

Desenvolvimento de algoritmos paralelizáveis para a atualização da COSSim

Contrato de cooperação 261/2021 entre a DGT e o ISA celebrado no âmbito do projeto POCI-05-5762-FSE-000368

Início: 23 de julho de 2021; duração: 12 meses

Entregável 1: Abordagem metodológica para deteção e classificação de alterações de ocupação do solo com base em imagens de satélite

Manuel Campagnolo, Bruno Barbosa e Daniel Moraes

23 de setembro de 2021

1. Glossário e abreviações	1
2. Tarefas e entregáveis	1
3. Produtos: estrutura, datas e legenda	2
4. Objectivos	4
5. Dados de input	5
6. Deteção de alterações: estado da arte	9

7. Estratégias para o problema COSSim	10
7.1 Deteção de alterações	10
7.2 Classificação	11
7.3 Transições válidas	12
8. Validação de resultados	14
9. Exemplos de detecção com CCDC	14
Primeiro exemplo: cortes em povoamentos de eucalipto	16
Segundo exemplo: limpeza de matos em povoamento de pinheiro manso	17
Terceiro exemplo: transição de agricultura anual para agricultura permanente	20
10. Referências	22

1. Glossário e abreviações

cdg: conjunto de dados geográficos (pode ser vectorial ou “raster”)

COS: Carta de Ocupação do Solo (na verdade, carta de uso do solo, vetorial)

COSSim : Carta de Ocupação do Solo Simplificada

COSSimRef : Carta de Ocupação do Solo Simplificada de referência (ano coincidente com a COS)

COSSimIntercalar: COSSim para os restantes anos

GEE: Google Earth Engine

HPC: High performance computing

OA: overall accuracy

SMOS: Sistema de Monitorização de Ocupação do Solo -- um sistema baseado em tecnologias digitais do espaço e inteligência artificial

S2: Sentinel 2/MSI

UP: unidade de paisagem - Portugal Continental está dividido em 14 UPs

2. Tarefas e entregáveis

De acordo com o contrato de colaboração, as tarefas e entregáveis são os seguintes.

Tarefa 1 – Elaboração da abordagem metodológica para deteção e classificação de alterações de ocupação do solo com base em imagens de satélite

Duração: Mês 1 – Mês 2

Entregável:

- E1 – Relatório com abordagem metodológica para deteção e classificação de alterações de ocupação do solo com base em imagens de satélite

Tarefa 2 – Definição de uma metodologia para deteção de alterações em áreas de floresta e mato com uma periodicidade de um ou dois meses

Duração: Mês 3 – Mês 5

Entregáveis:

- E2.1 – Algoritmo para deteção de alterações de ocupação do solo
- E2.2 – Relatório com a avaliação da qualidade dos resultados produzidos pelo algoritmo

Tarefa 3 – Definição de uma metodologia para classificação das alterações utilizando séries multitemporais de imagens do Sentinel 2 e integração de conhecimento pericial

Duração: Mês 7 – Mês 10

Entregáveis:

- E3.1 – Algoritmo para classificação de ocupação do solo
- E3.2 – Relatório com a avaliação da qualidade dos resultados produzidos pelo algoritmo

Tarefa 4 – Produção das versões finais dos algoritmos

Duração: Mês 11– Mês 12

- E4.1 – Algoritmo final para deteção de alterações de ocupação do solo
- E4.2 – Algoritmo final para classificação de ocupação do solo

3. Produtos: estrutura, datas e legenda

A COS é um produto vectorial de uso do solo. A COSSim é um produto raster (ou grid) de ocupação do solo. Distingue-se a COSSimRef produzida para o mesmo ano da COS e a COSSimIntercalar, produzida para os anos intercalares. A data de referência de um produto corresponde à data em que existe sobre o terreno o uso ou ocupação do solo que o produto representa.

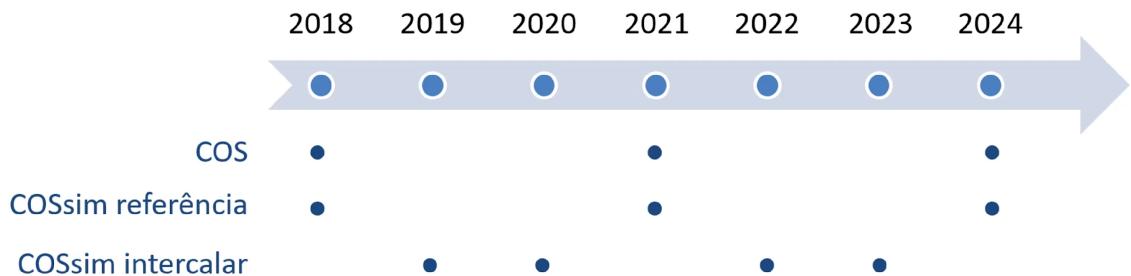


Figura 1: datas de referência dos produtos

	2021			2022			2023			2024			2025			2026		
	Jan	Fev-Nov	Dez	Jan	Fev-Nov	Dez	Jan	Fev-Nov	Dez	Jan	Fev-Nov	Dez	Jan	Fev-Nov	Dez	Jan	Fev-Nov	Dez
COS						2021											2024	
COSSim ref							2021										2024	
COSSim int	2020			2021			2022			2023			2024			2025		

↔ 3-4 anos sem atualização de certas classes
devido ao ciclo e calendário da COS

Figura 2: datas de referência dos produtos e datas de publicação dos produtos

O prefixo “sim” indica uma legenda simplificada, com um número de classes muito menor do que a legenda da COS.

2018	2020
Artificializado	Artificializado
Agricultura	Agricultura de outono/inverno
	Agricultura de primavera/verão
	Outras áreas agrícolas
Sobreiro e Azinheira	Sobreiro e Azinheira

Eucalipto	Eucalipto
Outras folhosas	Outras folhosas
Pinheiro bravo	Pinheiro bravo
Pinheiro manso	Pinheiro manso
Outras resinosas	Outras resinosas
Matos	Matos
Vegetação herbácea espontânea	Vegetação herbácea espontânea
Superfícies sem vegetação	Superfícies sem vegetação
Zonas húmidas	Zonas húmidas
Água	Água

Figura 3: Legendas propostas pela DGT para a COSSimRef2018 e para cada COSSimIntercalar (neste caso para 2020)

4. Objectivos

O objectivo da colaboração DGT/ISA é o desenvolvimento de algoritmos para atualização da COSSim. Dada a complexidade do problema e o risco de criar muitos falsos positivos, apenas as alterações com maior impacto e utilidade para os utilizadores deverão ser consideradas. Nesse sentido a DGT propõe os critérios e casos particulares listados abaixo como aproximação para o universo das localizações (pixels S2) a considerar na atualização da COSSim.

