

# **Análise do potencial de dados Sentinel-2 na classificação da ocupação do solo no controlo de subsídios agrícolas.**

**Inês Silva Justiniano Machado**

Dissertação para a obtenção do Grau de Mestre em  
**Engenharia Agronómica**

Orientador: Doutor Manuel Lameiras de Figueiredo Campagnolo

**Júri:**

Presidente: Doutora Maria do Rosário da Conceição Cameira, Professora Associada do Instituto Superior de Agronomia da Universidade de Lisboa.

Vogais: Doutor João Catalão Fernandes, Professor Associado com Agregação da Faculdade de Ciências da Universidade de Lisboa;

Doutor Manuel Lameiras de Figueiredo Campagnolo, Professor Associado do Instituto Superior de Agronomia da Universidade de Lisboa.



## **Agradecimentos**

Àqueles que me inspiraram e motivaram e que contribuíram direta ou indiretamente para a finalização desta dissertação, o meu muito obrigada.

Ao meu orientador, Professor Doutor Manuel Campagnolo, agradeço o seu apoio, orientação, total disponibilidade e incansável paciência. O seu interesse e partilha de conhecimentos foram essenciais para a realização desta dissertação.

Ao Instituto de Financiamento da Agricultura e Pescas, em particular ao Vitor Carmona e ao Manuel Camacho do departamento de controlo pela cedência de dados, disponibilidade e troca de impressões fundamentais para a compreensão deste estudo num contexto real.

Aos meus pais, Carlos e São, pela coragem, compreensão, incentivo, paciência e apoio demonstrado ao longo destes anos.

Aos meus irmãos Igor e Fábio, pela inspiração e por todos os momentos de boa disposição assim como de realização partilhados. Agradeço ainda a motivação, entusiasmo e apoio.

A todos os meus amigos que me acompanharam neste percurso agradeço a partilha de bons momentos e a mitigação dos maus. Em especial à Inês Almeida pelos longos anos de cumplicidade e por ter sempre acreditado em mim.

À minha sobrinha Sofia, pelos momentos de descontração, ternura e carinho mesmo em situações mais difíceis.

## Resumo

Os agricultores submetem anualmente as suas candidaturas aos subsídios agrícolas no âmbito da Política Agrícola Comum (PAC). Cabe às administrações dos Estados-Membros estabelecer um sistema de controlo fiável e eficaz para garantir a validade das declarações dos agricultores e o cumprimento dos critérios de elegibilidade. O controlo por deteção remota (CwRS) tem demonstrado ser um método importante no apoio aos controlos de superfície (OTSC) através da fotointerpretação assistida por computador (CAPI) para a identificação de culturas.

O principal objetivo deste estudo foi analisar o potencial das imagens Sentinel-2A (S2A) para a identificação das culturas, quanto ao tipo e à estação, enquanto combinadas com dados Landsat-7 (L7) e Landsat-8 (L8) aplicando algoritmo de classificação *Random Forest* (RF).

Para acompanhar a fenologia das culturas em diferentes fases, as imagens L7, L8 e S2A foram adquiridas entre outubro de 2015 e agosto de 2016 sobre Beja (Portugal).

Foram testados dois cenários diferentes pelo algoritmo RF para cada caso de identificação: usar apenas bandas espectrais ou usar as bandas mais os índices de vegetação (IV) - normalized difference vegetation index (NDVI), green normalized difference vegetation index (GNDVI), enhanced vegetation index (EVI), soil adjusted vegetation index (SAVI) e modified soil adjusted vegetation index (MSAVI).

O algoritmo RF foi calibrado e validado utilizando dados de controlo validados e fornecidos pelo Instituto de Financiamento da Agricultura e Pescas (IFAP). As classificações foram executadas usando uma abordagem assistida baseada em pixels e em objetos e as precisões foram obtidas a partir da validação cruzada 10-fold. A avaliação da precisão baseou-se na comparação entre dados da classificação e de referência. Os resultados mostraram que a melhor precisão global de 99.7% foi obtida na distinção entre estações quando se utiliza uma abordagem ao nível do pixel e os IV. Para a classificação de culturas a maior precisão global de 98.22% foi obtida quando se usa a mesma abordagem e a combinação entre bandas espectrais e IV. Na classificação das parcelas, obteve-se uma precisão global de 92.8%, sendo a precisão das culturas sempre maior ou igual a 79.2%, sendo as culturas de cereais que apresentam a maior confusão na classificação.

**Palavras-chave:** classificação de ocupação do solo, Sentinel-2A MSI, Política Agrícola Comum, *Random Forest*.

## **Abstract**

Farmers submit annually their applications requiring agricultural subsidies in the context of the Common Agriculture Policy (CAP). It's up to the member state administrations to set up a control system reliable and effective to ensure farmers' declarations validity and compliance with eligibility criteria. Control with Remote sensing (CwRS) has been revealed as an important method in supporting On-The-Spot Check (OTSC) work through computer-assisted photo interpretation (CAPI) for crop identification.

The main objective of this study was analysing the Sentinel-2A (S2A) imagery potential for crop and crop season identification while combining with Landsat-7 (L7) and Landsat-8 (L8) data using Random Forest (RF) classification algorithm.

To keep up with crops phenology at different stages, L7, L8 and S2A images were acquired between October 2015 and August 2016 over Beja (Portugal). Two different scenarios were tested for the classification for each identification case: using only spectral bands or using bands plus vegetation indices (VI) - the normalized difference vegetation index (NDVI), the green normalized difference vegetation index (GNDVI), the enhanced vegetation index (EVI), the soil adjusted vegetation index (SAVI) and the modified soil adjusted vegetation index (MSAVI).

A RF classifier algorithm was calibrated and validated using ground truth data provided by the Portuguese Control and Paying Agency. The classifications were performed using a supervised pixel and polygon-based approach and k-fold cross validation (k=10). Accuracy assessment was taken on by comparing predicted classification and reference data. The results have shown that the best overall accuracy of 99.7% is obtained in crop season classification when using a pixel-based approach and VI. For the crop classification the highest overall accuracy of 98.22% is obtained when using pixel-based approach and a combination between spectral bands with VI. For the classification of parcels, the overall accuracy is 92.8%, with the lowest accuracy being 79.2% for triticale.

**Keywords:** land cover classification, Sentinel-2A MSI, Common Agricultural Policy, Random Forest.

## Índice

Agradecimentos .....	i
Resumo.....	ii
Abstract.....	iii
Índice de figuras.....	vi
Índice de quadros.....	vii
Lista de abreviaturas .....	viii
1. INTRODUÇÃO .....	1
1.1. Política Agrícola Comum.....	1
1.2. Controlo de subsídios no âmbito da PAC.....	2
1.2.1. Controlo por deteção remota.....	3
1.3. Objetivos do estudo.....	5
2. UTILIZAÇÃO DE DETEÇÃO REMOTA PARA MONITORIZAÇÃO AGRÍCOLA: CLASSIFICAÇÃO DE CULTURAS.....	7
3. CONJUNTO DE DADOS UTILIZADOS .....	11
3.1. Caraterização da área de estudo .....	11
3.2. Dados de deteção remota utilizados .....	12
3.3. Parcelas agrícolas e classes de ocupação do solo .....	15
4. MÉTODOS.....	17
4.1. Pré-processamento.....	17
4.2. Nomenclatura.....	18
4.3. Índices de vegetação .....	18
4.4. Criação da base de dados utilizada no treinamento do modelo <i>random forest</i> .....	19
4.5. Análise da homogeneidade espectral nas parcelas.....	20
4.6. Classificação segundo o método Random Forest (RF) .....	21
4.7. Otimização: seleção de variáveis .....	22
4.8. Cenários da classificação.....	22
4.9. Validação e avaliação dos resultados da classificação.....	23
5. ANÁLISE E DISCUSSÃO DE RESULTADOS .....	25

5.1. Análise da homogeneidade espectral nas parcelas.....	25
5.2. Seleção de variáveis .....	25
5.3. Classificação ao nível do pixel .....	28
5.4. Classificação ao nível da parcela .....	29
6. CONCLUSÕES .....	33
BIBLIOGRAFIA .....	35

## Índice de figuras

<b>Figura 1.</b> Área de estudo e localização dos mosaicos Landsat 7, Landsat 8 e Sentinel-2A sobre o distrito de Beja a Sul de Portugal.....	12
<b>Figura 2.</b> Distribuição temporal das imagens multiespectrais utilizadas neste estudo e adquiridas entre outubro de 2015 e agosto de 2016 pelos respetivos sensores Landsat 7 (L7), Landsat 8 (L8) e Sentinel 2A (S2). Apenas as datas com baixa proporção de nuvens foi considerada sobre a área de estudo. ....	13
<b>Figura 3.</b> Calendário cultural 2015-2016. Inclui as fases de preparação do solo e os vários estados de desenvolvimento das culturas. ....	16
<b>Figura 4.</b> Esquemática da metodologia utilizada. ....	17
<b>Figura 5.</b> Nomenclatura de classificação. ....	18
<b>Figura 6.</b> Ilustração da organização dos dados. A tabela tem 4594 linhas e cada linha representa um pixel, com indicação da parcela a que pertence. As primeiras colunas indicam a localização do pixel, a parcela, e as variáveis que dependem da parcela (em particular, a cultura e a estação). As restantes colunas representam as variáveis espectrais, cada uma determinada pelo sensor, a data de aquisição da imagem e a banda ou índice considerado, havendo no total 386 colunas para as variáveis espectrais. Note-se que a tabela é construída após a reamostragem de todas as imagens para uma resolução espacial comum de 30 m.20	
<b>Figura 7.</b> Evolução da estimativa da precisão global obtida pela validação cruzada (10-fold) de acordo com o número de variáveis utilizado em cada ciclo de classificação. As variáveis foram previamente ordenadas por ordem decrescente de importância baseada nos valores MDG e foi removida a variável menos importante pelo que esta figura deve ser lida da direita para a esquerda. Qualquer que seja o dataset utilizado são necessárias no mínimo 7 variáveis para obter uma precisão global superior a 95%. Para obter uma precisão igual ou superior a 98% são necessárias pelo menos 24 variáveis e 63 para quando o dataset é constituído pelas refletâncias e índices de vegetação ou somente índices de vegetação respetivamente.....	26
<b>Figura 8.</b> Proporção das 40 variáveis mais importantes no ranking da MDG de acordo com o tipo de variáveis, refletâncias ou índices de vegetação, e o cenário.....	26
<b>Figura 9.</b> Distribuição temporal das variáveis selecionadas para a classificação relativamente (A) ao tipo de estação, utilizando apenas os IV; (B) ao tipo de cultura, utilizando a combinação bandas e IV, em relação a (C) que representa a totalidade de informação disponível. As 25 variáveis distribuídas temporalmente em (B) permitiram uma precisão superior a 95% enquanto que as 8 variáveis em (A) permitiram uma precisão global superior a 98%. ....	27
<b>Figura 10.</b> Proporção de pixels que foram classificados na classe mais frequente. ....	30
<b>Figura 11.</b> Proporção de pixels classificados por maioria por classe e por parcela.....	30



## Índice de quadros

<b>Quadro 1.</b> Dados de controlo no ano de 2016 e 2017 em Portugal. Estão representadas apenas as parcelas com área útil declarada (Manuel Camacho, IFAP, comunicação pessoal). .....	4
<b>Quadro 2.</b> Correspondência entre as bandas de cada sensor segundo a resolução espectral (nm) e espacial (m). Adaptado de Flood (2017). .....	13
<b>Quadro 3.</b> Nomes dos ficheiros input. Com yyymmdd a representar a data de aquisição, YYYYMMDD a data do ficheiro e RES a resolução. ....	15
<b>Quadro 4.</b> Atributos do conjunto de dados vetoriais fornecido pelo IFAP.....	15
<b>Quadro 5.</b> Dimensões das parcelas por classe de ocupação do solo consideradas neste estudo. Grande variabilidade no tamanho das parcelas. ....	16
<b>Quadro 6.</b> Descrição das variáveis extraídas para a construção da base de dados utilizada para a classificação de acordo com o seu tipo e origem.....	20
<b>Quadro 7.</b> Apresentação das regiões espectrais ou índices de vegetação a que correspondem às 25 e 8 variáveis seleccionadas correspondem à primeira e segunda coluna respetivamente. .....	27
<b>Quadro 8.</b> Precisão do utilizador (PU) e do produtor (PP) para cada classe e precisão global (PG) para o método de classificação Random Forest ao nível do tipo de cultura e da estação para ambos os datasets: B + IV (dataset que inclui refletâncias e índices de vegetação) e IV (dataset que inclui apenas índices de vegetação). Para cada caso foi usado o conjunto de variáveis ótimo originado pela MDG em cada classificação. A coluna Nvar representa o tamanho desse conjunto. ....	28
<b>Quadro 9.</b> Matriz de erro estimada por validação cruzada 10-fold resultante da aplicação do algoritmo RF aos 4594 pixels do dataset B + IV (dataset que inclui refletâncias e índices de vegetação). ....	28
<b>Quadro 10.</b> Matriz de erro média e valores da precisão do produtor (PP), do utilizador (PU) e global (PG) baseada no modelo RF final.....	31

## Lista de abreviaturas

<b>CAPI</b>	Fotointerpretação assistida por computador
<b>CE</b>	Comissão Europeia
<b>CwRS</b>	Controlo por deteção remota
<b>DR</b>	Deteção Remota
<b>EM</b>	Estado-Membro
<b>ESA</b>	European Space Agency
<b>EVI</b>	Enhanced Vegetation Index
<b>FEADER</b>	Fundo Europeu Agrícola de Desenvolvimento Rural
<b>FEAGA</b>	Fundo Europeu Agrícola de Garantia
<b>GNDVI</b>	Green Normalized Difference Vegetation Index
<b>HR</b>	High Resolution
<b>IFAP</b>	Instituto de Financiamento da Agricultura e Pescas
<b>IPMA</b>	Instituto Português do Mar e da Atmosfera
<b>IV</b>	Índices de Vegetação
<b>JRC</b>	Joint Research Centre
<b>L7</b>	Landsat-7 ETM+
<b>L8</b>	Landsat-8 OLI
<b>S2A</b>	Sentinel-2A MSI
<b>MDA</b>	Mean Decrease Accuracy
<b>MDG</b>	Mean Decrease Gini
<b>MSAVI</b>	Modified Soil-Adjusted Vegetation Index
<b>NDVI</b>	Normalized Difference Vegetation Index
<b>OP</b>	Organismo Pagador
<b>PAC</b>	Política Agrícola Comum
<b>PD</b>	Pagamentos Diretos
<b>PG</b>	Precisão global
<b>PP</b>	Precisão do produtor
<b>PU</b>	Precisão do utilizador
<b>RF</b>	<i>Random Forest</i>
<b>RPU</b>	Regime de Pagamento Único
<b>SAVI</b>	Soil-Adjusted Vegetation Index
<b>SCR</b>	Sistema de Coordenadas de Referência
<b>SIGC</b>	Sistema Integrado de Gestão e de Controlo
<b>SIP</b>	Sistema de Identificação Parcelar
<b>UE</b>	União Europeia

**USGS** United States Geological Survey  
**VHR** Very High Resolution



# 1. INTRODUÇÃO

De acordo com as projeções mais recentes das Nações Unidas, a população mundial deverá atingir 9.8 bilhões em 2050 (United Nations, Department of Economic Social Affairs, & Population Division, 2017). Se esta tendência se mantiver, estima-se que será necessário aumentar em quase 50% a produção agrícola mundial para atender às necessidades da população (FAO, 2017). Assim, garantir a segurança alimentar num contexto de crescimento populacional, de alterações climáticas e de aumento da pressão sobre os recursos naturais, torna urgente a conversão para sistema agroalimentar que contribua para um desenvolvimento ambiental, social e economicamente sustentável. Na União Europeia (UE) é a política agrícola comum (PAC) que desempenha um papel crucial na transição para uma agricultura mais sustentável. A atual vertente ecológica da PAC assenta na aplicação complementar de instrumentos políticos de incentivo à adoção de práticas agrícolas benéficas para o ambiente pelos agricultores.

