é então um indicador estatístico da qualidade dos dados em Detecção Remota em que quando maiores forem os desvios, melhores os dados para a classificação, uma vez que quanto maior, mais heterogeneidade existe e em, Detecção Remota, mais heterogeneidade pode significar uma classificação mais precisa, pois confere aos elementos da cena uma maior separabilidade espectral entre si.

O desvio padrão é dado pela raiz quadrada da variância que consiste no quadrado da soma da média subtraída a cada um dos valores $\sum(x_i - \bar{X})^2$ dividido por 1 subtraído ao total de pixels $n - 1$ e é dado por:

$$S = \sqrt{\frac{\sum(x_i - \bar{X})^2}{n-1}} \quad 10$$

A matriz de correlação é dada pelo coeficiente ρ de Pearson ou r , e analisa as correlações entre pares de bandas espectrais. Os valores variam entre -1 e 1, quanto mais próximo de das extremidades maior a correlação, no caso das bandas espectrais da mesma imagem, a correlação é sempre positiva. A correlação é dada pela covariância de de duas variáveis neste caso, duas bandas diferentes $n(\sum xy) - (\sum x)(\sum y)$ divididas pela raiz quadrada da multiplicação das variâncias de cada uma das variáveis. A função para a correlação de Pearson é:

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}} \quad 11$$

Em processamento digital de imagem, é geralmente analisada uma matriz de correlações entre as bandas, em que os pares de bandas mais úteis são aqueles que mais se aproximam do 0, ou seja, os que tenham correlações fracas que potencialmente, produzem uma maior variedade de informação espectral.

A análise destes parâmetros nas imagens permitiu selecionar as imagens de maior qualidade a utilizar na análise multitemporal da vegetação, sendo especialmente importante ter estes indicadores estatísticos para classificações supervisionadas, de forma a decidir as melhores bandas de entrada para os classificadores.

CAPÍTULO III

A análise multitemporal do coberto vegetal

*Este capítulo debruça-se sobre os métodos de obtenção do índice multitemporal de vegetação, incluindo a teoria e o workflow. Foi também feita a correlação, a analisar no capítulo VI, entre a presença de altos níveis de atividade clorofilina e as ocorrências de *Aedes albopictus*.*

Capítulo IV – A análise multitemporal do coberto vegetal

1. Os índices radiométricos e a sua importância

Os índices de estado de superfície fornecem informações sobre o espaço que podem ser utilizadas para a avaliação da adequabilidade de habitats para espécies de mosquitos.

Os índices espectrais de superfície são imagens que são geradas a partir de cálculos em imagens multiespectrais com o objetivo de demarcar um fenómeno específico enquanto atenuam o efeito de outros fatores. Para extrair estes índices, são feitas combinações através de operações matemáticas sobre bandas, geralmente adição, subtração e rácios entre bandas. O cálculo destes índices está ligado a bandas localizadas em partes específicas do espectro eletromagnético, ou seja, as bandas irão variar dependendo do sensor. Os índices têm a mais valia de ser pouco sensíveis ao efeito das sombras nos objetos que se querem estudar, pois são calculados, utilizando formulas normalizadas focadas na diferença de comportamento espectral entre duas ou mais bandas.

Estes índices geralmente vêm agrupados em quatro ou cinco classes definidas pelo fenómeno para o qual foram desenhados. Estes índices são frequentemente divididos em categorias, consoante a sua finalidade, e podem ser para análise de vegetação, de solo, de água, de geologia e de paisagem. A sua utilização tem sido abrangente desde a monitorização agrícola (Beeri & Peled, 2006; Kobayashi *et al.*, 2020) gestão florestal (Huete, 2012; Ye *et al.*, 2014), monitorização de áreas ardidas com utilização de índices de severidade como o NBR ou BAI (Tran *et al.*, 2018; Harris, Veraverbeke & Hook, 2011) ou gestão urbana através de índices de áreas construídas (built up indexes) (Ali, Hasim & Abidin, 2019; Yantao, Nguyen & Cheng, 2019).

Nesta dissertação pretende-se utilizar a vegetação como proxy para a distribuição potencial de *Aedes albopictus*. Para tal exploraram-se os índices de

vegetação ou fenologia para a escolha do mais adequado. Ainda foram testados índices de água, porém os resultados não foram satisfatórios.

Os estudos de propriedades fenológicas através da Detecção Remota são feitos sobre dados recolhidos por sensores de satélite. Estes sensores medem comprimentos de onda de luz absorvidos e refletidos pelo coberto vegetal. Os índices de vegetação, na sua maioria dependem da diferença resultante da absorvância das plantas nos comprimentos de onda da luz visível, especialmente no vermelho e a alta reflexão nos comprimentos de onda da luz quase infravermelha, chamado de próximo infravermelho, que é invisível aos olhos humanos (Brown, 2018).

Os índices de vegetação derivados da observação da terra têm uma grande importância em vários aspetos do estudo das dinâmicas planetárias, em especial, aquelas que afetam as populações, tais como monitorização da vegetação, estudos de seca, atividades agrícolas, modelação climática e hidrológica (UN-Spider, 2020). «Os índices de vegetação são amplamente reconhecidos como um indicador robusto e espacialmente explícito para aferir processos sociais-ecológicos, tais como a conversão do uso de habitats-terra (isto é, expansão urbana) ou rotação das culturas.» (Pasimeni *et al.*, 2019, p.472)

O Index database lista 139 índices de vegetação só para o sensor MSI a bordo do *Sentinel 2* (IDB⁸, 2020). Para a escolha do índice a utilizar procurou-se responder a duas perguntas: a primeira sendo se este índice é indicado para monitorizar o vigor da vegetação e a segunda se este índice já fora utilizado em estudos anteriores de modelação de habitats prováveis de mosquito. Vários índices devolvem o vigor e estado do coberto vegetal, porém aquele que mais tem sido mais utilizado na área de controlo de vetores é o NDVI.

⁸ O IDB é um repositório de índices de Detecção Remota com uma visão geral rápida dos quais os índices utilizáveis para determinado sensor, em <https://www.indexdatabase.de/>.

2. O Normalized Difference Vegetation Index

A aplicação e utilização do NDVI para a monitorização de *arbovírus* e outros patogénicos tem sido amplamente estudada, com especial destaque para a febre amarela e malária, com muitos autores a utilizar a vegetação como proxy para a localização de habitats de mosquito, isto acontece pois estudos anteriores demonstraram que existe uma elevada correlação entre valores altos de NDVI e a presença de algumas espécies de mosquitos (Kogan, 2020; Lopes *et al.*, 2006; Lourenço *et al.*, 2011). Sendo o *Aedes albopictus* um mosquito de buraco de árvore nativo das áreas tropicais do sudeste asiático, a vegetação continua a ser um fator preponderante na distribuição e atividade deste vetor.

O *Normalized Difference Vegetation Index*, atende a resposta espectral da vegetação que é mínima na banda do vermelho (RED) e máxima na banda do infravermelho próximo (NIR) (figura 16). A utilização do NIR permite avaliar a atividade clorofilina da vegetação, o que é bastante útil para diferenciar a vegetação fotossintética da vegetação seca e do solo, bem como outras ocupações do solo. O NDVI é obtido pela seguinte fórmula:

$$NDVI = \frac{NIR - RED}{NIR + RED} \quad 12$$

O NDVI tem vantagens, dado que é calculado através de um procedimento de normalização que é muito útil para mitigar os efeitos das sombras na iluminação dos objetos em regiões montanhosas, sombras de nuvens, e variações sazonais etc. e melhorar características espectrais que não são visíveis inicialmente (Lillesand, Kiefer & Chipman, 2004). O NDVI tem uma resposta sensível à vegetação verde mesmo para áreas cobertas de vegetação baixa.

Este índice mostra-se relacionado não só com a estrutura das copas e com a área de folha, mas também com a atividade fotossintética. Como falhas é apontado a alta sensibilidade do NDVI aos efeitos do brilho do solo, cor do solo, atmosfera, nuvem e sombra de nuvem, e sombra da cobertura das folhas e requer calibração de Deteção Remota (Xue & Su, 2017), problema que é tido em conta, e minimizado pelas calibrações

feitas sobre a imagem para a produção de imagens *Sentinel-2* L1C e L2A (ESA, 2019) e as correções atmosféricas realizadas anteriormente. Considera-se que NDVI fica saturado por vegetação densa ou um Índice de Área da Folha (LAI) elevado, o que é incorreto. A razão para a relação próxima entre o NDVI e LAI é porque todos os fótons visíveis são absorvidos quando LAI é elevado confirmando que o NDVI apresenta a capacidade fotossintética e produção primária em vez da LAI (Yengoh *et al.*, 2014). Esta saturação também não foi verificada no decorrer deste estudo.

O índice Radiométrico para a vegetação (NDVI), serviu então de proxy para avaliar a adequabilidade dos habitats, assumindo que as presenças dos altos níveis de atividade clorofilina também estão dependentes da presença de água.

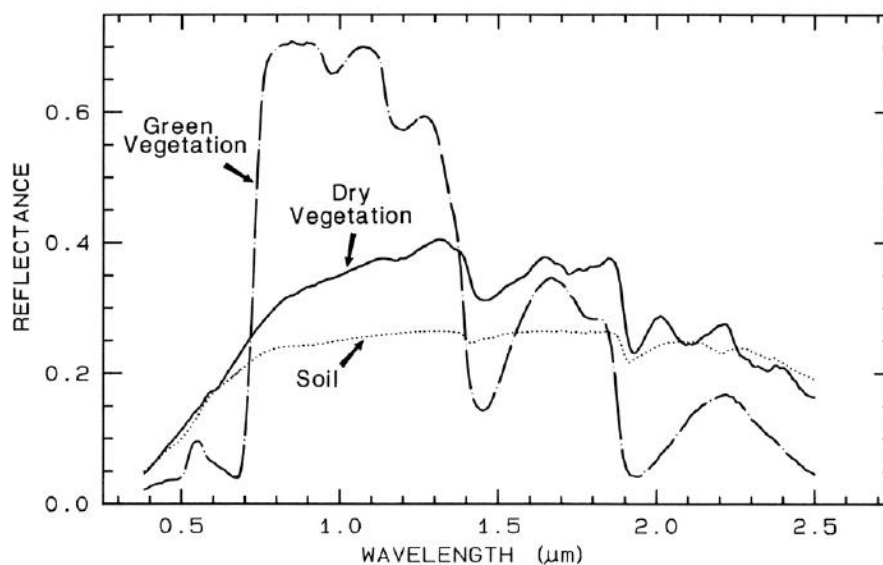


Figura 16 – Assinatura espectral da vegetação fotossintética (verde), vegetação não fotossintética (seca) e solo (Clark, 1999)

O estado da vegetação é particularmente importante na análise em contexto de habitats do vetor, isto pois a vegetação verde (ou fotossintética) que tem altos níveis de atividade clorofilina está associada a presença de água (margens de rios, lagos, áreas irrigadas) a par que a vegetação seca está associada a falta de água na maioria dos casos.

Em Itália, após 20 anos de presença de *Aedes albopictus*, a presença de espaços verdes, conjugada com a humidade relativa do ar, faz com que o mosquito sobreviva mais tempo (Romi *et al.*, 2008) e segundo a direção geral de saúde este pode ser encontrado em áreas com vegetação no interior e em redor de habitações (DGS, 2020),

o índice de vegetação poderá servir como ferramenta para mapear estas áreas em contexto urbano (Parques urbanos, vegetação urbana ornamental, ...).

2. O cálculo do índice e mapeamento multitemporal da vegetação

A fórmula de NDVI foi aplicada aos dados *Sentinel-2* MSI L2A de forma a produzir uma imagem com valores de NDVI correspondente a cada mês do ano. Os índices foram extraídos utilizando o *software* SNAP numa primeira instância para 2019, ano em que existe mais imagens disponíveis sem nebulosidade para a área de estudo, totalizando 31 imagens, das quais se selecionaram 12 imagens, uma por mês de forma a garantir uma distribuição simétrica ao longo do ano.

Para os anos de 2018 e 2017, pela existência de meses sem qualquer imagem não obstruída por nuvens, foram gerados NDVI para os meses existentes, com os quais se interpolaram dados para os meses sem dados através de uma média aritmética do mês anterior e do mês seguinte. Para o ano de 2016 por falta de imagens utilizáveis para toda a primeira metade do ano, considerou-se apenas a segunda metade para o cálculo do mapa de vegetação multitemporal. Numa segunda instância foram produzidos mapas sazonais, seguindo os passos que originaram os mapas anuais, estes permitem uma análise da evolução intra-anual da vegetação, utilizados para a classificação do uso do solo no capítulo IV.

A extração do índice de vegetação foi feita em SNAP, recorrendo ao processador temático de vegetação – NDVI *Processor*. Aplicou-se este operador sobre cada uma das imagens selecionadas. Esta seleção teve em conta o critério de escolha, sempre que disponível, imagens que não se afastassem muito em termos do dia do mês em que foram captadas pelo sensor, estas devem situar-se na segunda metade do mês a fim de garantir uma distância temporal equivalente entre elas.

O mapa de vegetação resulta da sobreposição de valores de NDVI de todas as imagens em cada um dos pixels, para isto optou-se pelo cálculo da média aritmética. Para os mapas anuais esta é uma média de todos os 12 meses que resulta em 4 nos

mapas anuais de vegetação. Para o sazonal utilizaram-se 3 meses para cada estação, por uma questão de coerência, o mês de dezembro foi incluído no outono e o mês de janeiro que dá início ao ano, é o primeiro mês de inverno, para a vegetação mensal, é calculada a média das imagens disponíveis para esse mês, quando existe mais do que uma imagem. Após o cálculo das médias para os anos e para as estações, foi feito o recorte da área de estudo sobre os rasters produzidos, para fazer face ao volume de rasters a recortar, o polígono que servirá para delimitar o recorte é o correspondente à área de estudo na figura 9. Esta operação é feita para os mapas anuais, sazonais e mensais.

Para obter mapas de vegetação, foi definido com base em literatura um limite (*Threshold*) sobre os valores de NDVI a partir do qual se considera haver presença de vegetação. Consideraram-se duas opções, $> 0,5$ e $> 0,3$, o primeiro limite surge na caracterização da vegetação urbana para Kuala Lumpur, Malásia, que está situada na área onde se considera que o mosquito é autóctone (Hashim, Latif & Adnan 2019), e a segunda tem sido o limiar definido para a delimitação da vegetação utilizada em modelação para a presença de outros vetores em outros estudos incidentes sobre a vegetação na Europa e na bacia do mediterrâneo (Ferraguti *et al.*, 2016; El-Gammal, Ali & Samra, 2014). Dada a forte adaptabilidade do *Aedes albopictus*, optou-se pelo limite de 0.3 pela adequabilidade ao bioma mediterrânico, onde a área de estudo se insere.

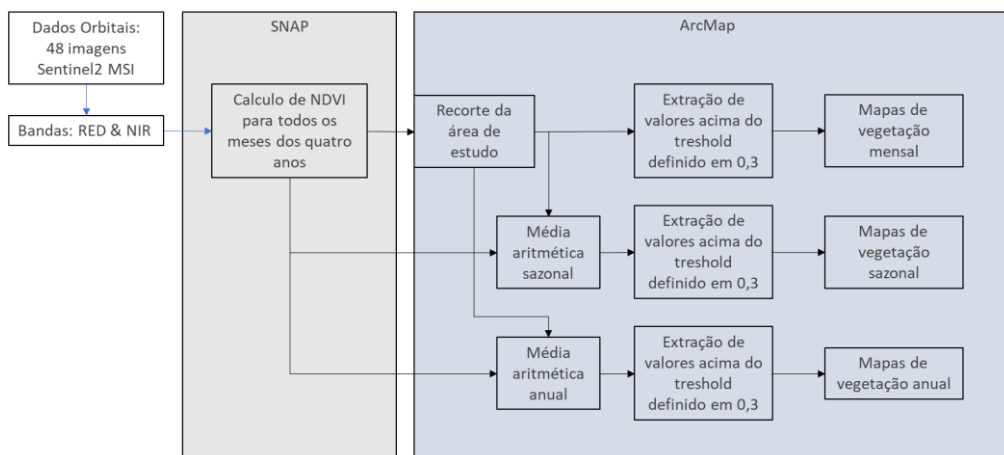


Figura 17 – Esquema de extração mensal, sazonal e anual de valores NDVI para a produção de mapas de vegetação

É importante seguir a ordem de processamento supramencionada a fim de evitar erros de cálculo. A extração dos valores acima de 0.3 NDVI só é efetuada após o

cálculo das médias com a finalidade de não perder dados relativos à vegetação de carácter temporário.

O exemplo mostrado na figura 18 representa a evolução sazonal da vegetação para o *Parc de la Ciutadella* e envolvente da cidade de Barcelona, é possível denotar o aumento da atividade clorofilina no verão em comparação com o inverno.

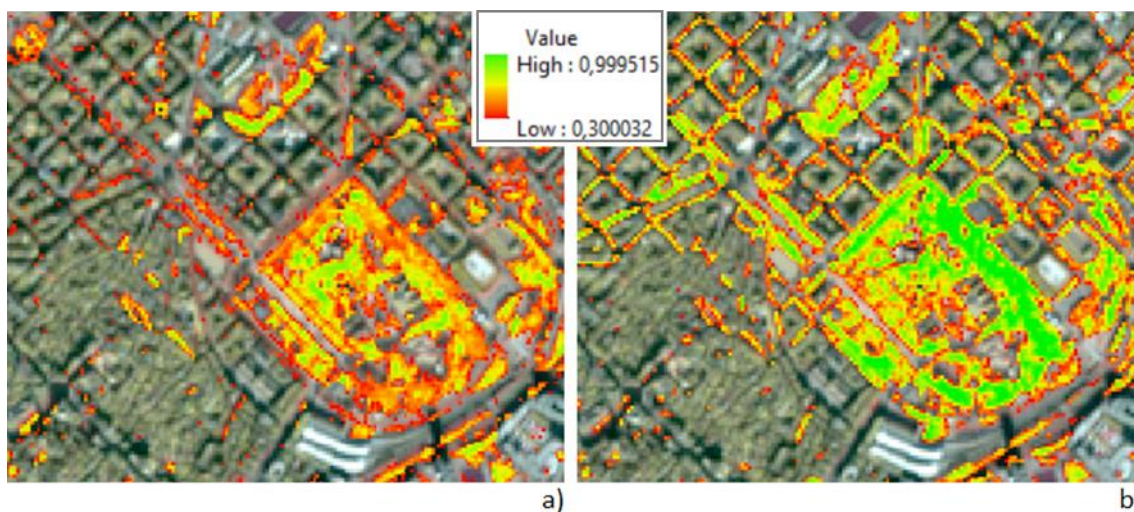


Figura 18 – Evolução do índice de vegetação acima do limiar proposto para o Parc de la Ciutadella e redondezas, entre meses de janeiro a) e julho b).

O mapeamento mensal e sazonal da vegetação permite verificar que a vegetação tem uma atividade clorofilina mais intensa no final da primavera e no verão, onde as temperaturas são mais elevadas e a irrigação artificial de verdes urbanos é amplamente utilizada.

3. A relação espacial entre os níveis de atividade clorofilina e ocorrência de *Aedes albopictus*

Relacionando os dados com as observações, detetou-se uma correlação entre a presença de vegetação e observações documentadas pela população, com base nos dados de observações de dados geográficos voluntários retirados da plataforma GBif, que compila dados de biodiversidade. Sabendo que «o mosquito tem um alcance de voo

limitado, cerca de 200 metros» (DGS, 2020) e de forma a avaliar a sua correlação com vegetação, foram agrupadas as ocorrências em frequências para intervalos de distância de forma a analisar o seu comportamento.

As classes de distância foram definidas em 0 para as observações foram realizadas dentro de um pixel com um valor de NDVI médio anual superior a 0.3. As restantes classes foram feitas consoante a distância sobre uma escala exponencial até se deixarem de verificar novas observações. Esta abordagem foi escolhida para validar os resultados em detrimento de um único buffer de 200 metros, pois para a área de estudo em questão, são residuais as áreas a mais de 200 metros de uma área com NDVI superior a 0.3.

O método para analisar esta relação foi o r de Pearson ou Coeficiente de correlação de momento do produto desenvolvido por Karl Pearson em 1948. Este método de correlação mede a relação linear entre duas variáveis e pode ter um valor entre -1 e 1, quanto mais próxima dos extremos mais forte a correlação, negativa ou positiva. Existem vantagens em utilizar o r de Pearson, principalmente por ser uma forma simples de avaliar a associação entre as variáveis vegetação e distância. A desvantagem é que apenas pode ser utilizado sobre variáveis lineares, no entanto esta desvantagem não se aplica a este estudo por ser o caso. O seu cálculo é dado pelo rácio das variações comuns a duas variáveis, ou seja, a covariância, quanto mais elevada esta for, maior foi a correlação (Chee, 2013).

Esta análise foi feita para todos os anos e também para um conjunto de pontos aleatórios para validação dos resultados. O número de casos em função da distância foi calculado em função da vegetação bem como as observações para esse ano, a correlação de Pearson foi também calculada para quantificar em número, por ano, influência provável que a distância da vegetação tem no número de casos. Aqui procura-se uma correlação negativa muito forte.

As distribuições em função da distância apresentar uma curva exponencial negativa. Isto é, nos primeiros metros se acumularem uma quantidade significativa de observações que irão baixando até se deixar de verificar a presença do mosquito. Partido desta análise, e de forma a complementar a análise estatística entre as

ocorrências e a distância à vegetação, foi calculada a regressão exponencial para os valores anteriores e para validação destes resultados repetindo-se o processo utilizando 860 pontos aleatórios, que corresponde ao número total de ocorrências pontos gerados aleatoriamente.

Este método tem problemas intrínsecos à análise estatística, em primeira instância o facto de correlação não significar dependência, isto ocorre, pois, apesar de duas variáveis seguirem tendências proporcionais ou proporcionalmente inversas uma da outra, estas podem ser ligadas a fatores externos que não são tidos em conta. Outro dos problemas prende-se com a recolha dos dados por ciência cidadã, neste caso é possível existir um desfasamento geográfico entre o local da ocorrência e o registo da mesma, o que poderia alterar os resultados. A utilização das observações confirmadas e prováveis traz um maior número de ocorrências, mas ao mesmo tempo adiciona erros potenciais causados pela possibilidade destas observações não se tratar de *Aedes albopictus*.

A ausência de observações de *Aedes albopictus* durante os meses de inverno, janeiro, fevereiro e março, o que indica que fatores climáticos têm efeito sobre a sua atividade, o que significa que, com as alterações climáticas, poderão existir melhores condições de propagação, e juntando ao facto de que este vetor procura humanos em especial, a urbanização e a criação de ilhas de calor urbano poderão ter um efeito na sua distribuição temporal, alargando o seu tempo de atividade nestes locais.

CAPÍTULO IV

A análise da ocupação do solo por *Object Based Image Analysis*

Este capítulo aborda a teoria e a prática para a classificação de imagem utilizando a análise orientada por objetos. São abordados os fundamentos teóricos e metodológicos, a segmentação de imagem, a recolha e validação de amostras teste, a extração das estatísticas para construção do modelo, a classificação dos segmentos e a validação temática. Para este capítulo foram testados dois algoritmos de segmentação e dois algoritmos de treino de classificador de vetores e foram escolhidos os que apresentaram melhores resultados.

Capítulo IV - A análise da ocupação do solo por *Object Based Image Analysis*

1. A Object Based Image Analysis.

A análise orientada por objetos (OBIA) pode ser aplicada a imagens obtidas por Detecção Remota para a criação de informação geográfica, a esta utilização dá-se o nome de GEOBIA. GEOBIA pode ser definida como «uma subdisciplina da Ciência da Informação Geográfica (*GIScience*) dedicada ao desenvolvimento de métodos automatizados para a divisão de imagens de Detecção Remota em objetos de imagem significativos, e avaliando as suas características através de escalas espaciais, espectrais e temporais, de modo a gerar novas informações geográficas em formato pronto para SIG.» (Hay & Castilla, 2008, p.77). A OBIA, ou neste caso e para objetos geográficos (GEOBIA) consiste numa categoria análise digital de imagem em Detecção Remota que estuda entidades geográficas, ou fenómenos presentes na imagem através da definição (segmentação) e análise de objetos ao invés de pixels individuais (Chen *et al.*, 2018 citado de Castilla & Hay, 2008; Blaschke, 2010).

Esta metodologia «combina segmentação (que é uma fase fundamental da abordagem) e classificação contextual. A segmentação divide a imagem em grupos homogêneos de pixels (segmentos), que são durante o processo de classificação semântica dispostos em classes baseadas nas suas características espectrais, geométricas, texturais e outras» (Veljanovski, Kanjir & Oštir, 2011), «o objetivo da classificação orientada por objetos da imagem é então associar estas primitivas com objetos do mundo real.» (Gao, 2009, p.421).

Desde cerca dos anos 2000 a GEOBIA surgiu como uma alternativa de análise de imagens em Detecção Remota, especialmente imagens de alta resolução espacial. Durante este período, os interesses de investigação demonstraram uma mudança do desenvolvimento dos fundamentos teóricos de OBIA para modelos de objetos geográficos avançados e a sua implementação numa grande variedade de aplicações no mundo real. Estas tendências puderam ser observadas em GEOBIA com aumento do volume e qualidade de dados, algoritmos de segmentação de imagem, e extração de

objetos geográficos com base estruturas de modelação (Chen *et al.*, 2018). O reconhecimento de objetos espaciais com recurso à OBIA tem verificado grandes avanços, com mais de 800 artigos reportados até ao ano de 2013 fazendo uso desta metodologia nas mais diversas aplicações, desde o mapeamento ambiental até à medicina (Aryal *et al.*, 2013).

Não independente dos desenvolvimentos teóricos e metodológicos, o advento de imagens de satélite de alta resolução e a variedade de fontes, contribuíram para desenvolver uma abordagem aos pixels segmentados usando limites de campo vetorial e atribuição de classes de ocupação do solo (Blaschke, Kelly & Merschdorf, 2016).

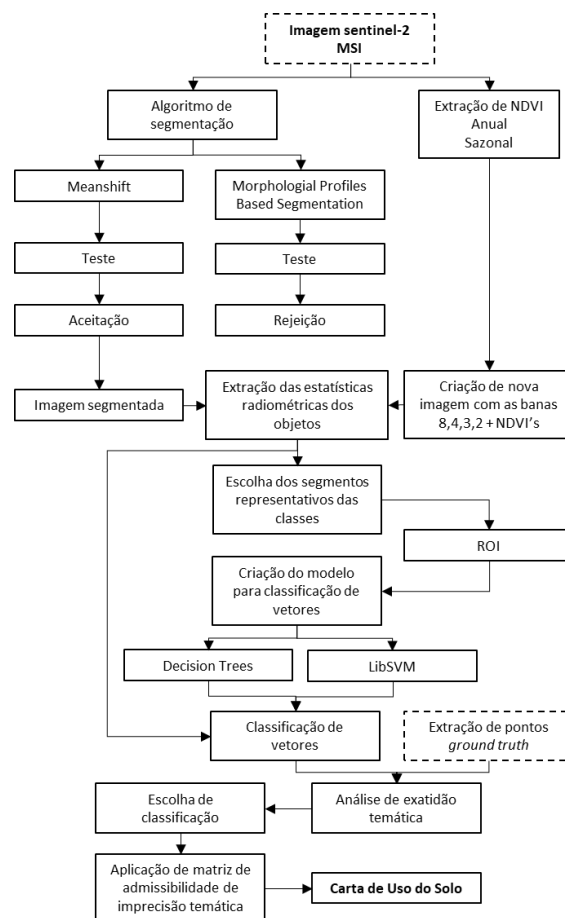
GEOBIA tem sido utilizada em toda uma grama de estudos e aplicações no mundo real, que vão desde agricultura (Solano, Fazio & Modica, 2019; Vogels *et al.* 2019), mapeamento de usos do solo (Silva & Bacani, 2017; Wezyk *et al.*, 2016) e com particular interesse para este estudo em conjunto com os usos do solo, deteção e identificação de corpos de água em contexto urbano (Kaplan & Avdan, 2017; Huang *et al.*, 2015).

