VALIDATION OF CCDC USE WITH SENTINEL-2 TIME SERIES FOR DEFORESTATION DETECTION IN PORTUGAL

Filipe Louro[1,2*], Hugo Costa[2], Mário Caetano[2]

[1] THE NAVIGATOR COMPANY, S.A. - Mitrena Aptd. 55 2910-738 Setúbal. Portugal.

[2] NOVA IMS Information Management School - Campus de Campolide, 1070-312 Lisboa. Portugal.

*Corresponding author
e-mail address: filipe.louro@thenavigatorcompany.com
Phone contact: +351 932009860

SUMMARY

Forest management planning and monitoring are complex tasks that require regular oversight. However, the widespread presence of smallholdings in Portugal and increasing rural abandonment make this task particularly challenging. The Continuous Change Detection and Classification algorithm is notable for enabling the analysis of trends in satellite image time series. While it has been widely used for continuous land cover change monitoring with Landsat data, it has seen limited application in Europe. The use of Sentinel-2 time series, offering higher spatial and temporal resolution and now covering a significant historical period, could be key to enabling CCDC for forest monitoring in Portugal and across Europe. This study aims to validate recently published parameter settings and processing optimizations for applying CCDC with Sentinel-2 data to detect deforestation events in Portugal. The methodology is compared against reference data on forestry activities provided by The Navigator Company. Results show strong agreement with the reference data, with eucalyptus stand harvest events detected with an F1-score of 0.86 and a detection lag of 19 days for 80%. There is no statistically significant difference in detection accuracy for smaller forest stands, suggesting the method is highly promising for monitoring in regions where smallholding forestry presents management challenges.

Keywords: Continuous Change Detection and Classification; Deforestation; Land Cover Monitoring; Sentinel-2; Sustainable Forest Management

INTRODUCTION

Forests account for about 30% of the world's surface area, being essential for some of humanity's basic needs such as food, clothing, medicine, shelter, and as a carbon sink [1]. Because they play an integral role in climate regulation and in maintaining ecological balances and carbon cycles, sustainable forest management is recognised as one of the most effective solutions for climate change mitigation and sustainable development [2]. In mainland Portugal, forest areas represent 69.4% of the territory and are mostly privately owned (only about 3% of the forest area in Portugal is public). They are characterised by a high prevalence of smallholdings, with 24% of forest areas being smaller than 2 ha and 30% between 2 and 10 ha [3]. Forest management planning is a complex problem that depends on different economic, environmental, and social criteria [4]. It largely depends on regular monitoring and follow-up which, until recently, in Portugal, was achieved almost exclusively through fieldwork by local technicians in the case of large companies, or by the landowners themselves. The prevalence of small-scale forest stands, implies additional challenges for monitoring and management. This paradigm is one of the main causes for the lack of forest stand management in Portugal, hindering the application of adequate silvicultural models and their monitoring.

Changes in land use and land cover can be studied at different scales and aim to accurately reflect phenomena or changes in the state of objects, whether they are natural phenomena or of anthropogenic origin. It is essential to understand the anthropogenic influence on land use and land cover changes, and how this contributes to such phenomena becoming more frequent and occurring at increasingly local scales [5, 6]. Satellite programmes such as Landsat and Sentinel, due to their availability, ease of access, and spatial and temporal resolutions (30 m and 8 days in the case of Landsat, and 10 m and 5 days in the case of Sentinel), and Google Earth Engine (GEE), which allows free processing of satellite datasets and readily available cloud computing, have enabled the expansion of the user community to process and analyse satellite observation time series [7]. Landsat time series, containing images from the last 50 years and being completely free since 2008, have been one of the standards for land cover change analyses, with a focus on forested areas[8]. Due to its more favourable spatial and temporal resolutions, Sentinel-2 time series are now reaching a more relevant temporal span, enabling their use for trend analysis and interpretation over time. Sentinel-2 imagery has been widely used for forest monitoring, from the simplest production of land use and land cover maps to conducting forest inventories, detecting invasive species, monitoring wildfires (and post-fire management). It has even been preferred over other higher-resolution datasets due to its free nature [9]. The use of remote sensing data to study land change, particularly satellite images, offers undeniable advantages such as wide coverage with long time series and easy processing, being one of the most relevant data sources for land change detection in recent decades [10].