## 1.1. Política Agrícola Comum

Criada em 1962 pela União Europeia (UE), a política agrícola comum (PAC) tem como objetivo principal assegurar aos cidadãos o aprovisionamento de alimentos seguros a preços acessíveis enquanto é garantido um nível de vida digno aos agricultores (Comissão Europeia, 2012). Desde então a PAC tem sido alvo de sucessivas reformas procurando adaptar-se aos novos desafios socioeconómicos, tecnológicos e ambientais com que se defronta o sector agrícola. A reforma de 1992 caracterizou-se essencialmente pela redução do apoio aos preços de mercado e pela introdução dos pagamentos diretos (PD) ligados à área cultivada para compensar a aproximação entre os preços comunitários e mundiais. Uma nova reforma da PAC em 2003, contou com a introdução do regime de pagamento único (RPU) dissociado da produção a favor do apoio aos rendimentos permitindo aos agricultores tornarem-se mais competitivos e direcionados para o mercado (Massot, 2018). Esta medida política protege os agricultores da volatilidade dos preços de mercado e compensa-os da baixa rentabilidade do setor agrícola face a outros setores económicos (Comissão Europeia, 2018a). Ainda nesta reforma foi adotado um conceito mais amplo de regras de condicionalidade (*cross-compliance*) que subordina os pagamentos únicos ao cumprimento obrigatório de uma série de critérios específicos no âmbito da conservação do ambiente, sanidade e bem-estar animal e segurança alimentar (Massot, 2018).

A etapa mais recente do processo de adaptação da PAC iniciou-se com a reforma de 2013 a qual ainda se encontra no seu período de aplicação. Apesar dos objetivos da PAC serem os mesmos – nomeadamente a produção alimentar viável, a gestão sustentável dos recursos naturais e alterações climáticas e o desenvolvimento territorial equilibrado em toda a UE

(Comissão Europeia, 2013b) –, foi necessária uma mudança substancial ao nível dos instrumentos políticos utilizados a fim de tornar uma PAC mais equitativa e equilibrada com promoção da atividade agrícola. Um novo sistema de PD foi introduzido tendo por base o pagamento sob a forma de apoio ao rendimento aplicável a todos os agricultores ativos por hectare elegível (regime de pagamento base) complementado por uma série de outros tipos de regimes de apoios (Comissão Europeia, 2013a). A grande inovação trazida a pretexto desta última reforma é a introdução dos pagamentos diretos por ecologização, mais à frente designado por *greening*, que tem como objetivo melhorar o desempenho ambiental da agricultura na UE. Esses pagamentos estão sujeitos a condições de elegibilidade relacionadas com a realização, pelos agricultores, de práticas agrícolas ambientalmente sustentáveis: manutenção de prados e pastagens permanentes existentes, diversificação de culturas e assegurar que pelo menos 5% das terras aráveis são superfícies de interesse ecológico. A estas práticas podem ainda ser acrescentadas outras práticas idênticas as quais deverão produzir um efeito no clima e no ambiente superior ou igualmente benéfico designadas por práticas equivalentes (Comissão Europeia, 2013a).

A novidade da reestruturação da PAC mais recente é o aumento do vínculo entre os dois pilares anteriormente existentes. O primeiro pilar financia integralmente as medidas de regulação dos mercados agrícolas e os PD aos agricultores que a cargo do FEAGA comporta respetivamente 10% e 70% do orçamento da PAC; o segundo pilar financia as medidas referentes a pagamentos no âmbito dos programas de desenvolvimento rural em regime de cofinanciamento que a cargo do FEADER e dos Orçamentos de Estado dos Estados-Membros (EM) comporta cerca de 20% do orçamento da PAC (Comissão Europeia, 2013b). Para 2018, o orçamento da UE prevê um total de 156.1 mil milhões de euros (a preços correntes) de verbas para esses dois fundos. No total, a PAC representa 37.6% do orçamento global da UE, i.e., 58.1 mil milhões de euros (Massot, 2018b).

## **1.2. Controlo de subsídios no âmbito da PAC**

Os pagamentos aos agricultores são geridos e controlados a nível nacional por cada membro da UE através de um organismo pagador (OP) nacional ou regional acreditado e designado pelos EM. Em Portugal, o OP responsável por desempenhar estas tarefas é Instituto de Financiamento da Agricultura e Pescas (IFAP). Os agricultores devem submeter anualmente os pedidos de ajuda junto dos respetivos OP, onde declaram a localização geográfica e a ocupação cultural explorada de todas as parcelas agrícolas das suas explorações (Comissão Europeia, 2013b). É da competência dos OP garantir a aplicação correta e equitativa dos fundos comunitário no âmbito da PAC e que quaisquer irregularidades nos pagamentos sejam detetadas e reavidas. Como consequência, os OP devem submeter todos os pedidos de ajuda a controlos administrativos e a controlos no local que, regra geral,

incidem apenas numa amostra mínima aleatória de 5% dos pedidos de ajuda por regime de ajuda à exceção do *greening* (Comissão Europeia, 2013b). Em termos práticos significa que para cada parcela agrícola declarada são verificados os seguintes elementos: localização, área e ocupação cultural (Loudjani, 2016). Estes controlos podem assumir a forma “clássica” de visitas ao terreno ou de aplicação da DR. A deteção de incumprimentos ou irregularidades significativas nas verificações no local implica o aumento da percentagem de controlos no ano seguinte até que seja atingido um nível admissível de conformidades dos pedidos de ajuda (Comissão Europeia, 2014).

O Sistema Integrado de Gestão e de Controlo (SIGC) é o sistema obrigatório mais importante na gestão e controlo dos pagamentos efetuados aos agricultores operado pelo OP no âmbito da PAC. O SIGC integra um sistema de identificação das parcelas agrícolas designado por Sistema de Identificação Parcelar (SIP), ou parcelário, cujo propósito é colaborar na gestão e no controlo dos PD e de outros regimes/medidas «superfícies» (relacionados com a área), e facilitar os controlos administrativos cruzados na deteção de declarações duplicadas ou excessivas (Comissão Europeia, 2015). O SIP consiste numa base de dados georreferenciada suportada por fotografias aéreas e imagens de satélite de alta precisão, que é utilizada na identificação dos limites e da respetiva ocupação/uso do solo de todas as parcelas agrícolas declaradas pelos agricultores e na quantificação da respetivas áreas elegíveis (Comissão Europeia, 2018b).

### **1.2.1. Controlo por deteção remota**

Desde 1993, a utilização do CwRS tem sido promovida pela CE como o método mais adequado para verificar se os subsídios são corretamente concedidos. O CwRS é executado através do uso de imagens de satélite que são adquiridas e submetidas a controlos de qualidade pelo Joint Research Centre e disponibilizadas posteriormente aos OP (Comissão Europeia, 2016b).

Presentemente, a técnica de CwRS oficialmente aceite no contexto da PAC é a fotointerpretação assistida por computador (CAPI). A CAPI consiste maioritariamente na interpretação visual de imagens multitemporais de alta resolução (HR) e de muito alta resolução (VHR) (ortofotos aéreas e orto-imagens de satélite com pixel de dimensão menor que 0.75 m). Os aspetos relacionados com a ocupação/uso do solo podem ser verificados por fotointerpretação de imagens de HR frequentemente apoiada por métodos de classificação de imagens automática ou semi-automática (assistida ou não assistida), não prescindindo, no entanto, da confirmação dos resultados através de uma inspeção visual. A verificação dos limites e da área das parcelas deve ser efetuada através do uso de imagens VHR. A forte necessidade de identificar o tipo de cultura ao nível da parcela surgiu com o aparecimento dos requisitos da diversificação de culturas estabelecidos no *greening*. Durante o processo

CAPI, caso sejam identificadas parcelas que levantem problemas ou dúvidas e que não permitam completar o diagnóstico, devem-se planejar visitas ao terreno como alternativa (Loudjani, 2016). No caso de ser necessário verificar situações passadas, o processo CAPI revela ser o processo mais adequado apesar de depender da disponibilidade de imagens de qualidade.

No

Quadro 1 são apresentados os números de controlos no local físicos e por DR realizados nos anos 2016 e 2017 em Portugal. Da totalidade dos controlos efetuados, apenas cerca de 28% das candidaturas foram submetidas a controlos por DR em 2016 e 23% em 2017 (dados provisórios).

**Quadro 1.** Dados de controlo no ano de 2016 e 2017 em Portugal. Estão representadas apenas as parcelas com área útil declarada (Manuel Camacho, IFAP, comunicação pessoal).

Controlos	Total beneficiários		Total parcelas	
	2016	2017*	2016	2017*
Físicos	10762	12256	117791	151989
Deteção remota	4169	3752	38334	27774
<b>Total</b>	<b>14931</b>	<b>16008</b>	<b>156125</b>	<b>179763</b>

(\*) dados provisórios

Os controlos no local, quer físicos quer por DR, capturam informação dependente da capacidade de interpretação visual dos foto-intérpretes o que resulta em informação subjetiva e irreproduzível. No CwRS, a utilização de um conjunto de imagens de satélite permite aos operadores aplicar o mesmo conceito, embora seja numa perspetiva e circunstâncias diferentes. O CwRS surge como uma das técnicas mais promissoras para redução dos encargos para os OP associados à natureza demorosa e onerosa das operações de controlo.

No contexto do *greening*, da condicionalidade e de algumas medidas de desenvolvimento rural, a monitorização torna-se crucial na verificação do cumprimento dos requisitos para a concessão de ajudas, o que implica a observação sistemática das parcelas agrícolas sujeitas a essas condições de elegibilidade com a consequência do acréscimo de encargos. Neste seguimento, o JRC publicou um documento de discussão no qual descreve conceitos e elementos que visam a substituição do corrente método de verificação no local por um novo sistema de monitorização que abrange a totalidade dos pedidos de ajuda, apelando à automatização dos métodos de DR. O principal objetivo desta abordagem é a simplificação e redução dos encargos com os controlos e consequente redução substancial de visitas ao terreno (Devos et al., 2017).



### 1.3. Objetivos do estudo

O principal objetivo deste estudo foi explorar a combinação de imagens multiespectrais e multitemporais do Sentinel-2A (MSI), Landsat-7 (ETM+) e Landsat-8 (OLI) na classificação automática da ocupação do solo no âmbito do controlo de subsídios no contexto da PAC. Paralelamente, pretendeu-se mostrar a forma como a classificação do tipo de cultura atempada e precisa pode auxiliar nos controlos de subsídios efetuados pelos OP, que através da redução do tempo despendido e dos custos que os controlos no local implicam, sugere um potencial do ponto de vista operacional na transição para um novo sistema de monitorização em larga escala atualmente em discussão.

#### Objetivos específicos

- Determinar a precisão da classificação usando apenas os índices de vegetação;
- Determinar se a combinação de bandas com os índices de vegetação melhora a precisão da classificação;
- Investigar a distribuição temporal/espectral das variáveis mais importantes selecionadas;
- Analisar se a precisão melhora quando a classificação é conduzida ao nível do pixel ou ao nível da parcela.

#### Questões

- Quais, os problemas associados à combinação de valores de refletância das imagens adquiridas pelos diferentes sensores: L7, L8 e Sentinel-2A?
- Quais são as combinações de sensores, datas, e bandas ou índices de vegetação que são mais relevantes para a identificação e discriminação de culturas?
- Qual a precisão da classificação da ocupação do solo resultante da aplicação dos métodos descritos?
- Como se pode tomar a decisão ao nível de cada parcela, e qual é a precisão da classificação das parcelas?



## **2. UTILIZAÇÃO DE DETEÇÃO REMOTA PARA MONITORIZAÇÃO AGRÍCOLA: CLASSIFICAÇÃO DE CULTURAS**

Na Europa, a monitorização agrícola desempenhada pelo JRC serve maioritariamente para diferenciar, identificar e medir as principais áreas de culturas produzidas, prever precocemente a produção e auxiliar no controlo de subsídios no contexto da PAC (Comissão Europeia, 2016a). O processo de monitorização das culturas requer séries temporais de imagens de alta resolução espacial e temporal para monitorizar todo o ciclo cultural para a identificação precisa das culturas.

