myFCT

Call Advanced Computing Projects (4th ed) - A0 Experimental Access (round D)

2024.07034.CPCA

Candidate

Manuel Campagnolo

Submission Date

31.05.2024 - 10:11

Overview

Ri and co-RI details

RESPONSIBLE INVESTIGATOR

Name	Manuel Lameiras de Figueiredo Campagnolo
E-mail	mlc@isa.ulisboa.pt
Gender	M
Birth Date	08-11-1965
Nationality	Portugal
Identification Document	Citizen Card
Identification Number	09943532
ORCID ID	0000-0002-9634-3061

RI INSTITUTION

Country	Portugal
Institution	Universidade de Lisboa (UL) - Instituto Superior de Agronomia - Centro de Estudos Florestais

Application details

GENERAL DESCRIPTION

Title

Monitoring vegetation losses over Portugal from Sentinel-2 imagery

Abstract

Monitoring land cover changes in Portugal in near real-time is crucial for effective planning and prevention public policies. This project focuses on identifying vegetation losses using satellite imagery. In Portugal, vegetation losses mainly result from forest fires and clear cuts but can also be due to urban expansion, agricultural development, and the recent growth of solar farms, among other causes.

The Directorate General for Territorial Development (DGT) is the national public agency responsible for land use and urban development policies. DGT provides multi-annual land use maps (COS) with a minimum unit area of 1 hectare and annual land cover maps (COSc) derived from ESA's Sentinel-2 imagery and other sources. However, the annual COSc products do not focus on change detection. For finer temporal analysis, particularly for vegetation losses, the complete collection of Sentinel-2 images needs to be explored, which offers a best-case temporal resolution of 5 days for clear images.

DGT and ISA, under cooperation contract 3044/2023, are collaborating to develop a continuous time vegetation loss product with a minimum unit area of 0.5 hectares. Towards that end, they have produced reference data sets to train and validate automatic change detection techniques.

The Continuous Change Detection (CCD) algorithm has been used in the context of the collaboration between ISA and DGT to estimate regular land cover patterns and identify anomalies indicating changes using the full Sentinel-2 archive. CCD has been tested for:
1. Optimal algorithm parameters;
2. Change detection accuracy;
3. Computational cost.

The accuracy results are promising, with errors below 20%. However, the computational resources required to apply the algorithm over Portugal are significant. The project aims to leverage advanced computing platforms to streamline the production of continuous time vegetation loss maps and reduce processing time.

Project scope

The problem of change detection from Sentinel-2 imagery has been explored by the team from ISA and DGT which includes PhD student Daniel Moraes at IMS/NOVA and Masters student Inês Silveira at ISA/ULisboa.

The research so far has been focused in establishing a first reference data set for Sentinel-2 tile 29TNE over the Center-West part of Portugal (Alves et al 2023, and Moraes et al 2024) and exploring Sentinel-2 bands and indices for change detection and classification (Alves, 2023). Several tests have been performed to determine the best set of parameters for the CCD algorithm over that same region and to obtain accuracy estimates for the CCD method (Moraes et al 2024). More precisely, that paper uses a reference data set with 300 photo-interpreted buffers to compare the results from CD and the actual changes identified by trained analysts from DGT.

References:

(Alves et al 2023) Alves, A.; Moraes, D.; Barbosa, B.; Costa, H.; Moreira, F.; Benevides, P.; Caetano, M. and Campagnolo, M. (2023). Exploring Spectral Data, Change Detection Information and Trajectories for Land Cover Monitoring over a Fire-Prone Area of Portugal. In Proceedings of the 9th International Conference on Geographical Information Systems Theory, Applications and Management - GISTAM, ISBN 978-989-758-649-1; ISSN 2184-500X, SciTePress, pages 87-97. DOI: 10.5220/0011993100003473

