



**FCTUC** DEPARTAMENTO DE ENGENHARIA CIVIL  
FACULDADE DE CIÊNCIAS E TECNOLOGIA  
UNIVERSIDADE DE COIMBRA

# **Drones e Modelos de Drenagem Urbana: Classificação Automática de Imagens para Identificação dos Usos do Solo**

Dissertação apresentada para a obtenção do grau de Mestre em Engenharia do Ambiente na Especialidade de Território e Gestão do Ambiente

Autor

**Fabíola Tatiana Gomes Freitas**

Orientador

**Professor Doutor Nuno Eduardo da Cruz Simões**

**Doutor João Paulo Correia Leitão**

Esta dissertação é da exclusiva responsabilidade do seu autor, não tendo sofrido correcções após a defesa em provas públicas. O Departamento de Engenharia Civil da FCTUC declina qualquer responsabilidade pelo uso da informação apresentada

**Coimbra, 29 de julho, 2016**

## AGRADECIMENTOS

Quero deixar o meu profundo agradecimento a todos os que de uma forma direta ou indireta fizeram parte do meu crescimento pessoal e acadêmico, em especial, para aqueles que contribuíram para a concretização desta dissertação.

Aos meus orientadores, o Professor Doutor Nuno Simões e o Doutor João Paulo Leitão pela orientação prestada, em que foi constante a disponibilidade e preocupação para auxiliar em todos os aspetos necessários.

Aos meus amigos pelo carinho, companheirismo e pelas aprendizagens que não se encontram em livros. Obrigada pelos fantásticos momentos vividos ao longo do meu percurso académico.

À família Garcia por todo o apoio, carinho e preocupação. Este agradecimento é pouco para tudo o que fizeram e fazem constantemente por mim.

Ao João Garcia pelos conhecimentos transmitidos sobre *Machine Learning*, essenciais ao desenvolvimento da presente dissertação. Quero também agradecer pela cumplicidade, carinho, paciência e motivação ao longo dos últimos cinco anos e meio.

Aos meus pais pelo amor incondicional e por tudo o que me providenciaram ao longo destes anos. Obrigada pelo constante incentivo para concretizar os meus sonhos e permitirem que os mesmos se realizem.

## RESUMO

As inundações são dos desastres naturais que mais pessoas afetam em todo o mundo, provocando, em regra, elevados prejuízos. O risco da sua ocorrência tem vindo aumentar devido à combinação das alterações climáticas com a constante modificação dos usos do solo.

Uma metodologia para estudar e prevenir estes fenómenos consiste em simular as condições de inundação através da modelação de drenagem urbana, em que através desta é recriada a resposta de uma bacia em condições de precipitação intensa. Para que tal seja reproduzido é necessária uma vasta quantidade de dados, tais como dados meteorológicos, geográficos, da rede de coletores, entre outros. Os elementos geográficos, tais como a distribuição das ocupações do solo e o perfil do terreno são dados importantes na modelação de drenagem urbana. A sua qualidade influencia os resultados da simulação pelo que é essencial uma escolha adequada dos mesmos, no entanto a sua aquisição poderá ser dispendiosa e requerer diversos recursos técnicos e humanos.

Os *drones* são veículos aéreos não tripulados que podem ser utilizados nas mais diversas áreas. O seu baixo custo e flexibilidade permite realizar diversas tarefas tais como entrega de encomendas, monitorização de incêndios, transporte de desfibriladores em emergências médicas, etc. Entre as diversas aplicações, estes equipamentos apresentam uma boa solução de aquisição de dados geográficos devido às razões acima apresentadas, como também à vantagem de poderem voar a baixas altitudes, permitindo a aquisição de imagens de elevada resolução.

Neste trabalho foram analisadas duas imagens de uma sub-bacia da cidade de Lucerna, captadas por *drone* e satélite, com intuito de proceder à identificação de áreas impermeáveis através de técnicas de imagem. Foi ainda analisado o potencial da utilização de ferramentas ArcGis e Google Earth Engine na análise de imagens.

**Palavras-chave:** drenagem urbana, *drones*, deteção remota, classificação

## ABSTRACT

Floods are one of the natural disasters that affect the most people around the world, causing, generally, high losses. The risk of occurrence has been increasing due to the combination of the constant change of land use and climate change.

A methodology to study and prevent these phenomena is to simulate flood conditions through urban drainage modeling, in which the response of a catchment to heavy rainfall conditions is recreated. In order to do this, a vast amount of data is required, such as meteorological and geographic data, the sewers network, and others. Geographic features, such as the distribution of land uses and the terrain slope profile data are important in urban drainage modelling. Their quality influences the simulation results. As such, it is essential to choose them properly them, however they can be expensive and require several technical and human resources.

Drones are unmanned Aerial vehicles that can be used in different applications. The low cost and flexibility of this equipment allows various tasks such as deliveries, fire monitoring, defibrillators transportation in medical emergencies, etc, to be done. Among the various uses, these devices offer a good solution for acquiring spatial data for the reasons aforementioned and also they have advantage of being able to fly at low altitudes, enabling acquisition of high resolution imagery.

In this study two images of Lucerne are analyzed, one captured by drone and the other by satellite, in order to proceed to the identification of impervious areas through imaging techniques. It was also analyzed the potential of using ArcGIS and Google Earth Engine tools for imagery analysis.

**Keywords:** urban drainage, *drones*, remote sensing, classification

## ÍNDICE

1	INTRODUÇÃO.....	1
1.1	Contexto e Motivação.....	1
1.2	Objetivos.....	2
1.3	Estrutura da Tese.....	2
2	ESTADO DA ARTE.....	3
2.1	Drenagem Urbana.....	3
2.1.1	Introdução.....	3
2.1.2	Efeitos da Urbanização na Drenagem Pluvial.....	4
2.2	Modelação Computacional.....	5
2.2.1	Modelos Hidrológicos.....	5
2.2.2	Modelos Hidráulicos e Drenagem Dual.....	8
2.2.3	Incertezas associadas aos modelos de drenagem.....	10
2.3	Deteção Remota.....	11
2.3.1	Técnicas de Deteção Remota.....	12
2.3.1.1	Tecnologia <i>LIDAR</i> .....	13
2.3.1.2	Fotografias aéreas.....	13
2.3.1.3	Termografia.....	14
2.3.2	Plataformas de deteção remota.....	14
2.3.2.1	Satélite.....	14
2.3.2.2	<i>Unmanned Aerial Vehicles</i> .....	15
2.4	Identificação dos Usos do Solo.....	16
2.5	Reconhecimento de padrões.....	17
3	METODOLOGIA.....	18
3.1	Caso de Estudo.....	18
3.2	Imagens.....	19
3.3	Técnicas de Análise de Imagens.....	20
3.3.1	<i>Software</i> .....	20
3.3.2	Procedimento.....	21
3.3.3	Algoritmos.....	23
3.3.3.1	Pré-avaliação.....	23
3.3.3.2	<i>Fast Naive Bayes</i> .....	24
3.3.3.3	Margin Support Vector Machine.....	25

---

3.3.3.4	Método da Máxima Verossimilhança .....	27
3.4	Avaliação da Precisão Global .....	27
4	RESULTADOS E DISCUSSÃO.....	31
4.1	Classificação com recurso ao Google Earth Engine .....	31
4.1.1	Imagem <i>Landsat</i> .....	31
4.1.2	Ortofotomapa.....	37
4.2	Avaliação feita com recurso ao ArcMap .....	47
4.2.1	Efeito da seleção da zona de treino .....	53
4.2.2	Efeito da seleção dos pontos de controlo.....	59
5	CONCLUSÕES E TRABALHOS FUTUROS .....	63
	REFERÊNCIAS BIBLIOGRÁFICAS .....	66
	ANEXOS .....	69

## ÍNDICE DE FIGURAS

Figura 2.1 - Efeito da urbanização na resposta hidrológica (Butler e Davies, 2011)	4
Figura 2.2- Exemplo de uma rede de drenagem 1D/1D (Butler and Davies, 2011)	10
Figura 2.3 - Rede de drenagem dual 1D/2D (Pedroso de Lima <i>et al.</i> , 2013)	10
Figura 3.1 – Wartegg, Lucerna (Air Luzern)	18
Figura 3.2 - Fotografia <i>Landsat</i> , Lucerna (GEE)	19
Figura 3.3 - Ortofotomapa de Lucerna (Leitão <i>et al.</i> , 2016)	19
Figura 3.4 – Pontos de controlo 1	22
Figura 3.5 - Exemplo do funcionamento do SVM (a) e Vetores de Suporte (b) (James, G. <i>et al.</i> , 2015)	26
Figura 3.6 – <i>Underfitting</i> , <i>Good fit</i> e <i>overfitting</i>	29
Figura 4.1 – Classificação obtida para o treino de 100 polígonos através do FNB	32
Figura 4.2 - Matriz de erro e precisão global: treino de 100 polígonos realizado com o FNB	32
Figura 4.3 - Classificação obtida para o treino de 100 polígonos através do <i>Margin SVM</i>	32
Figura 4.4 – Matriz de erro e precisão global: treino de 100 polígonos realizado com o <i>Margin SVM</i>	32
Figura 4.5 – Treino de 101 polígonos	33
Figura 4.6 - Classificação obtida para o treino de 101 polígonos através do FNB	34
Figura 4.7 - Matriz de erro e precisão global: treino de 101 polígonos realizado com o FNB	34
Figura 4.8 - Classificação obtida para o treino de 101 polígonos através do <i>Margin SVM</i>	34
Figura 4.9 - Matriz de erro e precisão global: treino de 101 polígonos realizado com o <i>Margin SVM</i>	34
Figura 4.10 - Classificação obtida para o treino de 500 polígonos através do FNB	36
Figura 4.11 - Matriz de erro e precisão global: treino de 500 polígonos realizado com o FNB	36
Figura 4.12 - Matriz de erro e precisão global: treino de 500 polígonos realizado com o <i>Margin SVM</i>	36
Figura 4.13 - Matriz de erro e precisão global: treino de 500 polígonos realizado com o <i>Margin SVM</i>	36

---

Figura 4.14 – Classificação obtida para o treino de 20 polígonos através do algoritmo FNB	37
Figura 4.15 - Matriz de erro e precisão global: treino de 20 polígonos realizado com o FNB	37
Figura 4.16 - Classificação obtida para o treino de 200 polígonos através do algoritmo FNB	39
Figura 4.17 - Matriz de erro e precisão global: treino de 200 polígonos realizado com o FNB	39
Figura 4.18 – Classificação obtida para o treino de 400 polígonos através do algoritmo FNB	40
Figura 4.19 - Matriz de erro e precisão global: treino de 400 polígonos realizado com o FNB	40
Figura 4.20 – Classificação obtida para o treino de 20 polígonos através do algoritmo <i>Margin SVM</i>	42
Figura 4.21 - Matriz de erro e precisão global: treino de 20 polígonos realizado com o <i>Margin SVM</i> .	42
Figura 4.22 – Classificação obtida para o treino de 100 polígonos através do algoritmo <i>Margin SVM</i>	43
Figura 4.23 - Matriz de erro e precisão global: treino de 100 polígonos realizado com o <i>Margin SVM</i>	43
Figura 4.24 – Classificação obtida para o treino de 300 polígonos através do algoritmo <i>Margin SVM</i>	44
Figura 4.25 Matriz de erro e precisão global: treino de 300 polígonos realizado com o <i>Margin SVM</i>	44
Figura 4.26 – Evolução da precisão global para todos os treinos realizados com o FNB e <i>Margin SVM</i> .	46
Figura 4.27 – Classificação obtida para o treino de 20 polígonos	47
Figura 4.28 – Classificação obtida para o treino de 50 polígonos	48
Figura 4.29 – Classificação obtida para o treino de 100 polígonos	49
Figura 4.30 - Evolução da precisão consoante o treino	51
Figura 4.31 – Evolução da precisão global dos três algoritmos	52
Figura 4.32 – Zona de treino B	53
Figura 4.33 – Classificação obtida para o treino de 50 polígonos com zona de treino B	54
Figura 4.34 – Classificação obtida para o treino de 100 polígonos com zona de treino B	55

---

Figura 4.35 – Zona de treino C	56
Figura 4.36 – Classificação obtida para o treino de 50 polígonos com zona de treino C	57
Figura 4.37 – Classificação obtida para o treino de 100 polígonos com zona de treino C	58
Figura A.1 – Bacia central da cidade Coimbra (Imagem gentilmente cedida pelo Prof. Doutor Nuno Simões)	69
Figura A.2 – Imagem <i>Landsat</i> , Coimbra (GEE)	69
Figura B.1 - Zona de Treino A	70
Figura C.1 – Classificação obtida para o treino de 300 polígonos através do algoritmo FNB	71
Figura C.2 – Matriz de erro e precisão global: treino de 300 polígonos realizado com FNB	71
Figura C.3 – Classificação obtida para o treino de 400 polígonos através do algoritmo <i>Margin SVM</i>	71
Figura C.4 – Matriz de erro e precisão global: treino de 400 polígonos realizado com o <i>Margin SVM</i>	71

## ÍNDICE DE QUADROS

Quadro 3.1 - Matriz de erro	28
Quadro 3.2 - Significado de Kappa. Adaptado de Viera e Garret (2015)	30
Quadro 4.1 – Precisão global obtida para o FNB para a classificação do ortofotomapa	40
Quadro 4.2 - Precisão global obtida para o <i>Margin</i> SVM para a classificação do ortofotomapa	44
Quadro 4.3 - Precisão global obtida para todos os treinos realizados com o FNB e <i>Margin</i> SVM	46
Quadro 4.4 - Resultados para o treino de 20 polígonos	48
Quadro 4.5 - Resultados para o treino de 50 polígonos	49
Quadro 4.6 – Resultados para o treino de 100 polígonos	50
Quadro 4.7 – Evolução da precisão de avaliação e de Kappa com o treino	50
Quadro 4.8– Percentagem de áreas impermeáveis e permeáveis	52
Quadro 4.9 – Resultados para o treino de 50 polígonos na Zona B	54
Quadro 4.10 – Resultados para o treino de 100 polígonos na Zona B	55
Quadro 4.11 – Percentagem de áreas impermeáveis e permeáveis	56
Quadro 4.12 – Resultados para o treino de 100 polígonos na zona B	57
Quadro 4.13 – Resultados obtidos para o treino de 100 polígonos na zona C	58
Quadro 4.14 - Precisão de avaliação obtida para os treinos de 50 e 100 polígonos nas diferentes zonas	59
Quadro 4.15 – Evolução da precisão global e K para todos os treinos feitos na zona A utilizando os Pontos de Controlo 2	59
Quadro 4.16 – Evolução da precisão e de K para os treinos de 20 e 50 polígonos realizados para as zonas A, B e C	60

## ABREVIATURAS

1D/1D – Modelo de drenagem dual unidimensional

1D/2D - Modelo de drenagem dual bidimensional

MDE – Modelo Digital de Elevação

ESRI – *Environmental Systems Research Institute*

FNB – *Fast Naïve Bayes*

GEE – Google Earth Engine

IA – Inteligência Artificial

LIDAR - *Light Detection and Ranging*

SVM – *Support Vector Machines*

UAV – *Unmanned Aerial Vehicle*

# 1 INTRODUÇÃO

O contínuo crescimento das cidades tem promovido a substituição dos solos permeáveis por superfícies impermeáveis. Como o próprio nome indica, a impermeabilização não permite que a água se infiltre no solo, ficando esta retida à superfície. Este fator aliado a precipitações intensas promove a ocorrência de inundações, aumentando a sua frequência e consequências. Neste âmbito, o modo de como os usos do solo se encontram distribuídos é um fator determinante no surgimento de inundações.

A detecção remota, que permite obter informações sobre a superfície terrestre, é definida como a ciência de adquirir, processar e interpretar dados ou informações sobre objetos sem haver contacto físico com os mesmos. A aquisição de dados do terreno pode ser efetuada por diversas fontes tais como satélites, *drones*, aviões, entre outros. Os *drones* apresentam diversas vantagens quando comparados com outras fontes de detecção remota, sendo mais leves, de custo competitivo e requerem menos recursos humanos e tecnológicos para aquisição de imagens. Estes equipamentos permitem obter imagens de elevada resolução, que podem ser utilizadas para reproduzir mapas de zonas impermeáveis utilizados nos modelos de drenagem urbana. Uma vez que não é razoável criar estes mapas de forma manual, é necessário recorrer a formas mais automatizadas e rápidas de geração de mapas. Uma solução que vai ao encontro desta necessidade é recorrer à utilização de algoritmos de *Machine Learning*.

*Machine Learning* é uma área da Inteligência Artificial (IA) que consiste em treinar os computadores através de exemplos para que estes possam executar tarefas sem serem explicitamente programados para tal. Os algoritmos, intitulados de classificadores, efetuam a aprendizagem dos dados de entrada e fazem a classificação com base no que aprenderam. Para o caso apresentado neste trabalho, os algoritmos aprenderão a distinguir as cores pertencentes quer a zonas permeáveis quer a zonas impermeáveis e gerarão mapas com base nas mesmas.

## 1.1 Contexto e Motivação

A motivação para realização deste trabalho é estudar a aplicação de novas tecnologias para a identificação de áreas permeáveis e impermeáveis, com intuito de melhorar a simulação de

inundações em meio urbano. É de interesse compreender o potencial do uso de imagens captadas a baixas altitudes e quais as vantagens e desvantagens quando comparadas à utilização de imagens de satélite, na modelação de drenagem urbana. Neste contexto, serão analisados os usos do solo de uma sub-bacia da cidade de Lucerna.

## 1.2 Objetivos

O principal objetivo deste trabalho é a identificação de áreas impermeáveis através de técnicas de análise de imagem. Além das novas tecnologias que permitem adquirir imagens, existem também novas ferramentas de análise de imagem tal como o Google Earth Engine (GEE). Assim sendo, um outro objetivo foi perceber quais as vantagens e inconvenientes da utilização do Google Earth Engine face a uma ferramenta tradicional de análise de imagem como o ArcGis, para efetuar estudos desta natureza. Para tal, é necessário ter em conta a facilidade de utilização e a gama de ferramentas de análise e processamento disponíveis.

## 1.3 Estrutura da Tese

No Capítulo 2, designado de Estado da Arte, é feita uma introdução sobre as inundações em meio urbano, abordando as suas causas, impactes na sociedade e quais as informações que são o ponto de partida para o seu estudo e prevenção. De seguida são apresentadas algumas noções de deteção remota, classificação dos usos do solo e a sua importância para modelação de inundações urbanas.

A metodologia utilizada neste trabalho encontra-se descrita no Capítulo 3. Neste é apresentado o caso de estudo, os algoritmos utilizados e a forma que será efetuada a classificação dos usos do solo.

No Capítulo 4 são expostos e analisados os dados resultantes da metodologia aplicada. A classificação dos usos do solo é analisada através da visualização de imagens assim como através de outros parâmetros numéricos. Nesta secção são também referidas as vantagens e inconvenientes dos dois *software* utilizados.

No último capítulo, são apresentadas as conclusões deste trabalho e são propostas algumas vertentes interessantes de serem investigadas em trabalhos futuros.