Um pressuposto básico em DR é que superfícies específicas, tais como rocha, água ou vegetação têm um comportamento característico de reflexão e, portanto, podem ser distinguidas pela sua assinatura espectral específica. Com base no mesmo conceito, a classificação por DR beneficia da diferença existente entre assinaturas espectrais de culturas diferentes para as distinguir. No caso das plantas, a diferença entre assinaturas espectrais depende da quantidade de biomassa acumulada, i.e., da fenologia, da arquitetura da copa e da disposição das folhas (Palchowdhuri et al., 2018).

As resoluções temporal e espacial oferecidas por sensores como o RapidEye (2008), GeoEye-1 (2008), WorldView-2 (2009), Landsat-8 (2013), SPOT-7 (2014), Sentinel-2A (2015) e Sentinel-2B (2017), a aquisição dos dados tem-se tornado mais rápida, atualizada e rentável (Palchowdhuri et al., 2018).

No entanto, existem várias limitações relacionadas com a discriminação entre culturas. Fatores agronómicos como variações na irrigação, fertilização, data de plantação, propriedades do solo, estado sanitário e as práticas culturais dificultam o processo de distinção entre culturas. A monitorização em paisagens heterogéneas é um processo complexo caracterizado pela dificuldade em separar espectralmente as culturas devido à baixa separabilidade entre classes e elevada variabilidade nas classes (Rodriguez-Galiano, Ghimire, et al., 2012).

As limitações mais comuns na classificação de culturas utilizando imagens de satélite estão relacionadas com a semelhança da refletância entre culturas diferentes e com a variabilidade espectral e espacial entre parcelas para a mesma cultura (Tatsumi et al., 2015).

Os índices de vegetação têm sido amplamente usados para analisar e monitorizar variações temporais e espaciais nas características das culturas. A sua utilização permite realçar variações subtis nas respostas espectrais de diferentes superfícies. A aplicação do rácio entre duas bandas espectrais diferentes por exemplo, evidencia variações nos declives das curvas

de refletância que de outra forma estariam mascaradas nomeadamente pelo brilho do pixel em cada uma das bandas. A utilização dos índices de vegetação permite ainda normalizar os valores, i.e., valores relativos (rácios), dos dados provenientes de diferentes sensores cujas condições de aquisição de imagem são diferentes.

O sucesso da identificação das culturas depende da combinação adequada de três elementos: resolução espectral, espacial e temporal dos sensores, e da escolha do método de classificação mais apropriado (Belgiu & Drăgu, 2016).

De entre vários algoritmos de classificação, o algoritmo *Random Forest* tem vindo a ser cada vez mais aplicado na classificação de ocupação do solo (Du et al., 2015; Gislason, Benediktsson, & Sveinsson, 2006; Rodriguez-Galiano, Ghimire, et al., 2012; Tatsumi et al., 2015).

Vários estudos foram feitos sobre a classificação de culturas no âmbito dos controlos de subsídios da PAC. Schmedtmann e Campagnolo (2015) propuseram um sistema de controlo automático para automatizar o processo CAPI num contexto operacional, com o objetivo de reduzir custos e tempo de controlo. Os autores recorreram a uma série temporal de imagens multiespectrais Landsat 7 ETM+ (L7) e ao algoritmo de classificação máquina de suporte vetorial (SVM - *Support Vector Machine*). O método apresentado permite escolher o nível de confiança na classificação automática das parcelas agrícolas. Os resultados demonstraram que, com um nível de confiança de 80%, mais de metade das parcelas foram automaticamente avaliadas com uma precisão total de 84%. Estrada *et al.* (2017) propuseram um método para a identificação automática e objetiva de elementos agrícolas nas parcelas baseada no processamento e na combinação de dados Sentinel-2 e LiDAR. O algoritmo para a identificação de padrões de irrigação das culturas baseada nos dados Sentinel-2 obtiveram a estimativa da precisão global mais alta acima dos 95% num dos locais de teste. Tais resultados indicam que a utilização de séries do Sentinel-2 e LiDAR podem ajudar na deteção de áreas que não são elegíveis para a atribuição de subsídios, auxiliar a validação cruzada e ser usadas na escolha de amostras, permitindo assim uma gestão fácil e economicamente eficiente da PAC. Filizzola *et al.* (2017) desenvolveram um algoritmo multitemporal e multiespectral tirando partido das imagens do Landsat TM e ETM+ para a discriminação entre terras aráveis e terras não aráveis no âmbito da PAC no sul de Itália. Neste estudo foi encontrado uma concordância máxima de 82% relativamente à identificação de terras aráveis. Sob o programa de controlos por DR, Palchowdhuri *et al.* (2018) avaliaram a importância de dados óticos na classificação de culturas multitemporal baseada em três imagens adquiridas pelos satélites WorldView-3 (VHR) e Sentinel-2 (HR). Os autores integraram pela primeira vez os algoritmos de classificação *Random Forest* (RF) e *Decision Tree* (DT) na avaliação da

importância de vários índices de vegetação espectrais na precisão dos resultados obtidos nas classificações. Com uma precisão global de 91% e coeficiente *kappa* de 0.90 estimados usando a combinação dos algoritmos de classificação RF e DT, os autores concluíram que a integração de imagens VHR e HR com de diferentes índices de vegetação pode ser implementado eficientemente para produzir mapas de culturas em larga escala mesmo com um conjunto de dados temporais limitados. Sitokonstantinou *et al.* (2018) implementaram e avaliaram uma classificação supervisionada baseada em objetos (parcelas) para produzir mapas precisos da ocupação do solo numa pequena zona agrícola em Navarra, Espanha. Os autores recorreram aos algoritmos de classificação SVM e RF e a séries temporais compostas por imagens de satélite Sentinel-2 e índices de vegetação. Os resultados obtidos mostraram que ambos os algoritmos de classificação oferecem resultados excelentes com precisões globais de 85.59% e 91.30% para a RF e a SVM respectivamente. No entanto, para o nível mais baixo da nomenclatura de ocupações culturais, i.e., tipo de cultura, a RF teve pior desempenho alcançado apenas uma precisão global de 0.78 (coeficiente *kappa* de Cohen) contra 0.87 resultante do algoritmo SVM. Os autores ainda aplicaram este método utilizando dados L8 destacando o desempenho superior do Sentinel-2 para este tipo de aplicação.



### 3. CONJUNTO DE DADOS UTILIZADOS

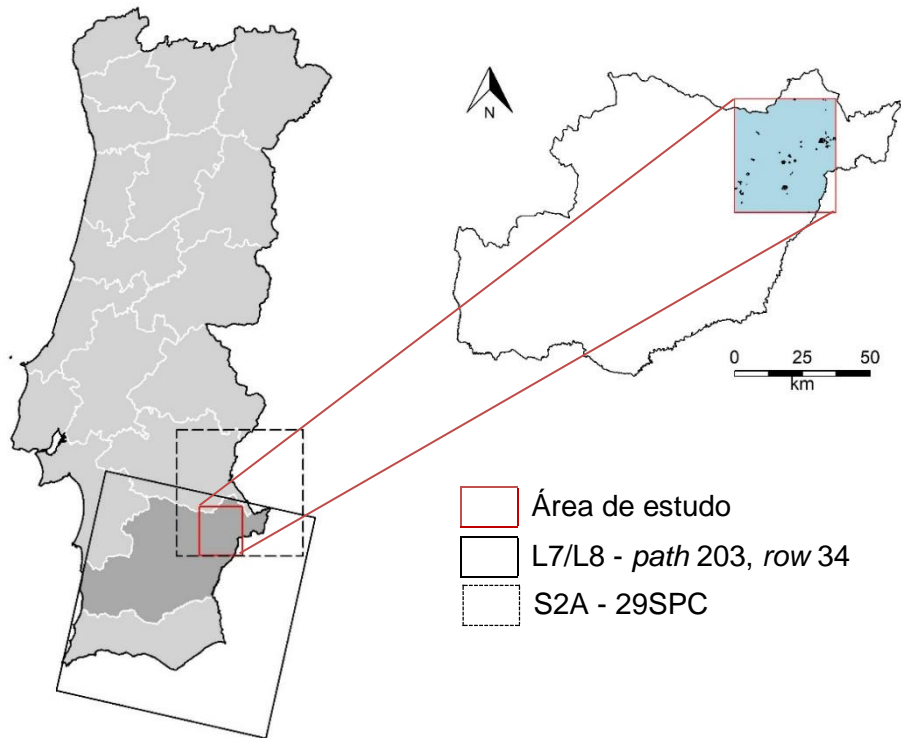
#### 3.1. Caracterização da área de estudo

A área de estudo localiza-se no distrito de Beja, sub-região Baixo Alentejo (NUT III) a sul de Portugal, demarcada entre as longitudes 7° 50' 15" W e 7° 12' 32" W e latitudes 37° 51' 16" N e 38° 15' 26" N (Figura 1). Nesta importante província agrícola em Portugal atravessada pelo rio Guadiana, a área de estudo ocupa uma área com cerca de 2464 km<sup>2</sup>.

Com caráter mediterrânico, o clima desta região é caracterizado segundo a classificação de Köppen como clima temperado com inverno chuvoso suave e verão seco e quente (Csa). Em média, a temperatura anual varia entre a mínima de 9.6°C e a máxima de 24.3°C e a precipitação anual acumulada é de 572 mm (IPMA, 2017; Portal do Clima, 2017).

Em conjugação com um relevo pouco acentuado, com a altitude a variar entre 25 m e 284 m acima do nível médio das águas do mar, característico das típicas planícies alentejanas, são proporcionadas condições para uma agricultura sobretudo de sequeiro em sistema extensivo. Os seus solos heterogéneos são ocupados predominantemente por prados temporários (culturas forrageiras), olival, trigo, pousio, pastagens permanentes, cevada, girassol e aveia. As culturas de regadio, cuja produtividade se via limitada pela disponibilidade hídrica nesta região, têm vindo a ganhar representatividade nos perímetros de Alqueva (EDIA, 2016; GPAa, 2005).

Com o principal objetivo de distinguir culturas de outono/inverno de culturas de primavera/verão, a região de Beja foi eleita para área de estudo não só por albergar um leque diversificado de culturas suficientemente caracterizador de ambos os grupos de estações, mas também por aí se localizarem explorações agrícolas geralmente de grandes dimensões. Segundo o Inquérito à Estrutura das Explorações Agrícolas (IEEA) de 2016 foi registado uma diminuição do número de explorações de 5.5% na região do Alentejo no período 2013-2016. Em relação à Superfície Agrícola Utilizada (SAU) contabilizou um aumento de 100 hectares face a 2013, representando 39.5% da superfície territorial portuguesa. No Alentejo a SAU superior a 50 hectares representa 90.16% da superfície total (INE, 2017).



**Figura 1.** Área de estudo e localização dos mosaicos Landsat 7, Landsat 8 e Sentinel-2A sobre o distrito de Beja a Sul de Portugal.

### 3.2. Dados de deteção remota utilizados

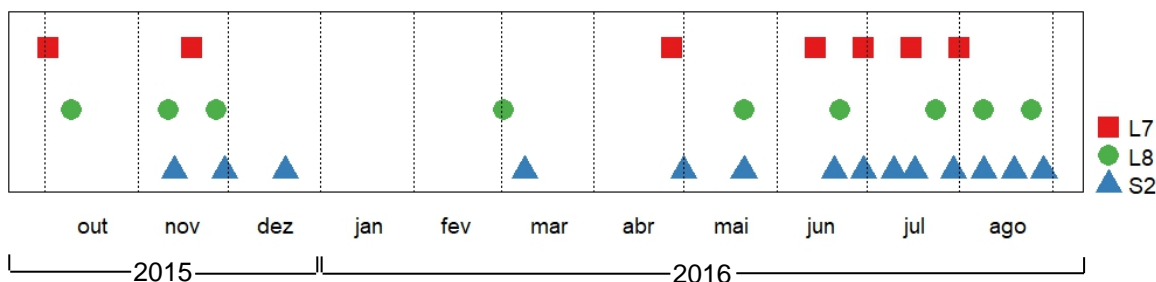
As imagens multiespectrais que integram a série temporal presente neste estudo foram adquiridas entre outubro de 2015 e agosto de 2016 correspondente ao período de desenvolvimento das culturas de outono-inverno e primavera-verão consideradas no mesmo. A seleção das imagens foi baseada na sua disponibilidade no período descrito anteriormente e na baixa percentagem de cobertura por nuvens sobre a área de estudo.

Foram adquiridas sobre a área de estudo sete imagens multiespectrais pelo Landsat 7 Thematic Mapper Plus ETM+ (L7), nove pelo Landsat 8 Operational Land Image OLI (L8), referenciadas no Worldwide Reference System (WRS-2) no *path* 203 e *row* 34, e catorze pelo Sentinel 2A MSI (S2A), referenciadas no sistema de grelha de mosaicos no mosaico 29SPC (Figura 2).

Todas as imagens foram adquiridas através da plataforma *open source* EarthExplorer da United States Geological Survey (USGS) à exceção de algumas imagens do S2A, para as quais ainda se recorreu à plataforma *open source* Copernicus Open Access Hub da European Spacial Agency (ESA) podendo também ser obtidos através da infraestrutura IPSentinel desenvolvida pela Direção-Geral do Território em parceria com o IPMA. Todos os conjuntos de dados geográficos utilizados estão associados ao sistema de coordenadas de referência



(SCR) *World Geodetic System (WGS84)* com a projeção *Universal Transverse Mercator* zona 29N (UTM zone 29N). O SCR é descrito pelo código EPSG:32629.



**Figura 2.** Distribuição temporal das imagens multiespectrais utilizadas neste estudo e adquiridas entre outubro de 2015 e agosto de 2016 pelos respectivos sensores Landsat 7 (L7), Landsat 8 (L8) e Sentinel 2A (S2). Apenas as datas com baixa proporção de nuvens foi considerada sobre a área de estudo.

A combinação de dados satélite de média resolução provenientes dos sensores heliosíncronos do L7, L8 e S2A promove o melhoramento da monitorização e o mapeamento da superfície da Terra ao oferecer resoluções espaciais e temporais respetivamente de 10 a 30 m e de 3 dias (Roy et al., 2014). Vários estudos têm demonstrado que a aquisição contínua de imagens melhora a análise e a compreensão da fenologia das culturas e dos respetivos processos biofísicos, nomeadamente a acumulação de biomassa (Gao et al., 2015; Testa, Mondino, & Pedroli, 2014).

