João Eduardo Albuquerque Martins Pereira Pires

Licenciado em Ciências da Engenharia Eletrotécnica e de Computadores

Machine Learning Aplicado a Imagens da Observação da Terra para o Estudo da Floresta

Dissertação para obtenção do Grau de Mestre em
Engenharia Eletrotécnica e de Computadores

Orientador: André Damas Mora

Professor Auxiliar, Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa

Júri:

Presidente: João Goes

Professor Catedrático, Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa

Arguente: José Manuel Fonseca

Professor Associado com Agregação, Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa

FACULDADE DE CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

Março, 2019

Machine Learning Aplicada a Imagens da Observação da Terra para o Estudo da Floresta

A todos os que me acompanharam nesta luta.

Agradecimentos

Começo por agradecer todo o empenho e apoio do meu orientador, Professor André Mora. A disponibilidade praticamente absoluta, a paciência para interrupções frequentes do seu trabalho para me guiar ao longo deste estudo. A boa disposição e simpatia com que sempre me tratou. A oportunidade que me deu ao integrar-me num projeto e a integração também no grupo de investigação CA3. Agradeço também a todo este grupo pela forma como me acolheu, e por sempre me convidar a participar nas reuniões relativas aos projetos de observação da Terra.

À Faculdade pela formação que me deu enquanto futuro engenheiro e por todas as oportunidades que me ofereceu enquanto aluno. Apesar de me ter cruzado com inúmeras pessoas que marcaram esse percurso, dirijo um agradecimento muito especial ao antigo Sr. Diretor, Professor Fernando Santana, que, apesar de não ter tido o privilégio de conhecer pessoalmente, muito me apoiou, de tal modo que sem a sua intervenção provavelmente hoje não estaria a finalizar a minha dissertação. Quero destacar também o Professor Raul Rato pela confiança que depositou em mim ao integrar-me num dos seus projetos de investigação. Por todo o conhecimento que me transmitiu tanto a nível intelectual como de experiência de vida. E, o principal, pela curiosidade que despertou em mim, levando-me à busca constante do saber. Agradeço à Fundação de Ciências e Tecnologia pelo apoio ao projeto FUELMON (PTDC / CCI-COM / 30344 / 2017).

À minha família, à minha mãe, Auristela, e ao meu pai, Eduardo, que nunca deixaram de acreditar e investir em mim quando normalmente ninguém o faria. Porque tomar as decisões corretas quando a vida segue o seu curso normal é fácil, quando surgem problemas sérios, a inteligência necessária para tomar a decisão mais benéfica a longo prazo revela-se uma tarefa extremamente difícil. À minha irmã, Mafalda, que sempre esteve do meu lado, sempre me apoiou e acompanhou em toda a minha vida. Por pensar comigo cada problema que surgia, procurando ambos sempre a solução mais adequada. Agradeço à minha namorada Zahara, pela paciência que demonstra todos os dias para lidar com a minha vida inesperada. Por escutar todas as minhas ideias, mesmo quando se tal se torna repetitivo. Por saber sempre como me ajudar quando surgem situações mais complicadas. Agradeço aos meus avós Helena, Silvestre, Lucília e Manuel pela crença absoluta que têm em mim, pelo apoio, e no caso da minha avó Helena que me acompanhou em tantas noites de estudo apenas para não estar sozinho, e o meu avô Silvestre que tem sempre uma habilidade especial para resolver aqueles problemas inesperados de última hora. Agradeço ao meu tio José e à minha madrinha Beatriz. Ao meu tio que tanto peso teve na minha decisão em seguir este caminho académico. À minha madrinha pela boa disposição que sempre me trouxe, mesmo nos momentos mais difíceis. Agradeço ao meu “tio” Jorge que para além de estar sempre presente nos momentos mais importantes, sempre teve as palavras certas, nos momentos certos, para comigo. Por fim, agradeço ao Sr. Arcebispo D. Zacarias Kamuwenho também pela amizade, e por toda a importância que teve tanto na minha vida, como no meu percurso académico.

À Srª. Diretora do Colégio Moderno, Dr.ª Isabel Soares. Se às vezes é difícil encontrar apoio na família, mais difícil é encontrá-lo fora. A crença em mim da Srª. Diretora, provavelmente maior do que a crença que eu tinha em mim mesmo, foi fundamental neste percurso que, na verdade, começou muito antes de eu começar o meu curso superior. Aos meus colegas de secundário, os quais destaco: Tiago Pequito, Tiago Martins, João Casteleira, Pedro Pinto, Manuel Gamito, Ricardo Gomes, Inês Araújo e Joana Henriques. Nestes cinco anos de curso tive ainda o privilégio de conhecer também inúmeras pessoas sendo as mais presentes André Torres, João Clemente, António Simões, Diogo Silva, Sara Martins, Elmarlon Pontes, David Pereira, Renato Ribeiro, António Borges e Conceição Teixeira. No estudo, nos trabalhos, juntos na mesma luta. Por fim, agradeço aos meus Professores do ensino secundário, Sr. Dr. Jorge Rio Cardoso e Sr. Dr. António Carlos Cortez. As suas aulas, as nossas conversas depois destas, o seu apoio, foram muito importantes para o meu crescimento e para a vontade de querer em todos os dias evoluir enquanto pessoa e enquanto profissional.

Resumo

A crescente observação da Terra por via de satélites tem permitido um estudo sobre o planeta mais completo e abrangente. A aquisição de imagens nos diferentes comprimentos de onda do espectro eletromagnético permite a extração de informação da superfície terrestre sem a necessidade de instalação de sensores, simplificando o processo de aquisição de dados. O aumento exponencial de dados disponíveis torna praticamente impossível a sua análise de forma humana, implicando o recurso a técnicas de classificação cada vez mais evoluídas.

O Instituto de Conservação da Natureza e das Florestas (ICNF) definiu como estratégia para prevenção e combate a incêndios florestais a implementação da rede primária de Faixas de Gestão de Combustível (FGC), nas quais é feita uma redução da biomassa criando uma barreira à progressão do fogo. Uma das responsabilidades do ICNF é garantir a sua manutenção, no entanto devido à extensão da rede necessita de uma ferramenta baseada em Sistemas de Informação Geográfica para detetar de forma automática intervenções nas FGC.

Nesta dissertação é proposto um conjunto de técnicas baseadas em imagens de observação da terra para identificar operações de manutenção nas FGC. A análise recorreu a imagens multiespectrais adquiridas pelo satélite *Sentinel 2*. Como pré-processamento destas imagens foi necessário corrigir o erro de georreferenciação e obter índices espectrais para extrair mais informação. Em seguida, foi estudado o comportamento destas FGC relativamente a uma intervenção comparando com a dinâmica em zonas de vegetação. Após o entendimento das dinâmicas espectrais utilizaram-se Redes Neuronais Artificiais para a deteção das intervenções. Antes da classificação foi feito um estudo para a escolha dos atributos a utilizar, com base na análise de dados referida no parágrafo anterior. Depois do treino do algoritmo e definição dos parâmetros foi testado numa região diferente tendo sido identificada corretamente a intervenção.

Palavras-chave: Observação da Terra; Faixas de Gestão de Combustível; Imagens Multiespectrais; Redes Neuronais; Sentinel 2.

Abstract

The increasing Earth observation using satellites allowed a study of the planet completer and more comprehensive. The image acquisition in the different wavelengths of the electromagnetic spectrum allows information extraction from the Earth's surface without the need of sensors simplifying the process of data acquisition. The exponential increase of available data makes almost impossible its analyses by humans, implying the use of classification techniques increasingly evolved.

The Institute for Conservation of Nature and Forest (ICNF, Portugal) defined as a strategy for prevention and forest fire fighting the implementation of the primary network of Fuel Management Bands (FGC), in which a reduction of the bio-mass is made creating a barrier to the progression of the fire. One of the responsibilities of the ICNF is to ensure its maintenance, however due to the extension of the network, it needs a tool based on Geographic Information Systems to automatically detect interventions in the FGC.

In this dissertation it is proposed a set of techniques based on Earth observation images to identify maintenance operation in the FGC. This analysis used multispectral images acquired with the satellite Sentinel 2. As a preprocessing of the images it is necessary to correct the georeferencing error and obtain the spectral indexes to extract more information. Then, the behavior of these FGC was studied in relation to an intervention comparing with the dynamics in vegetation zones. After the understanding of the spectral dynamics, Artificial Neural Networks were used to detect the interventions. Before the classification was made a study to choose the attributes to be used, based on the data analysis referred to in the previous paragraph. After the training of the algorithm and parameter definition was tested in a different region and the cutting operation was correctly identified.

Keywords: Earth Observation; Electromagnetic Spectrum; Machine Learning; Fuel Management Bands; Fires; Information; Sentinel 2.

Índice

Lista de Figuras

Lista de Tabelas

Lista de Acrónimos

ANN	Redes Neuronais Artificias (*Artificial Neural Networks*)
BAP	*Best Available Pixel*
CEOS	*Committee on Earth Observation Satellites*
CIVE	*Color Index of Vegetation Extraction*
CNES	*Centre National d'Etudes Spatiales*
CNN	*Convolutional Neural Networks*
CSV	*Comma Separated Values*
DT	Árvore de Decisão (Decision Tree)
DVI	*Difference Vegetation Index*
ETM+	*Enchanced Thematic Mapper Plus*
ESA	*European Space Agency*
EVI	*Enhanced Vegetation Index*
ExG	*Excess of Green*
ExGR	*Excess of Green minus Excess of Red*
ExR	*Excess of Red*
FCN	*Fully Convolutional Network*
FGC	Faixa de Gestão de Combustível
FIC	Faixa de Interrupção de Combustível
FIF	*Fuzzy Information Fusion*
FIS	Sstema de Inferência Difusa (*Fuzzy Inference System*)
FRC	Faixa de Redução de Combustível
GE	*Google Earth*
GRI	*Global Reference Image*
ICNF	Instituto de Conservação da Natureza e das Florestas
L1GS	*Level 1 Systematic Correction*

L1TP	*Level 1 Standard Terrain Correction*
MExG	*Modified Excess Green Index*
MNDWI	*Modified* NDWI
MODIS	*Moderate Resolution Imaging Spectroradiometer*
MSS	*Multispectral Scanner*
NASA	*National Aeronautics and Space Administration*
NDI	*Normalized Difference Index*
NDMI	*Normalized Difference Moisture Index*
NDWI	*Normalized Difference Water Index*
NDVI	*Normalized Difference Vegetation Index*
NMDI	*Normalized Multi-band Drought Index*
NOAA	*National Oceanic an Atmospheric Administration*
OLI	*Operation Land Imager*
QGIS	*Quantum GIS*
RFC	*Random Forest Classification*
RPFGC	Rede Primária de Faixas de Gestão de Combustível
RVI	*Ratio Vegetation Index*
TIRS	*Thermal Infrared Sensor*
TOA	Topo da Atmosfera (*Top-of-Atmosphere*)
SIG	Sistema de Informação Geográfica
SVM	*Support Vector Machines*

1 Introdução

1.1 Motivação

Todos os anos, os verões portugueses são caracterizados pelos incêndios. Ainda estão presentes na memória os incêndios de Pedrogão Grande em 2017 e (felizmente sem vítimas mortais) em Monchique o ano passado. Os valores de área ardida em relação à média europeia são bastante elevados. Observe-se a Figura 1.1 em que são comparados os valores absolutos de área ardida de Portugal como a média da União Europeia (UE) no período entre 2007 e 2017 (os seguintes países não são contabilizados por falta de dados: Bélgica, Dinamarca, Irlanda, Luxemburgo, Malta, Países Baixos, Reino Unido). Para além disso, verifica-se que no mesmo período Portugal é o país com maior valor médio de área ardida e de nº de incêndios, tendo valores bastante superiores à média da UE, veja-se a Figura 1.2.

Nas figuras referidas no parágrafo anterior apenas são considerados incêndios florestais. Estes definem-se como uma combustão limitada no espaço e no tempo afetando uma área florestal (segundo o Instituto Nacional de Estatística). Importa realçar que na Figura 1.2 observam-se valores absolutos, Portugal encontra-se acima de países com uma área bastante superior. Claro que o território nacional é bastante mais propício aos incêndios florestais que os restantes países em análise, nomeadamente a Espanha. Conhecendo estes dados torna-se imperativo a procura de soluções. Nas estratégias de combate aos incêndios, no decurso dos mesmos, o trabalho dos bombeiros é essencial. Contudo, uma atitude proativa (com vista à prevenção e preparação do combate) é tão ou mais importante do que os mecanismos reativos.

Neste âmbito o Instituto da Conservação da Natureza e das Florestas (ICNF) estabeleceu um conjunto de medidas preventivas, destacando-se as Faixas de Gestão de Combustível (FGC), que serão abordadas no próximo parágrafo. Após a implementação das FGC, estas requerem manutenção e vigilância constante do seu estado de conservação para garantir a sua eficiência. O crescente número de satélites de Observação da Terra (ver 2.3) e a disponibilidade

destas observações de forma gratuita e com uma periodicidade cada vez maior permite o desenvolvimento de técnicas de vigilância proativa em relação aos incêndios.

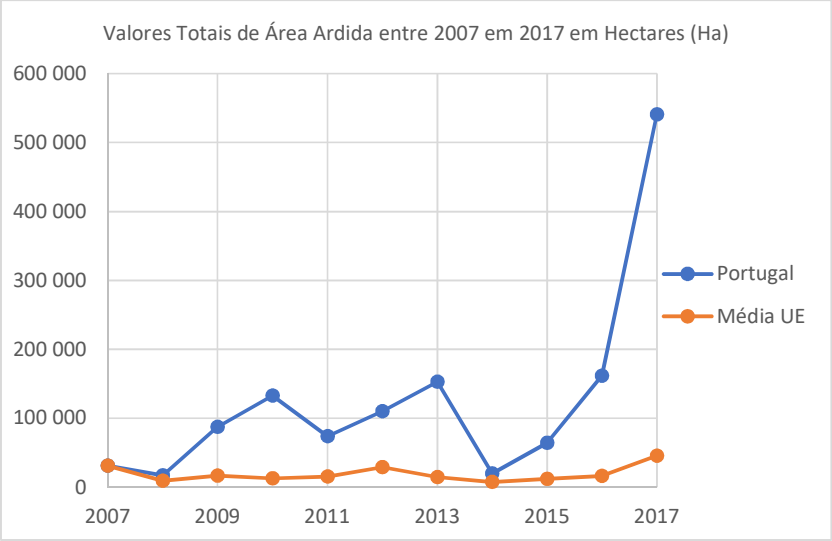


Figura 1.1: Valores Totais de Área Ardida entre 2007 e 2017 (fonte: PORDATA).

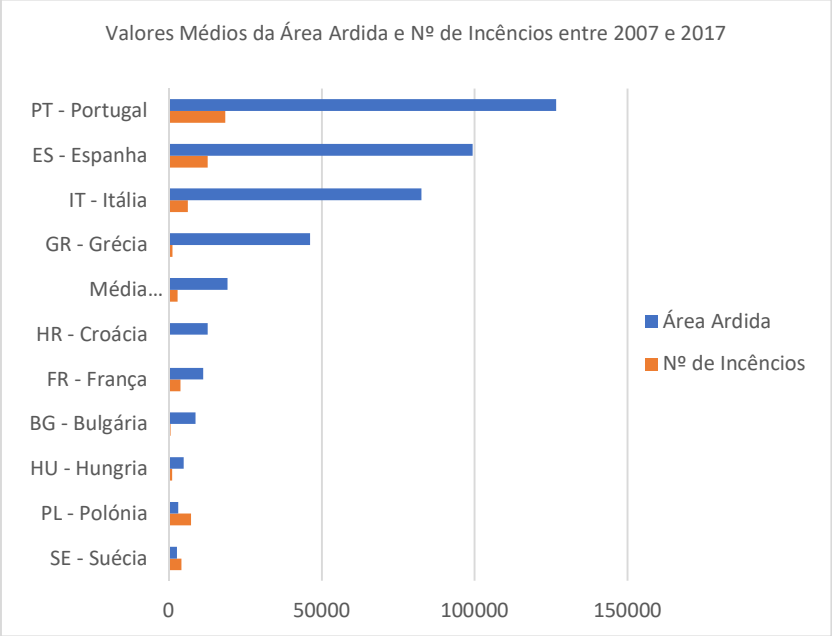


Figura 1.2: Valores Médios da Área Ardida e Nº de Incêndios entre 2007 e 2017 (fonte: PORDATA).

1.2 Faixas de Gestão de Combustível

No parágrafo anterior referiu-se a importância de medidas de prevenção para um combate aos incêndios mais eficiente. Neste plano o ICNF definiu como uma das estratégias a Rede Primária de Faixas de Gestão de Combustível (RPFGC) que pode ser consultada em (DPFVAP, 2014) A RPFGC consiste então no conjunto de todas a FGC. Até à data apenas 14% da RPFGC se encontra implementada (1600 km de 11 125 km).

As FGC têm como finalidade um maior controlo dos fogos durante o seu combate, diminuindo a área percorrida, reduzindo os efeitos de passagem dos incêndios (protegendo vias de comunicação, infraestruturas, povoamentos florestais, entre outros) e isolando possíveis pontos de ignição. Também permitirão às equipas de combate adotar táticas mais adequadas, o reabastecimento mais simples das viaturas e a melhoria das condições de posicionamento dos bombeiros. Para atingir estes objetivos a RPFGC foca-se na quebra da continuidade horizontal (distanciamento entre árvores) e vertical (desramação das árvores) dos combustíveis, moldando assim o comportamento de um incêndio através das FGC.

Tecnicamente, as FGC deverão ter uma largura mínima de 125 m, sendo constituídas por Faixas de Redução de Combustível (FRC) e Faixas de Interrupção de Combustível (FIC). Numa FRC existem dois tipos de áreas, uma em que a distância entre copas deve ser de 2 metros, assegurando a descontinuidade horizontal do estrato arbóreo seguida de uma área menos densa em que o espaçamento entre copas deve ser no mínimo de 4 metros. A desramação deve alcançar um mínimo de 4 metros de altura. Também a altura da vegetação deve ser regulada consoante a taxa de cobertura do solo, sendo inversamente proporcionais. Uma FIC deve encontrar-se junto à rede viária fundamental (tendo esta um mínimo de 5 metros) verificando uma largura de 10 metros para cada lado da via. A ilustração destas especificações pode ser vista na Figura 1.3.

O ICNF para o planeamento de aplicação das FGC deve observar vários aspetos. A RPFGC destina-se a incêndios de grandes dimensões e tem de ser feita de modo a oferecer melhores condições de segurança durante o combate aos incêndios. Regiões com um risco meteorológico elevado e/ou histórico de incêndios e regiões com valores socioeconómicos, paisagísticos e ecológicos devem também ser consideradas. Também devem ser instaladas em zonas nas quais se possam promover atividades que ajudem na sua sustentabilidade técnica e financeira. Deste modo, o ICNF para o planeamento de aplicação das FGC definiu uma estratégia de priorização dos terrenos. O cálculo do nível de prioridade de uma área depende dos seguintes fatores: valor económico a proteger; valor ecológico; valor da propriedade e perigosidade. Este cálculo deve ser feito num Sistema de Informação Geográfica (SIG, descritos em 2.4) resultando dele três classes de prioridade: Elevada, Média e Não Priorizada.

As FGC podem assumir um papel essencial na prevenção e também no combate aos incêndios. Contudo um dos grandes problemas associados às mesmas é a necessidade de manutenção regular e conhecimento do seu estado de conservação. Para tal, o ICNF, através da contratação de terceiros, executa operações de instalação e manutenção na RPFGC. A verificação de necessidade de manutenção e da correção das intervenções é feita através da observação no terreno e de imagens de satélite. Este processo, apesar de atingir os resultados

pretendidos, é demorado e, no caso das deslocações ao terreno, dispendioso. Claro que enquanto apenas se encontrar implementada 14% da RPFGC estes fatores negativos não serão muito sentidos, até porque as intervenções tendencialmente são programadas para diferentes datas. Contudo, se o plano se mantiver, no futuro terão de ser analisados 11 125 km. Na pior das hipóteses, a programação de intervenções para a totalidade do plano da RPFGC numa mesma data. Neste caso o processo de verificação seria extremamente dispendioso e/ou demorado.

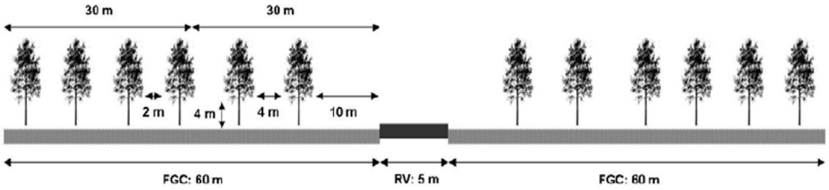


Figura 1.3: Secção transversal das FCG (DPFVAP, 2014)**.**

1.3 Sumário

A necessidade de uma organização territorial eficaz é um fator fulcral no desenvolvimento de uma nação. Um dos exemplos mais gritantes comprovativo desta premissa foi a civilização egípcia que graças às enchentes do Rio Nilo e ao terreno de que dispunha ser extremamente fértil, foi uma nação dominante durante um certo período da História. Nesta dissertação não se procura a organização do território para a produção agrícola, mas para prevenção de catástrofes. As guerras entre a natureza e a espécie humana têm sempre o mesmo vencedor. Lembre-se por exemplo o Sismo de 1755 em Lisboa, o Tsunami em 2004 na Indonésia, ou, novamente, o incêndio de Pedrogão em 2017.

O desenvolvimento técnico e científico do ser humano permite o desenvolvimento de métodos para a prevenção destas tragédias, não devendo apenas ser aplicado numa perspetiva guiada pelo instinto capitalista que apenas procura o lucro económico desprezando e desrespeitando tudo o resto. Muitas vezes as estratégias de proteção até já existem e até começaram por ser implementadas, mas pecam na sua fiscalização e manutenção regular. Felizmente hoje em dia as ferramentas de observação da Terra para além de serem cada vez mais sofisticadas, oferecendo cada vez mais informação, são também cada vez mais acessíveis (existindo muita informação gratuita). A par desta realidade também a evolução da aprendizagem automática e a existência de cada vez mais técnicas de extração de conhecimento de bancos de dados permite uma automatização mais eficiente dos processos de decisão com base em imagem. Também, como será visto em 2.4, os SIG possibilitam uma maior precisão nos dados obtidos pelos satélites.

1.4 Objetivos

Esta dissertação contribui para a automatização da monitorização das FGC, com recurso aos dados disponibilizados pelos programas de observação da Terra e à aprendizagem automática. Com a metodologia desenvolvida será possível identificar remotamente se uma operação de manutenção foi efetuada. Deste modo pretende-se automatizar o processo utilizado pelo ICNF (descrito no parágrafo anterior). Claro que a verificação humana será, tendencialmente, mais fiável que a verificação computacional, portanto pretende-se que sejam assinaladas todas as situações em que existam dúvidas na qualidade da manutenção adequada, sendo apenas estas as analisadas posteriormente com mais atenção. Procura-se agilizar o trabalho que já existe, não o substituir.

O primeiro objetivo é a definição das áreas de estudo, sendo preferencialmente escolhida uma área de interesse do ICNF e depois uma área próxima de forma a poder ser observada presencialmente. Depois desta escolha será definido o período de análise. Este estará dependente da disponibilidade de dados do satélite.

Após definida a área e período de estudo serão obtidos os dados que vão ser analisados. Começando pela obtenção de: observações relativas às áreas de estudo; vetores relativos às FGC; vetores relativos a zonas de vegetação.

Quando se iniciar a extração de dados serão desenvolvidas as ferramentas necessárias, compatíveis com o QGIS e com recurso ao *Python*, para automatização e otimização dos processos de análise. Pretende-se que o carregamento no visualizador das imagens seja automático juntamente com o recorte das zonas em análise e obtenção dos valores medidos pelas observações do satélite. Durante este processo é necessário ter uma etapa de georreferenciação da imagem face aos erros associados às imagens de observação da Terra. Note-se que devido à pequena área que representa uma FGC, um pequeno erro de geolocalização leva à perda de informação sobre esta. Para além disso introduz nas medições dinâmicas diferentes (dado que pode ser considerada num FGC alguns pixéis relativos a vegetação) podendo criar confusões no processo de deteção de mudanças.

Por fim, serão aplicadas várias técnicas de *Machine Learning*, desde a seleção de atributos, a classificação e a validação de resultados. Estes métodos vão permitir a deteção automática de uma intervenção numa FGC. Sintetizando o plano de trabalhos consiste em:

- Implementar ferramentas para a extração de dados partir das imagens de Observação da Terra;
- Estudo das bandas e dos índices para deteção de alterações no terreno;
- Correção de erros de georreferenciação das imagens;
- Seleção de atributos para algoritmos de *Machine Learning*;
- Deteção automática de intervenções nas FGCs.

1.5 Organização do Documento

Esta dissertação, para além do presente, é constituída por mais quatro capítulos, dos quais se segue uma breve descrição:

- ***Capítulo 2***: São apresentados os conceitos teóricos fundamentais para um entendimento global desta dissertação. O Remote Sensing é definido, bem como apresentado o Espectro Electromagnético e a sua importância. Seguidamente são apresentadas as diversas fontes de dados para o desenvolvimento na área de *Remote Sensing*, dando mais ênfase ao programa *Sentinel*. Termina-se o capítulo com a apresentação dos Sistemas de Informação Geográfica;
- ***Capítulo 3***: Realizou-se uma revisão da bibliografia atual, sendo apresentados os métodos tanto de extração de informação dos dados, como de extração de conhecimento aplicados nesta temática. São abordados os Índices Espectrais, o Registo de imagem e a Aprendizagem automática, estando incluidas todas as técnicas utilizadas no desenvolvimento deste trabalho;
- ***Capítulo 4***: Mostra todo o trabalho desenvolvido, a sequência de tarefas desde a obtenção dos dados até à deteção de uma intervenção, abordagem das etapas de pré-processamento incluindo a georrefrenciação das imagens;
- ***Capítulo 5***: Apresentação dos resultados da análise de dados, escolha de atributos e classificação. Justificação de todas as decisões tomadas;
- ***Capítulo 6:*** Este capítulo final é dedicado às conclusões obtidas nesta dissertação e revela também qual o caminho que será seguido na sequência deste estudo.

2 Enquadramento Teórico

Neste capítulo serão abordados os aspetos teóricos considerados mais importantes para o desenvolvimento deste estudo. Inicialmente será definido o conceito de *Remote Sensing*, a temática em que se enquadra esta dissertação. Em 2.2 e 2.3 serão abordados os tipos de dados que serão obtidos, a forma como é feita a sua aquisição (programas de observação da Terra que existem) e a informação que pode ser obtida. Por fim, em 2.4 são introduzidas as ferramentas que serão utilizadas no desenvolvimento desta dissertação.

2.1 O que é *Remote Sensing*?

2.1.1 Definição ESA

Segundo a *European Space Agency* (ESA), Remote Sensing consiste na aquisição e análise de dados daquilo que se pretende estudar, sem necessidade de um contacto direto. Este processo implica três elementos fundamentais:

- **Plataforma:** Onde se encontra o instrumento de aquisição de dados;
- **Objeto:** O que é observado;
- **Sensor:** Dispositivo que trata das medições.

2.1.2 Definição NOAA

O *National Oceanic an Atmospheric Administration* (NOAA) também oferece uma definição de *Remote Sensing*. O NOAA diz que se trata da ciência de obtenção de informação sobre objetos à distância, tipicamente através de satélites e aviões. O elemento fulcral são os sensores remotos que tratam da recolha de dados pela deteção de energia refletida ou emitida da Terra. Existem então dois tipos de sensores:

- **Passivos:** Respondem a impulsos externos obtendo a energia natural emitida ou refletida da superfície terrestre.

- **Ativos:** Utilização de estímulos internos para a recolha de dados. Por exemplo, a projecção de um laser para a superfície terrestre e medição do tempo de viagem do sinal.

Observando estas duas definições de duas diferentes organizações, pode ser concluído que *Remote Sensing* trata-se da ciência de estudo de um objeto remotamente recolhendo e analisando os dados através de um sensor. Estas técnicas são utilizadas para aplicações costeiras, de oceano, avaliação de risco e gestão de recursos naturais.

Nesta dissertação em que o alvo do estudo é a superfície terrestre a plataforma utilizada será o *Sentinel 2* e o respetivo sensor (abordado em 2.3.2). Apesar de não serem utilizados dados do *Landsat 8* é dada a devida relevância a este programa de observação da Terra. Para além da sua importância neste campo de estudos, também existe a possibilidade de utilização destes dados na continuidade deste trabalho (diminuindo, eventualmente, períodos mais alargados sem imagens do *Sentinel 2* devido às condições atmosféricas). Porém, a sua periodicidade de 16 dias, quando comparada aos 5 dias do *Sentinel 2* (que se deve ao facto de se tratar de uma constelação de dois satélites e o *Landsat 8* apenas utilizar um satélite) e à resolução espacial bastante superior (no *Landsat 8* as melhores bandas têm uma resolução espacial de 30 metros[1], no *Sentinel 2* têm 10 metros de resolução, ver Figura 2.1) leva a que neste estudo o *Landsat 8* seja preterido. Para futuro também será importante considerar o *Landsat 9*, cujo satélite entrará em orbita em dezembro de 2020.

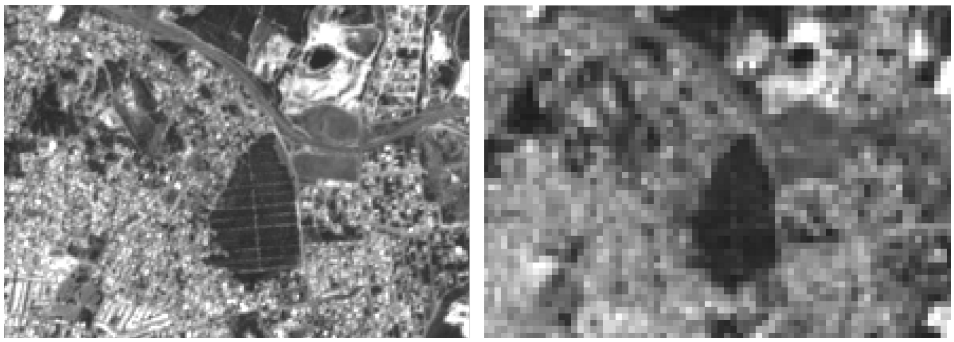


Figura 2.1: Comparação da resolução das imagens do Sentinel 2 (esquerda, 03/08/2017) com Landsat 8 (direita, 02/08/2017) ambas imagens da banda 4 (Vermelho) na região da Marisol.

2.2 Espectro Eletromagnético

A compreensão do Espectro Eletromagnético (daqui em diante referido apenas como Espectro) é essencial. Ao longo desta dissertação serão utilizadas Imagens Multiespectrais. Esta classe de imagens consiste na captura destas nas diversas regiões do Espectro. A exploração

[1] Note-se que a banda pancromática tem uma resolução espacial de 15 metros. Isto permite que aplicando técnicas de *pansharpening* aumente a resolução espacial de outras bandas.

das diferentes zonas deste permite obter um maior conhecimento através da imagem. Este género de dados tem desde aplicações militares a aplicações na saúde e claro, são bastante utilizados e importantes no *Remote Sensing*.

Em 1672, Isaac Newton apresenta a Nova Teoria sobre a Luz e as Cores. Nesta publicação a Luz foi então definida como um conjunto de raios difusos. Estas radiações foram associadas às diferentes cores observadas pelo olho humano. Contudo, o Espectro desta época estava incompleto, apenas era conhecida a sua zona visível (do Violeta ao Vermelho). Apenas em 1800 William Herschel publica as suas descobertas sobre os raios invisíveis do Sol. Herschel conduziu várias experiências com o propósito de calcular a temperatura das diferentes cores do Espectro conhecido. Porém verificou que depois do vermelho também existia radiação, denominando-a como radiação calorífica, atualmente a radiação infravermelha. Um ano depois Johann Ritter, ao tomar conhecimento das descobertas de Herschel, observou o que sucedia além do limite visível do violeta, descobrindo a radiação ultravioleta. Com estas três descobertas é definido então o conjunto de todas as radiações eletromagnéticas que constituem o Espectro.

A radiação eletromagnética consiste numa corrente de fotões que viajam à velocidade da luz, diferindo na quantidade de energia contida nos mesmos. Esta grandeza pode ser obtida através da equação de Planck-Einstein que relaciona a energia com o comprimento de onda (ou frequência) da radiação através da constante de Planck. O Espectro pode então ser dividido nos seguintes grupos, apresentados na Figura 2.2:

- Radio;
- Microondas;
- Infravermelho;
- Visível;
- Ultravioleta;
- Raios X;
- Raios Gamma.

Conclui-se que a região visível do Espectro é uma porção bastante pequena deste. A análise unicamente do espectro visível reduziria muito a informação que pode ser obtida. A radiação eletromagnética pode ser expressa em termos de energia, comprimento de onda e frequência. Dado que esta dissertação se encontra no âmbito de *Remote Sensing* a radiação será caracterizada pelo seu comprimento de onda, por ser esta a nomenclatura habitual.

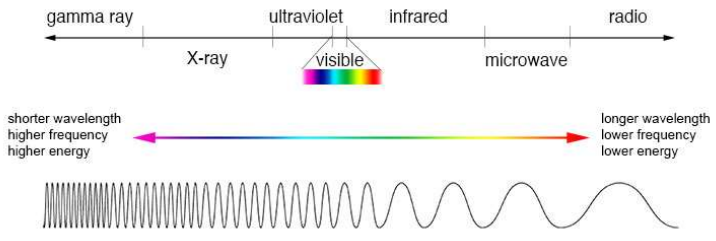


Figura 2.2: Regiões do Espectro Eletromagnético (https://imagine.gsfc.nasa.gov/).

2.2.1 Espectro e a Observação da Terra

Os grupos do espectro utilizados para a observação terrestre são: o visível, o infravermelho e as micro-ondas. Estas últimas no caso dos sensores ativos. Por causa das diferentes características de cada grupo, cada tipo de radiação fornece diferente informação. Em conjunto com este facto as diferentes coberturas da Terra absorvem e refletem as diferentes regiões do espectro de maneiras distintas. Esta informação pode ser sintetizada na assinatura espectral do que é observado. Uma assinatura espectral consiste nas medições de radiação refletida ou emitida para um determinado corpo ao longo dos diferentes comprimentos de onda do espectro. Na Figura 2.3 são apresentadas as assinaturas espectrais do solo, da água e da vegetação verde.

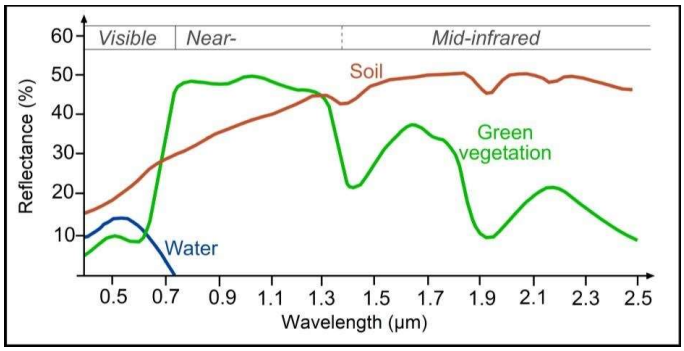


Figura 2.3: Assinaturas espectrais do solo (castanho), da água (azul) e da vegetação verde (verde) (https://grindgis.com/).

Consoante o que se pretende estudar algumas zonas do espectro podem revelar-se mais úteis do que outras. Na Tabela 2.1 são apresentadas as bandas utilizadas na observação da Terra. O estudo dos habitats subaquáticos e da vegetação marinha as bandas que se destacam são o Aerossol Costeiro e a Azul, dado que em boas condições conseguem obter informação a profundidade de 20 a 30 metros. A sensibilidade da banda Verde a superfícies de folhas permite a identificação de zonas de vegetação saudável. Não apenas esta última é importante no estudo da vegetação, mas também a banda Vermelha, Limite do Vermelho (*Red Edge*) e Próximo Infravermelho (NIR). Possibilitam um maior conhecimento sobre o tipo de solo, saúde da vegetação (através dos níveis de absorção da clorofila detetados) e da biomassa de uma região. A humidade do solo e irrigação do mesmo pode ser constatada pelas bandas de Ondas Curtas do Infravermelho (SWIR), sendo que estas têm a capacidade de transpor nuvens pouco densas. Para a deteção de nuvens a banda *Cirrus* é a mais indicada, pois os níveis de reflexão das nuvens são bastante elevados. Quando se pretendem fazer estudos da temperatura à superfície, monitorização de zonas vulcânicas recorremos às bandas Térmicas Infravermelhas. Estas, ao contrário das anteriores, recorrem à radiação emitida e não à refletida, sendo que têm resoluções menores. As bandas Pancromáticas têm como finalidade melhorar a resolução espacial das outras bandas. Estas captam o espectro visível numa única imagem (sendo sacrificada a resolução espectral) com resoluções mais elevadas do que as imagens de uma única banda. Aplicando técnicas de *pansharpening* obtêm-se resoluções espaciais mais elevadas. Por fim para detetar doenças em árvores causadas por insetos ou ter mais sensibilidade à época do ano recorre-se a banda do Amarelo. Contudo esta banda ainda é recente não existindo muitos sensores que façam a sua aquisição.

