CAPÍTULO 04

EFFICIENT USE OF LANDSAT-8 DATA TO MINIMIZE CLOUD COVERAGE IN THE ANALYSIS OF SOYBEAN, CORN, AND COTTON CULTIVATION

Márcia Helena Galina Dompieri
PhD in Geography
Brazilian Agricultural Research Corporation
E-mail: marcia.dompieri@embrapa.br

Pedro Alves Quilici Coutinho
Bachelor's Degree in Agronomy
Brazilian Agricultural Research Corporation
E-mail: pedro.coutinho@colaborador.embrapa.br

Marcos Aurélio Santos Silva
PhD in Computer Science
Brazilian Agricultural Research Corporation
E-mail: marcos.santos-silva@embrapa.br

ABSTRACT: This study aimed to manipulate databases from the Landsat-8 project to select orbital scenes with minimal cloud coverage aligned with the growth periods of cotton, corn, and soybean crops in the Cerrado biome. Employing an Extract, Transform, Load (ETL) approach, the research focused on extracting and transforming large datasets by integrating georeferenced information from IBGE and data from the Agricultural Risk Zoning (ZARC). The planting windows for the 2014 to 2019 harvests were identified with a 20% or lower climate risk, considering variables such as soil and climate conditions. The results indicated that the availability of images with low cloud coverage is inversely related to the planting windows, especially during critical periods. To streamline the processing, algorithms were developed in R and Python, and a script for automatic image downloading was implemented. Despite data standardization challenges, applying big data techniques proved essential for the analysis. This study enhances efficiency in agricultural planning and emphasizes the importance of further exploring remote sensing technologies in response to current food security and sustainability demands.

KEYWORDS: mapping, classification, spatial data, agricultural sustainability.

RESUMO: Este trabalho teve como objetivo a manipulação de grandes bases de dados do projeto Landsat-8 para selecionar cenas orbitais com menor cobertura de nuvens, correspondendo ao período de desenvolvimento das culturas de algodão, milho e soja no bioma do Cerrado. Utilizando uma abordagem de Extração, Transformação e Carregamento (ETL), o estudo focou na extração e transformação de grandes bases de dados, integrando informações georreferenciadas do IBGE e dados do Zoneamento Agrícola de Risco Climático

(ZARC). As janelas de plantio para as safras de 2014 a 2019 foram identificadas com um risco climático de 20% ou menos, considerando variáveis como solo e clima. Os resultados mostraram que a disponibilidade de imagens com baixa cobertura de nuvens é inversamente proporcional às janelas de cultivo, especialmente durante períodos críticos. Para facilitar o processamento, foram desenvolvidos algoritmos em R e Python, e um script para download automático de imagens com baixa cobertura de nuvens foi implementado. Apesar dos desafios na padronização dos dados, este estudo teve como objetivo contribuir para a eficiência do planejamento agrícola e destacar a importância de continuar a explorar tecnologias de sensoriamento remoto diante das demandas atuais por segurança alimentar e sustentabilidade.

PALAVRAS-CHAVE: mapeamento, classificação, dados espaciais, sustentabilidade agrícola.

1. INTRODUCTION

The evolution of Earth observation sensors, installed on space and aerial platforms, has generated an immense amount of data that is revolutionizing agricultural monitoring. Daily, this information feeds various applications in remote sensing, especially in precision agriculture, which benefits from the increased availability and diversity of sensors. This transformation is vital as it enables more efficient management of crops, promotes sustainable agricultural practices, and maximizes productivity.

The applications of remote sensing in agriculture are vast and varied. Monitoring techniques have proven effective in distinguishing crops, tracking plant growth, and estimating soil moisture. Additionally, the ability to obtain real-time information about crop conditions provides a solid foundation for strategic decisions related to food security and economic management. Early detection of stress in crops, for example, is crucial for farmers to react quickly and implement corrective actions that minimize losses.

However, challenges persist in the field of remote sensing, such as limitations in the spatial and temporal resolution of images and cloud interference. To address these barriers, various innovative approaches are being explored. Data fusion from multiple sources and the use of deep learning techniques have emerged as promising strategies to enhance the quality of the information obtained. Furthermore, integrating data from different sensors with local knowledge is essential for a more comprehensive and effective analysis.

In this context, this work aims to manipulate and analyze large databases from the Landsat-8 project, focusing on selecting orbital scenes with minimal cloud coverage during critical development periods for cotton, corn, and soybean crops in the Cerrado biome. The research seeks to implement a structured Extraction, Transformation, and Loading (ETL) approach to integrate and process georeferenced information, alongside data from the Agricultural Zoning of Climate Risk (ZARC) and metadata from IBGE, in order to identify planting windows with a climate risk of less than 20% between the harvests of 2014 to 2019. Additionally, the study aims to develop algorithms in R and Python to automate the download of images, allowing for a more efficient process for

monitoring crops and improving agricultural management practices in light of contemporary demands for food security and sustainability.