(Moraes et al, 2024) Moraes D., Barbosa B., Costa H., Moreira F.D., Benevides P., Caetano M., Campagnolo M. Continuous forest loss monitoring in a dynamic landscape of Central Portugal with Sentinel-2 data, (2024), International Journal of Applied Earth Observation and Geoinformation, 130,
DOI: 10.1016/j.jag.2024.103913

Graduate students:

PhD dissertation, advisors: Mário Caetano and Manuel Campagnolo
Daniel Moraes, with a FCT grant; Remote Sensing and Data Science; NOVA/IMS; Started in 2021; “Exploring change detection, classification variables and training data for satellite imagery-based continuous land cover monitoring”

Masters dissertation: adviser Manuel Campagnolo
Inês Silveira, Masters in Data Science at ISA/ULisboa; Started in 2023; "A reference data set for vegetation loss detection over Portugal"

Keywords

Land cover monitoring | remote sensing | change detection | time series

Domain

Natural Sciences

Area

Earth and Related Environmental Sciences

Subarea

Monitoring and Environmental Impact

COMPUTATIONAL WORK PLAN

Computational Tasks, milestones and deliverables

Tasks:

T1 (months 1 and 2): Upload input data and setup scripts to read, process and output estimated vegetation loss dates for a region of interest. The input data will be in a format that minimizes the access time. Each row of data corresponds to a time series of reflectance values for one pixel and one spectral band from the Sentinel-2/MSI sensor. Only required bands and cloud-free pixels will be included in the input. For a full tile, and before screening for clouds, the input data consists of several binary files in the Numpy 'npy' format, with a total size of approximately 0.8 Tb. The output for each pixel consists of a list of dates, and harmonic coefficients associated to each interval between dates.

T2 (months 2-5): Estimate processing time on the HPC platform per km2 and year. These results will guide the team into selecting candidate areas in Continental Portugal where the change detection algorithm will be applied. Lower estimated processing time will allow to expand the areas and analyze pixels with low probability of change. Higher estimated times will indicate that pixels with low prior probability of change should be discarded from processing.

Task 3 (months 4-6). Apply the algorithm to Continental Portugal to obtain a complete map of estimated dates of vegetation loss.

Milestones:

M1 (month 2): Upload input data and setup working environment

M2 (month 5): Estimate algorithm processing time on HPC platform

M3 (month 6): Create map of detection dates when vegetation loss occurred

Deliverables

D1 (month 3): report and video showing how to setup the working environment for the problem at hand. The report and video will be made available to the ISA and DGT teams that are collaborating towards the goal of creating near real time maps of vegetation loss for Portugal, but will also be made available to the public in general.

D2 (month 5): Report describing the estimates of processing time for candidate areas in Continental Portugal. In particular, the estimates will be stratified by regions, which vary according to data availability due to cloud cover.

D3 (month 6): Map with vegetation loss detection dates

TEAM

Team Members

Name	Institution	Task
Daniel Moraes (CO-RI) ORCID 0000-0002-4568-8182	Instituto Superior de Estatística e Gestão de Informação - NOVA Information Management School (NOVA IMS)	Task 1, task 2, task 3
Sara Caetano	Instituto Superior de Agronomia	Task 1, task 2, task 3
Inês Silveira	Instituto Superior de Agronomia	Task 1
Dominic Welsh	Instituto Superior de Agronomia	Task 2, task 3

PREVIOUS EXPERIENCE

Have you used HPC or HTC or Virtual Research Environment (Scientific Cloud) platforms?