As principais dinâmicas de alteração da COSSim podem ser caracterizadas de acordo com os seguintes critérios:

- Áreas florestais: alterações em áreas de uso florestal (não muda o uso)
- Áreas agrícolas: alterações em áreas de uso agrícola (pode ou não mudar o uso)
- Conversão e abandono do solo: muda a ocupação e o uso
- Não alterações potencialmente problemáticas: alterações espectrais não relacionadas com alterações temáticas
- Modos de alteração: a forma como as alterações acontecem ao longo do tempo

Para as áreas florestais, considere-se que existem duas possibilidades: perdas ou ganhos de vegetação, nos seguintes casos:

1. Perdas de vegetação (COSSimP), que podem corresponder a “Corte de ec, pb, pm”; Corte de Sb/Az; Fogo de ec, pb, pm; Fogo de Sb/Az; Corte de subcoberto em floresta aberta
2. Ganhos de vegetação em áreas cortadas ou recentemente ardidas: Novas herbáceas ou matos; Novo eucaliptal em antigo eucaliptal (regeneração); Novo eucaliptal em antigo eucaliptal (plantação); Novo eucaliptal onde não o era; Novo pinhal em antigo pinhal; Novo pinhal onde não o era; Novo Sb/Az onde não o era
3. Ganhos de vegetação por adensamentos (pinheiros e Sb/Az)

Para as áreas agrícolas consideram-se:

1. Rotação de culturas anuais (sequeiro, regadio, pousio)
2. Conversão de agricultura anual ou pastagens para permanente (intensiva)
3. Novas estufas
4. Alternância entre cultura anual e pastagem (outras áreas agrícolas)

Para conversão e abandono do solo:

1. Nova floresta em áreas não florestais (pode ser apanhado pela metodologia): Espécies de crescimento rápido (ec); Espécies de crescimento lento (pb, pm, Sb/Az)
2. Nova agricultura em áreas não agrícolas (não é apanhado pela metodologia): Agricultura anual ou pastagens; Agricultura permanente; Novas estufas
3. Abandono da agricultura (é apanhado pela metodologia): Novos matos; Novas herbáceas
4. Aumento ou redução de água (não é apanhado pela metodologia)
5. Novo artificializado (não é apanhado pela metodologia)

As não-alterações potencialmente problemáticas são

1. Outras Folhosas de folha caduca
2. Rios
3. Sombra
4. Neve

Finalmente, os modos de alteração a considerar são:

1. Alteração simples e directa - Ex: Floresta → sem vegetação
2. Alteração simples com classe transitória* - Ex: Mato → sem vegetação → artificializado (*Classes transitória pode ser, por exemplo, uma terraplenagem para construção ou agricultura intensiva)
3. Alterações múltiplas em anos consecutivos - Floresta → sem vegetação → herbáceas → mato

Espera-se da DGT:

1. **Dinâmica das transições: matriz de transição entre classes com indicação das áreas para se poder compreender a importância absoluta dos tipos de transições para o problema COSSim**

5. Dados de input

É fundamental para a colaboração entre a DGT e o ISA que haja uma partilha fácil de dados, e uma descrição clara dos dados de input do problema. Nesse sentido, faz-se abaixo uma sugestão das estruturas de dados que poderiam suportar os dados de input. Escolhe-se como exemplo, os dados para elaboração da COSSim2019, embora a estrutura seja idêntica para outros anos.

Embora se utilize o formato “grid” nesta seção, e dado que os produtos corresponderão espacialmente às imagens S2, com resolução de 10-20 m, o formato grid pode ser facilmente convertido em formato “raster” e vice-versa. Haverá um raster por cada campo da grid. Isto aplica-se nomeadamente à tabela PixelS2_2019 abaixo.

As tabelas que sugerimos incluir na base de dados como inputs da COSSim2019 (cdg que deveria ser disponibilizado em janeiro 2020) em formato “grid” seriam as seguintes. Note-se que as tabelas PixelS2_2019 e VALIDACAO serão organizadas e criadas pela DGT e pelo ISA em colaboração. As restantes tabelas correspondem aos conjuntos de dados públicos criados pela DGT, ICNF e IFAP, de que o ISA terá que ter acesso.

1. PixelS2_2019: cada linha corresponde a um pixel 10*10 m para Portugal Continental. Há aproximadamente 890 milhões de linhas ($890 \cdot 10^6 \cdot 100 \text{ m}^2 = 89000 \text{ km}^2$). Em particular, esta tabela vai incluir a informação da COSSim de referência 2018 . Será nesta tabela que serão também integrados os índices, datas, magnitudes, etc, resultantes do processamento das séries temporais de imagens S2. A descrição dos campos desta tabela é feita mais abaixo. Uma parte desta tabela está ilustrada na na [Seção “Exemplos”](#).
2. COS2018: cada linha corresponde a uma feature da COS2018. Idem para COS2015, etc
3. AA2018: cada linha corresponde a uma feature de AA2018 do ICNF. Idem para AA2017, etc
4. Parcelario2019: cada linha corresponde a uma feature do parcelário IFAP (disponibilizado em julho 2019). Idem para Parcelario2018, etc.
5. VALIDACAO: cada linha corresponde a uma observação (treino ou teste), o que tem sido designado pela DGT como “pontos de treino”. Os campos respectivos serão ID (inteiro), LAT (decimal), LONG (decimal), data real da alteração da ocupação do solo; ClasseCosSim1 (inteiro ou texto), ..., ClasseCosSim4, outros atributos a discutir

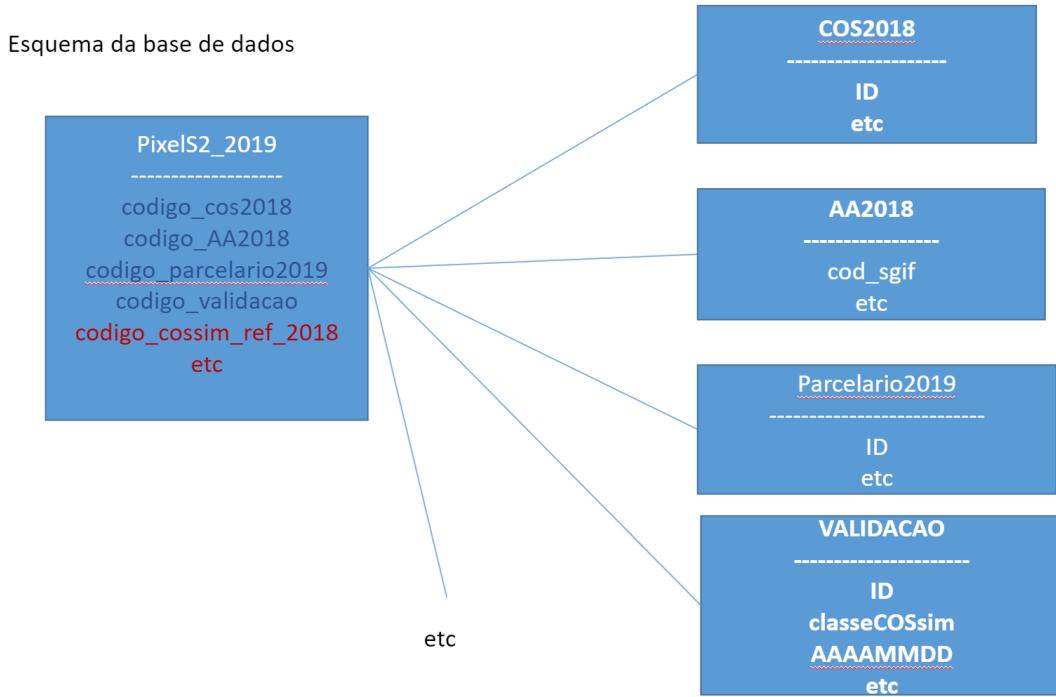


Figura 4: Esquema da base de dados a partilhar entre a DGT e o ISA

Os atributos (campos) de PixelS2_2019 deverão ser os seguintes:

- ID: ID do pixel (inteiro)
- LAT: Lat do pixel (decimal)
- LONG: Long do pixel (decimal)
- ID da unidade de paisagem UP (inteiro entre 1 e 14)
- codigo_cos2018: ID da feature da COS2018 onde está o centro do pixel (inteiro)
- dist_cos2018: distância em metros do (centro do) pixel ao limite da feature da COS2018 a que está associado (inteiro)
- codigo_AA2018: ID da feature de AA2018 onde está o centro do pixel (inteiro ou null)
- dist_AA2018: distância em metros do pixel ao limite da feature a que está associado (inteiro ou null)
- codigo_parcelario2019: ID da feature de parcelario2019 onde está o centro do pixel (inteiro ou null)

- dist_parcelario2019: distância em metros do pixel ao limite da feature a que está associado (inteiro ou null)
- codigo_validacao: ID da feature de VALIDACAO a que está associado (inteiro ou null)
- codigo_cossim_ref_2018: classe da COSSimref2018 (texto)
- conf_cossim_ref_2018: classe de confiança na classificação do pixel na cossim_ref_2018 (inteiro ou texto)
- outras variáveis que caracterizam o pixel e, em particular, indicadores, índices, datas, e outras variáveis sobre o pixel e a sua vizinhança resultante do processamento das séries temporais S2.