2. THEORETICAL FRAMEWORK

A range of Earth observation sensors embedded in space and aerial platforms generates vast databases every day, enabling applications in various research fields. Among these areas, agricultural remote sensing stands out, benefiting from the increasing launch of new sensors that expand observation possibilities and promote the success of precision agriculture (Huang et al., 2018). This technological evolution is crucial as it allows for a more efficient approach to crop monitoring.

In this context, Kingra et al. (2016) analyzed the various applications of remote sensing techniques in agriculture, including crop discrimination, growth monitoring, and soil moisture estimation, among others. Obtaining timely and reliable information about cultivated areas and their growth conditions can be extremely beneficial for producers and planners, facilitating strategic decisions regarding food security and economic management.

Additionally, Hazaymeh et al. (2016) addressed the methods used to monitor agricultural drought, highlighting the importance of remote sensing as a powerful tool for capturing spatial dynamics over large areas. They identified methods such as optical and thermal remote sensing, which have proven effective compared to in-situ approaches that, while accurate, do not offer the same spatial coverage.

Complementing these perspectives, Huang et al. (2016) discussed low-altitude remote sensing systems, focusing on detecting stress in crops, which is vital for precision agriculture. Early detection of stress allows for a quicker and more effective response from farmers, contributing to maximized productivity.

The review by Bégué et al. (2018) on remote sensing for mapping cultivation practices reinforces this idea, revealing that many investigations still rely on limited local data. This suggests a growing need for research that combines data from multiple sensors with expert knowledge for a more comprehensive and effective analysis.

Sanches et al. (2018) contributed a multi-temporal and multi-sensor reference database, which can be a valuable tool for mapping agricultural land use. With in-situ data collection and images from orbital sensors, it was possible to create a solid foundation for analyzing key annual crops, such as those occurring in Luís Eduardo Magalhães (LEM), a crucial agricultural area in Brazil.

Building on such advancements, Feyisaa et al. (2020) highlighted the limitations of accuracy in mapping agro-ecosystem complexes, particularly in small agricultural properties. They suggested that participatory classification and the use of vegetation indices could be effective strategies for addressing the diversity of cropping systems, emphasizing the importance of local community participation.

However, challenges such as spatial and temporal resolution and cloud coverage continue to be significant barriers. To address these limitations, scholars like Shen et al. (2021) tested space-time fusion models to improve the quality of NDVI data, demonstrating a continued commitment to innovation in remote sensing techniques.

Qin et al. (2021) explored the use of optical and thermal remote sensing for monitoring droughts, pointing out chlorophyll fluorescence as a promising indicator for early detection. Analysis of future directions, including the integration of multi-source data and deep learning techniques with knowledge transfer, suggests an evolutionary path for research in this area.

Ali et al. (2023) advanced further by employing a hyperspectral imaging system to identify different agricultural species. The use of advanced classifiers and dimensionality reduction shows promise for achieving high levels of accuracy in crop classification.

To conclude this chapter, it is essential to recognize that the integration of advanced remote sensing technologies not only enhances our understanding of agricultural dynamics but also significantly supports sustainable farming practices. By leveraging diverse datasets and innovative methodologies, researchers and practitioners can make informed decisions that ultimately contribute to food security and efficient resource management in agriculture.

3. METHODOLOGY

For planning, management, and monitoring projects at a regional scale that require medium spatial resolution images, the Landsat product portfolio is widely used. The launch of the Landsat-8 satellite not only continued the historical series of 40 years but also expanded data collection by including new spectral bands (443 nm, 1370 nm, 10895 nm, and 12000 nm) and improving the sensor's signal-to-noise performance and radiometric resolution (Roy et al., 2014; Vermote et al., 2016). This vast amount of data generated by the Landsat-8 sensors results in a large volume of metadata, essential for research and scene selection, reflecting the so-called "4Vs" (volume, velocity, variety, and veracity) that characterize the manipulation of large databases. This study focuses on the first stage of this process, addressing the extraction and transformation of massive datasets, both georeferenced and non-georeferenced, through automation in R and Python.