Wulder et al. [11] describe how recent technological advances, combined with the need to reduce costs, product latency, and human subjectivity, have over the past three decades led to a paradigm shift in the production and availability of land cover data. Data and products are increasingly broad and publicly available; systematically generated and more frequently updated; dependent on advanced and objective algorithms; easily accessible at low or no cost; and informed by time series, allowing the study not only of state but also of trends. Continuous Change Detection and Classification (CCDC) is a method that aims, for each pixel, to flag land cover changes in satellite image time series. It seeks to characterise intra-annual changes (usually due to temperature or precipitation fluctuations caused by the seasons) and gradual interannual changes (gradual trends associated with vegetation growth or climate variability) through the fitting of a harmonic regression and, subsequently, to flag abrupt changes (usually attributable to natural disasters or anthropogenic phenomena) by identifying values significantly different from those expected according to the established model [12, 13]. Figure 1 shows an example of CCDC behaviour, establishing regression lines for expected values and flagging a break when it identifies values significantly different from expected.

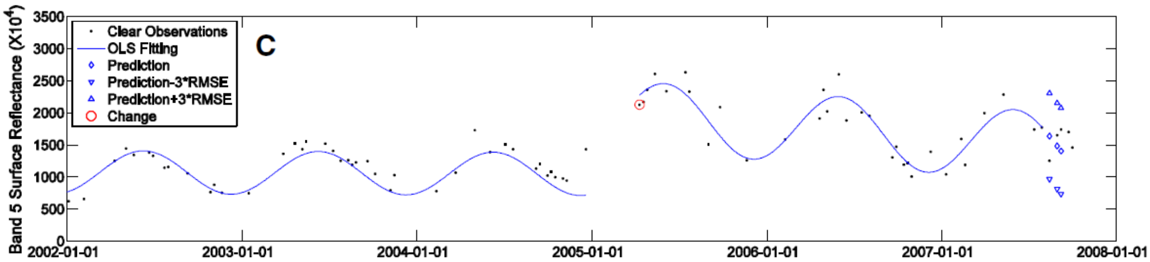

Figure 1. Example of CCDC application and break identification [12].

Since its original publication in 2014 and until the end of 2023, the method has been increasingly used in land cover change studies, mainly in the Amazon, China, and the United States of. Due to their longer time series (available since the 1970s), most case studies are based on Landsat data. As far as can be verified, the use of CCDC on Sentinel-2 time series is still in its early stages, but it can be an important tool for monitoring forest plots in Portugal. As with Landsat, access to its time series is free and available through Google Earth Engine; however, it has a higher spatial resolution, better suited to the small size of land parcels in Portugal.

In 2024, Moraes et al. [14] used 300 circular plots with a radius of 200 m to parameterise the CCDC method and identify land cover changes, using photointerpretation of orthophotos (25 cm) and Sentinel-2 images (10 m) to validate the resulting products. This work aims to assess the potential usefulness of Sentinel-2 imagery and the CCDC with the optimised parameterisation for Portugal to map deforestation actions in Portuguese forest stands.

This protocol, entirely dependent on free and widely available datasets and open-source software, could become a major asset for forest stand managers in Portugal as a relevant complement for monitoring smallholding areas (a large part of the national territory), contributing to more efficient and sustainable management and greater productivity.

EXPERIMENTAL

A map of land cover change detections between 2021 and 2023 in mainland Portugal was produced from a time series of Sentinel-2 satellite images using the Google Earth Engine (GEE) JavaScript API. This script used the dataset "Harmonized Sentinel-2 MSI: MultiSpectral Instrument, Level-2A", indexed as "S2_SR_HARMONIZED" in the GEE catalogue, to obtain a time series of Level-2A surface reflectance images (orthorectified and atmospherically corrected product – all reflectance values refer to the Bottom-of-Atmosphere) with a spatial resolution of 10 m and normalised values from 28/03/2017 [15, 16]. For each image, clouds and shadows were masked using the s2cloudless algorithm provided by ESA's Copernicus programme and Normalized Difference Vegetation Index (NDVI) [17] was calculated. This index is based on reflectance in the Red (B4) and Near-Infrared (B8) bands to identify the presence of vegetation according to Equation (1).

$$\mathrm{NDVI} = \frac{(\mathrm{B8} - \mathrm{B4})}{(\mathrm{B8} + \mathrm{B4})} \tag{1}$$

The instance of the CCDC algorithm available in GEE was applied using the parameters identified by Moraes et al. [14] as optimal for detecting land cover changes with Sentinel-2 time series in Portugal (lambda = 200, chi-square = 0.999, minYears = 1, and minObservations = 4), determining, for each pixel, the date of the last break. Subsequently, a majority filter was used and polygons for each area affected by a break detected on the same date were extracted. This identifies all contiguous pixels whose break date is the same and returns vector-format polygons. Finally, all polygons corresponding to dates outside the analysed period (2021 to 2023) were ignored, and a Minimum Mapping Unit (MMU) of 0.1 ha was defined.