## 2 ESTADO DA ARTE

### 2.1 Drenagem Urbana

#### 2.1.1 Introdução

A escassez de água potável e a inexistência de redes de drenagem promoveram, ao longo da História, propagação de diversas doenças. Esta relação já era relativamente conhecida há vários séculos, no entanto, só em meados do século XIX, é que foi comprovada por estudos efetuados pelo Dr. John Snow, que estudou a interdependência entre o surto de cólera, que propagava-se na cidade de Londres, e a água (Buttler and Davies, 2011). Este facto fomentou a preocupação sobre o acesso à água potável de qualidade por parte da população, assim como o encaminhamento e tratamento adequado de águas poluídas, de modo a evitar a propagação de doenças infecciosas através da água. Com a construção de redes de drenagem foram criadas melhores condições para garantia da saúde pública, promovendo melhores condições de higiene.

As redes de drenagem têm como objetivo recolher, transportar e rejeitar, em meios adequados, os efluentes indesejados provenientes de diversas atividades humanas assim como água proveniente da precipitação. Os primeiros efluentes mencionados designam-se de águas residuais e são provenientes de habitações, empresas, indústrias, etc. São caracterizados por conterem elevadas quantidades de matéria orgânica e/ou elementos nocivos para a saúde (Marques e Sousa, 2011). O processo de drenagem das águas residuais é importante para proteger não só a saúde pública como também o meio ambiente da poluição que poderá ser causada. Já as águas pluviais, provenientes da precipitação, poderão ser descarregadas num meio natural tais como cursos de água ou em solos permeáveis, ou estações de tratamento de águas residuais, dependendo do tipo de sistema de drenagem existente.

Os sistemas de drenagem podem ser unitários, separativos, pseudo-separativos ou mistos. As redes de drenagem unitárias são constituídas por uma única rede de coletores onde são recolhidas e transportadas as águas residuais e pluviais. Já os sistemas separativos caracterizam-se pela recolha e transporte dos efluentes residuais e das águas pluviais em redes independentes, não havendo qualquer ligação entre ambas. Uma rede mista é constituída pela conjugação dos outros sistemas acima descritos, onde parte do sistema é separativo, outra parte

é unitária. Os sistemas pseudo-separativos caracterizam-se pela admissão de águas pluviais dos pátios interiores ou jardim possa ser efetuada por coletores de águas residuais domésticas aquando a inexistência de coletores pluviais. Estes sistemas de drenagem artificiais têm como objetivo minimizar os impactes do Homem no ciclo hidrológico natural.

### 2.1.2 Efeitos da Urbanização na Drenagem Pluvial

A população residente em centros urbanos é cada vez maior, estando em 2015, 54% da população a viver nas cidades. É expectável que este valor aumente para 66% em 2050 (UNRIC, 2015). Este crescimento demográfico carece ser acompanhado pela expansão controlada dos meios citadinos. Este acompanhamento é realizado por diversas políticas do ordenamento do território, nas quais inserem-se as políticas dos usos do solo.

Como consequência direta, a crescente expansão das zonas urbanas tem contribuído para a alteração dos usos do solo. Parte do volume de água que outrora se infiltrava no solo ou que retornava à atmosfera, por via da evapotranspiração, escoava diretamente pelas superfícies impermeáveis. Estas superfícies, como o próprio nome indica, reduzem a infiltração, promovendo o aumento do escoamento superficial aquando a ocorrência de precipitações (Figura 2.1).

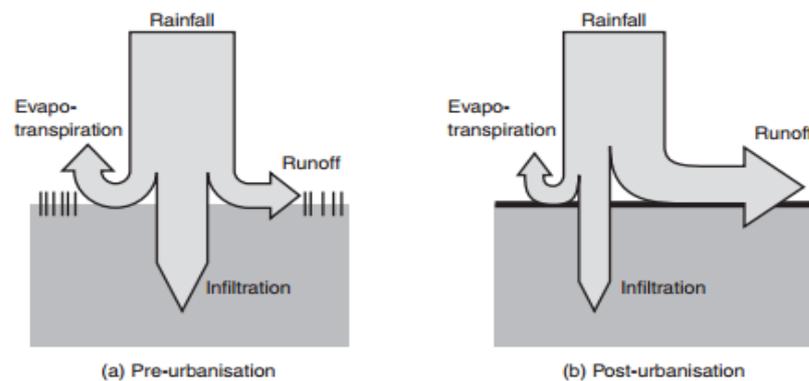


Figura 2.1 - Efeito da urbanização na resposta hidrológica (Butler e Davies, 2011)

Este facto, aliado ao subdimensionamento dos órgãos de entrada, como sarjetas e sumidouros, das redes de drenagem pluvial promovem o surgimento de inundações nos pontos baixos dos

meios urbanos<sup>1</sup>. Desta situação resulta o não aproveitamento integral da capacidade de transporte dos coletores. Os autores referem ainda que a deficiente manutenção das redes de drenagem, onde surgem obstruções e entupimentos promovem a ocorrência inundações urbanas. Estes fenómenos provocam prejuízos económicos, materiais, problemas sanitários, assim como, em casos mais graves, a perda de vidas humanas. Para prevenir tais eventos, é necessário encontrar soluções corretivas que permitam a rápida e eficiente recolha dos caudais pluviais.

Os sistemas de drenagem, mais concretamente a rede de drenagem pluvial, surgem assim, como já referido, da necessidade de minimizar os impactes da impermeabilização do solo no ciclo hidrológico natural, pelo que o correto dimensionamento do sistema é necessário para o bom funcionamento do mesmo. Para o dimensionamento e reabilitação dos sistemas de drenagem, são utilizados modelos de simulação.

## 2.2 Modelação Computacional

Na simulação de sistemas de drenagem urbana existem dois grupos de simulação: um referente à simulação do comportamento hidrológico/hidráulico do sistema e outro referente à presença de poluentes no sistema de drenagem e o seu impacte no meio recetor. Neste estudo são de interesse os modelos do primeiro tipo.

Os modelos de simulação hidrológica/hidráulica têm, em regra, na sua composição duas componentes: uma parte referente à transformação da precipitação em escoamento superficial e outra relativa ao comportamento do escoamento à superfície. A modelação hidrológica procede ao cálculo das perdas iniciais da precipitação e com os resultados obtidos é efetuada a distribuição temporal e espacial do escoamento superficial. A simulação do escoamento superficial, denominado por modelo hidráulico, com base nos resultados da primeira componente, faz a representação do escoamento nos coletores.

### 2.2.1 Modelos Hidrológicos

Em *software* de modelação hidrológica, os modelos mais utilizados são os modelos empíricos ou conceptuais. Os modelos empíricos não se apoiam em qualquer lei da física, pelo que a sua

---

<sup>1</sup> Lima, J., Marques, J, Sousa, J. (2013): Sistemas de Drenagem. In Hidrologia Urbana: Sistemas de drenagem de águas pluviais urbanas. Lima *et al.* Lisboa, Portugal.

formulação é obtida a partir de simples correlações matemáticas de dados experimentais. Os modelos conceituais são caracterizados pela utilização de relações entre uma variável independente e outras variáveis que descrevem o processo. Em regra, a simulação da precipitação-escoamento efetuada por modelos conceituais não é adequada, embora continuem a ser utilizados (Leitão *et al.*, 2008). A calibração destes modelos recorre à utilização de parâmetros sem significado físico. Dentro desta categoria insere-se o método racional, o hidrograma unitário e os modelos de reservatório linear, que são os métodos mais utilizados em modelação hidrológica (Leitão *et al.*, 2008).

### **Método Racional**

O método racional é dos modelos mais utilizados para a determinação de caudais de ponta de cheia em bacias hidrográficas de pequena dimensão, em a bacia é aproximadamente no máximo de 25 km<sup>2</sup> (Hipólito e Carmo Vaz, 2013). A utilização em pequenas bacias hidrográficas prende-se com as hipóteses consideradas neste modelo em que é assumido que a precipitação é uniforme em toda a bacia e a sua intensidade é constante durante todo o evento, igualando-se ao tempo de concentração da bacia. As limitações deste modelo são devidas ao facto de ser considerado que o coeficiente de escoamento é igual para todas as intensidades de chuvada, o que em boa verdade, este coeficiente aumenta com a intensidade de chuva. Também constitui como uma limitação a definição do coeficiente de escoamento que não representa, em bom rigor, a proporção entre o escoamento superficial e a precipitação total. Na verdade a proporção entre a ponta de cheia e a intensidade da chuvada é, em regra, superior à proporção entre o volume de escoamento e a precipitação (Lencastre e Franco, 1992). O método Racional é representado pela equação (1).

$$Q = C I A \quad (1)$$

Em que:

- Q é o caudal de ponta de cheia;
- C é o coeficiente de escoamento;
- I é a intensidade de precipitação;
- A é a área da bacia.

### **Método do Reservatório**

O Modelo do Reservatório é utilizado para a simulação do comportamento de solos aquando ocorrência de processos hidrológicos abaixo da superfície da bacia. Neste modelo não é

considerada a conservação da quantidade de movimento, uma vez que a resposta da bacia é instantânea. Assim sendo, este modelo apenas considera a conservação da massa. O Modelo do Reservatório é um modelo matemático cuja capacidade do reservatório corresponde à diferença entre a capacidade de campo<sup>2</sup> e o ponto de emurchecimento permanente<sup>3</sup>. O excesso de água que ultrapassa o volume de armazenamento, é escoada pela superfície e/ou percolada até aos aquíferos subterrâneos correspondendo ao volume de água que é descarregada do reservatório.

### **Hidrograma Unitário**

Segundo Hipólito e Carmo Vaz (2013), o método do hidrograma unitário é dos métodos mais utilizados para obtenção do escoamento resultante de uma precipitação intensa. O hidrograma unitário representa a transformação da precipitação útil em escoamento superficial, fornecendo informações sobre o caudal de ponta e do hidrograma de cheia. Quando o evento de precipitação tem uma duração definida, o hidrograma unitário caracteriza o hidrograma resultante do escoamento direto, originado por uma unidade de precipitação útil (Leitão *et al.*, 2008). Através desta metodologia, aquando ocorrência de eventos de precipitação com a mesma duração mas com intensidades diferentes, é possível obter os hidrogramas respetivos, tendo em conta o hidrograma unitário existente para a duração de chuvada em estudo.

Este modelo não deverá ser utilizado em bacias com dimensões superiores a 1000 km<sup>2</sup> (Hipólito e Carmo Vaz, 2013) ou que contenham regiões com características climáticas muito distintas, uma vez que poderá não respeitar os seguintes pressupostos do método: a distribuição da precipitação é uniforme em toda a extensão da bacia, a precipitação útil tem intensidade constante, o tempo de base do hidrograma do escoamento direto é constante e as características da bacia hidrográfica não variam.

Para uma bacia que cumpra os pressupostos acima descritos, existem dois postulados alusivos ao escoamento superficial direto:

- Para chuvadas de mesma duração, os hidrogramas com diferentes alturas, terão o mesmo tempo base e ordenadas proporcionais ao volume de precipitação útil - postulado da proporcionalidade;

---

<sup>2</sup> Quantidade de água presente no solo que este conseguiu reter contra ação gravítica.

<sup>3</sup> Teor de humidade no solo para o qual as plantas já não conseguem captar água e murcham de forma permanente, ainda que submetidas a uma atmosfera saturada.

- Aquando a sobreposição de hidrogramas de períodos de precipitação, contínuos ou isolados, o hidrograma resultante terá as suas ordenadas iguais à soma das ordenadas correspondentes dos hidrogramas relativos a cada um dos períodos, considerados isolados – postulado da sobreposição.

Resumindo, a partir do hidrograma unitário é possível determinar hidrogramas correspondentes a diferentes alturas de precipitação (postulado da proporcionalidade) ou a diferentes durações (postulado da sobreposição).

### **Diagrama Tempo-Área**

O diagrama Tempo-Área baseia-se no cálculo de como a área da bacia hidrográfica contribui para o escoamento ao longo do tempo, para uma chuvada contínua. Neste modelo as linhas de igual tempo de percurso de caudal são delineadas (isócronas). O maior tempo de percurso corresponde ao tempo de concentração da bacia. O formato do hidrograma do escoamento superficial é controlado pelo tempo de concentração, assim como pela curva tempo-área (Leitão *et al.*, 2008).

No *design* de novos sistemas, este modelo pode ser utilizado como extensão do método racional (Butler and Davies, 2011). Neste método, tal como no método racional, a área tem uma contribuição constante, mas, em boa verdade, a contribuição das áreas mais próximas da secção de referência ocorre primeiro do que a das áreas mais afastadas.

### **2.2.2 Modelos Hidráulicos e Drenagem Dual**

A modelação tem como finalidade a análise do comportamento de um sistema quando este é sujeito a diferentes condições. No caso da drenagem urbana, os modelos de simulação são utilizados para representar e simular o comportamento de sistemas de drenagem de águas pluviais em diversas situações correntes, assim como em situações extremas.

Os modelos de simulação hidráulica podem ser utilizados como instrumento de apoio aos decisores, que visam a facilitar análise e tomada de decisão. Estes modelos podem ser utilizados para a conceção de novos sistemas, preparação e planeamento de programas de reabilitação.

No dimensionamento de sistemas de drenagem, a simulação hidráulica é utilizada para encontrar uma solução que permita uma capacidade de resposta adequada da rede, quando

sujeitas a condições específicas e que, simultaneamente cumpra os requisitos necessários. Já a sua aplicação para sistemas existentes, o objetivo passa por perceber se o sistema necessita de melhoramentos, e em caso afirmativo, representa uma mais-valia na preparação e conceção de programas de reabilitação.

Para estudos relacionados com as inundações, estes modelos são importantes para perceber o como é que o sistema vai reagir perante a ocorrência de eventos de precipitações extremos. A análise irá permitir a tomada de decisões que tenham como visão evitar ou minimizar os impactes económicos, ambientais e sociais associados às inundações. No estudo destes eventos, importa perceber qual o caudal de escoamento no sistema, a profundidade e a extensão da sobrecarga e a superfície de inundação (Butler and Davies, 2011).

Um dos modelos hidrodinâmicos utilizado na modelação de sistemas de drenagem urbana são as equações de *Saint-Venant*. Estas equações são as mais utilizadas para descrever o comportamento de um escoamento variável em superfície livre<sup>4</sup>.

Na ocorrência de fenómenos de precipitação extremos, água no interior das condutas poderá atingir o topo do coletor, passando o escoamento a ocorrer em pressão, entrando os sistemas em carga. Quando isto ocorre, a capacidade limite do coletor foi atingida ou mesmo ultrapassada. O escoamento passa, então, a poder ocorrer em superfície livre em alguns segmentos do sistema e em pressão noutros. Em casos cuja capacidade hidráulica do sistema é mesmo ultrapassada, a água poderá sair do coletor e escoar pela superfície.

O conceito de drenagem dual prende-se com o facto dos sistemas de drenagem pluvial serem compostos por duas componentes distintas: o sistema “minor” e o sistema “major”. O sistema “minor” é composto por uma rede de coletores e o sistema “major” é composto por canais naturais e artificiais, por ruas, depressões de água e zonas de acumulação, etc. A modelação hidráulica baseada na drenagem dual faz a interação dos sistemas “minor” e “major”.

Existem duas abordagens distintas na drenagem dual: abordagem 1D/1D e 1D/2D. As equações de *Saint-Venant* são aplicadas a ambas abordagens, para obtenção do caudal, altura e velocidade de escoamento. No modelo unidimensional (Fig. 2.2), a superfície urbana é considerada como uma rede de canais abertos e de zonas de acumulação de água, formando um conjunto de canais e nós ligados ao sistema de coletores. Os sistemas “minor” e “major”

---

<sup>4</sup> Lima, J., Marques, J, Sousa, J. (2013): Sistemas de Drenagem. In: Hidrologia Urbana: Sistemas de drenagem de águas pluviais urbanas. Lima *et al.* Lisboa, Portugal.

estão interligados por aberturas que representam os elementos de entrada e saída de água dos coletores existentes na rede (Maksimovic *et al.*, 2009).

Nos modelos 1D/2D (Fig. 2.3), a rede de drenagem é repartida em pequenas células, cujas mesmas, são representadas no sistema de coordenadas (X,Y). As células contêm informações importantes sobre as suas características, tais como a área, coeficiente de rugosidade, altitude, entre outras. Este modelo representa o terreno de uma forma mais realista, uma vez que as superfícies de escoamento não são limitadas como nos modelos unidimensionais<sup>5</sup>, no entanto, quando comparados com os modelos bidimensionais estes modelos exigem maior tempo de cálculo assim como memória computacional (Leitão *et al.*, 2008).



Figura 2.2- Exemplo de uma rede de drenagem 1D/1D (Butler and Davies, 2011)

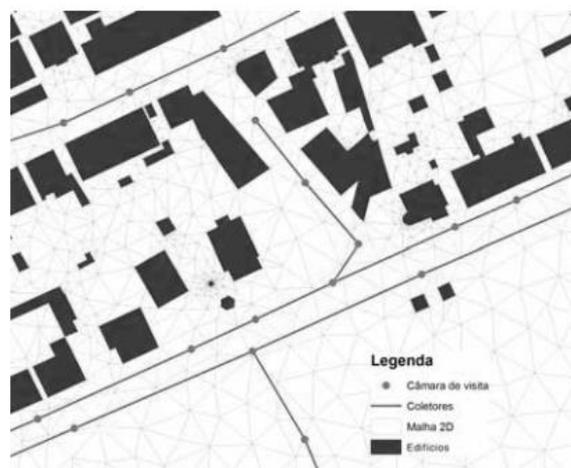


Figura 2.3 - Rede de drenagem dual 1D/2D (Pedroso de Lima *et al.*, 2013)

### 2.2.3 Incertezas associadas aos modelos de drenagem

Deletic *et al.* (2011) refere que as incertezas na modelação estão associadas a três categorias (Fig. 2.4): estrutura do modelo, dados de entrada e calibração.

<sup>5</sup> Lima, J., Marques, J, Sousa, J. (2013): Sistemas de Drenagem. In: Hidrologia Urbana: Sistemas de drenagem de águas pluviais urbanas. Lima *et al.* Lisboa, Portugal.

### Quadro 2.1 – Incertezas associadas à modelação

#### Estrutura do modelo

- Concetualização;
- Equações;
- Métodos numéricos e condições fronteira;
- Parâmetros do modelo.

#### Dados de entrada

- Dados de entrada;
- Parâmetros do modelo.

#### Calibração

- Parâmentos do modelo;
- Dados medidos;
- Algoritmo de calibração;
- Função objetivo.

Os modelos de drenagem, na sua maioria são determinísticos, no entanto a estes também estão associadas algumas incertezas, exemplo disso é o grau de simplificação dos modelos.<sup>6</sup> Um exemplo prático é o facto de, na drenagem dual, os dados de saída do modelo dependerem diretamente do rigor dos dados de entrada sobre o terreno. Nesta temática a deteção remota tem um papel fundamental na aquisição de fotografias de elevada resolução para extração dos diversos dados sobre a superfície.

## 2.3 Deteção Remota

Segundo Price e Vojinovic´ (2011) a deteção remota é definida como a ciência de adquirir, processar e interpretar dados ou informação sobre objetos sem haver contacto físico com os mesmos. As características de um objeto em estudo podem ser definidas por análise da interação entre a energia eletromagnética e o objeto. A referida interação entre a energia eletromagnética e um objeto, pode ser captada por um sensor remoto (*câmeras e scanners*) cuja função é detetar a radiação que é emitida ou refletida pelo objeto.

Um pressuposto utilizado na deteção remota é que a refletância espectral difere de objeto para objeto, sendo conseqüentemente diferente para os diversos tipo de cobertura do solo (Price e

<sup>6</sup> Lima, J., Marques, J, Sousa, J. (2013): Sistemas de Drenagem. In: Hidrologia Urbana: Sistemas de drenagem de águas pluviais urbanas. Lima *et al.* Lisboa, Portugal.