No

RELATED SCIENTIFIC OR INNOVATION PROJECT

Would this submission, in case of being approved, contribute to an existing project?	No

PROJECT DATA AND RESULTS

Dissemination of project results	Sharing HPC Expertise The use of a High-Performance Computing (HPC) platform to tackle the significant issue of land cover change monitoring across Portugal provides a valuable opportunity for disseminating HPC expertise. This initiative involves ISA, Portugal's largest institution for graduate and post-graduate education in Agricultural Sciences, and DGT, the national public agency responsible for land use and urban development policies. To further amplify this knowledge transfer, we will produce a video or a series of videos, possibly to be available at https://educast.fccn.pt/. These videos will detail the steps for preprocessing satellite imagery, uploading it to the HPC platform, setting up the working environment, and running the necessary tests and analyses. Vegetation Loss Maps The final product, maps depicting vegetation loss, is planned to be made freely available by DGT through the "Registo Nacional de Dados Geográficos" (National Geographic Data Register) at https://snig.dgterritorio.gov.pt/. This ensures that valuable data is accessible to researchers, policymakers, and the general public, fostering further research and informed decision-making in land use and environmental conservation.
Data management plan (DMP) delivery?	No

Technical requirements

COMPUTING MODELS

Do you need High Performance Computing (HPC) resources?	Yes
Do you need Scientific Cloud computing or VRE resources?	No

HIGH PERFORMANCE COMPUTING

Preferred HPC platform	Any platform
Number of simultaneous CPU cores needed	128
Number of total clock hours needed	400
Total CPU core.hours needed	51.200
RAM in GB per CPU core	0
Does your project require GPU usage?	No
Application Software - HPC	SoftwareList.xlsx Obtained on *30-05-2024* at *09:41:45*
Software scalability graph	—
Average CPU cores per job - HPC	0
Total quantity of disk, in GB, needed to support HPC runs or storage - HPC	1000

ADDITIONAL TECHNICAL QUESTIONS

Any special requirements for technical support, software or tools	We will need some technical assistance in order to setup the working environment.
Antecipated Internet bandwidth (Mbit/s) needed for the computational project	–
Estimated volume of Internet data to be transferred (GB) per month	–
Antecipated locations for high bandwidth Internet communications	Inside the Portuguese Academic Network (RCTS)
Needs tools for post-processing (visualization)?	No
Additional document	documents_that_support_this_proposal_S2change_pyccd_github.pdf Obtained on *30-05-2024* at *10:14:41*

COMPUTING RESOURCES

Justification of computational resources	Algorithm Input and Scope The PyCCD algorithm processes a 3-band Sentinel-2 time series for individual pixels. To monitor vegetation loss across Continental Portugal, we need to apply the algorithm to all Sentinel-2 pixels where such changes might occur, totaling approximately 120 million pixels. Performance Based on Previous Experiments Previous experiments indicate that the algorithm requires roughly 18 minutes to process a sample of 10,000 pixels (equivalent to 1 km²) over a 5-year time series on a conventional machine with 6 cores. These experiments involved preprocessing the input data to optimize reading efficiency and were conducted over a reference area, with multiple runs to establish confidence intervals for both classification accuracy and computation time. Extrapolation for Full Study Area Extrapolating these results to our study area, which includes 120 million pixels and extends over an 8-year time series, we estimate a computational requirement of approximately 34,560 core-hours, assuming linear scalability with the length of the time series. Additionally, we anticipate a 50% increase in processing time for tests planned in Task 2. Optimizing Computational Resources Given the pixel-by-pixel application of the algorithm, maximizing the number of cores used is preferable. Assuming access to an HPC platform with 128 cores, we estimate the total processing time for the entire territory of Portugal to be around 270 hours. Access to more cores would further reduce the computation time, underscoring the scalability and efficiency benefits of parallel processing for this task.
Usage starting date by	01-09-2024
Usage end date by	01-03-2025
Additional comments	The RI previously applied for and received the Advanced Computing Projects - A0 Access grant 2022.59005.CPCA.A0. The project, titled "Computing the Portuguese Wildland Urban Interface from Land Cover Maps," was initially intended to be supported by a research fellow from a separate project. However, due to administrative issues, this support was not realized within the grant period. For the current proposal, the team is fully assembled, and preliminary research is already in progress.
Considering article 23 of the regulation, do you accept that the following information may be used in publications related to the RNCA?	Yes