For this study, 300 eucalyptus forest stands (minimum forest management unit) managed by The Navigator Company (Navigator) were randomly selected, stratified across the company's administrative regions, each representing different forest stand management needs due to the potential productivity inherent to climate, soil type, or other constraints. This ensured a distribution with adequate representation of the eucalyptus production forest areas managed by the company. The studied stands have an average area of 41.56 ha, a median area of 27.19 ha. The smallest stand has an area of 0.36 ha, while the largest has 303.97 ha. This sample is representative of Navigator's holdings, both in terms of area distribution and geographic dispersion in Portugal. For these stands, queries were performed in Navigator's databases to select and identify the following situations occurring between 2021 and 2023:

(1) Harvesting actions: these records represent all interventions in which a forest stand underwent clearcutting for wood harvesting as raw material. For each stand, it was possible to identify (if it was harvested) the boundaries of the harvested area and the start and end dates of the intervention.

(2) Silvicultural actions: these records represent all actions resulting from the usual management of forest stands that could potentially cause false positives by leading *a priori* to a possible reduction of vegetation in the studied stand.

(3) Burnt areas – stands affected by wildfires.

The break map was first compared with Navigator's harvesting and burnt area records. True positives (TP) were defined as stands with harvesting actions or wildfires during the study period detected by the CCDC (i.e., situations where the produced map and reference data matched). For objective analysis, stands were considered TP if they met the following conditions:

(1) Break polygon(s) overlap at least 80% of the reference polygon area – 80% was chosen to account for potential information loss arising from the generalisation in the processing workflow. Due to the application of the majority filter and the established MMU, slight reductions in area are expected and should not be penalised.

(2) Break date within 60 days after the reference date – the application of the s2cloudless cloud

mask may result in missing samples in some locations, especially in winter. Thus, a delay of a few weeks in break detection by CCDC is acceptable.

True negatives (TN) were all stands with no harvesting records in the study period and no breaks detected by the CCDC.

For each stand, commission errors (false positives – FP) and omission errors (false negatives – FN) were quantified, and the product's accuracy and F1-score were calculated according to Equations (2 – 5) [18, 19]:

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} \tag{2}$$

$$Commission = \frac{FP}{(TP + FP)} \tag{3}$$

$$Omission = \frac{FN}{(TP + FN)} \tag{4}$$

$$F1 - score = \frac{(1 - Commission) \times (1 - Omission)}{(2 - Commission - Omission)} \times 2 \tag{5}$$

Similar criteria were used to study the influence of silvicultural actions on break detections by the CCDC.

RESULTS AND DISCUSSION

The product obtained was validated according to the criteria described for comparison with the reference data. For the 300 stands analysed, 82 TPs, 23 FPs, 3 FNs, and 192 TNs were quantified (Table 1). These results represent an accuracy of 0.91 and an F1-score of 0.86 (commission of 0.22 and omission of 0.04).

The results are similar when considering the area of the studied stands, with TPs having a total area of 3,793.97 ha, FPs 1,329.71 ha, FNs 132.29 ha, and TNs 7,212.51 ha, for a total of 12,468.48 ha (Table 2). These results represent an accuracy of 0.88 and an F1-score of 0.84 (commission of 0.26 and omission of 0.03).

Table 1. Harvesting detection by stand.

CCDC\Reference	Harvest	No harvest
Break detected	83	23
No break detected	3	192

Table 2. Harvesting detection by area

CCDC\Reference	Harvest	No harvest
Break detected	3794 ha	1330 ha
No break detected	132 ha	7213 ha

Among the 82 stands identified as TPs, 81 correspond to stands where harvesting operations took place and one to a stand affected by wildfire. Figure 2 shows one case where the detected cuts match the reference data well. In this example, the harvested area was correctly identified in the resulting product. Figure 3 illustrates the case where the damage caused by the wildfire of 5–9 August in Odemira was correctly detected by the CCDC on 7 and 12 August.