No Quadro 2 apresentam-se de modo grosseiro a correspondência entre as bandas de cada sensor utilizadas segundo a resolução espectral.

Para cada sensor, foram utilizados valores de refletância à superfície (Quadro 3) com o objetivo de minimizar possíveis diferenças provocadas pelas condições de observação dependente de cada sensor.

**Quadro 2.** Correspondência entre as bandas de cada sensor segundo a resolução espectral (nm) e espacial (m). Adaptado de Flood (2017).

Satélite	Landsat 7		Landsat 8		Sentinel 2A	
Sensor	ETM+		OLI		MSI	
Resolução	Espectral (nm)	Espacial (m)	Espectral (nm)	Espacial (m)	Espectral (nm)	Espacial (m)
Blue	1 (450–520)		2 (450–510)		2 (458–522)	10
Green	2 (520–600)		3 (530–590)		3 (543–578)	10
Red	3 (630–690)		4 (640–670)		4 (650–680)	10
Red-edge 1					5 (698–713)	20
Red-edge 2					6 (733–748)	20
Red-edge 3		30		30	7 (773–793)	20
NIR	4 (770–900)		5 (850–880)		8 (785–900)	10
NIR					8A (855–875)	20
SWIR 1	5 (1550–1750)		6 (1570–1650)		11 (1565–1655)	20
SWIR 2	7 (2090–2350)		7 (2110–2290)		12 (2100–2280)	20

Os valores de refletância à superfície disponibilizados nos produtos L7 resultaram da correção atmosférica aplicada às imagens Level-1 pelo *software* Landsat Ecosystem Disturbance Adaptive Processing System (LEDAPS). Esta ferramenta recorre a rotinas de correção atmosférica derivadas do Moderate Resolution Imaging Spectroradiometer (MODIS) e a modelos de transferência radiativa em Second Simulation of a Satellite Signal in the Solar Spectrum (6S) (USGS, 2018). Já nos produtos Landsat 8 OLI, os valores de reflectância à superfície resultaram da aplicação do Landsat Surface Reflectance Code (LaSRC). Este algoritmo, distintivamente do anterior, utiliza dados climáticos auxiliares do MODIS e um modelo de transferência radiativa específico (USGS, 2017).

Foram usadas seis bandas pertencentes à região espectral visível, infravermelho próximo e infravermelho médio com resolução espacial de 30 m. Adicionalmente, foi utilizada em separado a banda da Quality Assessment (QA) disponibilizada nos ficheiros. Esta contém informação acerca das condições atmosféricas, da superfície e do sensor que podem estar a afetar a utilidade geral de um dado pixel, permitindo aos utilizadores aplicar filtros ao nível do pixel.

Desde 31 de maio de 2003 o Scan Line Corrector (SLC) do L7 não se encontra operacional resultando na perda de cerca de 22% de informação da imagem. Porém, a qualidade da informação para os restantes pixels permanece inalterada atribuindo à imagem um certo grau de utilidade (Chander, Markham, & Helder, 2009). Apesar de ter tido algum impacto sobre a área de estudo foram considerados para a análise apenas os pixels cuja qualidade não foi afetada atribuindo um valor nulo aos restantes pixels.

A missão Sentinel-2 da ESA é constituída atualmente por dois satélites que operam em conjunto: o Sentinel-2A (lançado a junho de 2015) e o Sentinel-2B (lançado a março de 2017). Dispõe de um sensor MultiSpectral Instrument (MSI) com 13 bandas espectrais com boa resolução espacial (10, 20 ou 60 m consoante a banda) e média resolução temporal de 10 dias. Esta resolução temporal pode aumentar para 5 dias quando os dois satélites se encontram operacionais podendo mesmo atingir 2-3 dias nas latitudes médias. As imagens foram descarregadas como Level-1C traduzindo-se em imagens ortorretificadas de 100x100 km<sup>2</sup> na projeção UTM/WGS84 cujas medições radiométricas são fornecidas em valores de reflectâncias no topo da atmosfera (TOA).

A forma generalizada de apresentação dos nomes dos ficheiros *input* nos quais se baseia a base de dados utilizados para a classificação é descrita no Quadro 3.

**Quadro 3.** Nomes dos ficheiros *input*. Com *yyyymmdd* a representar a data de aquisição, *YYYYMMDD* a data do ficheiro e *RES* a resolução.

Dados input	Sensor	Bandas
LE07_L1TP_203034_yyyymmdd_YYYYMMDD_01_T1_sr_banda	ETM+	1,2,3,4,5,7, pixel_qa
LC08_L1TP_203034_yyyymmdd_YYYYMMDD_01_T1_sr_banda	OLI	2,3,4,5,6,7, pixel_qa
S2A_USER_MSI_L2A_TL_SGS_yyyymmddT182844_A005897_T29SPC_banda_RES	MSI	2,3,4,5,6,7, 8,8A,11,12
S2A_OPER_MSK_CLOUDS_MTI_yyyymmddT193334_A003709_T29SPC_B00_MSIL1C	MSI	-

### 3.3. Parcelas agrícolas e classes de ocupação do solo

Os dados das parcelas agrícolas controladas em 2016 para o distrito de Beja foram extraídos do SIP e fornecidos IFAP em formato vetorial (ficheiro *shapefile* do tipo polígono) projetado no sistema de coordenadas ETRS89/Portugal TM06 (EPSG:3763). Estes incluem os limites e respetiva ocupação do solo validada de cada parcela registada entre outro tipo de informação menos relevante. Os atributos associados a este conjunto de dados vetoriais utilizados neste estudo estão listados no Quadro 4.

**Quadro 4.** Atributos do conjunto de dados vetoriais fornecido pelo IFAP.

Atributo	Descrição	Tipo de variável
PAR_ID	Identificador da parcela	Numérico ( <i>double</i> )
OSA_ID	Identificador da subparcela	Numérico ( <i>double</i> )
OSA_CAMPAN	Ano de controlo dos subsídios	Texto
OSA_SUB_PA	Número da subparcela	Texto
COD_CON	Código da ocupação cultural controlada	Texto
COD_DESC	Nome da ocupação cultural controlada	Texto

O controlo de subsídios efetuado em 2016 significa que as culturas controladas nesse ano incluem as culturas de outono/inverno 2015 e 2016 e as culturas de primavera/verão de 2016, razão pela qual o período de aquisição das imagens de satélite se inicia em outubro de 2015.

Sendo o principal objetivo deste estudo a distinção entre culturas de outono/inverno e primavera/verão (no âmbito do *greening* na categoria de diversificação das culturas), o critério de escolha das culturas assentou sobre as culturas que se podem incluir nessas categorias e com maior representatividade na área de estudo. As culturas inicialmente consideradas foram a aveia para grão (designada mais à frente apenas por aveia), cevada, girassol, melancia, meloa, melão, milho, outras hortícolas, sorgo, tomate, trigo mole (designada mais à frente apenas por trigo) e tritcale.

De um total de 241 parcelas agrícolas contidas na área de estudo foram extraídas apenas 210 parcelas obedecendo aos seguintes critérios: a parcela tinha de possuir uma área superior a 1000 m<sup>2</sup> de forma a garantir que nela contivesse pelo menos um pixel, e a área ocupada pelas culturas tinha de ser suficientemente grande para que a amostra (pixels) fosse

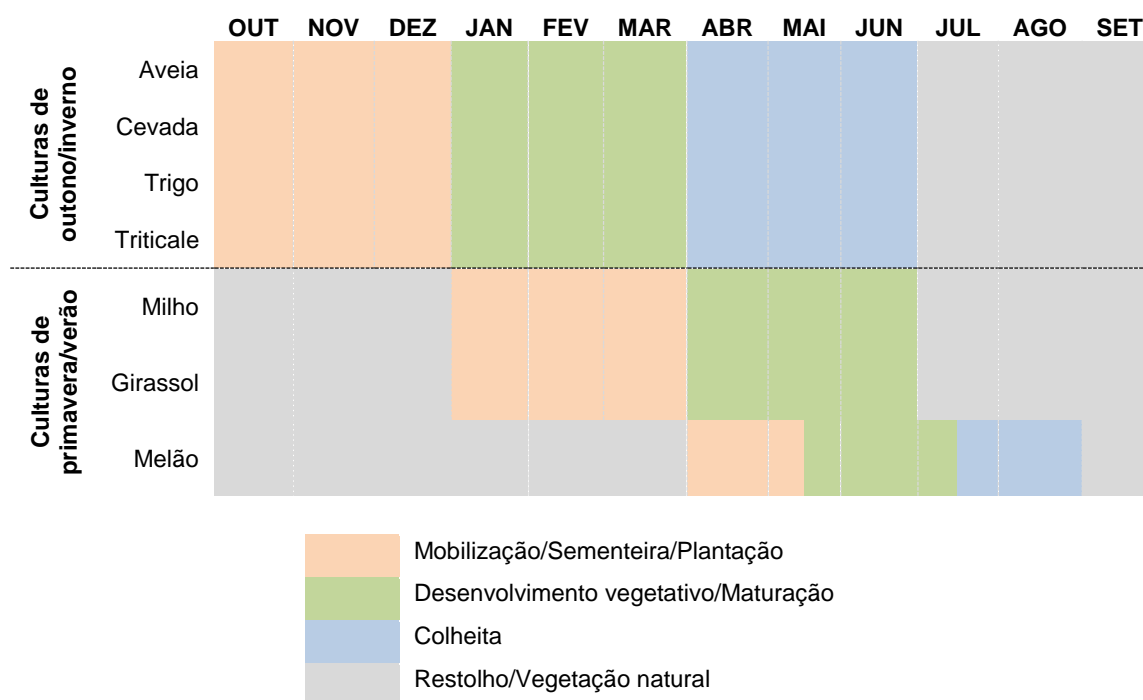
representativa dessa mesma cultura. Das doze culturas plantadas na área de estudo apenas as culturas apresentadas no Quadro 5 e Figura 3 foram usadas neste trabalho representando cerca de 1203 hectares da área de estudo.

**Quadro 5.** Dimensões das parcelas por classe de ocupação do solo consideradas neste estudo. Grande variabilidade no tamanho das parcelas.

Cultura	Nº Parcelas	Área total (ha)	Área média (ha)	Desvio padrão
Aveia	33	111.7	3.4	6.7
Cevada	43	328.1	7.6	10.1
Girassol	25	175.7	7.0	11.9
Melão	8	59.2	7.4	5.2
Milho	7	214.7	30.7	40.0
Trigo mole	63	224.6	3.6	6.3
Triticale	31	89.3	2.9	3.5

O facto de a área média por parcela sobre a área de estudo ser relativamente pequena com cerca de 5.7 hectares permite testar o desempenho dos diversos métodos de classificação e comparar entre sensores de diferentes resoluções espaciais a utilidade dos seus produtos.

A informação fenológica acerca das culturas consideradas neste estudo encontra-se sumariada na Figura 3 sob a forma de calendário.



**Figura 3.** Calendário cultural 2015-2016. Inclui as fases de preparação do solo e os vários estados de desenvolvimento das culturas.

## 4. MÉTODOS

A metodologia adotada neste capítulo está esquematizada na Figura 4.

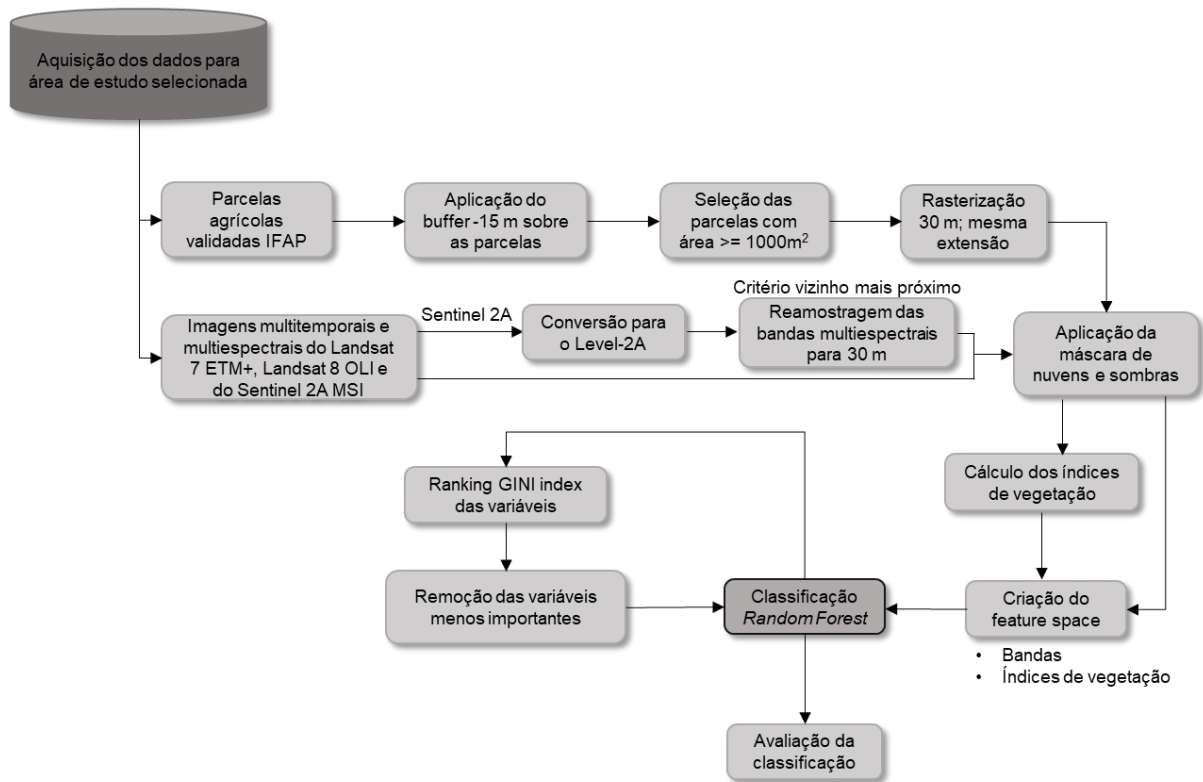


Figura 4. Esquematização da metodologia utilizada.