Vojinović, 2011). Consoante o objetivo de estudo das imagens, é necessária uma adequada seleção, uma vez que poderá ocorrer absorção de energia, quando utilizados tipos de bandas espectrais inadequados. A absorção caracteriza-se pela transformação de energia eletromagnética noutra tipo de energia por interação com a matéria.

O espectro eletromagnético engloba todas as radiações eletromagnéticas possíveis: Raios gama, raios X, radiação ultravioleta, radiação visível, radiação infravermelha, micro-ondas e ondas rádio. No entanto, nem todas são utilizadas na deteção remota sendo, pelo que as radiações utilizadas pertencem às regiões: da radiação ultravioleta, do visível, dos infravermelhos próximos, dos infravermelhos térmicos e micro-ondas (Price e Vojinović, 2011). Por exemplo, radiação dos infravermelhos próximos é frequentemente utilizada, em fotografia, para o estudo de zonas inundadas. A vantagem da utilização de radiação pertencente a esta região prende-se com o facto da refletância da água ser baixa, o que possibilita diferenciar áreas correspondentes a superfícies de água das áreas secas (Henriques, 1985). Fonseca e Fernandes (2004) referem igualmente que a água não reflete quase nenhuma radiação correspondente aos comprimentos de onda infravermelha.

Já a utilização de bandas de microondas apresenta vantagem na aquisição de imagens a qualquer independentemente das condições atmosféricas, não sendo as mesmas interferidas por nuvens, nevoeiro, poeiras e precipitação. (Fonseca e Fernandes, 2004).

A radiação do visível é dada pela combinação RGB (*red-vermelho, blue-azul, green-verde*) que permite representar os objetos na sua verdadeira cor. Na opinião da autora da presente tese, a utilização da radiação do visível é vantajosa para a classificação dos usos do solo como permeáveis, impermeáveis, estradas, telhados, vegetação, entre outros. Tal é explicado pelo facto de neste tipo de análise o interesse não é detetar teores de humidade, contaminações ou diversos tipos de vegetação e rochas mas sim diferenciar os tipos de uso do solo pelo que a verdadeira cor dos objetos é suficiente para a classificação.

### **2.3.1. Técnicas de Deteção Remota**

Uma vez que a recolha de dados do terreno é, por norma, uma atividade dispendiosa e requer alguns recursos humanos é necessário avaliar cuidadosamente as vantagens e desvantagens de cada uma das técnicas de deteção remota. São apresentadas três tipos de tecnologias de deteção remota: LIDAR, Termografia e Fotografias aéreas.

### 2.3.1.1 Tecnologia LIDAR

*Light Detection and Ranging* (LIDAR) ou *Airborne Laser Scanning* (ALS) é uma técnica de detecção remota econômica e precisa (Price e Vojinović, 2011). Esta tecnologia funciona através da emissão de um pulso de luz, cujo tempo de viagem combinado com a distância e os ângulos dos pulsos, permite obter informações tridimensionais precisas sobre a superfície do terreno. A tecnologia LIDAR é na maioria das vezes encontrada em aviões e helicópteros. A sua utilização torna-se impossível em *drones* pequenos devido ao seu peso e à elevada necessidade de energia (Leitão *et al.*, 2015).

Os modelos digitais de elevação (MDE)<sup>7</sup> obtidos por LIDAR podem ter uma precisão de afastamento de pontos até 0.3 metros e os pontos captados encontrar-se-ão num intervalo de 0.5 a 5 metros altura, com uma precisão vertical de 0.15 metros (Price e Vojinović, 2011). Estes autores referem ainda que o pulso de luz consegue alcançar uma distância superior a um quilómetro.

### 2.3.1.2 Fotografias aéreas

As fotografias aéreas têm como objetivo recolher dados sobre a superfície e são utilizadas para cartografia e para construção de ortofotomapas, que são mapas fotográficos, de onde se podem medir diretamente distâncias, devido estarem à escala. Contudo, é necessário salientar que as fotografias aéreas sem qualquer tipo de processamento não são mapas.

Estas são tiradas através de câmaras instaladas em plataformas aéreas, tais como aviões, helicópteros, *drones*, entre outros. O tipo de fotografia aérea mais comum é a fotografia vertical (Lillesand e Kiefer, 1987), que consiste na utilização dos eixos da câmara o mais verticalmente possível com a superfície. No entanto, é rara a fotografia que seja verdadeiramente captada na vertical, devido à rotação da câmara. Sucessivas fotografias são tiradas de modo a registar a superfície do terreno continuamente. Para conseguir uma correspondência aceitável entre imagens, deverá existir uma sobreposição frontal de 60% ou mais (Leitão *et al.*, 2015). Este autor afirma ainda que para terrenos complexos, o limite inferior aceitável da sobreposição frontal deve ser aumentado.

---

<sup>7</sup> Os modelos digitais de elevação são representações digitais da superfície às quais estão associadas variações da altitude numa área com coordenadas planimétricas (X,Y).

### 2.3.1.3 Termografia

A termografia é a técnica de detetar a temperatura dos corpos através da radiação infravermelha que um corpo emite. Este método apresenta como vantagens ser eficaz, rápido e preciso na monitorização espacial e temporal de gradientes de temperatura (Lima, R. *et al.*, 2015). As suas aplicações são vastas desde inspeções de infiltrações de água em edifícios, monitorização de parâmetros ambientais (Lillesand and Kiefer, 1987), reconhecimento do tipo de rochas (Sabins, 1997), fins militares (Lima, R. *et al.* 2015), entre outros. A aplicação desta técnica tem vindo aumentar na área de hidrologia devido ao decréscimo recente dos seus custos assim como devido há sua maior resolução e facilidade de transporte (Lima, R, 2015). Para utilização da termografia em análises do terreno, Henriques (1985) afirma que a temperatura da superfície do terreno depende essencialmente da inércia térmica, da rugosidade da superfície, textura do solo e da sua composição mineralógica e rugosidade da superfície. Estando a inércia térmica relacionada com o teor de humidade do solo, é possível avaliar o mesmo através das imagens, após devidamente processadas, que são captadas por detetores térmicos, encontrados a bordos de aviões e satélites.

### 2.3.2 Plataformas de deteção remota

Existem diversas plataformas que permitem o transporte de tecnologias de deteção remota tais como aviões, helicópteros, balões meteorológicos, entre outros. São dados dois exemplos desses equipamentos: Satélite e *Drone*.

#### 2.3.2.1 Satélite

Os satélites têm sido utilizados para avaliação de catástrofes naturais, devido à sua capacidade de quantificar fenómenos físicos associados com movimentos da superfície da terra (terramotos, inundações, tsunamis e deslizamentos de terras) e incêndios florestais. Os satélites podem ser compostos por dois tipos de sensores: ativos ou passivos. Sensores passivos de elevada resolução têm um desempenho positivo na avaliação de danos causados pelas inundações enquanto que os sensores passivos e os ativos têm sido usados para mapas de previsão de áreas vulneráveis a inundações.

Embora os satélites de elevada resolução espacial possam adquirir imagens detalhadas, estes são vulneráveis à cobertura de nuvens. Além disso, é também uma tecnologia dispendiosa e os períodos de atualização de dados por parte deste dispositivo, que deveriam ser uma vez por ano são na verdade, uma vez em cada três anos (Tokarczyk *et al.*, 2015).

### 2.3.2.2 *Unmanned Aerial Vehicles*

A palavra *drone* tem uma conotação negativa, uma vez que eram utilizados para fins bélicos. Como tal, surgiu a designação de *Unmanned Aerial Vehicle* (UAV) para desassociar a imagem do equipamento ao armamento. Apesar da negatividade associada à palavra *drone*, esta será utilizada para simplificação de leitura e para a uniformidade de texto.

As aplicações dos *drones* são diversas, desde a sua utilização para atividades de lazer, militar, empresarial como, por exemplo, é o caso da empresa Amazon que está a testar a entrega de encomendas por este tipo de veículos, e científico. Nesta última categoria inserem-se as utilizações de *drones* para levantamentos topográficos essenciais à modelação na drenagem urbana. Leitão *et al.* (2015), concluiu que a utilização destes equipamentos e do *software* inerentes aos mesmos apresentam resultados satisfatórios aquando aplicados à drenagem urbana.

Os *drones* podem efetuar voos de forma autónoma ou serem controlados remotamente ou então serem uma combinação dos dois. Estes são uma alternativa vantajosa, às outras tecnologias, devido terem um custo competitivo e serem mais flexíveis para captação de imagens (Tokarczyk *et al.*, 2015), principalmente para zonas de estudo de pequenas e médias dimensões (Leitão *et al.*, 2015). No entanto, para os *drones* de pequena dimensão, a autonomia, assim como a velocidade são baixas, o que impossibilita fotografar grandes áreas (Tokarczyk *et al.*, 2015).

No que refere à resolução de imagens, a elevada qualidade das imagens captadas pelos *drones*, cuja resolução de MDEs resultante pode ser inferior a 1 metro é uma vantagem para modelação de fenómenos de precipitação, no entanto, Tokarczyk *et al.* (2015), realça que deverá haver uma maior quantidade de pesquisa para determinar o verdadeiro benefício da tecnologia na modelação urbana.

As principais conclusões que Leitão *et al.* (2015) retirou da sua investigação sobre os *drones* são as seguintes:

- Voos que são efetuados a altitudes menores permitem obtenção de MDEs de melhor resolução;
- A realização do voo por parte dos *drones* em dias nublados permite obter melhores resultados na qualidade do MDE. Tal observação é provavelmente explicada devido à iluminação mais uniforme e à ausência de sombras.

Tokarczyk *et al.* (2015) e Leitão *et al.* (2015) utilizaram nas suas investigações um determinado tipo de *drone*, chamado eBee. Este *drone* de asas fixas foi construído pela Sensefly para uso profissional. Ambos recorreram a uma câmara compacta Canon IXUS 127 HS com 16 megapixéis para captação de imagens aéreas, referindo Leitão *et al.* (2015) que na utilização de uma melhor câmara era expectável observar maiores diferenças nas imagens, permitindo uma melhoria na identificação de pontos característicos.

## 2.4 Identificação dos Usos do Solo

É designado por usos do solo todas as modificações que são efetuadas no meio ambiente pelos humanos (Elkhrachy, 2015). Devido à crescente migração da população para as cidades, como referido no ponto 2.1.2, estes meios tendem a desenvolver-se de forma acompanhar esse acréscimo, através da sua expansão. Por consequência direta, a vegetação, os solos porosos e outras características ambientais que permitem a redução do escoamento são reduzidas ou mesmo eliminadas (Kaiser *et al.*, 1995).

Numa bacia fortemente urbanizada, para um mesmo volume de água o tempo de concentração da bacia hidrográfica diminui. Este parâmetro caracteriza-se pelo tempo ao qual toda a bacia está a contribuir para o escoamento, definido também como o tempo que uma gota proveniente da precipitação demora atingir uma secção de referência, partindo do ponto mais afastado desta. Num meio natural, os volumes de água são drenados lentamente através do solo, podendo atingir os aquíferos ou voltar a ressurgir à superfície sofrendo um atraso na chegada à secção de referência. Tal não acontece em áreas impermeáveis, em que o escoamento nestes tipos de superfície é mais rápido. Como consequência existe um aumento do caudal de ponta de cheia e uma redução temporal no qual este é atingido.

A modelação hidráulica para estudo de situações de inundação, além de necessitar de informação detalhada sobre a distribuição espacial da precipitação, como já referido anteriormente, carece também de informação pormenorizada sobre os usos do solo. Leitão (2009) refere que uma representação do solo realista é fundamental para modelação de processos superficiais, como é o caso do escoamento superficial. Os dados sobre o uso do solo podem ser recolhidos através de inspeções de campo, da interpretação de ficheiros que contenham um cadastro da área em estudo, ou através da classificação de dados adquiridos por deteção remota (Kaiser *et al.*, 1995).

## 2.5 Reconhecimento de padrões

*Machine Learning* é um ramo da ciência que tem como objetivo ensinar os computadores a pensar como os humanos. Com base nos dados de entrada, o computador aprende as informações recebidas e devolve resultados com base na sua aprendizagem. Esta área é muito útil para aplicação em projetos cuja elevada quantidade de dados impossibilite a sua análise manualmente.

Dada uma imagem para a classificação dos solos não é prático uma análise manual *pixel-a-pixel* pelo que será necessário recorrer a soluções mais automatizadas. O reconhecimento de padrões pode ser efetuado por técnicas de classificação, em que estas atribuem classes a um conjunto de dados em função de um conjunto de elementos padrão (Fonseca e Fernandes, 2004). A realização da classificação tem em conta o menor risco de erro de os dados pertencerem a uma determinada classe. Para a análise de imagens cujo objetivo é recolher informações sobre os usos do solo, a classificação dos píxeis é feita através do reconhecimento das suas bandas espectrais.

Existem duas categorias de métodos de classificação de imagens: os supervisionados (assistidos) e os não-supervisionados (não-assistidos). No primeiro tipo de classificação, o utilizador irá definir previamente as classes nas quais os píxeis se irão integrar, deverá existir também um conhecimento prévio da imagem. O classificador é treinado para definir corretamente as classes consoante com as classes previamente definidas. Tokarczyk *et al.* (2015) refere que a classificação supervisionada consiste em três passos:

- Extração de características de entrada da imagem;
- Treinar o classificador;
- Classificação de todos os píxeis utilizando o classificador previamente treinado.

Nos modelos de classificação não-supervisionada, a classificação é efetuada através da semelhança natural que exista entre um conjunto de píxeis. A similaridade pode ser determinada através de assinaturas espectrais, distâncias entre os centros de aglomerados e de outros critérios de aglutinação (Fonseca e Fernandes, 2004). Por não haver identificação *a priori* da semelhança entre elementos, estes métodos permitem compreender quais as causas intrínsecas em relação ao processo de um elemento pertencer a um determinado conjunto. Os algoritmos procuram assim minimizar a distância entre os elementos de um mesmo conglomerado e maximizar a distância entre elementos de conglomerados diferentes.

### 3 METODOLOGIA

Neste capítulo são apresentadas as técnicas de análise de imagem utilizadas para identificação de áreas impermeáveis. Os *software* utilizados foram o ArcGis e o Google Earth Engine (GEE). Para tal é apresentado um caso de estudo referente a uma das sub-bacias da cidade de Lucerna.

#### 3.1 Caso de Estudo

Lucerna é uma cidade suíça que em 2009 era povoada por 59 509 habitantes, sendo a mais populosa da Suíça Central. Sem contar com a área dos seus lagos, a cidade tem uma área de 29.1 km<sup>2</sup>, em que mais de metade da mesma era, em 2010, solos permeáveis, dos quais 28% eram terras agrícolas, 22.3% de floresta e 2.1% solo improdutivo, ocupando a zona urbana os restantes 47.7% do território (InfoSwiss). O caso de estudo escolhido para demonstrar as vantagens e inconvenientes da utilização de imagens de diferentes resoluções, na modelação de inundações em meio urbano, foi uma sub-bacia da cidade de Lucerna, Wartegg (Figura 3.1).



Figura 3.1 – Wartegg, Lucerna (Air Luzern)

Foi ainda utilizada uma imagem *Landsat* da bacia central de Coimbra (Apêndice A) para averiguar quais os algoritmos que seriam utilizados na presente tese. A escolha desta cidade para uma pré-avaliação dos algoritmos deve-se ao facto da autora conhecer a cidade, sendo mais fácil compreender se a classificação dos usos do solo está a ser feita de forma correta.

### 3.2 Imagens

As duas fotografias utilizadas referentes à sub-bacia de Lucerna são provenientes de fontes distintas. A Figura 3.2 foi captada por um satélite do programa da NASA *Landsat 7* e é datada de Março de 2016, cuja resolução é de 30 m, ou seja, cada píxel da imagem representa de lado 30 metros da superfície. Apesar da sua resolução grosseira, esta foi escolhida para análise uma vez que se encontra disponível de forma gratuita no GEE. Através da fotografia é possível perceber o contorno da bacia de Wartegg mas não são perceptíveis os pormenores da zona devido à sua grosseira resolução.

Já o ortofotomapa (Figura 3.3) permite reconhecer diversas características da zona, devido à sua elevada resolução de 5 cm. Este tipo de imagem tem a particularidade de poder ser utilizada como mapa uma vez que as deformações da imagem são retificadas com auxílio de MDEs. A Figura 3.3 foi registada por uma câmara de 16.1 megapixéis a bordo de um *drone* eBee (Leitão *et al.*, 2015). Quando ampliada, são observados diversos pormenores tais como pequenos arbustos, pessoas e pequenos elementos nos telhados, jardins, entre outros.

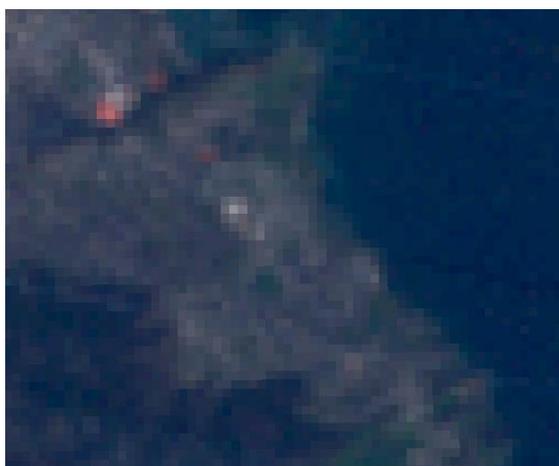


Figura 3.2 - Fotografia *Landsat*, Lucerna (GEE)

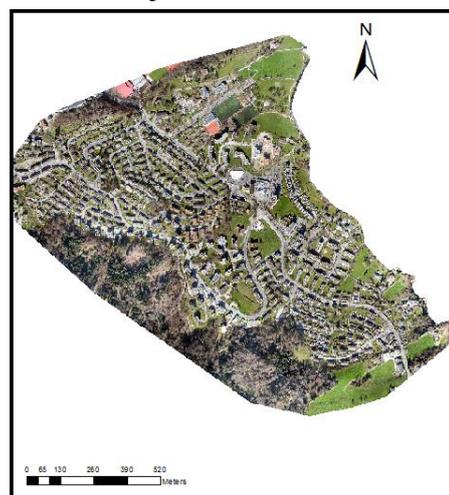


Figura 3.3 - Ortofotomapa de Lucerna (Leitão *et al.*, 2016)

Através destas imagens será efetuada a classificação dos usos do solo com base em duas classes distintas, permeável que engloba todos os espaços verdes e vegetação e impermeável que é caracterizada por estradas, telhados e outras superfícies impermeáveis.

### 3.3 Técnicas de Análise de Imagens

#### 3.3.1 Software

##### 3.3.1.1 Google Earth Engine

O Google Earth Engine (GEE) é uma plataforma que foi lançada pela Google em 2010 e que permite aos utilizadores explorar e analisar dados ambientais através de imagens. Contém uma vasta biblioteca de imagens, de todo o Mundo, captadas, maioritariamente, por satélites. As suas aplicações são variadíssimas, desde a monitorização da desflorestação global, deteção de águas superficiais, estimativa do risco de surtos de doenças infecciosas, monitorização do crescimento urbano, acidentes petrolíferos, entre outras.

O leque de imagens, datadas desde há cerca de 40 anos, pode ser utilizado de forma gratuita, o que permite ao público acesso a imagens que outrora seriam pagas. Nesta aplicação, além da utilização livre de imagens existentes na *Cloud*, os utilizadores poderão fazer *upload* dos seus próprios dados.