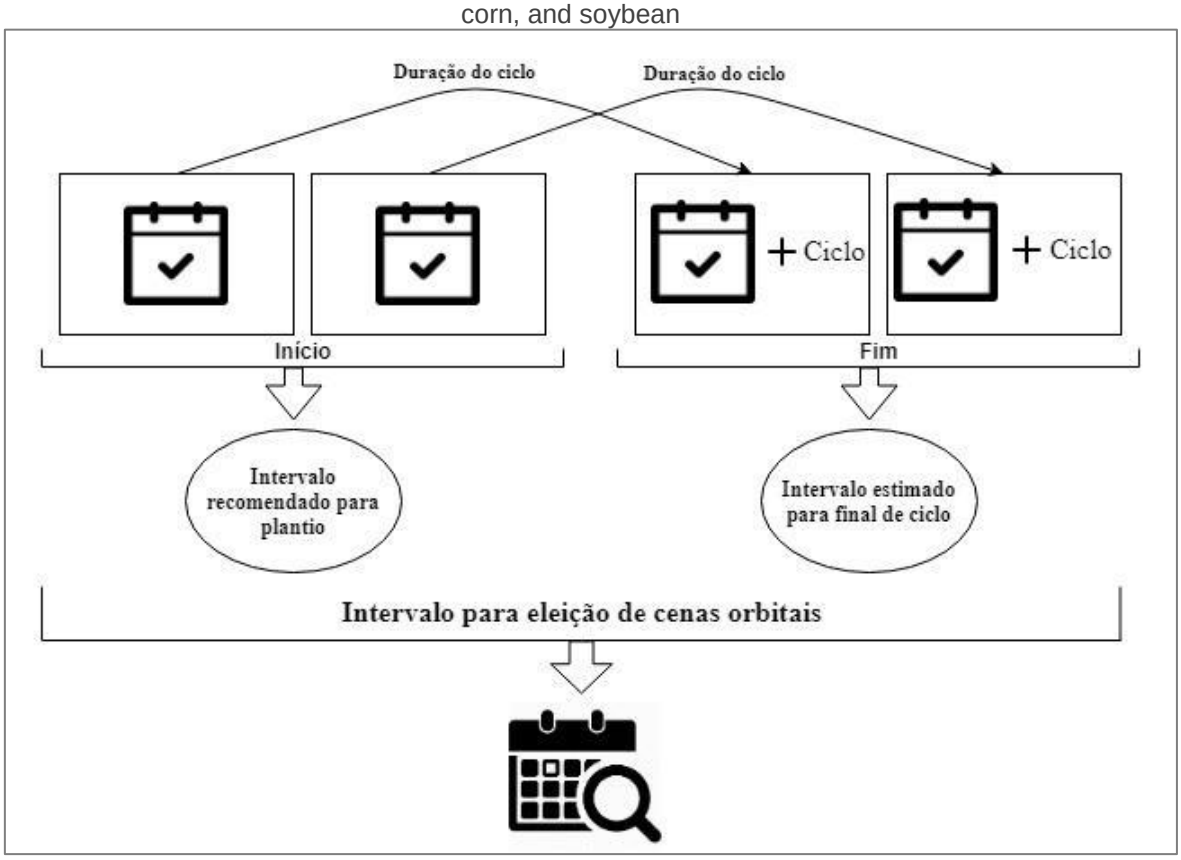
The study area was defined using the vector database of municipal boundaries provided by IBGE, encompassing all Brazilian municipalities within the Cerrado. For these municipalities, the planting windows for soybean, corn, and cotton crops from 2014 to 2019 were identified, with associated climate risk lower than 20%, using the Agricultural Zoning of Climate Risk (ZARC) (Assad et al., 2018) provided by MAPA (2018), whose analysis considered different soil types, climatological normals, and cultivar cycles to estimate crop development in each harvest.

Additionally, the vector database of Landsat-8 paths (Worldwide Reference System – WRS2) and the metadata of images associated with the selected tiles (USGS, 2018) were used to filter orbital scenes, taking into account atmospheric interference for the mosaic coverage of the area of interest and the time series. The compatibility between political-administrative boundaries and the boundaries of the Cerrado biome presented challenges, resulting in information gaps, especially in municipalities without intersections. To ensure the contiguity of the layer, it was necessary to select and incorporate these polygons.

The first product was a vector file containing the municipalities of interest, from which the municipal geocode from IBGE was extracted, serving as a primary key for subsequent joins with other databases (ZARC and the Landsat-8

metadata catalog). The ZARC database underwent filtering, selection, field standardization, and data completion through coding in R (RStudio). The selection of soybean, corn, and cotton cultivars for the 2014 to 2019 harvests was based on the planting windows defined by planting dates and climate risk, as illustrated in Figure 1.

Figure 1 – Schematic diagram used for calculating the planting window for herbaceous cotton, corn, and soybean



Source: created by the authors.