As tabelas da DGT (COS), ICNF (AA) e IFAP (parcelário) têm a sua própria estrutura. Será necessário ser possível cruzar a tabela Pixels2_2019 com qualquer uma delas.

Para referência ficam indicados os campos das tabelas COS2018 e Áreas ardidas 2018. É necessário identificar os campos da tabela dos parcelários (IFAP) para a qual não foi encontrada documentação.

Campos da tabela alfanumérica associada à COS2018

(http://mapas.dgterritorio.pt/atom-dgt/pdf-cous/COS2018/ET-COS-2018_v1.pdf)

Nome do campo Conteúdo Formato

- ID Identificador numérico único dos polígonos Long integer
- COS2015_n1 Código das classes de ocupação/uso do solo ao nível 1 String (?2018)
- COS2015_n4 Código das classes de ocupação/uso do solo ao nível 4 String (?2018)
- COS2018_Lg Descrição das classes de ocupação/uso do solo String
- AREA Área dos polígonos (ha) Double

Não foi encontrada documentação com a descrição dos campos da tabela de atributos dos cdg de áreas ardidas (ICNF). Na figura abaixo está uma porção da tabela que permite perceber como está estruturada.

	cod_sgif	cod_ncco	tipo	distrito	concelho	freguesia	local	data_inicio	data_fim	Ano	area_ha
1	TM218736	2018040031219	Florestal	Bragança	Torre de ...	Lousa	LOUSA (...)	2018-10-10 20:26:00.000	2018-10-11 04:30:00.000	2018	13,85153...
2	TM218720	2018040030679	Florestal	Bragança	Vinhais	Pinheiro ...	PINHEIR...	2018-10-06 13:25:00.000	2018-10-07 00:47:00.000	2018	65,04477...
3	TM218709	2018040030183	Florestal	Bragança	Macedo d...	Lagoa	LAGOA (...)	2018-10-02 11:51:00.000	2018-10-03 00:00:00.000	2018	77,94609...
4	TM218694	2018040029608	Florestal	Bragança	Vimioso	Argozelo	ARGOZEL...	2018-09-27 15:33:00.000	2018-09-28 02:15:00.000	2018	17,74186...
5	TM218680	2018040029068	Florestal	Bragança	Alfândeg...	Valverde	VALVERD...	2018-09-23 20:11:00.000	2018-09-25 09:30:00.000	2018	13,57645...
6	TM218662	2018040028387	Florestal	Bragança	Mirandela	Cedães	CEDAES ...	2018-09-17 15:57:00.000	2018-09-18 21:45:00.000	2018	21,4775938
7	TM21866	2018040004237	Florestal	Bragança	Vimioso	Caçarelhos	CACAREL...	2018-02-10 14:16:00.000	2018-02-10 17:35:00.000	2018	13,72487...

Figura 5: Exemplo de organização da tabela de atributos do cdg sobre Áreas Arditas (ICNF)

Espera-se da DGT:

1. Apoio na organização e criação das tabelas necessárias
2. Em particular, definição precisa e acesso à tabela VALIDACAO
3. Acesso aos dados do parcelário para o território nacional (pode estar dividido por UP ou outra estratificação)
4. COSSimRef2018. É preciso ter acesso à versão mais recente do tif completo para Portugal Continental (idem) caso não tenha sido ainda disponibilizado

6. Detecção de alterações: estado da arte

Dentro do contexto de modelos de monitorização, uma abordagem eficiente e pouco explorada consiste na monitorização da ocupação do solo através da atualização de uma cartografia de referência, classificando exclusivamente áreas onde é detectada alteração na ocupação (Goméz *et al.* 2016; Jin *et al.*, 2017; Pouliot *et al.*, 2014). A deteção de alterações tem sido realizada com métodos baseados na informação espectral, como diferenciação de imagens e thresholding, utilizando bandas ou índices como Normalized Difference Vegetation Index (NDVI) e Normalized Burn Ratio (NBR) (Jin *et al.*, 2013; Zhu, 2017). Em alternativa, podem ser usadas técnicas estatísticas auto adaptativas como a aplicação de Welch's t-test em séries temporais (Campagnolo *et al.*, 2019; Aubard *et al.*, 2020).

Abordagens alternativas propõem caracterizar séries temporais de um pixel com modelos estatísticos que permitem identificar alterações através de divergências relativamente ao padrão espectro-temporal de referência (Brown *et al.*, 2020).

Algumas técnicas de detecção de alterações aproximam a série temporal do sinal por uma linha poligonal por forma a identificar alterações de tendência, como por exemplo os algoritmos Vegetation Change Tracker (VCT) (Huang *et al.*, 2009), LandTrendr (Jin *et al.*, 2017) e Verdet (Hughes et al 2017) . Estes algoritmos tendem a ser aplicados a séries temporais longas, que são pré-processadas por forma a obter apenas uma observação por ano. Uma limitação dessas técnicas é não permitir modelar o comportamento intra anual das séries.

Em contrapartida, foram desenvolvidos outros algoritmos que decompõem o sinal em três termos: uma tendência, um termo periódico e ruído, o que possibilita incorporar o padrão de culturas anuais e outras ocupações do solo com variações bruscas em cada ano mas com um padrão

razoavelmente bem definido. Essas abordagens parecem ser as mais promissoras para o problema COSSim. Entre os algoritmos disponíveis, destacam-se:

1. BFAST Monitor (Verbesselt *et al.*, 2012) e suas adaptações (Lin *et al.*, 2020)
2. CCDC ou Continuous Change Detection (Zhu e Woodcock, 2014) <<Implements the Continuous Change Detection and Classification temporal breakpoint algorithm. This algorithm finds temporal breakpoints in an image collection by iteratively fitting harmonic functions to the data. Fit coefficients are produced for all input bands, but the bands used for breakpoint detection can be specified with the 'breakpointBands' argument.>>
3. EWMACD ou Exponentially Weighted Moving Average Change Detection (Brooks *et al*, 2014) <<This algorithm computes a harmonic model for the 'training' portion of the input data and subtracts that from the original results. The residuals are then subjected to Shewhart X-bar charts and an exponentially weighted moving average. Disturbed pixels are indicated when the charts signal a deviation from the given control limits.>>

A USGS desenvolveu com base no algoritmo CCDC o projecto Land Change Monitoring, Assessment, and Projection (LCMAP) para caracterizar anualmente a ocupação de solo e as alterações que ocorrem para todos o território dos EUA, com base no repositório de imagens U.S. Landsat Analysis Ready Data (ARD) com imagens desde 1984 (Brown *et al.*, 2020).

No entanto, não existem resultados de comparação desses algoritmos aplicados a imagens S2 e no quadro das classes de ocupação do solo da COSSim. Será necessário avaliar, adicionalmente, as bandas e índices espectrais mais adequados na realização da deteção, a exemplo do NDVI, NBR, ou Change Vector (Zhu, 2017).

7. Estratégias para o problema COSSim

7.1 Deteção de alterações

A deteção de alterações representa uma etapa essencial para a monitorização da ocupação do solo no âmbito do SMOS, uma vez que a abordagem proposta por esse sistema consiste em atualizar a cartografia através da classificação automática apenas em áreas onde são detetadas alterações na ocupação, mantendo a classificação do ano anterior em áreas onde não são detetadas alterações. Dessa forma, elimina-se a necessidade de efetuar uma nova classificação para todo o território a cada ano, portanto tornando o sistema mais eficiente. (Zhu *et al* 2020) obtêm valores de erros de omissão e de comissão ambos inferiores a 30% para deteção de alterações em 16 classes de ocupação do solo com imagens Landsat, com o algoritmo COLD, que é uma adaptação de CCDC para analisar “land disturbance”.