GEOBIA compreende quatro grandes etapas, a segmentação, treino, classificação e validação. Na primeira etapa é feita a escolha da imagem e a sua inspeção para aferir a sua qualidade, em seguida, a escolha do algoritmo e segmentação. Nesta dissertação realizaram se dois testes, o primeiro virado para as características espectrais para a formação dos grupos de pixéis que irão dar origem aos segmentos, *MeanShift*, e o segundo que adiciona a componente morfológica dos elementos da imagem para fazer a segmentação, o *Morphological Profiles Based Segmentation* ou MPBS.

A segunda etapa passa pela seleção de segmentos que irão servir como amostras de treino, é aqui que as classes irão ser determinadas. Até aqui o procedimento não difere muito do descrito em trabalhos referidos por Blaschke *et al.*, (2008) ou por Gao (2009). As diferenças surgem na escolha de programas, mas nuclearmente, o processo permanece o mesmo. Após este processo, retiram-se as estatísticas dos objetos, tanto os segmentos de amostra como a imagem segmentada no seu todo. As estatísticas servirão para treinar os classificadores.

Na terceira e quarta etapas é feita a classificação dos segmentos que compõem a totalidade da imagem, e sobre esta classificação é feita a validação temática dos resultados obtidos, comparando estes com amostras extraídas que correspondem ao uso verdadeiro do solo utilizando a matriz de confusão e dois medidores globais de precisão.

Os classificadores testados foram escolhidos através da análise exploratória entre um amplo leque de possibilidades. Para este estudo foram utilizados os algoritmos *Decision Trees* e o *LibSVM*. Selecionando o resultado com melhor precisão temática, a classificação foi melhorada por edição e depois elaborado o mapa de ocupação do solo com este. O mesmo procedimento foi repetido para os anos anteriores, de forma a testar a eficácia da utilização do modelo treinado com valores espectrais de uma imagem sobre outras imagens de outros anos. Antes de iniciar processo fez-se a fusão das bandas 8, 4,3 e 2 com uma resolução espacial de 10m com os NDVI anual e sazonais. Os procedimentos encontram se expostos no seguinte fluxograma.



Fluxograma 3 – Descrição dos procedimentos para a análise orientada a objetos.

Para a análise orientada por objetos (OBIA) foi utilizado o *software Monteverdi*, parte do *Orfeo Toolbox (OTB⁹)*, tradicionalmente utilizado como extensão para outros *softwares*, aqui foi testado como *standalone*, por ser suportado pelo SNAP, porém não estar configurado corretamente, o teste da sua capacidade torna-se um objetivo para a proposição da correção deste problema e melhoria do *software* para esta vertente. O *workflow* de OBIA foi também suportado parcialmente pelo *software ArcMap* para a extração e visualização dos resultados das diversas etapas que não são possíveis em OTB.

O objetivo do uso desta metodologia foi a produção de cartografia de uso do solo através de um meio semiautomático sobre imagens *Sentinel-2 MSI* para áreas suscetíveis à invasão pelo mosquito tigre asiático, em geral, são áreas complexas que englobam áreas urbanas e áreas naturais. OBIA tem se mostrado eficaz nas classificações que faz destas áreas sobre imagens *Sentinel-2 MSI* como demonstrado por Marangoz, Sekertekin & Skçin (2017). Este método também se provou bastante eficaz na detecção de água em áreas naturais (Kaplan & Avdan, 2017) e urbanas (Yılmaz *et al.*, 2019), que foi bastante útil para esta dissertação.

Para testar a melhoria da separabilidade espectral entre as classes, fundiu-se os dados obtidos pelo cálculo do NDVI para o ano bem como para as estações com a imagem contendo as bandas 8,4,3 e 2, o que irá permitir destacar a vegetação e outros elementos, apesar de esta abordagem ser mais utilizada para a componente agrícola (Castro *et al.*, 2018), outros autores utilizaram o NDVI nos conjuntos de regras para o classificador em classificações de áreas urbanas (Salehi *et al.*, 2012). Nesta dissertação foi testada a sua incorporação como banda (ou bandas no caso de multitemporal) auxiliar e aferir se são obtidas melhorias nas classificações.

As classes que se pretenderam destacar com esta abordagem são os “corpos de água”, “áreas húmidas”, “matos e descampados” e “vegetação”, todas as restantes

⁹ Orfeo ToolBox (OTB) é um projeto de código aberto para Detecção Remota de última geração. Construído sobre os ombros da comunidade geoespacial de código aberto, pode processar imagens óticas, multiespectrais e radar de alta resolução em grandes escalas. Em <https://www.orfeo-toolbox.org/>

serão agregadas e terão o nome de territórios artificializados. Na vegetação procurou-se extrair duas classes, arbórea e herbácea. Estas duas classes são constituídas por duas subclasses, na Vegetação herbácea, a uma corresponde à vegetação vigorosa, enquanto a outra a vegetação na etapa de crescimento ou na etapa de senescência. Na vegetação arbórea, a primeira corresponde a segmentos de floresta pura ou apenas com influência do solo, enquanto que a dois corresponde a árvores dispersas com influência de materiais artificiais, esta última classe surgiu da necessidade de se individualizar esta vegetação em particular e impedir que esta fosse agregada aos territórios artificializados, e surgiu como forma de tentar mitigar esse problema na classificação.

2. A segmentação de imagem

Um segmento ou objeto corresponde a um elemento vetorial fechado (polígono) que é realizado com base na homogeneidade dos espectrais dos pixels que se inserem dentro dos seus limites. O objetivo desta operação é dividir a imagem com base nas suas características, ou seja, digitaliza sobre a imagem com base na resposta espectral de um grupo de pixels com base na sua vizinhança. Ou seja, esta é afetada pelo número e características das bandas de entrada, da vizinhança e da tolerância de similaridade que o operador definir. Estes segmentos são resultado da aplicação de algoritmos sobre a imagem estes têm como tarefa básica «a fusão de elementos (de imagem) baseados em parâmetros de homogeneidade ou na diferenciação para as regiões vizinhas (heterogeneidade), respetivamente. Assim, os métodos de segmentação seguem os dois princípios fortemente correlacionados de vizinhança e de semelhança de valor» (Schiewe, 2002, p.2).

Para a segmentação da imagem foram testados dois algoritmos. Esta fase é apenas exploratória, e tem como objetivo a escolha de um algoritmo que se adeque aos objetivos pretendidos. Os algoritmos em questão diferem na forma como segmentam a imagem, os algoritmos escolhidos são o *meanshift* e o *morphological profiles based segmentation* (MPBS).

O MPBS é um algoritmo de segmentação que utiliza os residuais morfológicos entre a função de imagem original e a composição de uma granulometria construída em várias escalas e faz uso de operadores morfológicos clássicos, como a abertura e o fecho para construir o perfil morfológico. Os operadores matemáticos utilizados pelo algoritmo são a erosão e a dilatação e usam a definição de um grupo de pixels com uma forma particular a qual se dá o nome de elemento estruturante (EE).

A erosão pode ser utilizada para identificar os locais de imagem onde o EE se encaixa nos objetos presentes na imagem e no outro lado o operador de dilatação é utilizado para identificar os pixels onde o EE atinge os objetos na imagem, por objetos, definem se os grupos de pixels que o algoritmo considerou como elementos estruturantes (eg. Picos de relevo, vales, edifícios, ruas). Estes dois operadores podem ser combinados para definir outros operadores e desta forma cortar os picos do relevo topográfico que são menores do que os EE para encher os vales que são menores do que o EE (Tilton *et al.*, 2015; Yong Lv *et al.*, 2014).

Os testes preliminares feitos com este algoritmo permitiram resultaram na sua exclusão para este estudo, dado que foi fraco na segmentação dos elementos chave da cena, por exemplo a vegetação em contexto urbano não era isolada, enquanto que as áreas construídas atingiam o limite de segmentação, isto é, os elementos ficarem repartidos em vários segmentos, pelo que se optou por seguir com o *Mean Shift*, um algoritmo mais amplamente utilizado e que produz geralmente bons resultados.

O algoritmo *Mean Shift* é um algoritmo estatístico de agrupamento iterativo de características no espaço não paramétrica, ou seja, que não é afetada por pequenas variações de pontos e é uma poderosa técnica de análise de dados não supervisionada que não requer conhecimento prévio do número de clusters e não restringe a forma dos clusters. A premissa deste algoritmo é estimar o comportamento de agrupamentos de pontos amostrais e procurar a sua posição no espaço, isto é a posição com maior densidade de probabilidade, através de iterações que colocam esse ponto na sua posição algo, e repete o processo até todos os pontos estarem incluídos num certo local (Xia, *et al.*, 2020).

O algoritmo tem um princípio simples e uma velocidade de cálculo rápida. Normalmente, uns grandes números de pequenas regiões modais são formados após uma segmentação, levantando assim a hierarquia de análise de problemas do domínio pixel para o domínio da funcionalidade, o que simplifica o processamento subsequente (Hu, 2019).

A fase de agrupamento é dividida em duas. A primeira, conhecida como cluster baseado em pixéis mais próximos que os agrupa espacialmente em regiões homogêneas, de acordo com a sua proximidade espectral. O segundo clustering é um agrupamento baseado em segmentos que agrupa regiões espacialmente mais próximas formadas do primeiro agrupamento regiões homogêneas de acordo com a proximidade espectral.

A função multidimensional de kernel pode ser então expressa por:

$$K_{h_s h_r}(x) = \frac{c}{h_s^2 h_r^p} k\left(\left\|\frac{x^s}{h_s}\right\|^2\right) k\left(\left\|\frac{x^r}{h_r}\right\|^2\right) \quad 13$$

Onde h_s representa o coeficiente de largura de banda espacial e h_r o coeficiente de largura de banda espectral, então a informação dada por $k\left(\left\|\frac{x^s}{h_s}\right\|^2\right)$ representa a posição espacial, e $k\left(\left\|\frac{x^r}{h_r}\right\|^2\right)$ a posição espectral. K representa função de kernel, C a constante e p a dimensão da informação espectral no caso de Deteção Remota o número de píxeis que compõem as badas de entrada. Largura de banda espacial h_s é a distância espacial entre classes no domínio espacial, e indica o tamanho da janela espacial na segmentação. O atributo de largura de espectral h_r representa a diferença espectral entre classes no domínio espectral. O algoritmo de Mean-shift segmenta a imagem agrupando todos os pixels que estão mais perto de h_s no domínio espacial e h_r no domínio espectral e, em seguida, conecta os pontos de convergência. (Ming, *et al.*, 2015).

Os parâmetros de segmentação que podem ser variados para produzir resultados mais adequados são (Orfeo toolbox, 2019):

- *Spatial radius (spatialr)*: Raio espacial de vizinhança.

- *Range radius (ranger)*: Raio de alcance expresso na unidade radiométrica do espaço multiespectral.
- *Mode convergence threshold (thres)*: O esquema iterativo do algoritmo irá parar se o vetor de mudança média estiver abaixo deste limiar ou se o número de iteração atingir o máximo.
- *Maximum number of iterations (maxiter)*: Número máximo de iterações, o esquema iterativo do algoritmo vai parar se a convergência não for alcançada.
- *Minimum region size (minsize)*: Tamanho mínimo de uma região (em unidade pixel) em segmentação. Pequenos aglomerados serão fundidos com o aglomerado vizinho com a radiometria mais próxima. Se definido para 0 não é feito poda.
- *Simplify polygons (simplify.float)*: Simplifica os polígonos de acordo com uma determinada tolerância (em pixel). Esta opção permite reduzir o tamanho do ficheiro de saída.
- *Tiles size (tilesize)*: O tamanho das partições da imagem definidas pelo utilizador para segmentação por partes. O tamanho ideal da partição é selecionado de acordo com a RAM disponível.

O primeiro parâmetro, de vizinhança, implica que quanto maior for, maior foi a área de varrimento, o número tem de ser ímpar para ter um pixel central. O segundo, a tolerância de similaridade ou raio de alcance pode ser definida como a heterogeneidade que o operador considera tolerável, ou seja, o número de entrada vai corresponder ao número de níveis digitais que dentro do segmento se permitem ser diferentes. Ao escolher a tolerância de similaridade deve-se atender sempre à resolução radiométrica para se decidir o seu valor. Mas como não existe uma fórmula matemática, a fase de experimentação é fundamental, testando até se obterem resultados considerados ideais, para estes testes têm-se em conta o número de bandas utilizadas, «Quanto menor o limiar, mais homogêneos são os segmentos. Um limite maior causará um resultado de segmentação mais heterogêneo e generalizado.» (Clark Labs, 2020).

O número máximo de iterações representa o número de vezes que o cálculo foi repetido, quando mais forem, mais refinado será o resultado, porém terá custos

computacionais, por outro lado, se o limite de convergência estiver abaixo do valor estipulado pelo utilizador, o algoritmo irá terminar. O tamanho mínimo da região determina o tamanho mínimo do objeto a ser classificado e a simplificação de polígonos é utilizada para diminuir o espaço no disco, também poderá ser utilizado para obter uma imagem visualmente mais limpa dada a dimensão dos pixéis. Por fim o tamanho das partições implica decidir se o algoritmo deve tratar a imagem como um todo ou deverá fazer a segmentação por partes.

Como *input* foi utilizada a imagem com apenas as bandas de resolução nativa de 10m. Estas bandas como demonstrado em alguns estudos anteriores em OBIA (Marangoz, Sekertekin & Skçin, 2017; Belgiu & Csillik, 2018), são ideais para a segmentação de imagem *Sentinel-2* MSI e foram, através de testes, também aquelas que produziram melhores resultados de segmentação.

Após segmentada a imagem procedeu-se à validação dos segmentos através da análise visual em busca de sinais de sobre-segmentação ou sub-segmentação da imagem, se os resultados forem satisfatórios passa-se para a próxima etapa de OBIA, caso não sejam, volta-se a repetir o processo com novos parâmetros. Uma das preocupações foi que o limiar de tolerância de similaridade distinguisse edifícios industriais de cobertura não refletiva, como betão ou amianto do asfalto e também que fosse suficientemente detalhada para detetar, em aglomerações urbanas muito densas, objetos que compreendam vegetação.

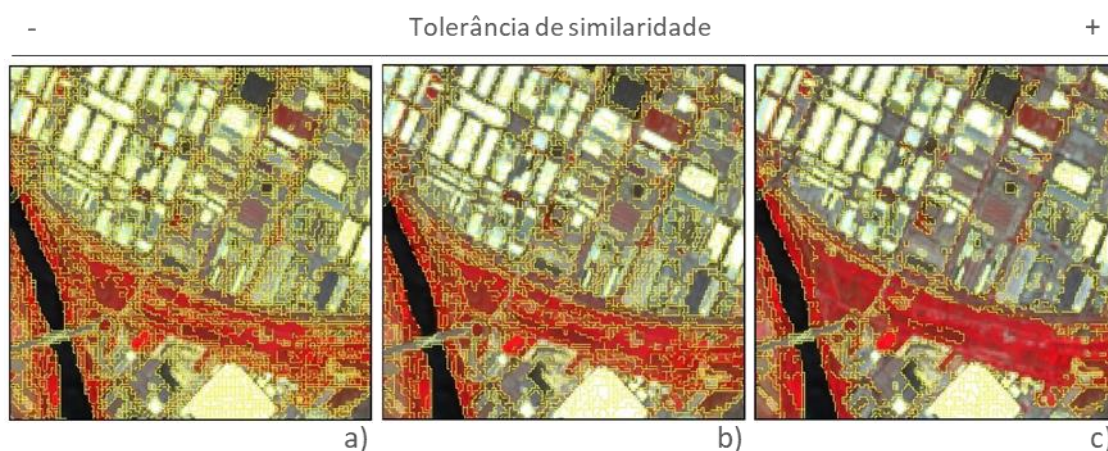


Figura 19 – Exemplos de segmentações, sobre-segmentação (a), segmentação ideal (b) e sub-segmentação (c) sobre a área industrial próxima do rio Llobregatt.

Esta validação resultou na necessidade de realizar várias tentativas, alterando parâmetros de segmentação. No total foram realizadas 7 segmentações com os parâmetros descritos na tabela a baixo. Foi necessário tomar nota dos parâmetros utilizados para esta área de estudo, dado que o processo foi repetido para teste sobre as imagens dos anos anteriores, mesmo que não sejam necessariamente são os mesmos valores, serviram de boa base de começo.

A fase de testes de segmentação é muito importante, já que não existe nenhuma formula matemática para definir que parâmetros vão produzir os melhores segmentos, é necessária esta fase de tentativa e erro. Cada um dos testes é inspecionado visualmente para analisar o que muda com a alteração de algum dos parâmetros e com isso chegar a melhor parametrização e consequentemente segmentação (Wang *et al.*, 2019).

Tabela 2 – Parâmetros teste para a segmentação de imagem utilizando o algoritmo MeanShift.

Parametro	Teste 1	Teste 2	Teste 3	Teste 4	Teste 5	Teste 6	Teste 7
Spatial radius	15	15	15	25	25	15	15
Range radius	250	350	450	400	350	350	150
Mode convergence treshold	0,1	0,1	0,1	0,1	0,1	0,1	0,1
Maximum number of iteractions	100	100	100	100	100	50	100
Minimum region size	1	2	2	5	5	4	4
8-neighbor connectivity	off	off	off	off	off	on	on
Stitch polygons	on	on	on	on	on	on	on
Minimum object size	1	1	1	5	5	4	4
Tiles size	1024	3605	1024	1024	450	1024	1024

A escolha da melhor combinação de parâmetros é feita à posteriori, ou seja, recorrendo ao método empírico de testar e avaliar resultados. Para averiguar a qualidade da segmentação é feita análise visual de imagem sobre a imagem segmentada com a imagem original e procurar visualmente semântica nos objetos. Se áreas homogéneas estiverem muito repartidas ou por outro lado áreas muito heterogéneas estiverem no mesmo segmento, repete-se a segmentação até obter um resultado ótimo, só aí se avançou para o processo de classificação automática supervisionada.

3. A recolha e validação de amostras

Amostras de treino ou *regions of interest* (ROI) consistem num conjunto de pixéis que servem de amostra espectral, para uma determinada ocupação do solo. Estas funcionam como amostras para a classificação, retirando dados estatísticos radiométricos destas regiões para depois inferir ao universo da imagem no processo de classificação. A sua qualidade define-se pela homogeneidade dos pixéis que esta inclui, e da divergência que esta assume em relação a outros ROI.

Foram extraídos segmentos, sobre os quais se alterou a tabela de atributos, isto é particularmente importante utilizando esta metodologia, pois é necessário, para a criação do modelo e da classificação final. Ao conceber cada classe, é necessário, adicionar na tabela de atributos o seu código respetivo, este deverá ser um número inteiro e deverá corresponder à sua classe. Para fazer corresponder criou se uma tabela de correspondências em que o código corresponde à classe, e no final do processo, juntar essa tabela aos valores previstos pelo algoritmo.

Recorreram se a 2 métodos para reconhecer e extrair as ROI, o primeiro envolve a utilização de imagens de muito alta resolução (ex. *Google Earth*, Ortofotomapas) para delimitar as áreas que irão servir de amostra para a classificação da imagem. Foram selecionados os segmentos mais representativos para cada uma das classes, o número mínimo de segmentos para cada amostra, exceto a classe Mar foi de 25, mas o objetivo para uma melhor precisão do algoritmo, foi a escolha de mais de 50 segmentos sempre que possível. A segunda fase consiste na validação no local, porém devido à pandemia Covid-19 tal não foi possível, pelo que se saltou este passo, realizando uma confirmação das classes no *Google Earth*.

Após se obterem e confirmarem as áreas de treino, estas foram validadas quanto à sua separabilidade espectral, isto é possível utilizando o algoritmo para a divergência transformada de Jeffries Matusita, que calcula a separabilidade espectral entre pares de amostras. De acordo com Jensen (2015) a divergência transformada de Jeffries-Matusita é uma transformação saturada da distância de Bhattacharyya que é dada por:

$$Bhat_{cd} = \frac{1}{8}(M_c - M_d)^T \left(\frac{V_c + V_d}{2}\right)^{-1} (M_c - M_d) + \frac{1}{2} \log_e \left[\frac{\frac{|V_c + V_d|}{2}}{\sqrt{|V_c| \cdot |V_d|}} \right] \quad 14$$

Em que as duas classes c and d são gaussianas e que M_c e M_d são as médias e as matrizes de covariância V_c e V_d estão disponíveis. Então a distância transformada de Jeffries-Matusita é dada por:

$$JM_{cd} = \sqrt{1 - e^{Bhat_{cd}}} \quad 15$$

Esses valores variam de 0 a 2 e indicam o quanto cada par de ROI's está estatisticamente separado. Valores maiores que 1,9 indicam que os pares de ROI têm boa separabilidade. Para pares de ROI com valores de separabilidade menores, devem se fazer alterações para melhorar a separabilidade, editando-as ou refazendo-as. Uma maior separabilidade aumenta as chances de uma melhor classificação (Richards, 2013). Para este passo, foi necessário recorrer ao *software* ENVI, que computa a separabilidade das amostras e devolve os valores de separabilidade entre cada uma das classes com cada uma das restantes classes.

Esta validação foi feita para o mesmo conjunto de amostras teste, utilizando três imagens diferentes, a primeira, contendo as bandas 8, 4, 3 e 2, a segunda adicionando o NDVI anual, e a terceira adicionando os 4 NDVI, um para cada uma das estações do ano, na tentativa de perceber o resultado que isto traz para a separabilidade espectral das amostras teste. O resultado ou a comparação entre cada uma das classes e todas as outras, totalizando 105 valores de separabilidade, para aferir melhorias na separabilidade compilou-se uma tabela para cada uma das três imagens (tabela 3), em que se compara quantas classes obtiveram valores inferiores a 1,6, entre 1,6 a 1,9, e superior a 1,9, esta última é o objetivo de separabilidade que se pretende.

O objetivo deste teste foi de aferir se é possível aumentar a separabilidade entre estas sem ter de alterar significativamente as amostras em si, dado que nem sempre é conveniente. Na tabela 3 é possível ver a melhoria da separabilidade nas amostras, o que significa que adicionar o NDVI não só beneficia a deteção da vegetação como também aumenta a separabilidade entre as outras classes.

Tabela 3 – Comparação dos valores agrupados de separabilidade das classes, para as combinações das bandas espectrais 8, 4, 3 e 2, estas combinadas com o NDVI anual, e as bandas espectrais com os 4 NDVI sazonais.

Separabilidade	B8432	B8432 + NDVI Ano	B8432 + NDVI Sazonal
0,0 - 1,6	11	4	0
1,6 - 1,9	27	21	11
1.9 - 2,0	67	80	94

Com base nestes resultados decidiu-se avançar para as classificações, isto tendo em conta que algumas classes não estão totalmente separadas entre si, mas que foram fundidas nos pós-classificação, como o caso dos edifícios com cobertura de baixa refletância com as vias de comunicação ou a vegetação arbórea florestal com a urbana. Em outras classes como a areia ou solo, que tem confusão espectral com algumas áreas urbanas, manteve-se, pois nem mesmo alterações substanciais nas amostras iriam produzir melhores resultados, dada a dificuldade em as separar no espectro eletromagnético. Existe confiança que a adição do NDVI também fez com que as sombras dos edifícios se separassem melhor dos corpos de água, minimizando erros.

4. A extração das estatísticas dos objetos

Este passo consiste numa operação em que se sobrepõe novamente os segmentos sobre a imagem e são calculadas as estatísticas para cada um deles. Este passo também foi feito para as ROI sobre o qual, e com base nas estatísticas foi possível criar um modelo de classificação de vetores, utilizado para classificar os segmentos totais baseada nestas.

A extração funciona sobrepondo os dados vetoriais sobre a imagem, e o *software* retira para cada um dos polígonos as estatísticas espectrais e morfológicas. Este processo foi feito para cada uma das três imagens, a imagem original com as bandas 8, 4, 3, 2, a imagem com o NDVI anual e a imagem com os quatro NDVI para cada uma das estações do ano. O resultado disto foi 3 ficheiros de ROI e três ficheiros de imagem segmentada.

As estatísticas foram devolvidas para cada segmento em cada uma das bandas de entrada, isto significa que cada segmento terá um conjunto de parâmetros estatísticos associado para cada uma das bandas espectrais, terá que se ter em atenção às bandas a utilizar na classificação. Os parâmetros estatísticos calculados foram a média (Mean), a mediana (Med), o desvio padrão (STD), a variância (Var), a curtose (Kurt) e a assimetria estatística (Skew), para além destas retira também informação morfológica dos segmentos, porém essa não foi incluída na fase de classificação.

5. O treino dos classificadores

As estatísticas extraídas para os objetos foram base onde o modelo de classificação operou. O *Software* traz consigo alguns modelos de classificação de vetores, na aplicação *Train Vector Classifier*, são eles: *LibSvm*, *Boost*, *Decision Tree*, *Artificial Neural Network*, *Normal Bayes*, *Random Forests*, *KNN*, *Shark Random forests* e *shark kmeans*. Para esta dissertação escolheu-se experimentar com dois destes modelos, o *Decision Trees* e o *LibSVM* e no final compará-los em termos de precisão temática.

O classificador *Decision trees* «é uma construção de árvore, descrito composto por um nó de raiz e uma série de nós internos e nós de folhas. Cada nó só pode ter um nó de pai e dois ou mais nó de criança. Os nós estão ligados uns aos outros por ramos. Cada nó interno corresponde a um grupo de propriedades ou propriedades de teste e cada lado corresponde a todos os valores possíveis de propriedade. O nó folha corresponde a um valor de propriedade de classe e diferentes nós de folhas podem corresponder ao mesmo valor de propriedade de classe.» (Jiang, *et al.*, 2011, p.354). O classificador *decision trees* presente no *OrfeoToolBox*, que foi utilizado para esta dissertação baseia-se nos algoritmos de classificação e árvore de regressão (CART) descritos em (Breiman, 1984) neste caso é uma árvore de classificação pois as classes são conhecidas à priori.

O funcionamento deste algoritmo é como uma árvore construída recursivamente, a partir do nó raiz. Todos os dados são usados para dividir o nó de raiz e seguidamente, para cada um dos nós, a regra de decisão ideal, ou seja, a melhor divisão "primária" é encontrada com base em critérios. Sendo que a estrutura de classificação definida pela árvore de decisão é estimada a partir de dados de treino ou amostras (Sharma, Ghosh & Joshi, 2013).

Uma Árvore de Decisão é construída a partir de um conjunto de amostras de treino, que consiste em objetos, cada um dos quais é completamente descrito por um conjunto de atributos e um rótulo de classe. Atributos são uma coleção de propriedades contendo toda a informação sobre um objeto. Ao contrário da classe, cada atributo pode ser ordenado (inteiro ou valor real) ou valores não ordenados (valor booleano) (Ghose *et al.*, 2010).

A forma como funciona o classificador é descrita em (Timofeev, 2004), como sendo construída de acordo com a regra de divisão, ou seja, a regra que executa a divisão da amostra de treino em partes menores e cada vez que isto acontece os dados são divididos em duas partes com a máxima homogeneidade que no caso da Detecção Remota e para esta dissertação em específico, trata-se da homogeneidade entre as estatísticas radiométricas dos segmentos.

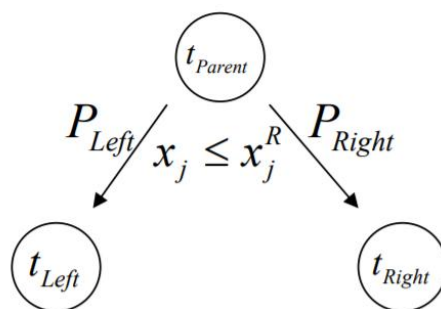


Figura 20 - O algoritmo de divisão das Árvores de classificação e regressão onde t_{parent} , t_{left} , t_{right} são respetivamente o nó progenitor, nó esquerdo e direito, x_j a variável j e x_j^R o melhor valor de divisão de x_j (Timofeev, 2004, p.9).