O GEE pode ser utilizado para investigação, ensino e outras atividades apenas exploratórias sem quaisquer custos para os utilizadores. No entanto, para aceder a este aplicativo é necessário uma pré-inscrição pelo que poderá nem sempre ser aceite pela Google.

Esta plataforma pode ser utilizada de duas formas: através do API ou do Explorer. A API permite aos utilizadores fazer inúmeras operações tais como processamento de imagens, classificação, deteção de mudanças, entre outras, através da compilação de códigos em JAVA e Python. Esta funcionalidade permite desenvolver algoritmos permitindo realizar vários tipos de análises geoespaciais. No entanto, para utilização desta ferramenta é recomendável que os utilizadores tenham ou adquiram conhecimentos básicos em programação.

O Explorer permite, com alguma facilidade, explorar a biblioteca de imagens e efetuar análises sobre as mesmas. Neste estudo, o Explorer foi utilizado para a identificação de zonas impermeáveis com base no ortofotomapa gerado através das fotografias obtidas com o UAV, quer com base na imagem do *Landsat*.

### 3.3.1.2 ArcGis

O ArcGis é um *software* de informação geográfica da *Environmental Systems Research Institute* (ESRI) que permite trabalhar com diversas informações geográficas como a criação e edição de mapas ou efetuar inúmeras tarefas de análise e processamento. O ArcMap é uma componente deste *software*, que permite criar e editar mapas assim como proceder a análises geoespaciais. Existem dois outros módulos: o ArcCatalog e o ArcToolbox. O ArcCatalog permite organizar e gerir vários tipos de informação geográfica, enquanto o ArcToolBox é um aplicativo que apresenta um conjunto de ferramentas de análise e processamento úteis aos diversos projetos, que são possíveis desenvolver no *software*.

Tal como o Google Earth Engine, este programa pode ser utilizado para diversos fins como gestão do território, defesa e segurança, transporte e logística, entre outros. No presente estudo, o ArcGis será utilizado para efetuar a identificação de áreas impermeáveis através de um ortofotomapa de Lucerna e das ferramentas existentes no *software*.

### 3.3.2 Procedimento

Como não é possível fazer análise manual de *pixel-a-pixel* das imagens para identificação dos usos do solo, os métodos classificativos de *Machine Learning* apresentam-se como uma boa solução. Como referido no capítulo 2, estes métodos estão divididos em duas categorias: supervisionados e não supervisionados. Neste estudo foram apenas utilizados classificadores supervisionados, cujos resultados provenientes dos mesmos baseiam-se em treinos fornecidos pelo utilizador.

Em ambos os *software*, o treino é feito de forma idêntica, no entanto, a forma de colocar os algoritmos em funcionamento e o modo de obter os resultados numéricos é diferente. Após abertura da imagem no *software* pretendido, foram criadas duas classes, uma com nome de Permeável e outra intitulada como Impermeável, tendo sido atribuída uma cor de identificação a cada uma das classes: verde e preto, respetivamente. Ao seleccionar cada uma das classes foram desenhados manualmente diversos polígonos sobre a imagem para que ficassem registados como pertencentes a essa categoria. Os polígonos, aqui, têm como intuito de recolher informações sobre a composição da cor dos píxeis dentro dos limites dos mesmos. Estes dados foram posteriormente utilizados na aprendizagem do algoritmo para a identificação dos usos do solo.

Foram utilizados diferentes números de elementos (polígonos) de treino para perceber qual a influência dos mesmos no cálculo da precisão global do algoritmo de classificação (conceito explicado no ponto 3.3.4). Para garantir que os resultados não fossem influenciados pela escolha aleatória de polígonos entre os diversos treinos, estipulou-se que seria feito primeiro o treino com maior número de polígonos, fazendo com que os seguintes possuam os polígonos do treino anterior. Ou seja, num treino de 400 polígonos, constariam 400 dos 500 polígonos usados no treino de 500, no treino de 300 constariam 300 polígonos do treino de 400 e assim sucessivamente.

No ArcGis é criado um ficheiro com os polígonos de treino e seguidamente o mesmo é colocado como *input* no algoritmo. O algoritmo cria uma imagem classificada a preto e verde, cores previamente escolhidas, a indicar quais as zonas impermeáveis e permeáveis, respetivamente. A esta imagem está associada uma tabela de atributos onde é possível consultar a percentagem de cada área pertencente a cada classe. Após a classificação, foram gerados 100 pontos, designados de pontos de controlo, a amarelo (Figura 3.4). Através da ferramenta “Extract Value to Points”, o classificador atribuiu uma classe a cada um destes pontos e através do conhecimento do utilizador da verdade do terreno, é possível construir uma matriz de erro e calcular a precisão global, conceitos posteriormente descritos.



Figura 3.4 – Pontos de Controlo 1

No GEE o treino é feito de forma idêntica. Após este estar concluído é apenas necessário escolher o algoritmo para fazer a classificação. Ao contrário do ArcGis, não é necessário

construir a matriz de erro e calcular a precisão global, uma vez que o *software* o faz automaticamente.

### 3.3.3 Algoritmos

Os métodos de aprendizagem utilizados nesta tese são mecanismos de inteligência artificial que, como já referido, por meio de uma aprendizagem de padrões é possível classificar imagens com base no que aprendeu. Existem inúmeros algoritmos que podem ser aplicados neste tipo de aprendizagens, no entanto não é fácil escolher quais os mais adequados para uma análise, como sugere o teorema *No free lunch* (James *et al.*, 2013). Este teorema refere que nenhum método é melhor que todos os outros sobre todos os conjuntos de dados possíveis. Ou seja, antes de se fazer qualquer experiência, não se pode afirmar qual o algoritmo que melhor fará a classificação. Por não ser fácil a escolha de um algoritmo, foi efetuada uma análise preliminar de uma imagem *Landsat* de Coimbra para determinar os algoritmos que devolviam melhores resultados. Como já referido anteriormente, a escolha da cidade prende-se com o facto de a autora conhecer a cidade, que é uma vantagem para compreender se a classificação dos usos do solo está a ser feita de forma correta. Considerou-se para esta fase o funcionamento dos algoritmos como modelo de caixa preta, em que apenas interessam os dados de entrada e saída e não como funcionam.

#### 3.3.3.1 Pré-avaliação

Por a cidade de Coimbra ser familiar, a análise preliminar foi feita no GEE como explicado no ponto 3.3.2 para uma imagem *Landsat* (A.2) da bacia central da cidade (A.1), com resolução de 30 m. Foram utilizados três conjuntos de polígonos: 300, 400 e 500. Posteriormente, para cada treino foram testados os seguintes algoritmos: *Fast Naive Bayes*, *GMO Entropy*, *Winnow*, *Perceptron*, *Pegasos*, *CART*, *Random Forests*, *IKPamir*, *Voting SVM* e *Margin SVM*.

Com base nesta análise escolheram-se os dois algoritmos com melhores resultados: *Fast Naive Bayes* e *Margin SVM* (Anexos C.1, C.2, C.3, C.4). Apesar do *Margin SVM* e o *Voting SVM* terem apresentado exatamente os mesmos resultados foi escolhido o primeiro devido a facilidade de encontrar informações sobre o mesmo. O terceiro algoritmo escolhido é um dos métodos mais populares na área de hidráulica (Tokarczyk, 2015), o Método da Máxima Verossimilhança.

### 3.3.3.2 Fast Naive Bayes

O *Fast Naive Bayes* é um classificador supervisionado de rápido desempenho baseado no teorema de Bayes. O objetivo deste algoritmo é calcular a probabilidade de um determinado objeto ter uma certa classificação, tendo em conta as suas características. De uma forma muito simples, se tivermos uma laranja, o classificador terá informação que é redonda, de cor alaranjada e tem um determinado diâmetro. Ao analisar uma outra peça de fruta, o FNB utilizará estas características de forma independente para o cálculo da probabilidade de essa peça ser, de facto, uma laranja. O suporte matemático deste teorema é descrito pelo teorema de Bayes dado pela equação (2) (Sá, M., 2011).

$$P(X_n|Y) = \frac{P(Y|X_n) \cdot P(X_n)}{P(Y)} \quad (2)$$

Sendo as variáveis:

- $P(X_n|X)$  a probabilidade  $X_n$  sabendo que ocorre  $X$ ;
- $P(X|X_n)$  a probabilidade de  $X$  sabendo que ocorre  $X_n$ ;
- $P(X_n)$  a probabilidade de  $X_n$ ;
- $P(Y)$  a probabilidade de  $Y$ .

$P(Y)$  é dado pela expressão matemática (3):

$$P(Y) = \sum_{i=1}^c P(Y|X_n) \cdot P(X_n) \quad (3)$$

O algoritmo é denominado de *Naive* devido ao mesmo considerar, como referido, que as variáveis são independentes, o que na prática representa uma simplificação de cálculos, equação (4), embora raramente se verifique a independência das variáveis. Além disso, esta simplificação é uma desvantagem em problemas complexos. No entanto, a utilização do FNB é vantajosa quando o tamanho do treino é pequeno, não melhorando significativamente os resultados quando este aumenta. Tal significa que a sua utilização é adequada à presente tese devido a grandes treinos serem considerados para milhares de polígonos.

$$P(X_1 \dots X_n | Y) = \prod_{i=1}^n P(X_i | Y) \quad (4)$$

Tendo em conta a probabilidade (4), o FNB procura encontrar a classe (Y) que maximiza a probabilidade de um conjunto de cores pertencer a dada classe, como indica a equação (5).

$$Y \leftarrow \arg \max_{y_k} P(Y = y_k) \cdot \prod_i (X_i | Y = y_k) \quad (5)$$

Este método coincide com o método da máxima verossimilhança quando os elementos da classe Y têm igual probabilidade de acontecerem. Por exemplo, para um problema de duas classes, em que cada uma tem 50% de probabilidade de acontecer apenas são utilizadas duas classes, em que a sua probabilidade P (Y=2) é de 50% para cada.

### 3.3.3.3 Margin Support Vector Machine

As *Support Vector Machines* (SVM) são uma técnica de aprendizagem que utiliza hiperplanos<sup>8</sup> como superfícies de separação entre classes. Para dois grupos lineamente separáveis, os objetos são dispostos acima ou abaixo do hiperplano, através das equações 6 e 7, respetivamente (Sá, M., 2011):

$$W \cdot X_i + W_o \geq 0 \text{ se } t_i = +1 \quad (6)$$

$$W \cdot X_i + W_o < 0 \text{ se } t_i = -1 \quad (7)$$

Em que:

- $X_i$  é a característica do objeto de entrada i;
- $W$  é o vetor normal ao hiperplano;
- $W_o$  representa uma constante;
- $t_i$  é a classe prevista para objeto i.

<sup>8</sup> Num espaço de n dimensões, o hiperplano representa um plano de dimensão n-1.

No entanto, mais do que um hiperplano é possível de ser delineado entre os objetos, como observamos na figura 3.5 (a), pelo que o algoritmo utiliza o que representa o ótimo. Este traduz-se naquele que tem uma margem maior aos pontos, dada pelos vetores de suporte (Figura 3.5 b). Esta distância ao hiperplano manifesta-se numa medida da confiança da previsão do classificador. O nome de *Margin* é baseado neste conceito de introdução de uma margem. Quando os objetos não são linearmente separáveis, o SVM penaliza os que se encontram no interior das margens definidas.

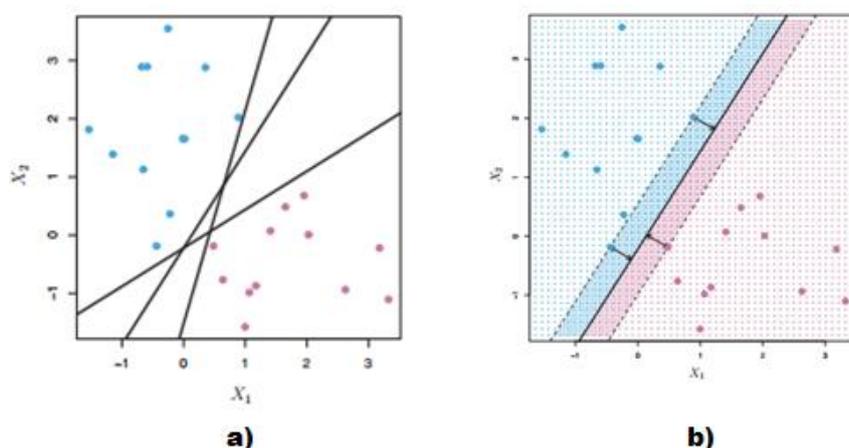


Figura 3.5 - Exemplo do funcionamento do SVM (a) e Vetores de Suporte (b) (James, G. *et al.*, 2015)

Neste trabalho, um modelo de SVM foi gerado, em que através do hiperplano e das respetivas margens, são separados os objetos pertencentes a cada uma das classes. A qualidade desta separação é dada pela precisão global. Seria interessante analisar graficamente a aprendizagem realizada pelo algoritmo, de forma idêntica à figura 3.5 no entanto, não foi possível devido à plataforma onde foi utilizado não dispor dessa opção.

A vantagem da utilização deste algoritmo é que consegue lidar bem com grandes conjuntos de exemplos. No entanto, para dados não linearmente separáveis, este método não se adequa bem.

### 3.3.3.4 Método da Máxima Verossimilhança

O Método da Máxima Verossimilhança é um algoritmo simples que assume que a probabilidade de cada característica da imagem pertencer a uma determinada classe segue uma distribuição normal. Em termos classificativos, o método utiliza a média e covariância das amostras para determinar a probabilidade de um píxel pertencer a uma determinada classe. Esta probabilidade é dada pela equação (8) (Hipólito e Vaz, 2013). O objetivo do algoritmo é maximizar esta função, a partir da qual foi calculada a probabilidade de um elemento da imagem pertencer à classe permeável ou à impermeável.

$$L(\theta|x) = \prod_{i=1}^n f(x_i|\theta) \quad (8)$$

Onde:

- $x$  são os parâmetros de entrada;
- $\theta$  representa o vetor de parâmetros da distribuição;
- $x_i$  representa o parâmetro de entrada  $i$ ;
- $L(\theta|x)$  é a função de verossimilhança;
- $f(x_i|\theta)$  é a densidade de probabilidade de  $x$  com parâmetros  $\theta$ .

Apesar de este classificador não ser eficiente para problemas com elevada dimensionalidade (Tokarczyk *et al.* 2015), com grandes quantidades de dados diferentes, adapta-se bem aos problemas desta investigação, uma vez que a classificação é baseada nas cores dos pixéis. Segundo o mesmo autor, na sua aplicação em imagens que apresentam doze ou mais dimensões, o seu desempenho diminui, pois a complexidade do problema aumenta.

Como já foi referido ao longo da metodologia, a todas as classificações realizadas pelos diversos algoritmos está associada a precisão global que permite avaliar a capacidade do classificador em aprender através dos exemplos que lhe são disponibilizados.

## 3.4 Avaliação da Precisão Global

A precisão global é um parâmetro que traduz a concordância entre a classificação efetuada pelo utilizador e a classificação realizada pelo classificador. De uma forma simples, são sintetizadas numa matriz de erro, ou de confusão, (Quadro 3.1) as amostras que de forma correta ou incorreta foram atribuídas às diferentes classes pelo algoritmo. Na diagonal da

matriz encontram-se as quantidades de amostras em que ambos os juízes, classificador e utilizador, concordam em cada classe. Fora destas células estão todos os outros valores que foram classificados de forma equívoca.

Quadro 3.1 - Matriz de erro

Verdade de campo	Classificador		
	Positivo	Negativo	Total
Positivo	Verdadeiro Positivo	Falso Negativo	$m_1$
Negativo	Falso Positivo	Verdadeiro Negativo	$m_2$
Total	$n_1$	$n_2$	$N$

Com estas informações a precisão global é calculada de forma simples, através da Equação (9).

$$Precisão\ Global = \frac{\sum \text{Elementos da diagonal}}{\text{Total de amostras}} \quad (9)$$

Este parâmetro é afetado pelo número de amostras de treino, pelo que um treino pequeno pode provocar *underfitting*. Este fenómeno caracteriza-se pelo algoritmo não ter capacidade suficiente para classificar através de uma pequena quantidade de padrões dados ou não tem a complexidade suficiente para compreender os padrões fornecidos (*Underfitting* na Figura 3.6). Também a situação de *overfitting* é desvantajosa para serem atingidos bons resultados da precisão global, na medida que o modelo se ajusta demasiado aos dados. Quando tal acontece, o classificador acaba por se adaptar demasiado a casos específicos apresentados no treino e não generaliza bem quando lhe é pedido para avaliar casos diferentes. O *overfitting* é graficamente representado pela última ilustração da Figura 3.6. Aprendizagem ideal é dada pela representação do meio, na Figura 3.6, em que o algoritmo consegue aprender de forma adequada, e a maioria da classificação é feita de forma correta.

No caso de utilização de algoritmos mais complexos como redes neuronais cujo tamanho da rede é fixo ao longo do treino a performance aumenta linearmente com o tamanho da rede (Kavzoglu e Mather, 2003). No entanto, em algoritmos simples como os utilizados na

presente tese, a evolução da precisão global não tende a seguir nenhuma tendência, uma vez que depende apenas dos treinos efetuados.

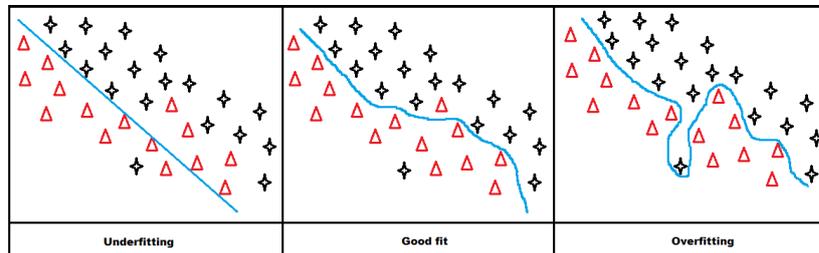


Figura 3.6 - *Underfitting, Good fit e overfitting*

### 3.5 Performance do classificador – Parâmetro Kappa

A precisão global de uma avaliação é um bom parâmetro para serem obtidas algumas conclusões sobre o desempenho do classificador. Através deste coeficiente é possível perceber se o classificador está aprendendo de forma correta os padrões que lhe são ensinados. No entanto, a precisão global não tem em conta a performance do classificador em relação ao acaso, isto é, não permite perceber se os resultados serão melhores do que os obtidos se a classificação fosse feita de forma aleatória. Kappa (K), equação (10), é o parâmetro que permite tirar conclusões sobre a melhoria dos resultados obtidos face ao que resultariam do acaso.

$$K = \frac{P(A) - P(E)}{1 - P(E)} \quad (10)$$

$P(A)$  representa a precisão global e  $P(E)$ , é a probabilidade de sucesso ao acaso, que para duas classes é de 0.5. No entanto, o seu cálculo é efetuado através do somatório da diagonal da matriz do produto dos totais de linha e totais de coluna sob o somatório dos totais de produto entre linhas e colunas. Para duas classes,  $P(E)$  é dado pela equação (11) (adaptado de Viera e Garret, 2005).

$$P(E) = \left[ \left( \frac{n_1}{N} \cdot \frac{m_1}{N} \right) + \left( \frac{n_2}{N} \cdot \frac{m_2}{N} \right) \right] \quad (11)$$

Onde:

- $n_1$  é o número total de amostras classificadas como positivas pelo classificador;
- $n_2$  é o número total de amostras classificadas como negativas pelo classificador;

- $m_1$  é o número total de amostras classificadas como positivas pelo utilizador;
- $m_2$  é o número total de amostras classificadas como negativas pelo utilizador;
- $N$  é o número total de amostras.