Para além da segmentação temporal do sinal realizada pelos algoritmos descritos acima, é importante considerar igualmente a segmentação espacial das deteções. Em conjunto, o objectivo é fazer uma segmentação espaco-temporal das deteções por forma a encontrar manchas espaciais (“patches”) de pixels que apresentam evidência de alteração em redor da mesma data. Uma proposta baseada em grafos para realizar a segmentação espaco-temporal de sinal MODIS no contexto de perdas de vegetação por fogos está descrita em (Campagnolo *et al* 2019). A mesma

metodologia foi usada com bons resultados em (Campagnolo et al 2021) para deteção de áreas ardidas no Cerrado brasileiro. O algoritmo atrás referido Verdet (Hughes et al 2017) é uma ferramenta que usa técnicas mais simples mas com o mesmo objetivo.

Outra estratégia que parece ser promissora é comparar a evolução temporal do sinal da mancha candidata a alteração com o comportamento de manchas vizinhas com o mesmo tipo de uso do solo tal como foi explorado em (Aubard et al 2000). Nesse artigo, aplicado à monitorização da limpeza de faixas de gestão de combustíveis (rede primária), foi mostrado que esse tipo de abordagem produz resultados com qualidade igual ou superior à classificação baseada em exemplos (supervisionada), permitindo desenvolver uma metodologia totalmente automática de detecção de alterações de ocupação do solo no contexto do problema tratado, em que existe uma delimitação a priori da região analisada (faixas).

Dada a dimensão dos dados (890 milhões de pixels de 10 m para Portugal Continental, e muitos dados auxiliares de input), é fundamental que os algoritmos de deteção de alterações sejam eficientes e paralelizáveis. Os algoritmos CCDC e EWMACD referidos atrás estão disponíveis e podem ser executados em Earth Engine ou Colab (usando o package ee) nos servidores da Google. Isso garante que são de facto algoritmos paralelizáveis que poderão no futuro ser executados em plataformas locais, na DGT ou em plataformas HPC portuguesas. Embora as experiências e testes a realizar no contexto do presente projecto de colaboração usem em grande medida os servidores da Google, o resultado será transponível para outras plataformas que suportem processamento em paralelo com enormes ganhos de tempo de processamento em comparação com algoritmos sequenciais.

7.2 Classificação

A deteção de alterações não é suficiente para o problema COSSim dado que é necessário afectar as alterações identificadas às classes da legenda de ocupação do solo.

No entanto, o output da detecção de alterações pode ser directamente usado para o problema de classificação. Por exemplo, o algoritmo CCDC (Zhu e Woodcock, 2014) inclui dois passos: o passo de detecção de alterações já referido acima, e um passo posterior que usa Random Forests para afectar cada observação, já caracterizada por um modelo com tendência e periodicidade, nomeadamente os coeficientes do modelo, a uma classe de ocupação do solo. Os resultados de (Zhu e Woodcock, 2014) sobre uma área que exibe baixa variabilidade espectral (New England) são muito bons, com OA de 90% e erros de comissão e omissão sempre inferiores a 20% para todas as classes consideradas. No entanto, estudos mais recentes aplicados a uma maior variedade de ocupações do solo e regiões nos EUA (Brown et al 2020) mostram que os valores de OA variam entre 73% e 87% para grandes regiões dos EUA para uma versão melhorada (Version 1) do algoritmo CCDC.

Para um problema semelhante ao COSSim, mas considerando uma legenda com apenas 7 classes e séries de imagens Landsat, (Li et al 2021) usam uma técnica (APL) para gerar muitos exemplos sintéticos. Com base nessas séries sintetizadas que procuram representar a variabilidade nos padrões associados às classes da legenda, (Li et al 2021) usam simplesmente uma medida de distância (twDTW) entre observações indeterminadas e os padrões de referência de observações com ocupação do solo definida, por forma a associar cada observação a uma classe. A metodologia também recorre a CCDC para detectar alterações, por forma a classificar as alterações nas classes

da legenda. Estes autores obtêm uma precisão global de 81% embora algumas classes como “croplands” ou “barren” possam ter erros estimados superiores a 50%.

A estratégia para identificação da ocupação a partir de séries temporais de imagens multiespectrais (e.g. S2) consiste portanto essencialmente na identificação de um padrão temporal associado a um tipo de ocupação do solo, possivelmente espacialmente consistente, incluindo a variabilidade desse padrão, e numa afectação de cada pixel ao tipo de ocupação do solo mais semelhante.

7.3 Transições válidas

Esse processo pode ser naturalmente estendido a padrões que resultam de uma sucessão temporal de tipos de ocupação do solo. A seguinte [Figura 6](#), retirada de (Zhu e Woodcock, 2014) ilustra o sinal espectral da transição entre tipos de ocupação do solo e é reveladora de alguns aspectos importantes:

1. Apenas é possível identificar a partir unicamente do sinal espectral, e com algum grau de certeza, o tipo de ocupação anterior e posterior se houver dados durante um intervalo de tempo suficientemente longo (dois anos no exemplo, antes da alteração e dois anos após a alteração). Isto significa que mesmo que possa ser identificada uma alteração em tempo quase real (um ou dois meses), a identificação da nova ocupação poderá não ser possível ou terá que ser baseada em dados auxiliares (por exemplo, uma detecção de alteração correspondente a uma área ardida pode ser completada por informação de fogos activos).
2. A escolha de bandas espectrais e índices é determinante para a identificação de transições entre determinados tipos de ocupação.
3. Na figura são indicados 5 tipos de ocupação do solo. Das 20 transições possíveis com alteração de ocupação do solo, a figura apenas ilustra 4 transições, que são as transições consideradas mais comuns. Esta redução do número de transições consideradas válidas ou relevantes é fundamental para que o problema COSSim se torne tratável.

O último ponto (3) discutido acima pode ser visto como o resultado da aplicação de “regras”, que excluem do universo de padrões temporais sob consideração, padrões não válidos ou pouco relevantes para o problema COSSim. Essas regras também podem ser descritas numa matriz de transição, com a indicação da verossimilhança da transição para cada par de ocupações do solo, e em função possivelmente de outros dados auxiliares (ver informação na [Secção sobre os dados de input](#) do problema COSSim). Caso a verossimilhança seja baixa, a transição pode ser descartada.

Espera-se da DGT:

1. Determinação de exemplos (localizações geográficas e identificação de ocupação do solo respectiva) onde possam ser recolhidos dados de treino, por forma a caracterizar o padrão espetro temporal das classes da legenda e das transições consideradas relevantes
2. Colaboração na identificação das transições válidas ou relevantes, em função dos requisitos dos utilizadores do produtos COSSim.