A homogeneidade máxima dos subnós é definida pela chamada função de impureza $i(t)$. Uma vez que a impureza do nó-mãe é constante para qualquer uma das

possíveis divisões a homogeneidade máxima dos nós de crianças esquerda e direita foi então equivalente à maximização da mudança da função de impureza $\Delta i(t)$.

$$\Delta i(t) = i(t_p) - E[i(t_c)] \quad 16$$

Onde t_p é o nó mãe e t_c é o subnó esquerdo e direito. Em seguida é necessário definir a função de impureza, duas são amplamente utilizadas, a função de divisão de Gini e a função de divisão de Twoing. Para a classificação experimentada nesta dissertação, o *software* utiliza a função de divisão de Gini que usa a seguinte função de impureza $i(t)$:

$$i(t) = \sum_{k \neq l} p(k|t)p(l|t) \quad 17$$

Onde k, l e t são os índices de classe e $p(k|t)$ a probabilidade condicional de classe k para o nó t . «A regra de divisão de Gini tenta encontrar a maior categoria homogénea dentro do conjunto de dados e isolá-la do restante dos dados. Os nós subsequentes são então segregados da mesma forma até que novas divisões não sejam possíveis» (Ghose *et al.*, 2010, p.94).

O procedimento divide novamente os nós esquerdo e direito e para cada novo nó o procedimento recursivo pode parar a divisão quando alguns critérios são cumpridos ou quando não é mais possível dividir o nó.

No *software* é possível definir alguns parâmetros para adequar a árvore de decisão às amostras recolhidas para produzir um modelo de classificação de objetos para a posterior classificação da imagem segmentada. Em baixo encontram-se listados os parâmetros e as descrições retiradas da documentação Orfeo Toolbox 7.1 (OrfeoToolbox, 2020).

- *Maximum depth of the tree* (Profundidade máxima da árvore): o algoritmo de treino tenta dividir cada nó enquanto a sua profundidade é menor do que a profundidade máxima possível da árvore. A profundidade real pode ser menor se os outros critérios de terminação forem cumpridos e/ou se a árvore for podada.

- *Minimum number of samples in each nod* (Número mínimo de amostras em cada nó): se o número de amostras num nó for menor do que este parâmetro, então o nó não foi dividido.
- *Termination criteria for regression tree* (Critério de finalização para a árvore de regressão): se todas as diferenças absolutas entre um valor estimado num nó e os valores das amostras de treino neste nó forem menores do que este parâmetro de precisão da regressão, então o nó não foi dividido mais.
- *Cluster possible values of a categorical variable into $K \leq \text{cat}$ clusters to find a suboptimal split* (Agregar possíveis valores de uma variável categórica em $K \leq$ aglomerados CAT para encontrar uma divisão sub-ideal): agrega possíveis valores de uma variável categórica em $K \leq$ aglomerados CAT para encontrar uma divisão sub-ideais.
- *Set Use1seRule flag to false* (Definir bandeira Use1seRule para falso): se for verdade, então uma poda foi mais dura. Isto tornará uma árvore mais compacta e mais resistente ao ruído dos dados de treino, mas um pouco menos precisa.
- *Set Truncate Pruned Tree flag to false* (Definir bandeira Truncar Arvore Podada para falso): se for verdade, então os ramos podados são fisicamente removidos da árvore.

O SVM utilizado para esta dissertação está incluído no LibSVM, presente no *software* OTB, esta biblioteca de *Support Vector Machines* tem sido desenvolvida desde o ano 2000 pela Universidade Nacional do Taiwan e é amplamente utilizada em *machine learning* e outras aplicações, como a Detecção Remota (Chang & Lin, 2019). O objetivo da SVM é produzir um modelo (com base nos dados de treino) que preveja valores alvo, neste caso a imagem segmentada, a partir apenas dos atributos dos dados de teste (Hsu *et al.*, 2003)

Support Vector Machines, são uma forma de *machine learning* ou modelos de aprendizagem supervisionados que são utilizados, à semelhança de *decison trees*, para a classificação e análise de regressão. Um algoritmo de formação SVM constrói um modelo classificador binário não probabilístico. Um modelo de SVM trata os exemplos

das categorias como uma representação de pontos no espaço, mapeados para que os exemplos das categorias separadas sejam divididos por uma clara lacuna o mais ampla possível para serem então mapeados nesse mesmo espaço e prever a sua categoria baseada no lado da lacuna em que se enquadram (Liu *et al.*, 2017).

Ajay *et al.* (2016) explica que o SVM constrói um hiperplano que separa as classes. Este algoritmo encaixa o classificador na sua margem máxima. Margem máxima garante mínimo erro de sobreencaxe. Em que o SVM pode ser compreendido se considerarmos duas classes de dados linearmente separáveis, x_i, y_i em que para cada vector x_i foi associado o nome da classe y_i . A equação para o hiperplano é dada por:

$$w^t x - y = 0 \quad 18$$

Onde $w = (w_1, w_2, w_3, \dots, w_n)^T$ e $x = (x_1, x_2, x_3, \dots, x_n)^T$ são vetores de características dimensionais no espaço real R^n . A função de decisão correspondente é dada por:

$$f(x) = \text{sign}(w^t x - y) \quad 19$$

Os dados são então mapeados em espaços dimensionais mais elevados utilizando *kernels* (K).

$$K(x_i, x_j) = \langle \varphi(x_i), \varphi(x_j) \rangle \quad 20$$

Em que φ é a função de mapeamento $\varphi: R^n \rightarrow R^k$ onde $\varphi(x) \in R^k$. Esta função também é utilizada para mapear os dados que não sejam não lineares.

O *LibSVM* contém diversas fórmulas para *support vector machine* e kernels, para esta dissertação foi utilizada a *kernel* linear, pois foi a que apresentou os melhores resultados na fase de testes com este classificador, a *kernel* linear é definida por:

$$K(x_i, x_j) = x_i^T x_j \quad 21$$

Dentro do *LibSVM* existem várias fórmulas para classificação, regressão e estimativa de distribuições. Para a dissertação foi o CSVC (*C-Support Vector Classification* ou Classificação do vetor de suporte C), descrito por Chang & Lin (2019) que descreve o funcionamento deste tipo de SVM.

Dado os vetores de treino $x_i \in R^n$, $i = 1, \dots, l$, em duas classes, e um vetor indicativo $y \in R^l$ como por exemplo $y_i \in \{1, -1\}$, o C-SVC resolve o seguinte problema de otimização primordial (Chang & Lin, 2019, citando Boser *et al.*, 1992; Cortes & Vapnik, 1995):

$$\min_{w, b, \xi} \frac{1}{2} w^T w + C \sum_{i=1}^l \xi_i \quad 22$$

$$\begin{aligned} \text{Sujeito a} \quad & y_i(w^T \phi(x_i) + b) \geq 1 - \xi_i \\ & \xi_i \geq 0, i = 1, \dots, l, \end{aligned}$$

Onde $\phi(x_i)$ mapeia x_i num espaço dimensional mais elevado e $C > 0$ é o parâmetro de regularização. Devido à possível alta dimensionalidade da variável vetorial w , normalmente é resolvido pelos autores com o seguinte duplo problema:

$$\min_{\alpha} \frac{1}{2} \alpha^T Q \alpha - e^T \alpha \quad 23$$

$$\begin{aligned} \text{Sujeito a} \quad & y^t \alpha = 0, \\ & 0 \leq \alpha_i \leq C, \quad i = 1, \dots, l, \end{aligned}$$

Onde se supõe a existência uma classe 1, em que $e = [1, \dots, 1]^T$ é o vetor de todos os uns, ou seja, a variável. Q é uma matriz semidefinida positiva l por l , $Q_{ij} \equiv y_i y_j K(x_i x_j)$ e $K(x_i x_j) \equiv \phi(x_i)^T \phi(x_j)$ é a função de *kernel*. Após ser resolvido utilizando a relação primal-dual, o w ótimo é:

$$w = \sum_{i=1}^l y_i \alpha_i \phi(x_i) \quad 24$$

E então a função de decisão é:

$$\text{sgn}(w^t \phi(x) + b) = \text{sgn} \left(\sum_{i=1}^l y_i \alpha_i K(x_i, x) + b \right) \quad 25$$

São armazenados $y_i \alpha_i \forall i, b$, nomes de legenda, que no caso do LibSVM pode ser qualquer número inteiro, vetores de suporte, e outras informações, como parâmetros de *kernel* para construir o modelo de previsão (Chang & Lin, 2019).

A escolha destes parâmetros aqui explicados teve por base o método empírico de tentativa e “erro”, neste caso, a análise dos resultados preliminares permitiu aferir que a melhor parametrização para a utilização do LibSVM, ou seja, a utilização do C *Support vector classification* com uma *kernel* linear.

As opções presentes na documentação para Orfeo Toolbox 7.1 para a parametrização do LibSVM (OrfeoToolBox, 2020) são:

- *SVM Kernel type* (Tipo de kernel SVM)
- *Linear*: não é feito mapeamento, esta é a opção mais rápida.
- *Gaussian radial basis function* (Função de base radial gaussiana): é uma função exponencial da distância euclidiana entre os vetores.
- *Polynomial* (Kernel polinomial): o mapeamento é uma função polinomial. O kernel é uma função de tangente hiperbólica dos vetores.
- *Type of SVM formulation* (Tipo de formulação SVM):
- *C support vector classification* (Classificação do vetor de suporte C): esta formulação permite uma separação imperfeita das classes. A penalidade é definida através do parâmetro de custo C.
- *Nu support vector classification* (Classificação do vetor de suporte Nu): esta formulação permite uma separação imperfeita das classes. A penalidade é definida através do parâmetro de custo Nu. Em comparação com C, Nu é mais difícil de otimizar, e pode não ser tão rápido.
- *Distribution estimation (One Class SVM)* (Estimativa de distribuição (SVM de classe única)): Todos os dados de treino são da mesma classe, SVM constrói um limite que separa a classe do resto do espaço de recurso.
- *Cost parameter C* (Parâmetro de custo C): os modelos SVM têm um parâmetro de custo C (1 por defeito) para controlar a compensação entre erros de treino e forçar margens rígidas.
- *Cost parameter Nu* (Parâmetro de custo Nu): na gama 0.1, quanto maior o valor, mais suave a decisão.
- *Parameters optimization* (Otimização de parâmetros).
- *Probability estimation* (Estimativa de probabilidade).

Na ótica de escolha do algoritmo a aplicar para as classificações finais, foram feitos alguns testes em que se variaram parâmetros para a construção do modelo de classificação de vetores, esta parametrização encontra se explícita na tabela 4 para a *Decision tree* e na tabela 5 para o *LibSVM*.

Para além dos testes apresentados nas tabelas, existiram testes preliminares cujo objetivo foi o domínio do *software* e a análise das suas capacidades, bem como a seleção dos dois algoritmos a testar para a classificação de objetos, para estes testes não existe uma tabela de parametrização, pois os parâmetros utilizados foram sempre os parâmetros padrão do *software*.

Tabela 4 – Parametrização para a criação do modelo de classificação de vetores utilizando o algoritmo *Decision Tree*.

Parâmetro	Bandas 8,4,3 e 2			Bandas 8,4,3 e 2 + NDVI		Bandas 8,4,3 e 2 + 4NDVI	
	Teste 1	Teste 2	Teste 3	Teste 4	Teste 5	Teste 6	Teste 7
<i>Maximum depth of the tree</i>	10	10	15	10	15	10	15
<i>Minimum number of samples in each node</i>	10	2	2	2	2	2	2
<i>Termination criteria for regression tree</i>	0,01	0,01	0,01	0,01	0,01	0,01	0,01
<i>Cluster possible values of a categorical variable into K</i>	10	5	10	5	10	5	10
<i>Set Use1seRule flag to false</i>	False	True	False	True	False	True	False
<i>Set TruncatePrunedTree flag to false</i>	False	False	False	False	False	False	False

Tabela 5 – Parametrização para a criação do modelo de classificação de vetores utilizando o treino de vetores LibSVM

Parâmetro	Bandas 8,4,3 e 2				Bandas 8,4,3 e 2 + NDVI			Bandas 8,4,3 e 2 + 4NDVI	
	Teste 1	Teste 2	Teste 3	Teste 4	Teste 5	Teste 6	Teste 7	Teste 8	Teste 9
<i>SVM Kernel Type</i>	Gaussian radial basis function	Gaussian radial basis function	Linear	Linear	Linear	Linear	Linear	Linear	Linear
<i>SVM Model Type</i>	Nu support vector classification	C support vector classification	C support vector classification	C support vector classification	C support vector classification	C support vector classification	C support vector classification	C support vector classification	C support vector classification
<i>Cost Parameter C</i>	1	1	1	1	1	1	1	1	1
<i>Cost Parameter NU</i>	0,5	0,5	0,5	0,5	0,5	0,5	0,5	0,5	0,5
<i>Parameters optimization</i>	False	False	False	True	True	True	True	True	True
<i>Probability estimation</i>	False	False	False	True	False	True	False	True	False

Apesar de nos primeiros dois testes de classificação com o *LibSVM* terem produzido resultados insatisfatórios após análise visual, o terceiro teste, com a *Kernel* linear produziu resultados que largamente ultrapassaram em qualidade os obtidos por *decision tree*, sendo essa a razão para a existência de mais testes de parametrização com este e a sua eventual escolha para a produção do mapa de classificação do solo.

Manteve-se registro dos parâmetros dos classificadores para poder recriar os resultados em futuras classificações. As tabelas anteriores dizem respeito à classificação feita para área de Barcelona para o ano de 2019, o modelo com melhor resultado para o algoritmo *LibSVM* foi aplicado aos testes de classificação para os anos de 2018, 2017 e 2016.

6. A classificação e nomeação de objetos

Esta é a etapa em que, com base no modelo, o *software* procura fazer a previsão das classes para os objetos (Segmentos). Para tal utilizou-se a aplicação *vector*

classifier que, segundo o *cookbook* da *Orfeo toolbox*, faz a classificação dos dados de acordo com o ficheiro do modelo produzido pelo passo anterior. A característica da saída de dados vetoriais contém as etiquetas de classe decididas pelo classificador (rótulo de classe máxima = 65535) (Orfeo toolbox, 2020). Estas etiquetas são números inteiros, cada um correspondente a uma classe, como para esta classificação foram recolhidas 15 classes, com um número associado à sua respetiva classe, foram utilizados 15 valores para a saída dos dados vetoriais.

Existem dois modos:

- Modo de atualização: que adiciona o campo 'cfield' contendo a classe prevista no ficheiro de entrada.
- Modo de escrita: copia os campos existentes do ficheiro de entrada no ficheiro de saída e adiciona o campo 'cfield' contendo a classe prevista.

O modo de escrita foi bastante útil, pois a repetição do processo não requer que se altere o ficheiro original, o que foi fulcral na fase de testes ter a opção de criação de um novo ficheiro de objetos classificados.

O resultado é um ficheiro shapefile, com todos os atributos estatísticos dos segmentos com uma coluna extra que contém o número de classe previsto. Como o *software* não faz a ligação entre o número de classe e a classe propriamente dita foi preparada foi feita a correspondência dos valores previstos com os nomes das classes para representação em mapa.

7. A validação temática

Para validação temática da classificação orientada por objetos recorreu se à Matriz de confusão, que identifica a natureza dos erros e as suas quantidades, pois é preciso ter uma base que sustente que a classificação corresponde ao máximo aquilo que se encontra no território estudado. Então para isto é necessário avaliar a exatidão temática. Isto consiste em comparar a classificação com dados geográficos reais para determinar a precisão da classificação, para tal é necessário normalmente recorrer a

trabalho de campo, ou a imagens aéreas de muito alta resolução com finalidade de recolher novas amostras sobre áreas diferentes que não se sobreponham às ROI para cada classe classificada. Estas novas amostras vão ser base para uma comparação que vai permitir avaliar qual é a precisão do classificador e das amostras utilizadas nessa classificação, no entanto é necessário atender a alguns aspetos antes de avançar para a validação em si.

Para a validação dos resultados recorrendo à matriz de confusão, nesta dissertação tiveram-se em conta as recomendações dadas por Congalton (1991), que explica os fatores a ter em consideração para dar à validação uma significância mais robusta. Estas recomendações incidem sobre a recolha das amostras no terreno, o esquema de classificação, a autocorrelação espacial, o tamanho da amostra e o esquema de amostragem descritos em seguida.

A recolha das amostras no terreno para a validação deve ser tida em conta na altura da validação temática, ou seja, a qualidade das amostras de validação que devem de ter em consideração a complexidade da imagem que se pretende classificar bem como o esquema de classificação. É importante de ter em conta para a escolha dos dados ground truth, por exemplo, para se obter dados de referência para uma classificação que pretenda apenas distinguir duas classes abrangentes é mais simples do que obter dados para classes mais específicas. Para a distinção entre urbano e florestal no primeiro caso, pode-se utilizar uma carta de ocupação do solo já existente, porém para a distinção entre tipos de plantações, que requerem medidas exatas no terreno para produzir amostras de qualidade para a validação.

O segundo parâmetro a ter em conta é o esquema da classificação, este deverá ser mutuamente exclusivo, dado pelo cálculo da equação da divergência transformada de Jeffries Matusita, e exaustivo, isto é, cada área a ser classificada deverá ficar apenas numa única classe, e todas as áreas devem ser incluídas na classificação. É também vantajoso se existir uma hierarquia, de modo a que algumas categorias possam ser agregadas para formarem classes mais abrangentes algo que acontece com as categorias na classificação, em que por exemplo, as classes vegetação arbórea 1 e vegetação arbórea 2 são agregadas para formar uma classe que abrange toda a

vegetação arbórea da presente na imagem classificada. Por fim, o esquema de classes deve poder ser logicamente explicado.

A «autocorrelação espacial diz-se que ocorre quando a presença, ausência ou grau de uma determinada característica afeta a presença, ausência ou grau da mesma característica nas unidades vizinhas» (Congalton, 1991, p.43, citado de Cliff and Ord, 1973). A autocorrelação afeta a suposição da independência da amostra e pode então ser responsável pela periodicidade nos dados que possam afetar os resultados de qualquer tipo de amostra sistemática o que presume que cada novo pixel não contribui com informações independentes, e por esta razão, a recolha de amostras para a validação terá sempre em conta a autocorrelação espacial, isto é, a não sobreposição entre amostras de treino e amostras de validação.

Para a escolha do tamanho da amostra de validação, o autor nota a importância que existe no balanço entre o que é estatisticamente sólido e o que na prática pode ser conseguido, então propõe através da utilização do método empírico a utilização de 50 amostras por classe, porém se a área exceder um milhão de acres (500km²) ou o número de classes for superior a 12, devem ser recolhidas pelo menos 75 amostras para cada classe. Porém este número não é fixo, podendo o operador optar por recolher mais amostras de classes mais importantes, ou menos amostras de classes que tendem a ser mais constantes como corpos de água, para esta dissertação, recolheram-se as amostras, com uma média de 76 por cada classe, favorecendo as classes mais representativas em termos de área, o número de amostras por classe está presente na tabela 3.

Embora a regra de ter 75-100 amostras por classe seja apenas uma abordagem empírica, deve ser favorecida. Este método prevê medidas suficientes em cada estrato para cálculos posteriores especialmente a matriz de confusão e o cálculo do índice de kappa que requerem um número suficiente de amostras (Banko, 1998).

Por fim o esquema de amostragem também deve ser tido em conta conforme as características da classificação, ainda segundo Congalton (1991), a opinião de como isto deverá ser feito é variável entre os investigadores, pois cada forma de recolha de

amostras de validação aleatória traz consigo problemas associados, especialmente no que toca à dimensão – representatividade – importância das classes.

Nesta dissertação optou se, no entanto, pela recolha manual através de análise visual de imagem sobre imagens de muito alta resolução para a recolha das amostras de validação, isto é particularmente vantajoso por permite extrair as amostras que mais se adequam ao esquema de classificação utilizado, e efetivamente mitigar o problema entre a dimensão que a classe ocupa em área e a sua representatividade amostral. Após todos os passos anteriores terem sido considerados, calculou-se as matrizes de confusão, precisão global e kappa.

Tabela 6 – Classes e número de amostras extraídas para validação temática

Classe	Amostras de validação
Vias	85
Vegetação Arbórea florestal	150
Vegetação herbácea vigorosa	75
Vegetação herbácea senescente	55
Vegetação Arbórea misturada	65
Urbano Residencial	151
Solo	70
Plásticos e pisos sintéticos	20
Matos e descampados	75
Mar	150
Rios, Lagos e Lagoas	50
Coberturas refletivas	75
Coberturas escuras	75
Áreas húmidas	25
Areal	35
Total	1156

A matriz de confusão parte da premissa que numa classificação perfeita, todos os pixels dessa classe em linha (classificados) corresponderão a ela em coluna (Reais). «Os valores das células diagonais fora da diagonal principal representam o número de pixels mal classificados sob a forma de erros de omissão e erros de comissão. Quanto maior for o valor dessas células ou mais células têm um valor que não zero, menor é a precisão da classificação» (Gao, 2009, p.513).

A Precisão do produtor quantifica quantos elementos identificados no terreno como uma certa classe o são também no mapa produzido, enquanto que o erro de omissão corresponde aos pixels que não o são. Já a precisão do utilizador quantifica elementos classificados no mapa, quais estão corretamente identificados de acordo com o terreno, ou seja, as amostras “reais” o que permite definir o erro de comissão como os elementos que estão classificados no mapa que não correspondem à realidade de acordo com a amostra real (Richards, 2013).

Para além desta análise existem os dois índices globais de precisão da classificação. O primeiro, a precisão global do mapa de classificação «é determinada dividindo o número total de pixels corretamente classificados (isto é, a soma da diagonal na matriz de erro) pelo número total de pixels na matriz de erro (N)» (Jensen, 2015, p. 569 citado de Story & Congalton, 1986; Congalton & Green, 2009) e é calculada através da soma do número de pixels classificados corretamente dividido pelo número total de pixels, o que resulta na seguinte expressão:

$$Precisão\ global = \frac{\sum_{i=1}^K x_{ii}}{N} \quad 26$$

O Coeficiente de Kappa analisa a matriz de confusão para produzir um índice de concordância tendo em conta erros de omissão e comissão, logo mais abrangente e utilizado. Este coeficiente varia entre 0 e 1, sendo que 1 é a concordância perfeita.

Este índice é dado pela seguinte expressão:

$$\hat{K} = \frac{\sum_{i=1}^r (X_{ii} - \sum_{ci} \sum_{ri})}{N^2 - \sum_{i=1}^r (\sum_{ci} \sum_{ri})} \quad 27$$

Em que N é o número total da amostra, r os valores em linha e em coluna, X_{ii} é o número de observações em linha i e coluna i , \sum_{ci} e \sum_{ri} correspondem ao total marginal da coluna e da linha respetivamente.

Os valores K superiores a 0.80 representam uma forte concordância ou precisão entre o mapa de classificado e a informação de referência no terreno ou *ground truth*. Valores de Kappa entre 0,40 e 0,80 representam um acordo moderado e por fim valores a baixo dos 0.40 representam uma fraca concordância (Jensen, 2015).

41°50'63"N

1°92'16"E

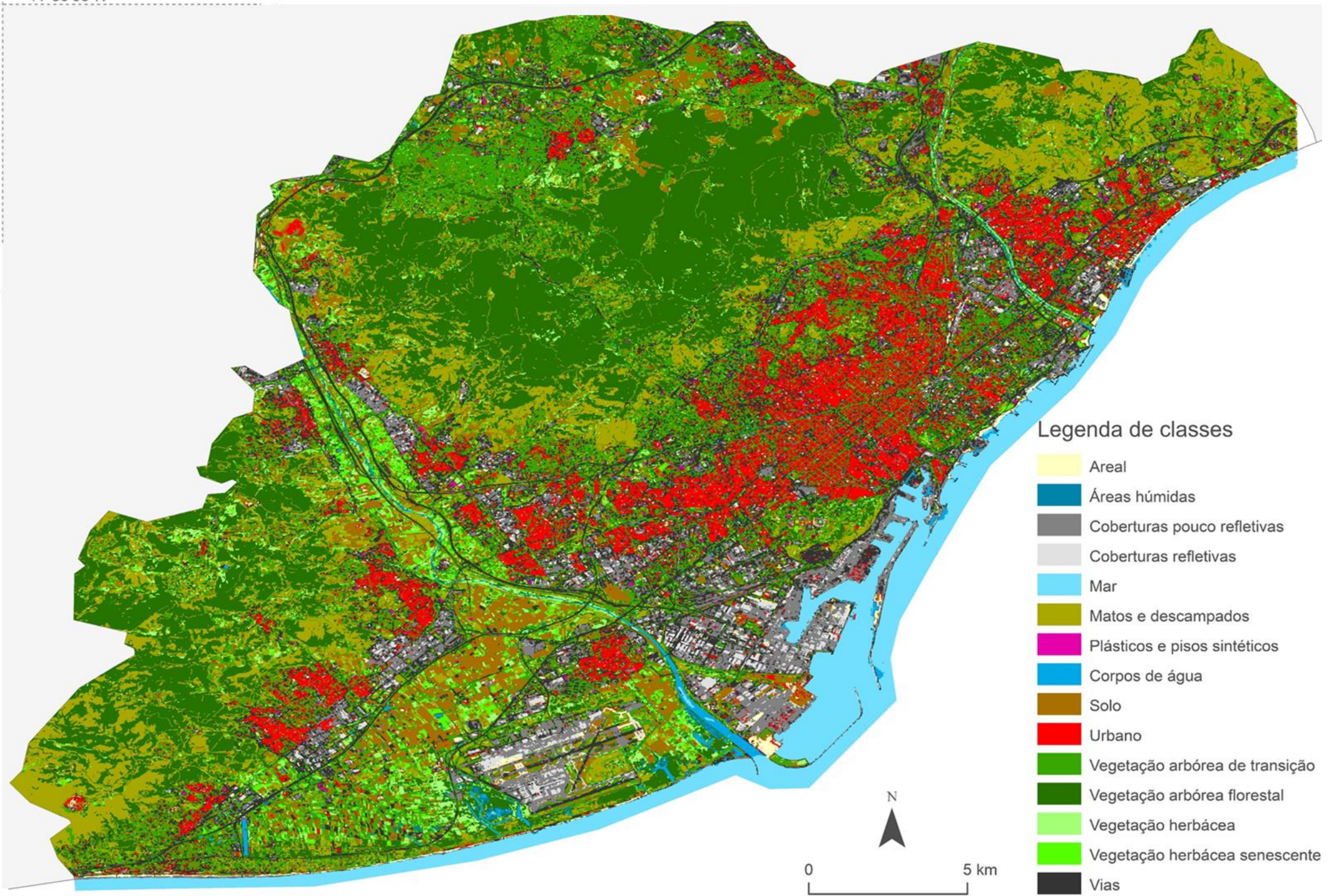


Figura 21 – Classificação escolhida após a aplicação dos métodos de OBIA. Originada pela aplicação do classificador LibSVM sobre as bandas 8,4,3 e 2 e os quatro NDVIs sazonais.

A classificação orientada por objetos resultou numa classificação da imagem constituída por vetores que correspondem a usos do solo, neste caso se chamou de classificação primária ou em bruto (figura 21), dado este mapa depois das validações temáticas contém ainda muitas imprecisões, apesar de um índice de *kappa* de 0.82, pelo que foi alvo de uma intervenção manual a que se chamou de operações de mitigação de imprecisão temática. As operações de mitigação visam evitar imprecisões que afetem as classes de uso do solo relevantes para este estudo.

8. A matriz de admissibilidade e mitigação da imprecisão temática

Para gerir a correção de imprecisões temáticas foi elaborada uma matriz, a que se chamou de matriz de admissibilidade de imprecisão temática (MAIT), esta tem duas funções, primeiramente serviu para a decisão sobre que classes agregar e a segunda a correção em si.