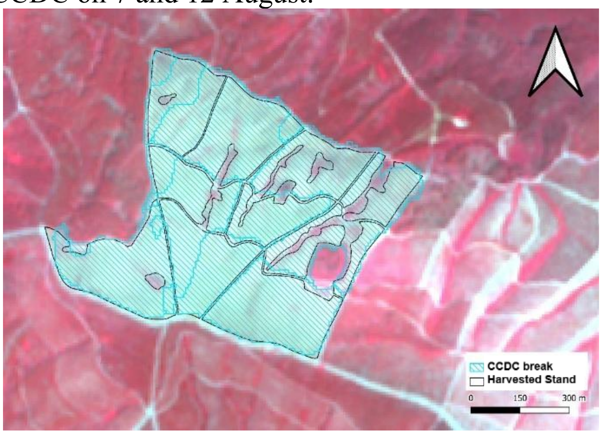

Figure 3. Example of CCDC break detection of a harvested stand on a false color Sentinel-2 image.

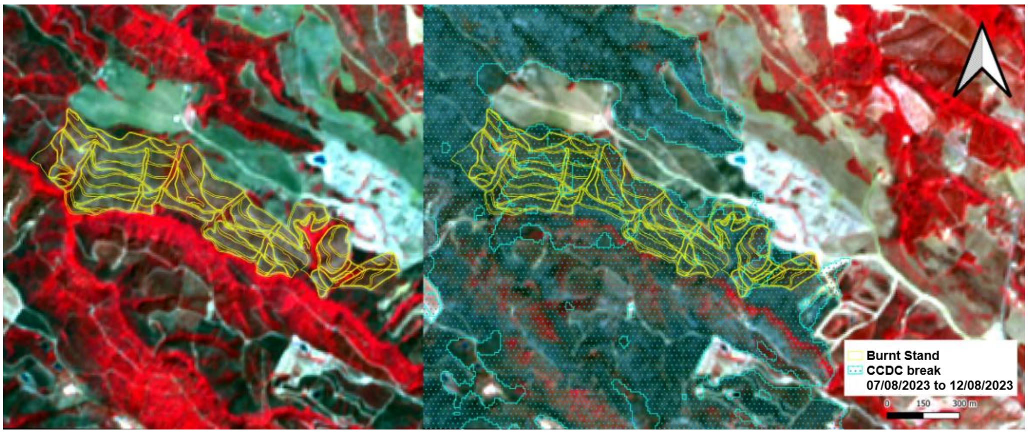

Figure 4. Example of CCDC break detection of a burnt stand on a false color Sentinel-2 image.

In addition to considering all harvesting actions, all records of silvicultural actions that could result in biomass reductions and/or represent abrupt changes in image reflectance were also considered. Within the universe of the 300 stands evaluated, 108 were identified with a history of operations in the study period, and for each of the silvicultural actions considered, it was quantified whether the operations gave rise to a break in the CCDC or not. Regarding spontaneous vegetation control actions, it was observed that in the 15 stands with records of manual clearing, there was one detection, and in the 42 stands where mechanical clearing operations were carried out, there were nine detections. These results are in line with expectations, as they reflect the severity of spontaneous vegetation and, consequently, the degree of human intervention. As for site preparation works for new planting, all were identified by the methodology used. These, together with harvesting, are the operations that cause the most significant changes. They normally involve the use of heavier machinery and can represent abrupt changes in the stands. Due to their nature, it is difficult to distinguish site preparation works from harvesting actions, as both cause a sharp biomass reduction. These cases, where silvicultural operations were identified as land cover changes by the CCDC, explain 20 of the 23 commission errors previously found. The possibility of maintenance operations being detected by the described protocol was anticipated, and these errors were therefore expected.

To study whether stand area could influence error production, and since a higher incidence of errors (mainly FNs) was expected in stands with smaller areas, the average and median stand areas were calculated for each scenario. The different results identified vary similarly, and a one-way ANOVA with a post-hoc Tukey HSD statistical test detected no significant differences between scenarios, as seen in ($p > 0.05$). Additionally, it was found that the FN with the smallest area was 22.63 ha; however, 36 smaller stands were identified as TPs (7 of them under 5 ha). From these results, it can be inferred with high confidence that the range of stand areas studied had no influence on this study and that no FN results were due to small areas.

