

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

Manuel Campagnolo, Bruno Barbosa e Daniel Moraes

maio de 2022

<b>1. Objectivo</b>	<b>0</b>
<b>2. Detecção em tempo quase real: últimos resultados com BDR640</b>	<b>1</b>
<b>3. Desenho da nova BD de referência BDR_TNE_300</b>	<b>2</b>
3.1 Amostras com alteração de ocupação	3
3.2 Amostras sem alteração de ocupação	4
<b>4. Algoritmo, resultados provisórios e próximos passos</b>	<b>4</b>

## 1. Objectivo

O objectivo geral 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.

No presente relatório (Tarefa 3) descrevem-se os resultados mais recentes da aplicação do CCDC à BD de referência pontual BRD640 . Descreve-se também o desenho da nova BD de referência BDR\_TNE\_300 e os resultados preliminares obtidos. Faz-se uma descrição sumária do algoritmo para classificação de ocupação do solo e dos próximos passos da análise.

## 2. Detecção em tempo quase real: últimos resultados com BDR640

Para verificar a possibilidade de o CCDC ser utilizado em *real time* estabelecemos a seguinte forma de análise dos dados de validação (640 pontos da DGT):

- Foram gerados modelos do CCDC estabelecendo datas de truncagem do final da série temporal de Sentinel-2 no último dia de cada mês entre 31-07-2019 a 31-04-2021.
- Os modelos foram avaliados a partir da data da alteração do analista Z. Nosso primeiro teste foi feito levando em consideração modelos de depois de 2 meses da alteração identificada, por exemplo, se o analista identifica a alteração no dia 15-08-2019 utilizamos o modelo com a truncagem no dia 31-10-2019 para validação e assim sucessivamente. Isto nos permite simular uma situação em tempo quase real para verificar se o modelo é capaz de identificar alterações recentes. Adicionalmente, verificamos diversas datas de diferença entre a data do analista para a data da truncagem do modelo: após 1, 2, 3, 4, 5 e 6 meses.
- Para identificar um ponto ‘verdadeiro positivo - vp’ são necessárias as seguintes condições:
  - a. O break identificado pelo CCDC deve estar entre +/- 30 dias de diferença para a data do Analista Z.
  - b. O break identificado pelo CCDC deve estar no modelo com mais  $n$  meses após a data do Analista Z (1, 2, 3, 4, 5 ou 6).

O resultado desta análise está na Figura 1. O CCDC mostra-se robusto para identificar os breaks na série temporal, seja com 1 mês após o evento ou 6 meses. Entretanto, cabe uma ressalva, como já havíamos mencionado no entregável 2, de acordo com o final da série de truncagem, alguns pontos podem deixar de ser verdadeiros positivos de um mês para o outro, mas isso ocorreu raramente no nosso universo de dados disponível.

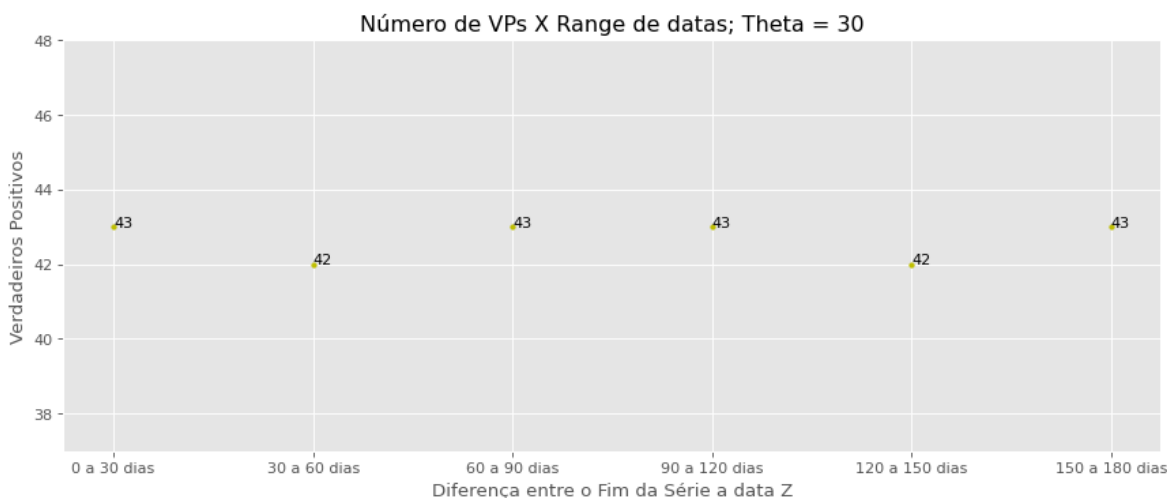


Figura : Identificação de alterações com  $n$  meses de dados Sentinel-2 após a data da alteração identificada pelo analista.