A elaboração de uma matriz de admissibilidade de imprecisão temática (tabela 9) é feita relacionando todas as classes e determinando, com base na investigação feita sobre o mosquito bem como as classes que se pretendam isoladas a admissibilidade ou inadmissibilidade de classificação incorreta de objetos.

A matriz é semelhante a uma matriz de confusão, mas trata se de uma matriz lógica, em que os parâmetros são definidos pelo utilizador, mas a sua construção parte do princípio de comparação de todas as classes com todas as classes, de forma a incrementar a eficiência na retificação de imprecisões graves.

A MAIT surgiu para a resolução de dois problemas, o primeiro são as imprecisões na classificação, que podem levar a erros na construção do modelo, podendo apontar áreas de transmissão potencial onde não existem ou omitir áreas onde estas realmente existem. Porém o fator tempo e poder de processamento têm que ser tidos em conta, e para tal, terá que haver um filtro de classes que seja necessária à sua correção ou não. Isto permite a minimização de falhas no modelo derivadas de imprecisões na classificação, enquanto alivia a carga temporal e de processamento.

Tabela 7 – Matriz de admissibilidade de imprecisão temática.

Classe	Vias	Vegetação Arborea1	Vegetação herbácea 1	Vegetação herbácea 2	Vegetação Arborea 2	Urbano	Solo	Plásticos e pisos sintéticos	Matos e descampados	Mar	Rios, Lagos e Lagoas	Coberturas refletivas	Coberturas pouco refletivas	Áreas húmidas	Areal
Vias	x	I	I	I	I	A	D	A	I	I	I	A	A	I	D
Vegetação Arbórea 1	I	x	D	D	A	I	I	I	D	I	I	I	I	D	I
Vegetação herbácea 1	I	D	x	A	D	I	D	I	D	I	I	I	I	D	I
Vegetação herbácea 2	I	D	A	x	D	I	A	I	A	I	I	I	I	D	I
Vegetação Arbórea 2	I	A	D	D	x	I	I	I	D	I	I	I	I	D	I
Urbano	A	I	I	I	I	x	D	A	I	I	I	A	A	I	D
Solo	D	I	D	A	I	D	x	I	A	I	I	I	I	I	D
Plásticos e pisos sintéticos	A	I	I	I	I	A	I	x	I	I	I	A	A	I	D
Matos e descampados	I	D	D	A	D	I	A	I	x	I	I	I	I	I	I
Mar	I	I	I	I	I	I	I	I	I	x	I	I	I	I	I
Rios, Lagos e Lagoas	I	I	I	I	I	I	I	I	I	I	x	I	I	D	I
Coberturas refletivas	A	I	I	I	I	A	I	A	I	I	I	x	A	I	D
Coberturas pouco refletivas	A	I	I	I	I	A	I	A	I	I	I	A	x	I	D
Áreas húmidas	I	D	D	D	D	I	I	I	I	I	D	I	I	x	I
Areal	D	I	I	I	I	D	D	D	I	I	I	D	D	I	x

As três classes foram definidas pelo grau de aceitação da confusão espectral, e podem ser definidas como:

- Admissível, que ocorre quando duas classes que se confundem espectralmente e originaram classificação de objetos uma na outra não afetem o resultado pretendido. Estes erros não serão alvo de retificação;
- Difícilmente aceitável que ocorre quando apesar de não comprometer o resultado final, a classificação errônea poderá levar a problemas indiretos na utilização da classificação;

- Inadmissível ocorre quando se compromete diretamente o resultado pretendido, um exemplo disto seria a mistura de classes entre urbano e corpos de água.

Utilizou-se a MAIT numa primeira instância para a fusão de classes que originou a classificação primária reclassificada (figura 22), e foi sobre as novas classes que foram feitas as correções à imprecisão temática, isto tornou o processo mais simples dada a dimensão da área de estudo.

41°50'63"N

1°92'16"E

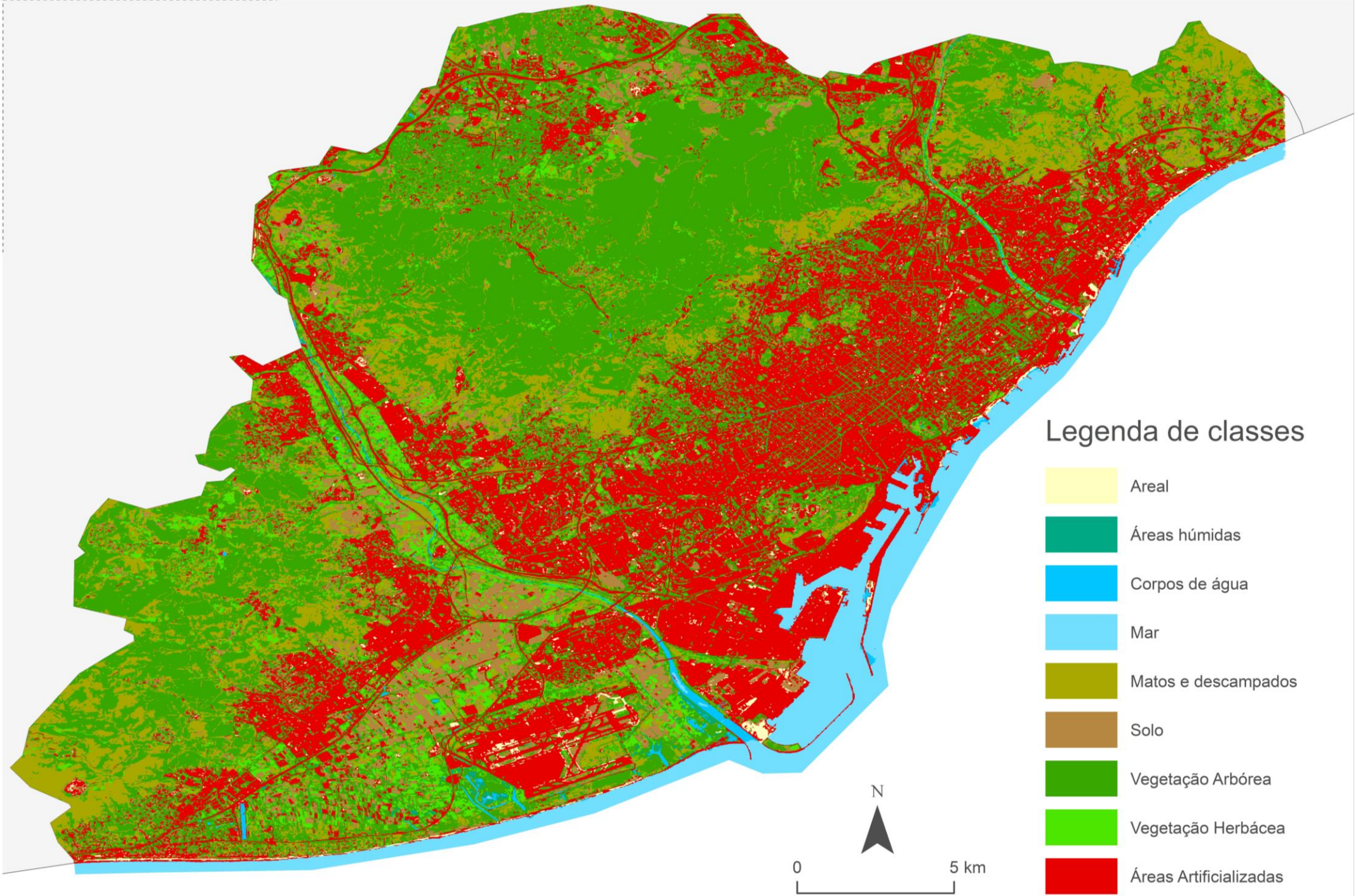


Figura 22 – Reclassificação originada pela fusão de classes cuja confusão espectral se admite e que se encontrem dentro da mesma temática para correção.

A Classificação originada pela aplicação da reclassificação através da fusão de classes, criou classes através da agregação das Vias, Urbano, Plásticos e pisos sintéticos, Coberturas refletivas e Coberturas pouco refletivas para criar a classe Territórios artificializados. Com as Classes vegetação arbórea 1 e 2 a classe Vegetação arbórea e as classes vegetação herbácea 1 e 2 deram origem à classe vegetação herbácea.

Obtendo uma classificação com a temática e as classes já definidas, foi feita a correção temática em maior detalhe. O método utilizado para a correção de imprecisões foi por decomposição de classes, isto significa que após a fusão de classes, foram gerados para cada classe um ficheiro vetorial, estes ficheiros vetoriais mais leves, permitiram um trabalho mais fluido mesmo para os mais pesados como áreas artificializadas e vegetação arbórea.

Por questões de tempo optou-se por não alterar a geometria dos polígonos apenas se alterando a classe do polígono na tabela de atributos. Isto reduziu as imprecisões temáticas de polígonos que tenham ficado na sua totalidade mal classificados, porém não reduziu as imprecisões geométricas resultantes da segmentação, podendo existir ainda polígonos que compreendam duas classes diferentes que no mundo real que deveriam estar separadas, mas não estão, neste caso atribuiu-se a classe mais relevante para a presença de *Aedes albopictus*.

As imprecisões de classificação são não um fenómeno incomum quando se trabalha com classificações automáticas, aliás, a imprecisão é sempre esperada, a análise do mapa anterior permite em primeira instância denotar a descontinuidade do rio El Llobregat, a subclassificação em áreas urbanas dispersas, áreas classificadas como áreas urbanas e vice-versa. Estas imprecisões originaram na escolha das estatísticas que alimentaram o classificador, a média e o desvio padrão. O resultado disto é que classes com assinaturas espectrais próximas, como areal e algumas áreas urbanas, ou então áreas húmidas e vegetação herbácea ou arbórea se confundam exemplificado no esquema seguinte.

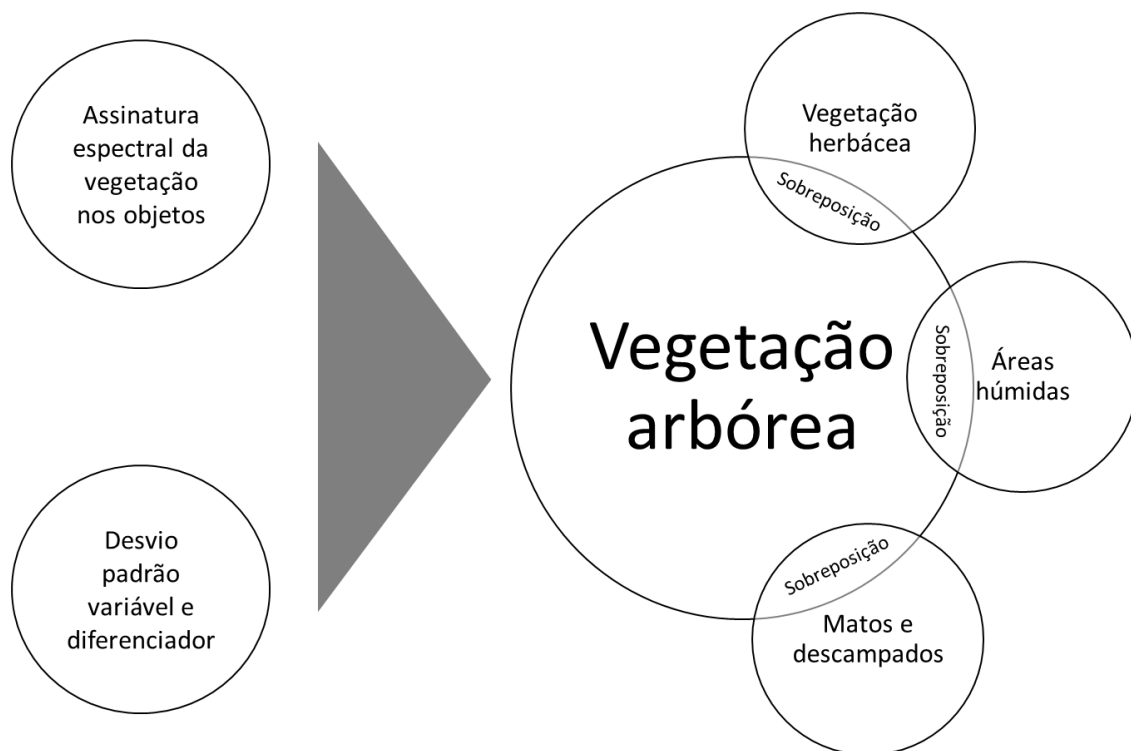


Figura 23 – Exemplificação de confusão por sobreposição de classes, esta sobreposição ocorre quando objetos que por natureza partilham uma assinatura espectral parecida, são geralmente diferenciados pelo desvio padrão, no entanto quando este também é similar, resultante da morfologia do terreno ou existência de misturas causadas por diferentes assinaturas no mesmo objeto, estas podem surgir sobrepostas e classificadas como uma classe diferente.

Exemplos práticos disto podem ser encontrados por toda a classificação, o esquema anterior exemplifica onde ocorreram mais imprecisões, estas imprecisões estão classificadas como dificilmente aceitáveis, o que significa que sempre que seja notada uma imprecisão clara, esta foi corrigida. Pontualmente quando em dúvida sobre a classificação de um objeto optou-se pela sua não alteração, ou seja, quando não houve a possibilidade de definir a classe correta, ou então quando foram objetos muito pequenos contíguos a objetos maiores da mesma classe em que tenham sido classificados.

Imprecisões consideradas inadmissíveis foram sempre corrigidas dentro do que foi possível, não existiu um critério para a sua não correção e o limite de correções efetuadas foi definido apenas pelo tempo disponível. Para auxiliar na correção recorreu

se ao mapa de usos e coberto do solo proveniente do *Departamento de Territorio y Sostenibilidad* (2019), este ficheiro contém classes que foram reclassificadas para as compatibilizar com a classificação produzida. A figura 24 dá um exemplo do impacto da mitigação de imprecisão temática para a foz do El Llobregat.

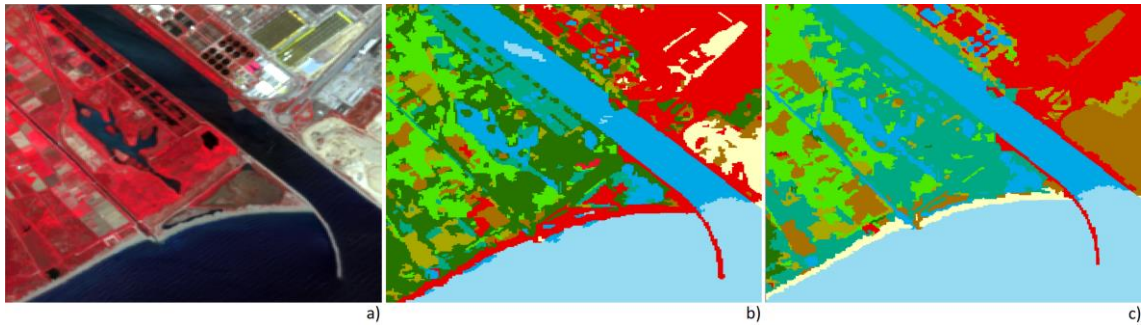


Figura 24 – Exemplo de correção de imprecisões temáticas para a área da foz do El Llobregat. As principais diferenças encontram se nas áreas de areal c) que tinham sido classificadas como urbano b), bem como áreas húmidas c) que foram consideradas vegetação arbórea b), ou áreas de solo a descoberto ou rocha c) classificadas como areal b). Esta área é bastante complexa com várias tipologias de uso do solo, bem como dinâmica, pois é afetada pelas marés, é possível denotar isto pelo corpo de água exacerbado em ambas as classificações b) e c) mas não presente na composição de imagem a).

As intervenções mais expressivas centraram se nas áreas urbanas dispersas ou com uma grande densidade de áreas verdes, nesta classificação torna se então possível identificar as localidades circundantes a Barcelona que estavam demasiado fragmentadas na classificação anterior, porém em áreas urbanas densas este problema não foi expressivo. A classe de areal também verificou grandes problemas de classificação, porém a sua solução foi bastante simples, fazendo uma seleção sobre as áreas à beira mar, foi possível separar aquilo que era areia de facto daquilo que eram áreas artificializadas, através de um *query* simples foi possível corrigir esta situação. Por fim, vegetação herbácea de solo, esta situação característica das áreas agrícolas tem na sua génese a variabilidade anual do estado da vegetação nestes espaços, assumiu que pelo menos durante uma parte do ano estas classes correspondiam à realidade, pelo que não se alterou a classificação atribuída ao objeto.

CAPÍTULO V

A Análise exploratória dos dados

Este capítulo dedica-se à análise e exploração com base nos resultados obtidos na extração do coberto vegetal bem como pela Object Based Image Analysis. É também estudada a relação espacial entre as variáveis as ocorrências de Aedes albopictus com a finalidade de se elaborar uma proposta de corredores de propagação potencial com base nas variáveis extraídas.

CAPÍTULO V: A análise exploratória dos dados

1. O conceito de análise exploratória de dados

Propor mapas de modelação espacial de probabilidades é sempre um trabalho complexo. No entanto, este trabalho pretende lançar as bases para a criação de cartografia de corredores de propagação potencial utilizando as variáveis extraídas por Detecção Remota trabalhadas em ambiente SIG. Neste caso, optou-se por deixar de fora as variáveis climáticas remetendo-as para trabalhos futuros e adicionaram-se as vias em vetorial pela sua maior precisão geométrica e temática.

No decorrer da revisão bibliográfica referente ao *Aedes albopictus*, sabe-se que este se adapta a ambientes com altos valores de atividade clorofilina e áreas urbanas. Sabe-se também que este pode ser transportado acidentalmente por veículos, pelo que as vias de comunicação poderão desempenhar um papel importante na sua disseminação. Então foi realizado um exercício de análise exploratória dos dados disponíveis.

A análise exploratória de dados espaciais (Exploratory Spatial Data Analysis, ESDA) é uma extensão da análise exploratória de dados, que se centra explicitamente nas características específicas dos dados geográficos. É uma técnica cada vez mais popular em Sistemas de Informação Geográfica (SIG) e permite aos utilizadores descrever e visualizar distribuições espaciais, identificar locais atípicos ou outliers espaciais, descobrir padrões de associação espacial, agrupamentos ou pontos quentes, e sugerir regimes espaciais ou outras formas de heterogeneidade espacial (Dall'èrba, 2009). Neste caso houve um enfoque na descoberta de padrões de associação espacial à semelhança do que foi feito com a associação entre a distância à vegetação e o número de ocorrências.

A ESDA é realizada sem conhecimento prévio de interações de processos ou padrões, e baseia-se em grande parte na perceção do observador utilizando métodos visuais e gráficos. É frequentemente utilizada como um primeiro passo para ajudar na

formação de hipóteses (Wagner, 2018, citado de Unwin, 1996; Anselin, 1999). Os métodos utilizados basearam-se na análise visual da distribuição de ocorrências, e a formulação de gráficos de regressão entre estas e a distância às variáveis consideradas, estes foram feitos em software não SIG, dado que os SIG estão muitas vezes mal equipados no que toca a muitas destas ferramentas, apesar de terem funcionalidades básicas para permitir a sua implementação (Haining & Wise, 1997) que se verificou nos softwares utilizados, o que significou a utilização de softwares de estatística para a análise preliminar de correlações entre variáveis, e apenas depois puderam ser feitas operações espaciais e representação em ambiente SIG.

2. Os corredores de propagação potencial

Existem vários modelos de representação de probabilidades, muitos bastante complexos, no entanto foi tomada uma abordagem mais simples e exploratória dos dados e das possibilidades. Para a proposta exploratória de corredores de propagação potencial seguiu-se uma lógica semelhante à produção de cartografia de risco, em que se tem em conta que a probabilidade de estabelecimento é tanto maior quanto mais adequadas forem as condições de estabelecimento e que para que se introduza uma espécie é necessário um agente de propagação, neste caso as vias de comunicação. A combinação de ambos junta a probabilidade de introdução por transporte acidental com a probabilidade de fixação da espécie em dado território.

Para a camada de probabilidade de estabelecimento, extraíram-se as distâncias contínuas em relação à vegetação e às áreas urbanas. Por motivos de prazos, não foram selecionadas apenas as áreas artificializadas habitadas, mas trabalhou-se com o total da classe extraída por OBIA.

Para a criação do mapa de distâncias que contenha a distância à vegetação, assumiu-se cada classe como zero, mantendo os valores de distância a partir destas, subtraindo então os dois mapas um pelo outro. Assumindo que, os valores de zero correspondem em cada um dos mapas de entrada, vegetação, ou áreas urbanas,

respetivamente, a subtração de uma imagem pela outra teria como resultado um mapa com distâncias para ambas as direções (figura 31). O resultando é mapa contínuo de distâncias a áreas de fronteira onde o coberto vegetal e áreas artificializadas se conjugam (figura 32). Para obter este mapa foi aplicada uma fórmula que visou manter as distâncias e eliminar valores negativos resultantes da subtração.

A fórmula é dada por:

$$\sqrt{(Distância\ à\ vegetação - Distância\ a\ áreas\ urbanas)^2} \quad 28$$

E o resultado pode ser simplificado pela figura seguinte.

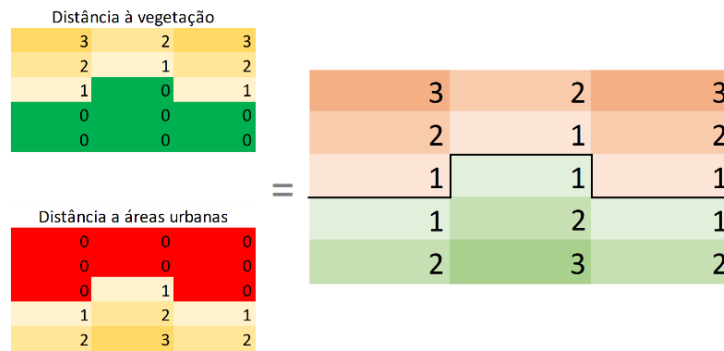


Figura 25 – Esquema de exemplo do resultado do cálculo da distância bilateral em relação à fronteira onde se conjugam as duas variáveis.

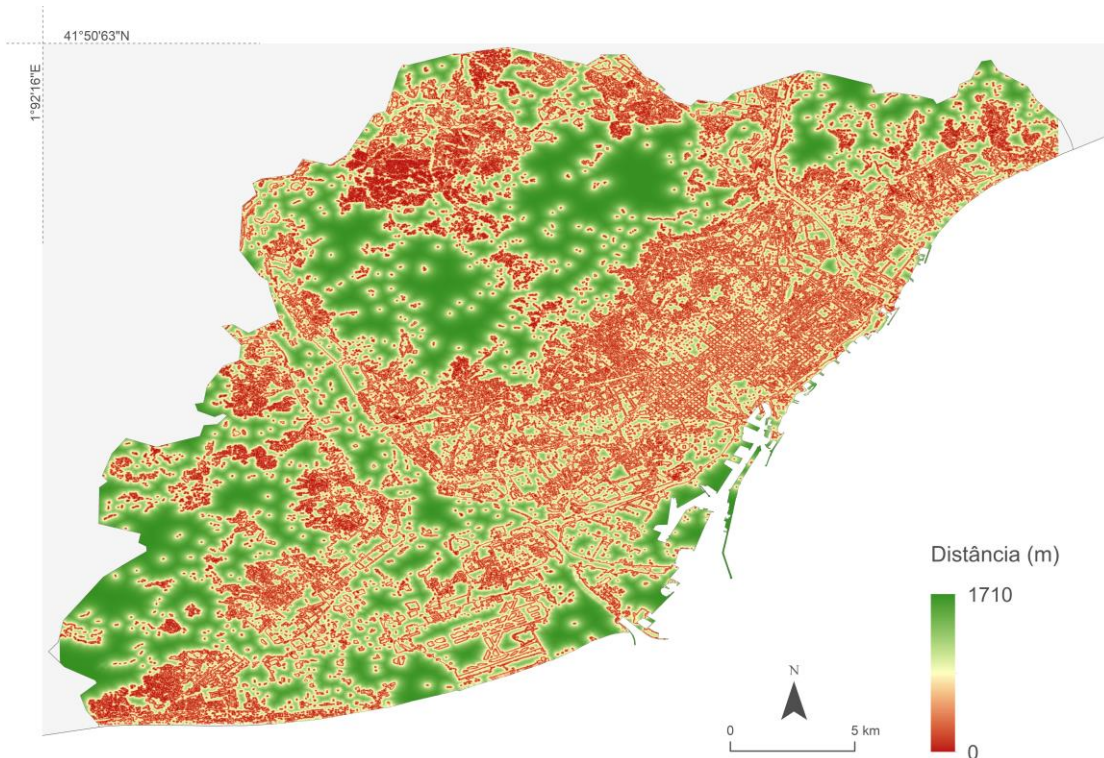


Figura 26 – Mapa de valores de distância contínua em relação à fronteira onde se intersectam as variáveis vegetação e áreas urbanas.

A camada das ocorrências foi depois utilizada para extrair os valores de distância, estes foram depois agregados em grupos de frequências, o que originou uma tabela de número de ocorrências por cada 10 metros de distância em relação a esta fronteira onde se conjugam as duas variáveis (tabela 14).

Tabela 8 – Frequência de ocorrências em função de classes de 10 metros de distância.

Distância (m)	Frequência (n)
10	392
20	167
30	109
40	69
50	42
60	34
70	26
80	4
90	7
100	2
110	1
120	3
130	1
140	3

Esta tabela de frequências foi então utilizada para calcular a regressão exponencial da frequência de ocorrências em função da distância à fronteira onde se encontram as áreas artificializadas e a vegetação. Como teste aos resultados, foram também gerados aleatoriamente mais 860 pontos e calculada a respectiva regressão que servirá de controlo aos resultados.

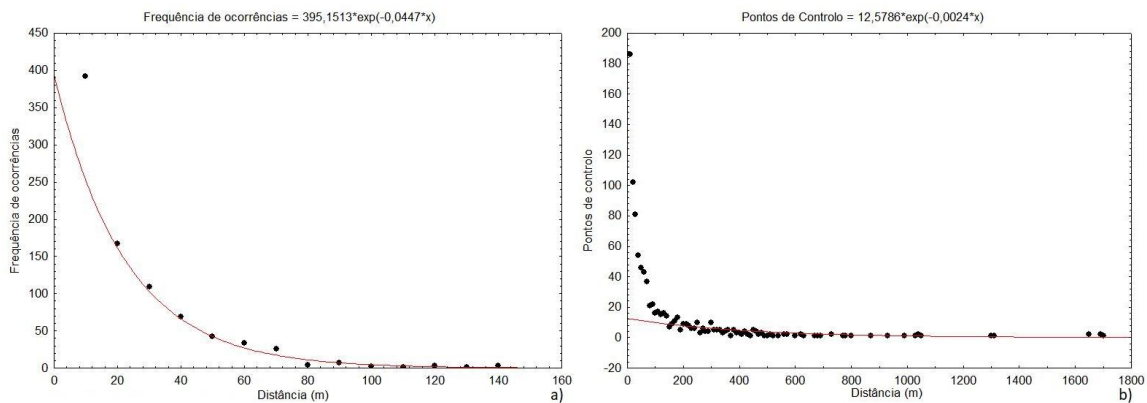


Figura 27 – Regressão exponencial do número de observações (a), e pontos de controlo (b) em função da distância a áreas urbanas e vegetação.

Para a regressão entre a distância e a frequência de ocorrências obteve-se $R^2=0,55$ e $r=-0,74$, com $p=0,0025$ e, num intervalo de confiança a 95%. Os resultados destas estatísticas significam a existência de uma correlação negativa forte entre a distância à fronteira, podendo-se rejeitar a hipótese nula, o que reforça a hipótese de que a distância a esta fronteira tem efeito sobre o número de ocorrências.

Os valores de R^2 demonstram que não existe um poder preditivo muito elevado, o que é explicado pela complexidade de previsão dos fatores que influenciam a distribuição de espécies como fatores ambientais como a temperatura e a precipitação bem como pela recolha dos dados que possivelmente se relacionará com a densidade populacional. Decidiu-se aceitar os resultados, mantendo reservas quando às conclusões.