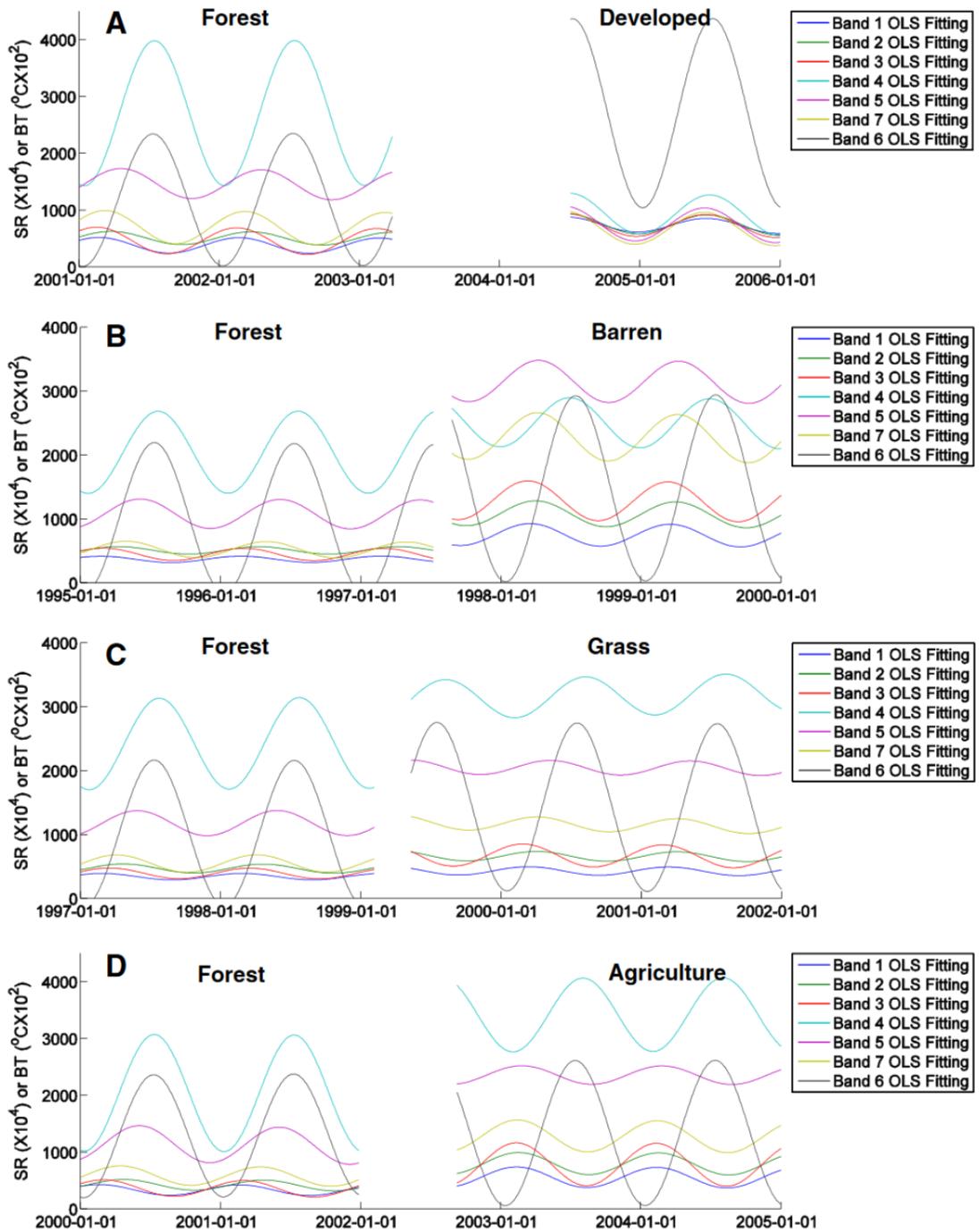


Fig. 8. Examples of the estimated time series models for all seven Landsat bands for the four most common land cover change in the study area.

Figura 6: reprodução da Figura 8 em (Zhu e Woodcock, 2014) que ilustra o padrão temporal esperado para cada banda Landsat 5 TM antes e depois da alteração de coberto.

8. Validação de resultados

Existem duas formas, possivelmente complementares, de validar os resultados:

1. Através de exemplos obtidos por observações de ortofotos e imagens de alta resolução, visitas ao campo, ou outros meios de confirmar a ocorrência de uma alteração da ocupação do solo. Nesse caso é possível comparar pelas técnicas habituais (baseadas na matriz de erro) a realidade com o resultado do algoritmo e proceder à validação dos resultados.
2. Usar situações controladas e gerar dados de input com ruído para quantificar a robustez dos algoritmos. A perturbação nos dados deverá ser calibrada a partir das características das imagens S2, como a presença de nuvens, sombras, e geometria de observação e de iluminação, assim como das características espectrais do coberto e em particular a variabilidade do sinal para cada uso do solo.

Espera-se da DGT:

1. Apoio na compilação dos dados de validação, em particular na elaboração e disponibilização da tabela VALIDACAO

9. Exemplos de detecção com CCDC

O algoritmo CCDC disponível no GEE analisa uma série temporal de imagens de Satélite no intuito de identificar quebras na variação harmónica do sinal espectral, seja ele a informação contida em apenas uma banda da imagem de satélite ou em algum índice composto por diversas bandas. O resultado obtido pelo algoritmo é uma imagem contendo uma série de informações para cada célula, são elas: a data onde foram detetadas quebras na variação harmónica, as datas de início e fim do segmento, a probabilidade de haver uma alteração e os coeficientes, a magnitude e o *root-mean-square error* para cada banda ou índice indicados na série temporal.

Este *output* do algoritmo é importante pois permite a construção de um banco de dados detalhado que pode ser manipulado em outros ambientes que não o GEE, fato que possibilita um maior acesso a informação da área de estudo almejada por quem não está habituado a este ambiente específico.

A princípio optamos por criar um banco de dados onde cada quebra detectada em uma célula corresponde à uma linha da tabela de análise. Nesta tabela, além das informações disponibilizadas no *output* do algoritmo, adicionamos as coordenadas do centróide da célula para manter a informação espacial. Também é possível adicionar novas informações que utilizem os dados existentes ou para complementar a informação existente. Na figura abaixo são apresentadas algumas linhas do banco de dados gerado.

No exemplo apresentado foram omitidas as colunas que indicam o início e o fim de cada segmento para melhorar a visualização dos dados. Aqui utilizamos os dados apenas do NDVI, porém caso se

queira utilizar outros índices ou mesmo bandas únicas serão criadas colunas informando os coeficientes, a magnitude e o *rmse* de cada uma das informações.

changeProb		ndvi_coefs	ndvi_magnitude	ndvi_rmse	tBreak	latitude	longitude	T_DATA
0	1	[0.723693351769531, 0, 0, 0, 0, 0, 0]		-0.697405	0.222545	1585567807406	40.165698	-7.317991 2020-03-30 11:30:07.406
0	0	[0.6422984209300222, 0, 0, 0, 0, 0, 0]		0	0.180194	0	40.165698	-7.317991 NaT
1	1	[0.723693351769531, 0, 0, 0, 0, 0, 0]		-0.697405	0.222545	1585567807406	40.165698	-7.317901 2020-03-30 11:30:07.406
1	0	[0.6422984209300222, 0, 0, 0, 0, 0, 0]		0	0.180194	0	40.165698	-7.317901 NaT
2	1	[0.6886840814536191, 0, 0, 0, 0, 0, 0]		-0.665743	0.234335	1585567807406	40.165698	-7.317811 2020-03-30 11:30:07.406

Figura 7: Estruturação do banco de dados do CCDC

Os valores do ID, na primeira coluna da esquerda, são replicados de acordo com o número de informações disponibilizadas na imagem de *output* para cada célula. Esta informação pode ser entendida como uma série de segmentos de reta com início e fim, caso haja uma quebra detetada novos segmentos de reta são criados . Na figura acima, a primeira linha da tabela indica uma alteração. O ‘tBreak’ contém a data, em milissegundos desde 1 de janeiro de 1970, em que foi detetada a quebra da variação harmónica. A coluna ‘T_DATA’, adicionada para facilitar a compreensão, traduz o ‘tBreak’ para ano, mês, dia e hora. No caso do NDVI, as magnitudes negativas indicam redução do índice, que podem estar associadas a remoção da vegetação e as magnitudes positivas indicam o aumento do índice, que podem estar associadas ao aumento da vegetação. A segunda linha, corresponde a um segundo segmento de reta, que foi criado pois foi identificada uma quebra, porém neste segundo segmento não foram identificadas quebras e isto é identificado pelo valor 0 do ‘tBreak’, é importante notar que a segunda linha apresenta os mesmos valores de ID, latitude e longitude, mas com alguns valores 0 visto que não houveram quebras. Desta forma as linhas onde não foram detetadas quebras podem ser eliminadas do banco de dados. Quanto maior o número de quebras identificadas no mesmo pixel, mais linhas corresponderão ao mesmo ID, porém cada linha contém um nível de informação existente na imagem resultante.

