MANIPULAÇÃO DE BASES MASSIVAS DE DADOS EM SENSORIAMENTO REMOTO AGRÍCOLA

Pedro Alves Quilici Coutinho[1], Márcia Helena Galina Dompieri[2], Hilton Luís Ferraz da Silveira[3], Paulo Roberto Rodrigues Martinho[4], Jaudete Daltio[5], Mário Balan[6]

Embrapa Territorial: [1]pedro.coutinho@colaborador.embrapa.br; [2] marcia.dompieri@embrapa.br; [3] hilton,ferraz@embrapa.br; [4]paulo.martinho@embrapa.br; [5]jaudete.daltio@embrapa.br, [6]mario.balan@colaborador.embrapa.br

RESUMO

Por meio do emprego de técnicas e processos relacionados à manipulação de bases massivas de dados, houve o tratamento de dados/metadados advindos de cenas orbitais e de três grandes complexos agropecuários: algodão herbáceo, milho e soja, com o objetivo de se eleger as cenas mais adequadas para o estudo. A delimitação da base político-administrativa municipal a partir do limite natural do bioma permitiu a seleção das janelas de cultivos em função da latitude dos municípios para as cadeias supracitadas, com base nas recomendações técnicas de plantio do Zoneamento Agrícola do Risco Climático (ZARC). Posteriormente, procedeu-se com operações de filtragem de metadados de cenas orbitais (Landsat8-OLI/TIRS). Foram utilizados os softwares ArcGis 10.6, R (R-Studio) e codificação em Python. Constatou-se que o emprego de técnicas adequadas na manipulação de bases massivas de dados permite a obtenção de informações mais consistentes, num tempo relativamente curto, para subsidiar a tomada de decisões em projetos.

Palavras-chave — Landsat-8, R, Python, Big-Data.

ABSTRACT

Using techniques and processes related to the manipulation of massive database, data/metadata from orbital scenes and three large agricultural complexes (herbaceous cotton, corn and soybean) were processed. The purpose was select more suitable scenes for the study. The delimitation of the municipal political-administrative from the natural limit of the biome allowed the selection of the windows of crops, according to the latitude of the municipalities for the chains, based on the technical recommendations of planting of the Agricultural Zoning of Climatic Risk (ZARC). Subsequently, operations were performed to filter metadata from orbital scenes (Landsat8-OLI / TIRS). We used the software ArcGis 10.6, R (R-Studio) and coding in Python. It was verified that the use of adequate techniques in the manipulation of massive database allows the obtaining of more consistent information to subsidize decisions in projects, in a relatively short time.

Key words — *Landsat-8, R, Python ,Big-Data.*

1. INTRODUÇÃO

Uma gama de sensores de observação da Terra embarcados em plataformas espaciais e aéreos geram, todos os dias, enormes bases de dados, que possibilitam aplicações nas mais variadas áreas de investigação. Uma destas áreas é o sensoriamento remoto agrícola que tem se beneficiado do crescente lançamento de novos sensores, permitindo maiores possibilidades de observação e o sucesso da agricultura de precisão [1]. No entanto, a cobertura de nuvens sobre os alvos agrícolas pode inviabilizar o trabalho e o processo seleção de cenas mais apropriadas torna-se um desafio.

Para projetos na escala regional de planejamento, gestão e monitoramento de alvos na superfície terrestre, que necessitam de imagens de média resolução espacial, o portfólio de produtos do projeto Landsat é largamente utilizado. O projeto Landsat-8 permitiu, além da continuidade da série histórica de 40 anos, a ampliação da obtenção dos dados ao incluir novas bandas espectrais (443 nm, 1370 nm,10895 nm e 12000nm) além de aperfeiçoar o desempenho sinal/ruído do sensor e a resolução radiométrica [2, 3].

Toda essa massa de dados gerado pelos sensores a bordo da plataforma Landsat-8 repercute no elevado volume de informações de metadados disponíveis para a pesquisa e seleção de cenas. Dentre os então chamados "4Vs" (volume, velocidade, variedade e veracidade) [4] que caracterizam as fases de manipulação de grandes bases de dados para entregar valor. Neste estudo procurou-se focar na primeira etapa desse processo que lida com a questão do volume, sobretudo quanto à extração e transformação de bases massivas (georreferenciadas e não georreferenciadas) por meio de um processo automatizado - codificação em R e Python.