### 4.1. Pré-processamento

As imagens S2A descarregadas no Level-1C de processamento foram posteriormente processadas para o Level-2A para obter o valor das refletâncias à superfície (BoA – Bottom of atmosphere) através do algoritmo *Sen2cor* versão 2.3.1. (fevereiro de 2017) disponibilizado na Science Toolbox Exploitation Platform (STEP) pela ESA.

As bandas do S2A adquiridas com uma resolução espacial de 20 m foram reamostradas para 30 m usando a mesma extensão do ficheiro vetorial das parcelas através da ferramenta Gdalwarp usando o método vizinho mais próximo (NN – Nearest Neighbor).

Os pixels afetados pelas nuvens e/ou sombras foram eliminados usando as bandas QA (quality assessment).

Depois de ser gerado um buffer com 15 m de distância para dentro dos limites das parcelas, foram selecionadas aquelas com área superior a 1000 m<sup>2</sup>. Este procedimento permitiu assegurar a seleção de pixels que contivessem apenas informação dentro da parcela, evitando a utilização de pixels com informação ruidosa que pudessem vir a afetar o desempenho da classificação automática.

## 4.2. Nomenclatura

A nomenclatura proposta obedece a uma hierarquia que representa a ocupação do solo a dois níveis. O primeiro nível é referente à estação em que uma determinada cultura é cultivada que se divide em 2 classes e o segundo nível refere-se ao tipo de cultura que se divide em 7 classes.

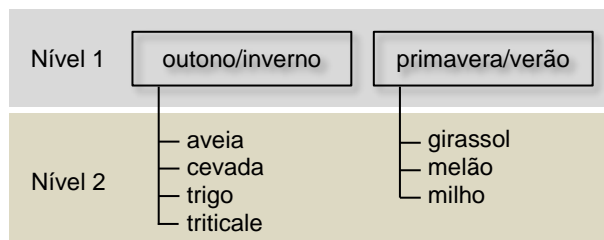


Figura 5. Nomenclatura de classificação.

## 4.3. Índices de vegetação

O recurso aos índices de vegetação neste tipo de estudos permite evidenciar o sinal emitido pela vegetação enquanto os efeitos do solo e da irradiação solar são atenuados (Hatfield & Prueger, 2010). Foram selecionados devido à sua importância na medição do vigor da vegetação, como por exemplo na percentagem de cobertura, biomassa, LAI, etc. A escolha dos mesmos recaiu sobre o tipo de culturas presentes na área de estudo ao longo do período de estudo. Exemplo disso é o SAVI o qual ajusta o impacto do solo na refletância permitindo facilitar a distinção entre culturas de primavera/verão e culturas de outono/inverno.

O normalized difference vegetation index (NDVI) (Rouse et al., 1974) é um dos IV mais utilizados na área da DR. É calculado a partir da normalização do rácio entre as bandas do vermelho e do infravermelho próximo demonstrado pela seguinte expressão:

$$\text{NDVI} = \frac{\rho_{\text{NIR}} - \rho_{\text{red}}}{\rho_{\text{NIR}} + \rho_{\text{red}}} \quad (4.1)$$

onde o  $\rho_{\text{NIR}}$  e  $\rho_{\text{red}}$  são as refletâncias na banda do infravermelho próximo e na banda do vermelho, respetivamente. O green normalized difference vegetation index (GNDVI) (Gitelson, Kaufman, & Merzlyak, 1996) é uma adaptação do NDVI que mede a refletância na banda do verde no lugar de medir na banda do vermelho. Este índice é calculado a partir da seguinte fórmula:

$$\text{GNDVI} = \frac{\rho_{\text{NIR}} - \rho_{\text{green}}}{\rho_{\text{NIR}} + \rho_{\text{green}}} \quad (4.2)$$

onde o  $\rho_{\text{NIR}}$  e  $\rho_{\text{green}}$  são as refletâncias na banda do infravermelho próximo e na banda do verde, respetivamente.

O objetivo do SAVI (Huete, 1988) é minimizar os efeitos do solo na refletância da vegetação onde a percentagem de cobertura vegetal é baixa. É no fundo uma modificação do NDVI no qual se introduziu um parâmetro de ajustamento  $L$  no seu cálculo:

$$SAVI = \frac{\rho_{NIR} - \rho_{red}}{\rho_{NIR} + \rho_{red} + L} \times (1 + L) \quad (4.3)$$

onde o  $\rho_{NIR}$  e  $\rho_{green}$  são as refletâncias na banda do infravermelho próximo e na banda do verde, respetivamente; o parâmetro  $L$  representa o ajustamento do impacto do solo e o seu valor é de 0.5 sob a maioria das condições ambientais.

O MSAVI resulta da modificação do SAVI com o objetivo de reduzir a influência do solo no SAVI e pode ser expresso por:

$$MSAVI = \frac{2 \times NIR + 1 - \sqrt{(2 \times NIR + 1)^2 - 8 \times (NIR - Red)}}{2} \quad (4.4)$$

O resultado da interação existente entre o solo e atmosfera é que a redução de um deles poderá aumentar o outro. Assim, Huete *et al.* (1997) desenvolveram um parâmetro que corrige em simultâneo os efeitos do solo e a da atmosfera. Esse parâmetro é o Enhanced Vegetation Index (EVI) e apresenta-se com a seguinte fórmula:

$$EVI = 2.5 \times \frac{\rho_{NIR} - \rho_{red}}{1 + \rho_{NIR} + 6 \times \rho_{red} - 7.5 \times \rho_{blue}} \quad (4.5)$$

onde  $L$  representa o parâmetro de ajustamento do solo sendo igual a 1; os valores 6 e 7.5 correspondem a parâmetros.

#### 4.4. Criação da base de dados utilizada no treinamento do modelo *random forest*

As variáveis que compuseram a base de dados estão descritas no Quadro 6 e dizem respeito a dois tipos de variáveis: refletâncias e índices de vegetação perfazendo um total de 386 variáveis. Os dados foram organizados numa tabela/matriz de dados numéricos constituído por  $n$  linhas e  $p$  colunas correspondente aos pixels e às variáveis respetivamente que é ilustrada na Figura 6.

**Quadro 6.** Descrição das variáveis extraídas para a construção da base de dados utilizada para a classificação de acordo com o seu tipo e origem.

Tipo/Origem	Landsat 7 ETM+	Landsat 8 OLI	Sentinel 2A MSI
Refletâncias	Bandas: B01, B02, B03, B04 (VIS), B05 (NIR), B07 (SWIR) × 7 imagens adquiridas (42 variáveis)	Bandas: B02, B03, B04 (VIS), B05 (NIR), B06, B07 (SWIR) × 9 imagens adquiridas (54 variáveis)	Bandas: B02, B03, B04 (VIS), B08, B8A (NIR), B05, B06, B07 (Red-Edge) e B11, B12 (SWIR) × 14 imagens adquiridas (140 variáveis)
Índices de vegetação	NDVI, GNDVI, EVI, SAVI, MSAVI × 7 imagens adquiridas (35 variáveis)	NDVI, GNDVI, EVI, SAVI, MSAVI × 9 imagens adquiridas (45 variáveis)	NDVI, GNDVI, EVI, SAVI, MSAVI × 14 imagens adquiridas (70 variáveis)

	culturã	estacao	x	y	osa.id	par.id	osa.sub.pã	cod.con	L7_203034_20151001_NDVI
16185	cevada	outono_inverno	622835	4198012	3349141	1438983	1	2	0.13088261
22589	melao	primavera_verao	619955	4199812	3379708	1438948	1	10	0.14405594
22469	melao	primavera_verao	620015	4199872	3379708	1438948	1	10	0.15566038
22588	melao	primavera_verao	619925	4199812	3379708	1438948	1	10	0.16254603
22577	melao	primavera_verao	620195	4199842	3379708	1438948	1	10	0.16326186
22598	melao	primavera_verao	620225	4199812	3379708	1438948	1	10	0.16892937
22531	melao	primavera_verao	620465	4199992	3379708	1438948	1	10	0.16803670
22604	melao	primavera_verao	619925	4199782	3379708	1438948	1	10	0.16293383
1107	trigo	outono_inverno	653795	4215652	3316217	1261493	1	1	0.15888651
9017	aveia	outono_inverno	651065	4217452	3339053	1434149	1	3	0.18351384
16175	cevada	outono_inverno	622835	4198102	3349141	1438983	1	2	0.11018182

**Figura 6.** Ilustração da organização dos dados. A tabela tem 4594 linhas e cada linha representa um pixel, com indicação da parcela a que pertence. As primeiras colunas indicam a localização do pixel, a parcela, e as variáveis que dependem da parcela (em particular, a cultura e a estação). As restantes colunas representam as variáveis espectrais, cada uma determinada pelo sensor, a data de aquisição da imagem e a banda ou índice considerado, havendo no total 386 colunas para as variáveis espectrais. Note-se que a tabela é construída após a reamostragem de todas as imagens para uma resolução espacial comum de 30 m.

#### 4.5. Análise da homogeneidade espectral nas parcelas

Cada parcela é descrita por um conjunto de pixels com valores de refletância sendo que cada pixel corresponde a uma observação realizada por um dado sensor em diversas bandas espectrais e datas. Por este motivo realizou-se um teste à homogeneidade espectral nas parcelas (Schmedtmann & Campagnolo, 2015) para despistar possíveis dificuldades que poderiam resultar da combinação de valores provenientes dos diferentes tipos de sensores utilizados neste estudo. Para cada combinação sensor, banda/índices de vegetação e data comparámos a variabilidade entre parcelas com a variabilidade nas parcelas (equação 4.6) e analisámos se a primeira era muito maior do que a segunda, o que seria expectável neste tipo de classificação em que as parcelas são homogéneas.



À totalidade dos pixels contidos nas parcelas foi aplicado o modelo ANOVA a dois fatores hierarquizados, onde o fator dominante e o fator subordinado são as culturas e as parcelas respetivamente, permitindo eliminar o efeito da cultura. Adicionalmente foi calculado o RMSE o qual mede o desvio padrão nas parcelas cuja fórmula é apresentada na equação 4.7.

$$F = \frac{\text{variância entre parcelas}}{\text{variância nas parcelas}} \quad (4.6)$$

$$\text{RMSE} = \sqrt{\text{variância nas parcelas}} \quad (4.7)$$

Um elevado valor de F e um baixo RMSE (equação 4.7) são indicadores de que as parcelas são homogéneas e de que a variabilidade entre os pixels da mesma cultura se deve sobretudo à diferença entre parcelas.

#### 4.6. Classificação segundo o método Random Forest (RF)

As classificações foram orientadas pelo algoritmo de classificação RF (Breiman, 2001). É uma técnica de aprendizagem cujo conceito passa por criar e combinar várias árvores de decisão para obter uma predição mais precisa e estável. A RF é um algoritmo de classificação assistida do *Machine Learning* amplamente utilizado com sucesso em diversas aplicações da DR nomeadamente em estudos relacionados com a classificação de culturas e de ocupação do solo (Gislason, Benediktsson, & Sveinsson, 2006; Rodriguez-Galiano, Chica-Olmo, et al., 2012; Rodriguez-Galiano, Ghimire, et al., 2012; Tatsumi et al., 2015). Esta técnica de classificação tem vindo a ganhar popularidade em aplicações que requerem o uso de imagens de elevada resolução espacial uma vez que conseguem lidar eficazmente com um grande volume de dados (Lebourgeois et al., 2017). Baseado num conjunto de árvores de decisão, este classificador usa o método *bootstrap aggregating*, i.e., *bagging*, para criar diferentes subconjuntos de objetos de treino e produzir diversas árvores em que cada fornece o resultado da classificação para os subconjuntos de objetos de teste. A classe atribuída ao objeto sob classificação é obtida como o resultado de um grande número de árvores de decisão individuais que obteve o número máximo de votos.

No atual estudo o modelo RF foi obtido através da função *randomForest* (Liaw & Wiener, 2002). Ao utilizador é dada a possibilidade de definir dois parâmetros: (1) o número de árvores de decisão a gerar (*n tree*), e (2) o número de variáveis a selecionar antes de cada divisão (*m try*) durante a expansão das árvores. Vários estudos demonstraram que é possível obter resultados satisfatórios com os valores padrão atribuídos aos parâmetros do modelo (Immitzer, Atzberger, & Koukal, 2012; Liaw & Wiener, 2002; Zhang & Roy, 2017). Desta forma, os valores padrão (Breiman, 2001; Liaw & Wiener, 2002) foram adotados e definidos 500 para

o número de árvores (*n<sub>tree</sub>*) e para o *mtry* a raiz quadrada do número total de variáveis utilizadas. Este algoritmo produz uma estimativa não enviesada (*unbiased*) do erro utilizando objetos (amostras) *out-of-bag*, i.e., objetos não estão incluídos no subconjunto de treino. Uma desvantagem da RF comparativamente ao método *bagging trees* é o facto de se tratar de uma abordagem “às cegas” uma vez que cada árvore não pode ser estimada separadamente (Prasad, Iverson, & Liaw, 2016). Em contrapartida, a RF disponibiliza várias métricas úteis para a interpretação do modelo. A importância das variáveis é avaliada com base no grau de decréscimo na precisão ou no índice GINI quando os dados são permutados aleatoriamente para aquela variável. Estes indicadores denominam-se por *Mean Decrease Accuracy* (MDA) e *Mean Decrease Gini* (MDG) respetivamente, e são frequentemente utilizados na seleção ou *ranking* de variáveis (Genuer et al., 2012; Gislason, Benediktsson, & Sveinsson, 2006; Immitzer, Atzberger, & Koukal, 2012).

#### **4.7. Otimização: seleção de variáveis**

As grandes dimensões do *feature space* requer na maioria das vezes a seleção de variáveis por ser uma forma de encurtar o tempo de computação e de reduzir a complexidade da classificação e/ou ainda melhorar as precisões da classificação através da redução do ruído causado por possíveis variáveis correlacionadas e redundantes.

O objetivo da seleção de variáveis é otimizar o resultado da classificação RF através da utilização do menor número de variáveis possível que contribua para a maximização da precisão da classificação. Esta fase de otimização foi baseada na análise da importância das variáveis, segundo a MDG, no resultado da precisão global da classificação. Para o processo de otimização foi usado o método de eliminação recursiva de variáveis com recurso à validação cruzada *10-fold*. Isto significa que em cada ciclo de aplicação é calculado o valor de MDG e a variável que apresentar o menor valor é eliminada do *dataset* e assim sucessivamente até sobrar apenas uma variável. Assim, a importância de cada variável pode ser dada pela ordem em que foi eliminada.