A informação disponibilizada em um banco de dados é eficiente pois com algumas *queries*, que utilizem ou não a informação espacial, é possível identificar o histórico de informações disponíveis para determinado local. Por exemplo, em qual trimestre foi verificado maior número de quebras da variação harmónica, como apresentado na figura abaixo Neste exemplo nota-se que foram detetadas mais alterações no segundo trimestre dos anos de 2019 e 2020.

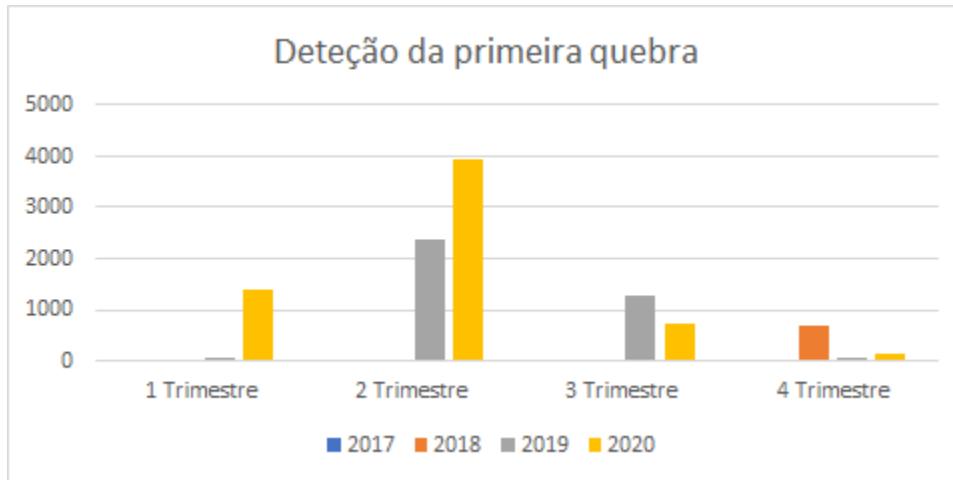


Figura 8: Concentração das alterações por trimestre na área do exemplo 1 abaixo.

Link para acesso via Colab:

<https://code.earthengine.google.com/b969f601c805ef062244a1407b022c3c>

1. Primeiro exemplo: cortes em povoamentos de eucalipto

Com o objectivo de ilustrar o processo de deteção de alterações, procedeu-se à aplicação do algoritmo CCDC em ambiente Google Earth Engine. Nesse caso, trata-se de uma região de floresta de eucalipto, onde houve perda de vegetação devido ao corte da floresta. A Figura 9 apresenta a localização de duas áreas onde houve corte e as suas respectivas séries temporais do NDVI. É possível observar que o corte na Área 1 (-7.3125315318, 40.1705863687) ocorre entre Maio e Junho de 2020, enquanto na Área 2 (-7.3089731226, 40.1744211549) ocorre entre Junho e Julho de 2019.



Figura 9(X): Áreas de corte de eucalipto e respectivos padrões temporais do NDVI.

A aplicação do algoritmo CCDC possibilita a caracterização da série temporal dos pixels em segmentos temporais estáveis, permitindo assim identificar automaticamente o início, fim e o momento em que há quebra no padrão temporal, como explicado atrás. A Figura 10 apresenta um mapa feito com base no atributo “tBreak” do *output* do algoritmo, que representa as datas das quebras dos segmentos temporais. Como explicado em detalhe no início desta secção, o algoritmo CCDC devolve possivelmente mais do que um “tBreak” por pixel. Neste exemplo, apenas se usa o

“tBreak” com a maior magnitude absoluta, descartando-se outros possíveis “tBreaks” por forma a simplificar a apresentação dos resultados. Apesar do CCDC fornecer informação do dia, mês e ano da quebra, o mapa apresentado consiste em uma simplificação, visando ilustrar apenas o ano da quebra. É possível observar que a informação apresentada no mapa acerca do momento em que ocorreram alterações na ocupação do solo devido ao corte da floresta está de acordo com o que se vê nas imagens Sentinel-2 da [Figura 10](#) e nas séries temporais exibidas na [Figura 9](#).

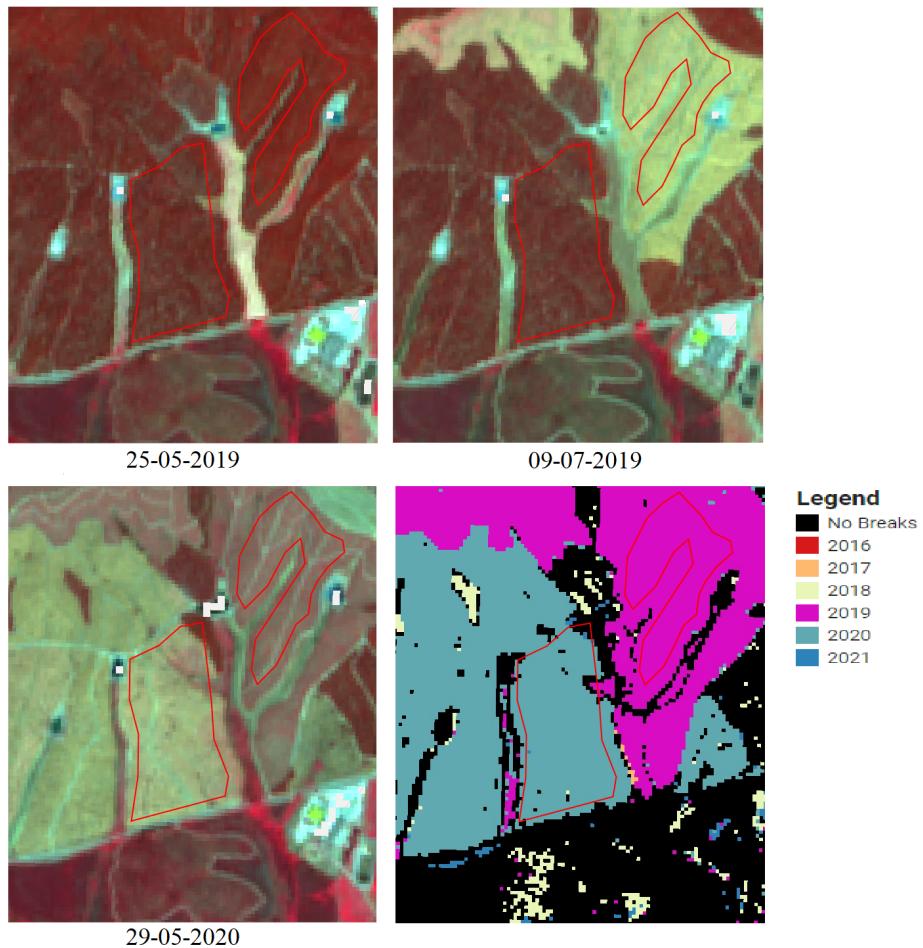


Figura 10: Imagens Sentinel-2 da região de estudo e mapa com data das deteções de quebras nos segmentos temporais, produzido com o output “tBreak” do CCDC (χ^2 Probability=0.95).

2. Segundo exemplo: limpeza de matos em povoamento de pinheiro manso

O segundo exemplo, coordenadas (-7.6678, 37.5686), ilustra as alterações identificadas em uma área que contém pinheiros mansos (Pm), nomeadamente um corte de mato em plantação de Pm. A figura abaixo ilustra o local analisado e o respectivo gráfico da variação média anual do NDVI no interior do polígono. A partir do gráfico pode-se verificar uma sazonalidade natural do sinal espectral que apresenta os picos entre os meses de novembro e março, enquanto os menores valores de NDVI são registrados nos meses de maio a setembro.