O escopo do presente trabalho foi a manipulação de bases de dados para a eleição de cenas orbitais advindas do projeto Landsat-8, com as menores coberturas de nuvens, coincidentes com o período de desenvolvimento fenológico das culturas do algodão herbáceo, milho e soja para os municípios brasileiros pertencentes ao bioma do Cerrado.

2. MATERIAIS E MÉTODOS

A partir da base de dados vetoriais georreferenciadas dos limites municipais disponibilizadas pelo IBGE [5] a área de estudo foi definida como a totalidade dos municípios brasileiros que tivessem inserção no Cerrado. Para estes municípios, as janelas de plantio para as três culturas selecionadas (soja, milho e algodão) nas safras consideradas (2014 à 2019) com risco associado menor que 20% foram identificadas por meio do Zoneamento Agrícola de Risco Climático (ZARC) [6], disponibilizado pelo MAPA[7]. Assim, em função dos diferentes tipos de solo, normais climatológicas, e ciclos de cultivares foi possível estimar o desenvolvimento das culturas dentro de cada safra.

Também foi empregada a base vetorial com rotas do satélite Landsat-8 (Worldwide Reference System –WRS2), assim como os metadados das imagens associados às quadrículas selecionadas [8], a partir da qual foi realizada a diretiva para a filtragem das cenas orbitais, que também considerou a porcentagem de interferência atmosférica (nuvens), para fins de cobertura do mosaico da área de interesse e da série temporal em questão.

Procedeu-se com o método para tratamento de grandes bases de dados, conhecido como ETL (Extração, Transformação e Carregamento), por meio da construção de algoritmos por meio das linguagens de programação R 3.5.0 e Python por meios dos IDES (*Integrated development environment*) Rstudio 1.1.453 e Eclipse, além do uso de softwares GIS (ArcGis v. 6), bibliotecas, pacotes e funções específicas.

3. RESULTADOS

3.1- Definição do limite político-administrativo municipal

Compatibilizar os limites político-administrativos e os limites da ocorrência de um fenômeno natural, como um bioma, nem sempre é uma tarefa trivial, mas torna-se necessária uma vez que as bases de dados oficiais são disponibilizadas considerando essa unidade básica de análise. A intersecção dos limites do Cerrado e dos municípios, resultou em um plano de informação com *gaps*, representados pelos municípios encravados sem interseção. A fim de garantir a contiguidade dessa camada, houve o processo de seleção e incorporação desses polígonos.

O primeiro produto correspondeu ao arquivo vetorial com os municípios de interesse na análise, a partir do qual foi extraído o campo correspondente ao geocódigo municipal definido pelo IBGE e que serviu como chave primária, para posterior junção com outras bases (ZARC, catálogo de metadados Landsat-8).

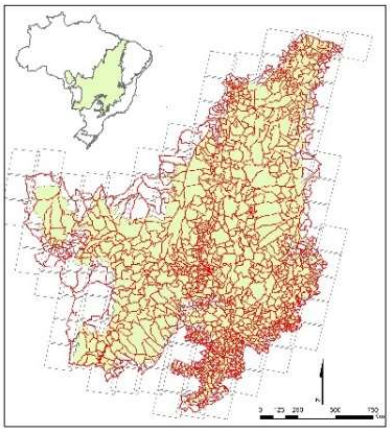

Figura 1 – Malha municipal com interseção do Bioma Cerrado e da grade de cenas Landsat-8

3.2- Composição do intervalo de cultivo em função da latitude para as cadeias agrícolas de interesse

A base de dados advinda do ZARC passou pelas operações de filtragem, seleção, padronização de campos e complementação de dados por meio de codificação no software R (RStudio). Houve a seleção dos cultivares de algodão, milho e soja para as safras 2014-2015; 2015-2016; 2016-2017; 2017-2018 e 2018-2019 e para os municípios selecionados na etapa anterior. Foram definidas as janelas de cultivo (calculadas a partir das datas de início e fim do plantio para um risco climático de 20%) de acordo com o ciclo fenológico de cada cultura, como ilustra a Figura 2.