The criteria defined for identifying TPs were also studied. All harvesting actions in the reference data identified as FNs are due to the absence, or residual detection, of breaks by the CCDC in the stand area. This means that the 60-day temporal window was adequate. More specifically, for each of the 82 TPs, the start and end dates of the reference disturbance (harvest or wildfire), as well as the dates of the first and last polygons detected by the CCDC, were examined. Figure 4 shows the lag (Δ) between these dates. Figure 5 also shows that both for the start and the end, the lags between the reference and detection behave similarly. Approximately 50% of detections occur within six days of the reference date and 80% within 19 days. Given the 5-day revisit time of Sentinel-2, it can be estimated that most disturbances are detected on the first or second satellite pass after the change, which would be expected in optimal situations. The detection of initial changes (start) seems to be more prompt, with the break with the largest lag detected 46 days later. On the other hand, some detections were up to 58 days after the intervention end date. These lags can be explained by the application of the cloud mask (clouds over the site in consecutive revisits may delay detection) or by the presence of residual vegetation in the stand immediately after cutting. Since only the wood is removed, the branches and leaves remaining on site may cause a less abrupt decrease in NDVI, delaying the break detection by the CCDC.

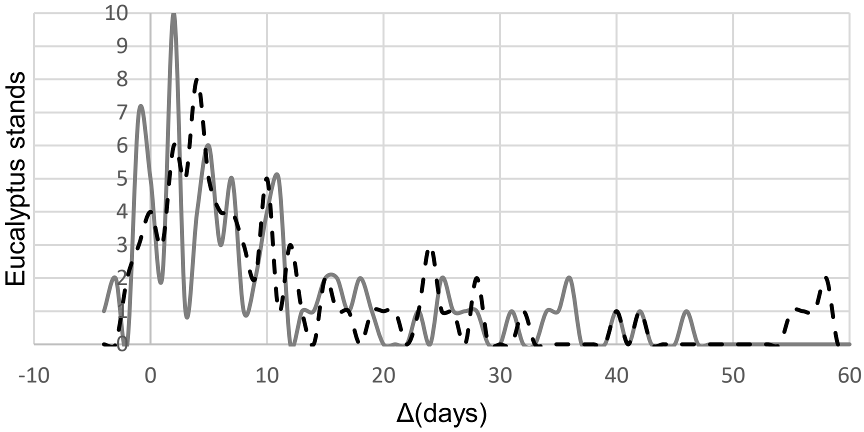

Figure 4. Number of days between the start date (solid line) and the end date (dash line) of disturbance reference and detection of disturbance by CCDC.

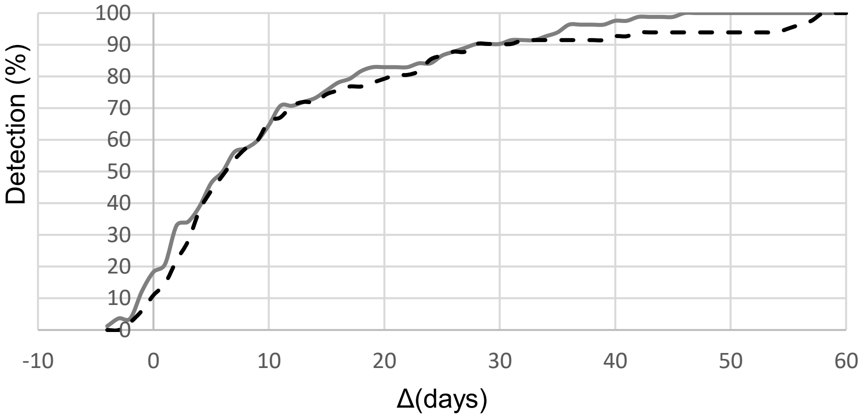

Figure 5. Percentage of start date (solid line) and end date (dash line) detection of disturbance by CCDC over time.

It is also interesting to note that some works were detected by the CCDC before the reference data recorded them. Among the 82 TPs, 10 stands had breaks detected by the CCDC before the harvest start date. These detections occurred up to four days before the reference date, i.e., on the Sentinel-2 pass immediately before the reference date. These may indicate preparatory works for the start of operations (such as opening clearings for unloading heavy machinery or for future stacking of harvested timber) or recording errors. Regarding the end of the intervention, there are also five break detections by the CCDC before the reference date which may represent situations where the last works were not detected or were lost in the map generalisation process because they were small areas, or they may be due to recording errors.