Tabela 2.1: Comprimentos de Onda relativos às bandas utilizadas em *Remote Sensing*.

BANDAS	INTERVALO DE COMPRIMENTOS DE ONDA (μm)
AEROSSOL COSTEIRO	0.430-0.450
AZUL	0.450-0.510
VERDE	0.530-0.590
VERMELHO	0.640-0.670
AMARELO	0.585-0.625
LIMITE VERMELHO	0.705-0.745
INFRAVERMELHO PRÓXIMO (NIR)	0.760-1.04
ONDAS CURTAS DO INFRA-VERMELHO (SWIR)	1.57-2.35
PANCROMÁTICA	0.500-0.680
CIRRUS	1.36-1.38
INFRAVERMELHO TÉRMICO	10.6-12.5

A combinação de várias bandas também permite a extração de mais informação. Para tal existem diversos índices dedicados às diferentes características que se pretendem detetar. Este assunto será aprofundado na 3.1.

2.3 Satélites de Observação da Terra

A 24 de Outubro de 1946 a Terra é fotografada pela primeira vez a partir do Espaço pelo míssil alemão V-2 (Figura 2.4). Apesar de se ter recorrido ao dispositivo germânico, esta imagem foi obtida pelos Estados Unidos da América (EUA). Em 1962 a União das Repúblicas Socialistas Soviéticas (URSS) lança o *Kosmos 4*, o seu primeiro satélite de *Remote Sensing*. Em 1964 é lançado pelos EUA o *Nimbus 1*, o primeiro satélite dedicado ao estudo climático. Em 1966 a França lança o *Diapason*, o seu primeiro satélite geodésico.

A observação da Terra a partir do espaço é uma possibilidade de aprendizagem sobre o planeta bastante enriquecedora. Torna possível uma cobertura no espaço e no tempo de uma forma simples, permitindo a análise de fenómenos a larga escala, chegando mesmo a lugares inacessíveis. A constatação de mudanças graduais (tais como as alterações ambientais) na superfície terreste também se torna percetível.

Neste âmbito a ESA desenvolveu um conjunto de programas na linha da Observação da Terra. Destes destacam-se o *Copernicus*, *The Living Planet* e o *The International Charter Space and Major Disasters*.

Figura 2.4: Primeira fotografia da Terra a partir do Espaço (https://gizmodo.com/).

2.3.1 Programa Copernicus

Copernicus é o programa mais ambicioso da ESA para a observação terrestre. Sucede ao *Global Monitoring for Environment and Security* (GMES) e fornece informação útil para a gestão ambiental, para a compreensão dos efeitos resultantes das alterações climáticas e para a segurança civil tendo as seguintes categorias de serviços:

- Gestão da terra;
- Ambiente Marinho;
- Atmosfera;
- Resposta de Emergência;
- Segurança;
- Alterações Climáticas.

Para o cumprimento destes objetivos, o programa *Copernicus* tem ao seu dispor a família dos satélites *Sentinel* (Figura 2.5). Cada um dos seus satélites tem uma finalidade específica, sendo então os seguintes,

- ***Sentinel 1***: capta imagens de radar sendo responsável pelas observações meteorológicas diurnas e noturnas;
- ***Sentinel 2***: trata da captação de imagens óticas de boa resolução para serviços da Terra;
- ***Sentinel 3A***: recolhe dados para os serviços da Terra e oceânicos;
- ***Sentinel 4 e 5***: fornecem os dados relativos à monitorização da composição da atmosfera pela órbita geoestacionária e polar;
- ***Sentinel 6***: carrega consigo altímetro para a medição do nível do mar para a oceanografia e estudos climáticos;
- ***Sentinel 5P***: o primeiro satélite em órbita para a monitorização da atmosfera terrestre.

Os *Sentinel 4*, *5* e *6* ainda não se encontram em órbita.

2.3.2 Sentinel 2

Dado o objetivo desta dissertação ser o estudo da superfície terrestre, o *Sentinel 2* assume uma maior importância. Na verdade, não é apenas um satélite, mas uma constelação de dois satélites (*2A* e *2B*) que cobrem toda a Terra com um período de 5 dias. Encontram-se a orbitar a Terra com uma altitude de $786\ Km$ e desfasados de $180°$. Cada uma das imagens obtidas abrange uma área de $100\ Km^2$. O *Sentinel 2A* foi lançado a 23 de junho de 2015 e o *Sentinel 2B* a 7 de março de 2017.

Os satélites encontram-se munidos com o *Multispectral Imager* (MSI). Este sensor tem as treze bandas espectrais presentes na Tabela 2.2 com diferentes resoluções espaciais. Este sensor tem uma resolução radiométrica de 12 *bits*, o que permite medir valores de reflectância[2] para todas as suas bandas no intervalo $[0,4095]$. O *Sentinel 2* tem como aplicações o estudo da saúde das plantas, as alterações terrestres, os corpos de água e o mapeamento em caso de desastres.

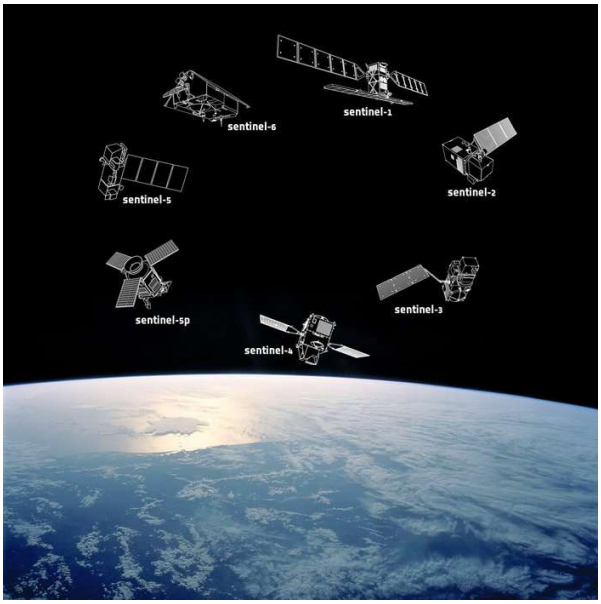


Figura 2.5: Família dos Sentinel (http://www.esa.int/).

No âmbito da saúde das plantas o seu desenho levou em consideração a necessidade de distinguir diferentes culturas, a necessidade de cálculo de diversos índices de vegetação, tais como a área de folhas, a clorofila destas e o seu conteúdo de água. Esta informação é essencial para a monitorização do crescimento da vegetação e muito útil na tomada de decisões sobre a gestão de culturas. As treze bandas disponibilizadas permitem uma análise bastante robusta da Terra.

[2] Esta medida consiste no rácio de radiação eletromagnética incidente com o fluxo que é refletido.

As alterações terrestres consistem no mapeamento das mudanças na cobertura da Terra ao longo do tempo. Permitem a sua análise e identificação se resultam de causas naturais ou humanas. Também é feita a distinção entre as diferentes classes de cobertura (floresta, cultivos, pastagem, superfícies aquáticas e cobertura artificial).

Relativamente ao estudo de corpos de água torna possível a captação de parâmetros de qualidade, tais como a concentração de clorofila à superfície, a deteção de florestas de algas nocivas e a turbidez[3] da água. Estes permitem avaliar se a água se encontra saudável e os seus níveis de poluição.

Este satélite pode ser útil para a compreensão e estudo da ocorrência de desastres, permitindo analisar o impacto dos incêndios e identificando através da imagem as áreas ardidas. Possibilita a identificação das alterações ao nível da cobertura terrestre. O conjunto de informação que o *Sentinel 2* permite obter (vegetação, corpos de água, etc), pode ajudar na monitorização da propagação de doenças (por exemplo, a malária). De qualquer modo, esta monitorização está fortemente dependente da inexistência de lacunas de dados por causa das condições atmosféricas, mantendo o período de observação de 5 dias.

Tabela 2.2: Bandas Espectrais *Sentinel 2*.

BANDAS *SENTINEL-2*	COMPRIMENTO DE ONDA (μm)	RESOLUÇÃO (m)	LARGURA DE BANDA (nm)
BANDA 1 – AEROSSOL COSTEIRO	0.443	60	20
BANDA 2 – AZUL	0.490	10	65
BANDA 3 – VERDE	0.560	10	35
BANDA 4 – VERMELHO	0.665	10	30
BANDA 5 – BORDA VERMELHA DE VEGETAÇÃO	0.705	20	15
BANDA 6 – BORDA VERMELHA DE VEGETAÇÃO	0.740	20	15
BANDA 7 – BORDA VERMELHA DE VEGETAÇÃO	0.783	20	20
BANDA 8 – NIR	0.842	10	115
BANDA 8A – NARROW NIR	0.865	20	20
BANDA 9 – VAPOR DE ÁGUA	0.945	60	20
BANDA 10 – SWIR – CIRRUS	1.375	60	20

[3] Corresponde à capacidade de absorção e reflexão de luz da água, permitindo o estudo da qualidade desta.

BANDA 11 – SWIR	1.610	20	90
BANDA 12 – SWIR	2.190	20	180

2.3.3 Landsat

O programa *Copernicus* da ESA não é único a dedicar-se ao estudo da superfície terreste. A NASA conjuntamente com a *United States Geological Survey* (USGS) tem o programa *Landsat*. Este programa tem também uma especial importância pois começou em 1972, nessa altura denominado por *Earth Resources Technology Satellite* (ERTS) e até hoje, com o sucessivo lançamento de novos satélites ao longo do tempo, permite ter uma base de dados de imagens da Terra bastante vasta. A cronologia deste programa pode ser vista na Figura 2.6.

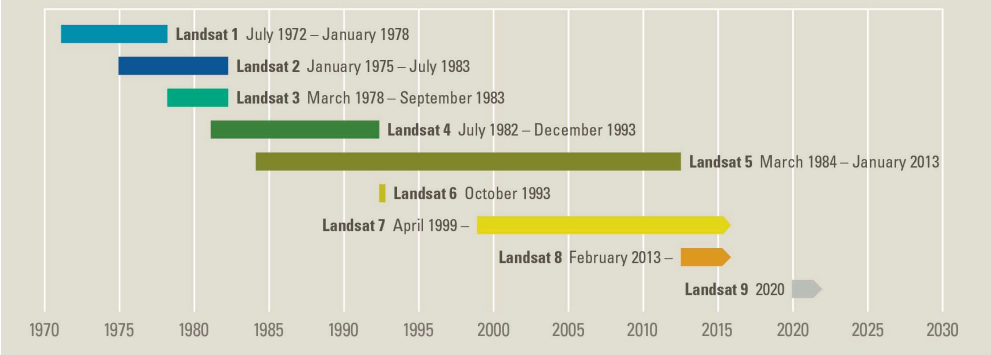


Figura 2.6: Cronologia do Programa *Landsat (https://landsat.usgs.gov).*

Este programa esteve a cargo de diferentes organizações, chegando a estar em mão privada e depois voltando para a alçada da NASA. Pode ser registada também a evolução dos sensores que estes satélites carregavam consigo. Começando com o *Return Beam Vidcon* (RBV) e *Multispectral Scanner* (MSS), sendo o MSS mais importante, dado que tinha a capacidade de aquisição de imagens em quatro bandas diferentes (Verde, Vermelho e duas de Infravermelho).

Apenas um dos satélites falhou a entrada na órbita, o *Landsat 6*, e atualmente existem dois satélites em operação, *Landsat 7* e *Landsat 8* munidos, respetivamente, com os sensores *Enchanced Thematic Mapper Plus* (ETM+) e *Operation Land Imager* (OLI) e *Thermal Infrared Sensor* (TIRS). Encontra-se também já previsto o lançamento do *Landsat 9* para 2020.

A plataforma mais atual do programa *Landsat*, como referido anteriormente, é o satélite *Landsat 8*. Este satélite faz a aquisição das bandas apresentadas na Tabela 2.3. Importa referir que as bandas 7 e 8 são bandas térmicas (responsabilidade do TIRS) e as restantes correspondem a imagens da superfície terrestre (responsabilidade do OLI). Estes últimos adquirem dados sobre a refletividade de diversos comprimentos de onda e os sensores TIRS captam o brilho emitido pela área em observação, sendo bastante úteis pois permitem a correção atmosférica das imagens. O *Landsat 8* tem um ciclo de dezasseis dias captando áreas de 185 Km por 180 Km. Uma das grandes vantagens é a calibração igual à existente no *Sentinel 2*.

Por fim, é importante fazer uma breve menção ao *Landsat 9*. Este satélite será colocado em órbita em dezembro de 2020. Esta plataforma será, no seu essencial, uma replicação de seu

antecessor. As bandas obtidas são as mesmas que o *Landsat 8* com as mesmas resoluções. Estes instrumentos têm um ciclo de vida de cinco anos, enquanto o TIRS apenas tinha um ciclo de vida de três anos. Outro detalhe é o facto de que o TIRS tinha um problema que reduzia a precisão das suas medidas, que foi corrigida no TIRS-2.

Tabela 2.3: Bandas Espectrais *Landsat 8*.

BANDAS *LANDSAT 8*	COMPRIMENTO DE ONDA (μm)	RESOLUÇÃO (m)
BANDA 1 – AEROSSOL COSTEIRO	0.435 - 0.451	30
BANDA 2 – AZUL	0.452 - 0.512	30
BANDA 3 – VERDE	0.533 - 0.590	30
BANDA 4 – VERMELHO	0.636 - 0.673	30
BANDA 5 – NIR	0.851 - 0.879	30
BANDA 6 – SWIR-1	1.566 - 1.651	30
BANDA 7 – TIR-1	10.60 - 11.19	100
BANDA 8 – TIR-2	11.50 - 12.51	100
BANDA 8A – SWIR-2	2.107 - 2.294	30
BANDA 9 – PAN	0.503 - 0.676	15
BANDA 10 – CIRRUS	1.363 - 1.384	30

2.3.4 Outros Satélites de Observação da Terra

Para além destes dois programas de observação da Terra existem muitos outros satélites concorrentes pelo mundo. Entre eles o *Tropical Rainfall Measuring Mission* (TRMM) da *Japan Aerospace Exploration Agency* (JAXA) e NASA para a monitorização e estudo da precipitação tropical. O *CloudSat* programa de uma parceria entre a NASA e o Canadá que mede a altitude e propriedades das nuvens de forma a relacionar estas propriedades com o clima. O *Jason-3* que junta quatro organizações, NASA, NOAA, *Centre National d'Etudes Spatiales* (CNES) e a *European Organization for the Exploitation of Meteorological Satellites* (EUMETSAT) que realiza medidas da altura da superfície oceânica. Destaca-se também o *Terra* da NASA que se encontra munido com o sensor *Moderate Resolution Imaging Spectroradiometer* (MODIS). Esta plataforma tem como finalidade o estudo da atmosfera, da terra, da neve, de forma a retirar conclusões sobre a ação humana e alterações climáticas. O instrumento tem uma resolução espectral de 36 bandas, porém a resolução espacial varia entre os $250\ m$ e $1000m$. A resolução temporal é de dois dias. Existe ainda uma vasta lista de satélites e programas de observação terrestre que não são mencionados nesta dissertação.

2.3.5 Produtos de Observação da Terra

As missões de Observação da Terra oferecem vários produtos com diferentes níveis de processamento sobre as imagens adquiridas pelas suas plataformas. Deste modo o *Committee*

on Earth Observation Satellites (CEOS), o organismo que tem como função coordenar as observações da Terra de modo a facilitar o acesso aos dados, normalizou as diversas categorias de produtos disponibilizados por estas missões.

2.3.5.1 Nível 0

Esta categoria representa os produtos com menor processamento, sendo os dados obtidos pelos sensores reconstruídos com resolução espacial total (Hagolle, 2014). Estes produtos contêm também toda a informação (meta dados) necessária para o processamento. No caso do *Sentinel 2*, o *Ground Segment* (este consiste nos elementos da base terrestre com a finalidade de gerir os satélites e distribuir os dados obtidos) tem como tarefas de processamento as seguintes: análise de telemetria para a deteção de erros; datação das linhas da imagem; extração da imagem em baixa resolução executando uma análise telemétrica a fim de verificar se os valores medidos se encontram nos intervalos esperados.

2.3.5.2 Nível 1

O processamento deste nível de produtos (que têm como base os dados de nível 0) não diferencia os pixéis consoante a sua natureza (nuvens, floresta, oceano), recebendo todos o mesmo tratamento. Os valores dos pixéis são expressos em unidades físicas. Estes produtos podem ser reamostrados para uma grelha cartográfica, caso não o sejam fornecem a informação necessária para estas operações (Tian, Zhang, Tian, & Sun, 2016).

O *Sentinel 2* subdivide esta categoria em três grupos, 1A, 1B e 1C. Os produtos 1A são constituídos pelos dados de Nível 0 com os pacotes de dados mais relevantes da missão descomprimidos (correção e calibração radiométrica, espectral e geométrica). A classe 1B já atravessa uma sequência de processamento mais complexa. Sobre as imagens são executados um conjunto de tarefas radiométricas; é feita a reamostragem geométrica entre a banda de referência (por omissão a Banda 4) e a *Global Reference Image*[4] (GRI). A compressão das imagens deste grupo recorre ao algoritmo JPEG2000 (jp2). O processamento de nível 1C consiste em correções radiométricas e geométricas que incluem a orto-retificação e registo espacial no sistema global de referência com precisão na ordem do sub-píxel. Neste nível são também geradas máscaras de nuvens e de terra/água. Apenas este último nível está disponível aos utilizadores. Ambos os produtos 1B e 1C correspondem à radiância do Topo da Atmosfera (TOA). Quando se refere à radiância do TOA, significa que os valores medidos correspondem aos valores emitidos pela atmosfera. Note-se que fatores climatéricos podem influenciar muito estas medições. Quando se apresentarem os produtos de nível 2, serão apresentadas as radiâncias de abaixo da atmosfera (BOA).

Quanto ao *Landsat 8* existem três tipos de produtos de nível 1 (USGS, 2018), *Standard Terrain Correction* (L1TP), *Systematic Terrain Correction* (L1GT) e *Systematic Correction* (L1GS). Todas estas categorias de produtos executam a calibração radiométrica. Os dois primeiros tipos de produtos nas suas operações de correção geométrica consideram o relevo, sendo

[4] A GRI corresponde a um conjunto de aquisições do *Sentinel 2* com a geometria já corrigida. Deste modo é então utilizada como uma referência absoluta no registo temporal e espectral das imagens deste satélite, ver (Déchoz et al., n.d.).

que o primeiro, L1TP, executa a orto-retificação tornando-se o produto de maior qualidade. O grupo de produtos L1GS executa as correções geométricas sem recurso aos dados do relevo.

2.3.5.3 Nível 2

Ao contrário do processamento indiscriminado de pixéis que acontece no nível 1, nesta categoria o tratamento destes diferencia-se consoante a sua natureza (Hagolle, 2014). Do lado do *Sentinel 2* existem os produtos 2A que executam essencialmente correções atmosféricas obtendo-se imagens de BOA a partir dos dados 1C. Para além disso são também disponibilizados outros produtos: um mapa do Aerosol Optical Thickness, mapa do Vapor de Água e um mapa de classificação de cena conjuntamente com indicadores de qualidade. No que toca ao *Landsat 8* existem dois produtos: *Surface Reflectance* e *Provisional Surface Temperature.* O primeiro é relativo à radiação solar refletida pela Terra e o segundo informa sobre a temperatura da superfície terrestre (lembre-se que o satélite da NASA dispõe de sensores térmicos).

2.3.5.4 Nível 3

Os produtos de nível 3 são obtidos a partir de dados de diferentes datas. Tal como no nível 2 o processamento dos pixéis pode variar consoante a natureza do mesmo. Estes produtos são compostos de dados de nível 2 de um determinado período de tempo (Hagolle, 2014). No caso do *Sentinel 2* existe o nível 3A que consiste em compostos mensais de imagens 2A livres de nuvens e sombras destas. Estes produtos são disponibilizados pelo THEIA *Land Data Center* (organismo francês com a finalidade de facilitar o acesso aos dados de observação da Terra). No programa *Landsat* os produtos de nível 3 ficaram disponíveis durante o corrente ano sendo estes: *Dynamic Surface Water Extent*, *Fractional Snow Covered Area* e *Burned Area.* O primeiro é relativo a existência e condições da água. O segundo indica a percentagem do pixel que corresponde a neve. O último, e mais interessante para este trabalho, classifica os pixéis relativamente à queima dos mesmos e assinala a probabilidade de incêndio na região que representam.

2.3.5.5 Nível 4

Esta última categoria de produtos é definida pelo CEOS, porém nenhum dos satélites aqui abordados oferece ainda produtos deste nível. Consistem então em modelos ou resultados obtidos a partir de dados do nível 3. Não são medidas diretas dos instrumentos, apenas derivam dessas medidas (Hagolle, 2014).

Nesta dissertação utilizaram-se os produtos de nível 1C. Sendo os produtos disponibilizados gratuitamente ao utilizador com menor processamento verificou-se a possibilidade de utilização destes para a resolução do problema.

2.4 Sistemas de Informação Geográfica

Em 1854 quando a cidade de Londres é atingida pela Cólera, *John Snow*, médico nascido em *York* e considerado o pai da epidemiologia, decide mapear os locais de surto desta doença infeciosa. Este mapeamento permitiu verificar a relação entre o sistema de canalização da capital britânica e a enfermidade. Para além da enorme importância que este fenómeno teve para a Saúde Pública, foi o primeiro passo na análise espacial. A contextualização de dados com

a sua localização geográfica permitiu a extração de uma informação mais rica do que a resultante apenas dos dados em bruto.

Em 1968, *Roger Tomlinson*, apresenta pela primeira vez o termo de "Sistema de Informação Geográfica" (SIG) em (Tomlinson, 1969). No âmbito de um projeto do governo canadiano foi implementada uma ferramenta computacional para o armazenamento e análise de dados na sua conjuntura geográfica. Esta ferramenta tinha como propósito o desenvolvimento e entendimento da terra, da água e dos recursos naturais.

Um SIG pode ser definido como uma ferramenta computacional para a análise, armazenamento, manipulação e visualização da informação geográfica num mapa. A perspetiva geoespacial que surge pela conexão dos dados e a geografia dos mesmos torna mais fácil o estudo e compreensão de determinados fenómenos. Por exemplo, é mais enriquecedor observar a localização geográfica num mapa do que apenas examinar as suas coordenadas numa tabela.

A análise espacial executada através de um SIG permite verificar e analisar alterações climáticas, catástrofes naturais, dinâmicas populacionais, entre outras manifestações, através da relação entre os dados e os seus atributos espaciais. Num SIG existem dois tipos de dados: *raster data* e vetores. Os dados *raster* consistem em imagens com os respetivos valores dos pixéis. Os vetores podem assumir a forma de pontos, linhas ou polígonos, tratando-se de tabelas com vários campos para os seus elementos, associando a cada elemento uma localização no mapa.

Estes sistemas têm aplicação na área ambiental, militar, de segurança pública, empresarial e muitas outras. Existem vários *softwares* GIS, sendo que nesta dissertação serão analisados o ArcGIS e o *Quantum GIS* (QGIS).

2.4.1 ArcGIS

O ArcGIS é um *software* líder do mercado, desenvolvido pela *Environmental Systems Research Institute* (Esri), para utilização governamental, comercial e industrial.

A visualização e gestão de dados é assegurada pelo ArcMap e ArcCatalog. O primeiro é o componente central do ArcGIS dado ser a ferramenta de criação de mapas e edição de dados permitindo assim ser realizada a análise espacial. O segundo consiste num explorador de ficheiros do formato SIG com a possibilidade de escrita de meta dados nesses ficheiros.

Uma das principais vantagens do ArcGIS em relação à concorrência é a capacidade de visualização 3D da Terra através dos instrumentos ArcGlobe e ArcScene. O primeiro dedica-se a regiões extensas, enquanto que o segundo se dedica a áreas de estudo mais pequenas.

A plataforma ArcGIS *Online* permite a obtenção um conjunto muito amplo de informação e dados, extremamente útil a qualquer um que procure fazer uma análise espacial mais rica e robusta.

A ArcToolbox oferece ao utilizador um vasto conjunto de ferramentas de geoprocessamento. Encontram-se opções para a modelação da superfície, análise e gestão recorrendo: a dados em 3D (*3D Analysis Tools*); instrumentos cartográficos dedicados à produção de mapas (*Cartography Tools*); atribuição de localizações aos respetivos endereços (*Geocoding Tools*);

ferramentas para a análise espacial estatística úteis na obtenção de distribuições espaciais, padrões e relações.

Para além do geoprocessamento no que toca à análise espacial também se encontram inúmeras funcionalidades. Podem ser destacadas a calculadora dedicada a dados *raster*; procura do caminho mais rápido utilizando ferramentas de custo e de distância Euclidiana; instrumentos dedicados à hidrologia, para a deteção da direção, acumulação e largura do fluxo; análise da radiação solar e o cálculo da geometria e estatísticas zonais.

Por fim, a grande desvantagem, o ArcGIS é um *software* pago com três níveis de licenças: básica, padrão e avançada. A licença básica e padrão não permitem o acesso total ao SIG.

2.4.2 QGIS

O QGIS é o grande rival do ArcGIS no que toca a *softwares* de informação geográfica. Este SIG surge em 2002 criado por Alaskan Sherman e a sua evolução é garantida através de voluntariado.

A janela principal nomeada por *Map Canvas* é responsável por todas as funcionalidades dedicadas aos mapas. Este *software* também traz consigo um explorador para a procura e gestão de dados SIG.

O QGIS oferece também ao utilizador uma quantidade enorme de *plug-ins* permitindo-lhe adquirir cada vez mais funcionalidades. Dentro desta imensa oferta podem ser destacados os seguintes:

- ***OpenLayers***: fornece o acesso a diversas camadas de mapas, tais como os mapas do Google e do Bing;
- ***Qgis2threejs***: apesar do QGIS não ter de origem a capacidade de visualização em três dimensões, com este *plug-in* torna-se possível;
- ***Semi-Automatic Classification***: implementado com o objetivo de classificar e processar imagens multiespectrais, sendo bastante útil em Remote Sensing.

As ferramentas básicas do QGIS consistem na adição, criação e edição de camadas. Quanto à etiquetagem destas camadas é disponibilizado um grupo bastante grande de opções e propriedades. Para além disso também é dada a possibilidade de exportação dos mapas nos mais variados formatos (PDF, JPG, TIF).

Ao nível do processamento estão imediatamente disponíveis mais de 600 ferramentas. Estas podem ser agrupadas da seguinte forma:

- ***Domínios específcos***: dedicados a determinadas matérias específicas, tais como a geoestatística, a hidrologia, entre outras;
- ***Imagem***: manipulação, análise, calibração, segmentação de imagens;
- ***Raster:*** operação sobre dados *raster*;
- ***Raster-Vector:*** conversão entre estes dois tipos de dados;
- ***Vector:*** operações sobre vetores, linhas e polígonos.

A possibilidade de automatização de um conjunto de operações de processamento também é possível. Por fim, ao contrário do ArcGIS trata-se de um *software* gratuito.

2.4.3 QGIS *versus* ArcGIS

Após terem sido apresentados os dois SIG importa fazer uma comparação entre ambos apurando qual mostra ser mais adequado no desenvolvimento desta dissertação.

Em primeiro lugar serão analisados os fatores positivos e negativos de cada um deles. Iniciando pelo QGIS, destaca-se o facto de ser *opensource* e consequentemente gratuito; um conjunto muito diversificado de opções no que toca às etiquetas a colocar nos mapas; aceita uma grande variedade de dados; muitas ferramentas de análise e estão disponíveis muitos *plug-ins*. Negativamente não se encontra disponível de raiz a integração a três dimensões; o modelador gráfico ainda tem alguns erros de implementação e não tem uma topologia automática para a correção de erros.

Observando agora o ArcGIS, este apresenta uma estrutura sólida para o geoprocessamento; muita simbologia disponível; muitas funcionalidades 3D; oferta de muita informação e dados *online*. Contudo, a nível da aceitação de dados é bastante restritivo, não é um *software* distribuído gratuitamente.

Quando comparados, o ArcGIS destaca-se pela sua simplicidade na adição de dados; o ArcCatalog tem maior variedade de opções que QGIS *Browser* (explorador do QGIS); a informação disponibilizada pelo ArcGIS *Online* é superior à disponibilizada pelo QGIS; apesar de um dos pontos fortes do QGIS serem os *plug-ins,* o ArcGIS tem uma variedade superior e *plug-ins* mais especializados; as ferramentas de geoestatística do ArcGIS disponibilizam um conjunto de explicações sobre os resultados atalhando o entendimento das mesmas, algo que o QGIS não oferece; a análise de rotas mais rápidas dada a maior informação disponível no ArcGIS é superior também; quando se pretende modelar e automatizar tarefas no ArcGIS é possível fazê-lo de forma mais intuitiva e enfrentando menos erros; as funcionalidades de três dimensões são superiores; as ferramentas de edição são ainda mais avançadas; a topologia de correção de erros do ArcGIS também tem um conjunto mais vasto de regras, sendo ainda mais interativo e a documentação disponível é mais extensa e aprofundada.

Quanto ao QGIS é superior ao ArcGIS no maior tipo de dados que aceita; não implica licenças tendo todas as suas ferramentas imediatamente disponíveis; na aplicação para Remote Sensing por causa do *plug-in Semi-Automatic Classification*; as ferramentas de *Geocoding* são melhores e a generalização de pontos, linhas e polígonos é mais eficiente no QGIS.

No que toca ao desenvolvimento de *plug-ins* e automatização de tarefas, uma das ferramentas mais poderosas de ambos os SIG é facto de compilarem código em *Python* o que permite a utilização de inúmeras bibliotecas que facilitarão a análise de dados e a extração de conhecimento.

2.4.4 Web SIG

No âmbito dos SIG existe também *software* que pode ser implementado nos navegadores de internet. Tratam-se normalmente de implementações com menos funcionalidades, mas mais preparadas para determinados objetivos, como o estudo de uma região específica, a

observação de mapas, definição de rotas, entre outros. Um Web SIG também contém um conjunto de ferramentas para a análise espacial e por vezes o facto de apenas ser preciso o navegador torna-o bastante mais acessível. Exemplos deste tipo de aplicações são o ArcGIS Online e EPIC WebGIS Portugal. Este último, dedicado ao território nacional, já fornece uma boa variedade de instrumentos para a visualização e estudo de Portugal, tendo sido realizado por uma equipa do Instituto Superior de Agronomia. Através dele podem ser recolhidas informações geológicas, do solo, da água, de vegetação, climatéricas, entre outras, essenciais para o ecossistema e gestão de recursos naturais.

2.5 Sumário

Atualmente existem inúmeras fontes de dados no âmbito da observação da Terra. Mas mais importante a disponibilidade dos mesmos com uma frequência cada vez maior, produtos com cada vez maior qualidade (a nível da resolução espacial, espectral e temporal) e de forma gratuita o que permite um desenvolvimento mais rápido. Os SIG permitem uma análise no contexto geográfico algo crucial quando se trata da observação da Terra. Também a possibilidade de automatizar processos permite uma análise mais eficiente do território e neste caso do estudo das FGC. A escolha do *software* prendeu-se com o objetivo de utilizar softwares livres, de forma a que a metodologia desenvolvida não estivesse dependente da compra de nenhuma licença. Deste modo foi escolhido o QGIS, tratando-se também de um dos SIG mais utilizados.

3 Estado de Arte

Este capítulo pode ser dividido em duas partes. Na primeira, presente em 3.1 e 3.2, são apresentados os índices de vegetação que serão estudados para uma deteção de intervenções mais eficaz. São também apresentados trabalhos que sem recurso à aprendizagem automática extraem informação das observações da Terra. A segunda parte do capítulo (em 3.3) dedica-se às técnicas de classificação que já são utilizadas com a finalidade de extração de conhecimento a partir das imagens multiespectrais. Também são mencionadas as formas de estimação do erro e validação dos algoritmos de *Machine Learning.*

3.1 Índices de Vegetação

Uma das grandes áreas de aplicação e investigação de *Remote Sensing* é o estudo da vegetação. Esta análise é feita com o recurso à informação da reflectância eletromagnética obtida através das copas das árvores utilizando sensores passivos. A reflectância depende das características químicas e morfológicas da vegetação observada. O espectro de luz utilizado nestas aplicações situa-se entre os $10\ nm$ e $1700\ nm$ de comprimento de onda. São então abrangidas as regiões de radiação ultravioleta, espectro visível e o espectro de radiação infravermelha.

A interpretação dos dados obtidos pelos satélites de observação, ao nível da vegetação, pode ser feita pelo cálculo dos índices de vegetação. Em (Xue & Su, 2017), (Hamuda, Glavin, & Jones, 2016) e (Mestre, Fonseca, & Mora, 2017), são apresentados vários índices. Estes podem ser agrupados de diferentes formas, sendo que nesta dissertação serão divididos entre os que utilizam dados do espectro visível e não visível e os que apenas necessitam do espectro visível. Nas próximas secções: R é a reflectância da banda vermelha; G é a reflectância banda verde e B a reflectância banda azul. As restantes abreviaturas foram apresentadas na Tabela 2.1.

3.1.1 Dados do espectro visível e não visível

Em (Xue & Su, 2017) são expostos os seguintes índices:

3.1.1.1 *Ratio Vegetation Indices*

Esta razão tem como base o princípio de que as folhas das árvores absorvem mais vermelho do que luz infravermelha, traduzindo-se matematicamente por,

$$R = \frac{R}{NIR} \quad \textbf{(1)}$$

Este índice é utilizado para a estimação e monitorização da biomassa verde. São alcançados melhores resultados em zonas com maior densidade de vegetação. Quando a premissa anterior não se verifica, o *Ratio Vegetation Index* (RVI) torna-se mais sensível aos efeitos atmosféricos resultando numa representação da biomassa menos correta.

3.1.1.2 *Difference Vegetation Index* e *Normalized Difference Vegetation Index*

O *Difference Vegetation Index* (DVI) pretende analisar o estado de saúde da vegetação. Para tal recorre-se à banda NIR, de forma a obter informação sobre a estrutura da folha e à banda R que é afetada essencialmente pela absorção da clorofila[5] (quanto maior, mais saudável se encontra a planta). As outras duas bandas visíveis não são utilizadas, pois, apesar de também serem sensíveis ao estado da clorofila, os valores de reflectância destas retratam outros fenómenos de absorção (atmosféricos, ou no caso da banda azul a absorção dos carotenoides[6]). Este índice é muito sensível à vegetação verde, mesmo em regiões em que a presença desta seja escassa. O cálculo é dado pela equação **(2)**.

$$DVI = NIR - R \quad \textbf{(2)}$$

Apesar da utilidade inegável deste índice, tem a grande desvantagem de não se encontrar normalizado. Este facto torna mais difícil a comparação entre valores obtidos para zonas diferentes e a perceção de um valor alto ou baixo do mesmo. Assim surge o *Normalized Difference Vegetation Index* (NDVI) que normaliza o resultado do DVI. Este trata-se do índice mais comum na estimação global e regional de vegetação. O NDVI não se relaciona apenas com a copa das árvores, mas também com a fotossíntese (devido ao estudo da clorofila). Ainda que já seja obtido um valor normalizado não resolve os problemas relativos ao brilho e cor do solo, os quais são refletidos nestes índices. Para além disso revela sensibilidade à existência de nuvens e às sombras das mesmas, degradando os seus resultados nestas condições. O NDVI é calculado pela equação **(3)**.

$$NDVI = \frac{NIR - R}{NIR + R} \quad \textbf{(3)}$$

---

[5] A clorofila é um pigmento verde que se encontra nas folhas responsável pela absorção de luz solar no processo de fotossíntese.

[6] Os carotenoides são pigmentos amarelos aos quais é conferido um papel secundário na fotossíntese dado que absorvem luz em diferentes comprimentos de onda diferentes da clorofila.

3.1.1.3 *Enhanced Vegetation Index* 1 e 2

O *Enhanced Vegetation Index* (EVI) foi um índice desenvolvido para o MODIS com o objetivo de apresentar melhorias em relação ao NDVI. O EVI destaca-se pela sua maior sensibilidade às alterações em regiões de biomassa bastante elevada; por uma maior robustez às condições atmosféricas e apresenta melhores resultados relativamente à presença de dossel[7] na imagem. Isto acontece porque ao contrário do NDVI, o EVI analisa a estrutura do dossel e fenomenologia[8], em vez de absorção da clorofila. O cálculo do EVI1 é dado pela equação **(4)**, em que A, $C1$, $C2$ e L são fatores a definir.

$$EVI1 = A \times \frac{NIR - R}{NIR + C1 \times R - C2 \times B + L} \quad \textbf{(4)}$$

Para o MODIS os valores definidos são respetivamente: 2.5; 6; 7.5 e 1. Para o uso no *Sentinel 2* os valores de reflectância devem ser divididos por 10000, dado que os produtos de nível 1C são multiplicados por esse fator. Um possível problema deste índice consiste no uso de três bandas espectrais. Para certas plataformas a banda azul pode não estar disponibilizada ou ter uma resolução espectral bastante menor. Deste modo desenvolveu-se o EVI2 que apenas necessita das bandas NIR e R. O cálculo é dado pela equação **(5)**.

$$EVI2 = 2.5 \times \frac{NIR - R}{NIR + 2.4 \times R + 1} \quad \textbf{(5)}$$

Este índice assume uma menor importância nesta dissertação dado que são disponibilizadas todas as bandas necessárias para o cálculo do EVI1.

