A SENTINEL-2 BASED CLOUD COVER RECURRENCE TOOL FOR QUICK ASSESSMENT OF BRAZILIAN LOCATIONS

Bruno Rech[1]; Jahdy Moreno Oliveira[1]; Vinícius Lima Guimarães[1]; Gilberto Ribeiro de Queiroz[1]; Thales Sehn Körting[1]; Cláudio Clemente Faria Barbosa[1]; Evlyn Marcia Leão de Moraes Novo[1]

[1] Earth Observation and Geoinformatics Division, National Institute for Space Research, São José dos Campos, SP, Brazil.
{bruno.rech, jahdy.oliveira, vinicius.lima, gilberto.queiroz, thales.korting, claudio.barbosa, evlyn.novo}@inpe.br
https://orcid.org/0000-0003-2787-2450; https://orcid.org/0000-0002-6842-258X; https://orcid.org/0000-0003-0686-8321; https://orcid.org/0000-0001-7534-0219; https://orcid.org/0000-0002-0876-0501; https://orcid.org/0000-0002-3221-9774; https://orcid.org/0000-0002-1223-9276

RESUMEN. Las aplicaciones de big data en observación de la Tierra son herramientas poderosas diseñadas para mejorar el almacenamiento, la organización y la gestión de datos de observación de la Tierra a gran escala. Por ejemplo, la plataforma Brazil Data Cube (BDC) proporciona cubos de datos de resolución media para el territorio brasileño. Basándose en los datos de nivel 2A del Sentinel-2 disponibles en el BDC, se desarrolló un algoritmo para la evaluación rápida de la dinámica de cobertura de nubes sobre una región y un período de interés, la Herramienta de Recurrencia de Cobertura de Nubes (CCRT), que genera información estadística básica y visualizaciones de los datos. El algoritmo se aplicó en el municipio de São Paulo para ilustrar sus funcionalidades.

Palabras-clave: Brazil Data Cube; Teledetección; Series temporales; Visualización de datos.

RESUMO. As aplicações de big data em observação da Terra são ferramentas poderosas projetadas para aprimorar o armazenamento, a organização e a gestão de dados de observação da Terra em larga escala. Por exemplo, a plataforma Brazil Data Cube (BDC) fornece cubos de dados de média resolução para o território brasileiro. Com base nos dados nível 2A do Sentinel-2 disponíveis no BDC, desenvolveu-se um algoritmo para avaliação rápida da dinâmica de cobertura de nuvens sobre uma região e um período de interesse, a Ferramenta de Recorrência de Cobertura de Nuvens (CCRT), que gera informações estatísticas básicas e visualizações dos dados. O algoritmo foi aplicado para o município de São Paulo para ilustrar suas funcionalidades.

Palavras-chave: Brazil Data Cube; Sensoriamento Remoto; Séries Temporais; Visualização de Dados.

ABSTRACT. Earth observation big data applications are powerful tools designed to enhance the storage, organization, and management of large-scale Earth observation data. For instance, the Brazil Data Cube (BDC) platform provides medium-resolution data cubes for the Brazilian territory. Based on Level-2A Sentinel-2 data available at BDC, we developed an algorithm for quick assessment of cloud cover dynamics over a region and a period of interest, the Cloud Cover Recurrence Tool (CCRT), which generates basic statistical information and data visualizations. The algorithm was applied to the municipality of São Paulo to illustrate its functionalities.

Keywords: Brazil Data Cube; Remote Sensing; Time Series; Data Visualization.

1. INTRODUCTION

One of the major challenges faced by the remote sensing community is dealing with large datasets, as the number of different open-access spaceborne and airborne products exceeds petabyte size (Amani *et al.*, 2020). In an effort to address such challenge, Earth observation big data applications emerge as a trustable and feasible way to gather and provide analysis-ready data, enabling users to get and compare products from multiple sensors across spatial, temporal, and spectral dimensions (Lucena *et al.*, 2020). They consist of collections of satellite imagery, co-registered, and cropped based on user-defined parameters (Lima *et al.*, 2021).

In this regard, the National Institute for Space Research (INPE) developed the Brazil Data Cube (BDC) platform, as part of the Environmental Monitoring of Brazilian Biomes project. BDC provides multidimensional data cubes acquired by different resolution sensors onboard Sentinel-2, CBERS-4/4A, Landsat 5/7/8/9 and AQUA/TERRA satellites (Marujo *et al.*, 2022). Through the SpatioTemporal Asset Catalog (STAC), BDC grants access to different cloud and cloud shadow masks, such as the scene classification layer (SCL), resulting from the Sen2Cor atmospheric correction module applied to Sentinel 2 Multispectral Instrument (MSI) images. STAC is a standardized application programming interface (API) that facilitates the discovery of catalogs, items, and assets within the data collection.