One of the most promising applications of this product is the possibility of post-validation of completed works. The gradual progress of operations can be analysed with very low delay compared to reality. In this case, although Navigator's weekly records only indicate that harvesting took place between 26/12/2022 and 20/02/2023, the product obtained allows the works carried out to be studied and even enables the extraction of quality and work rate metrics. In optimal circumstances and in periods without cloud cover, it is expected that CCDC could be used to validate forest harvesting operations without requiring a responsible technician to visit the site.

The results are extremely promising and represent an F1-score slightly higher than that presented by Moraes et al. in 2024 [14]. While Moraes et al. obtained an F1-score of 0.82, this approach, studying forest stands, produced an F1-score of 0.86. The commission error was similar in both cases (0.22 compared to 0.21 for Moraes et al.), but the omission error was much lower (0.04 compared to 0.15 for

Moraes et al.). These discrepancies can easily be explained by the different validation and accuracy assessment methodology. The present work is based on forest patches with known boundaries and seeks to match detected breaks to historical forest management records, while Moraes et al. considered each pixel individually. This procedure, combined with the majority filter applied and the defined MMU (necessary to generate a product more easily interpretable by the user), certainly explains these results, making the final product more suitable for use by forest technicians for forest stand management.

CONCLUSIONS

Overall, the method used resulted in a product that was highly suitable and closely matched the reference data. It was demonstrated that the use of the CCDC algorithm in combination with Sentinel-2 time series can be a highly reliable tool for monitoring eucalyptus production forest stands, allowing the identification and detection of abrupt land cover changes, particularly harvesting operations, with high accuracy. In the review of published works using CCDC, a reliance on Landsat time series was observed, as well as an absence of studies in Europe. The present study shows that the use of Sentinel-2 time series, although shorter than Landsat's, can be relevant in applying CCDC and can address monitoring needs where Landsat, due to its lower spatial resolution, is insufficient.

Among the 300 stands studied, 274 were correctly identified by the CCDC (81 harvests, 1 wildfire, and 192 without changes). Only 3 stands with harvesting records were not identified by the described protocol, and 23 were falsely flagged as having been subject to deforestation. Among the various reasons for the occurrence of these false positives, the nine stands that underwent site preparation for replanting are noteworthy.

It is important to highlight that, although the sampled reference data is representative of the forest estate managed by Navigator, they are not necessarily representative of the national forest landscape. Despite having studied stands with very small areas, only 10 of them are under 2 ha. The reality of forests in Portugal, particularly in the northern half, is regularly characterised by smallholdings, where rural plots often have areas below 2 ha, sometimes even smaller than 0.1 ha and widths of less than 10 m. This reality poses an extreme challenge, and high-quality results are not expected in such situations, even with the high spatial resolution of Sentinel-2.

In addition to enabling remote monitoring of its own forest stands, the possibility for Navigator to remotely monitor harvesting operations in third-party forest stands can also be highly useful in a Business Intelligence framework. By knowing the forest patches that traditionally supply the company's timber yards and using CCDC, it may be possible to anticipate timber deliveries from local producers or loggers. This intelligence could be an asset for stock and supply management, allowing the company to be less dependent on market price fluctuations and thus gain a significant competitive advantage.

It is also crucial to recognise the potential use of this methodology to address the growing need to ensure companies' compliance with sustainability commitments. One of the anticipated opportunities for such tools is to respond to the expected implementation of the European Union Deforestation Regulation (EUDR), which is expected to be enforced in early 2026. This regulation foresees a blockchain model intended to track forest products from harvest to final use, holding each actor in the supply chain responsible for carrying out due diligence and ensuring the correct provenance of each forest-derived product [20]. With the use of CCDC, companies such as Navigator can more easily monitor harvesting operations in third-party forest stands at the time of raw material acquisition. This way, they can ensure the provenance of each batch, validating whether the declared origin indeed corresponds to a forest stand harvested on the stated dates.

This validation demonstrates the suitability of the methodology, as well as the time series used, for detecting clearcuts in Portuguese forest areas and for the potential production of land cover change maps with a high degree of confidence.