3.1.1.4 *Normalized Difference Moisture Index* e *Normalized Difference Water Index*

O *Normalized Difference Moisture Index* (NDMI) é um índice que se concentra na água presente na vegetação. Para tal recorre à reflectância nas bandas NIR e $SWIR$. A primeira é utilizada para entender a estrutura das folhas que constituem a vegetação observada, a segunda remove os efeitos não associados ao conteúdo de água nos dosséis das plantas. Quanto ao *Normalized Difference Water Index* (NDWI), este não é um índice de vegetação, mas sim um índice hidrológico. Este índice, apesar do primeiro também ser bem sucedido nesta tarefa, é bastante utilizado para o mapeamento de corpos de água. Diferenciam-se por utilizarem bandas $SWIR$ em comprimentos de onda distintos e porque o NDWI utilizada a banda $Narrow\ NIR$ em vez da NIR. O primeiro utiliza a banda de $1600\ nm$ e o segundo de $1240\ nm$. Apesar desta distinção, e das capacidades também reconhecidas ao NDMI para a descoberta de corpos de água, muitas vezes o NDMI é referido como NDWI. Também o facto de o *Sentinel 2* para o cálculo do NDWI ter de recorrer à banda 10 que por causa da sua resolução espacial de $60\ m$ não permite

[7] O dossel consiste na vida animal e vegetal que está presente acima do chão da floresta, habitando nas folhagens das árvores, resultado da sobreposição destas.

[8] Também designada apenas por fenologia corresponde a comportamentos periódicos dos seres vivos consequência das condições ambientais. Este tema será melhor abordado na observação dos dados obtidos, sendo verificados fenómenos sazonais nesses dados.

os resultados mais precisos acaba-se por recorrer à banda 11 que se encontra no comprimento de onda utilizado para o NDMI. O cálculo destes índices é dado pela equação **(6)**.

$$NDMI = \frac{NIR - SWIR}{NIR + SWIR} \quad \textbf{(6)}$$

3.1.1.5 *Normalized Multi-band Drought Index*

O *Normalized Multi-band Drought Index* (NMDI) trata-se também de um índice que pretende estimar a quantidade de água presente numa região. Contudo a melhoria proposta consiste em considerar duas bandas $SWIR$ sendo uma relativa à vegetação e outra ao solo. Este detalhe permite uma maior sensibilidade do índice para o estudar períodos de seca. O NMDI é obtido pela equação **(7)**.

$$NMDI = \frac{Narrow\ NIR - (SWIR1 - SWIR2)}{Narrow\ NIR + (SWIR1 - SWIR2)} \quad \textbf{(7)}$$

Sendo que a $SWIR1$ é a banda no comprimento de onda dos $1640\ nm$ e a $SWIR2$ dos $2130\ nm$.

3.1.2 Dados apenas do espectro visível

Os índices abordados nesta secção recorrem exclusivamente a bandas visíveis ao espectro eletromagnético. Em (Hamuda et al., 2016) e (Mestre et al., 2017) são apresentados alguns destes índices de vegetação.

3.1.2.1 Normalized Difference Index

O *Normalized Difference Index* (NDI) é um índice que permite a distinção, numa imagem RGB, da planta do solo presente no fundo da imagem. O seu cálculo é apresentado na equação **(8)** produzindo uma imagem próximo do binário numa escala de cinzentos.

$$NDI = 128 \times \left(\frac{G - R}{G + R} + 1\right) \quad \textbf{(8)}$$

3.1.2.2 Excess of Green Index

Este índice é bastante utilizado para a separação de plantas de não plantas. Para tal são utilizadas as coordenadas cromáticas (calculadas através dos valores normalizados das bandas RGB). São produzidas então imagens próximas do binário com um bom contraste entre a planta e o solo. O *Excess of Green Index* (ExG) é obtido através da equação **(9)**,

$$ExG = 2g - r - b \quad \textbf{(9)}$$

Em que g, r e b são as coordenadas crómaticas.

3.1.2.3 Excess of Red Index

O índice *Excess of Red* (ExR), tal como o ExG, tem como objetivo a separação das plantas do fundo da imagem. Recorre ao facto de a retina humana ser mais sensível ao vermelho, calculando o seu excesso na imagem através da equação **(10)**,

$$ExR = 1.3R - G \quad \textbf{(10)}$$

3.1.2.4 Excess of Green minus Excess of Red Index

Apesar do ExR ser um índice de excesso de cor inferior ao ExG, ganha uma importância maior quando estes dois índices se combinam originando o *Excess of Green minus Excess of Red Index* (ExGR). A separação das plantas do solo e dos resíduos de fundo é responsabilidade do ExG, sendo função do ExR a eliminação do ruído de fundo (resultante do solo) onde pode existir material verde-vermelho. O método de cálculo está presente na equação **(11)**,

$$ExGR = ExG - ExR \quad \textbf{(11)}$$

3.1.2.5 Color Index of Vegetation Extraction

O *Color Index of Vegetation Extraction* (CIVE) aplicou-se no estudo do crescimento de cultivos. É baseado também na separação das plantas do solo. Este índice tem melhores resultados que os métodos do infravermelho próximo por causa da maior sensibilidade às áreas verdes. A fórmula de cálculo observa-se na equação **(12)**,

$$CIVE = 0.441R - 0.811G + 0.383B + 18.78745 \quad \textbf{(12)}$$

3.1.2.6 Modified Excess Green Index

Este método em relação aos referidos anteriormente tem como grande vantagem a sua robustez às alterações de iluminação em tempo real, sendo preferencial para ambientes em que a luminosidade não é controlada. O *Modified Excess Green Index* (MExG) tem como resultado uma imagem em escala de cinzentos facilmente binarizável. Tem uma elevada precisão na distinção da planta e do solo.

$$MExG = 1.262G - 0.884R - 0.311B \quad \textbf{(13)}$$

3.2 Aplicações de Remote Sensing

Nesta secção serão analisadas várias aplicações já existentes no âmbito de *Remote Sensing*. Note-se que neste campo existem essencialmente dois tipos de aplicações: as de mapeamento de zonas com uma característica especial e as de deteção de mudanças. Sendo que este estudo se enquadra no segundo grupo são apresentadas duas aplicações de carácter temporal e uma de mapeamento.

3.2.1 Mapeamento de Corpos de Água a partir de imagens do Sentinel 2

Uma das aplicações já existentes que poderá ter utilidade para o desenvolvimento desta dissertação é o mapeamento de Corpos de Água. Exemplos destes são os rios, reservas de água, lagos entre outros.

Em (Du et al., 2016) para a classificação de um Corpo de Água recorre-se ao cálculo de um indicador de água. Este indicador pode ser então dado pelo NDWI ou pelo MNDWI. O

segundo é mais eficaz sendo que substitui a utilização de uma banda NIR, por uma banda SWIR. Este facto é consequência destes corpos absorverem mais luz SWIR levando a valores positivos mais altos de MNDWI.

O cálculo do NDWI recorre ao valor de reflectância TOA da banda verde e da banda NIR. Note-se que este valor, dado que as duas bandas têm uma resolução de dez metros, tem também essa resolução.

No caso de utilização do MNDWI o primeiro obstáculo com que se depara é o facto da banda SWIR ter uma resolução de vinte metros e não de dez metros tal como a banda verde. Este detalhe leva a que possa ser calculado este índice com duas resoluções diferentes (dez e vinte metros). Para a resolução de vinte metros basta fazer *upscale* à imagem relativa à banda verde para o dobro, contudo os resultados obtidos são piores. No caso de se pretender obter resultados mais precisos recorre-se à operação de *downscale* à banda 11. Para tal, são necessários métodos de *Pan-Sharpening* adaptando a resolução da banda SWIR à banda verde.

Após a obtenção de um destes índices é necessário saber qual o valor de *threshold* que distingue pixéis correspondentes a água dos restantes. Para tal, nesta aplicação recorreu-se ao método de *Otsu*. Apesar da importância do mapeamento de corpos de água, o elemento diferenciador deste trabalho foram as técnicas de *Pan-Sharpening* utilizadas. Se se pretendia a deteção de corpos de água de reduzida dimensão, era importante a utilização da resolução de 10 m.

3.2.2 Deteção de Perturbações Florestais

Em (Hamunyela, Reiche, Verbesselt, & Herold, 2017) utilizaram-se dados do *Landsat 7* e *Landsat 8* para o estudo de alterações na floresta. O principal problema com que se depara neste processo são as falsas deteções. Existem, para dar maior robustez aos algoritmos, três métodos para a confirmação de uma perturbação,

- Duas anomalias consecutivas;
- Duas anomalias consecutivas em conjugação com a magnitude de mudança;
- Recurso à característica de espaço-tempo.

A partir das imagens das plataformas utilizadas é calculado o NDVI. O primeiro método consiste na definição de um valor de *threshold* e comparando-o com a magnitude da alteração do NDVI é detetada uma perturbação na floresta se isto ocorrer duas vezes consecutivas. Trata-se de um método extremamente simples, contudo insuficiente para descobrir pequenas alterações (fogueiras domésticas, pequenas expansões agrícolas).

O método seguinte, para além da deteção de duas anomalias consecutivas, apoia-se também na magnitude da mudança para a confirmação de uma perturbação.

Finalmente, o último método utiliza as características de espaço-tempo para implementar um cubo de dados de NDVI. Este cubo é constituído por várias camadas temporais, sendo obtida a variabilidade espacial em cada uma delas. Foi definido um conjunto de características espaciais e temporais. Obtém-se então a probabilidade de perturbação num pixel quando ocorrem duas anomalias. Se este valor for superior ou igual a 0,5, é então descoberta uma perturbação no pixel.

Em (Hamunyela et al., 2017) foi definido que uma perturbação consiste na perda de floresta e a área em estudo é a *Unesco Kafa Biosphere Reserve* localizada na Etiópia. Aplicaram-se o segundo e o terceiro método, tendo-se concluído que este último revelou uma maior precisão.

3.2.3 Deteção de Mudanças e Geração de Imagens sem Lacunas

Em (Hermosilla, Wulder, White, Coops, & Hobart, 2015) é descrita uma aplicação de deteção de mudanças numa determinada zona de observação e de lacunas de dados nas imagens. A segunda funcionalidade mencionada consiste no preenchimento de pixéis cujos dados são inexistentes ou danificados por causa das condições de aquisição. A plataforma utilizada neste estudo foi o satélite *Landsat 7* e o respetivo sensor *ETM+*.

A resolução do problema das lacunas começa com a criação de compósitos dos *Best Available Pixel* (*BAP*). Seguidamente é derivada uma série temporal dos valores dos pixéis que pode ser utilizada para a identificação, descrição e quantificação de mudanças e perturbações na área de estudo. Contudo estas séries podem ser estáveis, o que não levanta obstáculos a esta análise, ou irregulares, resultado das alterações na área de cobertura. Neste último caso as variações podem também ser consequência de ruído na aquisição de dados ou valores anómalos, fruto das condições de obtenção das imagens (Nebulosidade, variações na luz solar, entre outros). Antes da deteção de mudanças é necessário então a atribuição de valores a estes pixéis.

A criação das imagens *BAP*, parte da geração de compósitos de valores de confiança. Inicia-se com a derivação de um conjunto de métricas a partir da deteção de mudanças espectrais. Estas são depois utilizadas para a atribuição de valores de confiança. Numa segunda parte são identificados os valores que fogem à normalidade e substituídos pelos valores de confiança. A partir da seleção dos *BAP* são então obtidas as imagens baseadas em píxeis. Existem os seguintes três métodos para o preenchimento das lacunas de dados:

- ***Single-year:*** etiquetação dos píxeis sem dados adequados como *no data*;
- ***Multi-year:*** utilização de píxeis de anos anteriores para preencher as lacunas de dados;
- ***Proxy composites:*** recorre-se a toda a informação espectral da série de píxeis substituindo os valores em falta pelos mais similares espectralmente no espaço e no tempo.

Estes procedimentos diferem no tratamento dos píxeis sem observações adequadas. No primeiro verifica-se que a área coberta irá diminuir. O segundo pode ter impactos negativos no estudo da vegetação (análise de tendências). O terceiro leva a melhores resultados preenchendo a imagem com base no padrão temporal e comportamento inter-anual. Esta última técnica segue então os seguintes passos:

1) geração das imagens *BAP*;
2) localização dos píxeis com valores anómalos nos compostos anuais de imagens *BAP*;
3) etiquetação dos mesmos com *no data*;
4) análise de *breakpoint* para identificação de alterações espectrais;

5) estabelecimento de limites de interpolação temporal para determinar os valores de confiança aplicáveis.

Na geração de imagens *BAP* é calculado então um *score* com base no sensor, data de aquisição, distância a nuvens e sombras de nuvens e opacidade atmosférica. O pixel com *score* mais alto é então selecionado. Através das séries de píxeis são então detetados os *outliers* e assinalados como *no data* criando-se assim as lacunas de dados. Seguidamente verifica-se se estes píxeis são resultado do ruído ou se se tratam efetivamente de alterações. Estas lacunas são também utilizadas para localizar áreas em que persistentemente faltam dados. Finalmente, através do *Normalized Burn Ratio* (*NBR*) são detetados os *breakpoints* (píxeis classificados como *no data* correspondentes a efetivas alterações).

O objeto de estudo deste trabalho foi a região florestal canadiana, a província do Saskatchewan no período de 2000 a 2010. Repara-se que dada a análise temporal recorreu-se a dados dos anos 1998 a 2012. Verificou-se que durante este período existia uma perda entre 0.5% a 31% de dados, sendo estes recuperados pelas metodologias anteriormente referidas. Após esta fase foram então classificadas as alterações e descritas pelos seguintes pontos:

- Anos de persistência;
- Ano da mudança;
- Magnitude;
- Taxa de recuperação.

3.3 *Machine Learning*

Após a aquisição de dados e pré-processamento dos mesmos, o próximo passo consiste na criação de conhecimento (gerar conclusões novas). Este estudo consiste na procura de alterações no terreno, o caso específico em que uma zona de vegetação é cortada, sendo substituída por solo. Em 5.1 é apresentado o efeito deste fenómeno na informação que foi obtida. Agora serão abordadas várias técnicas de *Machine Learning* e aplicações já existentes em *Remote Sensing* com recurso a estes mecanismos.

Em qualquer método têm de ser definidos os atributos e as classes que serão utilizadas. Os atributos consistem nas medidas e informações observadas e as classes nas possibilidades de classificação de cada exemplo. Neste caso os atributos são constituídos pelos valores de reflectância obtidos pelo *Sentinel 2* e pelos índices espectrais calculados a partir das bandas. Trata-se de um problema binário (apenas com duas possibilidades de classificação) em que se procura detetar a existência ou não de uma intervenção na FGC. Um detalhe importante é que a utilização de todos os atributos disponíveis não leva obrigatoriamente a um conhecimento melhor. Em primeiro lugar cada atributo acrescenta ao problema uma dimensão, logo em termos de visualização a capacidade do ser humano limita-se às três dimensões. Cada dimensão nova aumenta a complexidade da classificação. Importa perceber se existe algum ganho no processo de classificação ou se é uma informação desnecessária. Por exemplo, atributos que evidenciem o mesmo comportamento perante todas as situações, atributos muito correlacionados, ambos transmitem a mesma informação, sendo a utilização destes, em simultâneo, desnecessária. Em 5.1 são definidos os grupos de atributos que apresentam um comportamento similar.

As técnicas de classificação podem ser divididas de diferentes formas. Uma das possíveis diferenciações consiste na existência de um conjunto pré-classificado durante a fase de treino. Caso este exista trata-se de aprendizagem supervisionada, caso contrário, o agrupamento em classes é responsabilidade dos algoritmos escolhidos, sendo uma aprendizagem sem supervisão. Nesta dissertação os conjuntos já se encontram pré-classificados. No primeiro grupo referido é possível subdividi-lo consoante a previsão resulte em valores discretos ou valores contínuos. O primeiro caso categoriza informação através de técnicas de classificação, a segunda devolve um número real através de algoritmos de regressão. Note-se que existem métodos que podem devolver os dois tipos de resultados.

Por fim, é muito importante perceber se o classificador obtido apresentará resultados positivos quando apresentados casos desconhecidos. Para tal é necessário estimar o erro do mesmo o que também será abordado neste capítulo.

3.3.1 Algoritmos de *Machine Learning*

Nesta secção serão abordadas várias metodologias de classificação juntamente com aplicações das mesmas em *Remote Sensing*. Serão abordados algoritmos difusos, de regras e de funções. Também será feita a distinção entre a classificação de imagem e de objeto, algo que assume um papel importante neste estudo.

3.3.1.1 Fuzzy Information Fusion

Em (Mora et al., 2017) analisou-se as consequências da alteração de utilização do solo causada pelo retorno da população ao distrito de Mandimba, em Moçambique, em 1992. As imagens utilizadas na classificação provêm do *Landsat 5* e *Landsat 7* dos anos 1989, 2002 e 2005.

Nesta análise é então aplicado método *Fuzzy Information Fusion* (FIF) para a classificação e comparado com outros dois métodos, Árvores de Decisão (DT) e Redes Neuronais Artificiais (ANN). O FIF trata-se de um método de inferência difusa reforçado por operadores de agregação no raciocínio de inferência. Um sistema de inferência difusa (FIS) consiste num modelo de regras, sendo estas descritas por operadores lógicos, estabelecendo relações entre conjuntos difusos. Estas regras podem ser dadas por especialistas ou por todas as combinações possíveis das variáveis de entrada. São constituídas por antecedentes (entradas) e consequentes (saídas). Um FIS é caracterizado por três processos:

1. *Fuzzification* das variáveis de entrada, consiste na representação destas no intervalo [0,1] por todos os conjuntos difusos;
2. Definição das regras difusas;
3. Inferência do esquema de seleção para as saídas.

Em (Mora et al., 2017) a metodologia de classificação de terrenos através de FFIS assemelha-se à de um FIS, acrescentando o reforço por operadores de agregação, sendo a seguinte:

1. Classificação difusa da informação espectral;
2. Criação de um conjunto de regras para as classes de cobertura terrestre;
3. Avaliação das regras com operadores de agregação.

O passo 1 consiste na geração de histogramas para os dados de treino e criação de funções de associação através do ajuste de funções gaussianas aos grupos do histograma. No segundo passo definem-se sete regras, uma por cada banda espectral de entrada, para cada classe. O último passo traduz-se no teste das regras para quatro diferentes operadores de agregação para o esquema de inferência.

O FIF foi comparado com as DT e ANN. O recurso a DT resulta na divisão dos dados em conjuntos mais pequenos com características similares até serem o mais homogéneos possível. Uma DT é constituída por nós (raiz e interior) que consistem numa condição que define qual o próximo nó e em folhas, os nós terminais de decisão. As ANN caracterizam-se pelo treino de neurónios para a deteção de padrões no conjunto de treino e posterior aplicação a dados desconhecidos. Numa ANN formada por várias camadas, a de entrada e a de saída são determinadas pelos dados a serem analisados, enquanto que a camada escondida decorre de um processo de tentativa-erro. Os pontos mais negativos de uma ANN é o seu longo período de aprendizagem, pouca transparência do método de aprendizagem e a tendência de *overfitting*[9].

3.3.1.2 *Random Forest Classification*

Em (Tian et al., 2016) classificaram-se as zonas húmidas do rio Ertix no Norte de Xinjiang, China. Utilizaram-se dados do *Pléiade-1B* (pertencente ao CNES) e do *Landsat 8*. O algoritmo de classificação utilizado foi o *Random Forest Classification* (RFC) e foi comparado com as ANN e as *Support Vector Machines* (SVM).

O RFC resulta da geração de várias árvores através de três variáveis de entrada. Cada uma das árvores depende dos valores de um vetor aleatório amostrado de forma independente com a mesma distribuição para todas as árvores da floresta (conjunto de todas as árvores) e independente dos anteriores. A classificação é feita de forma conjunta, sendo decidida pela maioria das árvores geradas. A equação **(14)** mostra a forma de cálculo do resultado da classificação:

$$H(x) = argmax_Y \sum_{i=i}^{k} I(h_i(X, \theta_k) = Y) \qquad \textbf{(14)}$$

em que $h_i(X, \theta_k)$ representa a classificação de cada árvore, $I(.)$ é a função *indicator* Y a variável de saída, θ_k um vetor aleatório para a geração das árvores e X o vetor de entrada.

Esta técnica foi comparada com as ANN e as SVM, a primeira já explicada em 3.3.1.1 e a segunda, de forma muito simplista, consiste na procura de um hiperplano que separe as regiões de classificação. Os resultados da RFC foram bastante superiores.

[9] Adaptação excessiva do classificador à informação conhecida, levando a uma classificação muito precisa desta, mas a um fraco desempenho perante informação desconhecida.

3.3.1.3 Fully Convolutional Network

Em (Fu, Liu, Zhou, Sun, & Zhang, 2017) aplicou-se o método *Fully Convolutional Network* (FCN) para a classificação de imagens de *Remote Sensing* de alta resolução (superior ou igual a 2 metros). O FCN é uma variante do método *Convolutional Neural Networks* (CNN).

Começando por apresentar o CNN, segundo (Längkvist, Kiselev, Alirezaie, & Loutfi, 2016), esta técnica é bastante útil no processamento de sinais naturais, dado o aproveitamento das suas propriedades estabelecendo conexões locais e pesos ligados sendo mais fácil o seu treino. Uma *Deep* CNN é constituída por uma pilha de CNN de baixo nível, implementando estruturas de mais alto nível. A última camada pode ser constituída por uma ou mais camadas totalmente ligadas, podendo ser adicionado à última camada um classificador. Cada camada CNN executa três passos: convolução; ativação não-linear e agrupamento. A convolução é feita para k mapas de características, sendo o seu cálculo apresentado na equação **(15)**,

$$f_{ij}^{k} = \sigma\left(\sum_{c}\sum_{a=0}^{n-1}\sum_{b=0}^{n-1}\omega_{abc}^{k}x_{i+a,j+b}^{c}\right) \quad \textbf{(15)}$$

sendo σ a função de ativação não linear, ω^{k} o filtro aplicado de tamanho $n \times n$ e a região local x de tamanho $m \times m$. A função de ativação mais usual é a tangente hiperbólica.

O FCN diferencia-se nas últimas camadas em que substitui as camadas totalmente ligadas por camadas totalmente convolusionadas. Podem ser aplicadas várias fórmulas de convolução, sendo que em (Fu et al., 2017) é utilizada a convolução *Atrous* descrita na equação **(16)**,

$$y[i] = \sum_{k=1}^{K} x[i + r.k]\,\omega[k] \quad \textbf{(16)}$$

sendo $x[i]$ o sinal de entrada e ω o *kernel* convolucional. Esta variante apresenta uma maior facilidade na implementação, dado que permite imagens de vários tamanhos e a manutenção de uma estrutura de duas dimensões, também reduzindo a complexidade. O FCN é também mais preciso e tem menos tempo de computação.

3.3.1.4 Redes Neuronais Artificiais

O último algoritmo de classificação que será abordado são as Redes Neuronais Artificiais. O objetivo de uma ANN consiste na imitação do método de aprendizagem do cérebro humano (refere-se aqui ao método que atualmente se pensa que este tenha). Para tal, a estrutura de uma ANN assemelha-se à do cérebro. Diferencia-se claro, porque apenas é treinada para a resolução de um conjunto limitado de problemas, sendo que não aprende a executar tarefas completamente novas. Trata-se de uma técnica de aprendizagem supervisionada e quanto aos valores de saída pode-se enquadrar em qualquer um dos grupos (classificação ou regressão).

Uma ANN, tal como apresentado na Figura 3.1, é constituída por várias camadas, sendo duas delas referentes às entradas e saídas e as outras constituintes do grupo de camadas escondidas. Não existem regras para a estruturação de uma ANN, contudo existem algumas normas habituais (não conduzem necessariamente a melhores resultados). A cada atributo dedica-se, habitualmente, um neurónio na camada de entrada. Cada classe é, normalmente,

representada por um neurónio na camada de saída (no caso de uma classificação binária, basta apenas um neurónio). O número de camadas escondidas e o número de neurónios entre cada uma delas pode variar. Os neurónios encontram-se completamente ligados, ou seja, a saída de um neurónio de uma camada está conectada com todos os neurónios da camada seguinte. Note-se que em determinadas arquiteturas um neurónio pode estar ligado não só à camada seguinte, mas também às posteriores, contudo a ligação não ocorre no sentido inverso. Para além disso, também pode ser utilizado um neurónio de *bias* nas diferentes camadas, permitindo uma melhor aprendizagem por parte da ANN.

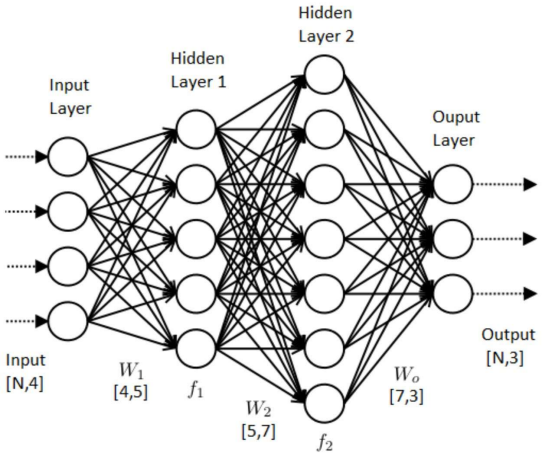


Figura 3.1: Representação gráfica de uma Rede Neuronal com duas camadas escondidas. Os neurónios são representados pelas circunferências e as ligações pelas setas. (fonte: https://medium.com/).

Cada neurónio computa o seu valor de saída consoante as suas entradas, peso associado (onde reside o conhecimento) e função de ativação. O valor que a função de ativação recebe é dado pela equação **(17)**, em que N corresponde ao neurónio em causa, i é relativo à entrada, w o peso associado à entrada e x o valor de entrada. Para esta função pode-se recorrer à função escalão, à função sinal, à função linear por partes, à função identidade, à função sigmoidal, entre outras.

$$Out_N = \sum_{i=1}^{k} w_{Ni} x_i \qquad \textbf{(17)}$$

Um dos pontos mais importantes e diferenciadores das ANN é o método de aprendizagem. Utilizando a retro propagação dos erros este algoritmo aprende com os próprios erros, adaptando os pesos da ANN ao problema. Um dos métodos utilizados é o gradiente descendente, apresentado em (E. Rumelhart, E. Hinton, & J. Williams, 1986). O termo de correção dos pesos é dado pela equação **(18)** em que E é o erro total dado pela equação **(19)** em que y é o valor obtido e d; t corresponde à época[10] de treino presente; α é um fator que decai

[10] Uma época corresponde a uma etapa de treino em que todo o conjunto é percorrido.

exponencialmente encontrando-se no intervalo $[0,1]$ que define a contribuição do gradiente atual na correção e ε a proporção do erro acumulado em uma época.

$$\Delta w(t) = -\varepsilon \frac{dE}{dw(t)} + \alpha \Delta w(t-1) \quad \textbf{(18)}$$

$$E = \frac{1}{2} \sum_i \sum_j \left(y_{i,j} - d_{i,j}\right)^2 \quad \textbf{(19)}$$

O segundo termo da equação **(18)** tem como responsabilidade a aceleração do processo de aprendizagem, o primeiro termo representa a correção pelo gradiente descendente. Note-se o seguinte detalhe, muitos dos métodos de retro propagação do erro efetuam a alteração dos pesos logo após a análise de um exemplo, neste procedimento essa correção só é realizada ao final de cada época.

A grande vantagem destes algoritmos é a capacidade que têm de encontrar padrões na informação a eles disponibilizada e a adaptação aos problemas em questão. Contudo esta mesma capacidade de adaptação, também característica do cérebro humano, tem as suas desvantagens. As ANN têm tendência a ter problemas de *overfiiting*. Outro fator menos positivo destes métodos é o problema dos mínimos locais. Durante o processo de aprendizagem por vezes as correções levam a piores classificadores, quando isto acontece a ANN, como não melhorou o resultado da classificação, interrompe o processo de aprendizagem, julgando ter chegado à solução mais adequada. Contudo por vezes se o treino continuasse seriam encontradas soluções melhores até atingir o mínimo global. Apesar desta desvantagem, muitas vezes os mínimos locais encontrados têm desempenhos semelhantes ao mínimo global, como é dito em (E. Rumelhart et al., 1986).

Em (Ahmed & Al Noman, 2015) é apresentada uma aplicação de ANN no âmbito de *Remote Sensing* em que utilizava como plataforma o *Quickbird* que fornece imagens em quatro bandas espectrais. A zona de estudo foi Pequim tendo como objetivo a classificação do terreno nas seguintes classes: Estrada; Água; Vegetação; Urbano e terra vaga. Foi então projetada uma ANN com uma camada escondida constituída por 24 neurónios. Os resultados obtidos foram positivos sendo este facto atribuído à capacidade das ANN terem uma dinâmica de adaptação aos problemas. Contudo existe um detalhe interessante de referir. Apesar de parecer que quanto mais camadas escondidas tivermos, melhor os resultados, isto não é verdade. Após a primeira camada escondida começam as surgir os *vanishing gradients*, isto é, as correções aos pesos dos neurónios das camadas mais próximas da entrada são de tal modo pequenas que não acrescentam conhecimento ao algoritmo.

3.3.2 Object-Based Classification

Em (Lopes, Fauvel, Girard, & Sheeren, 2017) classificaram-se pastagens do sudoeste francês através de imagens satélite *Formosat-2*. São adquiridas imagens em quatro bandas (R, G, B e NIR) com uma resolução de 8 metros. Recorre-se a SVM para a classificação, mas o ponto de diferenciação é o facto de se observarem objetos e não apenas pixéis. Para tal são escolhidas as pastagens que vão ser estudadas. São modeladas através de distribuições gaussianas, caracterizando o objeto por esta distribuição, pela média dos seus pixéis, covariância e variável de resposta. A classificação é feita através de funções de semelhança para a

comparação entre distribuições gaussianas, *mean map kernels,* sendo que foi utilizado *α-Gaussian Mean Kernel* apresentado.

3.3.3 Estimação do Erro

A estimação é uma parte crucial da aprendizagem automática. É nesta fase que se avalia a qualidade do classificador e se prevê a sua eficiência quando apresentado a exemplos desconhecidos. Existem vários métodos para efetuar esta avaliação, o mais simples de todos é o cálculo do erro por resubstituição. Porém, trata-se de uma estimativa demasiado otimista pois considera-se que o classificador já conhece todo o universo possível de exemplos. Outra forma já mais realista consiste em dividir o conjunto de dados num grupo de teste e conjunto de treino. Mas o problema reside no modo como é feita esta divisão sendo que pode ser obtida uma estimativa muito favorável ou o contrário. Um método mais realista é por exemplo a validação cruzada, em que o conjunto de dados é dividido num número definido de grupos. O classificador treina com todos grupos exceto um que está reservado para o teste. O processo repete-se até todos os conjuntos terem sido utilizados para o teste. O erro é a média dos erros de todos os classificadores gerados. Quando os conjuntos de dados são muito reduzidos muitas vezes é utilizado o caso específico do *Leave-one-out*, em que o conjunto de teste é constituído apenas por um exemplo.

Não só é importante estimar o erro do sistema de decisão, mas também é importante perceber que tipos de erros e por que razão ocorrem. Para tal o primeiro mecanismo disponível é a construção da matriz de confusão. Esta matriz permite observar que confusões foram feitas pelo algoritmo, que classe foi atribuída aos exemplos mal classificados. Para além disto, certos erros podem ser mais graves do que outros e o que se procura minimizar não é o erro relativo do classificador, mas a má classificação de uma determinada classe. Existem várias métricas que podem ser utilizadas com o objetivo de tornar mais sensível o classificador a um determinado tipo de erros. Por exemplo o *Recall* oferece uma noção dos falsos negativos (o significado de um falso negativo depende do problema em questão), quanto menor o seu valor mais erros de classificação ocorrem na classe positiva. A Precisão por sua vez pretende transmitir dentro dos casos classificados como positivos quantos o são realmente. O *F1-Score* é uma métrica que tenta balancear as duas métricas anteriores.

Outro ponto que também é importante numa estimação realista do erro é a divisão prévia do conjunto de dados em três grupos: treino, validação e teste. No primeiro é feito o treino, recorrendo a técnicas como a validação cruzada e testado no segundo. Repete-se o processo até serem alcançados resultados satisfatórios. Note-se que apesar do conjunto de validação nunca ser utilizado no processo de treino, como as alterações em função do mesmo, indiretamente ele entra no processo de treino. Por fim, existe o conjunto de teste no qual deve ser testado uma única vez o classificador final.

A fase de avaliação do desempenho do classificador pode ser muito sensível. A execução desta tarefa implica um nível muito alto de seriedade por parte do seu executor. Por mais métodos que existam (sendo já a validação cruzada um método muito robusto) é sempre possível adaptar a escolha dos conjuntos de forma a mascarar os resultados. Nunca a ética pode ser posta de parte em troca de um melhor resultado.

3.4 Sumário

Antes de partir para a aplicação de métodos de *Machine Learning*, existe muito conhecimento que pode ser extraído. A análise de dados por aplicação de índices espectrais é bastante útil para a perceção das dinâmicas do terreno observado na imagem. Para além disso, a utilização destes como atributos para a classificação revela-se importante. A utilização dos valores de reflectância sem qualquer tipo de tratamento fornece uma menor informação para a resolução do problema.

Não existe uma técnica de classificação que possa ser considerada ideal para todos os problemas. Cada caso tem de ser estudado individualmente e ser escolhido o método que se julgue levar a melhores resultados. Nesta dissertação escolheu-se como método as ANN. Apesar de não se procurar classificar a vegetação como ocorre em (Ahmed & Al Noman, 2015), o objetivo de deteção de alterações (neste caso a passagem de uma zona de vegetação a solo sem vegetação) também beneficiará da capacidade de adaptação das ANN ao problema em questão.

Sendo que o objetivo é a classificação de imagens é necessário definir se serão classificados pixéis ou objetos, como apresentado em 3.3.2. Como neste estudo pretende-se detetar a ocorrência de intervenções na faixa, serão classificados objetos. A estes será associado o valor médio dos atributos como será visto em 5.1.

No que toca à avaliação do classificador serão feitas as estimações com recurso à validação cruzada. Quanto às métricas será apresentado o erro relativo, a matriz de confusão e a partir desta calculado o *Recall* (considerandos os falsos negativos a deteção errada de ocorrência de corte). Também será observada a precisão do classificador.

4 Trabalho Desenvolvido

Neste capítulo é apresentado todo o trabalho desenvolvido. Começam por ser apresentadas as áreas de estudo, seguidamente é dito como é feita a obtenção de dados e partir destes a extração de informação, abordando a correção da georreferenciação. Depois disto segue-se a fase de estudo e seleção de atributos para a classificação, geração dos conjuntos de dados de treino e teste e o dimensionamento da ANN. Por fim, é apresentado o *plugin para QGIS* implementado para a análise do NDVI. Na Figura 4.1 é apresentado o diagrama de tarefas desde a obtenção de dados até à deteção de intervenções. Todas as fases pelas quais este processo passa serão explicadas ao longo deste capítulo.

4.1 Áreas de Estudo

Nesta dissertação foram definidas duas áreas de estudo. A primeira consistiu no conjunto de FGC de Serra de Aire e Candeeiros, e a segunda área, perto da Marisol (Almada), é uma zona arborizada com eucaliptos e com uma linha de alta tensão. Esta última apesar de não fazer parte da RPFGC encontra-se intervencionada para reduzir a biomassa acumulada por baixo da linha de alta tensão, criando uma faixa com uma largura de aproximadamente 40m. A localização de ambas pode ser observada na Figura 4.2. O período de estudo são os anos de 2017 e 2018 (note-se que no início de 2017 apenas havia um satélite da missão em órbita, sendo que a frequência das observações era menor).

Note-se que apesar de a RPFGC ter planeado um conjunto bastante grande de FGC, infelizmente a grande maioria delas ainda não se encontra instalada por falta de financiamento. Para além disso, as datas em que foram executadas operações de manutenção não estão disponíveis, tendo de ser obtidas junto do ICNF e nem sempre existem registos destes trabalhos, sendo esta uma das grandes dificuldades deste estudo.