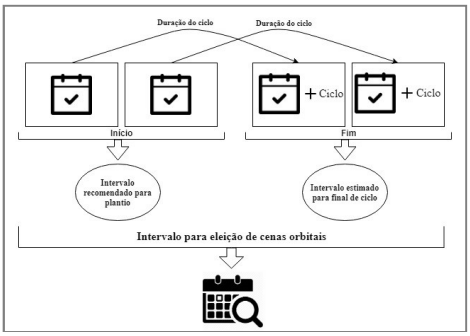

Figura 2 – Diagrama esquemático utilizado para o cálculo da janela de cultivo do algodão herbáceo, milho e soja

Para definição do ciclo dos cultivos, é sabido que inúmeros fatores contribuem para sua variação, como efeito de aproximação optou-se pela adoção da quantia média de dias com base na literatura consultada. No caso da soja adotou-se o período de 125 dias [9,10]; para o milho, 130 dias [11] e por fim, para o algodão herbáceo, 150 dias [12].

Como resultado, por meio da Figura 3, observa-se a composição das informações referente ao intervalo temporal recomendado para cada cultivo em função da latitude.

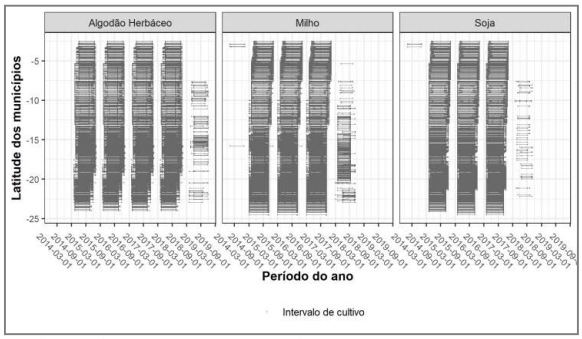

Figura 3 – Intervalo de cultivo calculado com base na recomendação de plantio pelo ZARC, para as culturas de algodão herbáceo, milho e soja

3.3- Filtragem das cenas orbitais e associação com a janela têmporo-espacial de cultivos

A base principal manipulada foi a das cenas orbitais referentes ao sensor OLI/TIRS. O mosaico de diversas cenas para uma área pré-definida permite uma representação sinótica de grandes extensões [13]. Preenchidos os requisitos quanto aos intervalos temporal (01/01/2014 – 20/09/2018), espacial (base municipal com cobertura do Bioma Cerrados) e de cobertura de nuvens (10%), houve a geração de uma base com os metadados das cenas, que foi submetida a operações de ajustes e filtragens.

Pela Figura 4, a partir do agrupamento mensal das cenas elegíveis, constata-se que cenas orbitais com menor cobertura de nuvens (0 a 2%) concentram-se entre os meses de maio a outubro.

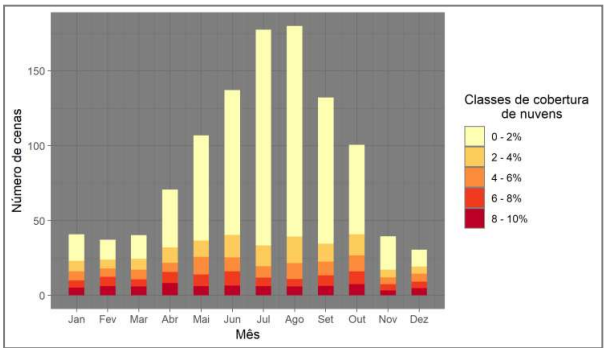

Figura 4 – Quantidade de cenas orbitais mensais para a área de estudo, em função da cobertura de nuvens

A plotagem dos dados associando a quantidade de imagens elegíveis da série temporal com as janelas de cultivo, obtidas pelo processamento dos dados do ZARC, fornece o quadro disponível do material adequado para o projeto em função da latitude (Figuras 5, 6 e 7). Nota-se que, a disponibilidade das cenas orbitais é inversamente proporcional à janela de recomendação de cultivos ao longo de quase toda a faixa latitudinal.