REFERENCES

1. G. S. Bhunia, P. K. Shit. Chapter 1 - Recent development and future challenges of geospatial approaches

for enhancing forest inventories, in: P. Kumar Shit (eds.), *H*, R. Pourghasemi, P. P. Adhikary, G. S. Bhunia, & V. P. Sati (Eds.), Forest Resources Resilience and Conflicts (pp. 3–16). Elsevier (2021).

2. Y. Cai, Q. Shi, X. Xu, X. Liu. A novel approach towards continuous monitoring of forest change dynamics in fragmented landscapes using time series Landsat imagery. *Int. J. Appl. Earth Obs. Geoinf.*, **118**: 103226 (2023).
3. ICNF. *6º Inventário Florestal Nacional - Relatório Final* (2019).
4. L. Diaz-Balteiro, C. Romero. Making forestry decisions with multiple criteria: A review and an assessment. *Forest Ecology and Management*, **255**(8): 3222–3241 (2008).
5. E. F. Lambin, P. Meyfroidt. Global land use change, economic globalization, and the looming land scarcity. *Proc. Natl. Acad. Sci. U.S.A.*, **108**(9): 3465–3472 (2011).
6. J. Townshend, C. Justice, W. Li, C. Gurney, J. McManus. Global land cover classification by remote sensing: present capabilities and future possibilities. *Remote Sens. Environ.*, **35**(2–3): 243–255 (1991).
7. V. J. Pasquarella, Arévalo, P., K. H. Bratley, E. L. Bullock, N. Gorelick, Z. Yang, R. E. Kennedy. Demystifying LandTrendr and CCDC temporal segmentation. *Int. J. Appl. Earth Obs. Geoinf.*, **110**: 102806 (2022).
8. C. E. Holden, C. E. Woodcock. An analysis of Landsat 7 and Landsat 8 underflight data and the implications for time series investigations. *Remote Sens. Environ.*, **185**: 16–36 (2016).
9. D. Phiri, M. Simwanda, S. Salekin, V. R. Nyirenda, Y. Murayama, M. Ranagalage. Sentinel-2 data for land cover/use mapping: A review, in: Remote Sensing (Vol. 12 (eds.), *Issue 14)*, MDPI AG (2020).
10. B. Liu, W. Song, Z. Meng, X. Liu. Review of Land Use Change Detection—A Method Combining Machine Learning and Bibliometric Analysis, in: Land (Vol. 12 (eds.), *Issue 5)*, Multidisciplinary Digital Publishing Institute (MDPI) (2023).
11. M. A. Wulder, N. C. Coops, D. P. Roy, J. C. White, T. Hermosilla. Land cover 2. *0. International Journal of Remote Sensing*, **39**(12): 4254–4284 (2018).
12. Z. Zhu, C. E. Woodcock. Continuous change detection and classification of land cover using all available Landsat data. *Remote Sens. Environ.*, **144**: 152–171 (2014).
13. Q. Zhou, Z. Zhu, G. Xian, C. Li. A novel regression method for harmonic analysis of time series. *ISPRS J. Photogramm. Remote Sens.*, **185**: 48–61 (2022).
14. D. Moraes, B. Barbosa, H. Costa, F. D. Moreira, P. Benevides, M. Caetano, M. Campagnolo. Continuous forest loss monitoring in a dynamic landscape of Central Portugal with Sentinel-2 data. *Int. J. Appl. Earth Obs. Geoinf.*, **130**: 103913 (2024).
15. ESA. *Sentinel-2 User Handbook*. In ESA Standard Document (2nd ed.). ESA Standard Document (2015).
16. GEE. *Harmonized Sentinel-2 MSI: MultiSpectral Instrument, Level-2A*. Earth Engine Data Catalog (2024a).
17. J. W. Rouse, R. H. Haas, J. A. Schell, D. W. Deering. Monitoring vegetation systems in the Great Plains with ERTS. *NASA Spec. Publ*, **351**(1): 309 (1974).
18. R. G. Congalton, K. Green. *Assessing the accuracy of remotely sensed data: principles and practices*. CRC press (2019).
19. Z. Zhu, J. Zhang, Z. Yang, A. H. Aljaddani, W. B. Cohen, S. Qiu, C. Zhou. Continuous monitoring of land disturbance based on Landsat time series. *Remote Sens. Environ.*, **238**: 111116 (2020).
20. M. Köthke, M. Lippe, P. Elsasser. Comparing the former EUTR and upcoming EUDR: Some implications for private sector and authorities. *For. Policy Econ.*, **157**: 103079 (2023)