Os resultados ao utilizar os 860 pontos de controlo mostraram que esta relação não resultante do acaso, mesmo que tal não signifique dependência. As estatísticas devolvidas para estes pontos são um $R^2=0,16$ e um $r=-0,4$, com $p=0,0003$ para um intervalo de confiança a 95%.

Os testes feitos com as regressões mostram que a mais explicativa dos dados é a regressão exponencial como pode ser observado na figura 33, pelo que se decidiu pela sua utilização para a produção de um mapa de valores previstos, este servirá como camada de probabilidade de estabelecimento de *Aedes albopictus*.

Para a criação de um mapa de valores previstos de ocorrências em função da distância à vegetação e áreas urbanas, utilizou-se então a função de regressão exponencial. Na função exponencial y (variável dependente) é dada pelo expoente x (variável independente) do coeficiente a multiplicado pelo declive elevado ao expoente e pode ser definida por:

$$y = a \times b^x \quad 29$$

Ou seja, a função de previsão de ocorrências em função da distância à fronteira é:

$$y = 395,1513 \times 0,0447^x \quad 30$$

Aplicou se esta função para geração de dados espaciais de previsão de ocorrências baseados na distância à fronteira onde se conjugam as variáveis, vegetação e urbano.

Para representação da probabilidade de ocorrências, a fim de criar os corredores de propagação potencial, foi então necessário criar classes em função dos resultados anteriores (figura 34). Para estas, reclassificaram-se os valores previstos em função da distância, utilizou se o princípio exponencial, em que a primeira classe terá cerca de metade dos valores previstos, a segunda terá $\frac{1}{4}$ e assim sucessivamente até à quinta classe, que terá o mesmo número de valores previstos, mas cobrirá uma área muito maior. Optou-se por atribuir graus de probabilidade à totalidade da área de estudo, dado que fatores externos como vento e transporte acidental podem quebrar a barreira de voo dos mosquitos, aumentando-a.

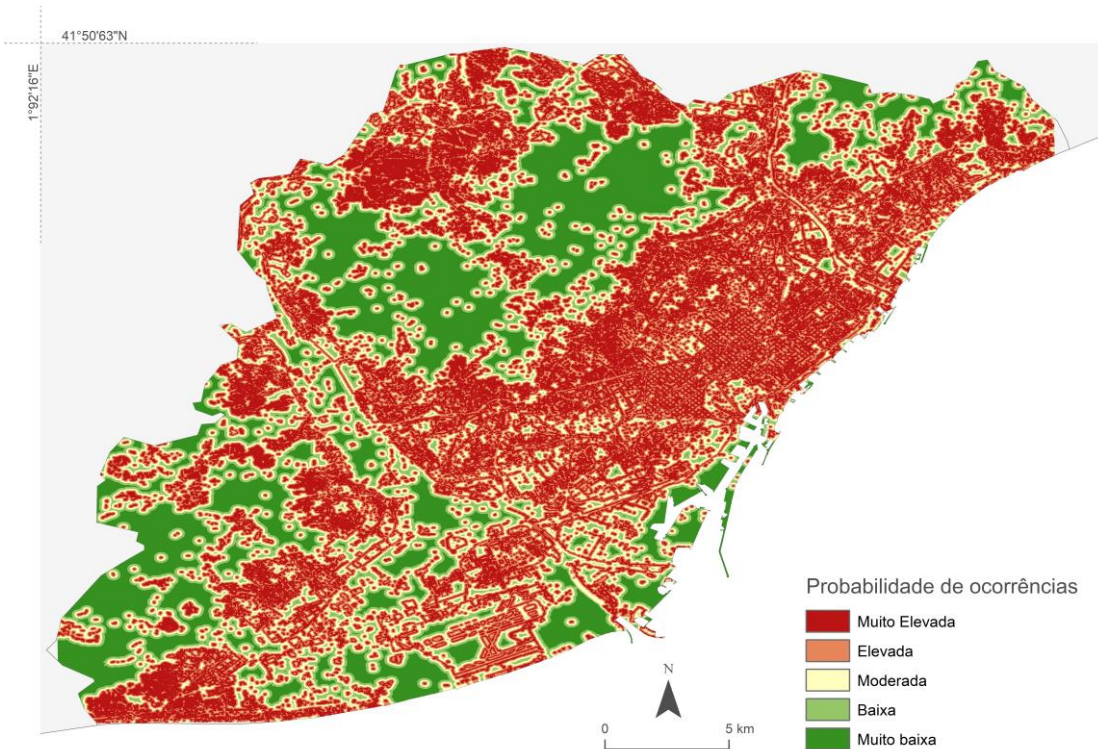


Figura 28 – Probabilidade de ocorrências em função da previsão obtida pela regressão exponencial entre as ocorrências e a distância às variáveis de habitat.

O mapa resultante representa a probabilidade de estabelecimento de *Aedes albopictus* em função da presença de vegetação em conjugação com áreas urbanas ajustado à sua distância de voo. Para a proposta de corredores de propagação, conjugou-se o fator de estabelecimento com o fator de transporte ou introdução.

O fator de introdução da espécie em novas áreas escolhido foram as vias de comunicação, pois as «infraestruturas de transportes também podem atuar como "corredores" para a dispersão natural da biodiversidade não nativa. Estas infraestruturas facilitam a dispersão de organismos não nativos, permitindo o seu movimento através de barreiras físicas e ambientais» (Ascensão, & Capinha, 2017, p.67).

Para a criação das classes para a distância em relação às vias utilizou se informação oficial das vias do *Departamento de Territorio y Sostenibilidad (2020)*, sobre a qual também se calculou a distância contínua em metros a partir destas. Aqui optou-se por não utilizar os resultados das vias extraídas por OBIA por motivos de precisão, dado que a classificação automática dificilmente diferencia vias de outras áreas

construídas ou por outro lado, pela resolução espacial do sensor, algumas vias não foram classificadas.

Para aferir se existe alguma relação entre as observações de *Aedes albopictus* e a presença de vias de comunicação foi também calculada a regressão entre as classes de distância às vias e a frequência de ocorrências, e o devido controlo com pontos aleatórios.

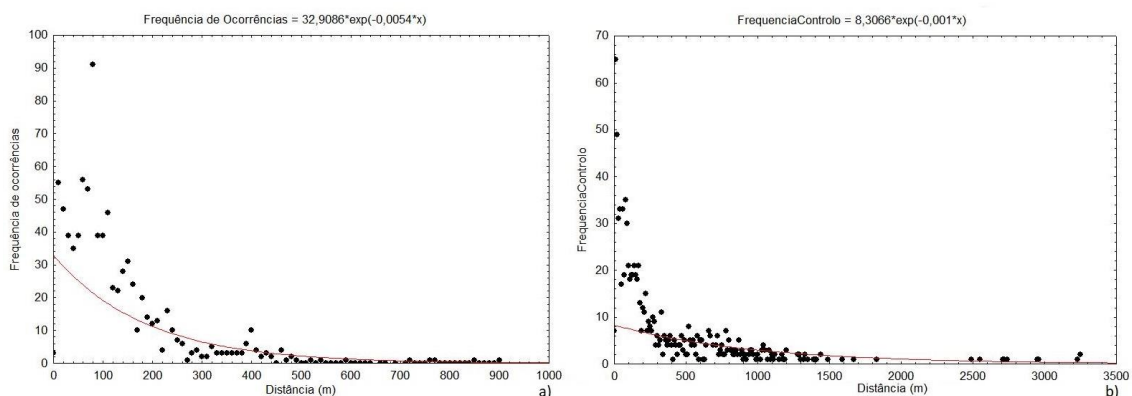


Figura 29 – Regressão exponencial do número de observações (a), e pontos de controlo (b) em função da distância a vias de comunicação.

Para a regressão entre a distância e a frequência de ocorrências obteve-se $R^2=0,48$ e $r=-0,70$, com $p=0,0000$ e, num intervalo de confiança a 95%. Para a regressão entre a distância e a frequência de pontos de controlo obteve-se $R^2=0,23$ e $r=-0,48$, com $p= 0,000000003$ e, num intervalo de confiança a 95%. Em ambos os casos, é rejeitada a hipótese nula, dando significância estatística aos resultados. Em relação às vias e obteve se um é a menor diferença entre as correlações entre variáveis e ocorrências/variáveis pontos de controlo.

Apesar da correlação ser elevada foi descartado o seu uso para a criação de um mapa de valores previstos, por motivos de índole teórica e surge do facto de esta prever muito para além dos 200 metros, pelo que se reclassificou diretamente sobre os valores de distância atribuindo-lhes um grau de probabilidade de introdução accidental. Gerando se classes de distância com base nos resultados da regressão exponencial. A classe 5 até 20 metros de distância a segunda até 60 metros, a terceira até 120 metros, a quarta até aos 200 metros (limite da capacidade de vôo natural) e a quinta todos os outros valores com a classe 1 (figura 36).



Figura 36 – Mapa de classes de distância às vias de comunicação como fatores de probabilidade de introdução de *Aedes albopictus*.

Para criar a proposta de corredores de propagação potencial, assumiu-se que os fatores de uso do solo/estado de superfície dão a probabilidade de fixação em função dos fatores ambientais e as vias de comunicação a probabilidade de introdução, dada pelas classes de distância a estas. Partiu-se então do princípio de que se o vetor é introduzido numa área menos favorável, a probabilidade deste se fixar é menor do que se este for introduzido numa área que reúna as suas condições de habitat.

Com apenas duas variáveis, optou-se por uma sobreposição não ponderada das camadas de probabilidade de ocorrências (em função dos fatores ambientais) e da probabilidade em função das classes de distância (em relação às vias de comunicação). Ambas as camadas de entrada já se encontravam classificadas de 1 que é muito elevado no caso da probabilidade e de menos que 20m no de distância às vias, a 5 que é muito baixa no caso da probabilidade e superior a 200m de distância em relação às vias, pelo que a sua combinação terá a mesma classificação, originando valores de probabilidade de propagação.

CAPÍTULO VI

A análise e discussão de resultados

*Este capítulo dedica-se à análise, interpretação e avaliação dos resultados obtidos na extração do coberto vegetal e os obtidos pela Object Based Image Analysis. É também estudada a relação espacial entre o coberto vegetal multitemporal com as ocorrências de *Aedes albopictus* de forma a estudar as correlações existentes entre estes. Analisam se também os resultados obtidos pela análise exploratória dos dados.*

CAPÍTULO VI: A análise e discussão de resultados

Os resultados da aplicação dos métodos de extração de informação foram ao encontro do esperado pelo que foi feita uma análise com base nestes resultados de forma a produzir informação pertinente da sua relação com a presença de *Aedes albopictus*.

Os resultados da análise multitemporal do coberto vegetal originaram um mapa de vegetação anual, quatro mapas de vegetação sazonal e 12 índices de vegetação mensais, o mesmo se obteve para os anos de 2018 e 2017. Por limitações de visibilidade em imagens do ano de 2016, a análise multitemporal do coberto vegetal originou um mapa de vegetação para um segundo semestre, dois mapas sazonais para o verão e outono e seis índices mensais.

O resultado da análise multitemporal da vegetação para produção cartográfica originou um mapa bastante preciso de presença de vegetação, contendo não só as grandes manchas florestais da área de estudo como também as pequenas áreas verdes presentes por toda a malha urbana da cidade de Barcelona. Estas pequenas áreas verdes que se distinguem das áreas urbanas foram fundamentais para perceber a utilidade em utilizar o NDVI no processo de OBIA de forma a isolar estas manchas da restante malha urbana que originou uma melhoria substancial na precisão temática da classificação devido ao seu poder diferenciador em algumas classes.

Os mapas extraídos para os quatro anos foram então utilizados para testar a correlação com as ocorrências anuais de *Aedes albopictus*. Optou por se incluir os anos anteriores para aumentar o tamanho da amostra, isto foi benéfico para a análise e um reforço para o cálculo da correlação. O mapa anual de 2019 foi também utilizado em conjunto com a classificação para a proposta de corredores de propagação potencial.

A análise do mapa originado (figura 25) mostra as áreas em que a média dos NDVI para os 12 meses é superior a 0,3. É possível distinguir os valores mais elevados, mais próximos a 0,9, em vegetação permanente, como as florestas e margens de cursos de água, assim como algumas áreas verdes de irrigação permanente, os valores mais baixo ou próximos do 0,3 são maioritariamente encontrados em áreas agrícolas e áreas de transição e urbanas, os valores centrais são maioritariamente atribuídos a matos.

41°50'63"N

1°92'16"E



Figura 31 – Mapa de vegetação para o ano de 2019 originado pelo cálculo da média para os doze meses do ano, e posterior extração de valores superiores ao limiar proposto.

Os resultados das correlações foram promissores e bastante consistentes para os quatro anos analisados, para além de mostrarem uma correlação negativa entre o número de ocorrências em relação com a distância à vegetação, o que significa que, para os 4 anos estudados, em todos houve uma tendência decrescente de observações quanto maior a distância entre as áreas com NDVI superior a 0,3 e o restante território. Estes resultados indicam que a média aritmética como medida para a extração do coberto vegetal em multitemporal foi uma boa escolha, no entanto não invalida testes futuros utilizado outras formas de análise multitemporal.

Apesar da forte correlação assume-se sempre que a correlação não significa dependência e que as variáveis que contribuem para a presença de *Aedes albopictus* são muitas, dada a elevada adaptabilidade da espécie em ambientes urbanos. A vegetação por si só não explica tudo, porém reforçam os resultados encontrados por outros autores que fizeram trabalhos similares com outras espécies de *Aedes* (Kogan, 2020; Lopes *et al.*, 2006; Lourenço *et al.*, 2011).

Quando confrontada a existência de vegetação no contexto urbano como parques, logradouros e a atividade do vetor é visualmente perceptível um certo grau correlação espacial. A sobreposição das observações de *Aedes albopictus* sobre o mapa de vegetação produzido para 2019 permite visualizar a vegetação e as ocorrências no espaço, nota-se uma porção significativa de ocorrências em algumas áreas verdes como exemplificado na figura 26.



Figura 32 – Exemplo de sobreposição entre as ocorrências e áreas de maior atividade clorofilina em contexto urbano para o Parc de la Ciutatella e áreas circundantes.

A distribuição do número de ocorrências em áreas de NDVI superior a 0,3 ou nas suas imediações, é superior em todos os anos esta a 50% de todas as ocorrências para uma distância de até 20 metros, com especial destaque para 2017 onde esta diferença foi mais demarcada. Para obter o valor médio fez-se a média para a vegetação e aplicaram-se as operações ao conjunto de dados para os 4 anos. O ano de 2018 como aquele em que existiram mais registos tem um peso não indicador da média para os 4 anos.

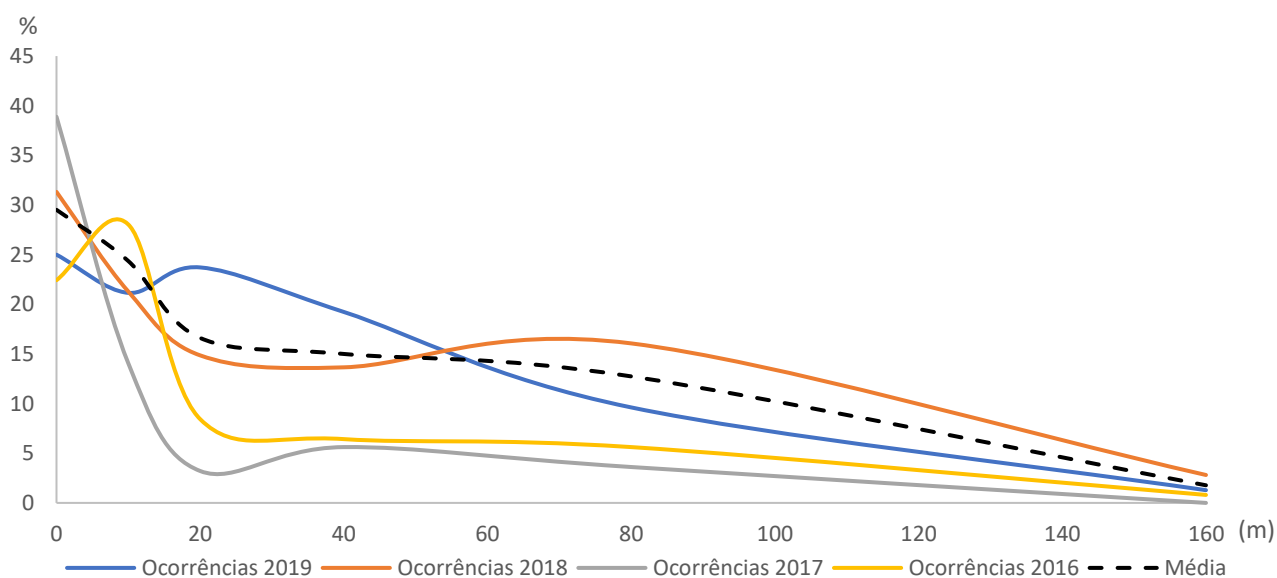


Figura 33 – Distribuição do número de ocorrências (%) em função da distância (m) à vegetação

As correlações foram feitas para todos os anos individualmente e em conjunto, a correlação maior é observada (tabela 8) para o ano de 2019, enquanto que 2017 apresenta a correlação mais fraca, porém ainda acima de - 0.75 pode se afirmar que esta ainda é uma correlação negativa forte. Para a média dos quatro anos, a correlação é de -0,83 ou negativa muito forte.

Tabela 9 – Coeficiente de Correlação de Pearson entre o número de ocorrências e a distância à vegetação, uma correlação negativa indica que quanto maior a distância menor o número de ocorrências.

Ano	2019	2018	2017	2016	Total
r Pearson	-0,98	-0,84	-0,76	-0,87	-0,93

Para complementar os estudos entre a relação entre as ocorrências de *Aedes albopictus* e a vegetação, foi ainda calculada a regressão exponencial entre estas e a distância à vegetação bem como o R^2 e a respetiva significância estatística p , repetindo-se o procedimento para 860 pontos de controlo gerados aleatoriamente de forma a aferir se os resultados derivam da existência de áreas vegetadas em praticamente toda a área de estudo.

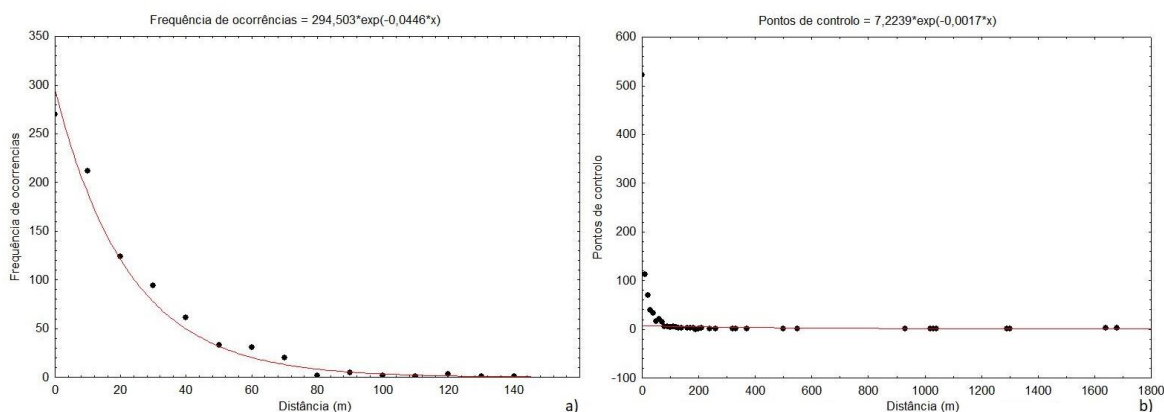


Figura 34 – Regressão exponencial de ocorrências a) e de pontos aleatórios b) em função da distância à vegetação.

Para a regressão entre a distância e a frequência de ocorrências a) obteve-se $R^2=0,69$, com $p=0,0001$, num intervalo de confiança a 95%. Isto significa que % das ocorrências pode ser explicada pela proximidade a vegetação, e que se pode rejeitar a hipótese nula neste caso; o que ajuda a suportar a hipótese que a presença de vegetação ter uma relação com o número de ocorrências.

Os resultados obtidos pela regressão entre a distância e a frequência de pontos de controlo são de $R^2=0,04$ e $r=-0,22$ com $p=0,19$, num intervalo de confiança e 95% demonstram uma clara baixa no valor de associação, com uma significância estatística que não permite rejeitar a hipótese nula. As diferenças substanciais entre as ocorrências e os pontos gerados aleatoriamente mostram que as relações observadas não se devem a aleatoriedade das ocorrências em função da existência de muitas áreas de vegetação com limiar de intensidade de NDVI superior a 0,3. Apesar disto, mantêm-se as reservas existentes em relação aos problemas associados aos dados.

Os resultados da análise orientada a objetos produziram uma classificação de uso do solo para as classes estipuladas com uma imprecisão aceitável para todas as classificações. A adição dos índices multitemporais de vegetação para as 4 estações originaram melhores resultados em termos de precisão temática, confirmando a sua utilidade na separabilidade entre classes amostrais, que pode ser aumentada via alteração das bandas de entrada.

Os resultados do índice de Kappa (tabela 9) variaram entre aproximadamente 0,65 com algoritmo *Decision Trees* sobre as bandas 8, 4, 3, 2 e NDVI anual e aproximadamente 0,84 para o classificador LibSVM sobre as bandas 8, 4, 3, 2 e os quatro NDVIs sazonais, pelo que se optou pela utilização deste último na proposta de corredores de propagação potencial. O índice de precisão global teve em geral resultados similares, variando cerca de 0,2 em todos as classificações (tabela 10).

Tabela 10 – Índice de Kappa para as classificações efetuadas

Kappa	Bandas 8,4,3,2	Bandas 8,4,3,2 + NDVI Ano	Bandas 8,4,3,2 + NDVI Sazonal
LibSVM	0.773042	0.809697	0.836971
Decision Trees	0.700326	0.654473	0.744392

Tabela 11 – Precisão Global para as classificações efetuadas

Overall Acc.	Bandas 8,4,3,2	Bandas 8,4,3,2 + NDVI Ano	Bandas 8,4,3,2 + NDVI Sazonal
LibSVM	0.792769	0.826279	0.85097
Decision Trees	0.72575	0.680776	0.766314

Foi ainda testada a aplicação do modelo originado pelas assinaturas espectrais das amostras recolhidas para o ano de 2019 no classificador para imagens segmentadas para outros anos. Resultando assim em mais três classificações para os anos de 2018, 2017 e 2016, para imagens do mesmo mês, de forma a minimizar diferenças entre as estatísticas espectrais que originaram o modelo e as estatísticas dos segmentos, pelo que se deu preferência a imagens de 22 de julho quando possível. Os resultados foram satisfatórios, porém observou se uma diminuição global das métricas de validação temática (tabela 11).

Tabela 12 – Índice de kappa e precisão global para os três anos anteriores utilizando LibSVM.

Métrica de precisão	2016	2017	2018
Kappa	0.705791	0.73469	0.741627
Overall Acc	0.731041	0.757496	0.763668

A classificação selecionada para a elaboração da proposta dos corredores de propagação potencial, com uma concordância temática muito forte, foi submetida a correção de imprecisões. Esta classificação do uso do solo resultante foi bastante precisa para algumas classes, porém a confusão espectral entre algumas é inevitável pelo que correções adicionais foram implementadas. Recorrendo à MAIT foi feita uma fusão de classes, foi possível visualizar onde o classificador foi preciso e onde ele falhou ao atribuir as classes corretas aos vetores. Estas falhas de classificação foram corrigidas através de edição manual.

Para tal, de acordo com a MAIT, foram corrigidas as imprecisões consideradas inadmissíveis e dificilmente admissíveis sempre que possível para a introdução da classificação no modelo. Este passo foi efetuado para a classificação do ano 2019, resultando num mapa de uso do solo (figura 29), em que as imprecisões temáticas que poderiam constituir um problema para o resultado dos corredores de propagação potencial foram corrigidas.

Após gerar este ficheiro foi feita um novo teste de precisão temática, desta vez sem recorrer ao cálculo do índice de *kappa*. Este foi feito ao sobrepor o resultado da classificação alvo da mitigação de imprecisão temática com o mapa do uso e coberto do solo oficial. Optou-se por utilizar este método por ser mais amplo, e permitir identificar e investigar as diferenças. Esta sobreposição foi uma operação aritmética, em que se converteram ambos os mapas em *raster* e se subtraiu um pelo outro, a leitura dos resultado segue a lógica de que qualquer número subtraído por si mesmo é zero, então, todas as classes zero resultantes correspondem a áreas correspondentes em ambas as classificações, enquanto que qualquer outro número, positivo ou negativo significa uma discordância. O resultado foi então um mapa de discordância entre classificações (figura 30):

41°50'63"N

1°92'16"E

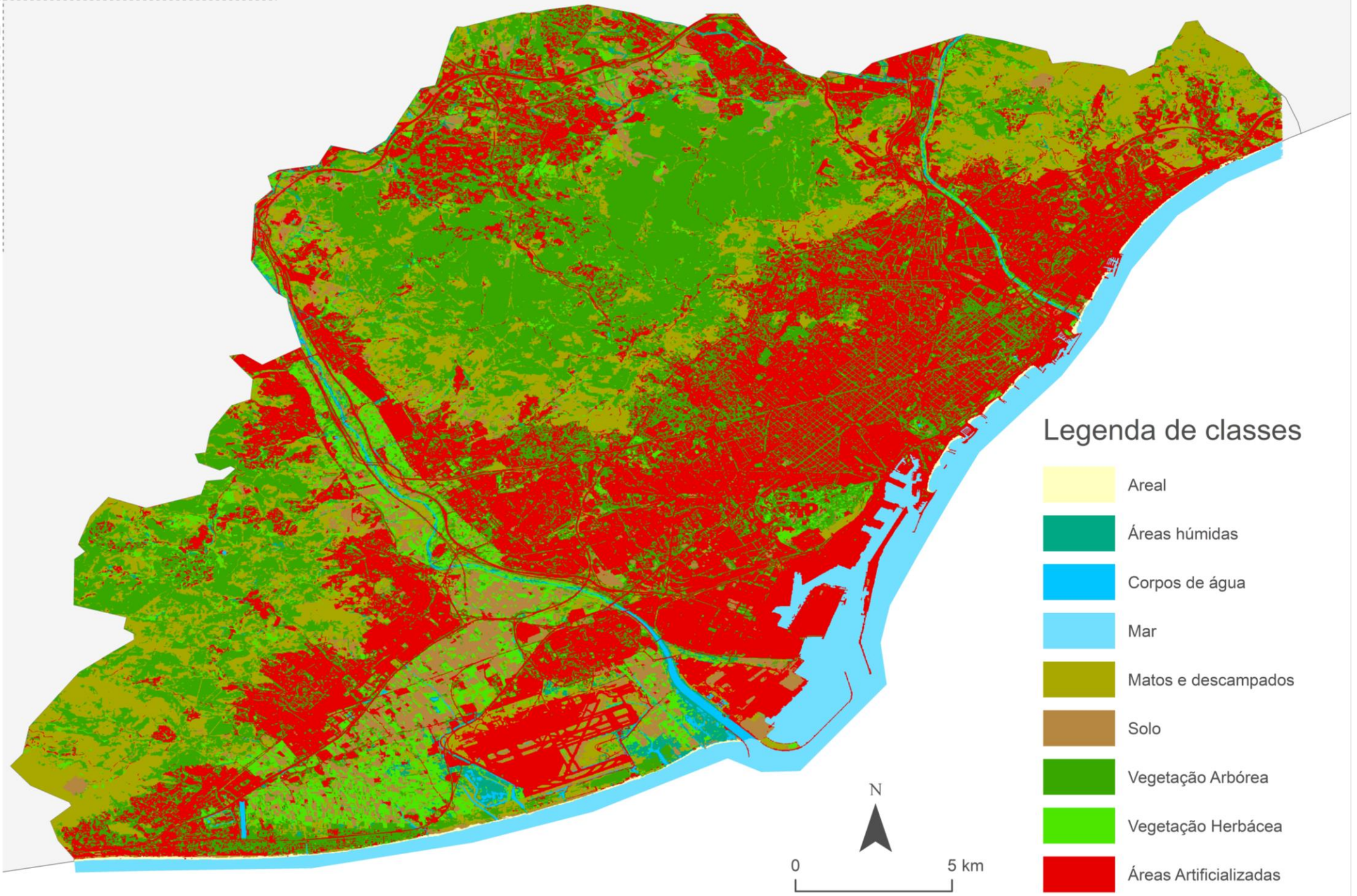


Figura 35 – Classificação do uso do solo, resultante da aplicação das medidas de mitigação de imprecisão temática auxiliadas por cartografia de referência para o ano 2019.