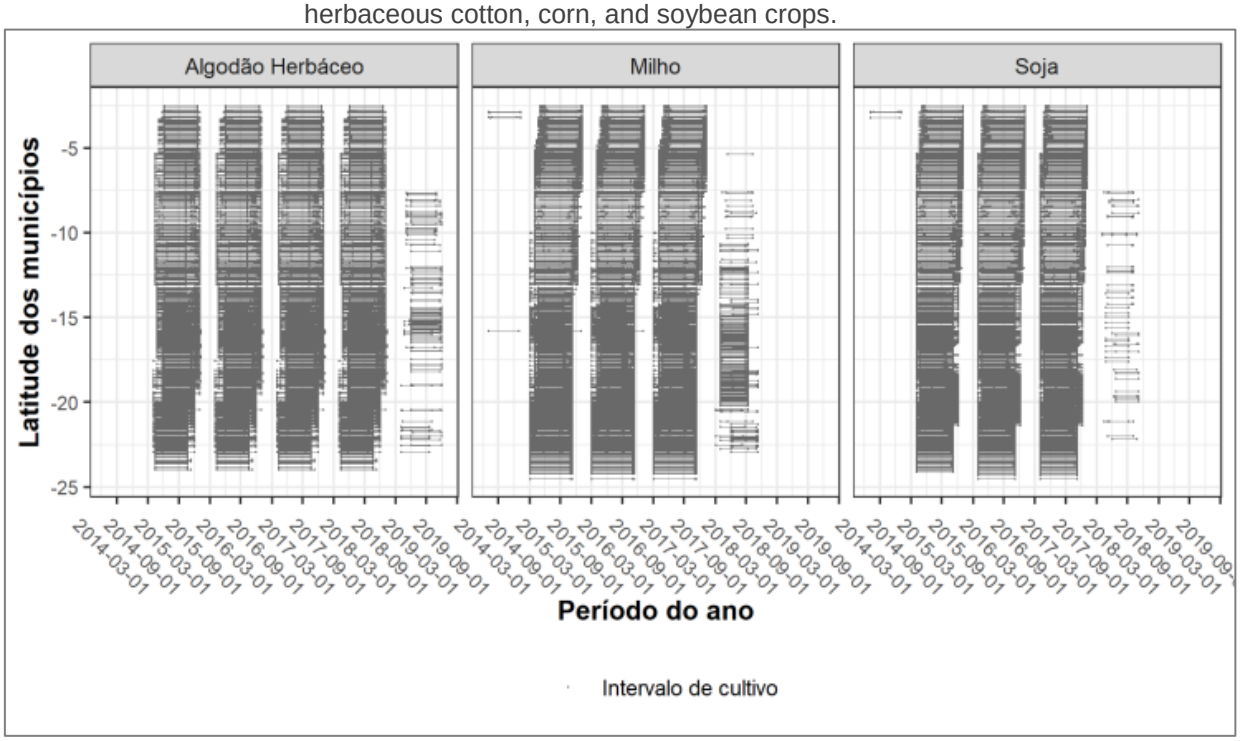
To determine the crop cycles, the average number of days based on the literature was considered: 125 days for soybean (Carneiro et al., 2014; Zito et al., 2017), 130 days for corn (Cruz et al., 2014), and 150 days for herbaceous cotton (Marur et al., 2003). Data processing followed the ETL (Extraction, Transformation, and Loading) method, using algorithms developed in R and Python through the R Studio and Eclipse IDEs, as well as ArcGIS software (Esri, 2018), along with specific libraries and functions.

4. RESULTS AND DISCUSSIONS

The main database manipulated was that of the orbital scenes related to the OLI/TIRS sensor. The mosaic of various scenes for a predefined area allows for representation of large expanses (Guimarães, 2015). After meeting the requirements regarding the time intervals (01/01/2014 – 20/09/2018), spatial coverage (municipal base within the Cerrado biome), and cloud coverage (10%), a database with the metadata of the scenes was generated, which underwent adjustment and filtering operations.

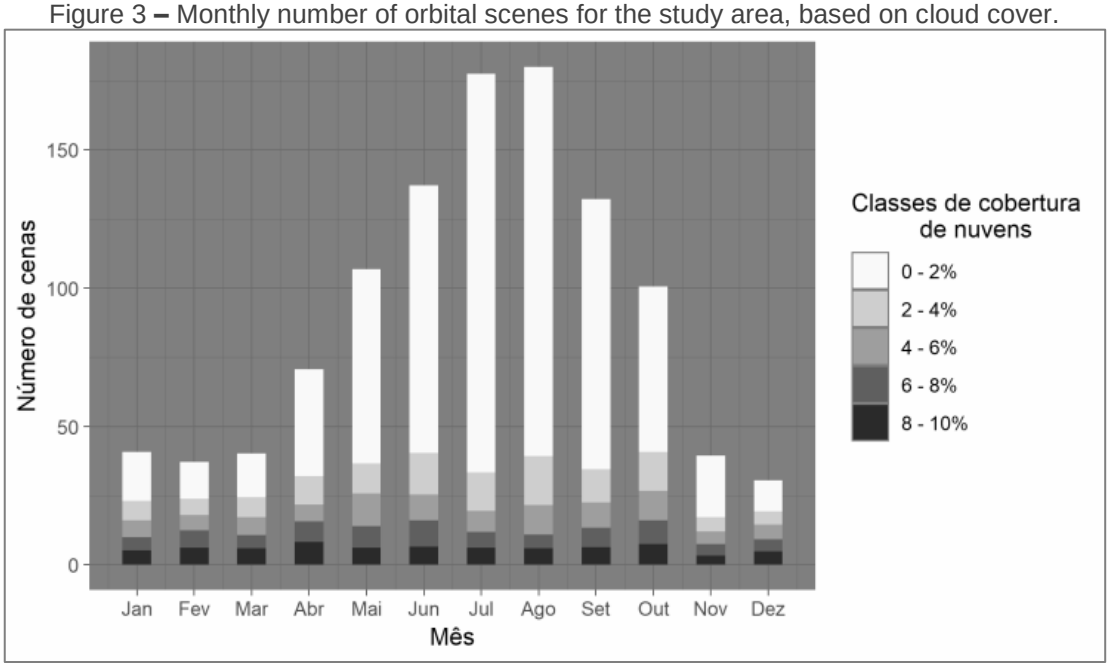
As an initial result, Figure 2 shows the composition of information regarding the recommended time interval for each crop based on latitude. Subsequently, Figure 3 indicates that orbital scenes with lower cloud coverage (0 to 2%) are concentrated between the months of May and October, based on the monthly grouping of eligible scenes.