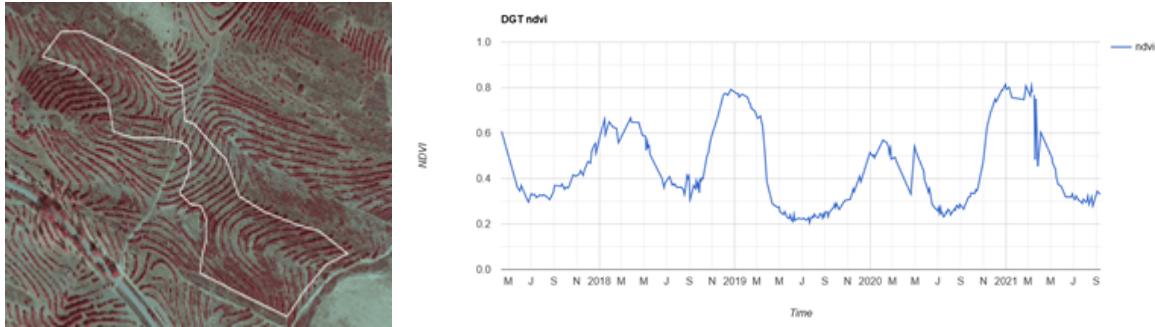


Figura 11: Área de alteração em pinheiro manso e variação média anual do NDVI.

Este exemplo é importante para salientar que o algoritmo CCDC registra as ‘quebras’ da série temporal nos dois sentidos, ou seja, quando indica alterações tanto para a remoção quanto o aumento da vegetação. A figura abaixo apresenta a data das alterações com a maior magnitude na área selecionada. Nota-se que na área de estudo a deteção de alterações deu-se principalmente no ano de 2020.

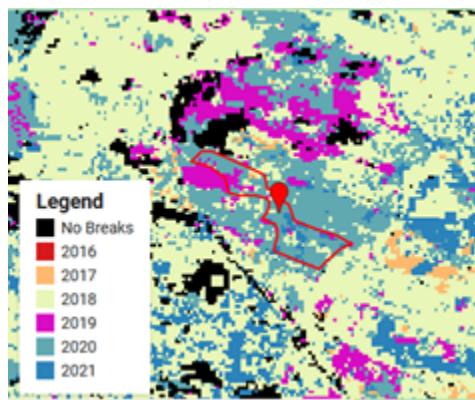


Figura 12: Mapa com data das deteções de quebras de maior magnitude nos segmentos temporais, produzido com o output “tBreak” do CCDC (χ^2 Probablility = 0.95)

Entretanto, no ponto vermelho localizado no centro da área de estudo, o algoritmo foi capaz de identificar duas quebras significativas na série temporal. A figura abaixo apresenta as imagens do satélite Sentinel 2A para as datas próximas das quebras detectadas pelo algoritmo. Na imagem (a) verifica-se a perda da vegetação na área de estudo em meados de abril de 2019, enquanto na imagem (b) é representado o mês de novembro de 2020, quando foi detectado o aumento da vegetação no local. A magnitude associada ao NDVI foi mais expressiva no mês de novembro, desta forma o mapa apresentado anteriormente trás esta informação.



a) Abril de 2019



b) Novembro de 2020

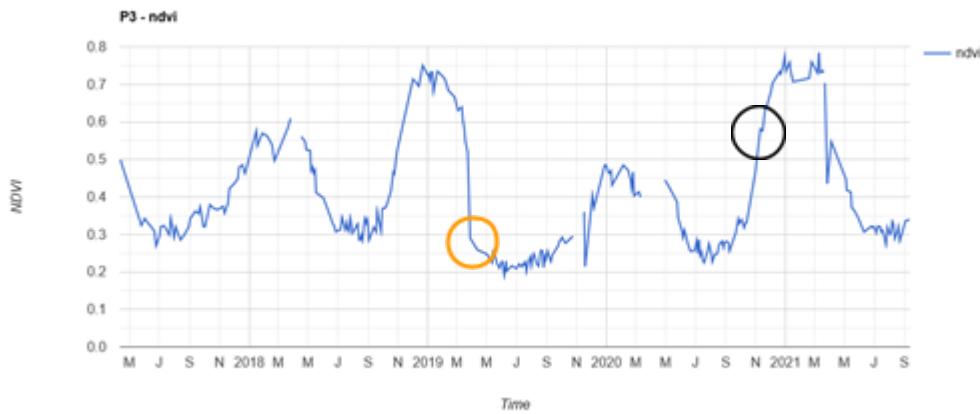


Figura 13: Imagem Sentinel 2A para a) abril de 2020 e b) novembro de 2020 e variação do NDVI no ponto em vermelho.

O gráfico exposto na Figura 13 acima apresenta a variação do NDVI no ponto em vermelho. Nele estão destacados os momentos em que o CCDC detectou as ‘quebras’ identificadas como alteração. O círculo laranja apresenta o momento em que o algoritmo identificou a perda da vegetação (12-04-2019) e o círculo preto apresenta o momento em que o algoritmo identificou ganho de vegetação (10-11-2020).

Link para acesso ao código via GEE:

<https://code.earthengine.google.com/b969f601c805ef062244a1407b022c3c>

3. Terceiro exemplo: transição de agricultura anual para agricultura permanente

Este exemplo ilustra uma situação em que o CCDC apresenta dificuldades em detetar corretamente as quebras na série temporal. Trata-se de regiões onde ocorre transição de agricultura anual para agricultura permanente. Na Área 1 (-8.2930482509, 37.9352483730) da Figura 14 há agricultura permanente de longa data e não se verifica na análise do NDVI nenhuma alteração significativa no padrão espectral ao longo dos anos. Entretanto, o CCDC indica haver uma quebra em Setembro de 2021. Essa situação, que representa um erro de comissão, pode ser explicada pela presença de ruído no final da série temporal, dessa forma dificultando o ajuste da regressão nesse período.