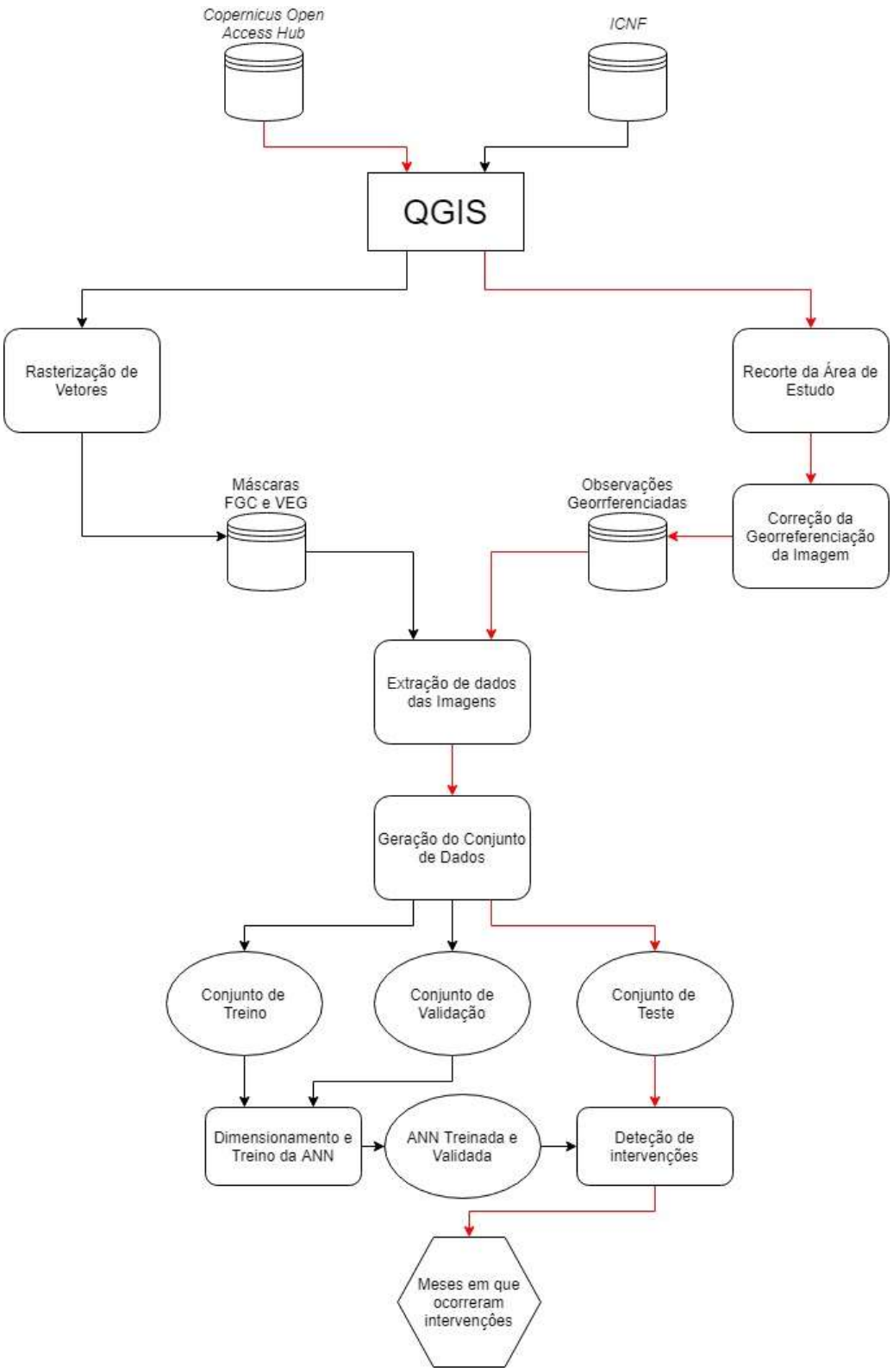


Figura 4.1: Esquema de execução desde a obtenção de dados até deteção dos meses relativos a intervenções.

Segundo o ICNF a vegetação de Serra de Aire é constituída por carvalhais de carvalho-cerquinho e carvalho-negral. Existem também pequenas zonas de azinheira, sobreiro, ulmeiro e castanheiros. Por fim, devido à ação humana surgiram várias zonas de vegetação espontânea (mato constituído por áreas arbustivas de carrasco e subarbustivas de alecrim) e vegetação não espontânea, sendo principalmente caracterizada por oliveiras. Quanto à Marisol trata-se de uma zona de muito menor complexidade sendo caracterizada por eucaliptos e alguns pinheiros.

A única operação de manutenção que se verificou em Serra de Aire e Candeeiros, durante o período de estudo, ocorreu em 27 de junho de 2017. Quanto à Marisol o corte terá sido feito em março de 2017, porém este corte é detetado pelos métodos descritos em 1.2. A zona da Marisol foi utilizada como zona de teste para os algoritmos desenvolvidos a partir da observação da zona de Serra de Aire e Candeeiros.

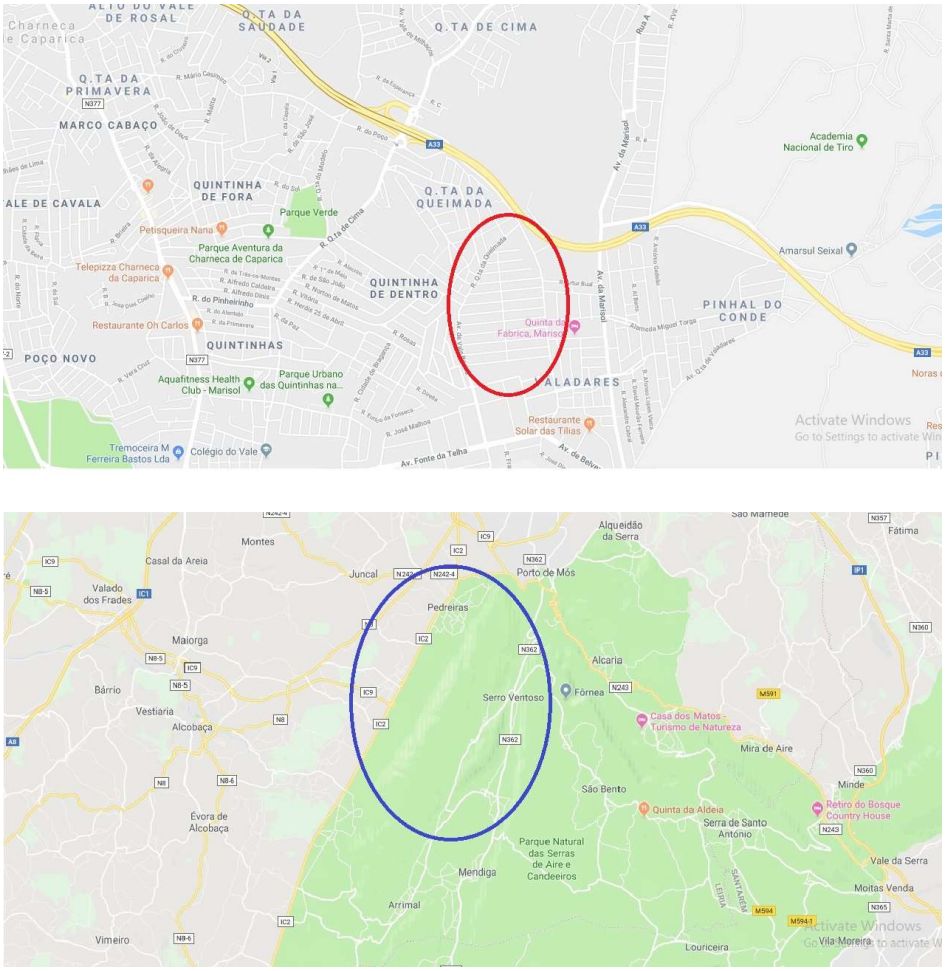


Figura 4.2: Localização das Áreas de Estudo. Em cima a vermelho Marisol, em baixo a azul Serra de Aire e Candeeiros.

4.2 Obtenção das Observações

Os dados utilizados neste estudo foram obtidos a partir do *Copernicus Open Access Hub* (https://scihub.copernicus.eu/). Obtiveram-se 265 observações correspondentes a produtos do nível 1C. Estas são relativas aos anos de 2017 e 2018 para as áreas de estudo. Destes dados, com recurso à ferramenta de pré-visualização disponibilizada pela ESA (no mesmo sítio) foram eliminadas aquelas que se encontravam inutilizadas em resultado das condições atmosféricas. Porém, algumas das imagens que passaram neste crivo ainda seriam suscetíveis de ser eliminadas. A verificação única e exclusiva do nível de nebulosidade disponibilizado por esta ferramenta pode eliminar observações que poderiam ser utilizadas. Repare-se que as nuvens presentes na imagem podem não afetar a região em estudo, podendo assim ser utilizadas. Observando na Figura 4.3 constata-se que as nuvens se encontram sobre o oceano estando ainda afastadas das zonas terrestres. Como consequência será realizada uma nova verificação antes da extração de dados. Esta verificação não é feita de forma automática, mas com recurso ao QGIS.

Dado o elevado número de observações (note-se que a cada uma das 265 observações correspondem 13 bandas) o processo de carregamento de imagens foi automatizado com recurso à integração de *Python* no SIG utilizado. São então definidas as bandas pretendidas (sendo que para esta avaliação recorreu-se à banda 4) e a pasta onde se encontram as observações. São então apresentadas as imagens no visualizador do QGIS. Seguidamente verificaram-se quais as observações que efetivamente não poderiam ser utilizadas. Do total foram eliminadas 78 observações, sendo o mês de fevereiro o mais afetado.

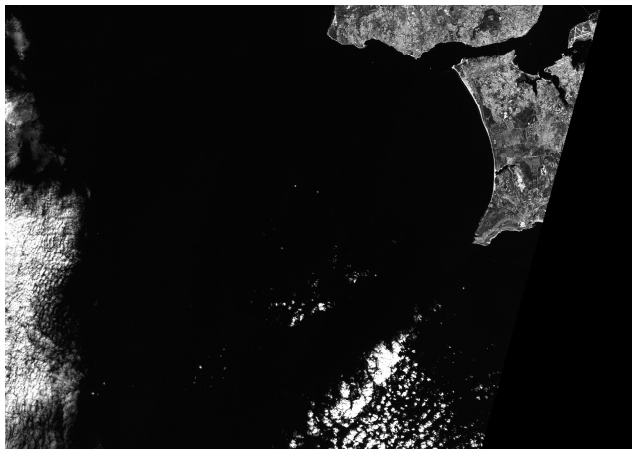


Figura 4.3: Observação não eliminada pela nebulosidade (08/01/2018 relativa a Marisol).

4.3 Obtenção das FGC

Após a aquisição de dados foi necessário obter a localização das FGC. A RPFGC encontra-se disponível ao público no sítio do ICNF (http://www2.icnf.pt) na forma de um ficheiro vetorial. Os atributos do ficheiro apresentados em (DPFVAP, 2014) não correspondem

totalmente à realidade. Para além da localização espacial das FGC também é fornecida a informação sobre qual a entidade responsável; a prioridade relativa à FGC e um campo indicativo do estado da FGC (se já se encontra instalada, se necessita de manutenção, ou se está instalada e operacional). Contudo, para várias FGC estes campos não se encontram preenchidos.

O primeiro contratempo deste estudo prendeu-se com a ausência de informação sobre as datas em que ocorreram as intervenções. Estas só foram obtidas entrando em contacto com o ICNF. Para além disso como referido em 1.2 só uma percentagem muito reduzida do plano da RPFGC se encontra já instalado.

Após identificadas as áreas de estudo em 4.1 foi preciso extrair do ficheiro vetorial disponibilizado pelo ICNF as FGC relativas às áreas de estudo desta dissertação. Para tal utilizaram-se duas ferramentas neste processo. Um *plugin* desenvolvido para o QGIS, *GEarthView*, que permite a sincronização entre o que se está a visualizar no QGIS com a aplicação *Google Earth* (GE). Permite a sobreposição de camadas do QGIS no GE e a marcação de pontos de um para o outro. Sendo que a identificação das FGC utilizada pelo ICNF não é a mesma usada no ficheiro vetorial (quando é indicado que a faixa pertence a Serra de Aire e Candeeiros nenhum dos campos do vetor contém esta informação), a utilização do GE permite localizar as FGC de forma mais simples e eficiente. A segunda ferramenta, já incluída de origem no QGIS (ferramentas dedicadas à *Attribute Table*), permite remover elementos dos vetores com recurso a instruções simples. Por exemplo, permitiu eliminar todas a FGC que não estão sobre a alçada do ICNF. Estas duas ferramentas permitem a extração das FGC em análise a partir do ficheiro disponibilizado pelo ICNF que inclui todas a FGC.

Para Serra de Aire e Candeeiros foram então selecionadas quatro FGCs, juntamente com três zonas de vegetação próximas destes locais para efeitos comparativos, visando a deteção de alterações (Figura 4.4). Note-se que não foi selecionada a FGC004, dado que se tratou de uma região em que visualmente não se identificou nenhum corte, e que inquirido o ICNF sobre esta questão foi transmitido que se tratava de uma FGC onde não ocorreu intervenção na data definida.

No caso da Marisol, sendo um caso de estudo fora das RPFGC, definiu-se uma FGC e uma zona de vegetação, sendo que ambos os vetores foram criados com recurso à ferramenta de criação vetorial no QGIS. São apresentados na Figura 4.5.

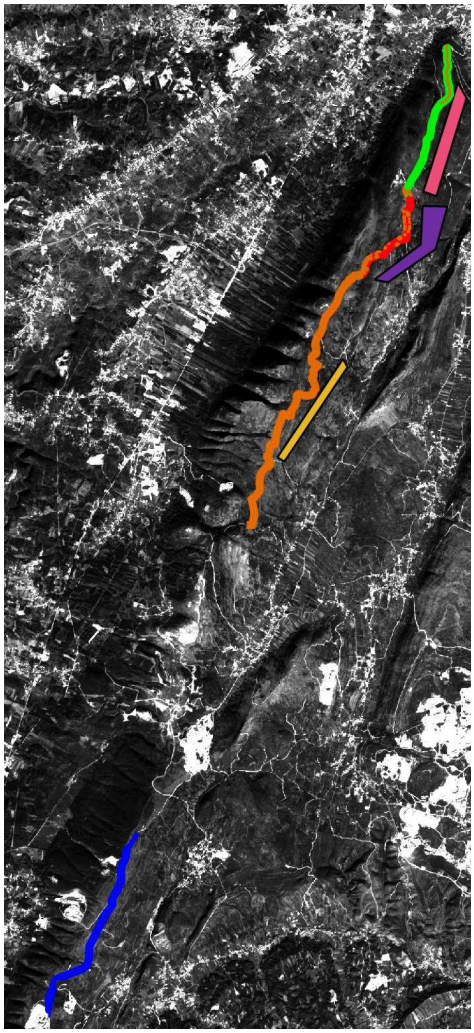


Figura 4.4: Zonas de Estudo de Serra de Aire e Candeeiros (Banda 4 - 15/01/2017): FGC001 (Laranja); FGC002 (Verde); FGC003 (Vermelho); FGC005 (Azul); VEG001 (Amarelo); VEG002 (Rosa); VEG003 (Roxo).

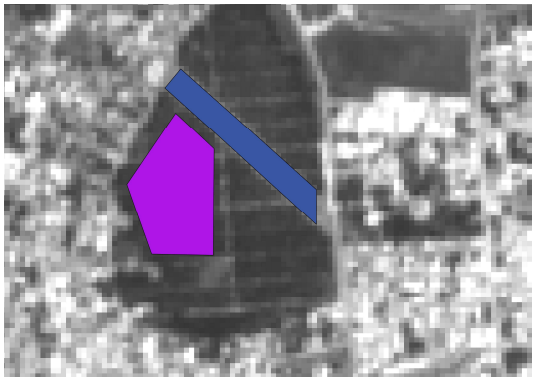


Figura 4.5: Zonas de Estudo de Marisol (Banda 4 – 10/05/2018): FGC001 (Azul); VEG001 (Roxo).

4.4 Registo de Imagem

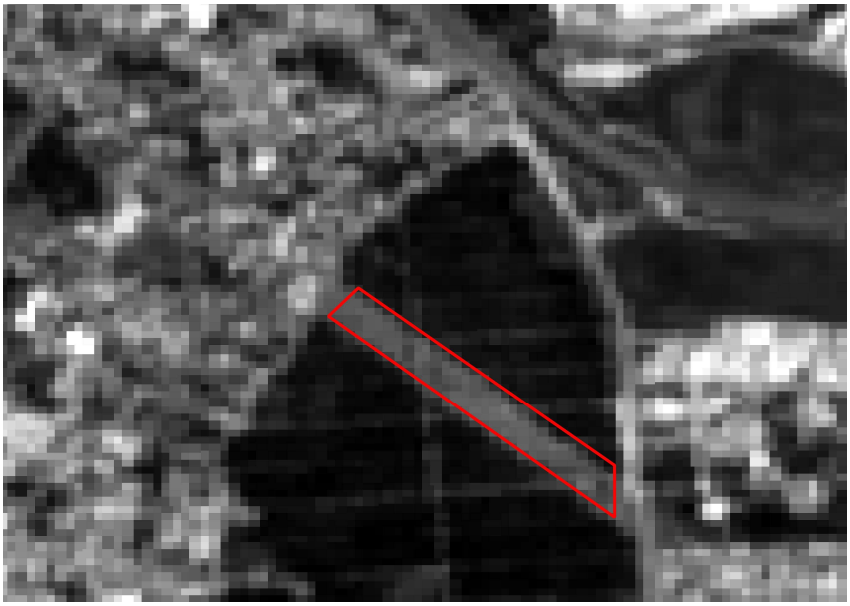
Um dos processos associados ao tratamento de imagens de satélite consiste na sua georreferenciação, ou seja, na associação de coordenadas geográficas aos pixéis da imagem. Contudo, apesar dos algoritmos de correção através da GRI (secção 2.3.5.2) e de calibrações regulares efetuadas pela ESA as imagens do Sentinel-2 de nível 1-C ainda podem conter pequenos desvios de georreferenciação (Clerc & MPC Team, 2019). Estes erros muito esporadicamente excedem os 1,5 pixéis, mas são frequentes ao nível do sub-píxel. Como pode ser observado na Figura 4.6, existem pequenos desvios nas observações relativamente à mesma referência, sendo neste caso um erro de 0,29 pixéis para baixo e 0,08 pixéis para baixo.

O facto de esta dissertação se concentrar no estudo de FGC, as quais têm um número reduzido de pixéis de largura, conduz a que o erro de georreferenciação possa ser bastante prejudicial na sua análise. Como as FGC se encontram normalmente junto de áreas de grande vegetação, caso ocorra um desvio de georreferenciação em relação à imagem de referência, pode ser incluída na FGC uma pequena área de vegetação, conduzindo a uma avaliação errada sobre o seu estado de conservação.

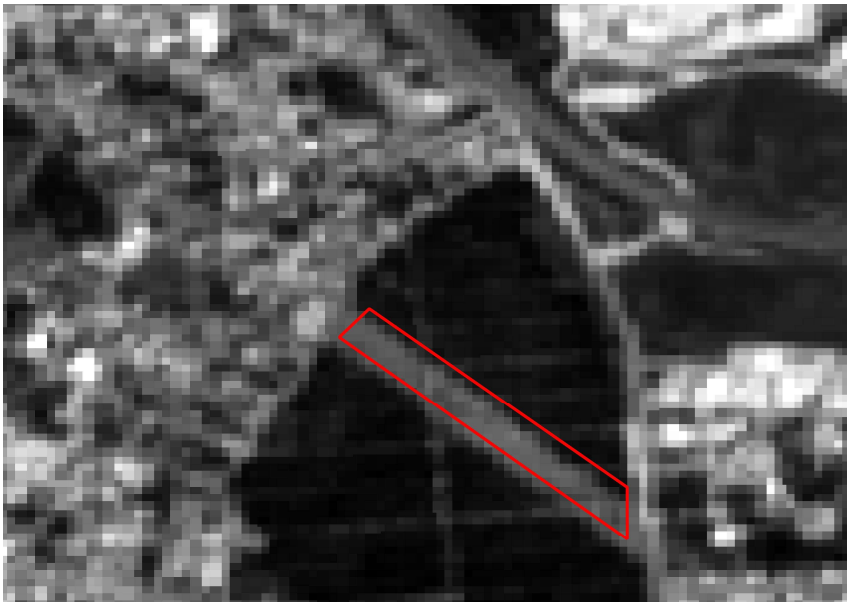
Torna-se imperativo uma etapa de pré-processamento das imagens para a correção destes desvios usando um algoritmo de registo de imagem. Neste estudo utilizou-se o algoritmo *Single-step DFT* proposto em (Guizar-sicairos, Thurman, & Fienup, 2008). Este algoritmo baseia-se no cálculo do valor máximo da correlação-cruzada entre as imagens através da *Fast Fourier Transform*. Este passo permite definir o ponto inicial de procura de pontos com correlações-cruzadas semelhantes. Seguidamente, com recurso ao método da *Discrete Fourier Transform* a duas dimensões é calculada a correlação cruzada *upsampled*[11] numa vizinhança de 1.5 pixéis em torno do ponto definido no passo anterior. Este valor é utilizado para procurar a correlação máxima entre as imagens. O *upsample* é que permite uma correção fina (ao nível do sub-pixel) da imagem. Note-se que apenas são encontrados erros de translação, não podendo ser usado este algoritmo para encontrar desvios rotacionais. O principal ganho deste método em relação ao que já existia, é no tempo de processamento, que é diminui em relação a técnicas anteriores.

Com os desvios calculados, efetuou-se a correção da georreferenciação das imagens. Note-se que apenas se utiliza a zona em análise, não sendo esta operação aplicada à imagem Sentinel-2 completa. Este facto prende-se com o elevado tempo de processamento que seria necessário para processar a imagem completa, sendo que os resultados obtidos não trazem ganhos a este processo. O processo de recorte da zona de interesse foi automatizado através de um *script* em *Python*, que recebendo a zona que se pretende analisar, recorta essa zona de todas as imagens *Sentinel 2* que estão a ser analisadas.

[11] O *upsample* é um método de processamento de sinal que aumenta a taxa de amostragem artificialmente.



a)



b)

Figura 4.6: Desvios geográficos das observações *Sentinel 2*; a) Banda 4 - 15/05/2018;
b) Banda 4 - 10/05/2018.

4.5 Extração de Dados

Depois das correções da geolocalização das imagens (claro que mesmo depois desta etapa, podem existir erros) inicia-se a fase de extração de dados. Nesta fase foi desenvolvido

um *plugin* para o QGIS para o cálculo do NDVI que será abordado em 4.9. Para otimizar o processo de extração de dados a partir dos ficheiros *raster* relativos às observações em conjunto com os vetores do ICNF foram desenvolvidos *scripts* em *Python* que executam automaticamente o procedimento, que será descrito nesta secção.

Em primeiro lugar foi necessária a conversão dos ficheiros vetoriais das FGC em ficheiros *raster*. Através da ferramenta de rasterização de vetores oferecida pelo QGIS são geradas máscaras binárias das FGC que serão utilizadas no processo de extração de dados. São preferidos os ficheiros *raster* dado que as imagens *Sentinel 2* pertencem a este tipo de dados, facilitando todas as operações necessárias. A partir das imagens já recortadas e georreferenciadas através de *scripts,* ou no caso do NDVI do *plugin* desenvolvido, são obtidos os valores médios de cada observação das bandas e dos índices espectrais. Como resultado é gerado um ficheiro *Comma-separated values* (CSV) para cada FGC, contendo uma listagem de todas as observações com data/hora, e valores médios, desvio padrões, máximos e mínimos de cada banda ou índice espectral em análise.

A associação do valor médio à região conduz a uma menor precisão nos resultados. Para medir a qualidade desta substituição foi calculada a flutuação relativa. Os valores obtidos estão presentes na Tabela 4.1. Segundo (Casquilho & Teixeira, 2011) este valor deve ser menor que 1 para a aproximação poder ser feita. Verifica-se que o valor médio deste parâmetro é de 0,19. Verificou-se também um detalhe interessante, a flutuação relativa nas zonas de vegetação é menor que nas zonas de FGC, sendo respetivamente 0,16 e 0,21. Este fenómeno surge porque no ficheiro vetorial contendo as FGC por vezes o desenho destas encontra-se um pouco deslocado incluindo também pequenas regiões de vegetação, como se pode ver na Figura 4.7 Aqui é possível verificar que o vetor contém zonas de vegetação nos limites laterais da FGC, sendo que a sua assinatura espectral é diferente da do solo levando a uma distribuição dos valores mais distante da média.

Figura 4.7: Exemplo de desvio entre desenho da FGC (linha a vermelho) e FGC real (imagem).

Tabela 4.1: Valores de Flutuação Relativa obtidos para os dados analisados.

		FLUTUAÇÃO RELATIVA								
	Zona	B02	B03	B04	B05	B07	B08	B8A	B11	B12
2017	FGC001	0.12	0.16	0.23	0.16	0.12	0.13	0.12	0.16	0.22
	FGC002	0.14	0.19	0.29	0.19	0.14	0.16	0.13	0.19	0.26
	FGC003	0.18	0.24	0.34	0.23	0.16	0.18	0.15	0.19	0.25
	FGC005	0.25	0.34	0.47	0.33	0.22	0.23	0.20	0.22	0.29
	VEG001	0.09	0.11	0.20	0.12	0.10	0.11	0.10	0.19	0.26
	VEG002	0.07	0.09	0.19	0.10	0.12	0.13	0.12	0.18	0.30
	VEG003	0.14	0.18	0.32	0.17	0.12	0.14	0.12	0.21	0.34
2018	FGC001	0.11	0.14	0.22	0.15	0.13	0.14	0.12	0.16	0.22
	FGC002	0.15	0.19	0.31	0.20	0.14	0.16	0.13	0.19	0.28
	FGC003	0.17	0.23	0.35	0.23	0.15	0.17	0.14	0.20	0.28
	FGC005	0.23	0.31	0.44	0.31	0.21	0.22	0.19	0.21	0.26
	VEG001	0.08	0.10	0.19	0.11	0.13	0.13	0.12	0.17	0.25
	VEG002	0.06	0.08	0.18	0.11	0.11	0.12	0.11	0.18	0.30
	VEG003	0.12	0.16	0.31	0.17	0.12	0.13	0.11	0.22	0.34

Uma vez que a periodicidade das observações de *Sentinel 2* desprovidas de nuvens não é constante e que no início do período de estudo (início de 2017) houve uma menor periodicidade (cerca de 1 observação por mês), houve necessidade de obter um vetor de observações com periodicidade constante para suavizar a evolução temporal dos dados e facilitar a análise de alterações. Neste estudo foi definido condensar os dados das observações para uma periodicidade mensal, sendo que em estudos futuros que incluam períodos com os dois satélites *Sentinel 2* em funcionamento deverá ser possível aumentar para uma frequência quinzenal. O método usado para fazer a condensação dos valores foi usar valor médio mensal. Este processo recorre mais uma vez ao *Python*. No caso de meses sem observações é calculado o valor médio entre a última observação do mês anterior e a primeira observação do mês seguinte. A flutuação média relativa é de 0.04 sendo novamente uma aproximação adequada (neste caso não são apresentados os valores individuais).

Para o tratamento de dados também foi necessária a normalização das medições da reflectância nas diversas bandas. O método utilizado foi o *Min-Max* que redimensiona os valores de x para uma escala no intervalo $[a, b]$ sendo feita pela equação **(20)**.

$$x' = a + \frac{(x - x_{min})(b - a)}{x_{max} - x_{min}} \qquad \textbf{(20)}$$

No caso *Sentinel 2* os valores máximos e mínimo de x serão os relativos à resolução radiométrica das bandas referida em 2.3.2. Na Figura 4.8 é apresentada a aplicação dos diversos índices espectrais estudados em 3.1 a uma região.

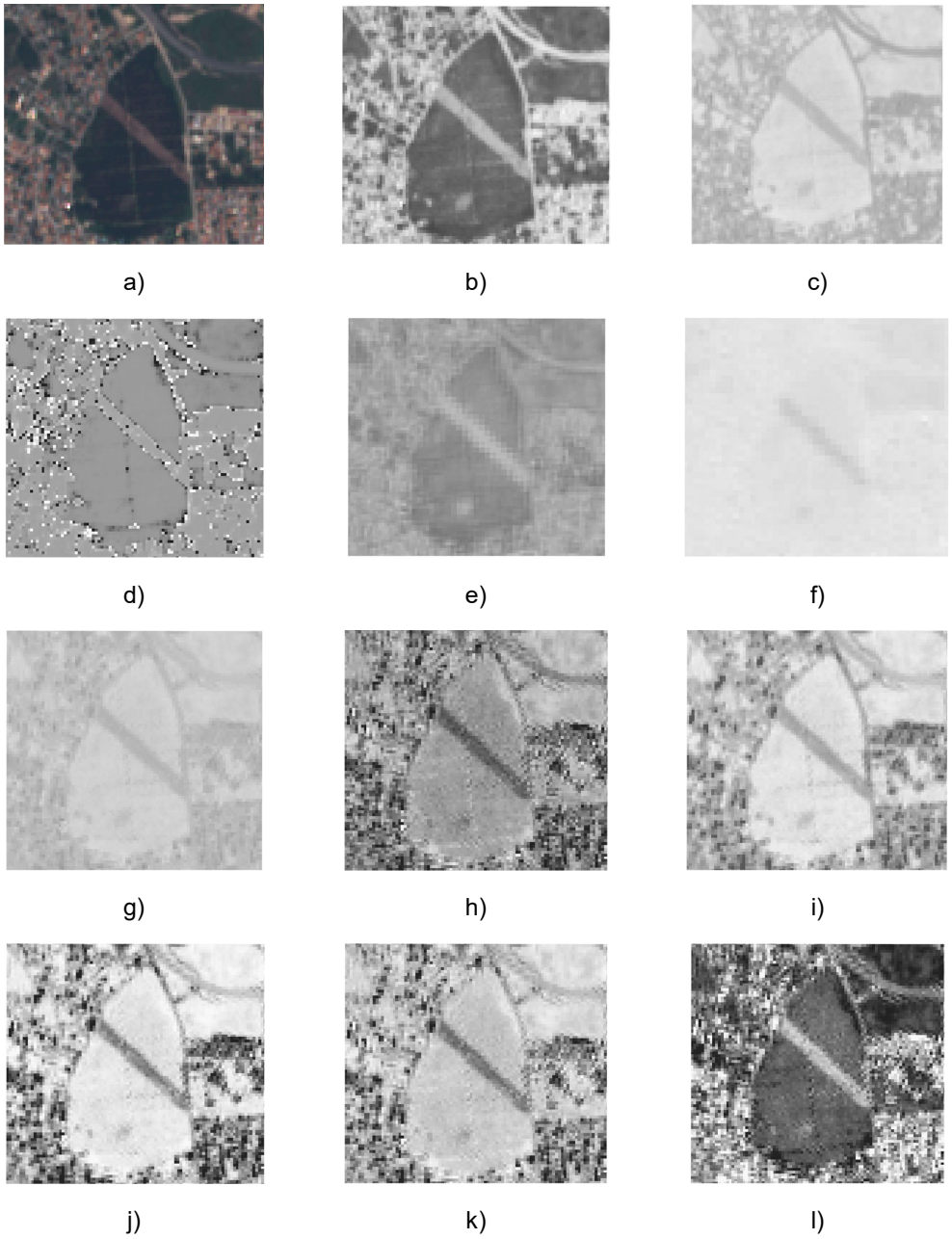


Figura 4.8: Aplicação de Índices Espectrais. a) Banda TCI; b) RVI; c) NDVI; d) EVI; e) NDMI; f) NMDI; g) NDI; h) ExG; i) ExR; j) ExGR; k) MExG; l) CIVE. As observações ocorreram na região de Marisol em 18/04/2018, sendo resultado dos dados do *Sentinel 2*.

4.6 Seleção de Atributos para a Classificação

Antes da fase de deteção de alterações, foi necessário identificar os atributos de entrada que iriam ser utilizados. Dentro do lote de atributos disponíveis para a classificação encontravam-

se os valores de reflectância das bandas 2, 3, 4, 5, 7, 8, 8A, 9, 11 e 12 (as restantes não foram analisadas dada a sua baixa resolução espacial) e os índices espectrais referidos em 3.1, com exceção do EVI2 e CIVE.

Cada atributo acrescenta uma dimensão ao problema a resolver e consequentemente uma maior complexidade no processo de aprendizagem. Ao mesmo tempo a utilização de características desnecessárias não traz qualquer benefício ao classificador. Estes factos tornam esta fase de seleção bastante importante, porque uma boa escolha dos atributos conduz a melhores resultados na classificação.

Para a seleção dos atributos recorreu-se a duas ferramentas do *Python*: o algoritmo *SelectKBest*, que seleciona um conjunto de atributos consoante a métrica de classificação escolhida e o cálculo da correlação de *Pearson*. Calculou-se então o valor F a partir de uma análise de variância. Este método compara cada atributo com a classe, sendo que consoante o número de atributos existentes e observações é obtido um valor crítico. Contudo este método não analisa a correlação dos atributos entre si, podendo selecionar dois atributos muito correlacionados que juntos não acrescentam informação ao problema. Desenvolveu-se então uma ferramenta que calcula as correlações de *Pearson* entre todos os atributos disponíveis para a classificação após a seleção de atributos. Na presença de variáveis com correlações muito altas é removida aquela que apresenta maiores valores em relação aos restantes atributos. Seguidamente observam-se quais os que têm menores níveis de semelhança, definindo-se quais os pares de variáveis que trazem mais informação para o classificador. Após esta análise são estabelecidos vários grupos de atributos. A partir de cada um deles são geradas várias ANN e calculado o valor médio do erro de resubstituição. O conjunto que obteve o melhor desempenho foi escolhido.

4.7 Conjunto de Treino e Conjunto de Validação

Sendo que a cada FGC corresponde um CSV individual o primeiro passo consistiu no agrupamento de todos estes ficheiros num só conjunto de dados com recurso a *scripts Python*. Com vista a uma estimação do erro mais robusta, foram definidos um conjunto de treino, um conjunto de validação e um conjunto de teste. A escolha de exemplos para cada um destes conjuntos foi feita com recurso ao método de separação de conjuntos da biblioteca *Sklearn* do *Python*.

Reservaram-se dois terços do conjunto total para treino e o restante para validação e a divisão foi feita de modo estratificado de forma a preservar a proporção das classes. Note-se que o número de exemplos que contêm intervenção constitui apenas 5% dos exemplos totais. Num processo simples de treino estes seriam interpretados como *outliers.* De forma a evitar que isto aconteça foi atribuído um peso a estes erros, tendo estes uma penalização cinco vezes superior em relação a outro erro possível. O conjunto de treino é então constituído por 112 exemplos, sendo que apenas 5 deles são relativos a intervenções e o conjunto de validação por 56 exemplos, sendo 3 deles classificados com intervenção.

O conjunto de teste foi relativo aos dados medidos para a região da Marisol. Este conjunto apenas foi testado no final do dimensionamento do algoritmo de decisão sem serem executadas adaptações para a obtenção de melhores resultados. Para além disso trata-se de uma

região diferente e pretendia-se verificar a aplicabilidade do classificador a áreas de estudo diferentes. Este conjunto é constituído por 48 elementos, sendo apenas um deles classificado como intervenção.

4.8 Dimensionamento e treino da ANN

Após a definição dos conjuntos de treino e de validação foi dimensionada a ANN. De forma a não realizar um processo fastidioso e pouco otimizado de tentativa e erro para escolher o número de camadas e neurónios em cada camada que mais se adequava ao problema, foram analisadas várias configurações possíveis através de um *script*. Nestas configurações foram contemplados os conjuntos de atributos selecionados de acordo com 4.6 e variou-se o número de neurónios de uma ANN com uma camada escondida e com duas camadas escondidas. Note-se que os pesos com que é inicializada a ANN são aleatórios, portanto procurou-se uma solução que independentemente dos pesos inicias convergisse, provavelmente, para bons classificadores. Quanto à camada de entrada definiu-se que será utilizado um neurónio por cada atributo e na camada de saída um neurónio, pois trata-se de uma classificação binária (FGC com ou sem intervenção). Os resultados deste estudo serão apresentados no capítulo de resultados.

Para este dimensionamento foi implementado um *script* que estima o erro do classificador, primeiro aplicando a validação cruzada ao conjunto de treino e o erro de resubstituição ao conjunto de validação. Posteriormente o algoritmo foi treinado com ambos os conjuntos referidos anteriormente e detetada a intervenção no conjunto de teste. Comparou-se este resultado com o que se observou visualmente e com os dados de análise obtidos. Também são apresentados em 5.3 os resultados das métricas de validação presentes em 3.3.3.

4.9 *Plugin* para QGIS de extração de NDVI

Foi desenvolvido um *plugin* para QGIS que teve como finalidade ser uma ferramenta que permite a extração de dados de forma automática e simples. A interface do *plugin* está apresentada na Figura 4.9.

Mediante a indicação de uma pasta através do botão *Load NDVI Folder* são obtidas todas as bandas necessárias para o cálculo do NDVI (banda 4 e banda 8). A opção *Load Mask*, carrega no visualizador do QGIS a máscara relativa à FGC ou zona de vegetação em análise. A opção *Crop Bands* recorta a zona que se pretende analisar (é dada pelo utilizador a máscara relativa) e prepara as imagens para a georreferenciação. Caso se pretenda estudar apenas de uma observação o carregamento de bandas pode ser feito com recurso ao *Load NDVI Bands*.