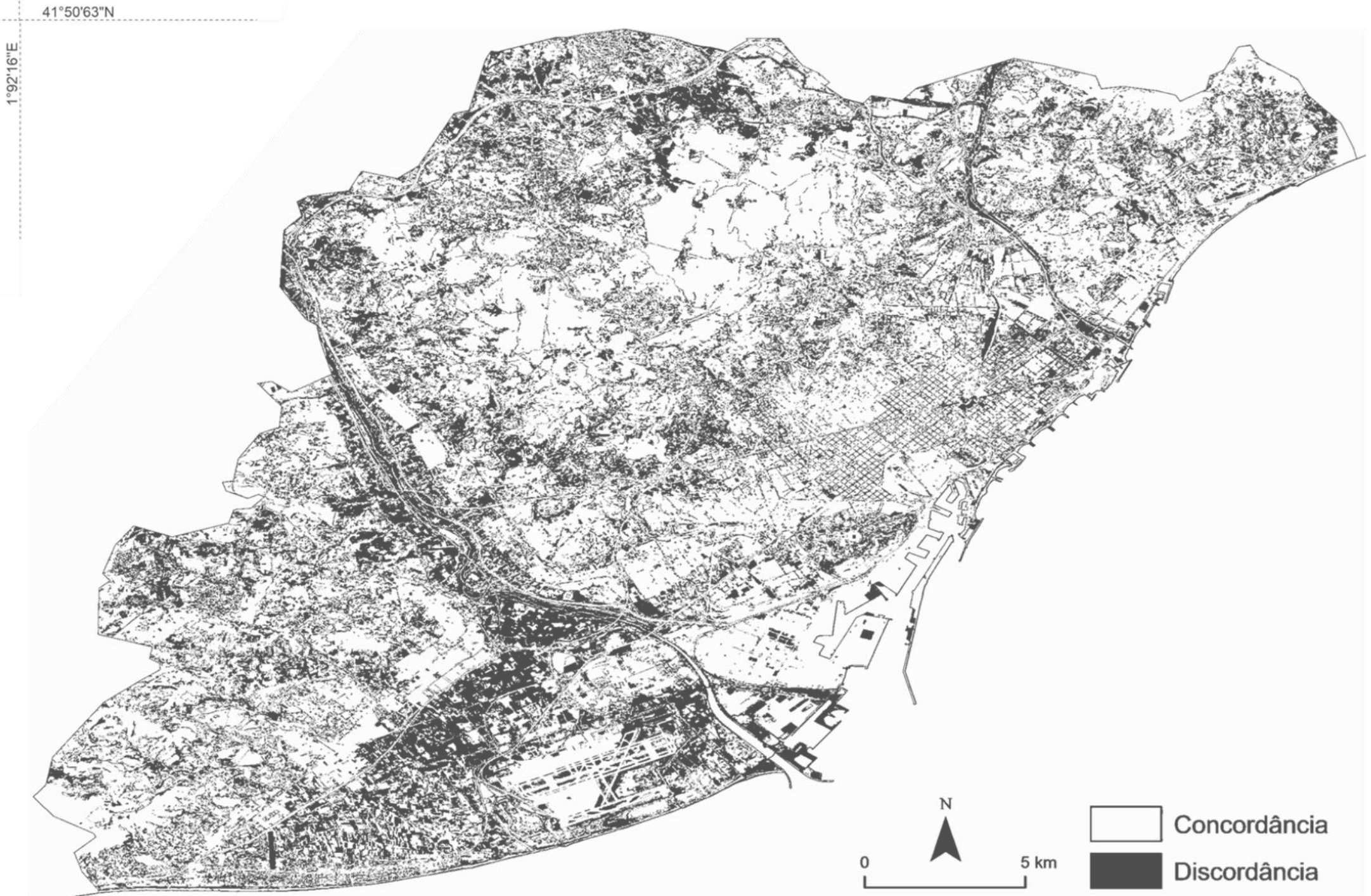


Figura 36 – Mapa de discordância temática e semântica entre a classificação do uso do solo produzida com o mapa de uso do solo descarregado do departamento de território e sustentabilidade da Catalunha.

O mapa produzido após a avaliação da imprecisão temática foi sujeito a alterações substanciais na distribuição das classes, o que pode ser observado pela variação da área ocupada por cada classe na classificação primária para a versão corrigida (tabela 12). Apenas a classe do Mar manteve a sua área que não corresponde à área apresentada nas figuras anteriores que diz apenas respeito à faixa litoral recortada pela área de estudo. As variações negativas ocorreram em apenas 3 classes, vegetação arbórea, solo e areal. O maior aumento percentual foi nas áreas húmidas que passaram a ocupar 6,3 km², houve uma particular dificuldade do para o algoritmo na sua classificação, pelo que se teve que recorrer a mais que uma informação externa para a correção, como o *Google Earth* ou cartografia oficial, isto acontece pela sua resposta espectral ser bastantes vezes confundível com a vegetação bem como apresentar também altos valores de NDVI.

Tabela 13 – Áreas em km² ocupadas por cada classe por estágio da classificação e respetiva variação

Classe	Classificação primária	Classificação corrigida	Variação
	Área Km ²	Área km ²	%
Areal	4,4	1,5	-65,5
Áreas. Artificializadas	144,0	175,3	21,7
Áreas húmidas	1,5	6,3	311,4
Corpos de água	2,8	3,1	11,9
Mar	28,3	29,3	3,5
Matos e descampados	62,9	83,4	32,5
Solo	39,7	24,1	-39,3
Vegetação Arbórea	197,3	150,2	-23,9
Vegetação Herbácea	21,3	29,0	36,3

A avaliação feita do mapa de discordância com classificação gerada e a obtida no portal do Departamento do Território e Sustentabilidade da Catalunha mostrou a existência de ainda muitas discordâncias, apesar de a concordância ter uma proporção maior (tabela 13). A explicação para esta discrepância, para além de potenciais imprecisões restantes, centra-se maioritariamente nas diferentes metodologias utilizadas para a produção das duas classificações dadas: o mapa oficial tem uma resolução de 1m enquanto o produzido tem 10m. A conversão da legenda também pôs o desafio de alguns usos do solo poderem compreender mais do que uma classe de

destino. Outro importante fator prendeu se com as áreas agrícolas das margens do El Llobregat e áreas a Este e noroeste do Aeroporto, onde se verificou a maior mancha, causada pela diferença entre decisões, no caso do mapa transferido optou se por deixar como vegetação herbácea, enquanto no mapa produzido decidiu-se orientar a classificação para a resposta espectral captada no momento, ou seja, áreas lavradas ou sem plantação naquele momento ficaram classificadas como solo.

Tabela 13 – Concordância entre a classificação descarregada e a classificação produzida.

	Nº de pixels	Área km2	Porcentagem
Concordância	3229534	322,95	68,47
Discordância	1486892	148,69	31,53

A criação e aplicação da matriz de admissibilidade de imprecisão temática foi sem dúvida uma mais valia para obter uma classificação com o máximo de viabilidade possível, em especial para classificações de áreas menores, no entanto a utilização desta matriz não foi levada ao seu máximo pela dimensão da área de estudo escolhida, porém a sua utilização foi um pilar na gestão do tempo para esta correção.

Os resultados da análise exploratória dos dados, os corredores de propagação potencial (figura 37), é um mapa também ele composto de 5 classes em que a primeira classe representa maior probabilidade de introdução e fixação e 5 menor probabilidade de introdução e fixação da espécie. O mapa de corredores de propagação potencial consiste então numa conjugação entre a probabilidade de introdução acidental de *Aedes albopictus* ao longo das vias de comunicação e a probabilidade de quando introduzido, tenha as condições de habitabilidade necessárias para a sua fixação.

41°50'63"N

1°92'16"E

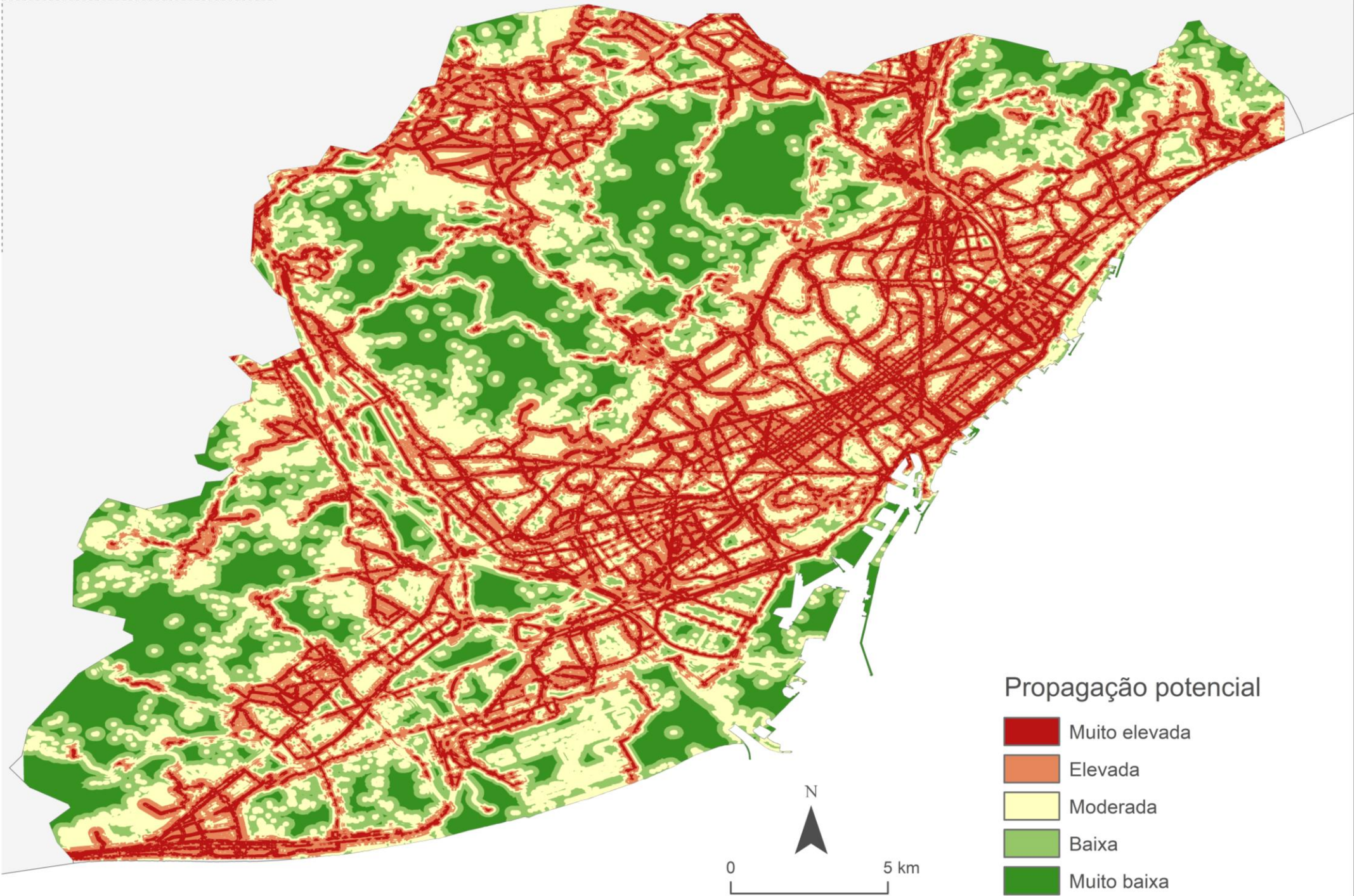


Figura 37 – Proposta de corredores de propagação potencial resultante da sobreposição não ponderada das de camadas de probabilidade de ocorrências de *Aedes albopictus*.

O mapa originado que indica o grau de probabilidade que o vetor tem de ser transportado e se estabelecer. Este mapa tem muitas similaridades com mapas de vias, pois a não atribuição de pesos de ponderação por falta de bibliografia de referência neste sentido levou a que as vias tivessem um peso total de cinquenta por cento.

Demarcam-se os padrões das vias de comunicação especialmente nas áreas urbanas com o nível mais alto de probabilidade, isto acontece pelo tipo de padrão de ocupação do solo em Barcelona, que combina os dois fatores, em especial para áreas urbanizadas com grande densidade de vias e comunicação e uma ampla presença de espaços verdes. Contrastando com estas, as áreas florestais, matos e agrícolas afastadas das vias e das áreas urbanas apresentam os valores de probabilidade mais baixos, no entanto não é impossível que o vetor não se estabeleça nestas áreas, a sua introdução e fixação é apenas mais improvável atendendo à conjugação das variáveis de introdução e estabelecimento utilizadas.

O resultado foi então o esperado e resulta da metodologia implementada, esta metodologia peca pela simplicidade, porém permite uma adaptação e monitorização passo a passo do efeito das variáveis que se utilizam neste tipo de análises, neste caso as 3 variáveis, sendo que duas delas se encontram colapsadas na mesma. O resultado foi altamente afetado por este colapso, caso contrário as vias deveriam ter um terço do peso ao invés de metade como aconteceu no resultado apresentado. Não obstante, a sua melhoria e posterior teste com novas variáveis ecológicas poderá criar resultados mais completos.

CONSIDERAÇÕES FINAIS E INVESTIGAÇÃO FUTURA

Este estudo debruçou-se sobre técnicas de extração de informação geográfica através da Detecção Remota de variáveis de coberto de superfície para inclusão em modelação de habitats prováveis, pelo que deverá ser replicável a outras áreas geográficas suscetíveis de ser invadidas por *Aedes albopictus*. São necessários outros testes a esta metodologia para verificar a sua replicabilidade, no entanto, sendo as únicas limitações a disponibilidade de dados de observações.

No âmbito do projeto TRIAD, financiado pela FCT, pretende-se replicar este estudo para áreas de Portugal suscetíveis ao desenvolvimento de populações de *Aedes albopictus*. Apesar de tanto a área de estudo deste trabalho como o território Português terem algumas similaridades do ponto de vista bioclimático, será sempre necessário algum ajuste caso-a-caso para otimizar os resultados, ajustando-os à realidade local e de cada escala de análise e âmbito geográfico. Como componente do projeto TRIAD, novas variáveis serão introduzidas para produzir um estudo em maior detalhe pelos diferentes parceiros deste estudo, portanto, a proposta de corredores de propagação potencial é exclusiva desta dissertação, podendo em toda a medida beneficiar da introdução de novas variáveis para a tornar mais robusta.

A extração do coberto vegetal sob a forma de índices de vegetação multitemporais é possível para praticamente qualquer área do mundo, desde que existam imagens sobre esses territórios de boa qualidade e sem nuvens. No entanto, serão sempre necessárias correções atmosféricas bem como o acautelar da dificuldade que é obter imagens de boa qualidade ao longo do ano nas regiões tropicais por exemplo.

A replicabilidade da Análise Orientada por Objetos tem mais desafios, uma vez que mesmo as mais pequenas alterações na imagem podem ter efeitos muito diferenciadores na segmentação e classificação, pelo que os parâmetros deverão ser sempre testados em novas áreas e situações.

A utilização do NDVI como *proxy* foi bem estudada, porém, a inclusão de outros índices de estado de superfície é sempre um fator positivo a acrescentar. Acabou por

não se utilizar índices de humidade devido à correlação espacial que este geralmente tem com o NDVI pela presença de água em plantas e vice-versa, porém, com um prazo maior este poderia ser estudado em maior detalhe para aferir as vantagens na sua utilização. O *Aedes albopictus* é bastante sensível à temperatura, pelo que seria bastante positivo a inclusão da temperatura de superfície a partir de imagens do *sensor Thermal Infrared Sensor (TIRS)* da plataforma *Landsat-8* para os estudos de habitats e áreas de atividade potenciais, o que é também possível de fazer em multitemporal com recurso a técnicas de Detecção Remota.

Quanto aos métodos utilizados, estes poderão ser sempre refinados para melhorar os resultados. Por exemplo a análise multitemporal utilizando a média aritmética beneficiaria de testes com outras estatísticas de forma a extrair o coberto vegetal multitemporal, mas que não foram realizadas neste estudo na ótica de otimizar o tempo e o poder de processamento disponíveis. O principal problema com utilização da média aritmética para os 12 meses do ano é que esta pode ser contestada pela ausência de observações para os meses de inverno, que pesam igualmente na média, dado que no inverno não existem observações, poderia assumir-se que contar estas imagens não seria o mais pertinente, porém mais investigação nos tempos de atividade do mosquito seria necessária e novos métodos de criação de informação desenhados para extrair as variáveis otimizadas à presença e atividade de *Aedes albopictus*.

A escolha de utilizar apenas uma imagem Sentinel-2 MSI mensal teve o intuito de minimizar os enviesamentos estatísticos resultantes da existência de mais imagens para uns meses que outros motivado por fatores meteorológicos, outra solução, e a mais robusta, seria realizar uma média aritmética dos NDVI's disponíveis para dado mês, e utilizar o resultado no cálculo do NDVI sazonal e anual, no entanto por limitações de capacidade computacional e de conexão à internet optou se pela imagem única, remetendo para trabalhos futuros a utilização do total das imagens disponíveis que cumpram os critérios de qualidade.

Para a extração dos usos do solo, a OBIA produziu uma classificação com uma muito boa precisão temática, porém a sua refinação é sempre possível e desejável, em especial na extração de amostras para treinar o algoritmo pois sabe-se que em *machine learning*, quanto mais amostras mais aprendizagem, logo maior a capacidade

diferenciadora do algoritmo. A inclusão/exclusão de estatísticas no treino também poderá oferecer diferentes resultados e melhorar o resultado. Os métodos de validação com a utilização do índice de *Kappa* e precisão global podem não dizer tudo sobre a qualidade da classificação por serem métodos estatísticos que recorrem a amostras, este problema poderá ser tão pior quando maior e mais complexa for a área de estudo.

Como forma de melhoria de resultados, que se observou nesta dissertação é a alteração dos dados radiométricos de entrada ao criar o modelo classificador. Para tal a inclusão de índices radiométricos multitemporais nos dados de entrada beneficia bastante a capacidade que o algoritmo tem de diferenciar as classes. Por exemplo, uma cultura de regadio tem um comportamento distinto no espaço, diferente de uma cultura de sequeiro se atendermos aos índices espectrais de vegetação e de presença de água. De forma a melhorar os resultados por extração automática, a exploração da inclusão destes índices multitemporais merece atenção e testagem futura.

A matriz de admissibilidade de imprecisão temática (MAIT) tem as suas limitações quando a área de estudo cobre territórios muito vastos, podendo ser otimizada para estudos mais detalhados sobre áreas menores em que seja possível a correção de elementos mal classificados, bem como a possibilidade de se alterar e dividir segmentos que contenham mais do que uma classe para melhorar ainda mais a classificação. A sua utilização em classificações de áreas de estudo extensas tem, no entanto, um impacto bastante expressivo ao mitigar a imprecisão temática.

No que diz respeito ao *software* utilizado, o SNAP, da Agência Espacial Europeia, demonstrou imenso potencial pela sua interoperabilidade com outros *software* e pelo seu carácter modular: é possível alterar o *software* com conhecimentos intermédios de programação para a introdução dos módulos para OBIA, porém o módulo seria a OTB, que tem uma melhor performance quanto é utilizado como *standalone*. A extração dos índices NDVI e testes com outros índices foram efetuados em SNAP, que demonstrou uma excelente responsividade. As operações de correção atmosférica e geométricas também foram feitas em ambiente SNAP, o que trouxe um novo grau de simplicidade e performance, abrindo a possibilidade ao utilizador de fazer estas operações sem requerer um grande poder de processamento e em relativamente pouco tempo. Por outro lado, as limitações prendem-se com a falha em incorporar todas as

funcionalidades da extensão OTB, mesmo que exista a possibilidade de as adicionar, seria um ponto positivo que a equipa responsável pelo desenvolvimento e atualização do *software* as disponibilizasse como extensão pré-programada para facilitar a utilização dos métodos de GEOBIA do início ao fim, dado que nem todos os utilizadores possuem conhecimentos suficientes para os programar no *software*. A visualização de objetos vetoriais em SNAP também se pode considerar um ponto fraco pela fraca responsividade do *software* quando se adicionam este tipo de ficheiros.

A utilização do *OrfeoToolBox monteverti* como *standalone* é uma ferramenta muito simples de utilizar e com uma responsividade bastante alta, sendo possível fazer a maior parte dos passos de OBIA, exceto a extração de amostras e visualização dos resultados, pelo que se teve de utilizar o *software* ArcMap para o efeito. Uma solução para a resolução deste problema é a utilização do *software open source Quantum Gis*¹⁰ com o módulo OTB, ou realizar o processo em *software* pago como *ArcMap* ou *eCognition*¹¹. A escolha do *software* tem então um peso reduzido, podendo dar-se preferência à performance, contando que os métodos de análise orientada por objetos e os algoritmos de um *software* seja replicável nos outros.

A relação de *Aedes Albopictus* com os elementos extraídos neste estudo necessitarão de ser replicados e comparados com outras variáveis para se definir quais terão, e a que nível, influência nas ocorrências, para se poder realizar um estudo mais aprofundado. A extração da ocupação do solo poderá então beneficiar da adição ou subtração de classes em novas classificações. O mapa produzido contém classes elaboradas pela sua separabilidade espectral de forma a otimizar os resultados, algo que poderá ter de ser alterado em estudos futuros.

Por fim, a proposta de corredores de propagação potencial mostra resultados promissores devido à sua simplicidade e potencial de aplicabilidade à escala regional,

¹⁰ O *Quantum Gis* (QGIS) é um software livre com código-fonte aberto, de informação geográfica que permite a visualização, edição e análise de dados georeferenciados, em <https://www.qgis.org/en/site/>.

¹¹ O *eCognition* é um software desenhado para automatizar a análise de dados geoespaciais desenvolvido pela *Trimble Geospatial* em <https://geospatial.trimble.com/products-and-solutions/ecognition>.

dando ainda espaço para melhorias através da consideração de variáveis adicionais, em especial indicadores de temperatura, que se têm mostrado um fator preponderante na atividade e adaptação de *Aedes albopictus*. Porém, as limitações dos dados disponíveis para as ocorrências e a falta de estudos para a parametrização dos pesos das variáveis no cálculo dos corredores coloca interrogações sobre a fiabilidade deste tipo de abordagem. A introdução das vias de comunicação teria beneficiado da atribuição de pesos consoante o tráfego médio, produzindo dessa forma classes ponderadas, não dando o mesmo valor a uma estrada terciária que se dá a uma via principal com muito mais tráfego. A validação do resultado utilizando dados da plataforma GBIF para 2020 foi gorada pela pandemia Covid-19 que inibiu a ciência voluntária e não produziu uma amostra de ocorrências significativa para se avançar com uma validação dos resultados.

A posterior realização de estudos sobre como esta espécie se está a adaptar aos novos territórios onde se estabelece é de suma importância. É necessário estudar os meios de disseminação e as condições de fixação para a criação de parâmetros mapeáveis para o auxílio às autoridades competentes nos seus esforços de contenção e mitigação de potenciais doenças que este mosquito seja potencial vetor.

Para finalizar, é também importante refletir sobre a importância deste estudo não só como exercício metodológico de ‘meios’ para atingir um determinado fim, mas também a relevância destas temáticas para a Geografia e, sobretudo, para a Gestão do Território, que são as áreas científicas de interesse deste trabalho.

As alterações climáticas representam o maior desafio para a humanidade no século XXI, para além toda a crescente fragilidade dos ecossistemas, questões de desenvolvimento e o aumento ou ressurgimento de algumas doenças endémicas, como a malária, representam uma necessidade acrescida de melhorar qualitativa e tematicamente os estudos de propagação de vetores, sobretudo com recurso a técnicas e tecnologias de observação da Terra.

De todos os tipos de dados geográficos, as imagens obtidas por Detecção Remota são talvez as mais democráticas: todo o Mundo tem cobertura e um varrimento periódico, com dados disponíveis gratuitamente não só para a comunidade científica, mas também para associações, mundo empresarial e o cidadão comum. Tendo em conta o carácter endémico geograficamente muito bem definido de algumas doenças

propagadas por mosquitos, é possível realizar estudos de previsão de propagação com alguma fiabilidade e poucos recursos disponíveis. Neste aspeto, os dados abertos, ferramentas de processamento *open source* e bases de dados de informação geográfica voluntária (*VGI – Voluntary Geographic Information*) são uma excelente oportunidade de democratizar estudos em países em desenvolvimento, ou com menos condições de estudos exaustivos de campo.

Para além disso, as alterações na localização de alguns ecossistemas potenciadas pelo aumento médio da temperatura do Planeta podem fazer ressurgir doenças propagadas por mosquitos a territórios onde já haviam sido erradicadas: a bacia do Mediterrâneo, outrora endémica em malária, terá tendência a tornar-se cada vez mais uma área suscetível ao surgimento deste tipo de doenças, especialmente nas planícies aluviais e estuários de alguns rios, como é exemplo o Estuário do Sado ou a Ria Formosa em Faro.

Em suma, esta dissertação de cariz metodológico tem também um significado mais abrangente para a ciência geográfica e gestão do território: contribuir para consolidar as bases para a modelação de dados, tendo em conta realidades espaciais do presente e passado, bem como uma melhor preparação para as questões de um futuro próximo.