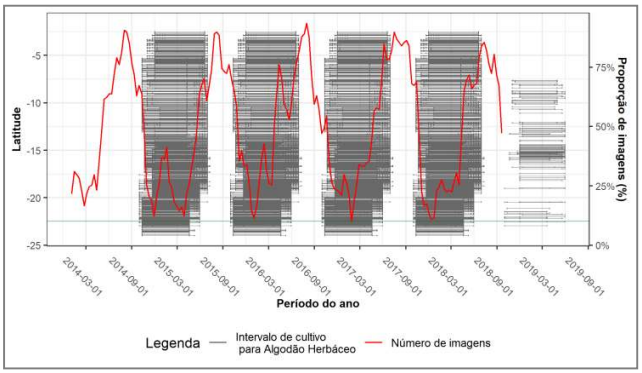

Figura 5 – Composição anual da disponibilidade de imagens com baixa cobertura de nuvens, em função do intervalo do algodão herbáceo

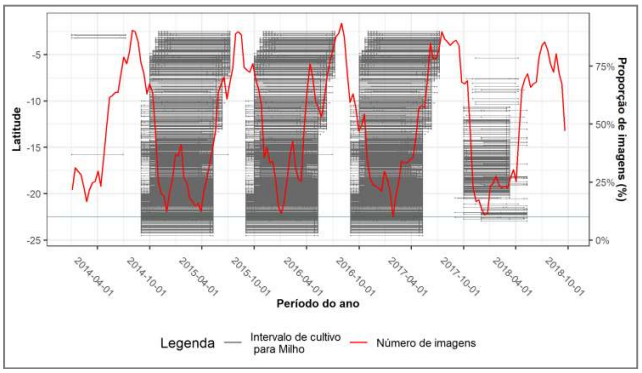

Figura 6 – Composição anual da disponibilidade de imagens com baixa cobertura de nuvens, em função do intervalo de cultivo do milho

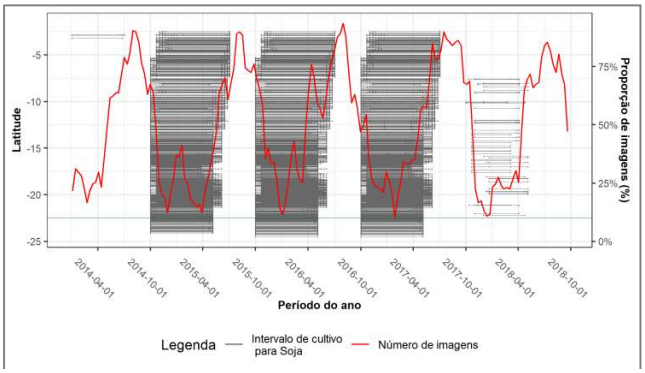

Figura 7 – Composição anual da disponibilidade de imagens com baixa cobertura de nuvens, em função do intervalo de cultivo da soja

Por fim, construiu-se uma codificação em python (pacote *landsatxplore*), com base no catálogo de metadados, para download automático das cenas a partir da plataforma da USGS (Earth Explorer). Uma vez com as imagens disponíveis, uma nova etapa de extração de dados deve se iniciar, por meio do pré-processamento, classificação e pós-processamento das cenas orbitais elegíveis.

4. DISCUSSÃO

No processo de tratamento inicial dos dados, a transformação se mostrou uma das fases mais trabalhosas na manipulação das bases, pois envolveu o diagnóstico de problemas, a padronização, normalização e agregação de variáveis, além de aplicação de filtragens e complementação de registros. Por conta da utilização de bases mistas (tabulares e vetoriais georreferenciadas) advindas de várias fontes, inúmeras desconformidades foram identificadas, como nomes de municípios despadronizados, dados incompletos e/ou inconsistentes.

Apesar dos esforços na homogeneização das bases, a aplicação das referidas técnicas advindas do paradigma *big data* se mostraram essenciais na geração de diretivas decisórias e consistentes, num tempo relativamente curto, para etapas posteriores do projeto. Inúmeros desafios e oportunidades envolvem o emprego das referidas técnicas, e deveriam ser mais explorados na área do sensoriamento remoto [14].

5. CONCLUSÕES

O emprego de técnicas adequadas na manipulação de bases massivas de dados permite a obtenção de informações consistentes num tempo relativamente curto, subsidiando tomada de decisões em projetos.

Os resultados mostraram que há desafios reais na utilização de produtos orbitais para a agricultura, sobretudo quando envolvem sensores passivos pois a necessidade de chuva para o desenvolvimento dos cultivos (algodão herbáceo, milho e soja) coincide com a época de maior interferência atmosférica. Assim, no período do ano em que há maior necessidade de produtos orbitais, ocorre também a menor oferta de cenas que satisfaçam os requisitos mínimos para o processo de extração de informações.