Sentinel-2 mission is constituted by a constellation of two identical satellites (Sentinel-2A and Sentinel-2B) in the same polar and Sun-synchronous orbit, aimed at global coastal and land observation. Both are equipped with MSI sensor, which acquires 13 bands in the visible, near infrared, and shortwave infrared spectral regions, providing a high spatial resolution (10 to 60 m) and a revisit time of 5 days at the Equator (Drusch *et al.*, 2012). Sen2Cor, meanwhile, is the algorithm used by the European Space Agency (ESA) for easy and efficient atmospheric correction of MSI images and generation of level 2A products (i.e., surface reflectance) (Gascon *et al.*, 2017; Sanchez *et al.*, 2020). Within Sen2Cor, SCL is a map input that allows one to distinguish between cloudy, clear, and water pixels (Gascon *et al.*, 2017), available at 20 m spatial resolution in BDC (Lucena *et al.*, 2020).

In optical remote sensing, cloud cover is one of the most important parameters to consider when searching and selecting satellite imagery, as it may preclude the visualization of the regions of interest (ROI) by the sensor (Bergsma; Almar, 2020; Tiede *et al.*, 2021). Therefore, providing users with additional information about cloud cover can offer a better assessment of its spatiotemporal dynamics. Given this perspective and the ease of using BDC data, Lima *et al.* (2021) developed a tool to evaluate cloud cover in Landsat images based on user-defined ROI and time parameters, which was called Brazil Data Cube Cloud Coverage (BDC3).

Inspired by BDC3 (Lucena *et al.*, 2020), in this study we developed the Cloud Cover Recurrence Tool (CCRT), a ready-to-use algorithm for evaluating cloud cover based on Sentinel-2 SCL bands. CCRT takes an ROI and a period of interest, and offers an overview of the cloud recurrence according to the input parameters. The municipality São Paulo was used to illustrate the algorithm applicability.

2. MATERIALS AND METHODS

The assessment of cloud cover recurrence was developed based on the Sentinel-2 SCL band, retrieve from BDC database. To obtain the metrics, we considered the class Cloud High Probability (pixel 9), which has an estimated accuracy of 85.4% (Gascon *et al.*, 2017). The data were accessed through the BDC-STAC API for Python.

From an ROI and a period of interest, the algorithm is capable of generating simple statistics related to the dynamics of cloud cover (recurrence). Figure 1 briefly illustrates the operation of the algorithm.