#### **4.8. Cenários da classificação**

Foram selecionados dois tipos de abordagens diferentes para a classificação da ocupação do solo: uma baseada em pixels e a outra baseada em objetos (parcelas). A classificação ao nível do pixel é um processo que atribui a cada pixel uma classe, tratando cada pixel como uma única unidade. Na classificação assistida, os dados de treino são selecionados a partir de amostras de pixels da imagem e usados como referência na classificação da ocupação do solo. O conjunto de pixels representativo de cada classe é a chave para a implementação da classificação assistida.

Na abordagem ao nível da parcela foi aplicado o mesmo princípio, com a diferença de ser ter procedido posteriormente à agregação de todos os pixels classificados por parcela. A classe atribuída a cada parcela foi a classe com a maior proporção de pixels nessa parcela.

#### 4.9. Validação e avaliação dos resultados da classificação

A precisão do método de classificação pode ser avaliada através de uma análise estatística sumária que compara a informação de referência e a informação resultante da classificação. A forma padronizada mais frequentemente utilizada para descrever a precisão é através da matriz de erro. É uma matriz quadrada de números dispostos em linhas e colunas, onde normalmente as colunas representam os dados de referência enquanto as linhas indicam os dados resultantes do processo de classificação. Estes números quantificam os objetos/pixels atribuídos a uma dada classe de ocupação do solo em relação à classe de referência, i.e., à classe efetivamente conhecida.

Esta matriz permite aceder eficazmente a precisão da classificação individual de cada classe descrita através dos erros de omissão (exclusão) e comissão (inclusão). Os erros de omissão ocorrem quando os pixels são incorretamente excluídos da classe a que realmente pertencem, enquanto os erros de comissão ocorrem quando os pixels são incorretamente incluídos numa classe a que não pertencem de verdade.

A partir da matriz de erro podem ser estimadas: a precisão global (PG) e ao nível da classe, a precisão do utilizador (PU) e a precisão do produtor (PP), através das equações 4.1 a 4.3. A PG é o quociente entre o número total de pixels corretamente classificados, i.e., a soma da diagonal da matriz de erro e o número total de pixels. A PU determina-se a partir do quociente entre o número de pixels corretamente atribuídos a determinada classe e o número total de pixels atribuídos, correta ou incorretamente, a essa classe. A PP é o quociente entre o número de pixels corretamente atribuídos a uma dada classe e o número total de pixels que efetivamente pertencem a essa classe (Congalton & Green, 2009).

$$PG = \frac{\sum_{i=1}^k n_{ii}}{n} \quad (4.1)$$

$$PU = \frac{n_{ii}}{n_i} \quad (4.2)$$

$$PP = \frac{n_{ii}}{n_{.j}} \quad (4.3)$$

onde  $k$  é o número de classes de ocupação do solo;  $n$  é o número total de elementos da matriz,  $n_{ii}$  é o número de elementos classificados corretamente numa dada classe, i.e., o valor na diagonal principal da matriz,  $n_{i.}$  é o número de elementos classificados na classe da linha

$i$  ( $i=1,2,\dots,k$ ) da matriz,  $n_{ij}$  é o número de elementos que pertencem à classe da coluna  $j$  ( $j=1,2,\dots,k$ ) da matriz.

Para lidar com o número limitado pixels por parcela foi aplicada a técnica de validação cruzada *k-fold* (Kohavi, 1995). Ao usar a validação cruzada é possível criar múltiplos conjuntos de dados de treino e de teste que permitem obter uma estimativa da precisão do classificador menos enviesada do que a obtida para a totalidade da amostra. Esta técnica envolveu a divisão aleatória do conjunto de objetos de referência em dez subconjuntos ( $k=10$ ) mutuamente exclusivos e do mesmo tamanho. Cada subconjunto (*fold*) foi usado para testar o modelo gerado a partir dos restantes nove *folds* pelo algoritmo de classificação. No final, as  $k$  matrizes de confusão resultantes de cada *fold* foram agregadas (somadas) numa única matriz a partir da qual se estimaram a PG, PU e PP. Por forma a obter resultados comparáveis a validação foi feita para a totalidade dos pixels para ambos os cenários da classificação: pixel e parcela.

## 5. ANÁLISE E DISCUSSÃO DE RESULTADOS

### 5.1. Análise da homogeneidade espectral nas parcelas

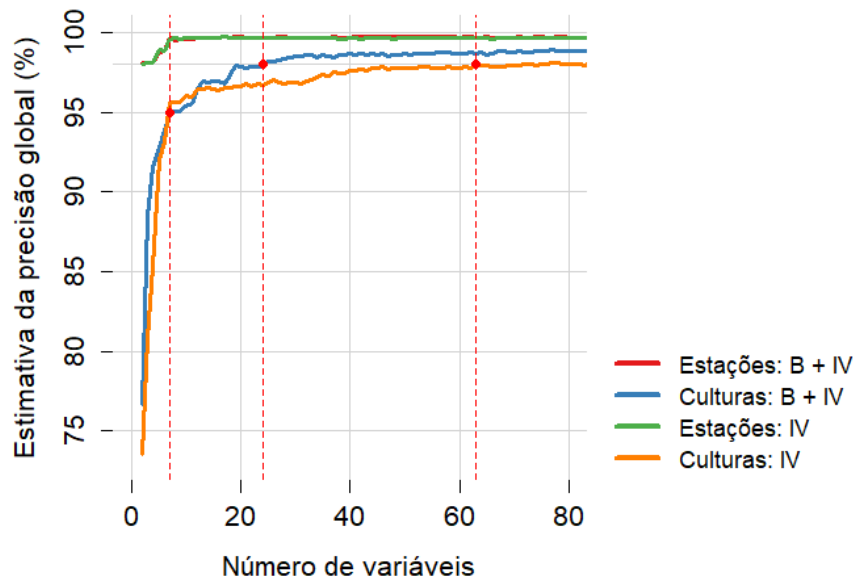
Os valores de F foram sempre altos, entre 53.7 e 318, para as combinações sensor, banda/índice de vegetação e data. O que significa que a variabilidade entre parcelas é pelo menos 53 vezes maior do que a variabilidade nas parcelas. Os valores do RMSE demonstraram que o desvio padrão global dos valores de refletância nas parcelas variam entre 0 e 0.045 (à exceção de 0.071) unidades de refletância para as mesmas combinações utilizadas na classificação. Tais resultados vêm comprovar a homogeneidade espectral nas parcelas e mitigar a problemática da compatibilidade da informação proveniente de diferentes sensores.

### 5.2. Seleção de variáveis

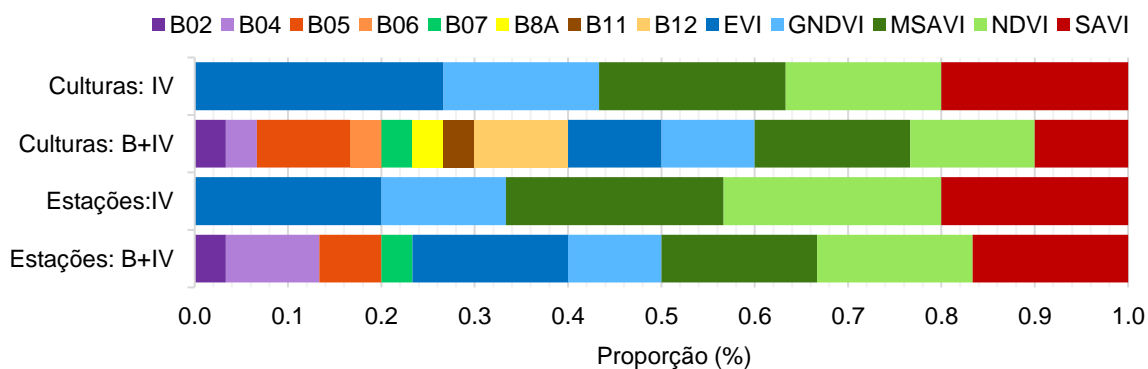
Muitas das variáveis presentes no *dataset* estão correlacionadas, são redundantes ou simplesmente não são informativas pelo que a sua utilização na classificação não melhora a precisão global. Isso mesmo é ilustrado pela Figura 7 **Figura 7**, que mostra a existência de alterações significativas na precisão global até a curva atingir a 8ª e 25ª variáveis relativamente ao nível da nomenclatura 1 e 2 respetivamente. Tal sucedido deve-se ao facto da RF ser relativamente insensível à presença de variáveis redundantes. No geral, os resultados demonstraram que o número ótimo de variáveis necessárias para alcançar a melhor precisão global aumenta com nível de nomenclatura adotado e com o grau de complexidade e de detalhe.

O processo de otimização permite uma redução de 97.8% e 93.2% do número de variáveis sem que isso tenha repercussões significativas na qualidade da classificação, o que significa ganhos relativamente ao tempo de computação quando o número de variáveis é demasiado grande.

Uma análise às proporções (%) de diferentes tipos de variáveis, refletância ou IV, entre as 40 variáveis mais importantes para cada nível de nomenclatura e cenário está apresentada na Figura 8. No que diz respeito ao tipo de dados, a Figura 8 mostra que os IV foram predominantes na discriminação tanto entre culturas como estações para o cenário B+IV, embora que com mais ênfase nesta última. Na distinção entre culturas notou-se que o top 40 das variáveis mais importantes é mais variado do que na distinção entre estações dada a complexidade da classificação.



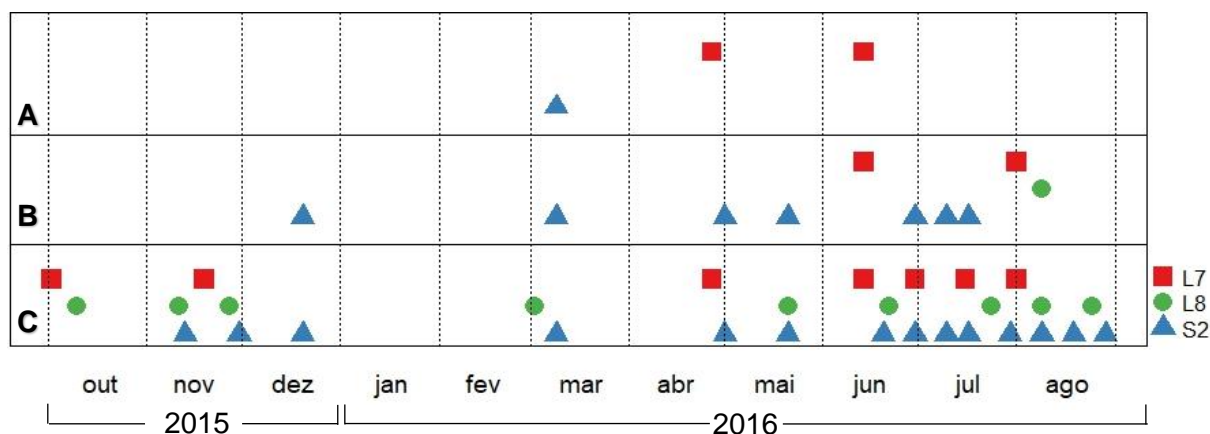
**Figura 7.** Evolução da estimativa da precisão global obtida pela validação cruzada (10-*fold*) de acordo com o número de variáveis utilizado em cada ciclo de classificação. As variáveis foram previamente ordenadas por ordem decrescente de importância baseada nos valores MDG e foi removida a variável menos importante pelo que esta figura deve ser lida da direita para a esquerda. Qualquer que seja o *dataset* utilizado são necessárias no mínimo 7 variáveis para obter uma precisão global superior a 95%. Para obter uma precisão igual ou superior a 98% são necessárias pelo menos 24 variáveis e 63 para quando o *dataset* é constituído pelas refletâncias e índices de vegetação ou somente índices de vegetação respetivamente.



**Figura 8.** Proporção das 40 variáveis mais importantes no ranking da MDG de acordo com o tipo de variáveis, refletâncias ou índices de vegetação, e o cenário.

A distribuição temporal das variáveis selecionadas anteriormente para a otimização da classificação é apresentada na Figura 9 em relação ao conjunto de variáveis inicial e com discriminação do tipo de sensor. As variáveis derivadas da série temporal do S2A predominaram na distinção entre culturas. Ao nível do tipo de cultura, as variáveis mais importantes estão distribuídas predominantemente entre maio e agosto, à exceção de dezembro e março (Quadro 7). Esse período coincide com a fase de senescência/colheita das

culturas de outono/inverno em contraste com as culturas de primavera/verão que ainda estão em fase de desenvolvimento ou de florescência. A utilização de uma série temporal que abrange todas as fases fenológicas, principalmente fases mais avançadas, das culturas de inverno e de verão permite melhorar a capacidade discriminatória dos classificadores devido ao facto de descrever períodos onde as diferenças no desenvolvimento vegetativo entre duas classes são máximas.



**Figura 9.** Distribuição temporal das variáveis selecionadas para a classificação relativamente (A) ao tipo de estação, utilizando apenas os IV; (B) ao tipo de cultura, utilizando a combinação bandas e IV, em relação a (C) que representa a totalidade de informação disponível. As 25 variáveis distribuídas temporalmente em (B) permitiram uma precisão superior a 95% enquanto que as 8 variáveis em (A) permitiram uma precisão global superior a 98%.

**Quadro 7.** Apresentação das regiões espectrais ou índices de vegetação a que correspondem às 25 e 8 variáveis selecionadas correspondem à primeira e segunda coluna respetivamente.

Bandas/IV	2015		2016				
	Dez	Mar	Abr	Mai	Jun	Jul	Ago
B2						▲	
B4					■		
B5					▲		
B6			▲				
B7	▲						
B8A				▲			
B11						▲	
B12					▲	▲	
NDVI		▲	■		■		●
GNDVI		▲	■		■	■	●
SAVI		▲	▲		■		●
MSAVI	▲	▲	▲		■	■	●
EVI		▲	▲		■		●

### 5.3. Classificação ao nível do pixel

A classificação foi realizada com o algoritmo RF aplicado a dois cenários distintos: (1) utilização da combinação de bandas multiespectrais com índices de vegetação e (2) utilização dos índices de vegetação apenas. As estatísticas relativas às PG, PP, PU apresentadas no Quadro 8 referem-se à classificação RF para ambos os níveis de nomenclatura: tipo de cultura e tipo de estação.