### 3. Desenho da nova BD de referência

#### BDR\_TNE\_300

A base de dados de referência anterior (BDR640) possui um número reduzido de unidades amostrais (UA) que apresentavam alteração na ocupação e refere-se a um período de análise limitado a aproximadamente 1 ano. Por isso, procedeu-se à criação de uma nova base de dados de referência que atendessem de forma mais adequada às necessidades deste projecto no âmbito da validação dos resultados da detecção de alterações.

A construção da nova base de dados de referência (BDR\_TNE\_300) visou identificar 300 UAs localizadas dentro da área do tile 29TNE a serem encaminhadas à DGT para preenchimento da informação de referência por meio de fotointerpretação. A construção foi feita de forma que metade da amostra correspondesse a exemplos em que há potencial alteração na ocupação e a outra metade a exemplos em que potencialmente não há alteração. Das 150 UAs correspondentes a potenciais alterações, 100 referem-se a potenciais alterações identificadas no período compreendido entre os anos agrícolas de 2018 a 2020 e 50 UAs a potenciais alterações durante o ano agrícola de 2021 (Tabela 1).

	Período de referência	Nº de unidades amostrais
Potencial alteração na ocupação	2018 - 2020	100
	2021	50
Sem alteração	2018 - 2021	150

Tabela 1: Distribuição das unidades amostrais e respectivos períodos de referência.

### 3.1 Amostras com alteração de ocupação

A nova base de dados de referência foi desenhada de forma a incluir unidades amostrais correspondentes a alterações na ocupação. Considerou-se que uma unidade amostral corresponde a uma alteração na ocupação quando há evidência de alteração, obtida através da comparação de produtos cartográficos e/ou do resultado da aplicação do CCDC.

As unidades amostrais que referem-se a potenciais alterações no período entre os anos agrícolas de 2018 a 2020 foram geradas a partir de uma combinação de processos. Em primeiro lugar, foram identificados os pixels em que houve perda de vegetação no período, com auxílio da COSsim de 2018 e 2020. Foram consideradas apenas as classes de ocupação do solo (Eucalipto, Pinheiro Bravo e Matos) e transições (para Matos, Vegetação Herbácea ou Superfície sem Vegetação) de interesse a este projecto, conforme ilustrado na Figura 2. A seguir, foram eliminados os grupos de pixels contíguos com área inferior a 0.5 ha, havendo em sequência um processo de erosão para a remoção de pixels na bordadura dos grupos. Além disso, foi computado o resultado do CCDC para os pixels restantes. Pixels em que o CCDC não indicou alteração na ocupação no período e pixels com magnitude da quebra positiva (i.e. indicando haver ganho de vegetação) foram excluídos. Dentre os pixels restantes foram, então, recolhidas 100 UAs através de uma amostragem aleatória estratificada por faixas de magnitude da quebra referente ao NDVI. Os estratos são apresentados na Tabela 2 abaixo.

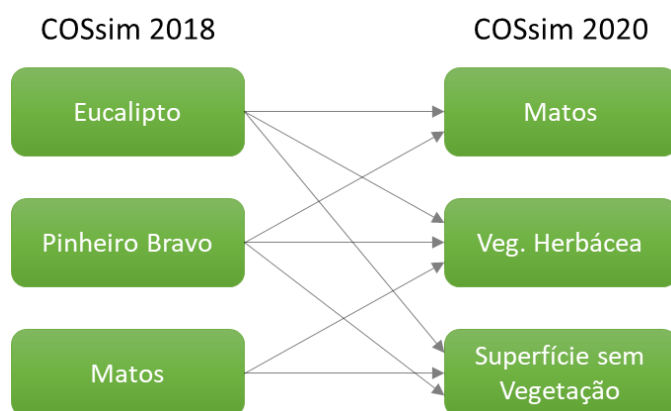


Figura 2: Classes de ocupação e transições consideradas na amostragem referente a 2018-2020.

Estrato	Magnitude da quebra ( $10000 \cdot \text{NDVI}$ )
1	$(-10000, -6000]$
2	$(-6000, -4000]$
3	$(-4000, -2000]$

4	(-2000, 0]
---	------------

Tabela 2: Estratos utilizados na amostragem estratificada.