O parâmetro Kappa tomará valores de -1 a 1, em que no intervalo -1 e 0 os resultados obtidos são piores do que seriam por acaso. O Quadro 3.2, demonstra o quão bons são os resultados segundo os seus intervalos.

Quadro 3.2 - Significado de Kappa. Adaptado de Viera e Garret (2005)

<b>Kappa</b>	<b>Concordância</b>
<b>&lt; 0</b>	Sem possibilidade de acordo
<b>0.01-0.20</b>	Pequeno acordo
<b>0.21-0.40</b>	Acordo justo
<b>0.41-0.60</b>	Acordo moderado
<b>0.61-0.80</b>	Acordo substancial
<b>0.81-0.99</b>	Quase um acordo perfeito

Todas as classificações e análises serão efetuados através dos *software* Google Earth Engine e ArcMap.

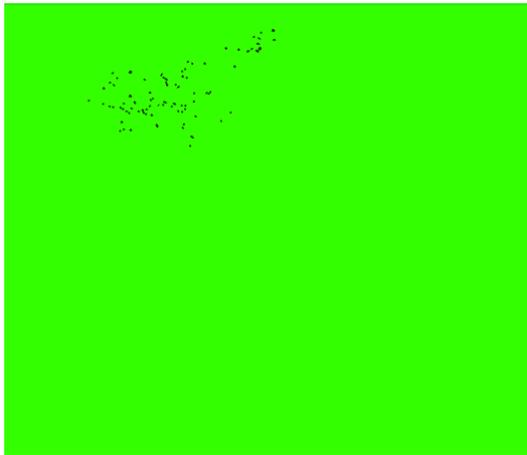
## 4 RESULTADOS E DISCUSSÃO

Como referido na metodologia, a análise efetuada para a imagem de *Landsat* foi realizada no GEE, por pertencer à biblioteca desta plataforma. O ortofotomapa da bacia de Wartegg foi analisado tanto no GEE como no ArcGis.

### 4.1 Classificação com recurso ao Google Earth Engine

#### 4.1.1 Imagem *Landsat*

Como referido no Capítulo 3, foram escolhidos dois algoritmos, o FNB e o *Margin SVM*, dos dez existentes na plataforma. Estes algoritmos foram executados no Explorer do GEE. Foram efetuados treinos de 20, 50, 100, 200, 300, 400 e 500 polígonos sobre a imagem *Landsat* para utilização na classificação com recurso ao FNB e *Margin SVM*. No entanto, para os treinos de 20, 50 e 100 polígonos não foram obtidos resultados para a classe “Impermeável” uma vez que o *software* indicava que não tinham sido detetados quaisquer polígonos para a mesma, classificando assim toda a imagem como permeável. Exemplo disso são as imagens obtidas através do treino de 100 polígonos quer pelo algoritmo FNB (Figura 4.1) quer pelo *Margin SVM* (Figura 4.3). A precisão global destes casos é de 100%, apesar de a classificação ser apenas permeável. Tal acontece devido não terem sido considerados quaisquer polígonos definidos na categoria “Impermeável”, como se pode ver na coluna #Points (Figura 4.2 e Figura 4.4) e que são visíveis nas Figuras 4.1 e 4.3, sendo que os únicos polígonos utilizados são apenas os da classe permeável. Nesta coluna é demonstrado que apenas foram utilizados para treino dois dos 50 polígonos permeáveis desenhados, enquanto nenhum polígono da classe impermeável é utilizado.



### Fast Naive Bayes

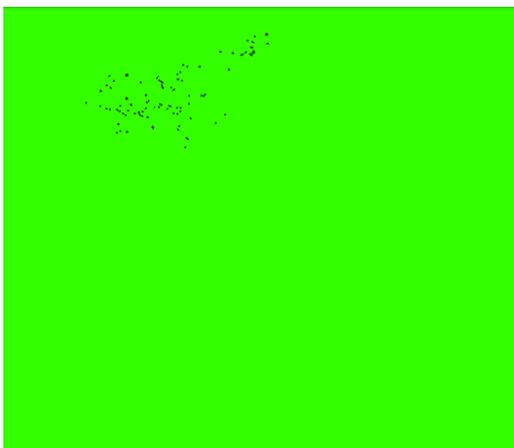
Model, trained Jul 28, 2016 at 12:45am

Overall validity: 100%

	# Points	Impermeável	Permeável
■ Impermeável	0	NaN%	NaN%
■ Permeável	2	0%	100%

Figura 4.1 – Classificação obtida para o treino de 100 polígonos através do FNB

Figura 4.2 - Matriz de erro e precisão global: treino de 100 polígonos realizado com o FNB



### Margin SVM

Model, trained Jul 28, 2016 at 12:45am

Overall validity: 100%

	# Points	Impermeável	Permeável
■ Impermeável	0	NaN%	NaN%
■ Permeável	2	0%	100%

Figura 4.3 - Classificação obtida para o treino de 100 polígonos através do *Margin SVM*

Figura 4.4 – Matriz de erro e precisão global: treino de 100 polígonos realizado com o *Margin SVM*

Embora a precisão global seja de 100% nas Figuras 4.1 e 4.3, esta não se traduz necessariamente num bom resultado. Por exemplo, se apenas for treinado um polígono cuja classe é “Permeável”, então o algoritmo concordará com a sua classificação uma vez que este é o seu único elemento de treino, pelo que aprenderá a informação extraída do mesmo.

Os resultados obtidos para os treinos de 100 polígonos, anteriormente apresentados, suscitaram questões sobre a forma como foram utilizados os polígonos por parte do GEE. Deste modo, foi feito o mesmo treino acrescentando um polígono grande na classe impermeável (Figura 4.5).

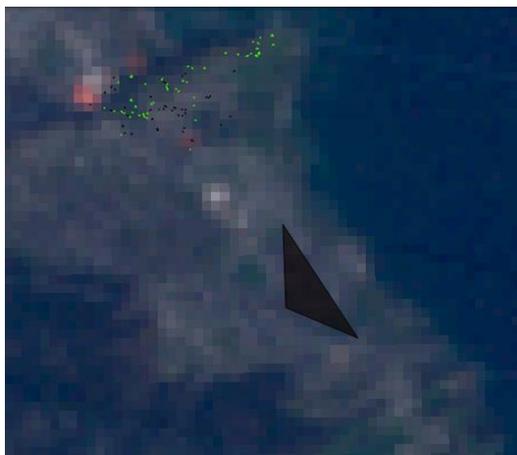


Figura 4.5 – Treino de 101 polígonos

Os resultados obtidos (Figura 4.6, Figura 4.7, Figura 4.8 e Figura 4.9) deste treino confirmam que nem todos os polígonos desenhados sobre a imagem são utilizados. As Figuras 4.6 e 4.8 mostram as classificações feitas pelo algoritmo, no entanto, estas não se assemelham à distribuição real dos usos do solo da bacia de Wartegg. As Figuras 4.7 e 4.9 já apresentam pontos para a classe “Impermeável” em “#Points”, perfazendo os mesmos um total de 54 pontos. Estes pontos foram todos identificados através da utilização do polígono grande a preto na Figura 4.5. O significado destes pontos e como se encontram distribuídos é desconhecido.

A razão apontada para que a classificação não tenha sido efetuada para os treinos de 20, 50 e 100 polígonos é que a classificação da imagem *Landsat* não funciona bem para polígonos tão pequenos. Como a resolução da imagem é baixa, o GEE não consegue retirar informação de polígonos tão pequenos, para efetuar o treino.

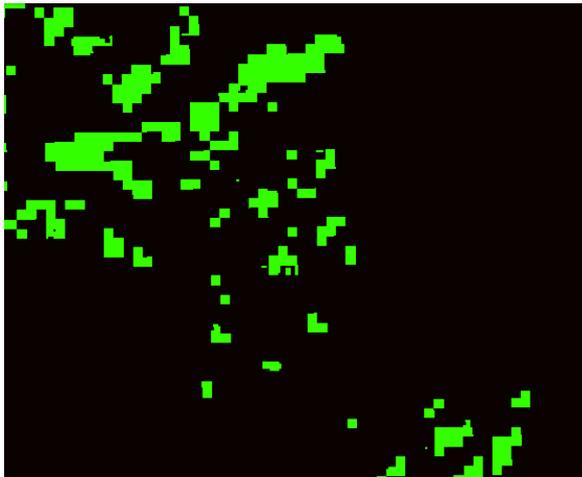


Figura 4.6 - Classificação obtida para o treino de 101 polígonos através do FNB

### Fast Naive Bayes

Model, trained Jul 29, 2016 at 1:09am

Overall validity: 96.43%

	# Points	Impermeável	Permeável
■ Impermeável	54	98.15%	1.85%
■ Permeável	2	50%	50%

Figura 4.7 - Matriz de erro e precisão global: treino de 101 polígonos realizado com o FNB

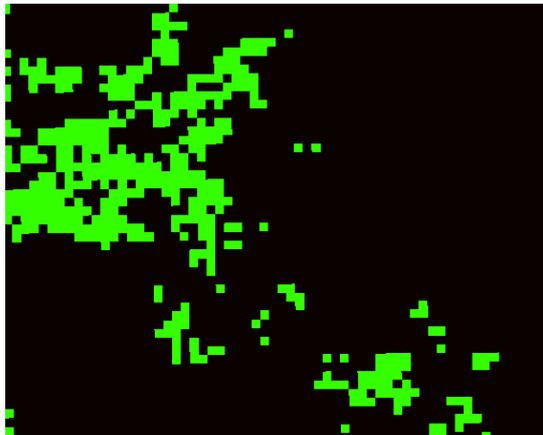


Figura 4.8 - Classificação obtida para o treino de 101 polígonos através do *Margin SVM*

### Margin SVM

Model, trained Jul 29, 2016 at 1:17am

Overall validity: 100%

	# Points	Impermeável	Permeável
■ Impermeável	54	100%	0%
■ Permeável	2	0%	100%

Figura 4.9 - Matriz de erro e precisão global: treino de 101 polígonos realizado com o *Margin SVM*

Para os restantes treinos de 200, 300, 400 e 500 polígonos foram obtidos resultados, pelo que deverão existir polígonos suficientemente grandes para que possa ser classificada a imagem. As Figuras 4.10 e 4.12 apresentam as imagens classificadas através do treino de 500 polígonos através dos algoritmos FNB e *Margin SVM*, respetivamente. Em ambos os treinos

efetuados, é utilizado 1 ponto impermeável e 6 pontos permeáveis (Figura 4.11 e Figura 4.13). Como é possível visualizar, como só é utilizado um ponto para a classe “Impermeável” e o algoritmo é treinado com base nessa informação, então a classificação está certa segundo o que o classificador treinou, sendo a sua precisão para esta classe de 100%.

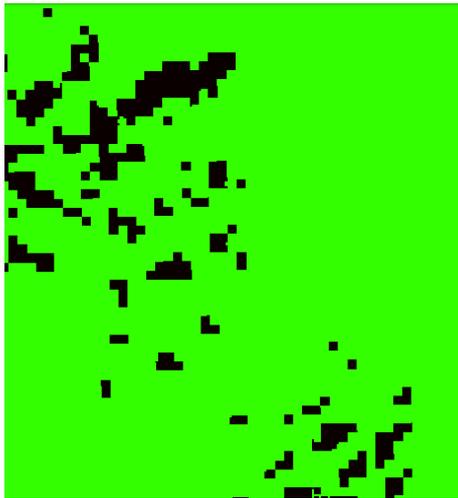


Figura 4.10 - Classificação obtida para o treino de 500 polígonos através do FNB

#### Fast Naive Bayes

Model, trained Jul 24, 2016 at 7:09pm

Overall validity: 71.43%

	# Points	Impermeável	Permeável
■ Impermeável	1	100%	0%
■ Permeável	6	33.33%	66.67%

Figura 4.11 - Matriz de erro e precisão global: treino de 500 polígonos realizado com o FNB

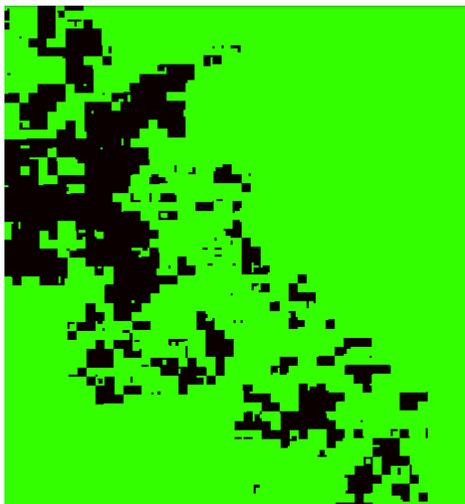


Figura 4.12 - Matriz de erro e precisão global: treino de 500 polígonos realizado com o *Margin SVM*

#### Margin SVM

Model, trained Jul 24, 2016 at 7:09pm

Overall validity: 100%

	# Points	Impermeável	Permeável
■ Impermeável	1	100%	0%
■ Permeável	6	0%	100%

Figura 4.13 - Matriz de erro e precisão global: treino de 500 polígonos realizado com o *Margin SVM*

A diferença de tamanhos nos polígonos depende essencialmente do objeto que está a ser classificado, pelo que, por exemplo, pequenos arbustos tenham pequenos polígonos, assim como telhados tenham treinos um pouco maiores. Pelo que se observa pelas matrizes de erro

anteriormente observadas, deverão existir um ou mais treinos em áreas permeáveis suficientemente grandes para que fosse possível ser extraída informação da mesma.

Recorda-se que os locais e tamanhos dos polígonos são os mesmos que os polígonos dos treinos efetuados no ArcGis. Tal era para que fosse possível comparar diretamente os resultados obtidos nos diferentes algoritmos, uma vez que os dados de entrada foram os mesmos. No entanto, é visível que o GEE não utilizou toda informação de entrada, pelo que se supõe que tenha a ver com o pequeno tamanho dos polígonos não permitir que seja extraída informação dos próprios. Para de facto poder ser afirmado que esta situação ocorre devido à resolução da imagem *Landsat*, serão efetuados os mesmos treinos no ortofotomapa.

#### 4.1.2 Ortofotomapa

As Figuras 4.14 e 4.15 exibem os resultados obtidos através do *Fast Naive Bayes* para o treino de 20 polígonos, efetuado sobre o ortofotomapa. Destaca-se desde logo a diferença da imagem obtida, quando comparada com os resultados referentes ao *Landsat*. Devido à elevada resolução do ortofotomapa, é notório que existe uma maior facilidade em descobrir erros na classificação, pelo facto de serem mais facilmente identificados os tipos de uso do solo. Para o *Landsat*, devido à sua resolução, os usos do solo são representados de forma grosseira.

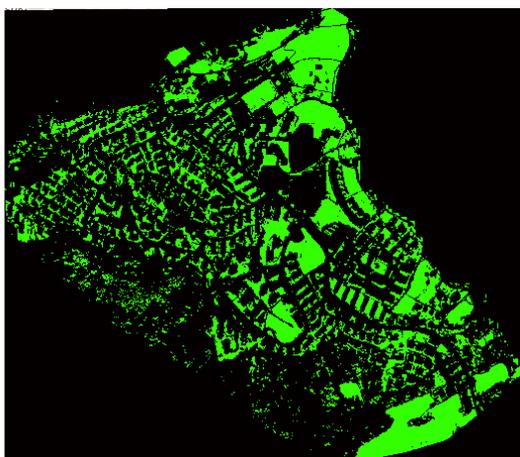


Figura 4.14– Classificação obtida para o treino de 20 polígonos através do algoritmo FNB

#### Fast Naive Bayes

Model, trained Jun 23, 2016 at 11:58am

Overall validity: 88.46%

	# Points	Impermeável	Permeável
■ Impermeável	46	95.65%	4.35%
■ Permeável	32	21.88%	78.13%

Figura 4.15 - Matriz de erro e precisão global: treino de 20 polígonos realizado com o FNB

Na Figura 4.16 é possível observar que as zonas de relvado são, na sua maioria bem identificadas como permeáveis, no entanto as zonas de árvores e arbustos são classificadas maioritariamente como impermeáveis. Embora o ortofotomapa tenha o formato apresentado na Figura 3.3, a zona exterior aos seus limites, foi classificada automaticamente no GEE. Não foi encontrada resposta para o sucedido, porém, este fator não tem influência nos resultados uma vez que fora das fronteiras não foram treinados quaisquer polígonos, pelo que não afeta os resultados da precisão global e da matriz de erro.

Enquanto para imagem de *Landsat* o maior número de pontos utilizados foi 7, na Figura 4.18 é possível ver esse número aumentou para 78, para o treino de 20 polígonos realizado com o ortofotomapa. Tal, é devido à resolução de imagem ser maior, em que na mesma é possível adquirir maior informação dos polígonos treinados. Neste treino, a precisão de avaliação é de 88.46%, o que significa que o classificador concorda em 88.46% do que foi identificado manualmente.

Uma vez que, para o ortofotomapa, um maior número de informação (número de pontos em “#Points”) foi utilizada para a classificação de imagens, são apresentados três treinos cujos resultados em termos de precisão global, foram mais significantes. Nestes já se inserem os resultados apresentados anteriormente, na Figura 4.14 e na Figura 4.15.

Para o treino de 200 polígonos, a quantidade de pontos utilizados na classificação foi de 642. O aumento deste número está relacionado com o facto de ser possível tirar maior quantidade de informação devido ao aumento do treino.

A precisão de avaliação diminuiu face à do treino de 20 polígonos, situando-se nos 85.67% (Figura 4.17), sendo que o classificador discorda em 11% das áreas que foram classificadas como impermeáveis pelo utilizador e em 17.42% das zonas identificadas como permeáveis. Na imagem (Figura 4.16), semelhantemente ao que foi verificado na Figura 4.14, as zonas de vegetação são maioritariamente classificadas como impermeáveis. A razão apontada para que tal ocorra é que nestas áreas encontram-se algumas cores térreas e esverdeadas, que são cores semelhantes às cores dos telhados com relvados. Devido a esses telhados estarem intensivamente treinados, o classificador identifica que a gama de cores da vegetação pertence à classe impermeável.

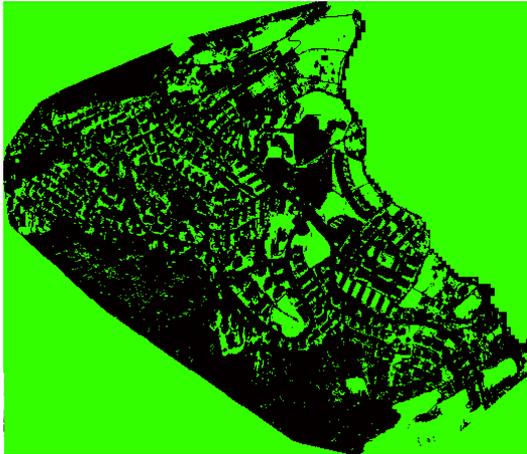


Figura 4.16- Classificação obtida para o treino de 200 polígonos através do algoritmo FNB

### Fast Naive Bayes

Model, trained Jun 23, 2016 at 11:35am

Overall validity: 85.67%

	# Points	Impermeável	Permeável
■ Impermeável	309	89%	11%
■ Permeável	333	17.42%	82.58%

Figura 4.17 - Matriz de erro e precisão global: treino de 200 polígonos realizado com o FNB

Na Figura 4.18, referente ao treino de 400 polígonos, é possível constatar que apesar de à medida que o treino aumenta, assim como a variedade de cores treinadas, inclusive as dos padrões de vegetação, o problema da classificação da vegetação como impermeável persiste. No entanto, pela razão anteriormente apontada, o algoritmo tende a classificar erroneamente objetos de cores semelhantes, tais como as zonas de arvoredo e telhados verdes.