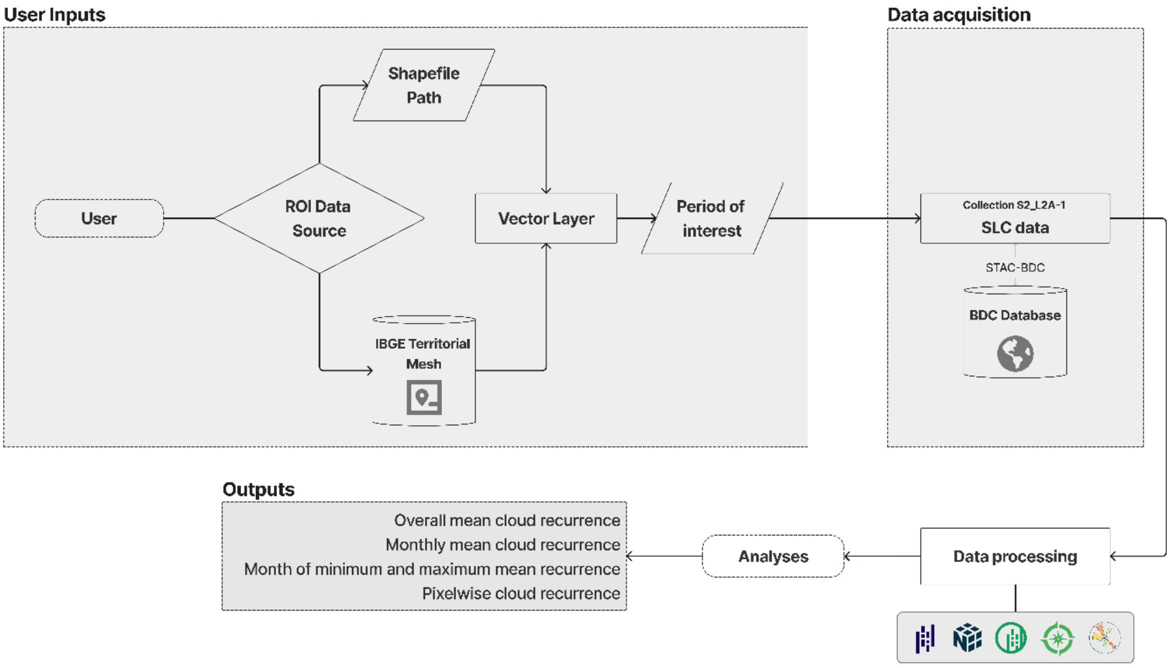


Figure 1. STAC-BDC data acquisition scheme.

The algorithm starts with the user informing whether the analysis will be carried out considering a user-defined ROI or a Brazilian municipality. In the latter case, the user is asked to inform the state and municipality name, and the data are automatically uploaded from the Brazilian Institute of Geography and Statistics (IBGE) database. In both options, the user needs to specify the period of interest. For periods that include two or more months, the user is asked whether a monthly analysis should be performed; otherwise, only an overall analysis is performed.

For ROIs covered by more than one tile, the processing is performed over each tile separately, and the data is merged at the end. In each scene, the proportion of no-data pixels within the ROI is evaluated; scenes with a proportion greater than 5% are discarded to avoid introducing errors in the algorithm.

Raster and vector data are processed according to the parameters defined by the user, and information on cloud cover is returned. The analysis provides information on overall and monthly (if feasible/required) cloud cover recurrence over the ROI, the months of maximum and minimum cloud cover recurrence (if feasible/required), and maps of pixelwise overall/monthly cloud cover recurrence.

3. RESULTS AND DISCUSSION

The Python code that constitutes the cloud cover recurrence tool is available at https://github.com/b-rech/cloud-cover. Several tests were performed to verify the functionality and the limitations of the algorithm. It allowed for the inclusion of warnings within the code so as to inform the user whether a processing may raise any issues regarding processing time and memory usage.

To demonstrate the functionality of the CCRT, we briefly present some results obtained for the municipality of São Paulo, which spans an area of 1500 km^2. The analysis considered the whole year of 2023, and the search in the BDC database yielded 294 scenes divided in 2 tiles. From that, 146 useful scenes were left after the no-data assessment. Analysis indicated that October had the highest mean cloud recurrence in 2023 (0.64 ± 0.11), while the lowest mean was recorded in September (0.17 ± 0.03). Figure 2 shows the monthly mean cloud recurrence over the year; it is one of the products generated by CCRT.

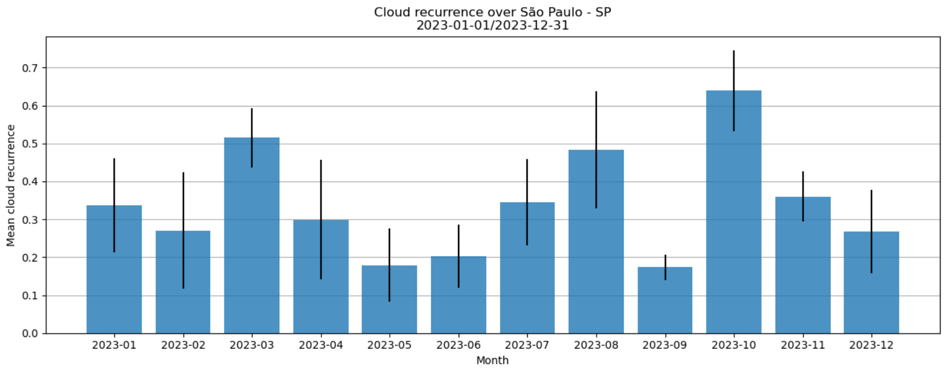
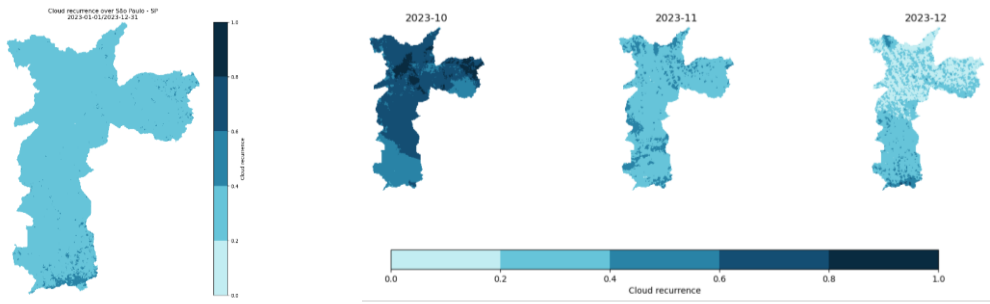


Figure 2. Monthly mean cloud recurrence in the municipality of São Paulo in 2023. Black lines indicate standard deviation.