Relativamente às unidades amostrais referentes a potenciais alterações durante o ano agrícola de 2021, o processo de amostragem foi distinto uma vez que não havia uma COSSim disponível para o ano em questão. Portanto, foram identificados na COSSim 2020 pixels das classes de ocupação de interesse (Eucalipto, Pinheiro Bravo e Matos) e foram eliminados os pixels em que o CCDC não indicou alteração ou indicou alteração com magnitude da quebra positiva no ano agrícola de 2021. Foram implementadas também as etapas de remoção de grupos de pixels contíguos com área inferior a 0,5 ha e erosão das bordaduras. Ao final, foram selecionadas UAs através de amostragem aleatória estratificada, utilizando o estratos da Tabela 2.

### 3.2 Amostras sem alteração de ocupação

As unidades amostrais consideradas sem alteração foram obtidas considerando os pixels em que a classe da COSSim de 2018 e 2020 se mantiveram iguais. Além disso, foram excluídos pixels em que o CCDC detectou alteração no período do ano agrícola de 2021. De forma semelhante às amostragens anteriores, foram implementados os processo de remoção de grupos de pixels com área inferior a 0,5 ha e erosão da bordadura. Ao final, foram selecionadas 150 UAs através de uma amostragem aleatória.

## 4. Algoritmo, resultados provisórios e próximos passos

O algoritmo de identificação de alterações e classificação está a ser escrito em python, de forma modular, em ambiente Google Colab.

O algoritmo faz a leitura da BD de referência e da colecção de imagens Sentinel 2/SR para a região de estudo e constrói uma tabela que inclui todos os eventos e todos os tBreaks identificados por CCDC e os coeficientes associados.

Logo que a BDR\_TNE\_300 estiver totalmente disponível, essa tabela completa para os 300 buffers de referência será gerada. Cada linha corresponde a um tBreak ou um pixel em que não é identificado nenhum tBreak. A tabela deverá ter por volta de 350000 linhas, uma parte das quais correspondem a eventos (alterações) com identificação da ocupação pré e pós evento.

Esse conjunto de dados permitirá então fazer o treino de classificadores usando como variáveis predictoras as classes existentes antes do evento e os parâmetros de CCDC. Esses resultados serão também usados para avaliar a aplicação de regras em classificação. Finalmente, os resultados poderão ser validados.

Dado que o BDR\_TNE\_300 é constituído por polígonos é possível analisar a qualidade espacial (e não apenas pontual) dos resultados do algoritmo. A Figura 3 mostra que, para esse caso particular,

se obtém um bom ajustamento espacial entre as regiões identificadas pelo CCDC e os polígonos de alteração em BDR\_TNE\_300.

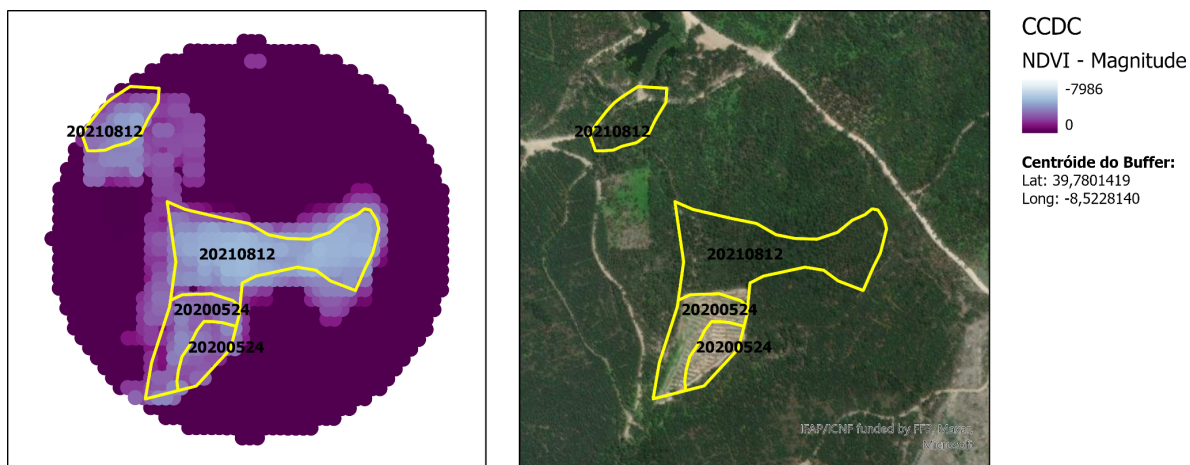


Figura 3: Um “buffer” de BDR\_TNE\_300 com 200 m de raio e o resultado de aplicação de CCDC. Do lado esquerdo está o resultado do CCDC (magnitude de tBreak) e do lado direito estão os polígonos indicando as localizações onde ocorreram alterações segundo os analistas.