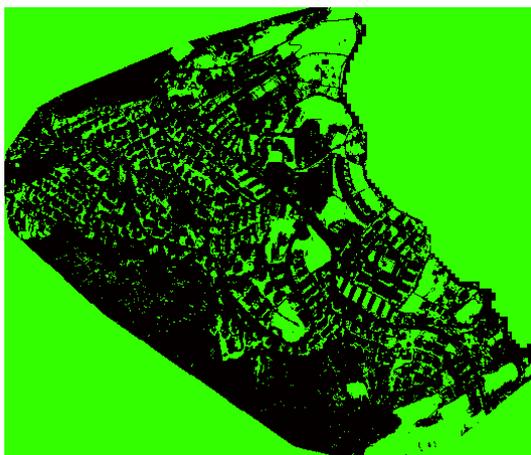


Figura 4.18 – Classificação obtida para o treino de 400 polígonos através do algoritmo FNB

### Fast Naive Bayes

Model, trained Jun 23, 2016 at 12:08pm

Overall validity: 83.43%

	# Points	Impermeável	Permeável
■ Impermeável	652	87.58%	12.42%
■ Permeável	670	20.6%	79.4%

Figura 4.19 - Matriz de erro e precisão global: treino de 400 polígonos realizado com o FNB

Embora não tenham sido analisados todos os resultados obtidos, é apresentado um quadro síntese (Quadro 4.1) dos valores da precisão global de todos os treinos.

Quadro 4.1 – Precisão global obtida para o FNB para a classificação do ortofotomapa

<i>Fast Naive Bayes</i>	
Nº de polígonos por treino	Precisão global (%)
20	88.46
50	84.51
100	84.05
200	85.67
300	83.64
400	83.43
500	85.23

O melhor resultado obtido para a precisão de avaliação na análise efetuada com o FNB é de 88.46%, correspondendo ao treino de 20 polígonos. Através deste facto, é possível concluir que para este exemplo em específico, o treino de 20 polígonos seria suficiente para serem obtidos resultados aceitáveis, se apenas fosse importante analisar a precisão global, todavia é necessário fazer uma análise da imagem. Como é possível ver na Figura 4.14, a maioria das zonas impermeáveis e áreas de relvados são bem identificadas porém, as regiões de

vegetação, tais como de arvoredo e arbustos, são classificadas como impermeáveis. Relativamente aos restantes resultados, apesar de haver algumas quebras na precisão de avaliação, não pode ser considerado que haja *overfitting* uma vez que as percentagens variam num intervalo pequeno entre 80% e 90%. A precisão global segue um comportamento estável com o aumento do número de polígonos de treino, em que o melhor resultado, como anteriormente mencionado, é obtido para o treino de 20 polígonos.

Tal como foi efetuado para análise do ortofotomapa recorrendo ao algoritmo FNB, são apresentados os três resultados mais relevantes para os treinos efetuados com o classificador *Margin SVM*.

Apesar de na Figura 4.20 as zonas de vegetação apresentarem uma melhoria na classificação, face ao mesmo treino realizado com o *Fast Naive Bayes*, ainda existem algumas zonas que continuam a ser classificadas como impermeáveis. Esta alteração visual comparada com a Figura 4.14 é explicada pela diferente aprendizagem que é feita pelos algoritmos, uma vez foram utilizados exatamente os mesmos polígonos.

Também na comparação da precisão da avaliação, quando comparada com a obtida com o FNB, para o mesmo treino, é ligeiramente superior, não sendo um aumento significativo. Os juízes concordam com a classificação “Impermeável” em todas as áreas treinadas como tal mas o classificador discorda em 25% dos polígonos classificados como permeável pelo utilizador.

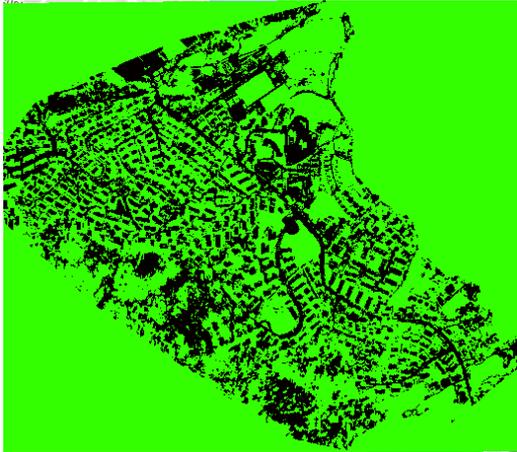


Figura 4.20 – Classificação obtida para o treino de 20 polígonos através do algoritmo *Margin SVM*

### Margin SVM

Model, trained Jun 23, 2016 at 11:58am

Overall validity: 89.74%

	# Points	Impermeável	Permeável
■ Impermeável	46	100%	0%
■ Permeável	32	25%	75%

Figura 4.21 - Matriz de erro e precisão global: treino de 20 polígonos realizado com o *Margin SVM*.

A Figura 4.22 apresenta alterações na classificação da zona de vegetação, sendo que esta encontra-se demarcada mais a preto do que a imagem resultante do treino de 20 polígonos (Figura 4.20). Tal deve-se à classificação errónea do algoritmo, explicada anteriormente.

Como é visível na coluna “#Points” da Figura 4.23, foram utilizados 301 pontos no treino de 100 polígonos. A precisão global referente a este treino diminuiu muito pouco quando comparada com o valor do mesmo parâmetro obtido para o treino de 20 polígonos.

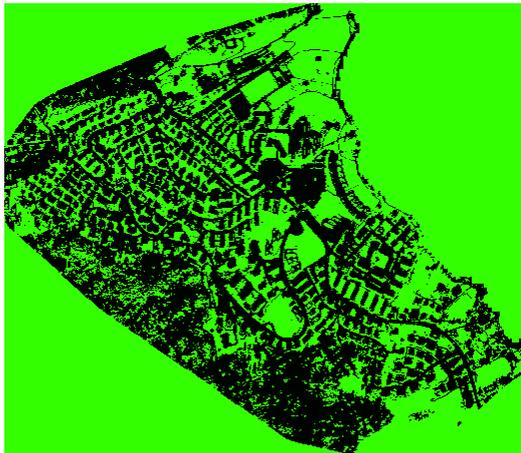


Figura 4.22– Classificação obtida para o treino de 100 polígonos através do algoritmo *Margin SVM*

### Margin SVM

Model, trained Jun 23, 2016 at 11:45am

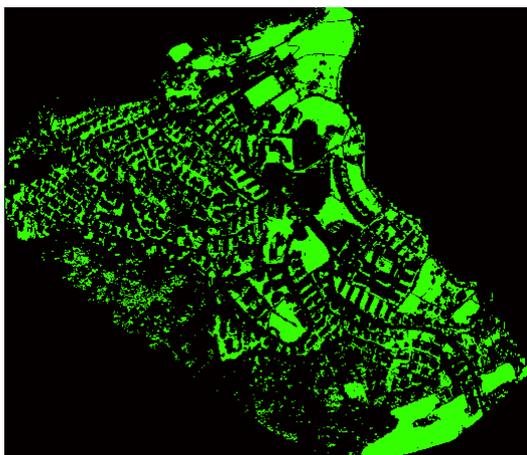
Overall validity: 88.37%

	# Points	Impermeável	Permeável
■ Impermeável	134	98.51%	1.49%
■ Permeável	167	19.76%	80.24%

Figura 4.23 - Matriz de erro e precisão global: treino de 100 polígonos realizado com o *Margin SVM*

Nos resultados seguintes não são observadas diferenças, além das zonas exteriores aos limites, entre as Figura 4.22 e 4.24, no entanto, a precisão de avaliação (Figura 4.25) decresceu face à anterior. Ambos os erros de classificação de zonas impermeáveis, quer permeável, são de 11.66% e 21.66%, respetivamente. Neste treino, o algoritmo foi treinado com a informação extraída de 923 pontos.

Como constatável, quanto maior o número de polígonos presentes num treino, maior o número de pontos em “#Points”. Isto ocorre devido ao maior número de polígonos permitir recolher um maior número de informação.



### Margin SVM

Model, trained Jun 23, 2016 at 11:23am

Overall validity: 83.64%

	# Points	Impermeável	Permeável
■ Impermeável	489	88.34%	11.66%
■ Permeável	434	21.66%	78.34%

Figura 4.24– Classificação obtida para o treino de 300 polígonos através do algoritmo *Margin SVM*

Figura 4.25 Matriz de erro e precisão global: treino de 300 polígonos realizado com o *Margin SVM*

Em suma, o Quadro 4.2 apresenta todos os valores da precisão global obtidos para todos os treinos efetuados com o *Margin SVM*.

Quadro 4.2 - Precisão global obtida para o *Margin SVM* para a classificação do ortofotomapa

<i>Margin SVM</i>	
Nº de polígonos por treino	Precisão global (%)
20	89.74
50	87.32
100	88.37
200	85.98
300	83.64
400	85.1
500	85.48

Ao analisar-se o Quadro 4.2 é possível perceber que, tal como o sucedido com o algoritmo FNB, o melhor resultado é alcançado com o treino de 20 polígonos. Este resultado é considerado um bom resultado na medida que tem uma precisão global elevada que é

confirmada pela representação gráfica na figura 4.20, dado que esta se aproxima à realidade da bacia.

Ao comparar a evolução da precisão global de ambos os algoritmos, observamos que os mesmos têm um comportamento quase estável (Figura 4.26). Entre os treinos de 20 e 50, o FNB apresenta uma queda de 4% nos resultados, enquanto o *Margin SVM* tem uma queda mais ligeira de 2%. Entre os treinos de 50 para 100 polígonos, o FNB continua a decrescer, embora pouco, não chegando a 1%, ao contrário do *Margin SVM*, que regista uma subida de igual valor. Entre os treinos de 100 e 200 polígonos, os papéis invertem-se sendo que a precisão de avaliação no FNB melhora, embora apenas em cerca de 1.62% e no *Margin SVM*, diminui. Já entre os treinos de 200 e 300 polígonos, os algoritmos comportam-se de forma idêntica, diminuindo ambos, sendo que é neste treino que se regista o pior resultado obtido através do *Margin SVM*, cujo valor já foi mencionado. Os resultados voltam a entrar em discordância entre os treinos de 300 para 400, estando a precisão de avaliação no FNB a diminuir, registando o seu pior resultado neste treino face a precisão obtida pelo *Margin SVM* que aumenta. Por fim, entre os treinos 400 e 500, ambos os treinos melhoram a precisão de avaliação, sendo que a melhoria obtida com o *Margin SVM* é mais contida.

Embora os gráficos apresentem-se quase estáveis (Figura 4.26), existem pequenas quebras. No entanto, estas não podem ser consideradas *overfitting*, por o intervalo em que varia a precisão global ser muito pequeno (entre 83 e 89%).

Quadro 4.3 - Precisão global obtida para todos os treinos realizados com o FNB e *Margin SVM*

Ortofotomapa		
Treino (Polígonos)	Precisão global FNB (%)	Precisão global SVM (%)
20	88.46	89.74
50	84.51	87.32
100	84.05	88.37
200	85.67	85.98
300	83.64	83.64
400	83.43	85.1
500	85.23	85.48

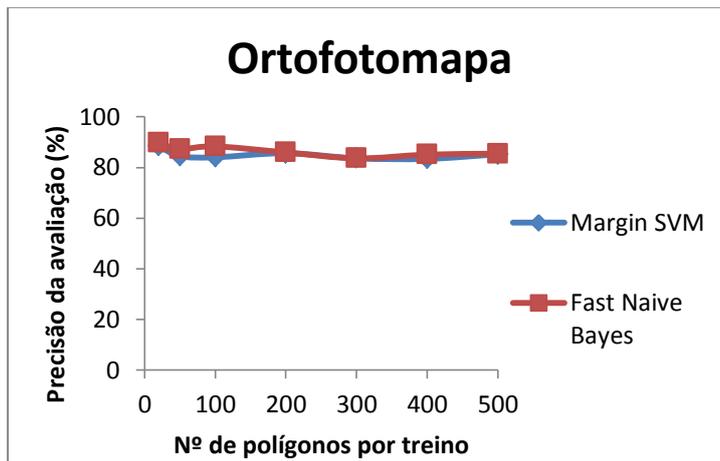


Figura 4.26 – Evolução da precisão global para todos os treinos realizados com o FNB e *Margin SVM*.

Conclui-se que a precisão global obtida pelos dois algoritmos, FNB e *Margin SVM*, para a classificação do ortofotomapa, tem um comportamento quase constante. As pequenas alterações nos resultados são referentes à confusão do algoritmo na classificação de determinadas cores, uma vez que as mesmas encontram-se em duas classes distintas.

Na generalidade, a classificação dos algoritmos é boa, representando bem a verdade do terreno, no entanto, as zonas de arvoredo e arbustos, são intensivamente classificadas como impermeáveis. É um facto que os classificadores tendem a classificar erroneamente objetos de cores semelhantes, tais como as zonas de arvoredo e telhados verdes.

O Explorer do GEE é uma plataforma de análise de imagens contendo uma vasta biblioteca de fotografias e de algoritmos, contudo, não oferece a possibilidade de realização de cálculos, pelo que na utilização deste *software*, não foi possível calcular o parâmetro K e obter uma estimativa sobre as percentagens de áreas permeáveis e impermeáveis.

Para perceber como a classificação de todos os treinos realizados é feita para o ortofotomapa através do Método da Máxima Verossimilhança, recorreu-se ao ArcGis, onde se encontra implementado este algoritmo. Também desta análise pretende-se perceber quais as vantagens e desvantagens na utilização deste *software* face à utilização do Google Earth Engine.

## 4.2 Avaliação feita com recurso ao ArcMap

A primeira imagem, Figura 4.27, foi conseguida através do treino de 20 polígonos. Conforme o que foi observado nos treinos efetuados com os outros algoritmos, a zona de vegetação é classificada, na sua maioria, como impermeável.

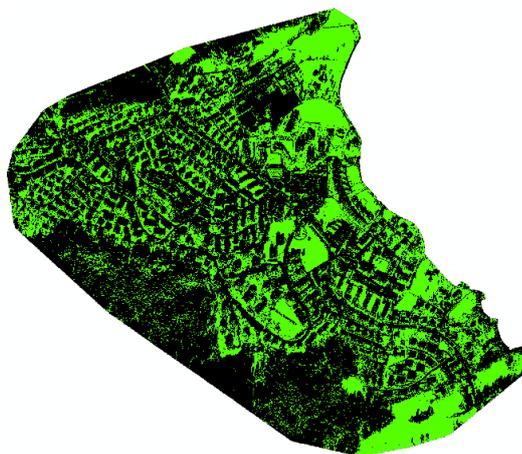


Figura 4.27 – Classificação obtida para o treino de 20 polígonos

Ao contrário do GEE, onde a matriz de erro é obtida diretamente de todos os polígonos desenhados, no ArcGis foram desenhados pontos de controlo para proceder ao cálculo da precisão global e de K. A forma de como os pontos de controlo foram criados encontra-se descrito na metodologia na secção 3.3.2.

No Quadro 4.4, é possível ver que o classificador encontra-se de acordo com a maioria dos pontos de controlo classificados pelo utilizador no entanto, existe desacordo em 23 pontos de um total de 100 pontos. Estes resultados representam uma precisão de avaliação de 77% o que, quando comparado com os resultados obtidos para os treinos de 20 polígonos realizados com o FNB e o *Margin SVM*, o resultado piorou significativamente.

Quadro 4.4 - Resultados para o treino de 20 polígonos

Treino de 20 Polígonos			
Classe	Classificador		
	Permeável	Impermeável	Total
Permeável	38	12	50
Impermeável	11	39	50
Total	49	51	100

Para o treino de 50 polígonos, a Figura 4.28 revela que a classificação sofreu uma melhoria no que diz respeito à zona de vegetação. Enquanto os campos de futebol, foram classificados como permeáveis, nos resultados obtidos com os algoritmos anteriores, neste treino e no treino de 20 polígonos efetuados com o Método da Máxima Verossimilhança, foram classificados como superfícies impermeáveis.

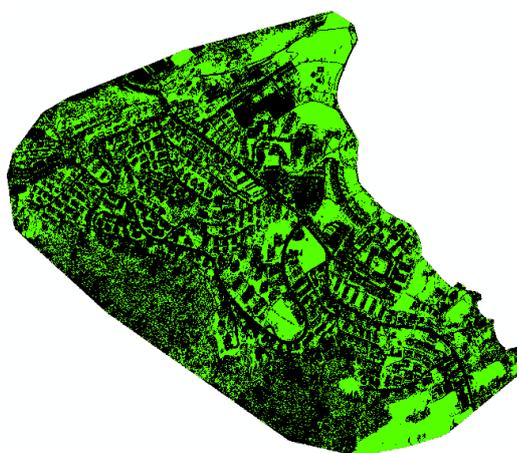


Figura 4.28– Classificação obtida para o treino de 50 polígonos

Em termos numéricos, o erro de classificação diminuiu, sendo que a discordância é referente a 12 polígonos no total (Quadro 4.5), sendo que a precisão de avaliação correspondente é de 88%. Com o aumento do número de polígonos de treino, houve um aumento significativo da precisão global, de 11%. Quando comparados com os resultados obtidos nos outros algoritmos, este apresentam melhores resultados, em que para o FNB e *Margin SVM*, as precisões globais eram, respectivamente, 84.51 e 87.32%.

Quadro 4.5 - Resultados para o treino de 50 polígonos

Treino de 50 Polígonos			
Classificador			
Classe	Permeável	Impermeável	Total
Permeável	41	9	50
Impermeável	3	47	50
Total	44	56	100

Em termos visuais, a Figura 4.29 assemelha-se mais à realidade do que à anterior, em que grande parte da vegetação é classificada como permeável.

A elevada resolução da imagem permite ao algoritmo classificar objetos de pequena dimensão que não têm qualquer tipo de interesse na classificação. Exemplo disso é o facto das linhas dos campos de futebol, de cor branca, terem sido classificadas impermeáveis devido à sua cor.

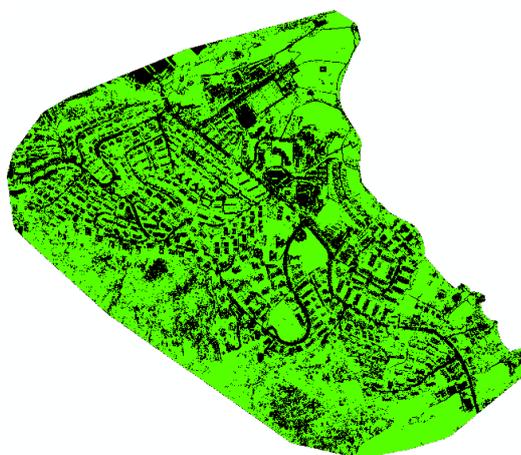


Figura 4.29 – Classificação obtida para o treino de 100 polígonos

Ao ser avaliada a concordância entre o utilizador e o classificador sobre as classes dos pontos de controlo, houve divergência no total de 15 pontos como é possível ver no Quadro 4.6. Com base nesta matriz, a precisão de avaliação é de 85%, o que caracteriza uma quebra de 3% face ao resultado anterior. Esta descida pode ser justificada pelo facto de o algoritmo ter classificado parte da zona de vegetação como permeável, e poderá, por esta razão, ter invertido a classificação de algumas zonas impermeáveis. Exemplo disso são alguns telhados e as estradas.