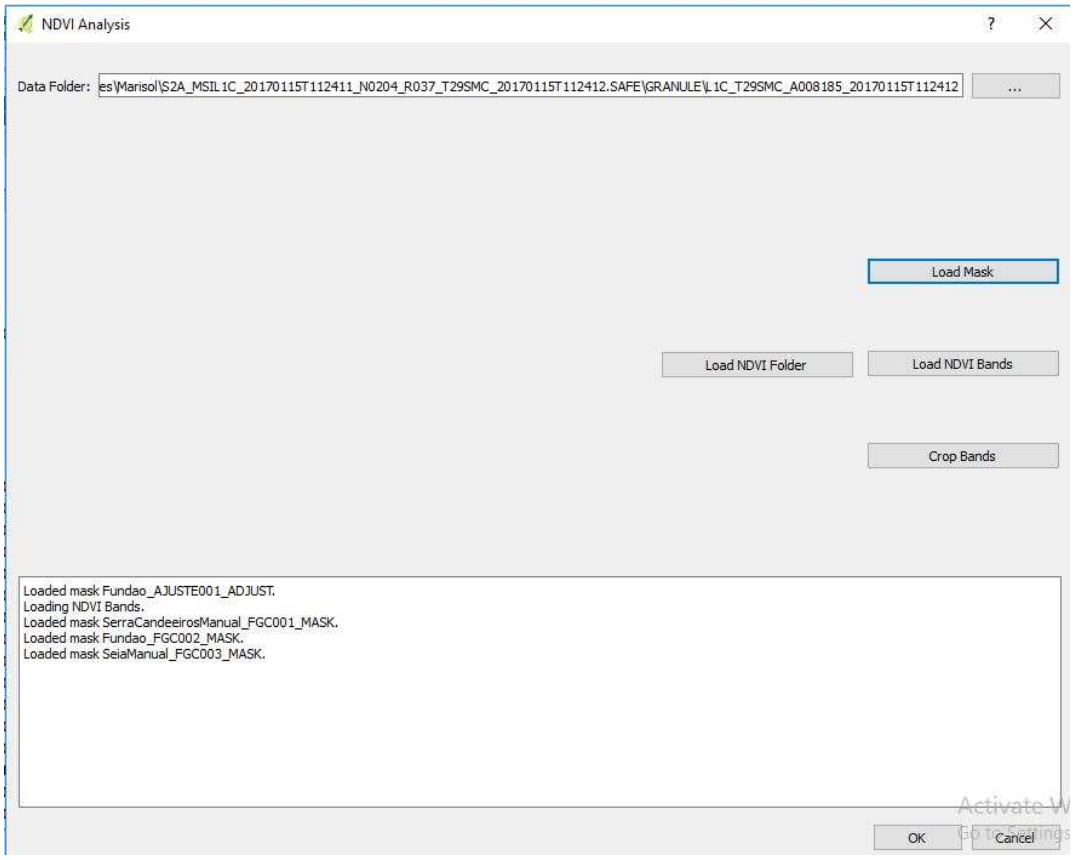


Figura 4.9: Interface do *Plugin* de análise do NDVI.

4.10 Sumário

Apresentado o trabalho desenvolvido um dos detalhes mais importantes verificados é a possibilidade de utilizar o QGIS para todos os pontos da deteção automática das intervenções nas FGC. O SIG permite uma automatização em conjunto com o *Python* dos processos, desde a aquisição de dados, pré-processamento dos dados, treino dos classificadores e finalmente extração de conhecimento.

As ferramentas desenvolvidas permitirão um conhecimento maior dos comportamentos da vegetação e de que forma se poderiam detetar as intervenções, como será visto no capítulo 5.

Por fim, foi desenvolvida a seguinte metodologia de trabalho que será implementada em futuras ferramentas para a deteção automática, apresentada na Figura 4.1 e sintetizada nos seguintes pontos:

1. Obtenção das imagens;
2. Obtenção dos vetores relativos às FGC e zonas de vegetação;
3. Rasterização de vetores;
4. Correção da georreferenciação das imagens;

5. Extração de dados das imagens, bandas e índices espectrais;
6. Geração de conjuntos de dados, executando a normalização de dados;
7. Divisão do anterior em conjunto de treino, validação e teste;
8. Dimensionamento e treino da ANN;
9. Classificação de dados com ANN treinada e validada.

Todas estas tarefas (exceto a primeira e a quarta) foram executadas com recurso ao QGIS e às potencialidades do *Python*. Note-se que depois de ter uma ANN já treinada os pontos 3 e 8 deixam de ser necessários, sendo seguido apenas o caminho assinalado a vermelho no diagrama da Figura 4.1.

5 Apresentação e Discussão de Resultados

Para demonstrar a validade das técnicas propostas para deteção de intervenções nas FGC e dos métodos de extração de dados neste capítulo serão expostos os resultados obtidos. Numa primeira parte é feita uma análise exaustiva das grandezas medidas e dos índices espectrais calculados para validar a escolha dos atributos para a classificação e quais as assinaturas espectrais importantes para entendimento do comportamento da vegetação e do solo. Em 5.2.1 são aplicados algoritmos de seleção de atributos para a classificação e discutida a seleção dos atributos. Em 5.2.2 é dimensionada a ANN que será utilizada para a classificação e em 5.3 mostram-se os resultados obtidos com a mesma.

5.1 Análise de dados em Serra de Aire e Candeeiros

A extração e análise de dados assumiu um papel muito importante nesta dissertação como observado no capítulo 4.5. São agora apresentados os resultados para os valores da reflectância das bandas em análise e dos índices espectrais escolhidos. Todos os gráficos não disponibilizados nesta secção estão presentes nos

Anexos.

5.1.1 Análise das Bandas do Espectro Electsromagnético

Primeiramente vão ser observados os valores das bandas e analisada a informação sobre a existência de um corte que pode ser extraída das mesmas. Nesta análise as bandas em estudo foram divididas nos quatro grupos definidos na Tabela 5.1. As bandas pertencentes ao mesmo conjunto evidenciam um comportamento semelhante em relação às operações de manutenção nas FGC. Deste modo por cada grupo serão apresentados os resultados apenas de uma das bandas em 2017 e 2018.

O Grupo 1 não apresenta grande sensibilidade às operações de corte da floresta como se pode ver na Figura 5.1 e na Figura 5.2. A fenologia das florestas encontra-se sempre

evidenciada na evolução anual dos valores medidos. Todas as bandas e índices que serão apresentados sofrem deste fénomeno. Pretende-se encontrar as bandas, que apesar deste facto consigam evidenciar uma maior alteração da sua assinatura espectral após as operações de manutenção. Verifica-se que apesar de uma alteração muito ténue após o corte é difícil definir se não são apenas fenómenos sazonais. Porém a determinação da altura do ano à qual pertence a observação pode assumir um papel muito importante na deteção de cortes. Uma consideração que deve ser feita é facto do corte estar registado na base de dados com a data de 27 de julho, dado que corresponde a uma época do ano em que surgem modificações nos valores de reflectância por causa dos fenómenos fenológicos. Estas alterações sazonais nos dados podem ser confundidas com operações de manutenção ou, podem mascarar intervenções que ocorreram durante este periodo.

Tabela 5.1: Grupos de Bandas.

GRUPOS	BANDAS
1	$B02\ e\ B03$
2	$B04\ e\ B05$
3	$B07, B08\ e\ B8A$
4	$B11\ e\ B12$

No Grupo 2 já é observável uma maior sensibilidade ao corte, sendo a banda 4 a mais sensível a este (Figura 5.3 e Figura 5.4). Constata-se também, observando as zonas VEG e os dados de 2018, o fenómeno da sazonalidade. No início e final do ano a reflectância é praticamente igual, revelando o seu comportamento periódico. Um detalhe importante para a deteção da existência de um corte no ano em análise é o facto de que se este existiu verifica-se uma variação dos valores medidos. Uma questão que surge a partir da análise do Grupo 2 é da evidência de alterações nos valores de reflectância antes da data oficial registada na base de dados para a operação de manutenção. Na realidade esta data representa o fim destas operações que não são feitas de forma instantânea, mas sim ao longo de um período de tempo. Neste caso torna-se claro que a operação terá tido início em junho e terminando no final do mês seguinte. Considerando este pormenor na fase de classificação será considerada a existência de corte em ambos os meses.

A deteção de cortes prende-se muito com a deteção de alterações no terreno, como o trabalho proposto por (Hamunyela et al., 2017), visto em 3.2.2. Isto implica a necessidade de informação temporal no processo de classificação. Serão utilizados então os valores de reflectância do mês anterior na perceção das FGC.

Apesar de já se encontrar uma alteração comportamental da reflectância medida pelo *Sentinel 2*, este fenómeno ainda se encontra muito mascarado pela fenologia das florestas.

A evolução temporal da banda 8, elemento do Grupo 3 é apresentada nas Figura 5.5 e Figura 5.6. Este grupo é fortemente afetado pela fenologia. Consegue-se observar uma descida

mais acentuada nos valores medidos no período de manutenção, porém não é um fenómeno evidente, parecendo ser resultado exclusivo das alterações estivais que ocorrem durante o ano. Note-se que individualmente as bandas podem não acrescentar muita informação para a resolução do problema proposto, contudo quando associadas a um índice espectral esta situação poderá ser diferente, como será observado na análise dos índices.

Finalmente nas Figura 5.7 e Figura 5.8 são apresentados os resultados para o Grupo 4. Este é o grupo em que as operações de manutenção são mais evidenciadas. Repare-se que os resultados são semelhantes aos do Grupo 2, porém as reflectâncias são maiores e, mais importante, a variação após a manutenção é superior.

Conclui-se que utilizando as bandas individualmente as que trazem maior informação sobre a ocorrência de manutenção numa FGC são as que compõem o Grupo 4.

Outro fenómeno bastante interessante é apresentado na Figura 5.9 e Figura 5.10. Observa-se em zonas de vegetação a B11 é sempre inferior à B08, contudo quando ocorre um corte a B11 superioriza-se a B08. Este facto resulta da diferente assinatura espectral que caracteriza o solo e a floresta. Este detalhe também pode ser bastante vantajoso para a deteção do corte. Em 2018 verifica-se que quando começa a crescer vegetação a relação entre estas bandas volta ao que era antes da operação.

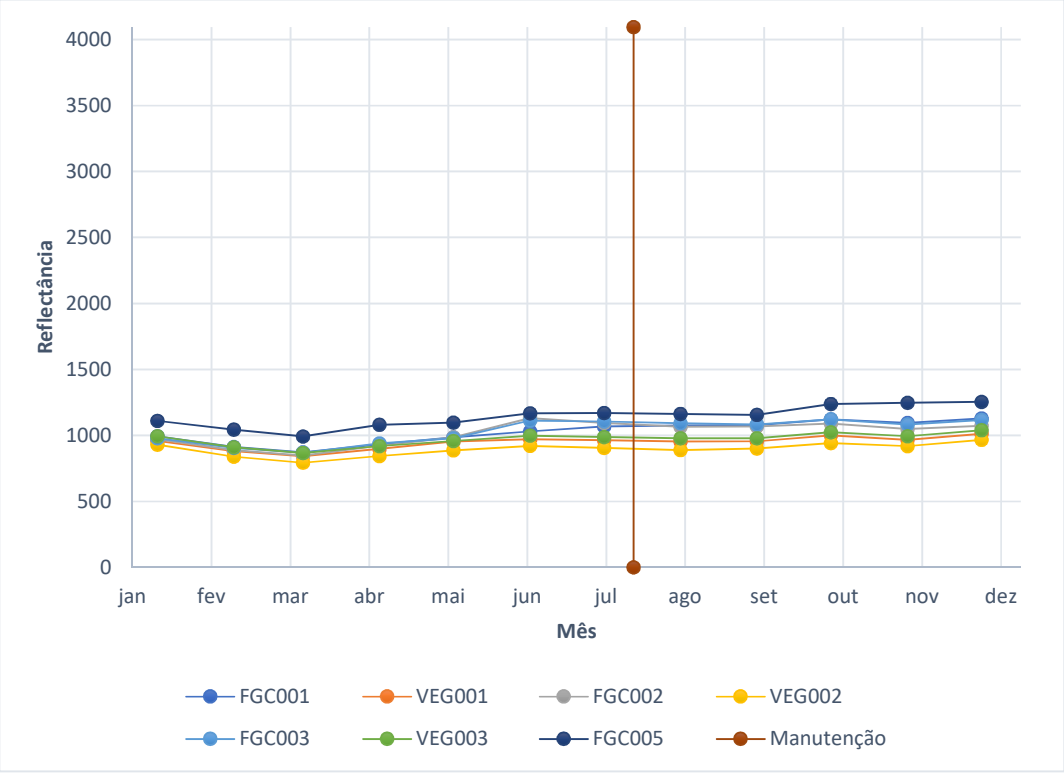


Figura 5.1: Evolução Temporal da Banda 2 em 2017.

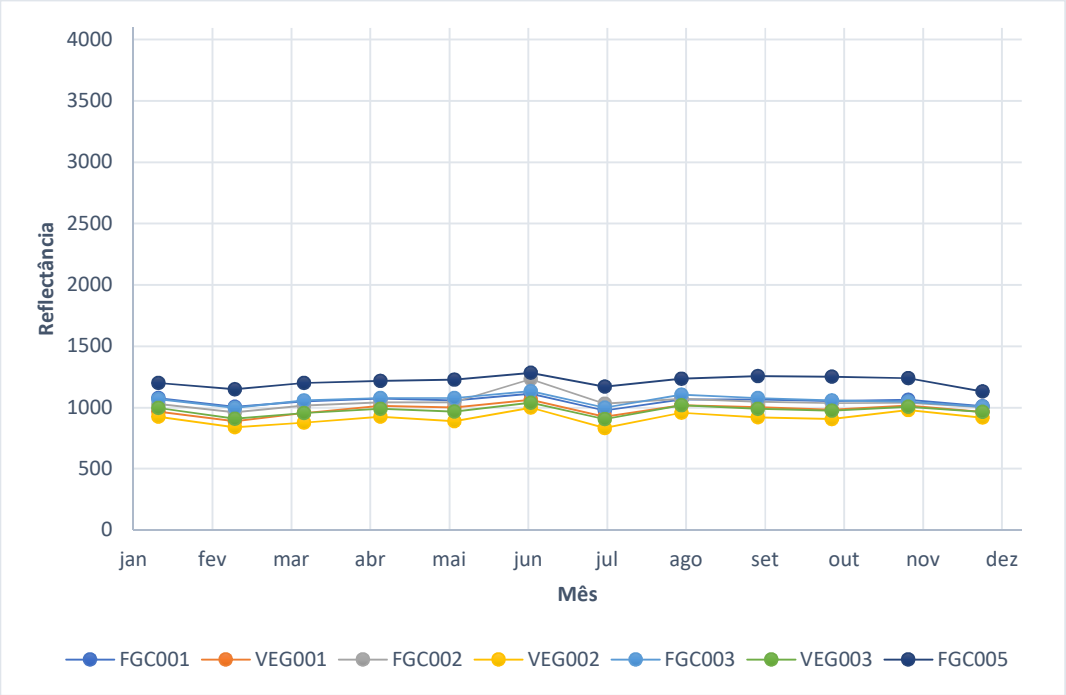


Figura 5.2: Evolução Temporal da Banda 2 em 2018.

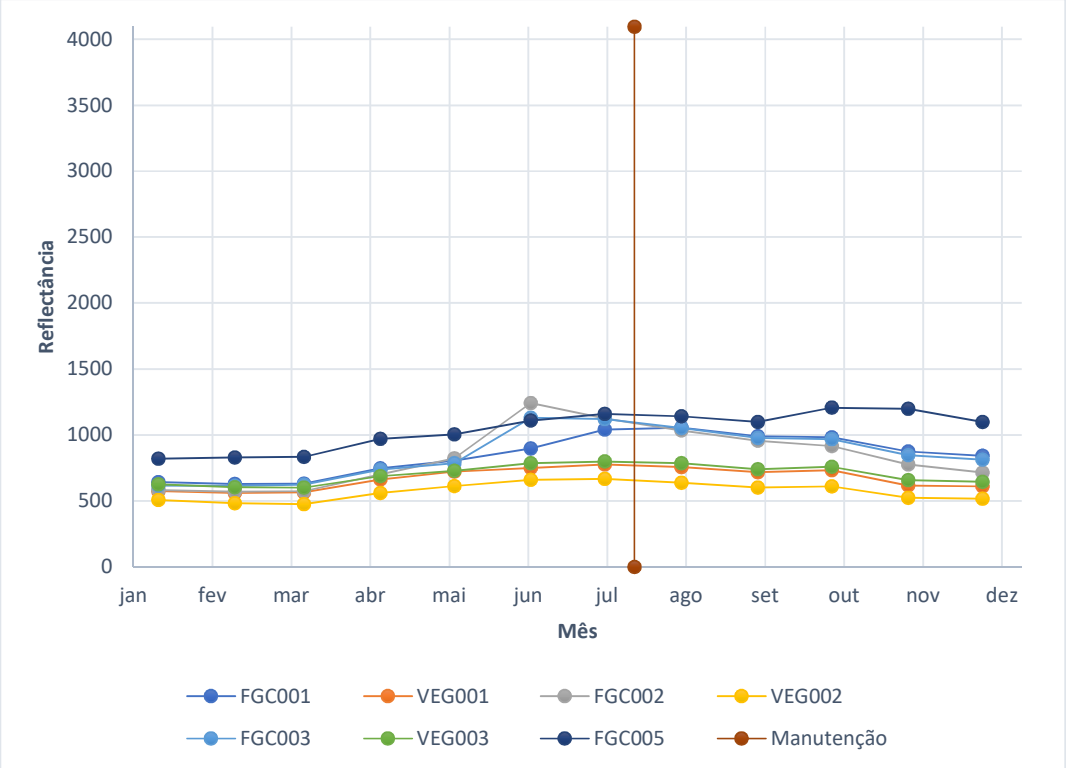


Figura 5.3: Evolução Temporal da Banda 4 em 2017.

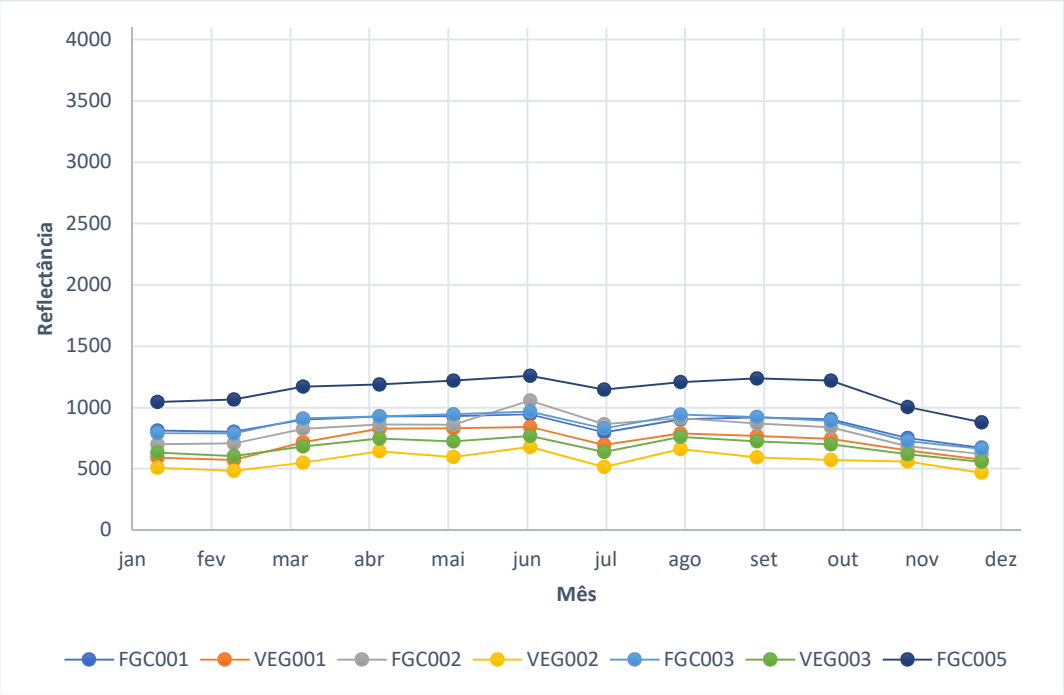


Figura 5.4: Evolução Temporal da Banda 4 em 2018.

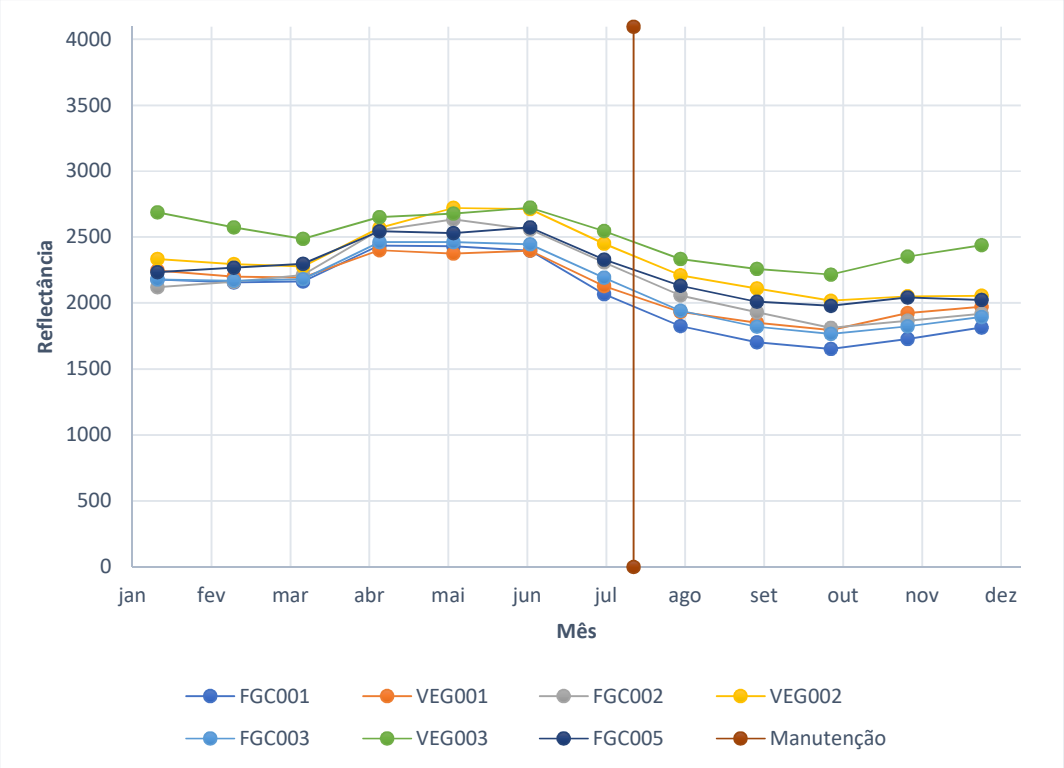


Figura 5.5: Evolução Temporal da Banda 8 em 2017.

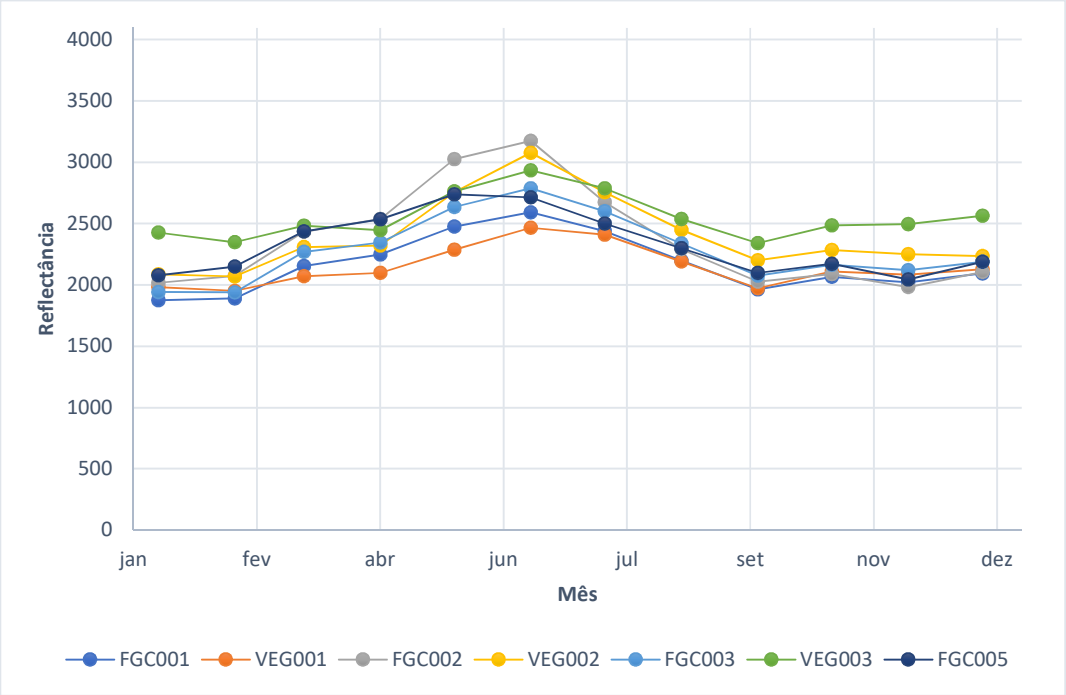


Figura 5.6: Evolução Temporal da Banda 8 em 2018.

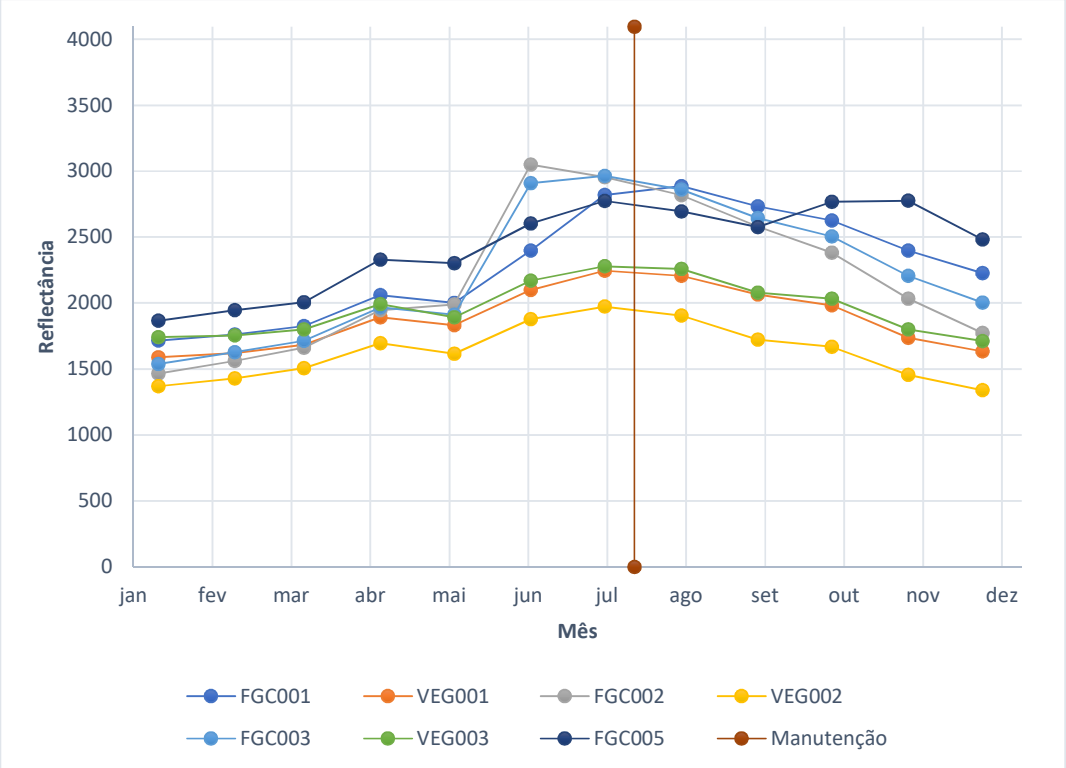


Figura 5.7: Evolução Temporal da Banda 11 em 2017.

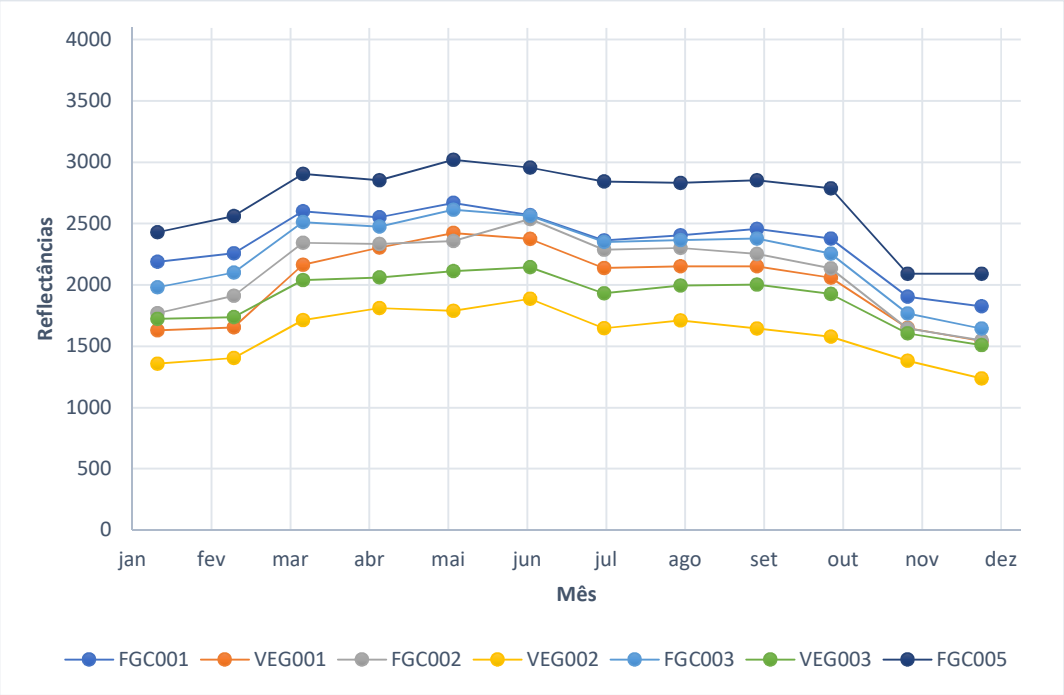


Figura 5.8: Evolução Temporal da Banda 11 em 2018.

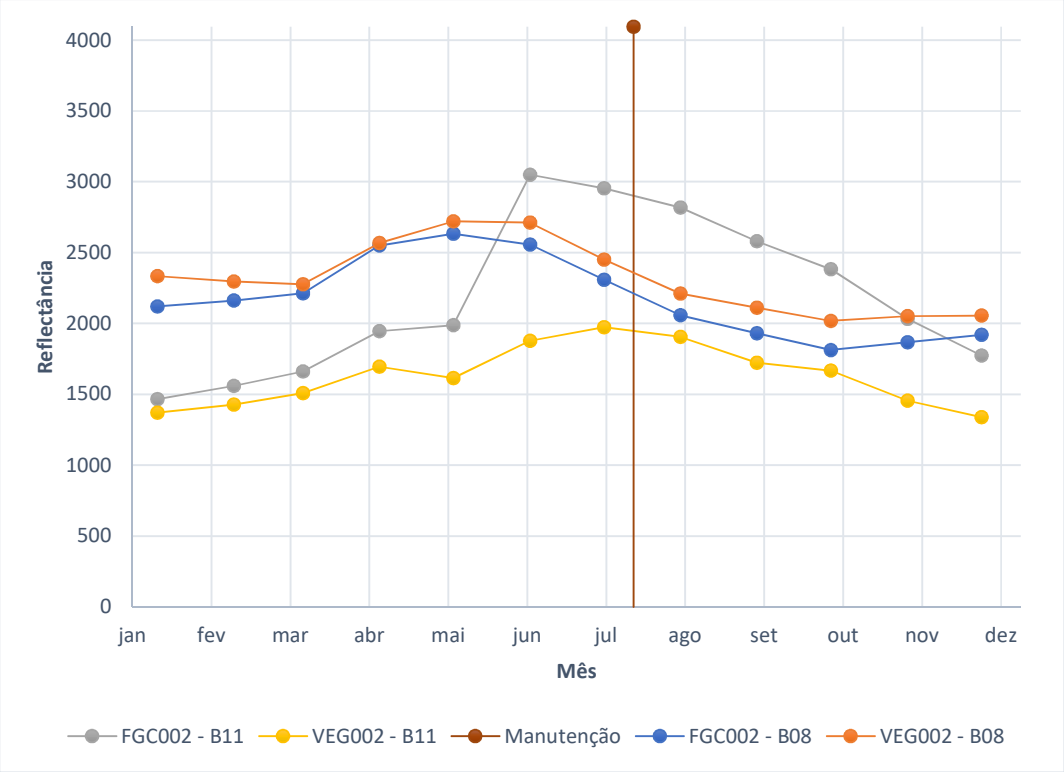


Figura 5.9: Evolução Temporal da Banda 8 e Banda 11 em FGC e VEG em 2017.

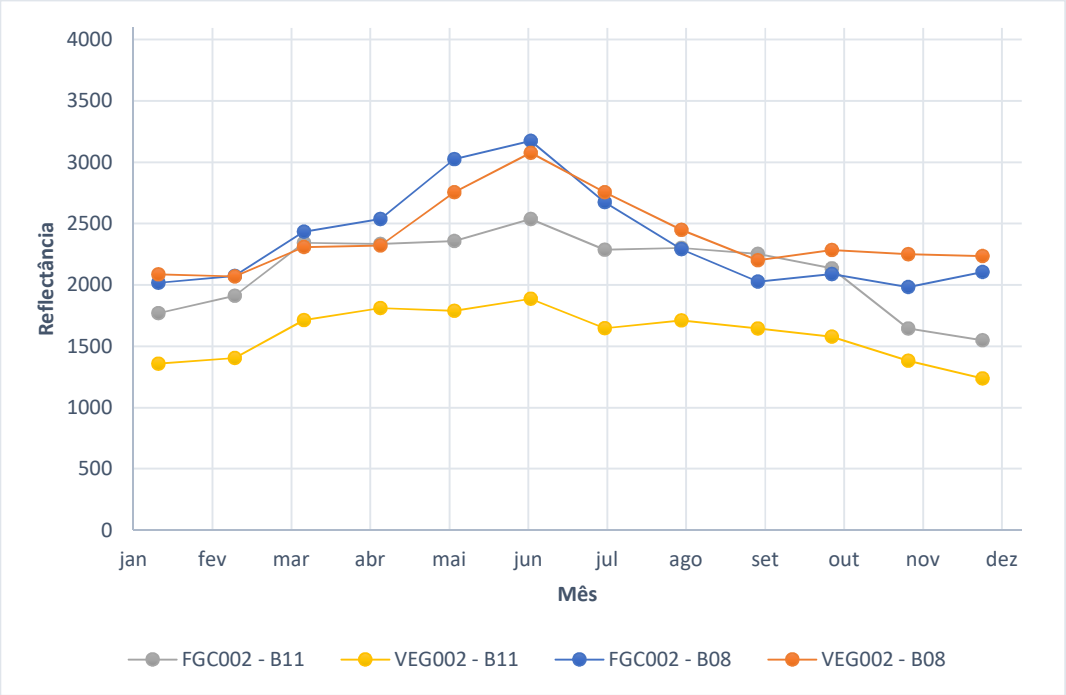


Figura 5.10: Evolução Temporal da Banda 8 e Banda 11 em FGC e VEG em 2018.

5.1.2 Análise de Índices Espectrais

Na análise dos resultados obtidos pelos índices espectrais foi adotada a mesma metodologia utilizada em 5.1.1. Dividiram-se em três grupos aqueles índices que apresentaram comportamentos semelhantes em relação a operações de manutenção (Tabela 5.2).

Tabela 5.2: Grupos de Índices com comportamentos semelhantes.

GRUPOS	BANDAS
5	$NDVI, NDI, NDMI, NMDI\ e\ EVI$
6	$RVI, ExG\ e\ MExG$
7	$ExGR\ e\ ExR$

O Grupo 5 (representado pelo NDI) apresentado na Figura 5.11 e na Figura 5.12 assume um comportamento similar ao do Grupo 2 em termos das variações temporais. A evidência da fenologia nestes índices é clara, sendo a intervenção facilmente confundida com esta. Mais uma vez a escolha da data de intervenção revela que pode induzir confusão no processo de deteção. Quanto ao NMDI, tratando-se de um índice muito associado à deteção de períodos de seca e ao estudo da água, praticamente não acrescenta informação para a deteção do corte.

No Grupo 6 (ver Figura 5.13, Figura 5.14, Figura 5.15 e Figura 5.16) a fenologia começa a assumir uma influência menor, estando mais presente no RVI, mas neste as variações resultantes da manutenção são muito evidentes. No MExG (a escala pode não parecer a mais adequada, contudo foi utilizada a mesma para todos os índices normalizados no intervalo $[-1,1]$ de forma a não induzir em erro), a fenologia assume um papel de menor importância, contudo as variações devido ao corte também não são muito elevadas.