Figure 2 – Cultivation interval calculated based on the planting recommendation by ZARC for herbaceous cotton, corn, and soybean crops.



Source: Research Results.

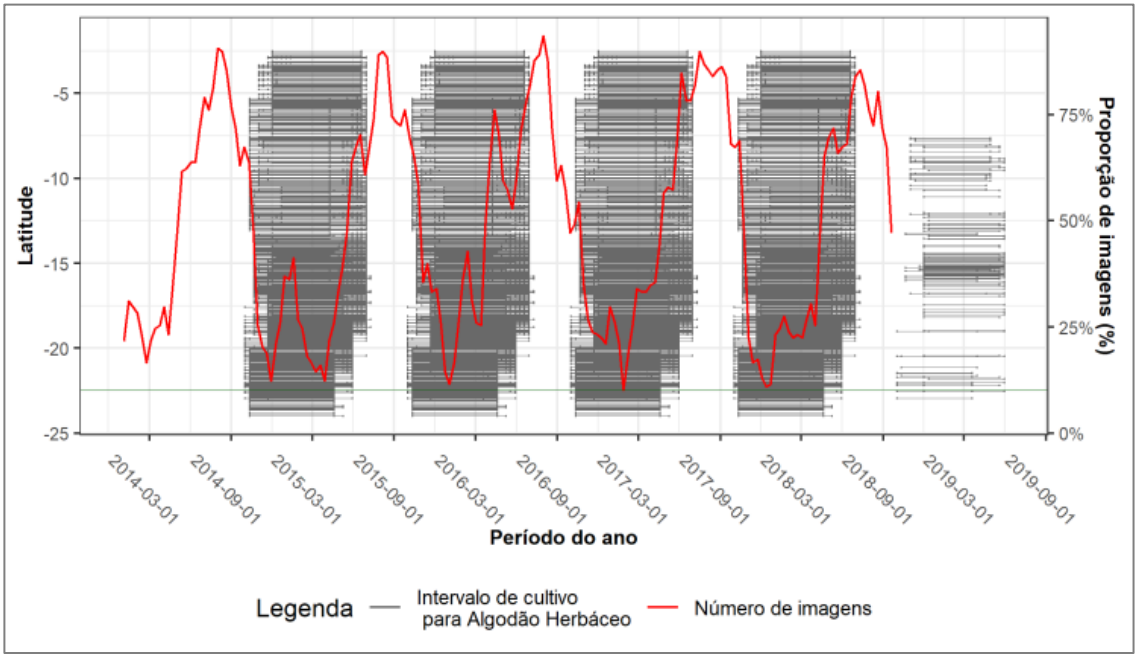
Figure 3 – Monthly number of orbital scenes for the study area, based on cloud cover.



Source: Research Results.

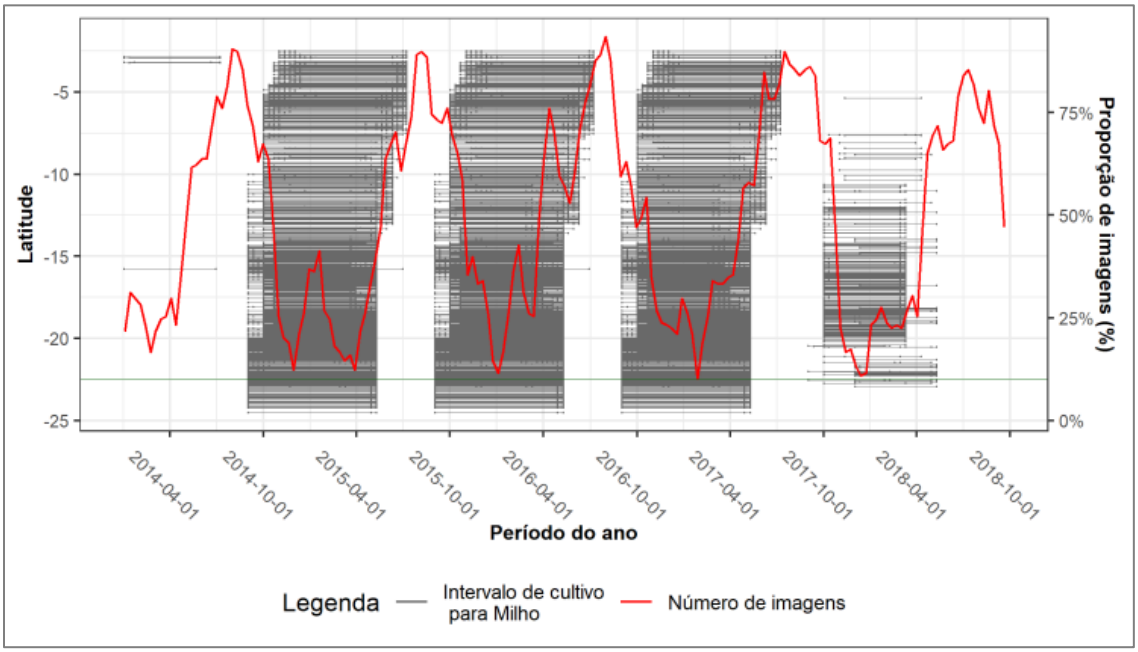
The plotting of data associating the number of eligible images from the time series with the growing windows obtained from the ZARC data processing provides an overview of the available material suitable for the project based on latitude (Figures 4 to 6). It is noted that the availability of orbital scenes does not coincide with the cultivation window across nearly the entire latitude range.