**Quadro 8.** Precisões do utilizador (PU) e do produtor (PP) para cada classe e precisão global (PG) para o método de classificação *Random Forest* ao nível do tipo de cultura e da estação para ambos os *datasets*: B + IV (*dataset* que inclui refletâncias e índices de vegetação) e IV (*dataset* que inclui apenas índices de vegetação). Para cada caso foi usado o conjunto de variáveis ótimo originado pela MDG em cada classificação. A coluna Nvar representa o tamanho desse conjunto.

	Nvar	Nº pixels	PU		PP		Precisão global		
			B + IV	IV	B + IV	IV	B + IV	IV	
<b>Culturas</b>		Aveia	609	96.38	90.64	96.06	97.70	<b>98.26</b>	<b>97.06</b>
		Cevada	1339	97.93	98.36	98.88	93.80		
		Girassol	686	99.42	98.69	99.13	99.27		
	25	Melão	404	99.26	99.26	99.50	99.50		
		Milho	490	99.79	98.98	98.98	99.59		
		Trigo	953	97.90	95.91	97.69	96.62		
		Triticale	113	98.17	92.92	94.69	98.13		
<b>Estações</b>		outono-inverno	3014	99.67	99.67	99.60	99.67	<b>99.52</b>	<b>99.70</b>
	8	primavera-verão	1580	99.24	99.75	99.37	99.75		

**Quadro 9.** Matriz de erro estimada por validação cruzada 10-*fold* resultante da aplicação do algoritmo RF aos 4594 pixels do *dataset* B + IV (*dataset* que inclui refletâncias e índices de vegetação).

	Classes de referência						
	Aveia	Cevada	Girassol	Melão	Milho	Trigo	Triticale
Aveia	585	8	1	0	0	12	1
Cevada	11	1324	4	1	0	10	2
Girassol	0	1	680	1	2	0	0
Melão	0	0	0	402	3	0	0
Milho	0	0	1	0	485	0	0
Trigo	11	6	0	0	0	931	3
Triticale	2	0	0	0	0	0	107

De uma maneira geral as culturas classificadas apresentam uma PP superior a 93.8% e uma PU superior a 90.1%, demonstrando um desempenho da classificação do tipo de culturas ao nível do pixel relativamente bom. A classificação do tipo de culturas apresentou melhores resultados para o cenário em que se utilizou a combinação entre as bandas multiespectrais e os índices de vegetação, em que se obteve uma PG de 98.26%. Com este resultado recomenda-se a escolha da utilização desta combinação (B+IV) para a classificação do tipo



de culturas. Apesar dos bons resultados da precisão, a matriz de erro estimada apresentada no Quadro 9 revela alguma confusão sobretudo na classificação de culturas de cereais, nomeadamente entre trigo e aveia, trigo e cevada, e aveia e cevada.

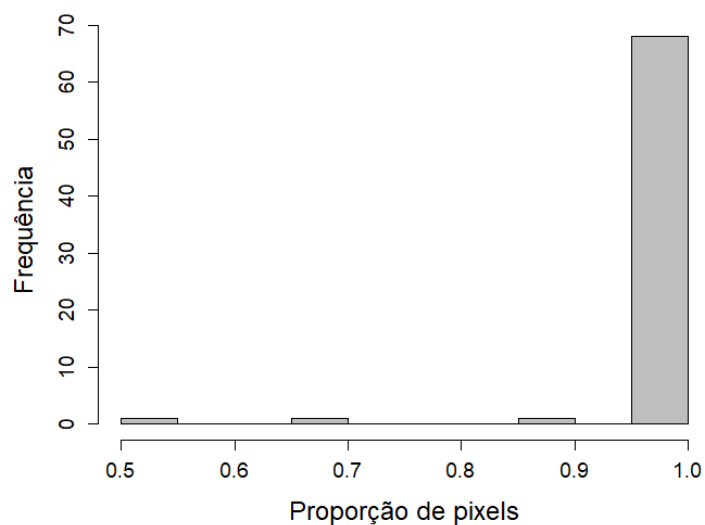
Quanto à classificação do tipo de estação as precisões não se revelaram diferentes perante os dois cenários distintos considerados. Pelo que a escolha do *dataset* que inclui apenas os índices de vegetação seria o suficiente para obter resultados considerados bastante bons sem que isso acarretasse custos no tempo de computação, uma vez que os índices de vegetação não são mais nem menos combinações matemáticas de várias bandas multiespectrais compiladas numa só banda sintética.

Tendo em conta o cenário que maximiza o desempenho da classificação, a cultura que apresentou piores precisões de classificação foi o triticale, que comparativamente com as restantes, obteve uma menor PU.

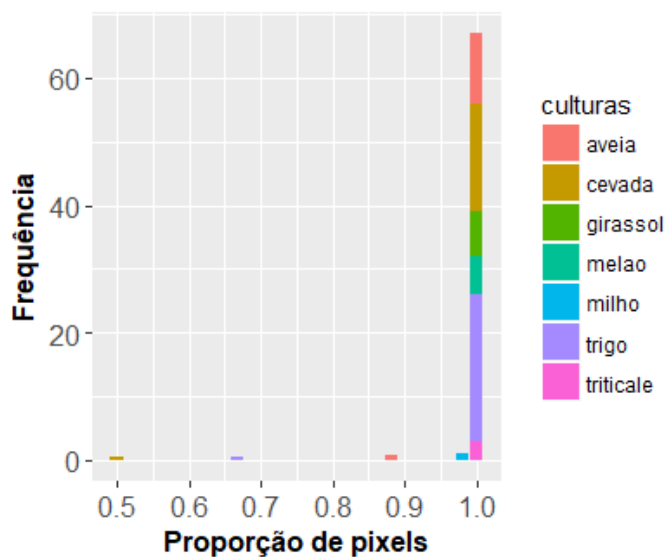
#### **5.4. Classificação ao nível da parcela**

Adicionalmente à abordagem da classificação ao nível do pixel realizou-se uma abordagem ao nível da parcela para o conjunto de culturas utilizando o *dataset* completo (B+IV). O objetivo principal deste tipo de abordagem é testar a aplicação da classificação num contexto operacional em que se pretende classificar cada parcela e não cada pixel. Note-se que, devido à diferença na dimensão das parcelas (Quadro 5), os valores de precisão ao nível do píxel e ao nível da parcela podem ser bastante distintos.

A proporção de pixels classificados na classe mais frequente por parcela alcançou os 100% em mais de 60 parcelas, num total de 69 parcelas (Figura 10). É de notar que a proporção mínima pela qual uma classe foi atribuída a uma dada parcela foi de 50%. Todas as classes, à exceção do milho, foram passíveis de serem escolhidas para a classificação das parcelas com proporções de 1 (Figura 11).



**Figura 10.** Proporção de pixels que foram classificados na classe mais frequente.



**Figura 11.** Proporção de pixels classificados por maioria por classe e por parcela

No Quadro 10 estão apresentados os resultados das precisões da classificação.

Da perspectiva do produtor constatou-se o seguinte:

- 89.6% das parcelas cultivadas com aveia foram corretamente classificadas, apesar de 7.4% das parcelas terem sido classificadas como trigo e 1.5% como cevada e triticale;
- 95.6% das parcelas cultivada com cevada foram corretamente classificadas, no entanto 3.3% das parcelas foram classificadas como trigo e 1.1% como aveia;
- Todas as parcelas com a cultura do girassol e melão foram bem classificadas;

- 90.1% das parcelas cultivadas com trigo foram corretamente classificadas, sendo que as restantes ora foram erradamente classificadas como aveia ora como cevada;
- A cultura do triticale foi a que apresentou o valor mais baixo, com 8.3% das parcelas atribuídas à cevada e ao trigo e 4.2% atribuídas à aveia, e de seguida foi a cultura do milho com 83.3% das parcelas bem classificadas.

Do ponto de vista do utilizador os resultados mostraram o seguinte:

- 100% das parcelas classificadas como girassol e milho foram bem classificadas;
- A cultura da cevada contou com o valor mais baixo da percentagem de parcelas bem classificadas (89.2%), sendo as restantes erradamente classificadas como aveia, trigo ou triticale;
- O melão conta com 96.3% das parcelas bem classificadas, com 3.7% das parcelas atribuídas ao milho;
- 91.4% e 90.5% das parcelas identificadas como trigo e triticale respetivamente são na realidade trigo e triticale, sendo as restantes parcelas aveia, cevada ou triticale e aveia, respetivamente.

A cultura do girassol foi quem devolveu os melhores resultados em ambas as precisões.

É de notar que as culturas com curvas de crescimento semelhantes, i.e., fenologia similar, tendem a ser mais passíveis de serem confundidas no processo de classificação por apresentarem comportamentos espectrais semelhantes. Posto isto, é claro que existe uma confusão apreciável da classificação entre culturas de cereais.

**Quadro 10.** Matriz de erro média e valores da precisão do produtor (PP), do utilizador (PU) e global (PG) baseada no modelo RF final.

	Classes de referência							PU (%)
	Aveia	Cevada	Girassol	Melão	Milho	Trigo	Triticale	
Aveia	121	2	0	0	0	4	1	94.5
Cevada	2	173	0	0	0	17	2	89.2
Girassol	0	0	78	0	0	0	0	100
Melão	0	0	0	52	2	0	0	96.3
Milho	0	0	0	0	10	0	0	100
Trigo	10	6	0	0	0	191	2	91.4
Triticale	2	0	0	0	0	0	19	90.5
<b>PP (%)</b>	<b>89.6</b>	<b>95.6</b>	<b>100</b>	<b>100</b>	<b>83.3</b>	<b>90.1</b>	<b>79.2</b>	<b>PG = 92.8</b>



## 6. CONCLUSÕES

Neste estudo pretendeu-se avaliar o potencial do Sentinel-2A na classificação da ocupação do solo no controlo de subsídios agrícolas. Adicionalmente explorou-se a possibilidade de combinar dados de deteção remota provenientes de diferentes sensores e investigou-se o potencial da integração de diferentes índices de vegetação para além dos valores de refletância na classificação de culturas.

A combinação de dados de deteção remota ao longo de uma série temporal provenientes de vários sensores só pode ser possível se for demonstrada a sua consistência ou se for adotada uma estratégia adequada para a integração desses dados (Lessio, Fissore, & Borgogno-Mondino, 2017). Com este objetivo foi aplicado um modelo ANOVA a dois fatores hierarquizados para testar a homogeneidade espectral das parcelas (Schmedtmann & Campagnolo, 2015). Os valores de F elevados obtidos comprovaram a homogeneidade das parcelas, i.e., não existe problemas significativos em combinar imagens de resoluções distintas, neste caso Landsat-7, Landsat-8 e Sentinel-2A.

O método de seleção de variáveis baseado na análise da importância das variáveis segundo o índice MDG permitiu escolher o melhor conjunto de variáveis que otimiza o resultado da classificação para cada nível hierárquico de ocupação do solo, estação e cultura. As variáveis mais importantes consideradas pelo algoritmo de classificação RF foram predominantemente derivadas do Sentinel-2A, apesar dos restantes sensores se revelarem também importantes, para o caso da classificação ao nível das culturas.

A abordagem da classificação ao nível da parcela permitiu averiguar a utilidade da classificação num contexto prático da verificação dos critérios de elegibilidade para a atribuição de subsídios. Para evitar sobre estimar o desempenho do classificador, a precisão da classificação automática foi avaliada por validação cruzada. Uma precisão global de 92.8% foi estimada para a classificação ao nível da parcela, tendo-se obtido valores sempre superiores a 79.2% para a precisão do produtor. A maioria dos erros de classificação concentraram-se entre culturas de cereais, i.e., culturas de outono/inverno.

Na classificação ao nível do pixel, para o conjunto de culturas, destaca-se o aumento da precisão global para 98.26% quando se combinam as bandas multiespectrais com os índices de vegetação.

As culturas que apresentaram piores precisões de classificação foram sobretudo a aveia e o triticale. Na distinção entre estações a confusão é quase nula obtendo precisões globais superiores a 99%. É de notar que a distinção entre estações é relevante para o controlo das

medidas de *greening*, e por isso este resultado sugere que algumas etapas da automatização da monitorização usando DR podem ser realizadas com elevada fiabilidade.

O facto de os resultados da classificação serem na generalidade francamente bons, com elevada precisão e pouca confusão entre a maioria das classes, é um grande incentivo à aplicação deste tipo de técnicas no controlo de subsídios de forma a agilizar e economizar custos associados ao presente procedimento.

É de referir que, dada a natureza dinâmica da agricultura e de outros fatores, os resultados que se obtiveram para este ano e local não podem ser diretamente extrapolados para outro ano ou local. Assim, julgou-se que não faria muito sentido comparar resultados usando apenas imagens de um único sensor com séries temporais de vários sensores (S2, L7 e L8), dado que os resultados desse eventual exercício para o conjunto de dados tratados neste trabalho não teriam grande relevância em outro ano ou local. De facto, uma das particularidades da DR com sensores óticos é depender muito da aleatoriedade das condições meteorológicas na altura da passagem do sensor. O aspeto mais crítico para o bom desempenho de um classificador baseado em séries temporais de imagens de satélite é a distribuição temporal/espacial das observações pois um período de revisita curto permite identificar melhor as diferentes fases de desenvolvimento/crescimento que são muitas vezes impercetíveis quando se utilizam dados esporádicos, e por isso a disponibilidade de imagens é mais relevante do que a resolução espacial ou espectral das mesmas no quadro deste trabalho. Pelas razões descritas atrás, decidiu-se avaliar a potencialidade dos métodos baseados em DR utilizando toda a informação disponível e, nomeadamente, usando sempre a combinação dos três sensores considerados.

O grau de contribuição das imagens Sentinel-2A na agilização das operações de controlo no âmbito do *greening*, diversificação de culturas, terá que ser avaliado em maior profundidade, em colaboração com o IFAP. Este trabalho sugere, no entanto, que existe um grande potencial para maior integração de técnicas baseadas em DR na operacionalização do processo de controlo.