No Grupo 7, cujos resultados são exibidos na Figura 5.17 e na Figura 5.18, surge uma característica bastante interessante. Apesar de valores muito baixos, verifica-se um comportamento das zonas de vegetação bastante constante, praticamente sem relação com a fenologia. Apesar das variações relativas ao corte não serem muito grandes, existe uma alteração comportamental nos índices. Esta modificação pode desempenhar um papel positivo na perceção da intervenção.

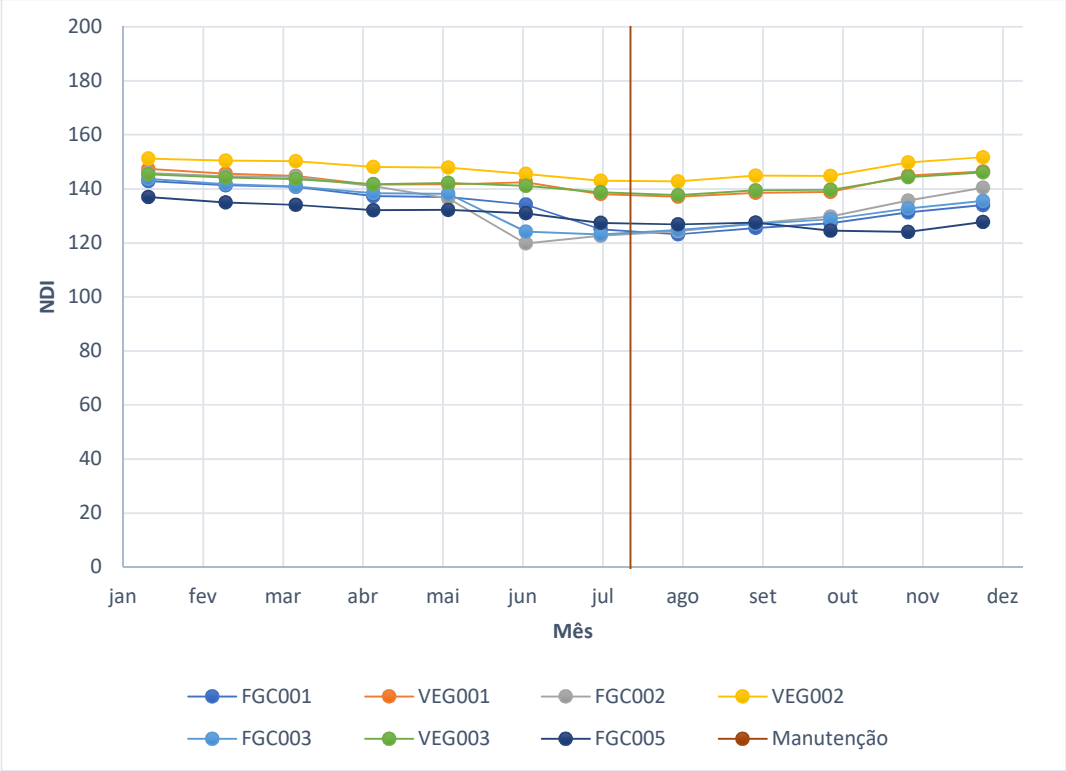


Figura 5.11: Evolução Temporal do NDI em 2017.

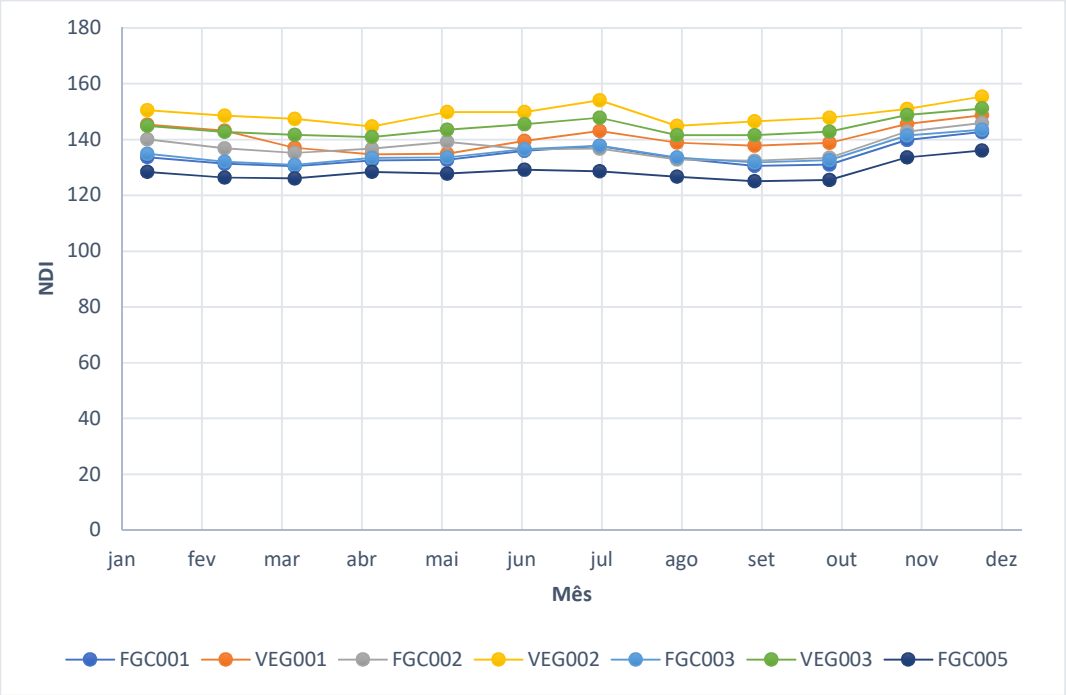


Figura 5.12: Evolução Temporal do NDI em 2017.

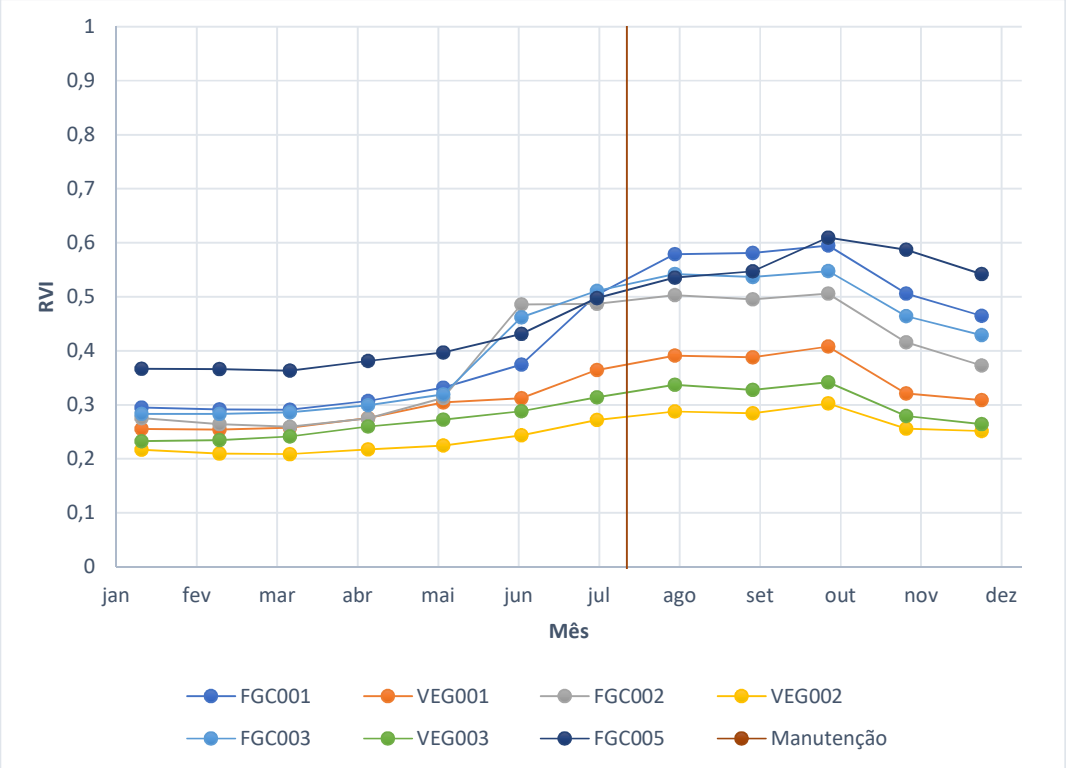


Figura 5.13: Evolução Temporal do RVI em 2017.

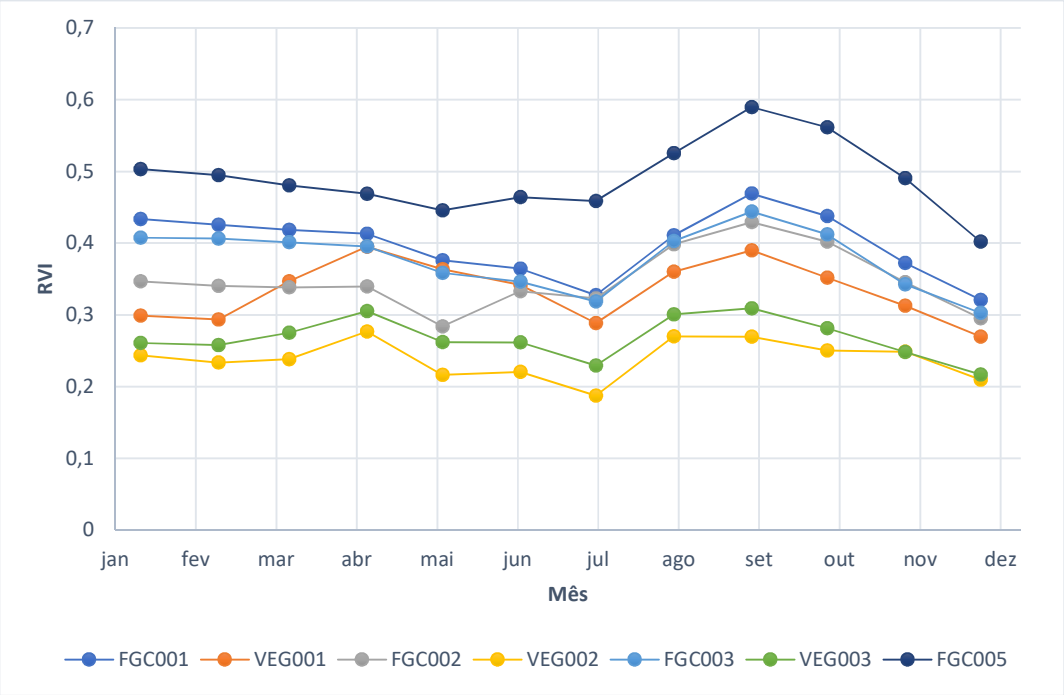


Figura 5.14: Evolução Temporal do RVI em 2018.

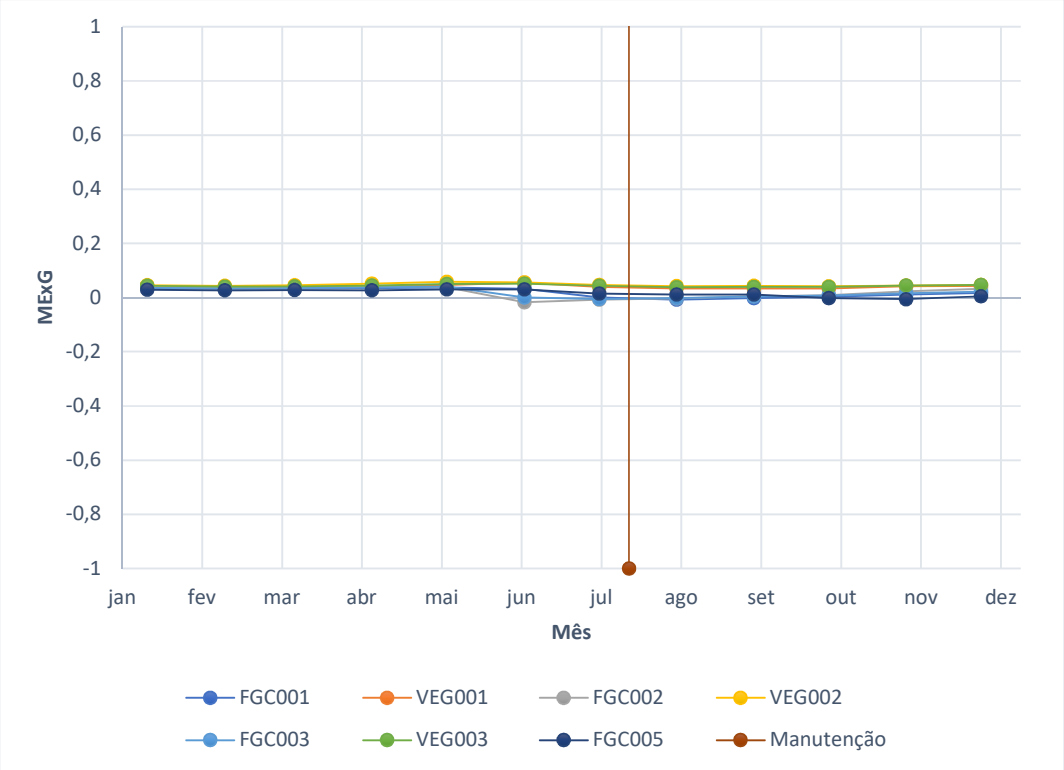


Figura 5.15: Evolução Temporal do MExG em 2017.

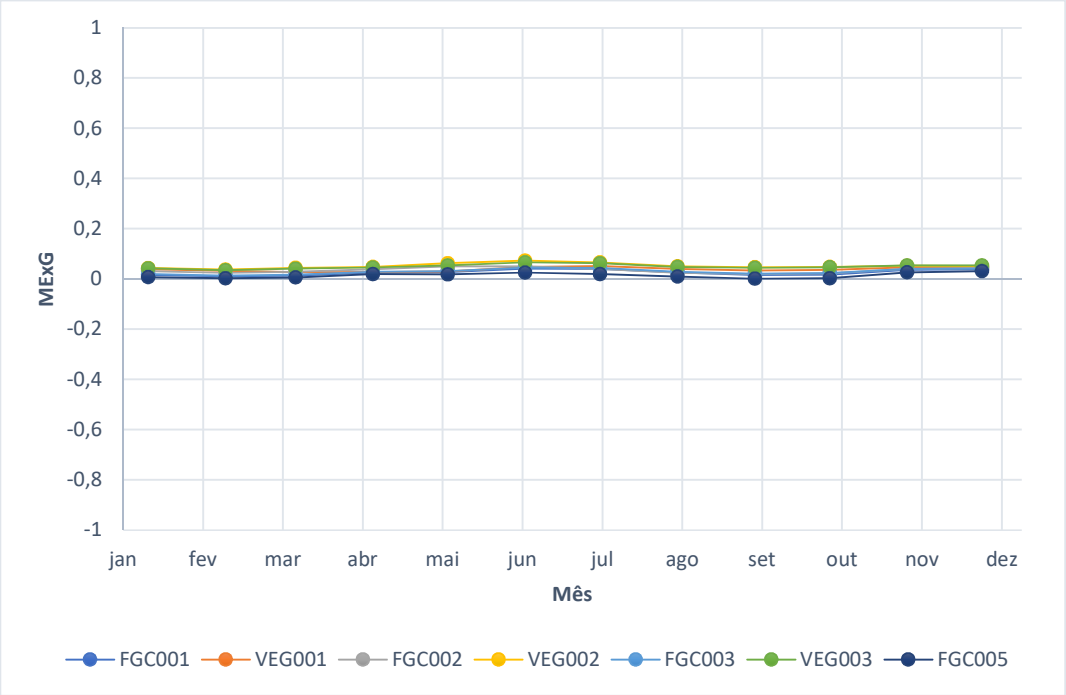


Figura 5.16: Evolução Temporal do MExG em 2018.

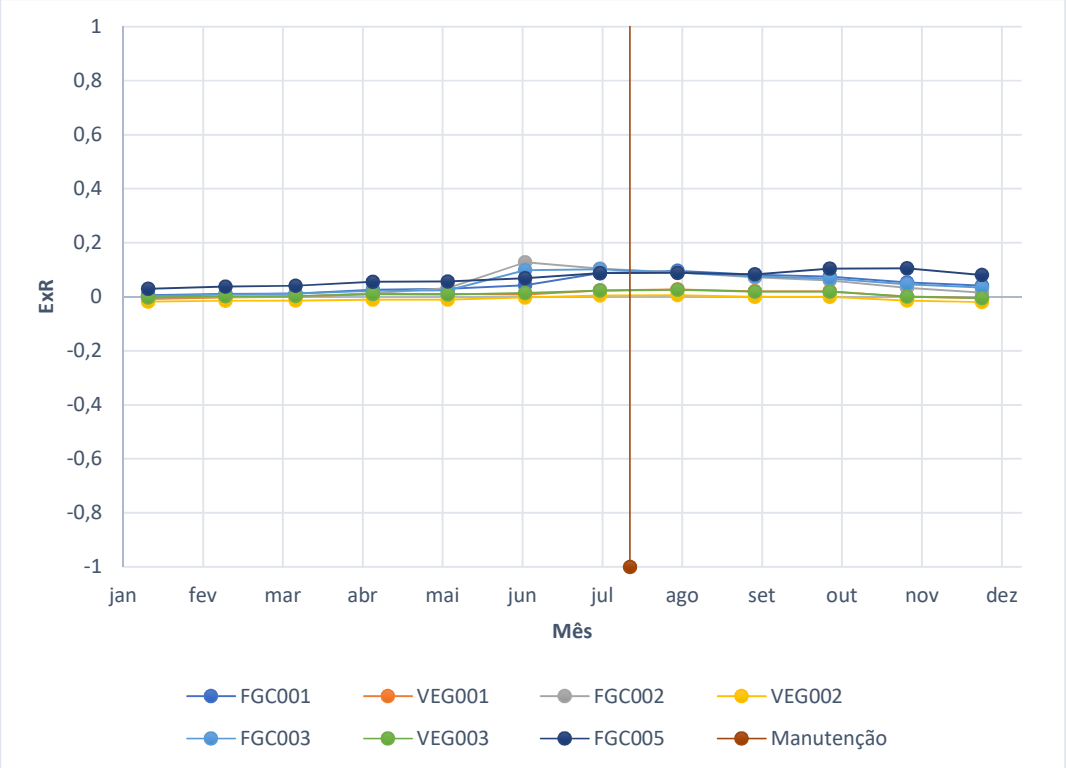


Figura 5.17: Evolução Temporal do ExR em 2017.

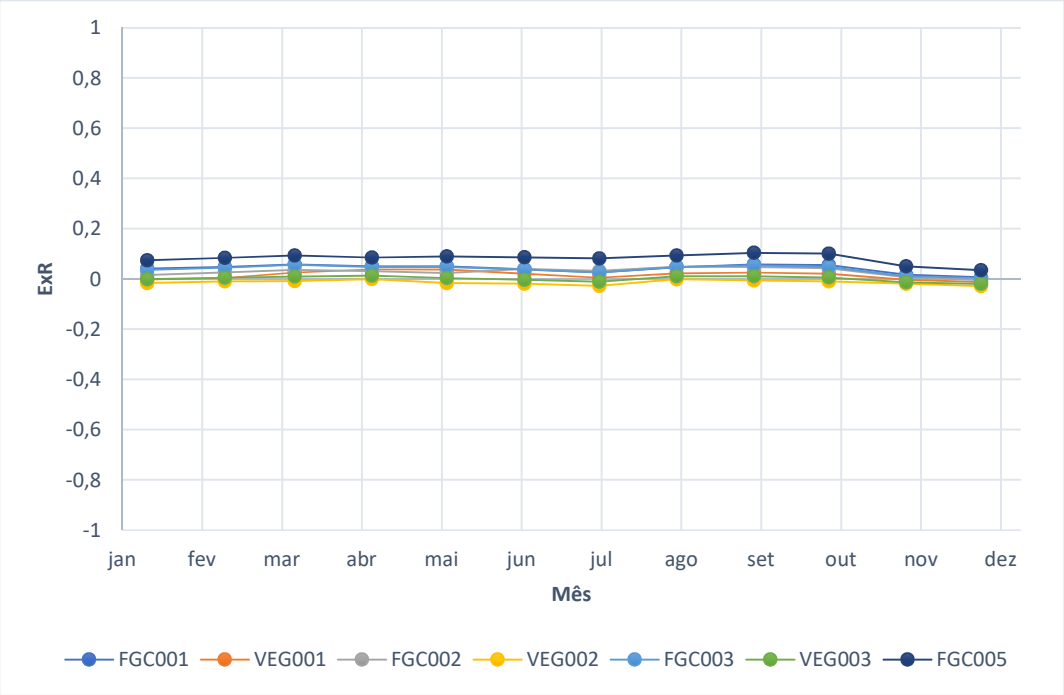


Figura 5.18: Evolução Temporal do ExR em 2018.

5.2 Resultados das Técnicas de *Machine Learning*

5.2.1 Seleção de Atributos

Após se ter observado o comportamento das bandas e dos índices relativamente às operações de manutenção é importante escolher quais os atributos a utilizar no classificador. Considerando como atributos os valores normalizados de reflectância de todas as bandas estudadas e os valores dos índices apresentados em 5.1, os dez atributos com maior relação com as intervenções são:

- Bandas: B03, B04, B05, B11, B12;
- Índices: NDVI, NDI, ExR, ExGR, MExG.

Seguidamente obtiveram-se as correlações entre todos estes atributos. O valor absoluto de correlação mais alto obtido foi entre a B11 e B12 de 0,986. A utilização de ambas provavelmente não trará ganhos significativos ao problema. Dado que têm uma resolução espacial idêntica e que B11 apresenta valores de correlação mais baixos com os restantes atributos, a B12 foi descartada do processo de deteção. A B03 e B05 apresentam uma correlação também bastante alta de 0,947. A primeira revela uma maior diferença para com as restantes variáveis, sendo a B05 eliminada. A correlação da B03 e B04 é de 0,938, mas B04 é eliminada pelos critérios usados anteriormente. Abordando agora os índices, o ExGR e o MExG têm um valor de correlação de 0,972 e avaliando a correlação com os atributos ainda não eliminados, o MExG é escolhido em detrimento do ExGR.

Observando agora os valores mais baixos de correlação, verifica-se que a B03 e o MExG apresentam o resultado mais baixo, 0,382, sendo importante a presença de ambos os atributos no processo de classificação. Quanto a B11 os resultados em relação aos atributos restantes, apresenta resultados piores sendo considerada apenas em conjunto com a B03. Definiram-se então os conjuntos de atributos presentes na Tabela 5.3. Foram testados e escolhido o conjunto que obteve melhores resultados de classificação. O teste consistiu no treino de uma ANN com apenas uma camada escondida em que o número de neurónios nesta varia no intervalo $[5, 100]$ e observou-se a evolução do erro de resubstituição mostrado nas Figura 5.19 e Figura 5.20 (os gráficos encontram-se separados para facilitar a leitura).

Tabela 5.3: Conjuntos de Atributos para a classificação.

CONJUNTO	ATRIBUTOS
1	B03, MExG, ExR, NDI, NDVI
2	B03, B11, MExG, NDI
3	B03, MExG, NDI, NDVI
4	B03, MExG, ExR, NDVI
5	B03, ExR, MExG, NDI
6	B03, MExG, NDVI
7	B03, MExG, NDI

Verifica-se que os conjuntos com melhor desempenho são o 1, o 4 e o 5. Contudo o grupo 4 precisa de mais neurónios, tornando-se mais suscetível de ocorrerem situaçãos de *overfitting*. Comparado os outros dois (1 e 5), o conjunto 5 necessita de menos atributos, sendo a solução escolhida por ser mais simples. Relembre-se que para além das medidas presentes também são utilizada as medidas do mês anterior, existindo então um total de oito atributos de entrada na ANN.

5.2.2 Dimensionamento da ANN

Definido qual o conjunto a utilizar, a próxima etapa consiste no dimensionamento da ANN. Definiu-se que na camada de entrada será utilizado um neurónio por cada atributo de entrada e um neurónio na camada de saída, pois trata-se de uma classificação binária (utilizaram-se as configurações habituiais). Executou-se um teste similar ao efetuado em 5.2.1, inicialmente utilizando apenas uma camada escondida. Observando novamente a Figura 5.20, verifica-se que a partir dos 40 neurónios já são obtidas taxas de erro bastante baixas. Para verificar se era positivo colocar uma segunda camada escondida testou-se uma rede neuronal variando os neurónios da primeira camada entre [20,40] e na segunda camada entre [2,30]. Os resultados estão presentes na Figura 5.21. Observa-se que os resultados obtidos não apresentam melhorias, não trazendo qualquer benefício para este problema a implementação de

mais uma camada escondida. Note-se que por causa dos *vanashing gradients* não são apresentados os resultados para ANN com mais do que duas camadas escondidas. A ANN a utilizar será então constituída por uma camada de entrada com 8 neurónios, uma camada escondida com 40 neurónios e uma camada de saída com 1 neurónio.

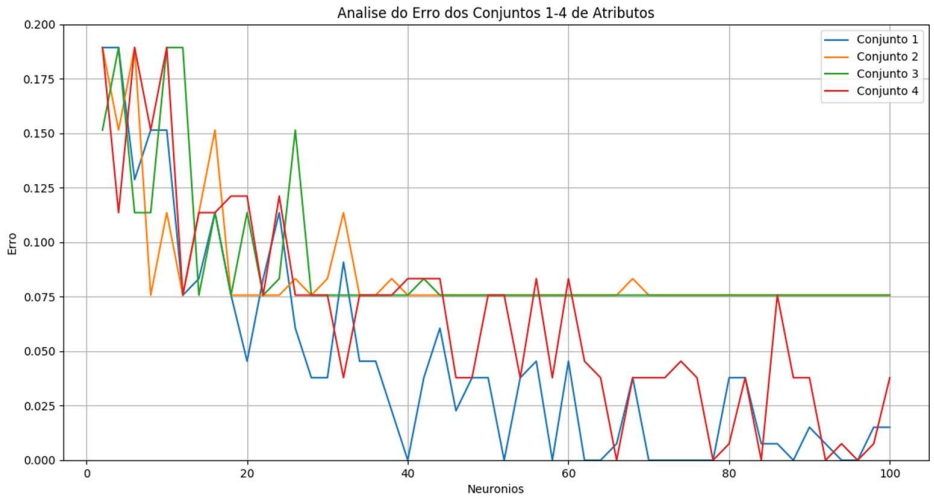


Figura 5.19: Análise do Erro de Resubstituição dos conjuntos de atributos 1-4.

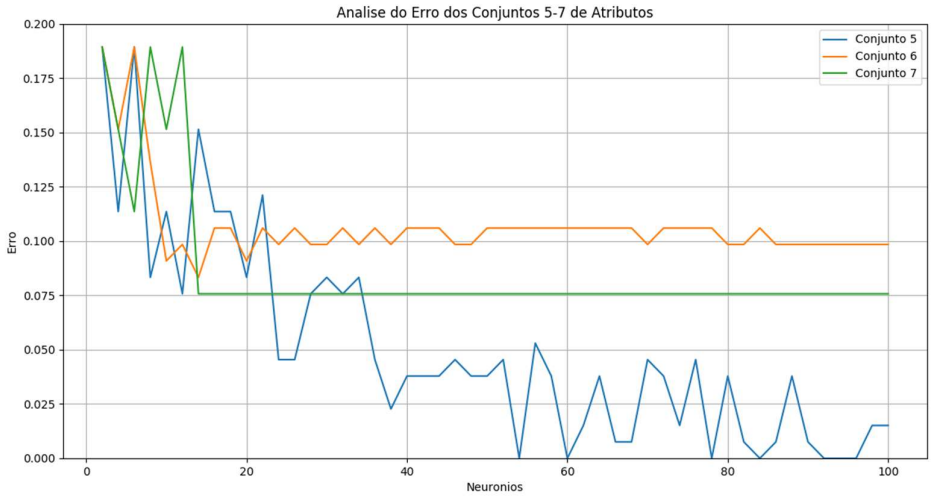


Figura 5.20: Análise do Erro de Resubstituição dos conjuntos de atributos 5-7.

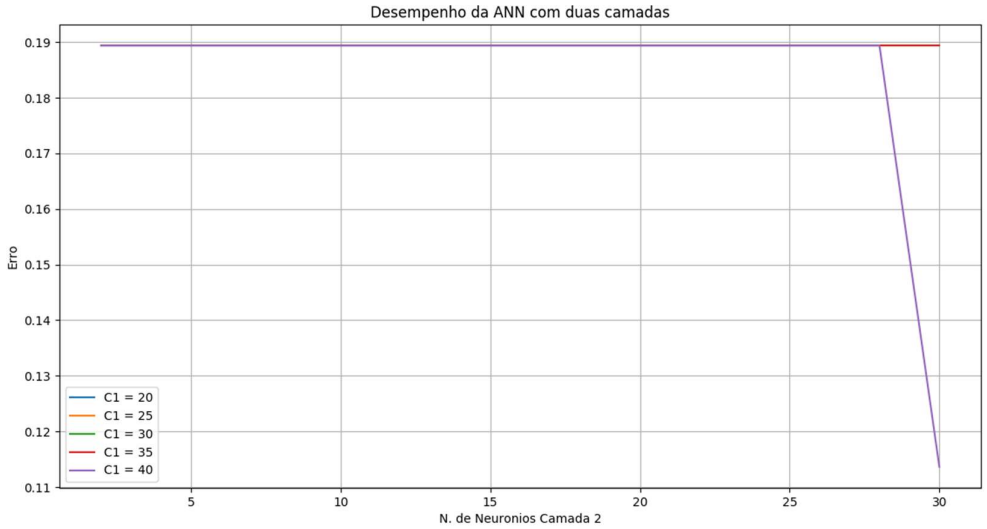


Figura 5.21: Desempenho da ANN com duas camadas.

5.3 Resultados da Classificação

A ANN utilizada para classificação usa a função sigmoidal como função de ativação e o gradiente descendente como método de retro propagação do erro. A taxa de aprendizagem é adaptativa e é utilizada uma tolerância de 0,001. Para a validação cruzada foi então definido que o conjunto de treino seria dividido em 5 grupos. Sendo que o conjunto de treino contém 112 exemplos, gerando em cada interação conjuntos de treino de 90 exemplos e um conjunto de validação de 22 exemplos. Apesar de normalmente se dividir o conjunto de dados em 10 grupos, devido à reduzida dimensão deste conjunto utilizaram-se apenas 5 grupos. A percentagem de erro obtida com esta técnica é de 4%.

Depois deste teste foi realizada a deteção no conjunto de validação. Para esta avaliação foram treinados 10 classificadores com recurso ao conjunto de treino. Classificou-se o conjunto com recursos a todas as redes neuronais geradas, obtendo-se um erro médio de 4,4%. Contudo é importante referir que os erros têm diferentes importâncias. Quando não é detetada uma operação de corte, é menos grave do que a deteção de um corte quando este não existiu. Porém, dentro deste segundo tipo de erro, a situação mais grave são os falsos positivos, ou seja, quando se detetam intervenções numa data em que estavam programadas e que estas não existiram ou foram mal-executadas. Este último erro nunca se registou nos classificadores gerados. Contudo, por vezes, são detetados cortes quando estes não existiram nem estavam planeados (ou seja, segundo o ICNF não foram requisitadas operações na área e data em causa). Também quando o classificador consegue detetar mais cortes corretamente acaba por detetar mais cortes não planeados e que efetivamente não existiram. Observando o *Recall* relativo à existência de corte, este teve um valor médio de 77%. Este valor pode parecer baixo, contudo é importante relembrar que no conjunto de validação apenas existem três exemplos em que existe operação de manutenção, o que significa que é comum falhar uma destas observações. Quanto à Precisão os

valores são mais baixos. Existem vários exemplos que são detetados como corte, não sendo. Contudo estes erros não surgiram em datas em que existia previsão de manutenção. O *F1-Score* obtido é de 66%. Este valor revela claramente que existe alguma dificuldade em classificar um dos exemplos de corte. Isto acontece, porque o corte é dito ser em junho e julho, mas estas são as datas de início e fim, não sendo disponível a data em que efetivamente ocorreu o corte. Quando utilizamos os classificadores para classificar o conjunto de treino e conjunto de validação o erro médio é de 2,2%, o *Recall* de 87,5%, a Precisão de 74,1% e o *F1-Score* de 79,5%. Basicamente o que acontece é que o exemplo que se encontra no conjunto de validação é complicado de classificar, sendo que normalmente este é o único falso negativo que ocorre. Quanto à Precisão existem alguns cortes que são detetados quando não existiram, com a atenuante de se verificarem em datas que tal não estava programado.

Por fim, foi classificado o conjunto de teste. Este conjunto é bastante mais reduzido dado apenas ser definida uma FGC e uma área VEG, sendo estudados os anos de 2017 e 2018. O conjunto é constituído por 48 exemplos (em média duas observações por mês, sendo representativo), contudo apenas um dos exemplos é classificado com a ocorrência de uma intervenção (semelhante ao que sucedeu em relação a Serra de Aire e Candeeiros). Note-se que a pré-classificação deste conjunto foi feita por observação das imagens de satélite da região durante estes dois anos. Definiu-se então que as operações ocorrem em março de 2018, uma altura em que os valores das bandas e índices ainda não começaram a decrescer por causa da fenologia, não mascarando a intervenção, como se pode observar na Figura 5.22 e Figura 5.23. Verifica-se uma alteração comportamental do NDVI mais clara do que no caso de Serra de Aire e Candeeiros. Este resultado também tem uma importância acrescida dado que este último conjunto apenas foi classificado uma vez, não tendo influência no processo de treino. Também se trata de uma região diferente da utilizada durante a fase treino, o que revela a possível aplicabilidade deste método a outras regiões.

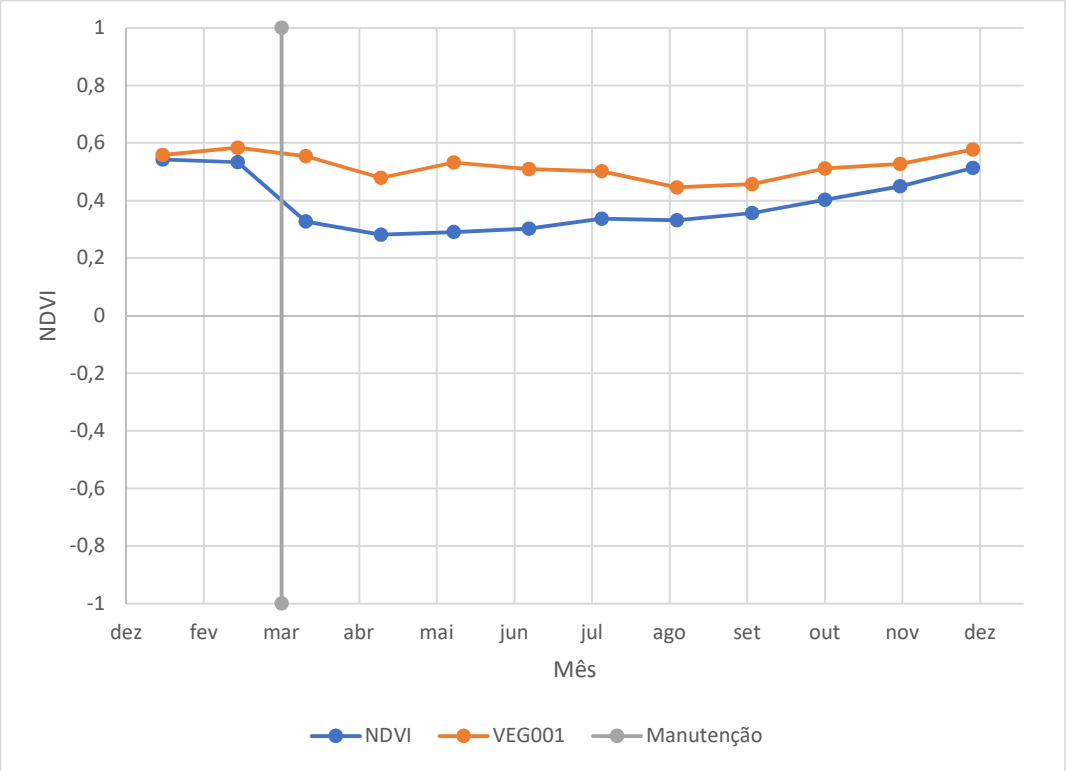


Figura 5.22: Evolução Temporal do NDVI em Marisol em 2018.