REFERÊNCIAS BIBLIOGRÁFICAS

- Ali, M. I., Hasim, A. H., & Raiz Abidin, M. (2019). Monitoring the Built-up Area Transformation Using Urban Index and Normalized Difference Built-up Index Analysis. *International Journal of Engineering*, 32(5), 647–653.
<https://doi.org/10.5829/ije.2019.32.05b.04>
- Agencia Estatal de Meteorología, (2020). Valores climatológicos normales Barcelona Aeropuerto 1981-2010.
- Ajay, A., Deepa Merlin Dixon, K., Sowmya, V., & Soman, K. P. (2016). Aerial image classification using GURLS and LIBSVM. *International Conference on Communication and Signal Processing, ICCSP 2016*, (fevereiro de 2018), 396–401.
<https://doi.org/10.1109/ICCSP.2016.7754165>
- Alto, B. W., & Juliano, S. A. (2001). Precipitation and temperature effects on populations of *Aedes albopictus* (Diptera: Culicidae): implications for range expansion. *Journal of Medical Entomology*, 38(5), 646–656.
<https://doi.org/10.1603/0022-2585-38.5.646>
- Aranda, C., Eritja, R., & Roiz, D. (2006). First record and establishment of the mosquito *Aedes albopictus* in Spain. *Medical and Veterinary Entomology*, 20(1), 150–152.
<https://doi.org/10.1111/j.1365-2915.2006.00605>
- Aryal, J. (2013). Advancing in spatial object recognition: GEOBIA approach. Hobart, Australia: AGILE 2013.
- Ascensão, F., & Capinha, C. (2017). Aliens on the move: Transportation networks and non-native species. Em *Railway Ecology*. https://doi.org/10.1007/978-3-319-57496-7_5
- Banko, G. (1998). A review of assessing the accuracy of and of methods including remote sensing data in forest inventory. *International Institute for Applied Systems Analysis, Interim Report IT-98-081*, Laxenburg, Austria.
- Beerli, O., & Peled, A. (2006). Spectral indices for precise agriculture monitoring. *International Journal of Remote Sensing*, 27(10), 2039–2047.
<https://doi.org/10.1080/01431160612331392950>

- Belgiu, M., & Csillik, O. (2018). *Sentinel-2* cropland mapping using pixel-based and object-based time-weighted dynamic time warping analysis. *Remote Sensing of Environment*, 204(setembro de 2017), 509–523.
<https://doi.org/10.1016/j.rse.2017.10.005>
- Blaschke, Thomas, Lang, S., & Hay, G. J. (2008). *Object-Based Image Analysis: Spatial Concepts for Knowledge-Driven Remote Sensing Applications*, (janeiro de 2008), V–VIII. <https://doi.org/10.1007/978-3-540-77058-9>
- Blaschke, T. (2010). Object based image analysis for remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing*, 65(1), 2–16.
<https://doi.org/10.1016/j.isprsjprs.2009.06.004>
- Blaschke, T., Kelly, M., & Merschdorf, H. (2016). Object-Based Image Analysis: Evolution, History, State of the Art, and Future Vision.
- Gordon, A. D., Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). Classification and Regression Trees. *Biometrics*. <https://doi.org/10.2307/2530946>
- Brown, J. (2018). NDVI, the Foundation for Remote Sensing Phenology. Acedido a 20 de agosto no website da USGSem https://www.usgs.gov/land-resources/eros/phenology/science/ndvi-foundation-remote-sensing-phenology?qt-science_center_objects=0#qt-science_center_objects
- Caminade, C., Medlock, J. M., Ducheyne, E., McIntyre, K. M., Leach, S., Baylis, M., & Morse, A. P. (2012). Suitability of European climate for the Asian tiger mosquito *Aedes albopictus*: Recent trends and future scenarios. *Journal of the Royal Society Interface*, 9(75), 2708–2717. <https://doi.org/10.1098/rsif.2012.0138>
- Canada, N. R. (2020). Digital Image Processing. Acedido a 19 de agosto de 2020 em <https://www.nrcan.gc.ca/maps-tools-publications/satellite-imagery-air-photos/remote-sensing-tutorials/image-interpretation-analysis/digital-image-processing/9279>
- Capinha, C., Essl, F., Seebens, H., Moser, D., & Pereira, H. M. (2015). The dispersal of alien species redefines biogeography in the Anthropocene. *Science*.
<https://doi.org/10.1126/science.aaa8913>
- Capinha, C., Rocha, J., & Sousa, C. A. (2014). Macroclimate Determines the Global

- Range Limit of *Aedes aegypti*. <https://doi.org/10.1007/s10393-014-0918-y>
- Caraballo, H., & King, K. (2014). Emergency department management of mosquito-borne illness: malaria, dengue, and West Nile virus. *Emergency Medicine Practice*.
- Chang, C. C., & Lin, C. J. (2019). LIBSVM: A Library for support vector machines. Department of Computer Science National Taiwan University, Taipei, Taiwan. <https://doi.org/10.1145/1961189.1961199>
- Chee, J. D., & Queen, T. (2013). Pearson ' s Product-Moment Correlation: Sample Analysis. (janeiro de 2013).
- Chen, G., Weng, Q., Hay, G. J., & He, Y. (2018). Geographic object-based image analysis (GEOBIA): emerging trends and future opportunities. *GIScience and Remote Sensing*, 55(2), 159–182. <https://doi.org/10.1080/15481603.2018.1426092>
- Chen, J.A., Wang, H., & Bjørnstad, O. N. (1996). Occurrence and spread in Italy of *Aedes albopictus*, with implications for its introduction into other parts of Europe. *Journal of the American Mosquito Control Association*, 12(2 Pt 1), 177–183.
- Childs, L. M., El Moustaid, F., Gajewski, Z., Kadelka, S., Nikin-Beers, R., Smith, J. W., ... Johnson, L. R. (2019). Multi-scale models and data for infectious diseases: A systematic review. <https://doi.org/10.7717/peerj.7057>
- Clark, R. N. (1999). Spectroscopy of Rocks and Minerals and Principles of Spectroscopy. Manual of Remote Sensing. In *Remote Sensing for the Earth Sciences: Manual of Remote Sensing* (Vol. 3).
- Cleckner, H. L., Allen, T. R., & Scott Bellows, A. (2011). Remote sensing and modeling of mosquito abundance and habitats in Coastal Virginia, USA. *Remote Sensing*, 3(12), 2663–2681. <https://doi.org/10.3390/rs3122663>
- Congalton, R. G. (1991). A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sensing of Environment*, 37(1), 35–46. [https://doi.org/10.1016/0034-4257\(91\)90048-B](https://doi.org/10.1016/0034-4257(91)90048-B)
- Castro, A. I., Six, J., Plant, R. E., & Peña, J. M. (2018). Mapping crop calendar events and phenology-related metrics at the parcel level by object-based image analysis (OBIA) of MODIS-NDVI time-series: A case study in central California. *Remote*

- Sensing*, 10(11). <https://doi.org/10.3390/rs10111745>
- Clark Labs (2020). Segmentation and Segment-based Classification. Acedido a 5 de abril de 2020 no website de clark Labs: <https://clarklabs.org/segmentation-and-segment-based-classification/>
- Claxton, N., & Macias, K. (2016). Zika, dengue and chikungunya prevention community module. Retrieved from www.climatecentre.org
- Dall'erba, S. (2009). Exploratory Spatial Data Analysis. *International Encyclopedia of Human Geography*, 683–690. <https://doi.org/10.1016/B978-008044910-4.00433-8>
- Delatte, H., Desvars, A., Bouétard, A., Bord, S., Gimonneau, G., Vourc'h, G., & Fontenille, D. (2010). Blood-feeding behavior of *Aedes albopictus*, a vector of chikungunya on la réunion. *Vector-Borne and Zoonotic Diseases*, 10(3), 249–258. <https://doi.org/10.1089/vbz.2009.0026>
- Departamento de Territorio y Sostenibilitat (nd). Grafo de infraestructuras terrestres. Retiraro a 17 de agosto de 2020 de https://territori.gencat.cat/es/01_departament/12_cartografia_i_toponimia/bases_cartografiques/infraestructures_mobilitat/infraestructures_terrestres/graf_infraestructures_terrestres/
- Departamento de Territorio y Sostenibilitat (2019). Mapa d'usos i cobertes del sòl. Descarregado a 10 de setembro de 2020 de https://territori.gencat.cat/ca/01_departament/12_cartografia_i_toponimia/bases_cartografiques/medi_ambient_i_sostenibilitat/bases_miramont/territori/mapa-dusos-i-cobertes-del-sol/
- Dieng, H., Saifur, R. G. M., Hassan, A. A., Che Salmah, M. R., Boots, M., Satho, T., ... Abubakar, S. (2010). Indoor-breeding of *Aedes albopictus* in northern peninsular Malaysia and its potential epidemiological implications. *PLoS ONE*, 5(7). <https://doi.org/10.1371/journal.pone.0011790>
- Direção Geral de Saúde (DGS). (n.d.). *Aedes albopictus* - Informação geral sobre o mosquito *Aedes albopictus*. Acedido a 21 de janeiro de 2020 em <https://www.dgs.pt/paginas-de-sistema/saude-de-a-a-z/aedes-albopictus->

mosquito.aspx

ECDC; EFSA. (2019). *Aedes albopictus* - current known distribution: August 2019.

Acedido a 21 de janeiro de 2020 no website da ECDC:

<https://www.ecdc.europa.eu/en/publications-data/aedes-albopictus-current-known-distribution-august-2019>

ECDC; EFSA. (2020). *Aedes albopictus* - current known distribution: May 2020. Acedido

a 17 de julho de 2020 no website da ECDC:

<https://www.ecdc.europa.eu/en/publications-data/aedes-albopictus-current-known-distribution-may-2020>

El-Gammal, M. I., Ali, R. R., & R. M. Abou Samra. (2014). NDVI Threshold Classification

for Detecting Vegetation Cover in Damietta Governorate, Egypt. *Journal of*

American Science, 10(8), 108–113. <https://doi.org/10.1088/1361-6560/aad43f>

El-naggar, A. N., Elbanna, S. M., Kaiser, M. F., & Refaat, M. (2017). Mosquito larval

habitat mapping using remote sensing and GIS for monitoring the filarial infection

regions in Alkorin village, Sharkia Governorate (Egypt). *International Journal of*

Mosquito Research, 4(4), 135–139.

El-Zeiny, A., El-Hefni, A., & Sowilem, M. (2017). Geospatial techniques for

environmental modeling of mosquito breeding habitats at Suez Canal Zone, Egypt.

Egyptian Journal of Remote Sensing and Space Science, 20(2), 283–293.

<https://doi.org/10.1016/j.ejrs.2016.11.009>

European Centre for Disease Prevention and Control. (n.d.). *Aedes albopictus* -

Factsheet for experts. Acedido a 21 de agosto de 2020 no website da ECDC:

<https://www.ecdc.europa.eu/en/disease-vectors/facts/mosquito-factsheets/aedes-albopictus>

European Space Agency. (2020.). Sen2Cor | STEP. Acedido a 31 de dezembro de 2019

em Step.Esa.int website: <https://step.esa.int/main/third-party-plugins-2/sen2cor/>

European Space Agency. (2018a). Level-2A Algorithm - *Sentinel-2* MSI Technical Guide -

Sentinel Online. Acedido a 17 de março de 2020 em Earth.Esa.int website:

<https://earth.esa.int/web/Sentinel/technical-guides/Sentinel-2-msi/level-2a/algorithm>

- European Space Agency. (2018b). User Guides - *Sentinel-2* MSI - Level-2 Processing - *Sentinel* Online. Acedido a 17 de março de 2020 em *Sentinel*.Esa.int website: <https://Sentinel.esa.int/web/Sentinel/user-guides/Sentinel-2-msi/processing-levels/level-2>
- European Space Agency. (2019). User Guides - *Sentinel-2* MSI - Level-2A Product - *Sentinel* Online. Acedido a 31 de dezembro de 2019 em Earth.ESA.int website: <https://earth.esa.int/web/Sentinel/user-guides/Sentinel-2-msi/product-types/level-2a>
- ESRI. (n.d.). What is image classification? - ArcMap Documentation. Acedido a 19 de agosto de 2020 em <https://desktop.arcgis.com/en/arcmap/latest/extensions/spatial-analyst/image-classification/what-is-image-classification-.htm>
- ESRI. (n.d.). Indices gallery—ArcGIS Pro | Documentation. Acedido a 19 de agosto de 2020 em https://pro.arcgis.com/en/pro-app/help/data/imagery/indices-gallery.htm#ESRI_SECTION1_C0443AF631EB4BECBB8E3A0D8B9DD5AA
- Fatiha, B., Abdelkader, A., Latifa, H., & Mohamed, E. (2013). Spatio temporal analysis of vegetation by vegetation indices from multi-dates satellite images: Application to a semi arid area in ALGERIA. *Energy Procedia*, 36(December), 667–675. <https://doi.org/10.1016/j.egypro.2013.07.077>
- Ferraguti, M., Martínez-De La Puente, J., Roiz, D., Ruiz, S., Soriguer, R., & Figuerola, J. (2016). Effects of landscape anthropization on mosquito community composition and abundance. *Scientific Reports*, 6(June), 1–9. <https://doi.org/10.1038/srep29002>
- Ferrão, J. L., Niquisse, S., Mendes, J. M., & Painho, M. (2018). Mapping and modelling malaria risk areas using climate, socio-demographic and clinical variables in Chimoio, Mozambique. *International Journal of Environmental Research and Public Health*, 15(4), 1–15. <https://doi.org/10.3390/ijerph15040795>
- Fischer, D., Thomas, S. M., Niemitz, F., Reineking, B., & Beierkuhnlein, C. (2011). Projection of climatic suitability for *Aedes albopictus* Skuse (Culicidae) in Europe under climate change conditions. *Global and Planetary Change*, 78(1–2), 54–64.

<https://doi.org/10.1016/j.gloplacha.2011.05.008>

- Foley, D. H., Rueda, L. M., & Wilkerson, R. C. (2007). Insight into global mosquito biogeography from country species records. *Journal of Medical Entomology*, 44(4), 554–567. <https://doi.org/10.1603/0022-2585>
- Franke, J., Gebreslasie, M., Bauwens, I., Deleu, J., & Siegert, F. (2015). Earth observation in support of malaria control and epidemiology: MALAREO monitoring approaches. *Geospatial Health*, 10(1), 117–131. <https://doi.org/10.4081/gh.2015.335>
- Franklinos, L. H. V., Jones, K. E., Redding, D. W., & Abubakar, I. (2019, September 1). The effect of global change on mosquito-borne disease. *The Lancet Infectious Diseases*, Vol. 19, pp. e302–e312. [https://doi.org/10.1016/S1473-3099\(19\)30161-6](https://doi.org/10.1016/S1473-3099(19)30161-6)
- Gao, J. (2009). *Digital Analysis of Remotely Sensed Imagery* (1st ed.; T. U. of Auckland, Ed.). Auckland: The McGraw-Hill Companies, Inc.
- Garabed, R. B., Jolles, A., Garira, W., Lanzas, C., Gutierrez, J., & Rempala, G. (2020). Multi-scale dynamics of infectious diseases. *Interface Focus*, 10(1), 20190118. <https://doi.org/10.1098/rsfs.2019.0118>
- Garira, W. (2018). A primer on multiscale modelling of infectious disease systems. *Infectious Disease Modelling*, 3, 176–191. <https://doi.org/10.1016/j.idm.2018.09.005>
- Garira, W. (2020). The research and development process for multiscale models of infectious disease systems. *PLoS Computational Biology*, 16(4). <https://doi.org/10.1371/journal.pcbi.1007734>
- Gascon, F., Bouzinac, C., Thépaut, O., Jung, M., Francesconi, B., Louis, J., ... Fernandez, V. (2017). Copernicus Sentinel-2A calibration and products validation status. *Remote Sensing*, 9(6). <https://doi.org/10.3390/rs9060584>
- GEO. (2017). In support of the Earth Observations 2030 Agenda for Sustainable Development. Acedido a 14 de abril de 2020 em <http://www.geoportal.org>
- Gimnig, John & Hightower, Allen & Hawley, William. (2005). Application of geographic

information systems to the study of the ecology of mosquitoes and mosquito-borne diseases. Wageningen UR Frontis Series. 9.

Global Invasive Species Database (2020) Species profile: *Aedes albopictus*.

Descarregado de <http://www.iucngisd.org/gisd/speciesname/Aedes+albopictus> on 22-12-2020.

Global Biodiversity Information Facility. (n.d.). Ocorrências de *Aedes albopictus*

Acedido a 3 de agosto de 2020 em <https://www.gbif.org/dataset/33614778-513a-4ec0-814d-125021cca5fe>

Ghose, M. ., Pradhan, R., & Sushan, S. (2010). Decision Tree Classification of Remotely Sensed Satellite Data using Spectral Separability Matrix. *International Journal of Advanced Computer Science and Applications*, 1(5).

<https://doi.org/10.14569/ijacsa.2010.010516>

Gomes, E. (2010). *Risco potencial de transmissão de malária em Portugal Continental* (Universidade Nova de Lisboa). Acedido em

<https://run.unl.pt/handle/10362/4156>

Gomes, E., Capinha, C., Rocha, J., & Sousa, C. (2016). Mapping risk of malaria transmission in mainland Portugal using a mathematical modelling approach.

PLoS ONE. <https://doi.org/10.1371/journal.pone.0164788>

Charles, J. & Godfray, H. (2013). Mosquito ecology and control of malaria. *Journal of Animal Ecology*. <https://doi.org/10.1111/1365-2656.12003>

Guo, F., Lenoir, J., & Bonebrake, T. C. (2018). Land-use change interacts with climate to determine elevational species redistribution. *Nature Communications*, 9(1), 1–7.

<https://doi.org/10.1038/s41467-018-03786-9>

Global Invasive Species Database (2021) Species profile: *Aedes albopictus*. Acedido a 18 de março de 2021 de <http://www.iucngisd.org/gisd/species.php?sc=109>

Haining, R., & Wise, S. (1997). Exploratory spatial data analysis, NCGIA Core curriculum in GIScience. *NCGIA Core Curriculum in GIScience*. Acedido em

<http://www.ncgia.ucsb.edu/giscc/units/u128/u128.html>

Harris, S., Veraverbeke, S., & Hook, S. (2011). Evaluating spectral indices for assessing

- fire severity in chaparral ecosystems (Southern California) using modis/aster (MASTER) airborne simulator data. *Remote Sensing*, 3(11), 2403–2419.
<https://doi.org/10.3390/rs3112403>
- Hänninen, S. (2019). Showing Drought Stress in Sacramento Parks. Acedido a 17 de agosto de 2020 em <https://slideplayer.com/slide/16979756/>
- Hashim, H., Abd Latif, Z., & Adnan, N. A. (2019). Urban Vegetation Classification With Ndvi Threshold Value Method With Very High Resolution (Vhr) Pleiades Imagery. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-4/W16(Outubro), 237–240.
<https://doi.org/10.5194/isprs-archives-xlii-4-w16-237-2019>
- Hay, G. J., & Castilla, G. (2008). Geographic Object-Based Image Analysis (GEOBIA): A new name for a new discipline. *Geographic Object-Based Image Analysis (GEOBIA)*, (Janeiro de 2008), 75–89. <https://doi.org/10.1007/978-3-540-77058-9>
- Henrich, V., Krauss, G., Götze, C., Sandow, C. (n.d.). IDB - Index DataBase. Acedido a 20 de agosto de 2020 em <https://www.indexdatabase.de/>
- Hu, H. (2019). Intelligent Identification of Survival Environment Information of Typical Endangered Animals. *Revista Científica, FVC-LUZ /*, XXIX(4), 827–839.
- Huang, X., Xie, C., Fang, X., & Zhang, L. (2015). Combining Pixel-and Object-Based Machine Learning for Identification of Water-Body Types from Urban High-Resolution Remote-Sensing Imagery. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 8(5), 2097–2110.
<https://doi.org/10.1109/JSTARS.2015.2420713>
- Huete, A. R. (2012). Vegetation Indices, Remote Sensing and Forest Monitoring. *Geography Compass*, 6(9), 513–532. <https://doi.org/10.1111/j.1749-8198.2012.00507.x>
- Huggett, R.J. (2004). *Fundamentals of Biogeography* (2nd ed.). Routledge.
<https://doi.org/10.4324/9780203356586>
- Hsu, C.-W., Chang, C.-C., & Lin, C.-J. (2003). *A Practical Guide to Support Vector Classification*. National Taiwan University, Taipei 106, Taiwan.

- Jiang, L., Wang, W., & Yang, X. (2011). Classification Methods of Remote Sensing Image Based. *IFIP International Federation for Information Processing*, 1, 353–358.
https://doi.org/10.1007/978-3-642-18333-1_41
- John R. Jensen. (2015). *INTRODUCTORY DIGITAL IMAGE PROCESSING A Remote Sensing Perspective* (4th ed., p. Education, Ed.). South Carolina: University of South Carolina.
- Joshi, Chudamani & De Leeuw, Jan & Van Duren, Iris. (2004). Remote sensing and GIS applications for mapping and spatial modeling of invasive species. *ISPRS*. 35.
- Kamal, M., Kenawy, M. A., Rady, M. H., Khaled, A. S., & Samy, A. M. (2018). Mapping the global potential distributions of two arboviral vectors *Aedes aegypti* and *Aedes albopictus* under changing climate. *PLoS ONE*, 13(12), 1–21.
<https://doi.org/10.1371/journal.pone.0210122>
- Kaplan, G., & Avdan, U. (2017). Object-based water body extraction model using *Sentinel-2* satellite imagery. *European Journal of Remote Sensing*, 50(1), 137–143.
<https://doi.org/10.1080/22797254.2017.1297540>
- Kennedy, D., & Marjorie Lucks. (1999). Rubber, Blight, and Mosquitoes: Biogeography Meets the Global Economy. *Environmental History*, 4(3), 369-383. Retrieved April 7, 2021, from <http://www.jstor.org/stable/3985132>
- Kobayashi, N., Tani, H., Wang, X., & Sonobe, R. (2020). Crop classification using spectral indices derived from *Sentinel-2A* imagery. *Journal of Information and Telecommunication*, 4(1), 67–90.
<https://doi.org/10.1080/24751839.2019.1694765>
- Kogan, F. (2020). *Remote Sensing for Malaria*. <https://doi.org/10.1007/978-3-030-46020-4>
- Kraemer, M. U. G., Sinka, M. E., Duda, K. A., Mylne, A. Q. N., Shearer, F. M., Barker, C. M., ... Hay, S. I. (2015). The global distribution of the arbovirus vectors *Aedes aegypti* and *Aedes albopictus*. *ELife*, 4(Junho de 2015), 1–18.
<https://doi.org/10.7554/eLife.08347>
- Kyalo, J. M. (2020). Neurophysiology Research The use of remote sensing and GIS in public health: a case study: geo-spatial risk analysis for schistosomiasis in a

- marginalized community, Kenya. Em *Neurophysiol Res. Osaka: Japan Webinar on Public Health*.
- Lillesand, T. M., Kiefer, R. W., & Chipman, J. (2004). Remote sensing and image interpretation [Teledetección e interpretación de imágenes]. Em *Remote sensing and image interpretation*. <https://doi.org/10.2307/634969>
- Liu, P., Choo, K. K. R., Wang, L., & Huang, F. (2017). SVM or deep learning? A comparative study on remote sensing image classification. *Soft Computing*, 21(23), 7053–7065. <https://doi.org/10.1007/s00500-016-2247-2>
- Lopes, N., Nozawa, C., & Linhares, R. E. C. (2014). Características gerais e epidemiologia dos arbovírus emergentes no Brasil. *Revista Pan-Amazônica de Saúde*, 5(3), 55–64. <https://doi.org/10.5123/s2176-62232014000300007>
- Lopes, P., Lourenco, P., Seixas, J., Sousa, C., Novo, T., Rodrigues, J., & Almeida, A. P. G. (2006). Modelling patterns of mosquito density based on remote sensing images. *Global Developments in Environmental Earth Observation from Space*, (Janeiro), 251–258.
- Lourenço, P. M., Sousa, C. A., Seixas, J., Lopes, P., Novo, M. T., & Almeida, A. P. G. (2011). Anopheles atroparvus density modeling using MODIS NDVI in a former malarious area in Portugal. *Journal of Vector Ecology*, 36(2), 279–291. <https://doi.org/10.1111/j.1948-7134.2011.00168.x>
- LSA-SAF. (2019). Normalized Difference Vegetation index. Acedido a 22 de julho de 2020 em <https://landsaf.ipma.pt/en/data/catalogue/>
- Main-Knorn, M., Pflug, B., Louis, J., Debaecker, V., Müller-Wilm, U., & Gascon, F. (2017). *Sen2Cor for Sentinel-2*. <https://doi.org/10.1117/12.2278218>
- Manorenjitha, M. & Zairi, J., (2015). The Adaptation of Field Collected *Aedes aegypti* (L.) and *Aedes albopictus* (Skuse) in Laboratory Condition. *International Journal of Life Science and Medical Research*. <https://doi.org/10.5963/lsmr0504001>
- Marangoz, A. M., Sekertekin, A., & Akçin, H. (2017). Analysis of Land Use Land Cover Classification Results Derived From *Sentinel-2* Image. *17th International Multidisciplinary Scientific GeoConference*.

- Marechal, F., Ribeiro, N., Lafaye, M., & Güell, A. (2008). Satellite imaging and vector-borne diseases: The approach of the French National Space Agency (CNES). *Geospatial Health*, 3(1), 1–5. <https://doi.org/10.4081/gh.2008.226>
- Mayer, B, & Kylling, A. (2005). Technical note: The libRadtran *software* package for radiative transfer calculations-description and examples of use. In *Atmos. Chem. Phys* (Vol. 5). Acedido em www.atmos-chem-phys.org/acp/5/1855/SRef-ID:1680-7324/acp/2005-5-1855EuropeanGeosciencesUnion
- Medeiros-Sousa, A. R., Fernandes, A., Ceretti-Junior, W., Wilke, A. B. B., & Marrelli, M. T. (2017). Mosquitoes in urban green spaces: Using an island biogeographic approach to identify drivers of species richness and composition. *Scientific Reports*, 7(1), 1–11. <https://doi.org/10.1038/s41598-017-18208-x>
- Millington, Andrew & Walsh, Stephen & Osborne, Patrick. (2001). GIS and Remote Sensing Applications in Biogeography and Ecology. 10.1007/978-1-4615-1523-4.
- Ming, D., Li, J., Wang, J., & Zhang, M. (2015). Scale parameter selection by spatial statistics for GeOBIA: Using mean-shift based multi-scale segmentation as an example. *ISPRS Journal of Photogrammetry and Remote Sensing*, 106, 28–41. <https://doi.org/10.1016/j.isprsjprs.2015.04.010>
- Monteiro, A. T., Gonçalves, J., Fernandes, R. F., Alves, S., Marcos, B., Lucas, R., ... Honrado, J. P. (2017). Estimating invasion success by non-native trees in a national park combining worldview-2 very high resolution satellite data and species distribution models. *Diversity*, 9(1). <https://doi.org/10.3390/d9010006>
- Morgan, K., O’Loughlin, S. M., Mun-Yik, F., Linton, Y. M., Somboon, P., Min, S., ... Walton, C. (2009). Molecular phylogenetics and biogeography of the Neocellia Series of Anopheles mosquitoes in the Oriental Region. *Molecular Phylogenetics and Evolution*, 52(3), 588–601. <https://doi.org/10.1016/j.ympev.2009.01.022>
- Mueller-Wilm, U. (2018). *Sen2Cor Configuration and User Manual*.
- Müller-Wilm, U., Louis, J., Richter, R., Gascon, F., & Niezette, M. (2013). *SENTINEL-2 LEVEL 2A PROTOTYPE PROCESSOR: ARCHITECTURE, ALGORITHMS AND FIRST RESULTS*. Acedido em <https://www.libradtran.org>
- Muralli, M. (2015). Applications of remote sensing and Geographic Information System

- (GIS) in Archaeology. Em A. S. Gaur (Ed.), *Recent Researches on Indus Civilization & Maritime Archaeology in India* (pp. 219–223). New Delhi.
- Neteler, M., Roiz, D., Rocchini, D., Castellani, C., & Rizzoli, A. (2011). Terra and Aqua satellites track tiger mosquito invasion: Modelling the potential distribution of *Aedes albopictus* in north-eastern Italy. *International Journal of Health Geographics*, 10(August). <https://doi.org/10.1186/1476-072X-10-49>
- Norris, D. E. (2004). Mosquito-borne Diseases as a Consequence of Land Use Change. *EcoHealth*, 1(1), 19–24. <https://doi.org/10.1007/s10393-004-0008-7>
- OrfeoToolbox. (2017). VectorClassifier - Vector Classification. Acedido a 24 de julho de 2020 no website CookBook-6.2: https://www.orfeo-toolbox.org/CookBook-6.2/_sources/Applications/app_VectorClassifier.rst.txt
- OrfeoToolbox. (2019). Segmentation. Acedido a 10 de março no website Orfeo ToolBox 7.0.0 documentation em https://www.orfeo-toolbox.org/CookBook/Applications/app_Segmentation.html
- OrfeoToolbox. (2020). TrainImagesClassifier. Acedido a 22 de julho no website Orfeo ToolBox 7.0.0 documentation em https://www.orfeo-toolbox.org/CookBook/Applications/app_TrainImagesClassifier.html#id7
- Yilmaz, E., Varol, B., Topaloglu, R., & Sertel, E. (2019). Object-based classification of Izmir Metropolitan City by using *Sentinel-2* images. *Proceedings of 9th International Conference on Recent Advances in Space Technologies, RAST 2019*, 407–412. <https://doi.org/10.1109/RAST.2019.8767781>
- Parselia, E., Kontoes, C., Tsouni, A., Hadjichristodoulou, C., Kioutsioukis, I., Magiorkinis, G., & Stilianakis, N. I. (2019). Satellite Earth Observation data in epidemiological modeling of malaria, dengue and West Nile Virus: A scoping review. *Remote Sensing*, 11(16). <https://doi.org/10.3390/rs11161862>
- Pasimeni, M. R., Valente, D., Semeraro, T., Petrosillo, I., & Zurlini, G. (2018). Anthropogenic landscapes. *Encyclopedia of Ecology*, 4, 472–481. <https://doi.org/10.1016/B978-0-12-409548-9.10602-5>
- Pinho, R. F. da C. (2009). *Aplicação de técnicas de Detecção Remota para a identificação de lineamentos estruturais num sector do noroeste de Portugal e do sudoeste da*

Galiza. Faculdade de Letras da Universidade do Porto.