## BIBLIOGRAFIA

- Belgiu, M., & Drăgu, L. (2016). Random forest in remote sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing*, 114, 24–31. <https://doi.org/10.1016/j.isprsjprs.2016.01.011>
- Breiman, L. (2001). Random forests. *Machine Learning*, 45, 5–32. <https://doi.org/10.1023/A:1010933404324>
- Chander, G., Markham, B. L., & Helder, D. L. (2009). Summary of current radiometric calibration coefficients for Landsat MSS, TM, ETM+, and EO-1 ALI sensors. *Remote Sensing of Environment*, 113(5), 893–903. <https://doi.org/10.1016/j.rse.2009.01.007>
- Comissão Europeia. (2012). A Política Agrícola Comum - A história continua. Obtido 23 de Março de 2018, de [http://ec.europa.eu/agriculture/50-years-of-cap/files/history/history\\_book\\_lr\\_pt.pdf](http://ec.europa.eu/agriculture/50-years-of-cap/files/history/history_book_lr_pt.pdf)
- Comissão Europeia. (2013a). Regulamento (UE) N.º 1307/2013 do Parlamento Europeu e do Conselho de 17 de dezembro de 2013 que estabelece regras para os pagamentos diretos aos agricultores ao abrigo de regimes de apoio no âmbito da política agrícola comum e que revoga o Regulamento (C. Obtido 16 de Abril de 2018, de <https://eur-lex.europa.eu/legal-content/PT/TXT/PDF/?uri=CELEX:32013R1307&from=PT>
- Comissão Europeia. (2013b). Regulamento (UE) N.º 1306/2013 do Parlamento Europeu e do Conselho de 17 de Dezembro de 2013 relativo ao financiamento, à gestão e ao acompanhamento da Política Agrícola Comum e que revoga os Regulamentos (CEE) n.º 352/78, (CE) n.º 165/94, (CE) n.º 2799/98. Obtido 16 de Abril de 2018, de <https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2013:347:0549:0607:pt:PDF>
- Comissão Europeia. (2014). Regulamento de Execução (UE) N.º 809/2014 da Comissão de 17 de julho de 2014 que estabelece as normas de execução do Regulamento (UE) n.º 1306/2013 do Parlamento Europeu e do Conselho no que diz respeito ao sistema integrado de gestão e de controlo, às me. Obtido 3 de Abril de 2018, de <https://eur-lex.europa.eu/legal-content/PT/TXT/PDF/?uri=CELEX:32014R0809&from=EN>
- Comissão Europeia. (2015). Summary Report 28th Rural Development Committee 18 november 2015. Obtido 24 de Abril de 2018, de <https://ec.europa.eu/agriculture/sites/agriculture/files/committees/rural-development/0028.pdf>
- Comissão Europeia. (2016a). Agricultural Monitoring. Obtido 16 de Maio de 2018, de <https://ec.europa.eu/jrc/en/research-topic/agricultural-monitoring>
- Comissão Europeia. (2016b). Earth Observation. Obtido 24 de Abril de 2018, de <https://ec.europa.eu/jrc/en/research-topic/earth-observation>
- Comissão Europeia. (2018a). Agriculture and rural development - Direct Support. Obtido 23 de Março de 2018, de [https://ec.europa.eu/agriculture/direct-support\\_en](https://ec.europa.eu/agriculture/direct-support_en)
- Comissão Europeia. (2018b). Integrated Administration and Control System (IACS). Obtido 24 de Março de 2018, de [https://ec.europa.eu/agriculture/direct-support/iacs\\_pt](https://ec.europa.eu/agriculture/direct-support/iacs_pt)
- Congalton, R. G., & Green, K. (2009). *Assessing the accuracy of remotely sensed data: principles and practices (2ª)*. Boca Raton, Estados Unidos da América: CRS Press.
- Devos, W., Fasbender, D., Lemoine, G., Loudjani, P., Milenov, P., & Wirnhardt, C. (2017). *Discussion document on the introduction of monitoring to substitute OTSC*.

<https://doi.org/10.2760/258531>

- Du, P., Samat, A., Waske, B., Liu, S., & Li, Z. (2015). Random Forest and Rotation Forest for fully polarized SAR image classification using polarimetric and spatial features. *ISPRS Journal of Photogrammetry and Remote Sensing*, 105, 38–53. <https://doi.org/10.1016/j.isprsjprs.2015.03.002>
- EDIA. (2016). Anuário agrícola de Alqueva 2016. Obtido 14 de Dezembro de 2017, de [http://www.alqueva.com.pt/folder/galeria/ficheiro/56\\_anuario\\_agricola2016\\_kwvvp6cslw.pdf](http://www.alqueva.com.pt/folder/galeria/ficheiro/56_anuario_agricola2016_kwvvp6cslw.pdf)
- Estrada, J., Sánchez, H., Hernanz, L., Checa, M., & Roman, D. (2017). Enabling the Use of Sentinel-2 and LiDAR Data for Common Agriculture Policy Funds Assignment. *ISPRS International Journal of Geo-Information*, 6, 255. <https://doi.org/10.3390/ijgi6080255>
- FAO. (2017). *The future of food and agriculture - Trends and challenges*. Rome.
- Filizzola, C., Corrado, R., Falconieri, A., Faruolo, M., Genzano, N., Lisi, M., ... Tramutoli, V. (2017). On the use of temporal vegetation indices in support of eligibility controls for EU aids in agriculture. *International Journal of Remote Sensing*, 00(00), 1–27. <https://doi.org/10.1080/01431161.2017.1395973>
- Flood, N. (2017). Comparing Sentinel-2A and Landsat 7 and 8 using surface reflectance over Australia. *Remote Sensing*, 9(7), 1–14. <https://doi.org/10.3390/rs9070659>
- Gao, F., Hilker, T., Zhu, X., Anderson, M., Masek, J., Wang, P., & Yang, Y. (2015). Fusing Landsat and MODIS Data for Vegetation Monitoring. *IEEE Geoscience and Remote Sensing Magazine*, 3(3), 47–60. <https://doi.org/10.1109/MGRS.2015.2434351>
- Genuer, R., Poggi, J., Tuleau-malot, C., Genuer, R., Poggi, J., Variable, C. T., ... Tuleau-malot, C. (2012). Variable selection using Random Forests To cite this version :, 31(14).
- Gislason, P. O., Benediktsson, J. A., & Sveinsson, J. R. (2006). Random forests for land cover classification. *Pattern Recognition Letters*, 27, 294–300. <https://doi.org/10.1016/j.patrec.2005.08.011>
- Gitelson, A. A., Kaufman, Y. J., & Merzlyak, M. N. (1996). Use of a Green Channel in Remote Sensing of Global Vegetation from EOS-MODIS, 58(3), 289–298.
- GPAa. (2005). Plano de intervenção para a zona de Alqueva - Caracterização da zona de Alqueva. Obtido 14 de Dezembro de 2017, de [http://sir.dgadr.gov.pt/images/conteudos/gpaa/tomos/tomo\\_2.pdf](http://sir.dgadr.gov.pt/images/conteudos/gpaa/tomos/tomo_2.pdf)
- Hatfield, J. L., & Prueger, J. H. (2010). Value of using different vegetative indices to quantify agricultural crop characteristics at different growth stages under varying management practices. *Remote Sensing*, 2(2), 562–578. <https://doi.org/10.3390/rs2020562>
- Huete, A. R. (1988). A Soil-Adjusted Vegetation Index (SAVI). *Remote Sensing of Environment*, 25(3), 295–309.
- Huete, A. R., Liu, H. Q., Batchily, K., & Leeuwen, W. Van. (1997). A Comparison of Vegetation Indices over a Global Set of TM Images for EOS-MODIS. *Remote Sensing of Environment*, 59(3), 440–451.
- Immitzer, M., Atzberger, C., & Koukal, T. (2012). Tree species classification with Random forest using very high spatial resolution 8-band worldView-2 satellite data. *Remote Sensing*, 4, 2661–2693. <https://doi.org/10.3390/rs4092661>



- INE. (2017). Análise sumária da evolução das características estruturais das explorações agrícolas - Inquérito à Estrutura das Explorações Agrícolas 2016. Obtido 14 de Dezembro de 2017, de [http://www.gpp.pt/images/Agricultura/Estatisticas\\_e\\_Analises/Estatisticas/AnaliseEstruturaExplAgricolas2016.pdf](http://www.gpp.pt/images/Agricultura/Estatisticas_e_Analises/Estatisticas/AnaliseEstruturaExplAgricolas2016.pdf)
- IPMA. (2017). Normais Climatológicas. Obtido 14 de Dezembro de 2017, de <https://www.ipma.pt/pt/oclima/normais.clima/>
- Kohavi, R. (1995). A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection. Em *International Joint Conference on Artificial Intelligence* (Vol. 14, pp. 1137–1145). <https://doi.org/10.1067/mod.2000.109031>
- Lebourgeois, V., Dupuy, S., Vintrou, É., Ameline, M., Butler, S., & Bégué, A. (2017). A combined random forest and OBIA classification scheme for mapping smallholder agriculture at different nomenclature levels using multisource data (simulated Sentinel-2 time series, VHRS and DEM). *Remote Sensing*, 9(3), 1–20. <https://doi.org/10.3390/rs9030259>
- Liaw, a, & Wiener, M. (2002). Classification and Regression by randomForest. *R news*, 2, 18–22. <https://doi.org/10.1177/154405910408300516>
- Loudjani, P. (2016). TECHNICAL GUIDANCE FOR ON-THE-SPOT CHECKS ( OTSC ) AND AREA MEASUREMENT REGULATION ( EU ) 2015 / 2333 The-Spot Checks, (809).
- Massot, A. (2018). Política Agrícola Comum (PAC) - Instrumentos da PAC e respetivas reformas. Obtido 23 de Abril de 2018, de <http://www.europarl.europa.eu/factsheets/pt/sheet/107/instrumentos-da-pac-e-respetivas-reformas>
- Palchowdhuri, Y., Valcarce-Diñeiro, R., King, P., & Sanabria-Soto, M. (2018). Classification of multi-temporal spectral indices for crop type mapping: a case study in Coalville, UK. *Journal of Agricultural Science*, 156(1), 1–13. <https://doi.org/10.1017/S0021859617000879>
- Portal do Clima. (2017). No Title. Obtido 14 de Dezembro de 2017, de <http://portaldoclima.pt>
- Prasad, A. M., Iverson, L. R., & Liaw, A. (2016). Newer Classification and Regression Tree Techniques : Bagging and Random Forests for Ecological Prediction, 9(2), 181–199. <https://doi.org/10.1007/S10021-005-0054-1>
- Rodriguez-Galiano, V. F., Chica-Olmo, M., Abarca-Hernandez, F., Atkinson, P. M., & Jeganathan, C. (2012). Random Forest classification of Mediterranean land cover using multi-seasonal imagery and multi-seasonal texture. *Remote Sensing of Environment*, 121, 93–107. <https://doi.org/10.1016/j.rse.2011.12.003>
- Rodriguez-Galiano, V. F., Ghimire, B., Rogan, J., Chica-Olmo, M., & Rigol-Sanchez, J. P. (2012). An assessment of the effectiveness of a random forest classifier for land-cover classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, 67, 93–104. <https://doi.org/10.1016/j.isprsjprs.2011.11.002>
- Rouse, J. W., Haas, R. H., Schell, J. A., & Deering, D. W. (1974). Monitoring vegetation systems in the great plains with ERTS. Em S. C. Freden, E. P. Mercanti, & M. A. Becker (Eds.), *Proceedings of the Third ERTS Symposium* (pp. 309–317). Washington DC, USA: NASA.
- Roy, D. P., Wulder, M. A., Loveland, T. R., C.E., W., Allen, R. G., Anderson, M. C., ... Zhu, Z. (2014). Landsat-8: Science and product vision for terrestrial global change research.

*Remote Sensing of Environment*, 145, 154–172.  
<https://doi.org/10.1016/j.rse.2014.02.001>

Schmedtmann, J., & Campagnolo, M. L. (2015). Reliable crop identification with satellite imagery in the context of Common Agriculture Policy subsidy control. *Remote Sensing*, 7(7), 9325–9346. <https://doi.org/10.3390/rs70709325>

Sitokonstantinou, V., Papoutsis, I., Kontoes, C., Arnal, A. L., Andrés, A. P. A., & Zurbano, J. A. G. (2018). Scalable parcel-based crop identification scheme using Sentinel-2 data time-series for the monitoring of the common agricultural policy. *Remote Sensing*, 10(6), 911. <https://doi.org/10.3390/rs10060911>

Tatsumi, K., Yamashiki, Y., Canales Torres, M. A., & Taïpe, C. L. R. (2015). Crop classification of upland fields using Random forest of time-series Landsat 7 ETM+ data. *Computers and Electronics in Agriculture*, 115, 171–179. <https://doi.org/10.1016/j.compag.2015.05.001>

Testa, S., Mondino, E. C. B., & Pedroli, C. (2014). Correcting MODIS 16-day composite NDVI time-series with actual acquisition dates. *European Journal of Remote Sensing*, 47(1), 285–305. <https://doi.org/10.5721/EuJRS20144718>

United Nations, Department of Economic Social Affairs, & Population Division. (2017). *World Population Prospects: The 2017 Revision, Key Findings and Advance Tables*. (No. ESA/P/WP/248).

USGS. (2017). Product guide - Landsat 8 surface reflectance code (LaSRC) product. Obtido 18 de Dezembro de 2017, de [https://landsat.usgs.gov/sites/default/files/documents/lasrc\\_product\\_guide.pdf](https://landsat.usgs.gov/sites/default/files/documents/lasrc_product_guide.pdf)

USGS. (2018). Product guide - Landsat 4-7 surface reflectance (LEDAPS) product. Obtido 20 de Fevereiro de 2018, de [https://landsat.usgs.gov/sites/default/files/documents/ledaps\\_product\\_guide.pdf](https://landsat.usgs.gov/sites/default/files/documents/ledaps_product_guide.pdf)

Zhang, H. K., & Roy, D. P. (2017). Using the 500 m MODIS land cover product to derive a consistent continental scale 30 m Landsat land cover classification. *Remote Sensing of Environment*, 197, 15–34. <https://doi.org/10.1016/j.rse.2017.05.024>