Figure 4 – Annual composition of the availability of images with low cloud cover, based on the interval for herbaceous cotton.



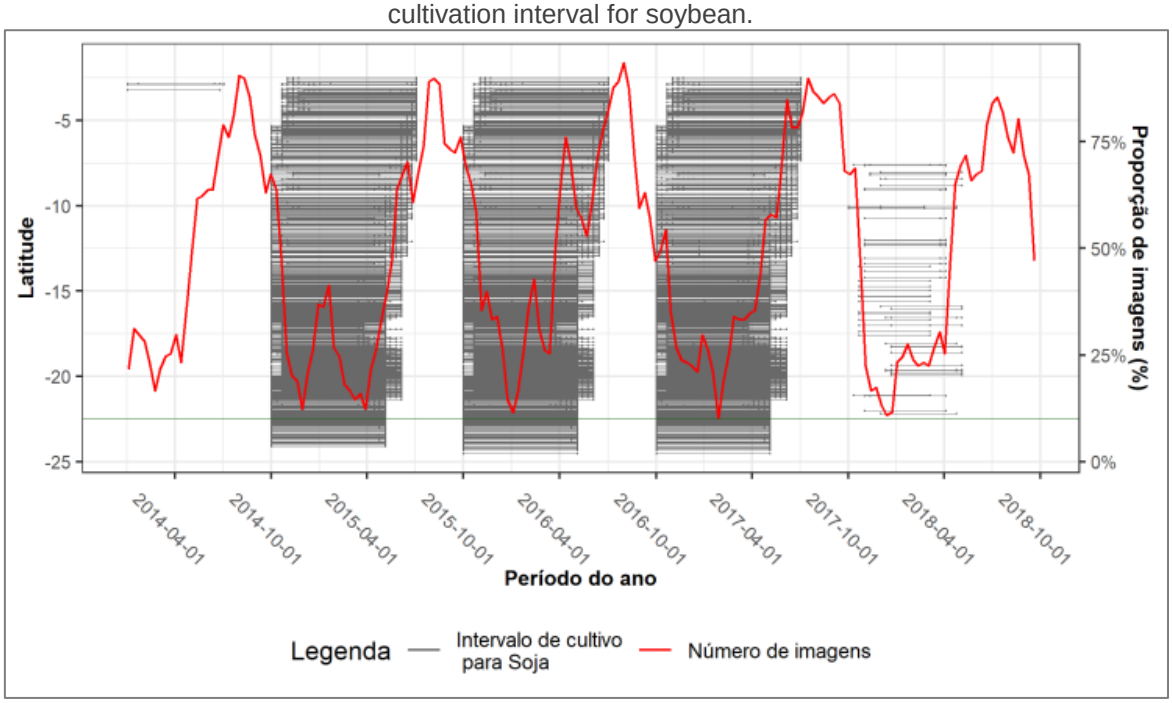
Source: Research Results.

Figure 5 – Annual composition of the availability of images with low cloud cover, based on the cultivation interval for corn.



Source: Research Results.

Figure 6 – Annual composition of the availability of images with low cloud cover, based on the cultivation interval for soybean.



Source: Research Results.

A Python script (using the landsatxplore package) was developed for the automatic download of scenes from the USGS (Earth Explorer) platform based on the metadata catalog. With the available images, a new stage of data

extraction will begin, involving pre-processing, classification, and post-processing of the eligible orbital scenes.

In the initial data treatment process, transformation proved to be one of the most labor-intensive phases in manipulating the databases, as it involved diagnosing issues, standardizing, normalizing, and aggregating variables, as well as applying filters and completing records. Due to the use of mixed databases (tabular and georeferenced vector data) from various sources, numerous discrepancies were identified, such as non-standardized municipality names, incomplete and/or inconsistent data.

Despite the challenges in homogenizing the databases, the application of the aforementioned techniques derived from the big data paradigm proved essential in generating consistent and decisive directives in a relatively short time for subsequent stages of the project. Numerous challenges and opportunities surround the use of these techniques and should be further explored in the field of remote sensing (Mingmin, 2016).