Comparada com os treinos de 100 polígonos, através dos outros algoritmos, precisão global obtida através do Método da Máxima Verossimilhança é ligeiramente melhor do que a obtida pelo FNB (84.05%) e pior do que a do *Margin SVM* (88.35%)

Quadro 4.6 – Resultados para o treino de 100 polígonos

Treino de 100 Polígonos			
Classificador			
Classe	Permeável	Impermeável	Total
<b>Permeável</b>	40	10	50
<b>Impermeável</b>	5	45	50
<b>Total</b>	45	55	100

Em síntese, o quadro 4.7 exhibe todas as precisões de avaliação, calculadas para os treinos efetuados no ArcGis, assim como o parâmetro K. Os melhores resultados numéricos obtidos são referentes ao treino de 50 polígonos e os piores referentes aos treinos de 200 e 500 polígonos, cujas precisões de avaliação são de 82%.

Quadro 4.7 – Evolução da precisão de avaliação e de Kappa com o treino

Treino (Polígonos)	Precisão (%)	K
<b>20</b>	<b>77</b>	<b>0.54</b>
<b>50</b>	<b>88</b>	<b>0.76</b>
<b>100</b>	85	0.70
<b>200</b>	82	0.64
<b>300</b>	85	0.70
<b>400</b>	83	0.66
<b>500</b>	82	0.64

Segundo o Quadro 3.2, apresentado para o treino de 20 polígonos existe um acordo moderado, em que a classificação obtida representa um resultado melhor em 54% face ao que seria obtido aleatoriamente. Para os restantes valores de K, existe um acordo substancial. Todos estes valores de K são bons uma vez que representam que ao fazer treinos, estão a ser melhoradas as classificações face ao acaso.

Para perceber a relação entre a precisão global e K, foram compilados os seus valores na Figura 4.30. Neste gráfico, é possível constatar, como já referido, que o melhor resultado é referente ao treino de 50 polígonos. Do treino de 20 polígonos para o de 50 polígonos, existe uma melhoria significativa de 11%, contrariamente aos restantes resultados cujas variações

são pequenas. Apesar da precisão global ter um maior valor para o treino de 50 polígonos, quando comparado com o treino de 100 polígonos, na verdade a imagem obtida através deste último (Figura 4.29) apresenta uma maior aproximação visual à realidade, uma vez que esta sub-bacia de Lucerna tem grandes áreas permeáveis.

Ainda na Figura 4.30 observa-se que o K tem o mesmo comportamento que a precisão global, sendo também o seu melhor valor referente à classificação obtida pelo treino de 50 polígonos. Tal significa que o resultado é 76% melhor do que os que seriam alcançados aleatoriamente.

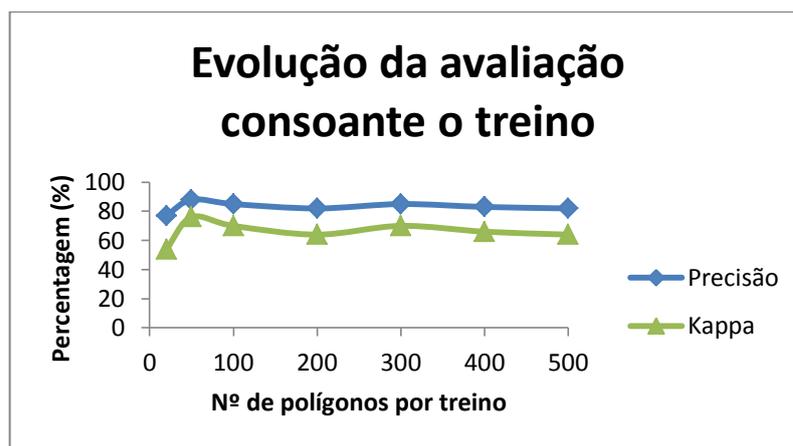


Figura 4.30 - Evolução da precisão consoante o treino

A Figura 4.31 permite fazer a comparação do comportamento da precisão global das classificações obtidas através dos algoritmos FNB, *Margin SVM* e Método da Máxima Verossimilhança.

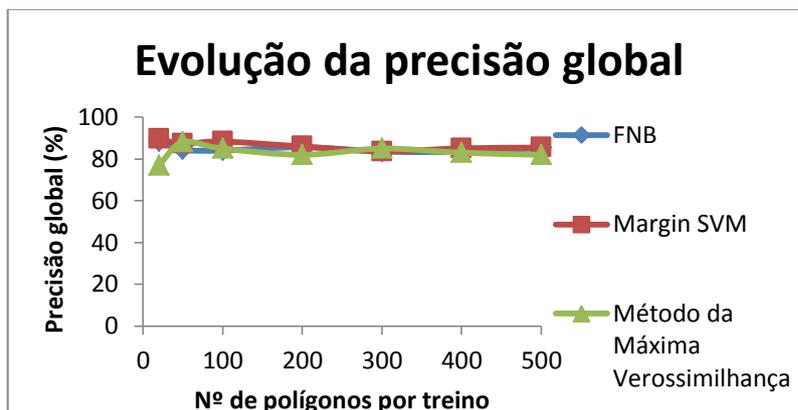


Figura 4.31 – Evolução da precisão global dos três algoritmos

Como é possível observar, apesar de algumas variações visíveis no gráfico, os algoritmos tendem todos a ter um comportamento estável. Este facto mostra a robustez destes algoritmos perante quantidade de parâmetros de entrada diferentes.

Embora não tenha sido possível saber através das análises efetuadas no GEE as percentagens de áreas, o ArcGis permite-o, sendo que para o Método da Máxima Verossimilhança, as percentagens de áreas permeáveis e impermeáveis encontram-se referidas no Quadro 4.8.

Quadro 4.8– Percentagem de áreas impermeáveis e permeáveis

Treino Polígonos	% Área	
	Impermeável	Permeável
20	64.3	35.7
50	61.7	38.3
100	<b>32.7</b>	<b>67.3</b>
200	67.9	32.1
300	65.3	34.7
400	61.6	38.4
500	<b>54.7</b>	<b>45.3</b>

Para a maioria dos treinos, a percentagem de área impermeável encontra-se entre os valores 61 e 68%, no entanto, no treino correspondente a 100 polígonos a área permeável é apenas de 32.%, constituindo, assim, num resultado estranho. A mesma situação irregular volta ocorrer para o treino de 100 polígonos onde a área permeável é cerca de 67% enquanto, comparado com os restantes resultados a área permeável situa-se ente 32 e 39%. Não foi encontrada

nenhuma justificção para estes resultados irregulares. Para o treino de 500 polígonos as percentagens de áreas impermeáveis e permeáveis, encontram-se fora do intervalo de 61 a 68%, sendo respetivamente, 54.7% e 45.3%.

#### 4.2.1 Efeito da seleção da zona de treino

Semelhante ao que ocorreu com os treinos efetuados no Explorer do Google Earth Engine, a zona de vegetação é na maioria das vezes classificada como impermeável. Para investigar como este fenómeno é influenciado pelos locais escolhidos para treino, foi treinada uma área diferente, incluindo um maior número de polígonos na zona de vegetação. Para tal recorreu-se apenas a dois treinos, o de 50 polígonos, que obteve os melhores resultados e treino de 100 polígonos cujas percentagens de áreas, de cada classe, eram anómalas. Este último treino foi escolhido para perceber se a irregularidade nos resultados repete-se. A região de treino utilizada está identificada a vermelho na Figura 4.32, a qual será referida como zona de treino B.



Figura 4.32 – Zona de treino B

A Figura 4.33 apresenta a classificação obtida para o treino de 50 polígonos efetuado na zona B. Apesar do treino ter sido mais intenso na zona de vegetação, parte desta ainda continua a ser classificada como impermeável. No entanto, quando comparada com a Figura 4.28, percebe-se que houve uma melhoria de classificação das zonas de vegetação.

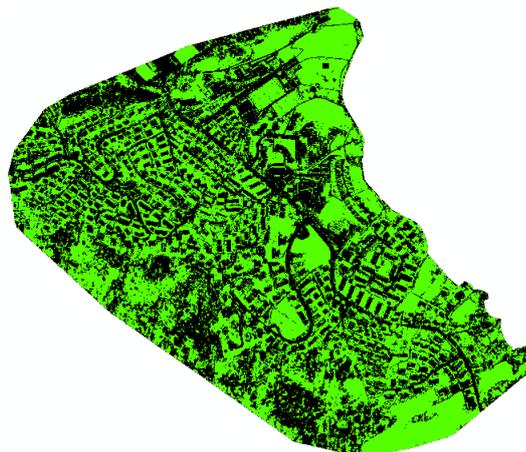


Figura 4.33 – Classificação obtida para o treino de 50 polígonos com zona de treino B

Mais uma vez, é evidente, através do Quadro 4.9, que a grande discordância entre o utilizador e o classificador foi referente à classificação de zonas permeáveis, discordando ambos em 12 pontos. Com estes resultados foi calculada a precisão de avaliação que tem valor de 83%. Quando comparada ao treino de 50 polígonos efetuado na zona de treino A (Anexo B.1), a precisão de avaliação diminuiu em 5%.

Quadro 4.9 – Resultados para o treino de 50 polígonos na Zona B

<b>Treino de 50 Polígonos na zona B</b>			
	Classificador		
<b>Classe</b>	Permeável	Impermeável	Total
<b>Permeável</b>	38	12	50
<b>Impermeável</b>	5	45	50
<b>Total</b>	43	57	100

Na realização do treino de 100 polígonos na zona B, a classificação (Figura 4.34) da zona de vegetação apresenta uma melhoria, no entanto a precisão de avaliação diminuiu 3%. Foram 11 os pontos de desacordo para as zonas permeáveis e 9 para as zonas impermeáveis, como demonstra o Quadro 4.10.

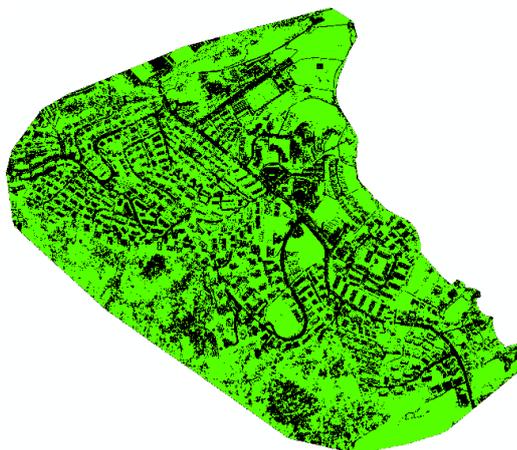


Figura 4.34 – Classificação obtida para o treino de 100 polígonos com zona de treino B

Quadro 4.10– Resultados para o treino de 100 polígonos na Zona B

Treino de 100 Polígonos na zona B			
Classe	Classificador		Total
	Permeável	Impermeável	
Permeável	39	11	50
Impermeável	9	41	50
Total	48	52	100

Em ambos os treinos quando comparados com os resultados das Figuras 4.28 e 4.29, referentes aos treinos de mesma dimensão, efetuados na zona A, as precisões globais diminuíram 5% cada. A razão apontada é a mesma que referida ao longo desta análise, devido à confusão por parte do algoritmo. Em termos de áreas, o treino de 50 polígonos, realizado na zona B, apresenta alguma diferença nas percentagens das mesmas. A área permeável que outrora era de 61.7% é agora de 46.6% (Quadro 4.4) , enquanto a área impermeável, neste caso, aumentou para 53.4%.

No treino de 100 polígonos, torna-se a constatar que a área impermeável é aproximadamente 34% e a percentagem de área permeável 66%. No entanto, tal como referido no igual treino de realizado na zona A, não foi encontrada nenhuma resposta para o sucedido.

Quadro 4.11– Percentagem de áreas impermeáveis e permeáveis

Treino na zona B	% Área	
	Impermeável	Permeável
50	46.6	53.4
100	33.7	66.3

Para se compreender se a precisão de avaliação é influenciada significativamente pela localização dos polígonos, foram feitos dois outros treinos de 50 e 100 polígonos situados na Zona de treino C, identificada pela Figura 4.35.



Figura 4.35 – Zona de treino C

Uma vez que ao longo de toda análise que tem vindo sendo feita, as zonas de vegetação encontram-se em foque pelas razões apontadas ao longo desta discussão, intensificou-se o treino nestas áreas. O primeiro resultado, Figura 4.36, foi obtido através da realização de um treino de 50 polígonos.

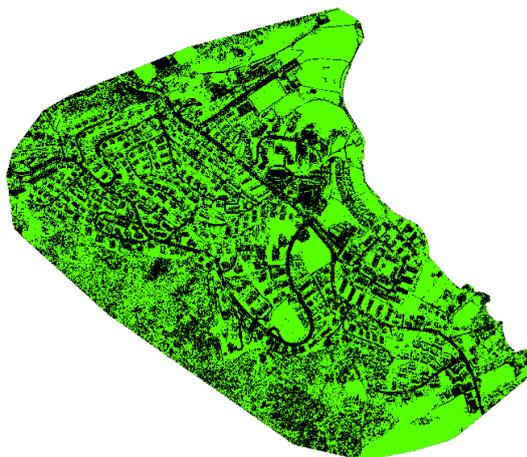


Figura 4.36 – Classificação obtida para o treino de 50 polígonos com zona de treino C

Como é possível observar na Figura 4.36, o aumento da intensidade de treino nas regiões de vegetação apresentou uma melhoria na classificação daquelas áreas contudo, ainda existe uma grande quantidade de zonas de permeáveis a serem classificadas como impermeáveis.

O Quadro 4.12 mostra que a discordância entre juízes é referente a 17 pontos de controlo, pelo que o maior erro de classificação é dado pela identificação de zonas impermeáveis como permeáveis. A precisão de avaliação resultante desta classificação é de 83%.

Quadro 4.12 – Resultados para o treino de 100 polígonos na zona B

Treino de 50 Polígonos			
Classificador			
Classe	Permeável	Impermeável	Total
Permeável	43	7	50
Impermeável	10	40	50
Total	53	47	100

A Figura 4.37 é resultante da classificação feita através do treino de 100 polígonos efetuado na zona de treino C. Nesta imagem é visível que a classificação da zona de vegetação piorou.



Figura 4.37 – Classificação obtida para o treino de 100 polígonos com zona de treino C

Através do cálculo da diagonal do Quadro 4.13, foi obtida uma precisão global de 85%, que representa uma melhoria face ao valor do mesmo parâmetro obtido para o treino de 100 polígonos na Zona B.

Quadro 4.13– Resultados obtidos para o treino de 100 polígonos na zona C

Treino de 100 Polígonos			
Classificador			
Classe	Permeável	Impermeável	Total
Permeável	43	7	50
Impermeável	8	42	50
Total	51	49	100

No quadro 4.14, encontra-se a precisão de avaliação para os treinos de 50 e 100 polígonos nas zonas A, B e C. O objetivo é tentar perceber a influência da localização dos treinos na precisão.

Quadro 4.14 - Precisão de avaliação obtida para os treinos de 50 e 100 polígonos nas diferentes zonas

Precisão Global (%)		
Treino	50	100
Zona A	88	85
Zona B	83	80
Zona C	83	85

Através da análise do Quadro 4.14 conclui-se que a zona de treino não influencia muito na precisão global, o que mostra a boa capacidade de generalização do algoritmo. No entanto, uma outra hipótese levantada é a escolha dos pontos de controlo terem influencia na precisão. Para ser averiguada tal hipótese, foi criado um novo *shapefile* com 100 pontos de controlo, designados pontos de controlo 2. Estes pontos foram estabelecidos manualmente em que foi dada especial importância à localização de muitos destes nas zonas de vegetação.

#### 4.2.2 Efeito da seleção dos pontos de controlo

Uma vez que as imagens são as mesmas que anteriormente apresentadas, apenas foram recalculados a precisão global e K. O Quadro 4.15 dispõe os resultados referentes às classificações efetuadas através da zona A. A precisão desceu significativamente, o que vem a reforçar a influência das zonas de vegetação nos resultados. Nos resultados obtidos através dos primeiros pontos de controlo, os piores resultados da precisão e de K são 77% e 54% e os melhores 88% e 76%, respetivamente. Pelo que através destes novos pontos de controlo esses valores desceram significativamente, sendo que o pior e melhor resultado da precisão global são 66% e 73% e de K, 32% e 46%, respetivamente.

Quadro 4.15 – Evolução da precisão global e K para todos os treinos feitos na zona A utilizando os Pontos de Controlo 2

Treino	Precisão (%)	K (%)
20	68.00	36.00
50	69.00	38.00
100	73.00	46.00
200	67.00	34.00
300	66.00	32.00
400	68.00	36.00
500	68.00	36.00

Os resultados obtidos através dos primeiros pontos controlo, indicavam que a melhor precisão global atingida era referente ao treino de 50 polígonos, enquanto nesta segunda avaliação, o melhor resultado obtido é alusivo ao treino de 100 polígonos, como é possível constatar quer no Quadro 4.15. Através deste quadro é possível observar, tal como referido anteriormente, que K acompanha o comportamento da precisão global, sendo que quando a precisão aumenta, K aumenta, assim como quando a precisão diminui, K diminui.

Foi feita a mesma análise, comparando os resultados obtidos através dos treinos de 50 e 100 polígonos para as diferentes zonas. Os valores da precisão global e de K encontram-se referenciados no Quadro 4.16.

Quadro 4.16 – Evolução da precisão e de K para os treinos de 20 e 50 polígonos realizados para as zonas A, B e C

	Treino	Precisão (%)	K
<b>A</b>	20	68	0.36
	50	69	0.38
<b>B</b>	20	69	0.38
	50	65	0.30
<b>C</b>	20	72	0.44
	50	71	0.42

É conclusivo que os resultados, quer da precisão global quer de K, não divergem significativamente entre treinos localizados em zonas diferentes da imagem. No entanto, quando utilizados dois conjuntos diferentes de pontos de controlo, é possível verificar que estes têm uma grande influência nos resultados, pelo que colocar pontos de controlo, neste caso, nas zonas de vegetação tende a que o algoritmo classifique erradamente, baixando o valor da precisão. Este facto também mostra que é importante usar um conjunto de pontos de controlo que represente bem a região a ser classificada, de modo a que a precisão global seja bem concordante com a qualidade dos resultados obtidos.

Num breve resumo de todos os resultados apresentados, é possível constatar que a imagem *Landsat* não é adequada para esta análise, uma vez que a sua resolução grosseira não permite representar adequadamente os usos do solo. No que se refere aos algoritmos foi ainda observado que a melhor aprendizagem foi feita pelo *Margin SVM*, uma vez que na maioria dos casos, o algoritmo concorda com todas as classificações feitas pelo utilizador. Porém, é importante referenciar que o cálculo da precisão no GEE e no ArcGis é feito de maneira

diferente. No primeiro o cálculo é feito sobre os resultados da classificação sobre todos os polígonos usados para o treino. No último, a precisão é calculada através de pontos de controlo. Este último método é mais correto, pois permite ter uma base de comparação igual independentemente do número de polígonos, dando a possibilidade de uma melhor comparação entre resultados.