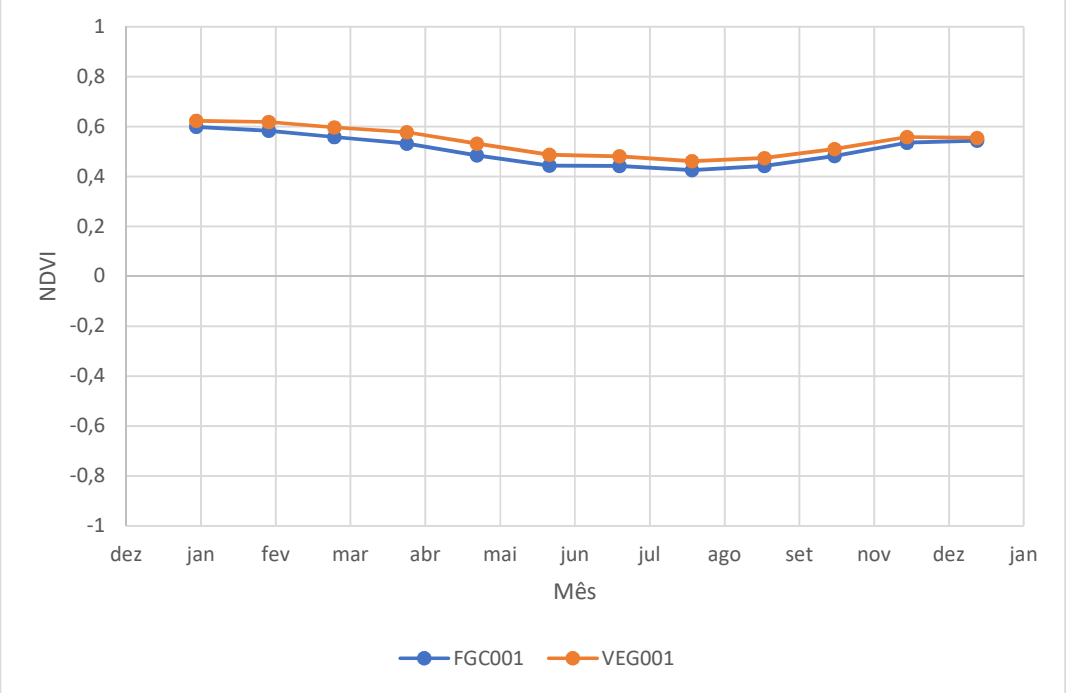


Figura 5.23: Evolução Temporal do NDVI em Marisol em 2017.

6 Conclusões e Trabalho Futuro

6.1 Conclusão

Em qualquer problema de extração de conhecimento, a análise prévia dos dados é um ponto bastante importante. A primeira fase desta dissertação centrou-se no entendimento dos dados, tendo-se verificado vários aspetos interessantes. Em primeiro lugar a sensibilidade das medições à fenologia da vegetação. Este efeito verifica-se ao longo do ano na maioria dos índices espectrais abordados em 3.1 e também nas bandas em dois grandes períodos. Um período relativo ao Inverno em que existe mais água nas regiões observadas e o Verão com níveis mais baixos de água. Sendo que o espectro na região do infravermelho é muito sensível a este fenómeno. Outro pormenor é o facto da transição entre estes períodos ter um comportamento similar ao de uma operação de intervenção. Como consequência índices que à partida seriam mais importantes, como o NDVI ou EVI, revelam-se ineficazes na deteção de intervenções. Note-se que se as operações forem executadas nestes períodos de transição[12] correm o risco de ser mascaradas pelos efeitos sazonais. Isto foi o que aconteceu no caso de Serra de Aire e Candeeiros, cujo corte foi realizado entre junho e julho.

Desta primeira análise foi possível concluir que os atributos fundamentais para a classificação têm como base o espectro visível. Os índices ExR e MExG, por exemplo, são bastante robustos sendo pouco afetados pela fenologia detetando com mais eficácia as intervenções. Outro aspeto a realçar é a localização temporal das operações. No primeiro caso de estudo apesar da operação estar registada em julho, pelas análises presentes em 5.1, verifica-se que o corte se terá iniciado em junho. Em junho e julho verificam-se duas dinâmicas um pouco diferentes o que pode originar confusão no classificador durante o processo de treino.

[12] Um período de transição é a fase do ano em que ocorrem alterações fenológicas na vegetação refletindo-se na reflectância medida.

Verificou-se que a resolução temporal, especialmente quando o segundo satélite da constelação *Sentinel 2* entrou em orbita, permite a deteção em tempo útil destas alterações. O mesmo sucede com a resolução espacial. Apesar de ao contrário de muitas das aplicações de *Remote Sensing* que estudam áreas com vários pixéis de informação, a resolução espacial das plataformas pode tornar-se um fator de exclusão, o que não ocorre com este caso. Porém, quanto à possibilidade de utilização de dados do *Landsat 8*, a resolução espacial de 30m pode não ser suficiente. Note-se que este último tem uma resolução temporal muito menor (15 dias), sendo que utilizado sozinho não é suficiente para uma deteção das operações em tempo útil.

A reduzida extensão das FGC e a ocorrência de pequenos erros de georreferenciação nas imagens obtidas pelo *Seninel 2* exposta nesta dissertação comprovou a necessidade de uma etapa no pré-processamento dos dados que corrija estes desvios. Caso contrário ocorrerão discrepâncias nos dados induzindo o classificador em erro.

A presente dissertação propunha-se identificar de forma automática intervenções nas FGCs, através da análise de imagem e com recurso a técnicas de aprendizagem automática. Em 5.3 verificou-se que se obtiveram resultados razoáveis no cumprimento deste objetivo. Os erros surgiram em transições de períodos identificados como corte ou quando como acontece em Serra de Aire e Candeeiros em que a intervenção ocorre durante dois meses, sendo que por vezes só é assinalado um dos meses. Contudo é de realçar que no caso de teste, a zona da Marisol, em que o corte foi executado em março, fora de um período de transição e com uma localização temporal melhor definida, a operação foi corretamente identificada, sem que o conjunto de exemplos tenha sido utilizado em qualquer fase de treino ou ajuste do classificador.

O último ponto a referir é o facto de não existir muita informação disponível quanto às datas das intervenções nas FGCs. A informação disponibilizada pelo ICNF já informa sobre o plano de localização das FGCs, se estão instaladas e se necessitam de manutenção, porém as datas de intervenção que permitirão classificar os conjuntos de treino não se encontram disponíveis publicamente. Relativamente a Serra de Aire e Candeeiros a data foi obtida junto do ICNF, mas esta informação só existe para algumas FGCs. Isto leva a conjuntos de dados mais reduzidos, podendo diminuir a qualidade dos classificadores. Em suma, os pontos a assinalar são os seguintes:

- A análise de dados desempenhou um papel crucial na resolução do problema;
- As plataformas (conceito definido em 2.1) existentes já fornecem dados suficientes para a resolução do problema;
- As intervenções nas FGCs não devem ser executadas em periodos de transição;
- O acesso alargado às datas de intervenção permitirá gerar conjuntos de treino mais completos.

6.2 Trabalho Futuro

Após este estudo existem várias tarefas a realizar na perspetiva de melhoramento do mesmo e de obtenção de mais conhecimento. Inicialmente é importante aplicar os mesmos métodos, mas com recurso aos produtos *Sentinel 2* de nível 2A. Verificar-se-á assim se existem vantagens na utilização destes produtos neste problema. Seguidamente deve-se avaliar a

aplicabilidade das imagens fornecidas pelo *Landsat 8* como forma de colmatar as falhas resultantes do desperdício de imagens, especialmente no Inverno, em consequência da nebulosidade que então ocorre. Note-se também que o *Landsat 9* será lançado no final de 2020, e como fornecerá dados praticamente iguais aos do seu antecessor, é importante avaliar esta última questão.

A automatização do processo de eliminação de observações por causa das nuvens, durante este estudo foi feito de forma manual.

A classificação por pixel também seria interessante. Este método pode permitir verificar a existência de focos dentro das FGCs que necessitam de maior atenção. Por exemplo, se é criado um caminho que conecte as duas zonas de vegetação separadas pela FGC.

A realização de conjuntos de dados maiores permitindo um treino mais adequado e uma validação mais precisa. Neste estudo detetaram-se intervenções, também será interessante verificar quando estas voltam a ser necessárias. Uma continuação deste trabalho será a análise de evolução da vegetação e quando será preciso ser cortada. Para tal seria interessante a avaliação do problema através de classificadores difusos em que a classificação com recurso a níveis de pertença dará uma maior sensibilidade necessário ao problema.

Por fim, a implementação de um *plugin* para o QGIS que execute todas estas tarefas e classifique as regiões pedidas.

Sumariamente o trabalho futuro pretende resolver os seguintes pontos:

- Utilização de produtos de nível 2A do *Sentinel 2*;
- Analisar a aplicabilidade das observações do programa *Landsat*;
- Automatização do processo de eliminação de observções com nuvens;
- Aplicar a classificação por pixel;
- Elaboração de novos conjuntos de treino;
- Deteção de zonas com necessidade de manutenção;
- Extensão de um *plugin* em QGIS.

Referências

Ahmed, B., & Al Noman, A. (2015). Land Cover Classification for Satellite Images based on Normalization Technique and Artificial Neural Networl (pp. 26–27).

Clerc, S., & MPC Team. (2019). *S2 MPC - L1C Data Quality Report - ESA*.

Déchoz, C., Poulain, V., Massera, S., Languille, F., Greslou, D., Lussy, F. De, … Europe, D. (n.d.). Sentinel 2 Global Reference Image.

DPFVAP. (2014). Manual de rede primária, 1–39.

Du, Y., Zhang, Y., Ling, F., Wang, Q., Li, W., & Li, X. (2016). Water bodies' mapping from Sentinel-2 imagery with Modified Normalized Difference Water Index at 10-m spatial resolution produced by sharpening the swir band. *Remote Sensing*, *8*(4). https://doi.org/10.3390/rs8040354

E. Rumelhart, D., E. Hinton, G., & J. Williams, R. (1986). Learning representations by back-propagating errors. *Nature*, *323*, 533–536. Retrieved from https://www.iro.umontreal.ca/~pift6266/A06/refs/backprop_old.pdf

Fu, G., Liu, C., Zhou, R., Sun, T., & Zhang, Q. (2017). Classification for High Resolution Remote Sensing Imagery Using a Fully Convolutional Network. *Remote Sensing*, *9*(12), 498. https://doi.org/10.3390/rs9050498

Guizar-sicairos, M., Thurman, S. T., & Fienup, J. R. (2008). Efficient subpixel image registration algorithms, *33*(2), 156–158.

Hagolle, O. (2014). The product level names, how they work ? Retrieved from http://www.cesbio.ups-tlse.fr/multitemp/?p=3202

Hamuda, E., Glavin, M., & Jones, E. (2016). A survey of image processing techniques for plant extraction and segmentation in the field. *Computers and Electronics in Agriculture*, *125*, 184–199. https://doi.org/10.1016/j.compag.2016.04.024

Hamunyela, E., Reiche, J., Verbesselt, J., & Herold, M. (2017). Using space-time features to improve detection of forest disturbances from Landsat time series. *Remote Sensing*, *9*(6), 1–17. https://doi.org/10.3390/rs9060515

Hermosilla, T., Wulder, M. A., White, J. C., Coops, N. C., & Hobart, G. W. (2015). An integrated Landsat

time series protocol for change detection and generation of annual gap-free surface reflectance composites. *Remote Sensing of Environment*, *158*, 220–234. https://doi.org/10.1016/j.rse.2014.11.005

Längkvist, M., Kiselev, A., Alirezaie, M., & Loutfi, A. (2016). Classification and segmentation of satellite orthoimagery using convolutional neural networks. *Remote Sensing*, *8*(4). https://doi.org/10.3390/rs8040329

Lopes, M., Fauvel, M., Girard, S., & Sheeren, D. (2017). Object-Based Classification of Grasslands from High Resolution Satellite Image Time Series Using Gaussian Mean Map Kernels. *Remote Sensing*, *9*(12), 688. https://doi.org/10.3390/rs9070688

Mestre, D., Fonseca, J. M., & Mora, A. (2017). Monitoring of in-vitro plant cultures using digital image processing and Random Forests. In *ICPRS - IET 8th International Conference on Pattern Recognition Systems*. https://doi.org/10.1049/cp.2017.0137

Mora, A., Santos, T., Łukasik, S., Silva, J., Falcão, A., Fonseca, J., & Ribeiro, R. (2017). Land Cover Classification from Multispectral Data Using Computational Intelligence Tools: A Comparative Study. *Information*, *8*(4), 147. https://doi.org/10.3390/info8040147

Tian, S., Zhang, X., Tian, J., & Sun, Q. (2016). Random Forest Classification of Wetland Landcovers from Multi-Sensor Data in the Arid Region of Xinjiang, China. *Remote Sensing*, *8*(12), 954. https://doi.org/10.3390/rs8110954

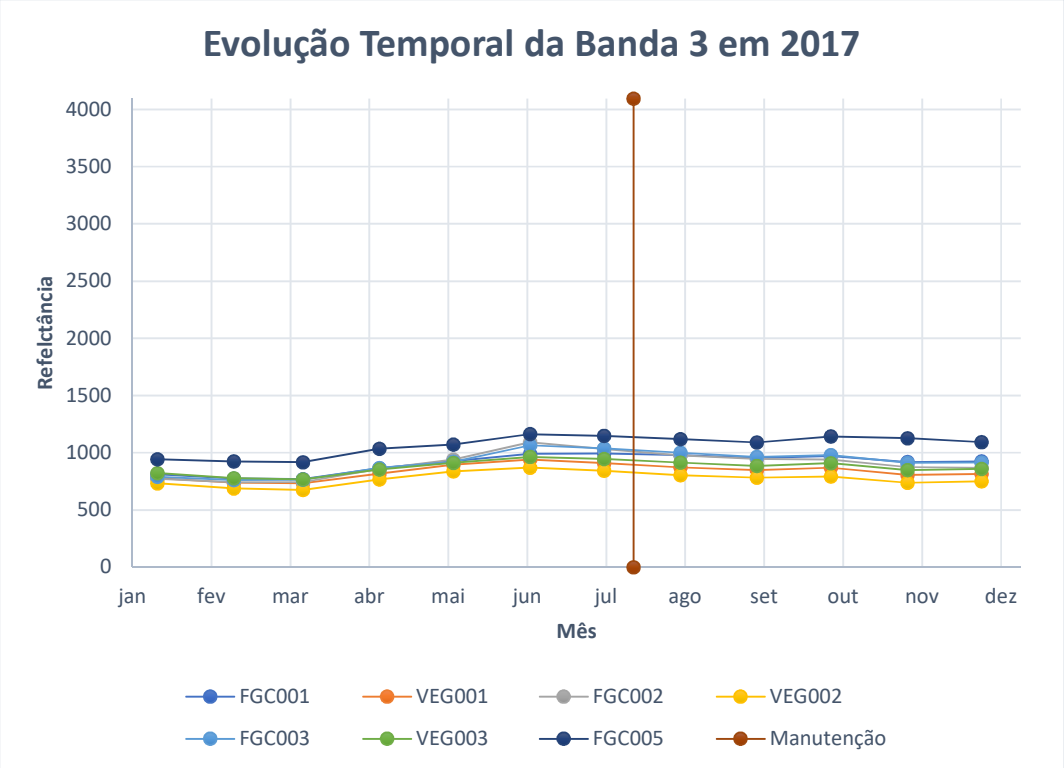
Tomlinson, R. F. (1969). A Geographic Information System for Regional Planning. *Journal of Geography*, *78*(1), 45–48. https://doi.org/10.5026/jgeography.78.45

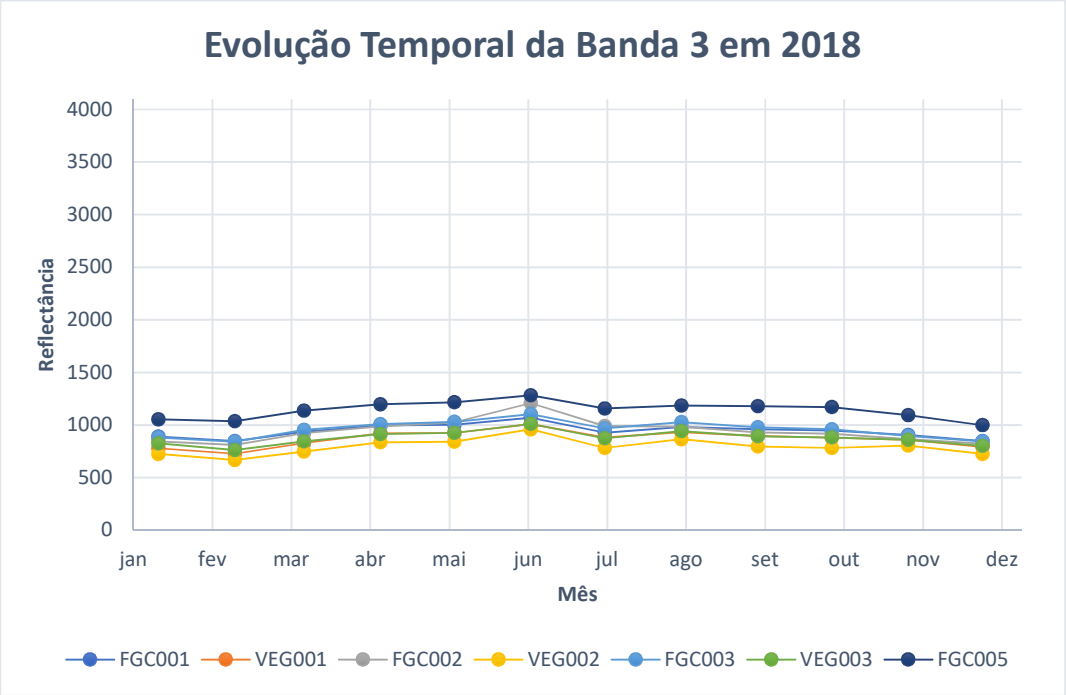
USGS. (2018). Landsat Levels of Processing.

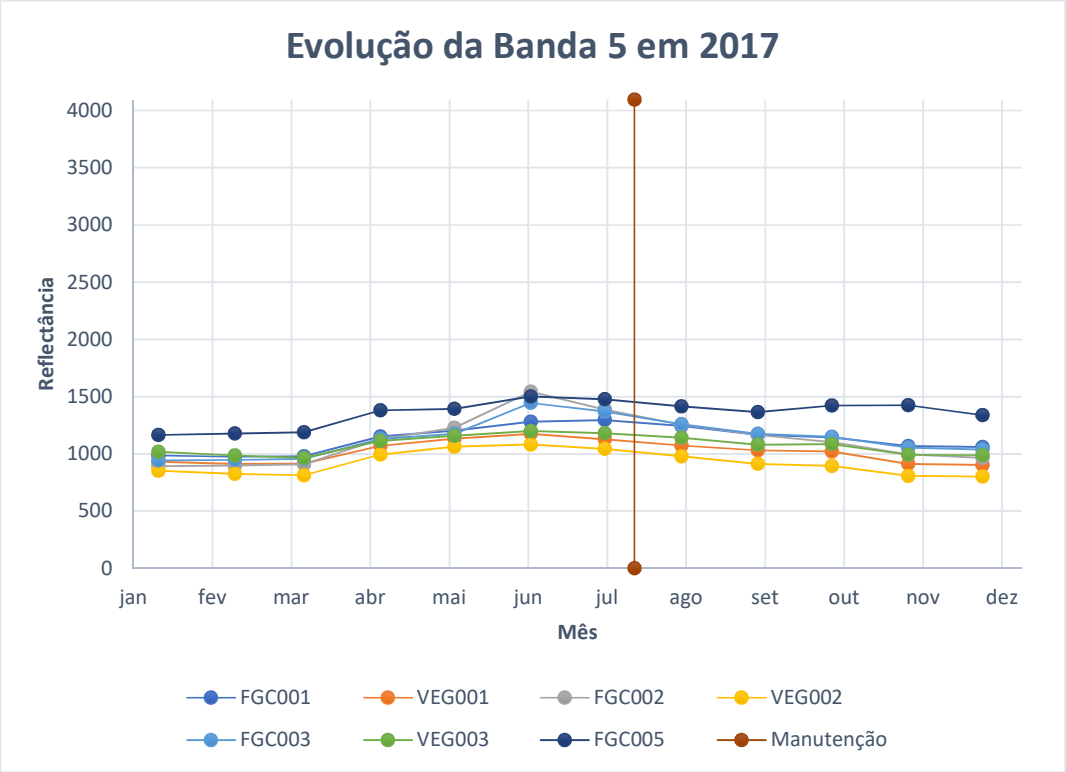
Xue, J., & Su, B. (2017). Significant remote sensing vegetation indices: A review of developments and applications. *Journal of Sensors*, *2017*. https://doi.org/10.1155/2017/1353691

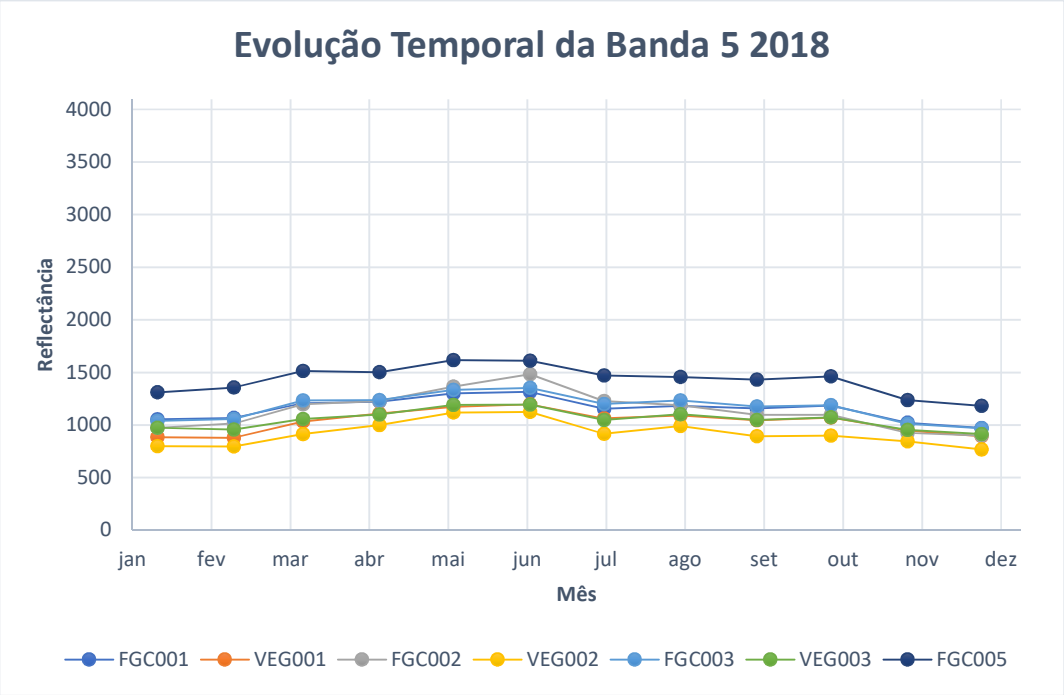
Anexos

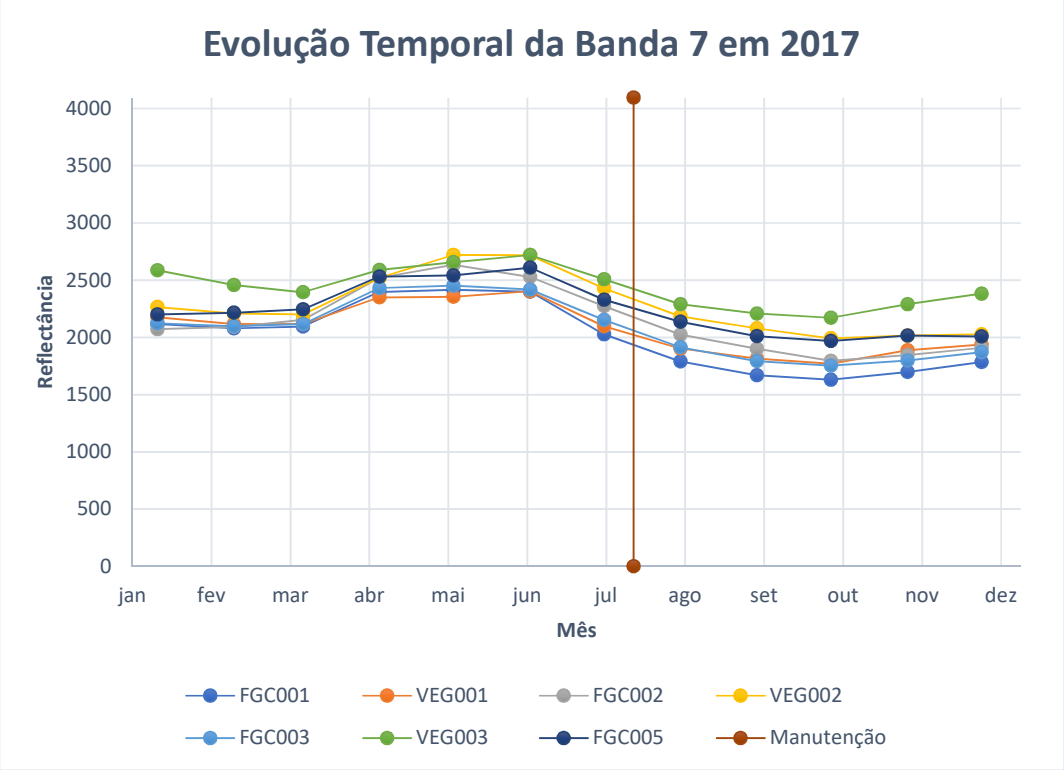
Reflectâncias das Bandas e dos Índices não apresentados:

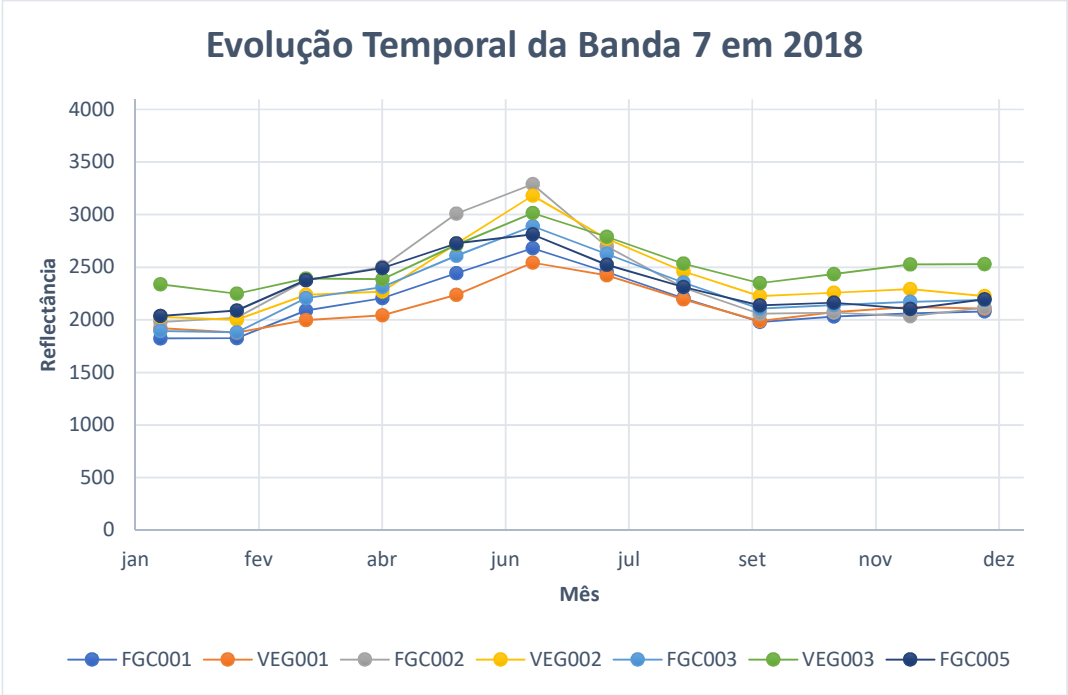


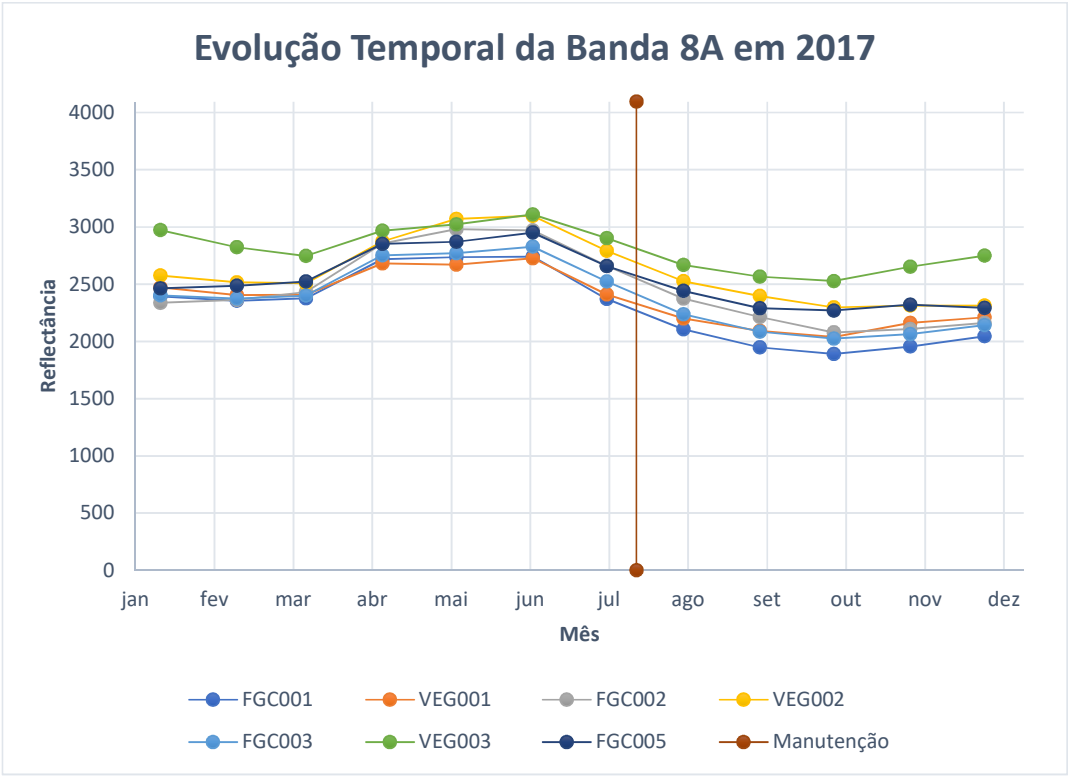

Evolução Temporal da Banda 3 em 2018
Reflectância
4000
3500
3000
2500
2000
1500
1000
500
0
jan
fev
mar
abr
mai
jun
jul
ago
set
out
nov
dez
Mês
FGC001
VEG001
FGC002
VEG002
FGC003
VEG003
FGC005

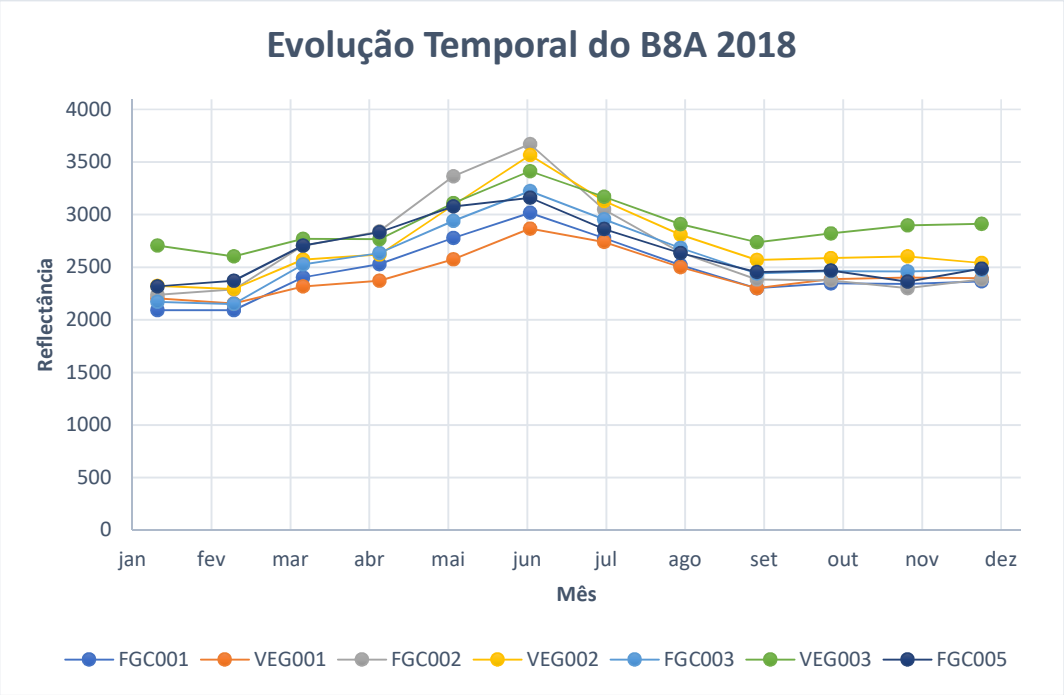

Evolução da Banda 5 em 2017
Reflectância
4000
3500
3000
2500
2000
1500
1000
500
0
jan
fev
mar
abr
mai
jun
jul
ago
set
out
nov
dez
Mês
FGC001
VEG001
FGC002
VEG002
FGC003
VEG003
FGC005
Manutenção

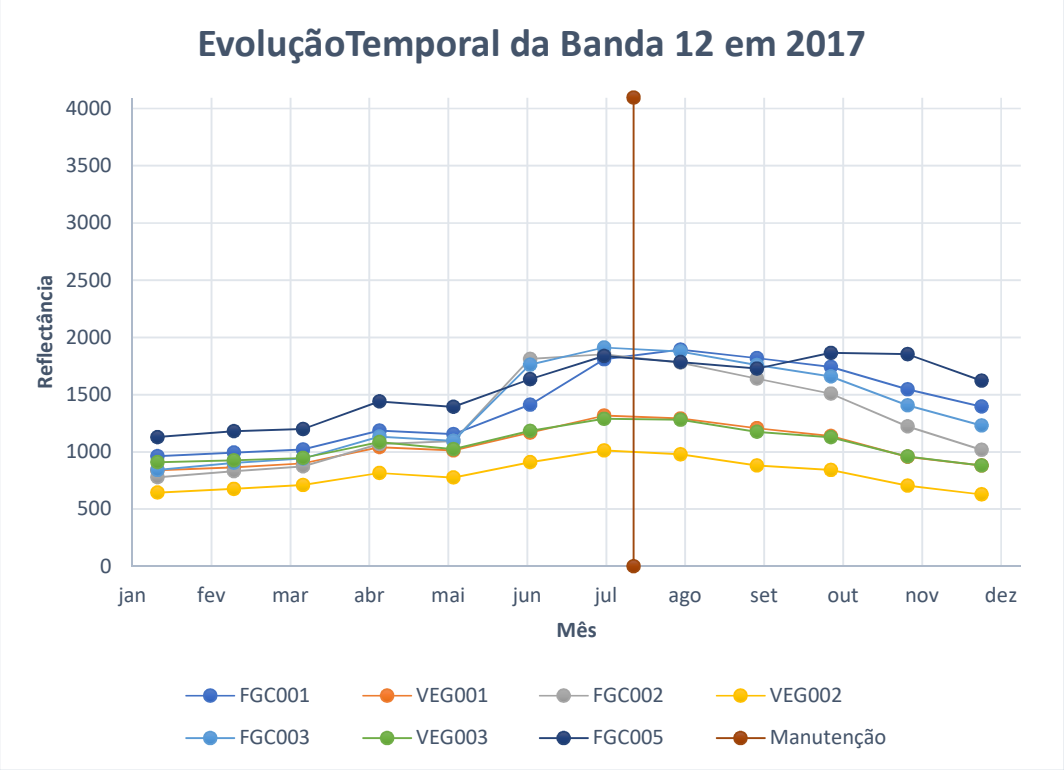

Evolução Temporal da Banda 5 2018
Reflectância
4000
3500
3000
2500
2000
1500
1000
500
0
jan
fev
mar
abr
mai
jun
jul
ago
set
out
nov
dez
Mês
FGC001
VEG001
FGC002
VEG002
FGC003
VEG003
FGC005