5. CONCLUSION

The conclusion of this work highlights the importance of systematically manipulating large databases, particularly in the context of the Landsat-8 project, for monitoring cotton, corn, and soybean crops in the Cerrado biome. Through a structured Extraction, Transformation, and Loading (ETL) approach, it was possible to identify and select orbital scenes with the least cloud cover, coinciding with critical phenological development periods of the crops. This process is essential, as atmospheric conditions can significantly impact image quality, directly affecting analysis by specialists and agricultural planning.

The results obtained demonstrate an inverse relationship between the availability of images with low cloud cover and the growing windows for the analyzed crops. It was observed that, across nearly the entire considered latitude range, the supply of orbital scenes is limited during the most critical planting and development periods. This fact emphasizes the importance of efficient data processing techniques that can address various challenges, such as inconsistencies in metadata and the heterogeneity of data sources. The use of automation algorithms developed in R and Python proved crucial for conducting

analyses in a reduced timeframe, enabling quicker and more informed decision-making.

Additionally, the implementation of a script for the automatic download of scenes from the USGS (Earth Explorer) platform represents an advancement that aids workflow in subsequent stages of pre-processing, classification, and post-processing of images. The initial manipulation of data proved complex, requiring standardization, normalization, and completion of variables.

Finally, the research suggests that the techniques and methods presented here have the potential to positively impact agricultural planning, promoting a more efficient and sustainable management of crops. The challenges and opportunities that arise from applying big data technologies in remote sensing should be further explored, underscoring the need for a commitment to research and development in this increasingly relevant field in light of climate change and the growing demand for food. Thus, this work paves the way for future investigations that can enrich our understanding of the interaction between technology, the environment, and agricultural practices.

REFERENCES

ALI, Imran; MUSHTAQ, Zain; ARIF, Shahid; ALGARNI, Abdul D.; SOLIMAN, Nasser F.; EL-SHAFAI, Wael. *Hyperspectral images-based crop classification scheme for agricultural remote sensing*. Computer Systems Science & Engineering, 2023. DOI: 10.32604/csse.2023.034374.

ASSAD, Edson D.; MARIN, Fabiano R.; PINTO, Hugo S.; ZULLO JÚNIOR, João. *Zoneamento agrícola de riscos climáticos do Brasil: base teórica, pesquisa e desenvolvimento*. Informe Agropecuário, Belo Horizonte, v. 29, n. 246, p. 47-60, set./out. 2008.

BALA, Mohd; BOUSSAID, Othmane; ALIMAZIGHI, Z. *A Fine-Grained Distribution Approach for ETL Processes in Big Data Environments*. Data & Knowledge Engineering, v. 111, p. 114-136, 2017.

BÉGUÉ, Agnès; DÉROST, S.; MARCHAND, P.; LÉONARD, J.; CANTARINO, J. Remote sensing and cropping practices: a review. *Remote Sensing*, v. 10, p. 99, 2018. DOI: 10.3390/rs10010099. Disponível em: www.mdpi.com/journal/remotesensing. Acesso em: 10 ago. 2018.

CARNEIRO, Gabriel E. de S.; CAVALCANTE, Rafael A.; BASSO, Carlos; OLIVEIRA, Alexandre R.; FREITAS, Ricardo A. *Cultivares de soja: macrorregiões 1, 2 e 3 Centro-Sul do Brasil*. Londrina: Embrapa Soja, 2014. 60 p.

CRUZ, José C.; PEREIRA FILHO, Ismael A.; SIMÃO, Edson de P. *478 cultivares de milho estão disponíveis no mercado de sementes do Brasil para a safra 2014/2015*. Sete Lagoas: Embrapa Milho e Sorgo, 2014. 35 p. (Embrapa Milho e Sorgo. Documentos, 167).

ECLIPSE FOUNDATION. Eclipse IDE for Java Developers. Disponível em: https://www.eclipse.org. Acesso em: 10 ago. 2018.

ESRI. ArcGIS Desktop: Release 10.6. 2017. Disponível em: https://www.esri.com/en-us/arcgis/products/arcgis-desktop/overview. Acesso em: 10 ago. 2018.

FEYISAA, Gudina Legese; MAZAR, Elham G.; AHMED, Alhaji; KABIR, Abdul; FAKHREDDINE, Rabih; BATAINEH, Ahmad; AL-MASRI, Mohamad. Characterizing and mapping cropping patterns in a complex agro-ecosystem: an iterative participatory mapping procedure using machine learning algorithms and MODIS vegetation indices. *Computers and Electronics in Agriculture*, v. 175, 2020. DOI: https://doi.org/10.1016/j.compag.2020.105595.

GUIMARÃES, Daniel P.; LANDAU, Edson C.; SOUSA, Diego L. *Mosaicos de imagens Landsat-8 dos estados brasileiros*. Sete Lagoas: Embrapa Milho e Sorgo, 2015. 38 p. (Embrapa Milho e Sorgo. Documentos, 186).

