APPLICATIONS

6 MC Manuel Campagnolo

Calls Applications

Application successfully submitted.

Call Advanced Computing Projects (5th ed) - A1 Development Access (round A)
2024.10004.CPCA

• **Submitted** **22.10.2024 - 16:01**

Export to PDF

Overview

Go to
- → Ri and co-RI details
- → Application details
- → Technical requirements

Ri and co-RI details

RESPONSIBLE INVESTIGATOR

CiênciaID	7F18-3B3C-06BB
Name	Manuel Lameiras de Figueiredo Campagnolo
E-mail	mlc@isa.ulisboa.pt
Institution	Universidade de Lisboa (UL) - Instituto Superior de Agronomia - Centro de Estudos Florestais
File gotten via CIÊNCIAVITAE	–

Application details

GENERAL DESCRIPTION

Title Monitoring vegetation losses over Portugal from Sentinel-2 imagery with change detection and deep learning

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 s

for Territorial Development (DGT) is the national publ 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 cloud free 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.

Data processing will encompass two major tasks. At the pixel level, we will perform continuous change detection using the CCD algorithm that has been adapted to identify anomalies indicating changes using the full Sentinel-2 archive. At the spatial level, we will apply deep-learning segmentation algorithms to delineate vegetation loss areas following an hybrid approach that combines multi-spectral Sentinel-2 imagery with the outputs of CCD.

This project will build on the results of the on-going project 2024.07034.CPCA.

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 Dominic Welsh 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 an extensive reference data set derived from The Navigator Company records for Continental Portugal (Silveira, 2024), and fine-tuning and testing the PyCCD algorithm for change detection (Moraes et al 2024).

Currently, under project 2024.07034.CPCA, we're in the process of bench-marking the change detection algorithm. The algorithms for the image segmentation part of the project are being developed and will be tested, fine-tuned, and bench-marked in the context this application.

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

(Silveira, 2024) Silveira, Inês, Development of a forest clear-cut reference database for Continental Portugal at Sentinel-2 resolution, Masters dissertation in Data Science, ISA/ULisboa (submitted in October 2024). Adviser: Manuel Campagnolo

Current graduate students:
PhD dissertation, advisors: Mário Caetano (DGT) and Manuel Campagnolo (ISA/ULisboa)
Daniel Morais, 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
Dominic Welsh, Masters in Data Science at ISA/ULisboa; Started in 2024

Keywords change vegetation imagery classification

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 1-3). This tasks will extend the work that is being currently done in the context of project 2024.07034.CPCA. The goal is to extend the application of the change detection algorithm CCD at the pixel level to obtain "near real time" (NRT) updates of change detection temporal breaks. This task will include computation experiments to assess improvements in [illegible] that would correspond to reprocessing the whole time series in NRT.

Application successfully submitted.

T2 (months 1-8). Implement hybrid classification algorithm that combines combines multi-spectral Sentinel-2 imagery with the outputs of CCD to create a vegetation loss map for Continental Portugal for a single date based on the past imagery for Sentinel-2. This is the part of the project that requires GPU hardware since this is a image segmentation problem that can be addressed with deep learning models like U-nets, and its more recent developments, or visual transformers. The hybrid approach has been applied successfully to change detection over Landsat satellite imagery (Zhao et al. 2024) .

T3 (months 7-12) Improve T2 to address NRT classification. The goal is to be able to update the in NRT vegetation loss maps without having to re-train the classifier from scratch. Transfer learning will be used to address this problem.

Milestones:
M1 (month 2). Setup working environment for deep learning algorithms using GPU hardware.
M2 (month 8). Create a single date vegetation loss map for Continental Portugal
M3 (month 12). Create a succession of multi-dates vegetation loss maps for Portugal from updated Sentinel-2 imagery

Deliverables:
D1 (month 3) Report describing the algorithm and bench-marking results for NRT change detection at the pixel level with CCD
D2 (month 8) Report describing bench-marking results for training and fine-tuning for distinct regions of Portugal the hybrid algorithm described in task 2. Vegetation loss map in raster or vector format for a single date.
D3 (month 12) Report describing algorithm and bench-marking for updating in near real time the hybrid algorithm for vegetation loss mapping. Vegetation loss maps for a time interval in raster or vector format.