Additionally, the algorithm provides the spatial distribution of cloud recurrence over the ROI for each month and/or for the entire period, as can be seen in Figure 3.



(a) (b)

Figure 3. Maps generated in CCRT for a) overall and b) monthly cloud recurrence analyses. Only the last three months are shown in b for the sake of brevity, but the tool produces a figure with all the months analyzed.

4. CONCLUSIONS

The Cloud Cover Recurrence Tool (CCRT) is based on the Sentinel-2 SCL band, and can produce an assessment of cloud recurrence considering the delimitation of Brazilian municipalities or other user-defined areas. It proved to be a valuable and promising application of the Brazil Data Cube dataset. The tool can provide an understanding of cloud cover recurrence, giving to the user easy and intuitive visualization of its spatial and temporal patterns. The algorithm can be continuously improved, and future enhancements should include the implementation of other statistical analyses, the minimization of computation time (which can allow the assessment of larger areas), and the use of data from other sensors.

REFERENCES

AMANI, M. *et al*. **Google Earth Engine Cloud Computing Platform for Remote Sensing Big Data Applications:** A Comprehensive Review. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, v. 13, p. 5326-5350, 2020. DOI: 10.1109/JSTARS.2020.3021052.

BERGSMA, E. W. J.; ALMAR, R. **Coastal coverage of ESA' Sentinel 2 mission**. Advances in Space Research, v. 65, n. 11, p. 2636-2644. 2020. DOI: 10.1016/j.asr.2020.03.001.

BOYD, C. E. Water Quality. Cham: Springer International Publishing, 2020. E-book. ISBN 9783030233341. Retrieved from: https://doi.org/10.1007/978-3-030-23335-8.

CHEN, J. *et al*. **Remote Sensing Big Data for Water Environment Monitoring:** Current Status, Challenges, and Future Prospects. Earth's Future, v. 10, n. 2. 2022. DOI: 10.1029/2021ef002289.

DRUSCH, M. *et al*. **Sentinel-2: ESA's Optical High-Resolution Mission for GMES Operational Services**. Remote Sensing of Environment, v. 120, p. 25-36. 2012. DOI: 10.1016/j.rse.2011.11.026.

GASCON, F. *et al*. **Copernicus Sentinel-2A calibration and products validation status.** Remote Sensing, v. 9, n. 6, p. 584. 2017. DOI: 10.3390/rs9060584.

LIMA, T. M. A. *et al*. Automated cloud coverage analysis with Brazil Data Cube. In: BRAZILIAN SYMPOSIUM ON GEOINFORMATICS (GEOINFO). **Proceedings XXII GEOINFO**, November 29 to December 02, 2021, São José dos Campos, SP, Brazil. INPE, 2021. Online. ISSN 2179-4847

LUCENA, F. R. S. M. *et al*. Brazil Data Cube Cloud Coverage (BDC^3) Viewer. In: BRAZILIAN SYMPOSIUM ON GEOINFORMATICS (GEOINFO). **Proceedings XXI GEOINFO**, November 30 to December 03, 2020, São José dos Campos, SP, Brazil. INPE, 2020. Online. ISSN 2179-4847.

MARUJO, R. F. B. *et al*. Generating Analysis Ready Data Collections for Brazil. In: INTERNATIONAL GEOSCIENCE AND REMOTE SENSING SYMPOSIUM (IGARSS). V 42. 2022. DOI: 10.1109/IGARSS46834.2022.9884104.

SANCHEZ, A. H. *et al*. **Comparison of cloud cover detection algorithms on sentinel-2 images of the Amazon tropical forest**. Remote Sensing, v. 12, n. 8, p. 1284, 18 abr. 2020. DOI: 10.3390/rs12081284.

TIEDE, D., SUDMANNS, M., AUGUSTIN, H., BARALDI, A. **Investigating ESA Sentinel-2 products' systematic cloud cover overestimation in very high-altitude areas.** Remote Sensing of Environment, v. 252, p. 112163, 2021. DOI: 10.1016/j.rse.2020.112163.