Como é possível perceber, embora o *Fast Naive Bayes* se comporte como o Método da Máxima Verossimilhança para duas classes, os valores são ligeiramente diferentes. Dando destaque aos treinos de 20 polígonos realizados através de ambos os algoritmos, é notória a diferença de 77% para 88.46% para o Método da Máxima Verossimilhança e FNB, respetivamente. Uma justificação plausível é que os algoritmos usam aproximações nos cálculos, daí as pequenas diferenças entre os valores apresentados pelos mesmos, não explicando, contudo, a grande diferença na precisão global obtida para o treino de 20 polígonos.

A precisão global calculada, através de uma classificação efetuada no ArcGis não é significativamente influenciada pela localização do treino mas sim significativamente influenciada pela localização dos pontos de controlo. Já no GEE, a precisão global é afetada pelo tamanho dos polígonos de treino sendo que a sua dimensão deve ser adequada à resolução de imagem com que se está a trabalhar. Ou seja, no caso do *Landsat*, para que o algoritmo consiga tirar informação sobre as classes é necessário que o tamanho do treino seja suficientemente grande para retirar a informações sobre assinatura espectral dos seus píxeis. Enquanto para uma imagem de elevada resolução, é possível adquirir diversas informações através de polígonos pequenos, uma vez que estes cobrem mais que um píxel.

Por fim, foi feita uma análise comparativa das vantagens e desvantagens do ArcGis e do GEE aquando a sua utilização para técnicas de análise de imagens.

### 4.3 Comparação de *software*

Ao trabalhar com dois *software* diferentes foi possível tirar algumas conclusões sobre as potencialidades e desvantagens da utilização de cada um. O Explorer do GEE, por ser uma ferramenta recente, não se encontra plenamente desenvolvida, impossibilitando o acesso a diversas informações que poderiam ser obtidas através das análises, tais como as percentagens de cada tipo de área. Esta informação teria sido útil para comparação com as percentagens de áreas obtidas através do ArcGis. Uma outra lacuna é também a escassa informação oficial sobre a plataforma, que para um iniciante é complicado compreender o funcionamento e o

significado de certos elementos. A falta de informação é também relevante aquando a ocorrência de um problema interno com código 500, ao qual não se encontra qualquer explicação anexa ao erro, não possibilitando identificar e solucionar a causa do mesmo. Este sinaliza que o algoritmo não consegue proceder à classificação embora não explique as causas do problema. Contudo, o facto de existir uma biblioteca de imagens e algoritmos é vantajoso, dando alguma variedade de escolha ao utilizador. No entanto, para análise desenvolvida na presente tese, a maioria das imagens das disponibilizadas gratuitamente no GEE têm uma resolução algo grosseira, sendo que as poucas com boa resolução não se encontram disponíveis para o caso de estudo escolhido. Esta plataforma é de livre acesso permitindo a qualquer acesso à vasta quantidade de fotografias que outrora seriam pagas.

Ao contrário do Google Earth Engine, o ArcGis é um *software* comercial desenvolvido com inúmeras ferramentas de análise. A desvantagem deste programa face ao Google Earth Engine é que não possui uma biblioteca de imagens, pelo que o utilizador terá que dispor previamente de uma imagem para trabalhar. Outro inconveniente deste *software* é a impossibilidade de aceder ao projeto a partir de um outro computador via internet, sendo uma vantagem do GEE, que promove a facilidade de manusear o projeto em qualquer computador com acesso à Internet, necessitando apenas que o utilizador introduza o seu correio eletrónico, na plataforma, para efetuar o *Login*.

Para o estudo efetuado nesta tese, a autora defende que é mais vantajoso recorrer ao ArcGis, quando se dispõe previamente de uma imagem geoespacial. A razão apontada é devido ao ArcGis ser uma ferramenta comercial da qual existe muita informação que permite ao utilizador aprender a trabalhar com a mesma. Para este tipo de análise, as imagens do GEE são demasiado grosseiras pelo que é preferível o utilizador procurar uma imagem de resolução adequada ao seu estudo e proceder ao mesmo no ArcGis.

## 5 CONCLUSÕES E TRABALHOS FUTUROS

### 5.1. Conclusões

Para suprimir as necessidades do Homem, as cidades têm-se expandido de forma rápida e consequentemente os solos que outrora eram permeáveis foram substituídos por zonas impermeáveis. Esta mudança nos usos do solo tem vindo a alterar o ciclo natural da água que proporciona, aquando ocorrência de precipitações, maiores caudais à superfície, aumentando a frequência e magnitude de inundações. Para poder-se prevenir adequadamente estes fenómenos é importante adquirir informações sobre os usos do solo. A deteção remota vem neste sentido auxiliar os especialistas a obter dados sobre a caracterização dos usos do solo, podendo estes ser adquiridos por diversos equipamentos tais como *drones*, satélites, aeronaves tripuladas, entre outros. Nesta tese foram utilizadas fotografias aéreas captadas por satélite e por *drone*, para Lucerna. Com o auxílio do ArcGis e do Google Earth Engine foi efetuada a classificação dos usos do solo, de onde foram retiradas diversas conclusões.

As imagens *Landsat* não são apropriadas para a identificação dos usos do solo para áreas pequenas, uma vez que não permite a classificação adequada das mesmas. Este tipo de fotografia é mais apropriado quando o objetivo é analisar os usos do solo de grandes áreas, tais como de países.

Ao serem treinadas cores muito semelhantes em classes diferentes, o classificador tende a confundir-se, classificando de forma errada alguns usos do solo. Este fator tem implicações no cálculo da precisão global, uma vez que ao classificar de forma indevida, o algoritmo tende a não concordar com o utilizador na definição de alguns pontos de controlo, havendo uma diminuição da precisão. No presente estudo, estas situações foram verificadas para as cores de vegetação. Ao serem treinados os telhados com relvados como impermeáveis, tendo estes cores semelhantes às das zonas de vegetação, o algoritmo classificou essas mesmas áreas como impermeáveis. No entanto, não só o treino de objetos de cores idênticas afeta a precisão assim como a escolha dos pontos de controlo também. Ao comparar-se os dois conjuntos de pontos de controlo escolhidos, verificou-se que ao localizá-los em zonas problemáticas como as áreas de vegetação, a precisão diminuiu significativamente. Tal acontece devido à explicação anterior referente à confusão por parte do algoritmo. No entanto, apesar dos pontos localizados na vegetação contribuírem de forma negativa para os resultados, a escolha dos

pontos de controlo deverá ser sempre representativa, quer em número, quer em cores da imagem, só assim poderá obter-se resultados de precisão confiáveis.

Todos os algoritmos apresentaram resultados de precisão global aceitáveis ou até mesmo bons. No entanto o valor deste parâmetro não se traduz necessariamente na melhor classificação, sendo necessário para cada caso analisar o valor da precisão global e a imagem classificada. O *Margin SVM* mostrou ter uma maior capacidade de aprendizagem e adaptação aos elementos de treinos, tendo este algoritmo tido os melhores resultados de precisão global face aos restantes.

Relativamente à análise de funcionamento dos dois *software* utilizados neste estudo, apesar do Explorer do GEE ser de fácil iteração, existem poucas informações associada às suas componentes. O Explorer do GEE apresenta como desvantagem não possuir ferramentas de cálculo, efetuando apenas análise de imagens, pelo que para a realização do cálculo do parâmetro K e percentagem de áreas permeáveis e impermeáveis será necessário recorrer a outras soluções. Para este problema o ArcGis apresenta-se como uma boa opção na medida que possui inúmeras ferramentas necessárias que permitem fazer diversas análises. No entanto, a desvantagem deste *software* face ao GEE é a lentidão de processamento dos algoritmos, que demoram algum tempo a devolver os resultados. Já nas análises efetuadas no Explorer do GEE, o processamento é feito em segundos.

## 5.2 Trabalhos Futuros

Desde o início da realização deste trabalho foi constatado que a temática de utilização de técnicas de análise de imagem para identificação dos usos do solo é vasta, existindo inúmeras vertentes interessantes de serem exploradas. Como analisado ao longo do Capítulo 4, a confusão do algoritmo na classificação de objetos de cores semelhantes demonstrou ter influência nos resultados. Uma proposta para um trabalho futuro passaria por utilizar o cadastro dos edifícios, representados por polígonos de treino para avaliação o seu efeito na resolução deste problema.

No âmbito da qualidade das imagens, como já foi referido, a fotografia captada por *drone* é de elevada resolução, onde constam objetos pequenos que são irrelevantes para a classificação dos usos do solo, tais como carros, pessoas entre outros. Uma experiência interessante seria proceder à comparação de duas imagens de resolução elevada de uma zona, no entanto uma dessas tivesse uma resolução ligeiramente menor de modo a não aparecerem pequenos

objetos. Assim, perceber-se-ia qual a influência de pequenos objetos no treino, classificação e nas percentagens de áreas.

A classificação dos solos pode subdividir-se em tantas classes quanto as que se queira. Neste trabalho foram utilizadas apenas duas classes, permeável e impermeável, devido serem os tipos de áreas que mais influência têm nos modelos de drenagem urbana. No entanto, poderiam ter sido utilizadas diversas categorias do uso do solo. Uma classificação baseada em diversas classes, como por exemplo, estradas, edifícios, vegetação, entre outras, permitiria perceber quais os usos do solo predominantes em Coimbra. Um trabalho futuro que se baseasse no levantamento destes usos, seria de interesse não só para planeamento do território mas também para auxílio das entidades competentes na prevenção de inundações urbanas.

## REFERÊNCIAS BIBLIOGRÁFICAS

Abreu, A. (2014, 15 de Fevereiro). *Drones para todo o serviço*. Revista Expresso, pp 50-54.

Butler, D. e Davies, J. W. (2011). “Urban Drainage”. Spon Press, Londres e Nova Iorque.

Elkhrachy, I., (2015). “Flash Flood Hazard Mapping Using Satellite Images and GIS Tools: A case study of Najran City, Kingdom of Saudi Arabia (KSA)”. *Egyptian Journal of Remote Sensing and Space Science*, Vol. 18, Issue 2, pp. 261-278.

Deletic, A., Dotto, C. B. S., McCarthy, D. T., Kleidorfer, M., Freni, G., Mannina, G., Uhl, M., Henrichs, M., Fletcher, T. D., Rauch, W., Bertrand-Krajewski, J. L., Tait, S. (2011). *Assessing uncertainties in urban drainage models*. Physics and Chemistry of the Earth.

Fonseca, A., Fernandes, J. (2004). “Detecção Remota”. Lidel, Lisboa.

Henriques, A. G. (1985). “Avaliação dos recursos hídricos de Portugal Continental: contribuição para o ordenamento do território”. IED, Lisboa.

Hipólito, J. R., Vaz, A. C. (2011). “Hidrologia e Recursos Hídricos”. Instituto Superior Técnico Press, Lisboa.

INE. [http://censos.ine.pt/xportal/xmain?xpid=CENSOS&xpgid=censos\\_quadros](http://censos.ine.pt/xportal/xmain?xpid=CENSOS&xpgid=censos_quadros). Censos de 2001. Instituto Nacional de Estatística.

James, G., Witten, D., Hastie, T., Tibshirani, R. (2015) “An Introduction to Statistical Learning with Applications in R” Springer, Nova Iorque.

Kaiser, E. J., Godschalk, D. R., Chapin, E. S. (1995). “Urban land use planning”. University of Illinois, Chicago.

Küng, O., Strecha, C., Beyeler, A., Zufferey, J. C., Floreano, D., Fua, P., Gervais, F. “The accuracy of automatic photogrammetric techniques on ultra-light UAV imagery.” <http://www.int-arch-photogramm-remote-sens-spatial-inf-sci.net/XXXVIII-1-C22/125/2011/>. Último acesso: 17 de Março de 2016.

Maksimovic C., Prodanovic D., Boonya-Aroonnet S., Leitão J. P., Djordjevic S., Allitt R. (2009). “Overland flow and pathway analysis for modelling of urban pluvial flooding”. *Journal of Hydraulic Research*, Vol. 47, pp: 512-523.

Mesev, V. (2007) “Integration of Gis and Remote Sensing”. Wiley. Inglaterra.

Leitão, J., Matos, J. S., Prodanovic´, Marksimovic´, C. (2008). “Modelos de simulação do escoamento superficial em meio urbano - potencialidades e limitações”. 13º Encontro de Saneamento Básico, Covilhã.

Leitão, J. P., Moy de Vitry, M., Scheidegger, A., Rieckermann, J. (2015). Assessing the quality of Digital Elevation Models obtained from mini-Unmanned Aerial Vehicles for overland flow modelling in urban areas. *Hydrology and Earth System Sciences Discussions*, Vol. 12, Issue 6, pp. 5629–5670.

Lencastre, A., Franco, F. M. (1992). “Lições de Hidrologia”. Serviços editoriais da Universidade Nova de Lisboa, Almada.

Lillesand, T. H., Kiefer, R. W. (1987). “Remote Sensing and Image Classification”. Wiley, Nova Iorque.

Lima, J. P., Martins, C. M. T., Mendes, M. G. T., Abreu, J. M., Almeida, J. P. L., Lima, I. P. (2013). “Hidrologia Urbana: Sistemas de drenagem de águas pluviais urbanas.” ERSAR e Universidade de Coimbra. Lisboa.

Lima, R. P., Abrantes, J. R.C.B., Lima, L. M. P., Lima, M. I. P. (2015). “Using thermal tracers to estimate flow velocities of shallow flows: Laboratory and field experiments.” *Journal of Hydrology and Hydromechanics*, Vol 3, pp. 255–262.

Sabins, F. F. (1997). “Remote Sensing. Principles and Interpretation”. W. H. Freeman. Nova Iorque.

Marques, J. A. A. S., Sousa, J. J. O. (2011). “Hidráulica urbana: Sistemas de abastecimento de água e de drenagem de águas residuais. Imprensa da Universidade de Coimbra, 3ª ed., Coimbra.

Price, R. K., Vojinovic´, Z. (2011). “Urban Hydroinformatics: Data, Models and Decision Support for Integrated Urban Water Management”. IWA, Londres.

Sá, J. P. M. (2001). "Pattern Recognition: Concepts, Methods and Applications". Springer, Porto.

Simões, N. E. (2012). "Urban Pluvial Flood Forecasting". Tese de Doutorado, Imperial College London. Londres

Sousa Pinto, N. L, Holtz, A. C. T., Martins, J. A., Gomide, F. L. S. (1990). Hidrologia Básica. Editora Edgard Blücher Ltda. São Paulo, Brasil.

Tokarczyk, P., Leitão, J. P., Rieckermann, J., Schindler, K., Blumensaat, F. (2015). "High-quality observation of surface imperviousness for urban runoff modelling using UAV imagery". Hydrology and Earth System Sciences Discussions, 19(10), 4215-4228.

Viera, A. J., Garrett, J. M. (2005). "Understanding Interobserver Agreement: The Kappa Statistic". Family Medicine Vol. 37, Issue 5, pp. 360-363.

## ANEXOS

### Anexo A: Localização da bacia central de Coimbra e respetivo *Landsat*



Figura A.1 – Bacia central da cidade Coimbra (Imagem gentilmente cedida pelo Prof. Doutor Nuno Simões)

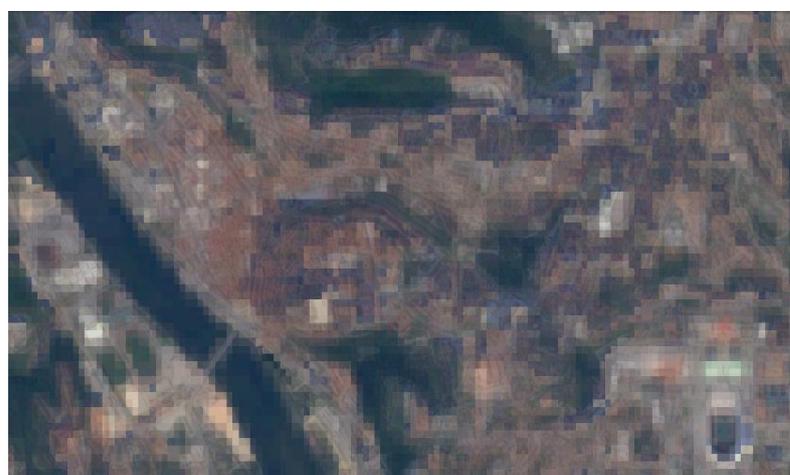


Figura A.2 – Imagem *Landsat*, Coimbra (GEE)

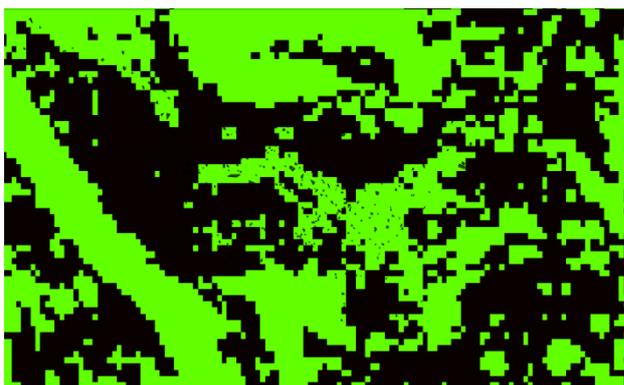
## **Anexo B:** Zona de treino A



Figura B.1 – Zona de Treino A

## **Anexo C:** Principais resultados para a classificação da imagem *Landsat* correspondente a Coimbra

Neste anexo são apresentados apenas os melhores resultados obtidos com cada algoritmo para a classificação dos usos do solo da bacia central de Coimbra. Nas Figuras C.1 e C.3 estão representadas as classificações obtidas com o algoritmo FNB e *Margin SVM*, respetivamente. De um modo geral, estas imagens representam bem a realidade da bacia, ainda que as mesmas tenham algumas diferenças entre si. A precisão global em ambas é boa.



### Fast Naive Bayes

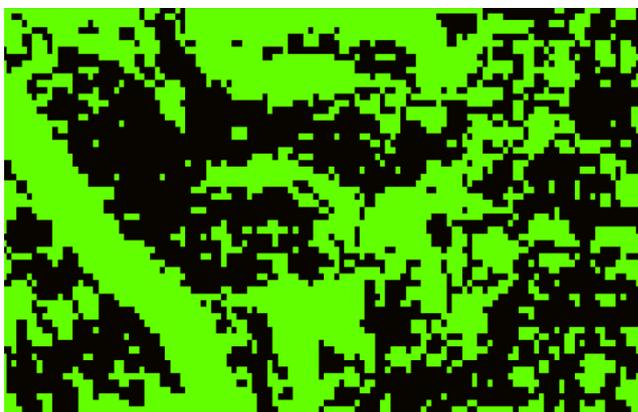
Model, trained Jul 2, 2016 at 8:33pm

Overall validity: 93.33%

	# Points	Impermeável	Permeável
■ Impermeável	8	100%	0%
■ Permeável	7	14.29%	85.71%

Figura C.1 – Classificação obtida para o treino de 300 polígonos através do algoritmo FNB

Figura C.2 - Matriz de erro e precisão global: treino de 300 polígonos realizado com o FNB



### Margin SVM

Model, trained Jul 3, 2016 at 10:40am

Overall validity: 100%

	# Points	Impermeável	Permeável
■ Impermeável	5	100%	0%
■ Permeável	7	0%	100%

Figura C.3 - Classificação obtida para o treino de 400 polígonos através do algoritmo *Margin SVM*

Figura C.4 - Matriz de erro e precisão global: treino de 400 polígonos realizado com o *Margin SVM*