Reference:
(Zhao et al 2024) Zhao, Lyu, Zhang, Pang, Zhang, A fast hybrid approach for continuous land cover change monitoring and semantic segmentation using satellite time series, International Journal of Applied Earth Observation and Geoinformation, 134, 2024, https://doi.org/10.1016/j.jag.2024.104222.

TEAM

Team Members

Name	Institution	Task	Curriculum Vitae
Daniel Moraes	Instituto Superior de Estatística e Gestão de Informação - NOVA Information Management School (NOVA IMS)	Task 1	–
Sara Caetano	Instituto Superior de Agronomia	Task 1	–
Dominic Welsh	Instituto Superior de Agronomia	Task 1	–

PREVIOUS EXPERIENCE

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

Selected options HPC

Description of previous work and platforms used

We are currently executing A0 project 2024.07034.CPCA. We are able to upload input data and setup the working environment, and we have starting performing computational experiments.

The initial CCD model was set up and ran on INCD's CPUs, processing images from Sentinel-2, specifically the tile T29TNE from 2017 to 2023. This processing of 241,000 pixels took about 85 CPU hours. For the entire territory, it would take 42,500 CPU hours to run the model for the whole country.

In order to make the processing faster, we are exploring using MPI to run our jobs on multiple nodes.

Do you had a CPCA approved project? Yes

Previous CPCA projects reference 2024.07034.CPCA.A0

Types of access in previous CPCA projects A0

Operational centers already used INCD

Final report of previous advanced computing projects –

RELATED SCIENTIFIC OR INNOVATION PROJECT

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

Technical requirements

COMPUTING MODELS

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

Application successfully submitted.

HIGH PERFORMANCE COMPUTING

Preferred HPC platform	Deucalion, GPU partition Cirrus and Stratus (Cloud), operated by INCD Deucalion, any partition Deucalion, ARM partition Deucalion, x86 partition
Number of simultaneous CPU cores needed	96
Number of total clock hours needed	2000
Total CPU core.hours needed	192.000
RAM in GB per CPU node	4
Does your project require GPU usage?	Yes
Number of GPUs needed	1
Number of total clock hours needed	1500
Total GPU.hours needed	1.500
Please state which type of GPUs do you need	Nvidia
Application Software - HPC	SoftwareList 5ºCPCA_V1.xlsx Obtained on *22-10-2024* at *15:52:21*
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	The technical support in the current A0 project is very adequate to our needs.
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	–
Needs tools for post-processing (visualization)?	No
Additional document	–

COMPUTING RESOURCES

Justification of computational resources	CPU: Algorithm Input and storage requirements The CCD 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. For s we estimate 0.8 Tb of storage. We need extra storage for the outputs, so our total estimate is 1000 Gb of storage. 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. This estimated are currently being reviewed in the context of project 2024.07034.CPCA 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 for one single run of the CCD algorithm over all pixels. Near real time We extrapolate our estimates for a single date output to a temporal succession of outputs, updated in near real time with the availability of new Sentinel-2 imagery. For the tests we intend to perform and the final Optimizing Computational Resources Given the pixel-by-pixel application of the algorithm, maximizing the number of cores used is preferable. In the context of project 2024.07034.CPCA we are in the process of exploring MPI and adapting our algorithm to distribute the computation by several nodes to have access to a larger number of cores. GPU: GPU requirements for training and a single date The tests we have performed in a local machine (GPU Nvidia RTX 4090) suggest that our processing can be performed with 16 of GPU RAM since the batch size we used in our runs was around 32. If we extrapolate our tests to the whole Continental Portugal area where potential vegetation loss can occur (approximately 120 million Sentinel-2 pixels), we estimate that we will need approximately 300 GPU-hours. Extrapolation for multiple dates For the tests we plan to do for near real time, training will have to be extended to multiple dates. Although the algorithm will be developed to fine tune previous output models, we estimate that we will need 4 times more resources for task 3 (~1200 GPU .hours) than for task 2 (~300 GPU-hours)
Usage starting date by	01-01-2025
Usage end date by	31-12-2025
Additional comments	–
Considering article 23 of the regulation, do you accept that the following information may be used in publications related to the RNCA?	Yes

Application successfully submitted.