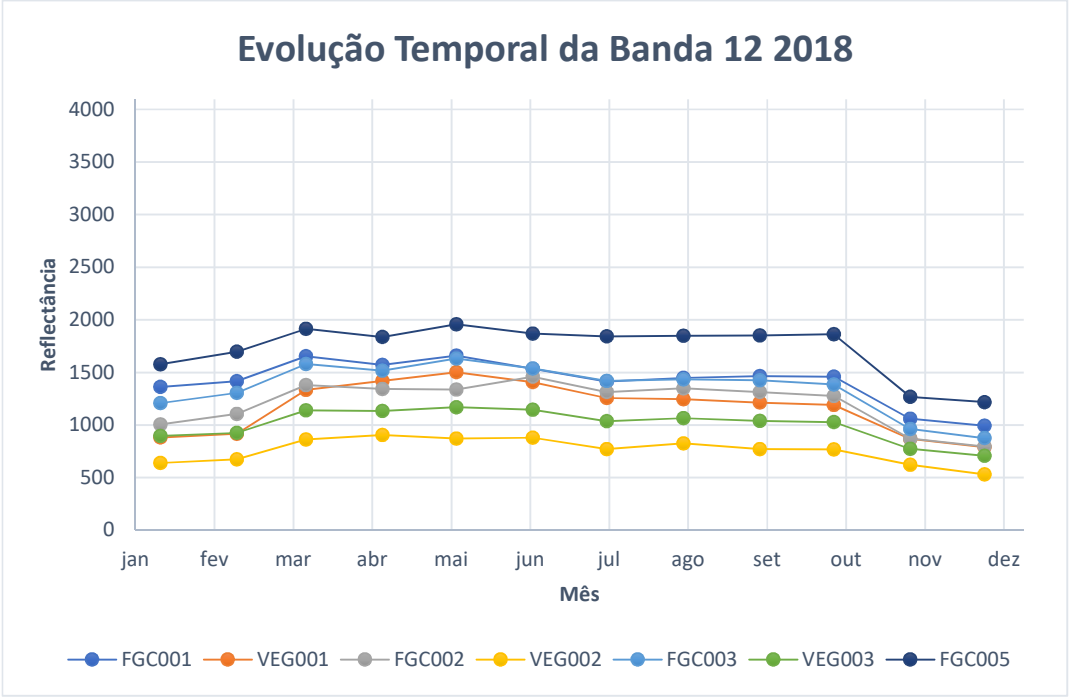

Evolução Temporal da Banda 7 em 2017
Reflectância
4000
3500
3000
2500
2000
1500
1000
500
0
jan
fev
mar
abr
mai
jun
jul
ago
set
out
nov
dez
Mês
FGC001
VEG001
FGC002
VEG002
FGC003
VEG003
FGC005
Manutenção

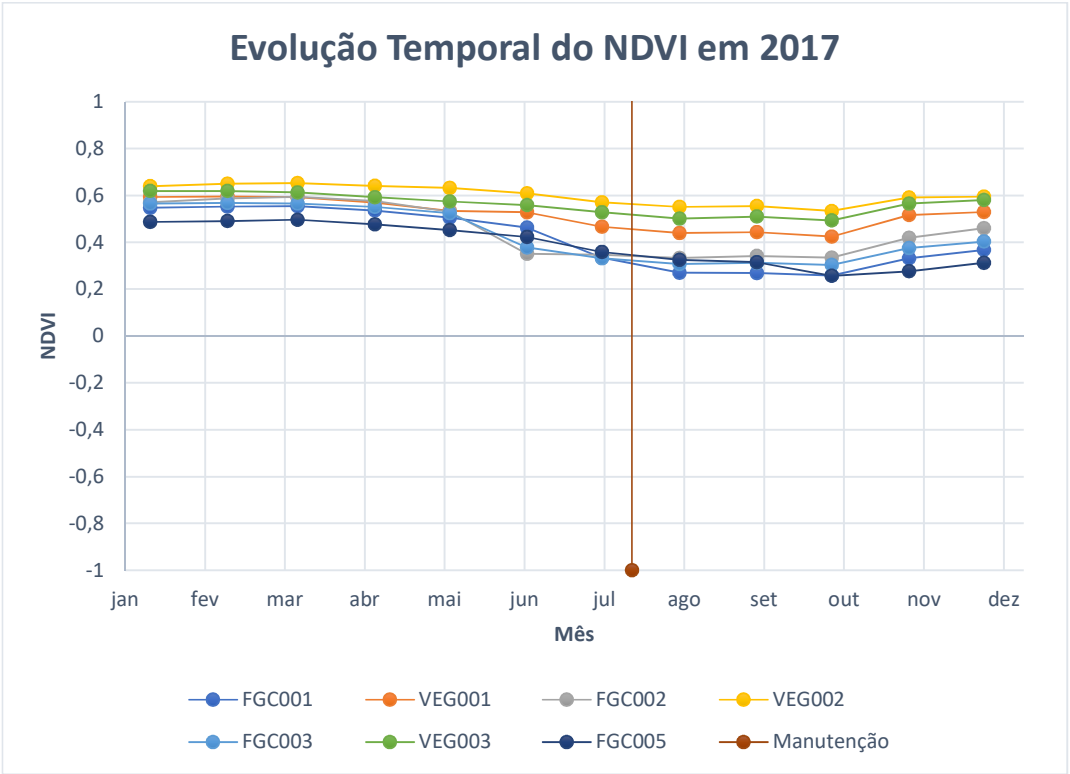

Evolução Temporal da Banda 7 em 2018
Reflectância
4000
3500
3000
2500
2000
1500
1000
500
0
jan
fev
abr
jun
jul
set
nov
dez
Mês
FGC001
VEG001
FGC002
VEG002
FGC003
VEG003
FGC005

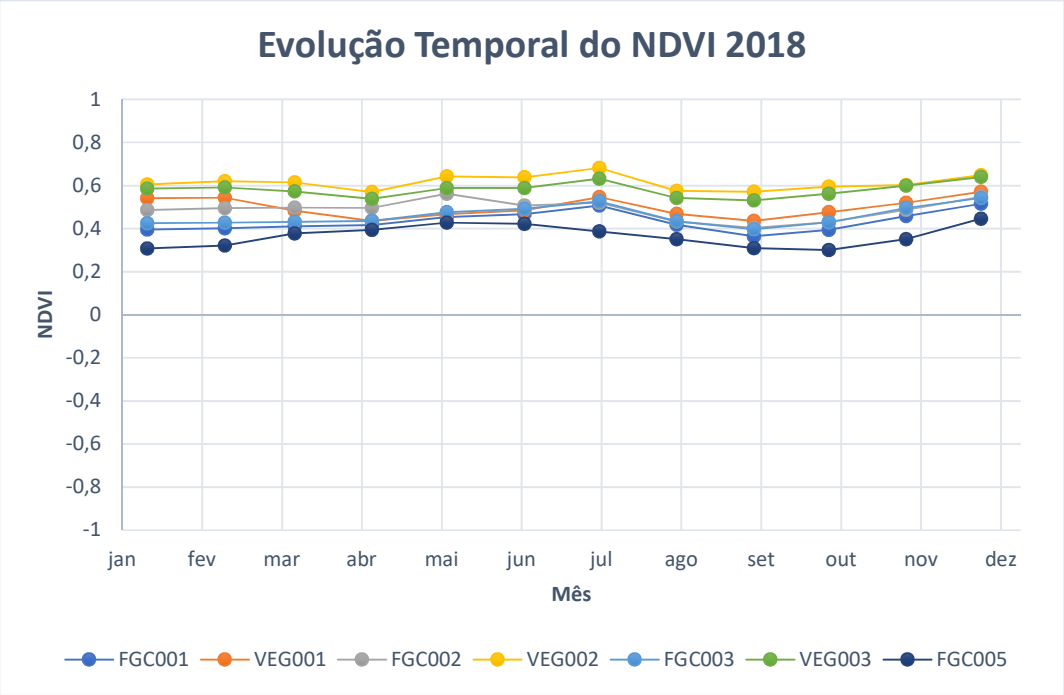

Evolução Temporal da Banda 8A em 2017
Reflectância
4000
3500
3000
2500
2000
1500
1000
500
0
jan
fev
mar
abr
mai
jun
jul
ago
set
out
nov
dez
Mês
FGC001
VEG001
FGC002
VEG002
FGC003
VEG003
FGC005
Manutenção

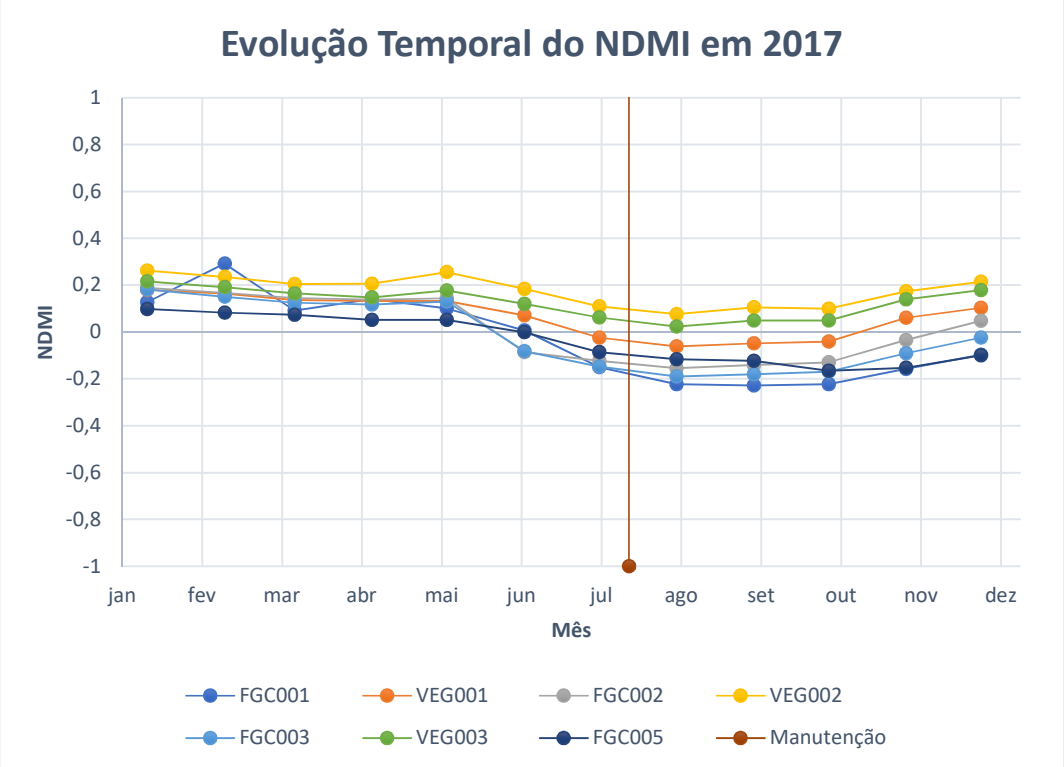

Evolução Temporal do B8A 2018
Reflectância
4000
3500
3000
2500
2000
1500
1000
500
0
jan
fev
mar
abr
mai
jun
jul
ago
set
out
nov
dez
Mês
FGC001
VEG001
FGC002
VEG002
FGC003
VEG003
FGC005

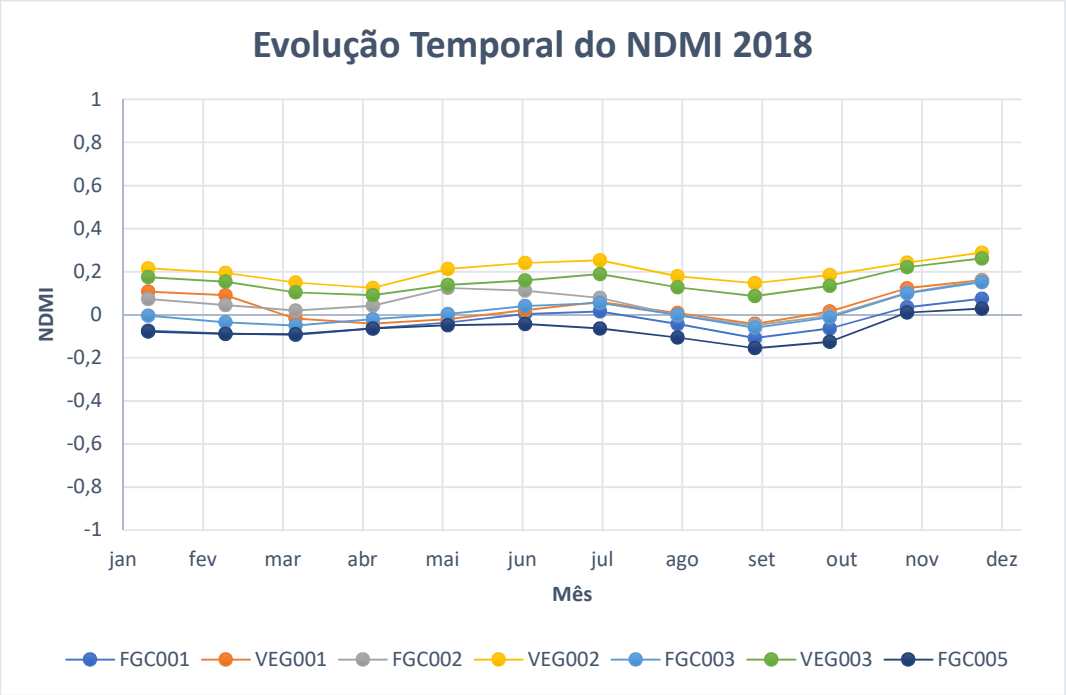

EvoluçãoTemporal da Banda 12 em 2017
Reflectância
4000
3500
3000
2500
2000
1500
1000
500
0
jan
fev
mar
abr
mai
jun
jul
ago
set
out
nov
dez
Mês
FGC001
VEG001
FGC002
VEG002
FGC003
VEG003
FGC005
Manutenção

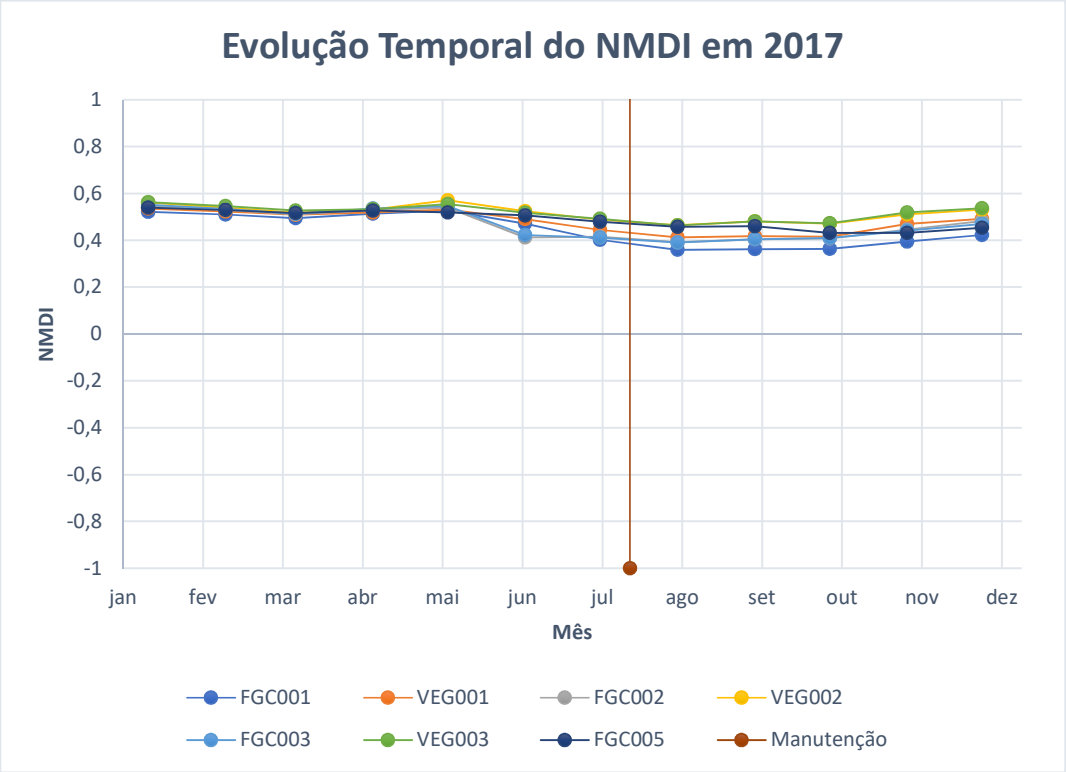

Evolução Temporal da Banda 12 2018
Reflectância
4000
3500
3000
2500
2000
1500
1000
500
0
jan
fev
mar
abr
mai
jun
jul
ago
set
out
nov
dez
Mês
FGC001
VEG001
FGC002
VEG002
FGC003
VEG003
FGC005

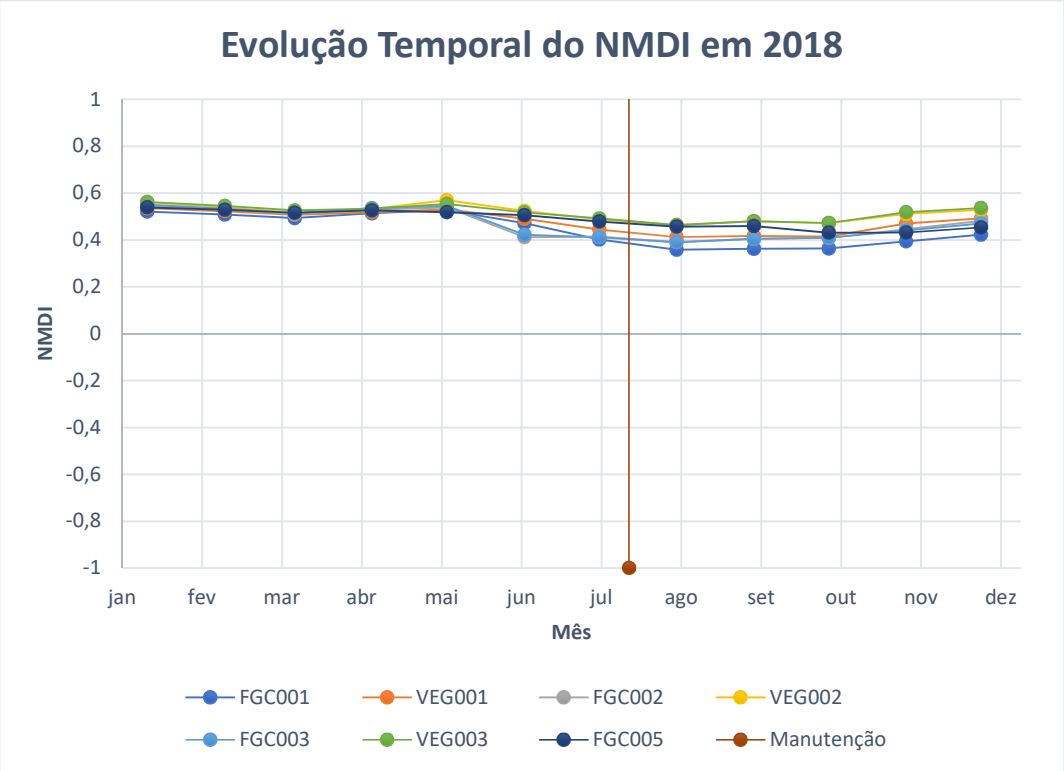

Evolução Temporal do NDVI em 2017
NDVI
1
0,8
0,6
0,4
0,2
0
-0,2
-0,4
-0,6
-0,8
-1
jan
fev
mar
abr
mai
jun
jul
ago
set
out
nov
dez
Mês
FGC001
VEG001
FGC002
VEG002
FGC003
VEG003
FGC005
Manutenção

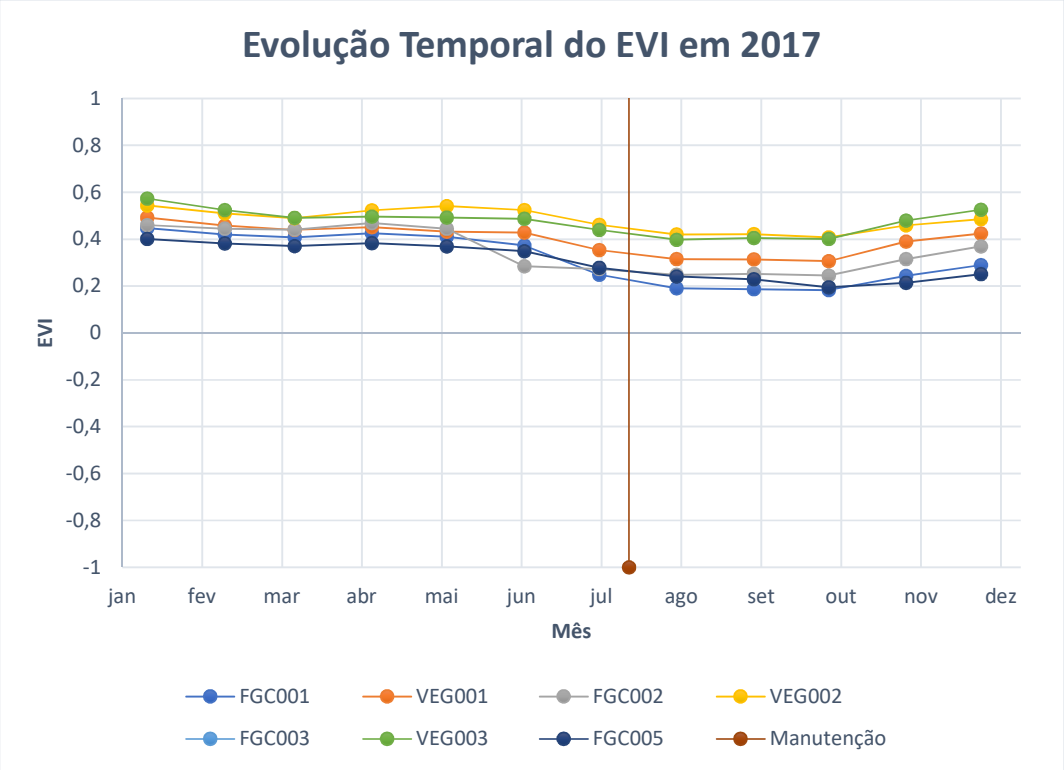

Evolução Temporal do NDVI 2018
1
0,8
0,6
0,4
0,2
0
-0,2
-0,4
-0,6
-0,8
-1
NDVI
jan
fev
mar
abr
mai
jun
jul
ago
set
out
nov
dez
Mês
FGC001
VEG001
FGC002
VEG002
FGC003
VEG003
FGC005

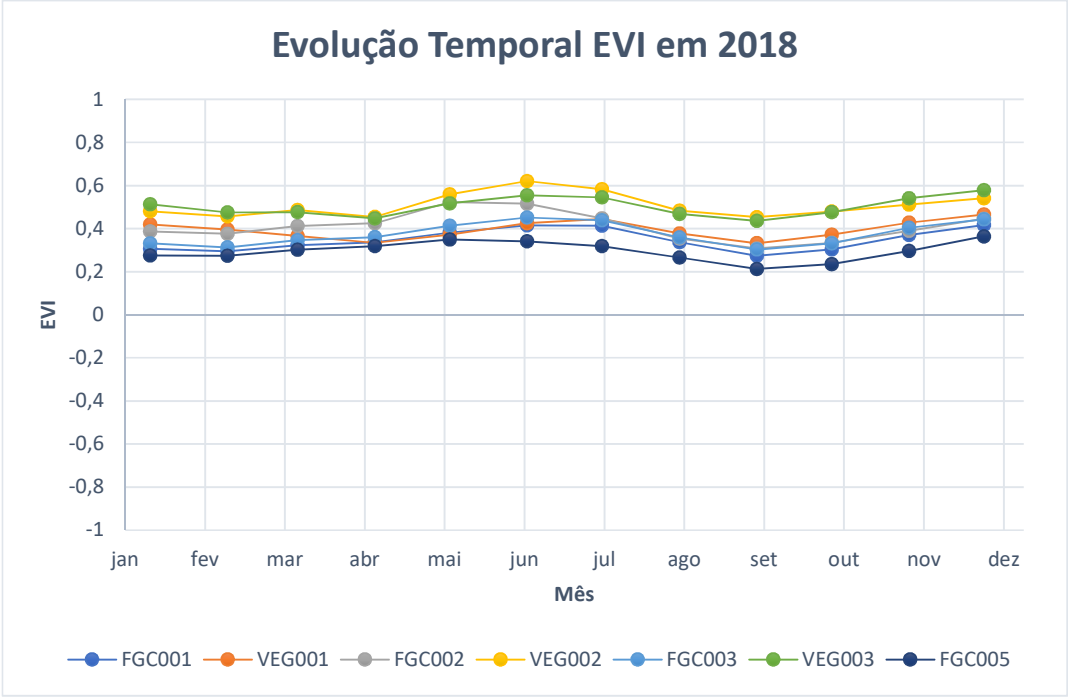

Evolução Temporal do NDMI em 2017
1
0,8
0,6
0,4
0,2
0
-0,2
-0,4
-0,6
-0,8
-1
NDMI
jan
fev
mar
abr
mai
jun
jul
ago
set
out
nov
dez
Mês
FGC001
VEG001
FGC002
VEG002
FGC003
VEG003
FGC005
Manutenção

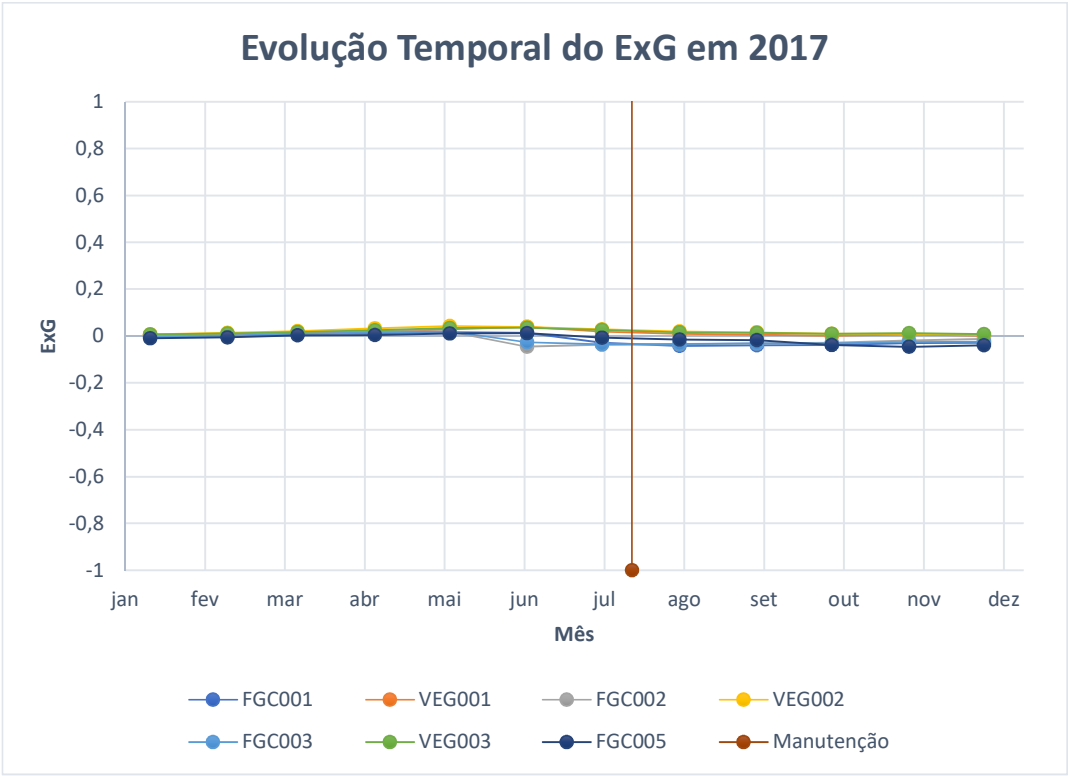

Evolução Temporal do NDMI 2018
1
0,8
0,6
0,4
0,2
0
-0,2
-0,4
-0,6
-0,8
-1
NDMI
jan
fev
mar
abr
mai
jun
jul
ago
set
out
nov
dez
Mês
FGC001
VEG001
FGC002
VEG002
FGC003
VEG003
FGC005

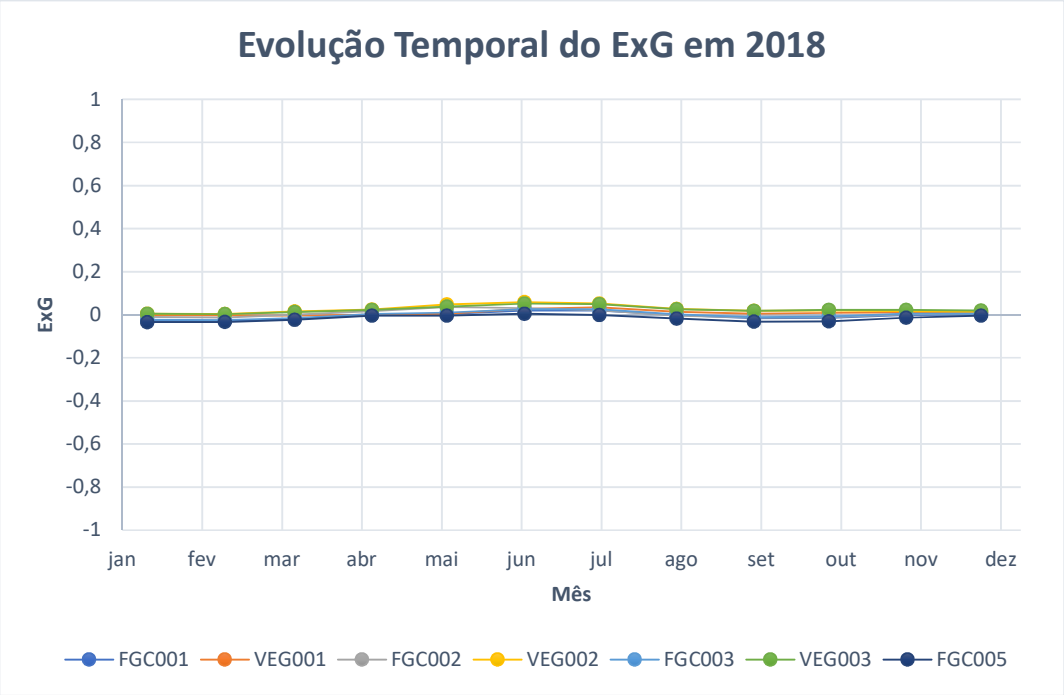

Evolução Temporal do NMDI em 2017
1
0,8
0,6
0,4
0,2
0
-0,2
-0,4
-0,6
-0,8
-1
NMDI
jan
fev
mar
abr
mai
jun
jul
ago
set
out
nov
dez
Mês
FGC001
VEG001
FGC002
VEG002
FGC003
VEG003
FGC005
Manutenção

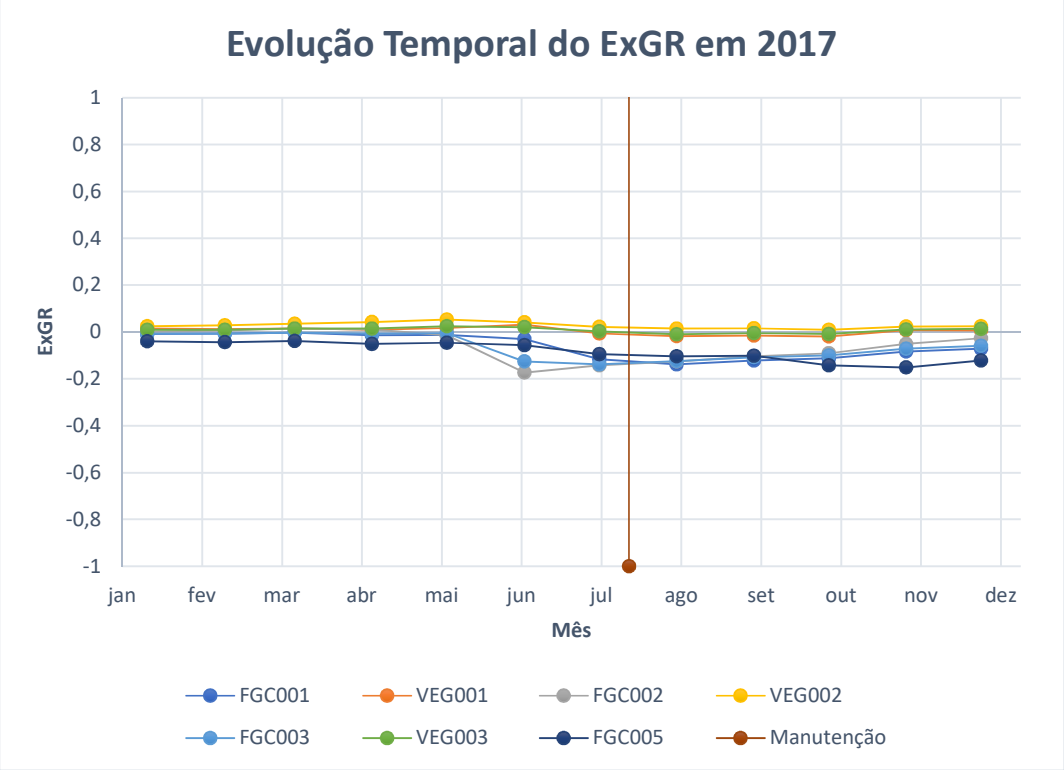

Evolução Temporal do NMDI em 2018
NMDI
1
0,8
0,6
0,4
0,2
0
-0,2
-0,4
-0,6
-0,8
-1
jan
fev
mar
abr
mai
jun
jul
ago
set
out
nov
dez
Mês
FGC001
VEG001
FGC002
VEG002
FGC003
VEG003
FGC005
Manutenção

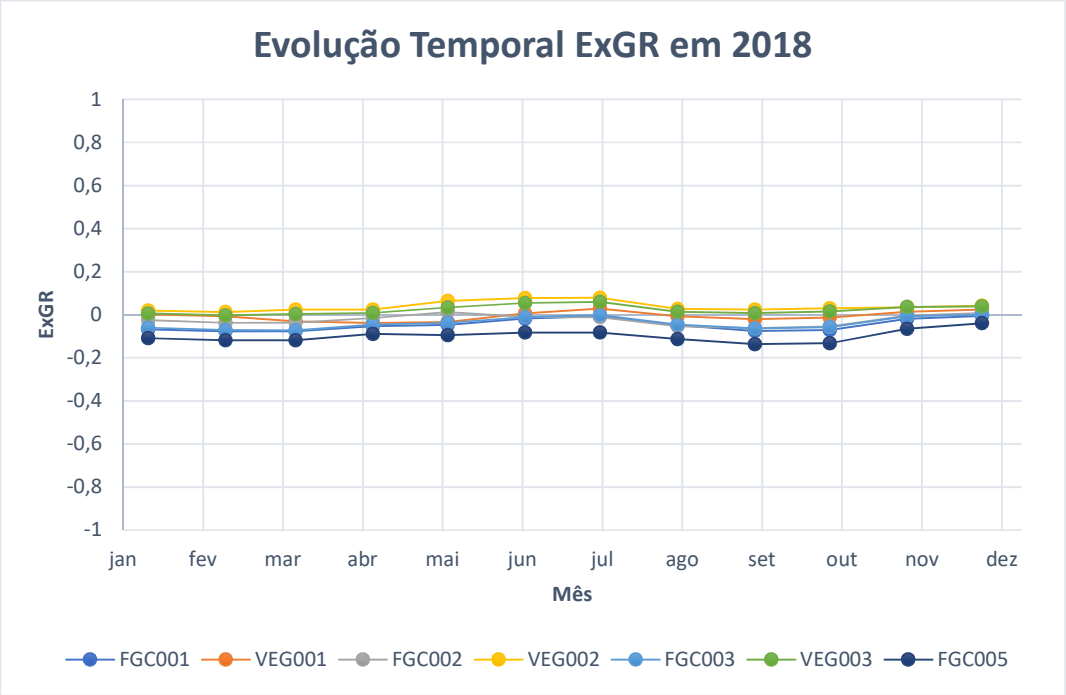

Evolução Temporal do EVI em 2017
EVI
1
0,8
0,6
0,4
0,2
0
-0,2
-0,4
-0,6
-0,8
-1
jan
fev
mar
abr
mai
jun
jul
ago
set
out
nov
dez
Mês
FGC001
VEG001
FGC002
VEG002
FGC003
VEG003
FGC005
Manutenção


Evolução Temporal EVI em 2018
EVI
1
0,8
0,6
0,4
0,2
0
-0,2
-0,4
-0,6
-0,8
-1
jan
fev
mar
abr
mai
jun
jul
ago
set
out
nov
dez
Mês
FGC001
VEG001
FGC002
VEG002
FGC003
VEG003
FGC005


Evolução Temporal do ExG em 2017
ExG
1
0,8
0,6
0,4
0,2
0
-0,2
-0,4
-0,6
-0,8
-1
jan
fev
mar
abr
mai
jun
jul
ago
set
out
nov
dez
Mês
FGC001
VEG001
FGC002
VEG002
FGC003
VEG003
FGC005
Manutenção


Evolução Temporal do ExG em 2018
ExG
1
0,8
0,6
0,4
0,2
0
-0,2
-0,4
-0,6
-0,8
-1
jan
fev
mar
abr
mai
jun
jul
ago
set
out
nov
dez
Mês
FGC001
VEG001
FGC002
VEG002
FGC003
VEG003
FGC005


Evolução Temporal do ExGR em 2017
ExGR
1
0,8
0,6
0,4
0,2
0
-0,2
-0,4
-0,6
-0,8
-1
jan
fev
mar
abr
mai
jun
jul
ago
set
out
nov
dez
Mês
FGC001
VEG001
FGC002
VEG002
FGC003
VEG003
FGC005
Manutenção


Evolução Temporal ExGR em 2018
1
0,8
0,6
0,4
0,2
0
-0,2
-0,4
-0,6
-0,8
-1
ExGR
jan
fev
mar
abr
mai
jun
jul
ago
set
out
nov
dez
Mês
FGC001
VEG001
FGC002
VEG002
FGC003
VEG003
FGC005