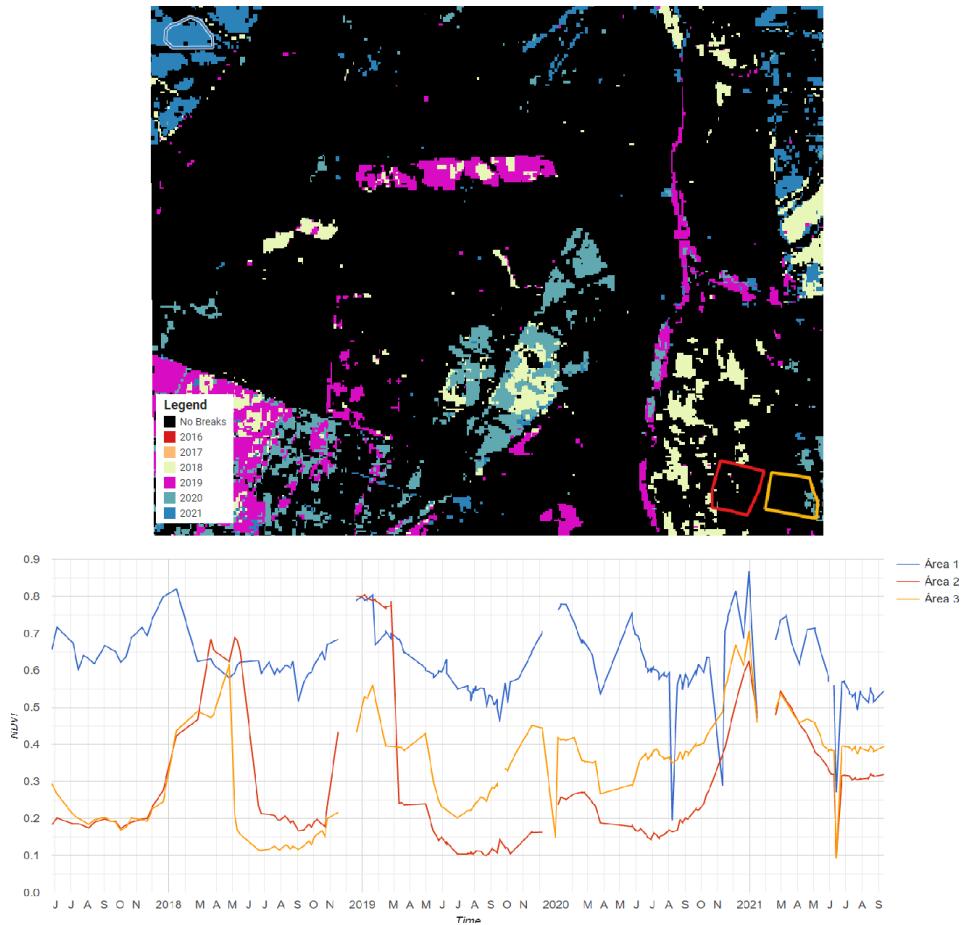


Figura 14: Quebras de maior magnitude identificadas pelo CCDC e padrão espectro-temporal do NDVI em áreas de transição de agricultura anual para permanente. Área 1 (NW) representa cultura permanente de longa data onde não houve transição.

Na Área 2 (-8.2647104798, 37.9168437832) ocorre a transição de agricultura anual para permanente. Análise das imagens Sentinel-2 revela que houve um corte entre Fevereiro e Março de 2019 para a introdução da cultura permanente. No entanto, o algoritmo falha em detetar qualquer quebra na série temporal. Nesse caso, é ilustrada a dificuldade de detetar quebras devido à nova ocupação do solo ainda não apresentar um padrão espectro-temporal definido.

Na Área 3 (-8.2617933810, 37.9165039060) é possível observar a transição de agricultura anual para permanente em 2018, mas que não é identificada pelo CCDC. Nesse caso, apesar de ser possível verificar a mudança no sinal do NDVI no gráfico, o algoritmo não conseguiu identificar o padrão espectral do olival adulto (semelhante ao da Área 1).

Portanto, esse exemplo ilustra as dificuldades encontradas em identificar quebras em que a nova ocupação do solo ainda não possui um padrão espectro-temporal bem definido. Essa situação resulta, inclusivamente, da falta de uma série temporal mais longa que permita identificar tendências tanto a partir do começo da série quanto após uma alteração na ocupação.

10. Referências

Aubard, V., Pereira-Pires, J.E., Campagnolo, M.L., Pereira, J., Mora, A. and Silva, J., “Fully Automated Countrywide Monitoring of Fuel Break Maintenance Operations”, *Remote Sensing*, vol 12, no. 18, pp. 2879, 2020. <https://doi.org/10.3390/rs12182879>

Brooks, E.B., Wynne, R.H., Thomas, V.A., Blinn, C.E. and Coulston, J.W., “On-the-fly massively multitemporal change detection using statistical quality control charts and Landsat data”. *IEEE Transactions on Geoscience and Remote Sensing*, 52(6), pp. 3316-3332, 2014, doi: 10.1109/TGRS.2013.2272545.

Brown, J.F., Tollerud, H.J., Barber, C.P., Zhou, Q., Dwyer, J.L., Vogelmann, J.E., Loveland, T.R., Woodcock, C.E., Stehman, S.V., Zhu, Z., Pengra, B.W., Smith, K., Horton, J.A., Xian, G., Auch, R.F., Sohl, T.L., Sayler, K.L., Gallant, A.L., Zelenak, D., Reker, R.R. and Rover, J., “Lessons learned implementing an operational continuous United States national land change monitoring capability: The Land Change Monitoring, Assessment, and Projection (LCMAP) approach”, *Remote Sensing of Environment*, vol. 238, pp. 111356, 2020, <https://doi.org/10.1016/j.rse.2019.111356>.

M.L. Campagnolo, D. Oom, M. Padilla, J.M.C. Pereira, “A patch-based algorithm for global and daily burned area mapping”, *Remote Sensing of Environment*, Volume 232, 2019, 111288, <https://doi.org/10.1016/j.rse.2019.111288>.

M.L. Campagnolo, R. Libonati, J.A. Rodrigues, J.M.C. Pereira, “A comprehensive characterization of MODIS daily burned area mapping accuracy across fire sizes in tropical savannas”, *Remote Sensing of Environment*, Volume 252, 2021, 112115, <https://doi.org/10.1016/j.rse.2020.112115>.

Gómez, C., White, J.C. and Wulder, M.A., “Optical remotely sensed time series data for land cover classification: A review”, *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 116, pp. 55-72, 2016, <https://doi.org/10.1016/j.isprsjprs.2016.03.008>.

Hughes, M.J., Kaylor, S.D. and Hayes, D.J., “Patch-based forest change detection from Landsat time series”. *Forests*, 8(5), p.166, 2017, <https://doi.org/10.3390/f8050166>.

Jin, S., Yang, L., Danielson, P., Homer, C., Fry, J. and Xian, G., “A comprehensive change detection method for updating the National Land Cover Database to circa 2011”, *Remote Sensing of Environment*, vol. 132, pp. 159-175, 2013, <https://doi.org/10.1016/j.rse.2013.01.012>.

Jin, S., Yang, L., Zhu, Z. and Homer, C., “A land cover change detection and classification protocol for updating Alaska NLCD 2001 to 2011”, *Remote Sensing of Environment*, vol. 195, pp. 44-55, 2017, <https://doi.org/10.1016/j.rse.2017.04.021>.

Li, Congcong , George Xian, Qiang Zhou, Bruce W. Pengra, “A novel automatic phenology learning (APL) method of training sample selection using multiple datasets for time-series land cover mapping”, *Remote Sensing of Environment*, Volume 266, 2021, 112670, <https://doi.org/10.1016/j.rse.2021.112670>.

Lin, Y., Zhang, L., Wang, N., Zhang, X., Cen, Y. and Sun, X., “A change detection method using spatial-temporal-spectral information from Landsat images”, International Journal of Remote Sensing, vol. 41, no. 2, pp. 772-793, 2020, <https://doi.org/10.1080/01431161.2019.1648905>.

Poulit, D., Latifovic, R., Zabeic, N., Guindon, L. and Olthof, I., “Development and assessment of a 250 m spatial resolution MODIS annual land cover time series (2000–2011) for the forest region of Canada derived from change-based updating”, Remote Sensing of Environment, vol. 140, pp. 731-743, 2014, <https://doi.org/10.1016/j.rse.2013.10.004>.

Verbesselt, J., Zeileis, A. and Herold, M., “Near real-time disturbance detection using satellite image time series”, Remote Sensing of Environment, vol. 123, pp. 98-108, 2012, <https://doi.org/10.1016/j.rse.2012.02.022>.

Zhu, Z. and Woodcock, C.E., “Continuous change detection and classification of land cover using all available Landsat data”, Remote sensing of Environment, vol. 144, pp. 152-171, 2014, <https://doi.org/10.1016/j.rse.2014.01.011>.

Zhu, Z., “Change detection using landsat time series: A review of frequencies, preprocessing, algorithms, and applications”. ISPRS Journal of Photogrammetry and Remote Sensing, vol. 130, pp. 370-384, 2017, <https://doi.org/10.1016/j.isprsjprs.2017.06.013>.

Zhe Zhu, Junxue Zhang, Zhiqiang Yang, Amal H. Aljaddani, Warren B. Cohen, Shi Qiu, Congliang Zhou, “Continuous monitoring of land disturbance based on Landsat time series”, Remote Sensing of Environment, Volume 238, 2020, 111116, <https://doi.org/10.1016/j.rse.2019.03.009>.