HAZAYMEH, Khaled; HASSAN, Quazi K. Remote sensing of agricultural drought monitoring: a state of art review. *AIMS Environmental Science*, v. 3, n. 4, p. 604-630, 2016. DOI: 10.3934/environsci.2016.4.604. Recebido em: 01 jul. 2016. Aceito em: 26 set. 2016. Publicado em: 29 set. 2016. Disponível em: http://www.aimspress.com/journal/environmental.

HUANG, Yanbo; HAO, Jian; YANG, Yanfeng; YAN, Jianjun; ZHOU, Wensheng. Development and evaluation of low-altitude remote sensing systems for crop

production management. *International Journal of Agricultural and Biological Engineering*, v. 9, n. 4, p. 1-10, jul. 2016. Disponível em: http://www.ijabe.org. Acesso em: 10 ago. 2018.

KINGRA, Pritpal K.; MAJUMDER, Debashis; SINGH, Surinder P. *Application of remote sensing and GIS in agriculture and natural resource management under changing climatic conditions*. Agricultural Research Journal, v. 53, n. 3, p. 295-302, set. 2016. DOI: 10.5958/2395-146X.2016.00058.2.

MARUR, Carlos J.; RUANO, O. *Escala do Algodão*. Informe da Pesquisa, v. 105, n. 1, p. 1-4, 2003.

MEHMOOD, Maryam; AZHAR, Muhammad; SHAH, Syed A.; ALHUSSIENI, Ali A.; AHMED, Faisal; ASHRAF, Muhammad. Remote sensing image classification: a comprehensive review and applications. *Mathematical Problems in Engineering*, v. 2022, Article ID 5880959, 24 p., 2022. DOI: 10.1155/2022/5880959.

MINISTÉRIO DA AGRICULTURA, PECUÁRIA E ABASTECIMENTO (MAPA). Zoneamento Agrícola do Risco Climático. Disponível em: http://indicadores.agricultura.gov.br/zarc/index.htm. Acesso em: 01 jul. 2018.

CHI, Mingmin A.; WEI, Xiaodong; ZHANG, Zhiqiang; LI, Hongli; ZHANG, Lei; WANG, Jian. Big Data for Remote Sensing: Challenges and Opportunities. *IEEE*, v. 104, n. 11, p. 2207-2219, 2016.

PYTHON SOFTWARE FOUNDATION. Python Language Reference, version 3.x. Disponível em: https://www.python.org/. Acesso em: 10 ago. 2018.

QIN, Qiming; DONG, Wei; ZHANG, Xu; LIU, Qi; TANG, Jian; SUN, Huilin. Optical and thermal remote sensing for monitoring agricultural drought. *Remote Sensing*, v. 13, n. 24, p. 5092, 2021. DOI: https://doi.org/10.3390/rs13245092.

ROY, David P.; AVERY, Scott; BAE, Dae-Kwang; BOUFFARD, J. D.; BRODRICK, A.; DUNN, C. E.; HANKS, C.; HURT, C.; IYER, H.; KIM, S.; KUKAL, M.; MAGEE, J. W.; MARTIN, M.; NARAYANAN, M.; PEZZO, S.; PINKER, R.; REESE, J. J.; RIEHL, J.; RUBIN, S.; SILVA, C.; SMITH, L. C.; STANTON, S.; STROM, J.; TOMPKINS, E.; VALENTINE, H.; WISDOM, S. Landsat-8: Science and product vision for terrestrial global change research. *Remote Sensing of Environment*, v. 145, p. 154-172, 2014.

R CORE TEAM. R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing, 2023. Disponível em: https://www.R-project.org. Acesso em: 10 ago. 2018.

R STUDIO, INC. RStudio: integrated development environment for R. 2023. Disponível em: https://www.rstudio.com. Acesso em: 10 ago. 2018.

SANCHES, Igor D.; REIS, Eduardo D.; ZUIN, Ana; BARRETO, Carlos A.; GOMES, Carlos R.; MELO, Rogério S.; TAVARES, Claudete M.; PEREIRA, Luiz H. L. *LEM benchmark database for tropical agricultural remote sensing application*. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, v. XLII-1, p. 1-8, 2018. ISPRS TC I Mid-term Symposium “Innovative Sensing – From Sensors to Methods and Applications”, 10–12 out. 2018, Karlsruhe, Alemanha.

UNITED STATES GEOLOGICAL SURVEY (USGS). *Landsat 8 OLI/TIRS Scene*. [Imagem]. Disponível em: https://earthexplorer.usgs.gov/

SHEN, Yonglin; KANG, Shoubo; KONG, Jinsong; ZHANG, Yuying; LI, Jian; LI, Xiaohong; LIU, Yichao; LIU, Xiangyun; LI, Zhaoyang. A Gaussian kernel-based spatiotemporal fusion model for agricultural remote sensing monitoring. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, v. 14, p. 3533, 2021.

Agência Brasileira ISBN
ISBN: 978-65-83227-08-9