- Van Der Putten, W. H., Macel, M., & Visser, M. E. (2010, July 12). Predicting species distribution and abundance responses to climate change: Why it is essential to include biotic interactions across trophic levels. *Philosophical Transactions of the Royal Society B: Biological Sciences*, Vol. 365, pp. 2025–2034.
<https://doi.org/10.1098/rstb.2010.0037>
- Ryan, S. J., Carlson, C. J., Mordecai, E. A., & Johnson, L. R. (2019). Global expansion and redistribution of Aedes-borne virus transmission risk with climate change. *PLOS Neglected Tropical Diseases*, 13(3), e0007213.
<https://doi.org/10.1371/journal.pntd.0007213>
- Richards, J. A. (2013). *Remote sensing digital image analysis: An introduction*. In *Remote Sensing Digital Image Analysis: An Introduction*.
<https://doi.org/10.1007/978-3-642-30062-2>
- Richter, R., Louis, J., Niezette, M., Quality, C. D., & Manager, A. (2011). *Sentinel-2 MSI-Level 2A Products Algorithm Theoretical Basis Document*.
- Rocchini, D., Andreo, V., Förster, M., Garzon-Lopez, C. X., Gutierrez, A. P., Gillespie, T. W., ... Neteler, M. (2015). Potential of remote sensing to predict species invasions: A modelling perspective. *Progress in Physical Geography*, 39(3), 283–309.
<https://doi.org/10.1177/0309133315574659>
- Roiz, D., Eritja, R., Melero-Alcibar, R., Molina, R., Marques, E., Ruiz, S., ... Lucientes, J. (2007). Distribución de Aedes (Stegomyia) albopictus (Skuse , 1894)(Diptera , Culicidae) en España. *Boletín de La Sociedad Entomológica Aragonesa*, (40), 523–526.
- Roiz, D., Eritja, R., Molina, R., Melero-Alcibar, R., & Lucientes, J. (2008). Initial Distribution Assessment of *Aedes albopictus* (Diptera: Culicidae) in the Barcelona, Spain, Area. *Journal of Medical Entomology*, 45(3), 347–352.
[https://doi.org/10.1603/0022-2585\(2008\)45](https://doi.org/10.1603/0022-2585(2008)45)
- Romi, R., Toma, L., Severini, F., & Luca, M. Di. (2008). Twenty years of the presence of *Aedes albopictus* in Italy—From the annoying pest mosquito to the real disease vector. *Eur Inf Dis*, (January), 98–101.

- Salehi, B., Zhang, Y., Zhong, M., & Dey, V. (2012). Object-based classification of urban areas using VHR imagery and height points ancillary data. *Remote Sensing*, 4(8), 2256–2276. <https://doi.org/10.3390/rs4082256>
- Silva, Laís & Bacani, Vitor, (2017). Classificação orientada a objeto (GEOBIA) aplicada à análise das alterações do uso e cobertura da terra da Bacia Hidrográfica do Rio da Prata em Jardim e Bonito - MS. *GEOGRAFIA* (Londrina). 26. 05. 10.5433/2447-1747.2017v26n1p05.
- Sirami, C., Caplat, P., Popy, S., Clamens, A., Arlettaz, R., Jiguet, F., ... Martin, J. L. (2016). Impacts of global change on species distributions: obstacles and solutions to integrate climate and land use. *Global Ecology and Biogeography*, 26(4), 385–394. <https://doi.org/10.1111/geb.12555>
- Solano, F., Di Fazio, S., & Modica, G. (2019). A methodology based on GEOBIA and WorldView-3 imagery to derive vegetation indices at tree crown detail in olive orchards. *International Journal of Applied Earth Observation and Geoinformation*, 83, 101912. <https://doi.org/10.1016/j.jag.2019.101912>
- Sousa, A., & Silva, J. (2011). *Fundamentos Teóricos de Detecção Remota*. Universidade de Évora.
- Schaffner, F., Hendrickx, G., Ducheyne, E., Medlock, J. M., & Avenell, D. (2009). Development of *Aedes albopictus* risk maps. *ECDC, Technical Report*.
- Schiewe, J. (2002). SEGMENTATION OF HIGH-RESOLUTION REMOTELY SENSED DATA - CONCEPTS , APPLICATIONS AND PROBLEMS. *Symposium on Geospatial Theory, Processing and Applications*, (June), 1–6.
- Sharma, R., Ghosh, A., & Joshi, P. K. (2013). Decision tree approach for classification of remotely sensed satellite data using open source support. *Journal of Earth System Science*, 122(5), 1237–1247. <https://doi.org/10.1007/s12040-013-0339-2>
- SUHET. (2013). *Sentinel-2 User Handbook*. European Space Agency.
- Takken, W., & Berg, H. van den. (2019). *Manual on prevention of establishment and control of mosquitoes of public health importance in the WHO European Region (with special reference to invasive mosquitoes)*. World Health Organization Regional Office for Europe.

- Tiago, P., Pereira, H. M., & Capinha, C. (2017). Using citizen science data to estimate climatic niches and species distributions. *Basic and Applied Ecology*.
<https://doi.org/10.1016/j.baae.2017.04.001>
- Timofeev, R. (2004). *Classification and Regression Trees (CART) Theory and Applications*. Humboldt University. <https://doi.org/10.1111/j.1440-1746.1998.tb00550.x>
- Tilton, J. C., Aksoy, S., & Tarabalka, Y. (2015). Image segmentation algorithms for land categorization. *Remotely Sensed Data Characterization, Classification, and Accuracies, 1*, 317–342. <https://doi.org/10.1201/b19294>
- Tran, B. N., Tanase, M. A., Bennett, L. T., & Aponte, C. (2018). Evaluation of spectral indices for assessing fire severity in Australian temperate forests. *Remote Sensing*, 10(11). <https://doi.org/10.3390/rs10111680>
- UN-Spider. (n.d.). Data application of the month: Vegetation indices. Acedido a 20 de agosto de 2020 no portal UN-SPIDER Knowledge em <http://www.un-spider.org/links-and-resources/data-sources/daotm/daotm-vegetation>
- Veljanovski, T., Kanjir, U., & Oštir, K. (2011). Object-based image analysis of remote sensing data. *Geodetski Vestnik*, 55(04), 641–664.
<https://doi.org/10.15292/geodetski-vestnik.2011.04.641-664>
- Waldock, J., Chandra, N. L., Lelieveld, J., Proestos, Y., Michael, E., Christophides, G., & Parham, P. E. (2013). The role of environmental variables on *Aedes albopictus* biology and chikungunya epidemiology. *Pathogens and Global Health*, 107(5), 224–241. <https://doi.org/10.1179/2047773213Y.0000000100>
- Vogels, Marjolein & De Jong, Steven & Sterk, Geert & Douma, Harke & Addink, Elisabeth. (2019). Spatio-Temporal Patterns of Smallholder Irrigated Agriculture in the Horn of Africa Using GEOBIA and *Sentinel-2* Imagery. *Remote Sensing*. 11. 143. [10.3390/rs11020143](https://doi.org/10.3390/rs11020143).
- Wang, Y., Qi, Q., Liu, Y., Jiang, L., & Wang, J. (2019). Unsupervised segmentation parameter selection using the local spatial statistics for remote sensing image segmentation. *International Journal of Applied Earth Observation and Geoinformation*, 81(May), 98–109. <https://doi.org/10.1016/j.jag.2019.05.004>

- Wagner, K. (2018). Geographic Information Systems and Glacial Environments. *Past Glacial Environments: Second Edition* (Vol. 14). <https://doi.org/10.1016/B978-0-08-100524-8.00015-4>
- Tijssen, M. *Mapping land use and landcover change in Kalmar, Sweden, using object-based change detection*. University of Salzburg
- Weng, Q. (2010). Remote Sensing and GIS Integration. *Theories, Methods and Applications* (1st ed.). McGraw-Hill.
- Wezyk, Piotr & Hawryło, Paweł & Szostak, Marta & Pierzchalski, Marcin. (2016). Using Geobia and Data Fusion Approach for Land use and Land Cover Mapping. *Quaestiones Geographicae*. 35. 93-104. 10.1515/quageo-2016-0009.
- Xia, Z., Chunkang, Z., Li, H., & Zhu, L. (2020). A Road Extraction Method Based On Naked High Resolution Remote Sensing Image. <https://doi.org/10.5194/isprs-archives-XLII-3-W10-671-2020>
- Xue, J., & Su, B. (2017). Significant remote sensing vegetation indices: A review of developments and applications. *Journal of Sensors*, Vol. 2017. <https://doi.org/10.1155/2017/1353691>
- Yantao, X., Nguyen, T., & Cheng, L. (2019). European Journal of Remote Sensing Preliminary comparative assessment of various spectral indices for built-up land derived from Landsat-8 OLI and *Sentinel-2A* MSI imageries Preliminary comparative assessment of various spectral indices for built-up land. *European Journal of Remote Sensing*, 52(1), 240–252. <https://doi.org/10.1080/22797254.2019.1584737>
- Ye, W., Li, X., Chen, X., & Zhang, G. (2014). A spectral index for highlighting forest cover from remotely sensed imagery. *Land Surface Remote Sensing II*, 9260, 92601L. <https://doi.org/10.1117/12.2068775>
- Yengoh, G. T., Dent, D., Olsson, L., Tengberg Compton, A. E., & Tucker, J. (2014). *The use of the Normalized Difference Vegetation Index (NDVI) to assess land degradation at multiple scales: a review of the current status, future trends, and practical considerations*.
- Yong. Z., Zhang, P., Benediktsson, J. A., & Shi, W. Z. (2014). Morphological profiles

based on differently shaped structuring elements for classification of images with very high spatial resolution. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 7(12), 4644–4652.

<https://doi.org/10.1109/JSTARS.2014.2328618>

Zou, L., Miller, S. N., & Schmidtman, E. T. (2006). Mosquito Larval Habitat Mapping Using Remote Sensing and GIS: Implications of Coalbed Methane Development and West Nile Virus. *Journal of Medical Entomology*, 43(5), 1034–1041.

<https://doi.org/10.1093/jmedent/43.5.1034>

Lista de figuras

Figura 1 - NDVI à escala regional extraído do sensor Metop AVRRR (LSA SAF, 2019) para a África subsaariana, e figura - 2, NDVI extraído à escala urbana extraído obtida por imagem aérea de muito alta resolução (Hänninen, 2019) para o parque do capitólio em Sacramento	6
Figura 2 – Esquema da estrutura da dissertação	10
Figura 3 – Distribuição global de ocorrências de <i>Aedes albopictus</i> 1990 - 2020 (adaptado de GBIF , 2020).....	15
Figura 4 – Distribuição potencial baseada em fatores biofísicos para o ano de 2018 (adaptado de Kamal, et al. 2018).....	16
Figura 5 - Envelope bioclimático para <i>Aedes albopictus</i> (gama nativa e invasiva), derivado por sobreposição geograficamente explícita de registos de presença com variáveis bioclimáticas anuais (Fisher et al. 2011).	18
Figura 6 - Mapa da distribuição conhecida de <i>Aedes albopictus</i> na Europa a nível regional, agosto de 2019 (ECDC & EFSA, 2019).	20
Figura 7 - Mapa da distribuição conhecida de <i>Aedes albopictus</i> na Europa a nível regional, maio de 2020 (ECDC & EFSA, 2020).	21
Figura 8 - Processo de Detecção Remota passiva. (Muralli, 2015, p. 220)	22
Figura 9 – Localização da área de estudo de Barcelona e arredores, no nordeste da península Ibérica.	26
Figura 10 – Gráfico Termopluiométrico para as normais climatológicas do observatório Fabra (período de referência 2007-2016) – (Servei meteorològic de Catalunya , 2017)	27
Figura 11 – Contextualização das ocorrências dentro da área de estudo, sobre composição colorida (cor verdadeira) Sentinel-2 MSI. As ocorrências de mosquitos distribuídas pela área de estudo irão permitir confrontar e validar resultados.....	28
Figura 12 – Níveis de processamento de imagens Sentinel-2 MSI, desde o nível 0 até ao nível 1C (SUHET, 2015)	34
Figura 13 – Exemplo de transformação de uma imagem nível 1C (a) para uma imagem nível 2A (b) da área do estuário do rio El Llobregat. As imagens Nível	

1C também podem ser referidas como TOA image data, e as imagens de nível 2A como BOA Image data.....	39
Figura 14 - Comparação entre histogramas de bandas diferentes, histograma a) corresponde à banda 4 (Vermelho) e o histograma b) corresponde a banda 8 (PIV). Os histogramas permitem avaliar a distribuição de níveis digitais na imagem, neste exemplo, o PIV mostra mais frequências em níveis mais altos pela influência da vegetação presente na cena	42
Figura 15 – Comparação entre o brilho de duas bandas com médias diferentes das localidades de Sant Andreu de la barca, Castellbisbal e Martorell e áreas circundantes, a imagem a) corresponde à banda 4 e a imagem b) à banda 8	42
Figura 16 – Assinatura espectral da vegetação fotossintética (verde), vegetação não fotossintética (seca) e solo (Clark, 1999)	48
Figura 17 – Esquema de extração mensal, sazonal e anual de valores NDVI para a produção de mapas de vegetação	50
Figura 18 – Evolução do índice de vegetação acima do limiar proposto para o Parc de la Ciutadella e redondezas, entre meses de janeiro a) e julho b).....	51
Figura 19 – Exemplos de segmentações, sobre segmentação (a), Sub segmentação (b) e segmentação ideal (c) sobre a área industrial próxima do rio Llobregatt.....	63
Figura 20 - O algoritmo de divisão das Árvores de classificação e regressão onde t_{parent} , t_{left} , t_{right} são respetivamente o nó progenitor, nó esquerdo e direito x_j a variável j e x_j^R o melhor valor de divisão de x_j (Timofeev, 2004, p.9).....	69
Figura 21 – Classificação escolhida após a aplicação dos métodos de OBIA. Originada pela aplicação do classificador LibSVM sobre as bandas 8,4,3 e 2 e os quatro NDVIs sazonais.	82
Figura 22 – Reclassificação originada pela fusão de classes cuja confusão se admite e que se encontrem dentro da mesma temática para correcção.....	86
Figura 23 – Exemplificação de confusão por sobreposição de classes, esta sobreposição ocorre quando objetos que por natureza partilham uma assinatura espectral parecida, são geralmente diferenciados pelo desvio	

padrão, no entanto quando este também é similar, resultante da morfologia do terreno ou existência de misturas causadas por diferentes assinaturas no mesmo objeto, estas podem surgir sobrepostas e classificadas como uma classe diferente.....	88
Figura 24 – Exemplo de correção de imprecisões temáticas para a área da foz do El Llobregat. A principal diferença prende-se à limpeza de áreas de areal c) que tinham sido classificadas como urbano b), bem como áreas húmidas c) que foram consideradas vegetação arbórea b), ou áreas de solo a descoberto ou rocha c) classificadas como areal b). Esta área é bastante complexa com várias tipologias de uso do solo, bem como dinâmica, pois é afetada pelas marés, é possível denotar isto pelo corpo de água exacerbado em ambas as classificações b) e c) mas não presente na composição de imagem a).....	89
Figura 25 – Esquema de exemplo do resultado do cálculo da distância bilateral em relação à fronteira das duas variáveis.	93
Figura 26 – Mapa de valores de distância contínua em relação à fronteira onde se intersectam as variáveis vegetação e áreas urbanas.....	94
Figura 27 – Regressão exponencial do número de observações (a), e pontos de controlo (b) em função da distância a áreas urbanas e vegetação.	95
Figura 28 – Probabilidade de ocorrências em função da previsão obtida pela regressão exponencial entre as ocorrências e a distância às variáveis de habitat	97
Figura 29 – Regressão exponencial do número de observações (a), e pontos de controlo (b) em função da distância a vias de comunicação.	98
Figura 30 – Mapa de classes de distância às vias de comunicação como fatores de probabilidade de introdução de <i>Aedes albopictus</i>	99
Figura 31 – Mapa de vegetação para o ano de 2019 originado pelo cálculo da média para os doze meses do ano, e posterior extração de valores superiores ao limiar proposto	102
Figura 32 – Exemplo de sobreposição entre as ocorrências e áreas de maior atividade clorofilina em contexto urbano para o Parc de la Ciutatella e áreas circundantes.....	103

Figura 33 – Distribuição do número de ocorrências (%) em função da distância (m) à vegetação	104
Figura 34 – Regressão exponencial de ocorrências a) e de pontos aleatórios b) em função da distância à vegetação	105
Figura 35 – Classificação do uso do solo, resultante da aplicação das medidas de mitigação de imprecisão temática auxiliadas por cartografia de referência para o ano 2019.....	108
Figura 36 – Mapa de discordância temática e semântica entre a classificação do uso do solo produzida com o mapa de uso do solo descarregado do departamento de território e sustentabilidade da Catalunha.....	109
Figura 37 – Proposta de corredores de propagação potencial resultante da sobreposição não ponderada das de camadas de probabilidade de ocorrências.....	112

Lista de tabelas

Tabela 1 - Bandas espectrais do sensor MSI a bordo da missão Sentinel-2 (S2A & S2B) (ESA, 2015)	32
Tabela 2 – Parâmetros teste para a segmentação de imagem utilizando o algoritmo MeanShift	64
Tabela 3 – Comparação dos valores agrupados de separabilidade das classes, para as combinações das bandas espectrais 8,4,3 e 2, estas combinadas com o NDVI anual, e as bandas espectrais com os 4 NDVI sazonais.	67
Tabela 4 – Parametrização para a criação do modelo de classificação de vetores utilizando o algoritmo Decision Tree.....	75
Tabela 5 – Parametrização para a criação do modelo de classificação de vetores utilizando o treino de vetores LibSVM.	76
Tabela 6 – Classes e número de amostras extraídas para validação temática	80
Tabela 7 – Matriz de admissibilidade de imprecisão temática.....	84
Tabela 8 – Frequência de ocorrências em função de classes de 10 metros de distância	94

Tabela 9 – Coeficiente de Correlação de Pearson entre o número de ocorrências e a distância à vegetação, uma correlação negativa indica que quanto maior a distância menor o número de ocorrências	104
Tabela 10 – Índices de Kappa para as classificações efetuadas.....	106
Tabela 11 – Precisão Global para as classificações efetuadas	106
Tabela 12 – Índice de kappa e precisão global para os três anos anteriores utilizando LibSVM	107
Tabela 13 – Áreas em km ² ocupadas por cada classe por estágio da classificação e respectiva variação	110
Tabela 14 – Concordância entre a classificação descarregada e a classificação produzida.	111

Lista de fluxogramas

Fluxograma 1 – Passos para análise da dinâmica no espaço e no tempo das condições de propagação do vetor.....	8
Fluxograma 2 – Estrutura do modelo de uvspec (adaptado de Mayer & Kylling, 2005)	37
Fluxograma 3 – Descrição dos procedimentos para a análise orientada a objetos.....	57

ANEXOS

Anexo I – Dados Sentinel-2 MSI

Data	Referência
13/04/2019	S2A_MSIL2A_20190413T104021_N0207_R008_T31TDF_20190413T141321
25/10/2019	S2B_MSIL2A_20191025T104029_N0213_R008_T31TDF_20191025T140913
28/01/2019	S2B_MSIL2A_20190128T104309_N0211_R008_T31TDF_20190128T134652
27/02/2019	S2B_MSIL2A_20190227T104019_N0211_R008_T31TDF_20190227T165024
24/03/2019	S2A_MSIL2A_20190324T104021_N0211_R008_T31TDF_20190324T192037
23/05/2019	S2A_MSIL2A_20190523T104031_N0212_R008_T31TDF_20190523T135708
27/06/2019	S2B_MSIL2A_20190627T104029_N0212_R008_T31TDF_20190627T135004
22/07/2019	S2A_MSIL2A_20190722T104031_N0213_R008_T31TDF_20190722T113040
26/08/2019	S2B_MSIL2A_20190826T104029_N0213_R008_T31TDF_20190826T140844
20/09/2019	S2A_MSIL2A_20190920T104021_N0213_R008_T31TDF_20190920T120330
24/11/2019	S2B_MSIL2A_20191124T104329_N0213_R008_T31TDF_20191124T135417
24/12/2019	S2B_MSIL2A_20191224T104349_N0213_R008_T31TDF_20191224T123641
28/01/2018	S2A_MSIL1C_20180128T104311_N9999_R008_T31TDF_20200214T191605
23/01/2018	S2B_MSIL1C_20180123T104329_N9999_R008_T31TDF_20200214T195656
18/04/2018	S2A_MSIL2A_20180418T104021_N0207_R008_T31TDF_20180418T125356
22/06/2018	S2B_MSIL1C_20180622T104019_N0208_R008_T31TDF_20180622T132235
12/07/2018	S2B_MSIL2A_20180712T104019_N9999_R008_T31TDF_20200214T212101
06/08/2018	S2A_MSIL2A_20180806T104021_N0208_R008_T31TDF_20180806T142805
20/09/2018	S2B_MSIL2A_20180920T104019_N0208_R008_T31TDF_20180920T153221
25/10/2018	S2A_MSIL2A_20181025T104121_N0209_R008_T31TDF_20181025T135730
24/11/2018	S2A_MSIL2A_20181124T104341_N0211_R008_T31TDF_20181124T120446
24/12/2018	S2A_MSIL2A_20181224T104441_N0211_R008_T31TDF_20181224T120348
12/07/2017	S2A_MSIL1C_20170712T104021_N9999_R008_T31TDF_20200215T163608
13/01/2017	S2A_MSIL1C_20170113T104401_N9999_R008_T31TDF_20200215T123551
22/02/2017	S2A_MSIL1C_20170222T104031_N9999_R008_T31TDF_20200215T121229
14/03/2017	S2A_MSIL1C_20170314T104011_N9999_R008_T31TDF_20200215T115529
13/04/2017	S2A_MSIL1C_20170413T104021_N9999_R008_T31TDF_20200215T133223
23/05/2017	S2A_MSIL1C_20170523T104031_N9999_R008_T31TDF_20200215T144226
12/06/2017	S2A_MSIL1C_20170612T104021_N9999_R008_T31TDF_20200215T152709
10/09/2017	S2A_MSIL1C_20170910T104021_N9999_R008_T31TDF_20200215T175949
25/10/2017	S2B_MSIL1C_20171025T104109_N9999_R008_T31TDF_20200215T222546
19/11/2017	S2A_MSIL1C_20171119T104331_N9999_R008_T31TDF_20200215T185218
24/12/2017	S2B_MSIL1C_20171224T104429_N9999_R008_T31TDF_20200215T225849
07/06/2016	S2A_MSIL1C_20160607T104032_N0202_R008_T31TDF_20160607T104026
17/07/2016	S2A_MSIL1C_20160717T104032_N0204_R008_T31TDF_20160717T104833
26/08/2016	S2A_MSIL1C_20160826T104022_N0204_R008_T31TDF_20160826T104023
05/09/2016	S2A_MSIL1C_20160905T104022_N0204_R008_T31TDF_20160905T104245
24/11/2016	S2A_MSIL1C_20161124T104352_N0204_R008_T31TDF_20161124T104816
24/12/2016	S2A_MSIL1C_20161224T104442_N0204_R008_T31TDF_20161224T104438

Anexo II – Parâmetros de segmentação Morphological Profiles Based Segmentation

Parâmetros	T 1	T 2	T 3	T 4	T 5	T 6	T 7	T 8
Profile Size	5	10	10	5	5	20	30	25
Initial Radius	1	5	1	5	1	1	1	1
Radius Step	1	5	1	1	5	1	1	1
8 neighbor conectivity	off	off	on	on	on	off	off	off
Minimum object size	1	1	1	1	1	1	5	5
Tiles Size	1024	1024	1024	1024	1024	3800	1024	1024

Anexo III – Tabela de frequências de ocorrências em função da distância à vegetação e à fronteira de conjugação da vegetação e áreas urbanas.

Classes de distância	(F)Vegetação	(F) Aleatório Vegetação	(F) Fronteira urbano/vegetação	(F) Aleatório Urbano/vegetação
0	270	523	78	44
10	212	113	314	142
20	124	70	167	102
30	94	40	109	81
40	61	33	69	54
50	33	17	42	46
60	31	21	34	43
70	20	15	26	37
80	2	6	4	21
90	5	5	7	22
100	2	4	2	16
110	1	6	1	17
120	3	4	3	15
130	1	2	1	16
140	1	2	3	14
150	0	2	0	7
160	0	2	0	9
170	0	2	0	11
180	0	0	0	13
190	0	1	0	5
200	0	0	0	9
210	0	3	0	0
240	0	1	0	0
260	0	1	0	0
320	0	1	0	0
330	0	1	0	0
370	0	1	0	0
500	0	1	0	0
550	0	1	0	0
930	0	1	0	0
1020	0	1	0	0
1030	0	1	0	0
1040	0	1	0	0
1290	0	1	0	0
1300	0	1	0	0
1640	0	2	0	0
1680	0	2	0	0

Anexo IV - Frequência de ocorrências em função da distância às vias de comunicação

Distância	(F) Ocorrências	(F) Aleatório	Distância	(F) Ocorrências	(F) Aleatório	Distância	(F) Ocorrências	(F) Aleatório
0	3	7	450	0	4	950	0	3
10	55	65	460	4	4	960	0	3
20	47	49	470	1	6	970	0	2
30	39	31	480	2	3	990	0	1
40	35	33	490	1	5	1000	0	2
50	39	17	500	0	2	1020	0	1
60	56	33	510	0	2	1030	0	3
70	53	19	520	1	8	1040	0	4
80	91	35	530	0	5	1050	0	3
90	39	30	540	1	4	1070	0	1
100	39	21	550	0	5	1080	0	3
110	46	18	560	0	4	1090	0	1
120	23	19	570	0	2	1100	0	2
130	22	19	580	0	6	1110	0	1
140	28	21	590	1	1	1120	0	2
150	31	19	600	0	5	1150	0	1
160	24	18	610	0	5	1170	0	2
170	10	21	620	0	1	1180	0	1
180	20	13	630	0	1	1190	0	1
190	14	7	640	0	4	1200	0	3
200	12	12	660	0	7	1280	0	2
210	13	11	670	0	6	1290	0	2
220	4	15	690	0	4	1300	0	1
230	16	7	710	0	4	1320	0	1
240	10	9	720	1	6	1330	0	2
250	7	8	730	0	2	1350	0	1
260	6	7	740	0	3	1390	0	1
270	1	10	750	0	4	1400	0	1
280	3	9	760	1	2	1410	0	1
290	4	4	770	1	2	1440	0	2
300	2	6	780	0	7	1490	0	1
310	2	4	790	0	3	1590	0	1
320	5	5	800	0	5	1670	0	1
330	3	11	810	0	3	1830	0	1
340	3	2	820	0	3	2490	0	1
350	3	6	830	0	2	2550	0	1
360	3	5	840	0	2	2710	0	1
370	3	4	850	1	3	2720	0	1
380	3	5	860	0	2	2740	0	1
390	6	6	870	0	5	2950	0	1
400	10	4	880	0	2	2960	0	1
400	10	4	890	0	3	3230	0	1
410	4	1	900	1	1	3230	0	1
420	2	5	910	0	2	3250	0	2
430	3	4	920	0	1			
440	2	2	940	0	2			