6. REFERÊNCIAS

[1] HUANG, Y.; CHEN, Z.; YU, T.; HUANG X., GU, X. Agricultural remote sensing big data: management and applications, **Journal of Integrative Agriculture**, v. 17, n. 9, 2018, p. 1915-1931.

[2] ROY D.P., WULDER M.A., LOVELAND T.R., WOODCOCK C.E., ALLEN R.G., ANDERSON M.C., HELDER D., IRONS J.R., JOHNSON D.M., KENNEDY R., SCAMBOS T.A., SCHAAF C.B., SCHOTT J.R., SHENG Y., VERMOTE E.F., BELWARD A.S., BINDSCHADLER R., COHEN W.B., GAO F., HIPPLE J.D., HOSTERT P., HUNTINGTON J., JUSTICE C.O., KILIC A., KOVALSKYY V., LEE Z.P., LYMBURNER L., MASEK J.G., MCCORKEL J., SHUAI Y., TREZZA R., VOGELMANN J., WYNNE R.H., ZHU Z. Landsat-8: Science and product vision for terrestrial global change research**, Remote Sensing of Environment**, v. 145, 2014, p. 154-172

[3] VERMOTE E., JUSTICE C., CLAVERIE M., FRANCH. B. Preliminary analysis of the performance of the Landsat 8/OLI land surface reflectance product, **Remote Sensing of Environment**, v. 185, 2016, p. 46-56

[4] BALA M., BOUSSAID O., ALIMAZIGHI. Z. A Fine-Grained Distribution Approach for ETL Processes in Big Data Environments, **Data & Knowledge Engineering**, v. 111, 2017, p. 114-136

[5]INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (IBGE). Disponível em: https://mapas.ibge.gov.br/bases-e-referenciais/bases-cartograficas/malhas-digitais Acesso em: 13 jul. 2018

[6] ASSAD, E. D.; MARIN, F. R.; PINTO, H. S.; ZULLO JÚNIOR, J. 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.

[7] 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. de 2018

[8] UNITED STATES GEOLOGICAL SURVEY (USGS). **EarthExplorer.** Disponível em: <https://earthexplorer.usgs.gov> Acesso em: 10 Ago de 2018

[9] CARNEIRO, G. E. de S.; PIPOLO, A. E.; MELO, C. L. P. de; LIMA, D. de; FOLONI, J. S. S.; MIRANDA, L. C.; PETEK, M. R.; BORGES, R. de S.; GOMIDE, F. B.; DALBOSCO, M.; DENGLER, R. U. **Cultivares de soja: macrorregiões 1, 2 e 3 Centro-Sul do Brasil**. Londrina: Embrapa Soja, 2014. 60 p.

[10] ZITO, R. K.; MELLO FILHO, O. L. de; PEREIRA, M. J. Z.; MEYER, M. C.; HIROSE, E.; FIDELIS, A. C.; NUNES JUNIOR, J.; VIEIRA, N. E.; SEII, A. H.; PIMENTA, C. B.; SANCHEZ, I.; MOREIRA, A. J. A.; NUNES, M. R.; DESSIMONE, M. G. L.; SENE, B. R. A. de; SOUSA, R. C.; NEIVA, L. C. S. **Cultivares de soja: macrorregiões 3, 4 e 5 Goiás e Região Central do Brasil.** Londrina: Embrapa Soja, 2017. 48 p.

[11] CRUZ, J. C.; PEREIRA FILHO, I. A.; SIMÃO, E. 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).

[12] MARUR, C. J.; RUANO, O. (2003). Escala do Algodão. **Informe da Pesquisa**, v. 105, n.1, p.1-4, 2003

[13] GUIMARAES, D. P.; LANDAU, E. C.; SOUSA, D. L. **Mosaicos de imagens Landsat-8 dos estados brasileiros**. Sete Lagoas: Embrapa Milho e Sorgo, 2015. 38 p. (Embrapa Milho e Sorgo. Documentos, 186).

[14] MINGMIN CHI, A.; PLAZA, J. A. ; BENEDIKTSSON, J. A.; ZHONGYI SUN, J. A.; JINSHENG SHEN, J. A.; YANGYONG ZHU, J. A. **Big Data for Remote Sensing: Challenges and Opportunities**. *IEEE*, v. 104, n. 11, 2016, p. 2207-2219