

Um ambiente para análise exploratória de grandes volumes de dados geoespaciais: explorando risco de fogo e focos de queimadas

Vitor Gomes^{1,2}, Gilberto Ribeiro de Queiroz², Karine Ferreira²,
Luciane Yumie Sato², Rafael Santos², Fabiano Morelli²

¹Instituto de Estudos Avançados – Departamento de Ciência e Tecnologia Aeroespacial
CEP 12.228-001 – São José dos Campos – SP – Brasil

² Instituto Nacional de Pesquisas Espaciais
Caixa Postal 515 – CEP 12.227-010 – São José dos Campos – SP – Brasil

vitor@ieav.cta.br, {gilberto.queiroz, karine.ferreira}@inpe.br
{luciane.sato, rafael.santos, fabiano.morelli}@inpe.br

Abstract. *This paper presents an environment for exploratory analysis of large volumes of geospatial data. We have employed innovative technologies that support the storage, access, and exploratory analysis of geographic data as multidimensional raster and vector representation. In addition, the environment architecture is structured so that other phases of the research cycle can be carried out by reusing part of the established infrastructure. A preliminary analysis with meteorological data and vegetation fires was conducted in this environment to show its viability. We show also the results of the analysis indicating a variation in the fire risk values associated with the vegetation fires when segregated by biome, region or climatic season.*

Resumo. *Este trabalho apresenta um ambiente para análise exploratória de grandes volumes de dados geoespaciais. Utilizamos tecnologias inovadoras que suportam o armazenamento, o acesso e a análise exploratória de dados matriciais e vetoriais. Além disso, a arquitetura do ambiente é estruturada para que outras fases do ciclo de pesquisa possam ser realizadas reaproveitando parte da infraestrutura estabelecida. Uma análise preliminar com dados meteorológicos e focos de queimadas foi realizada neste ambiente para mostrar sua viabilidade. Apresentamos os resultados da análise, os quais indicam uma variação nos valores de risco de incêndio associados aos focos de queimadas quando segregados por bioma, região ou estação climática.*

1. Introdução

Nos últimos anos, a quantidade disponível de dados geoespaciais tem crescido. Se por um lado a disponibilidade dessas informações tem permitido que novos avanços científicos e tecnológicos possam acontecer, o armazenamento, o acesso e a utilização desses dados de forma eficiente representa um desafio devido a sua multidimensionalidade, densidade e seu grande volume.

Nas áreas de Observação da Terra (EO, do inglês *Earth Observation*) e Meteorologia, a maioria dos métodos de análise de dados atualmente é baseada em arquivos.

Em aplicações que demandam grandes volumes de dados, os usuários precisam obter centenas (ou milhares) de arquivos, para que um programa possa extrair e armazenar as informações relevantes na memória do computador ou em arquivos intermediários. Com o aumento da quantidade de dados, o uso desse tipo de abordagem será cada vez menos eficaz, dependendo maior tempo com a organização dos dados. De maneira geral, esta prática tem colocado limites severos sobre os usos científicos de dados de EO [Camara et al. 2016, Cudre-Mauroux et al. 2009].

A busca por atender essas demandas tem estimulado o desenvolvimento de novas tecnologias. Algumas soluções, como extensões espaciais para Sistemas Gerenciadores de Banco de Dados (SGBD) [Ramsey 2016, Ramsey et al. 2016] e SGBD Matriciais (SGBD-M) [Baumann et al. 1998, Cudre-Mauroux et al. 2009] visam atender o armazenamento e gestão de grandes volumes e/ou estruturas de dados específicas. Outras, estão voltadas para a forma de acesso aos dados, com o uso de interfaces padronizadas [OGC 2017] ou abordagens específicas [Queiroz et al. 2015, Vinhas et al. 2016]. Ainda, há soluções que integram tecnologias para fornecer plataformas para a análise de dados de EO, como Google Earth Engine [Gorelick et al. 2017], a qual utiliza a infraestrutura do Google para o armazenamento e processamento dos dados, e arquiteturas baseadas em ferramentas abertas [Camara et al. 2016].

Na arquitetura apresentada por [Camara et al. 2016], os dados são armazenados em um SGBD-M e são usadas ferramentas de análise estatística, produzindo uma solução onde os algoritmos são executados juntamente aos dados, aproveitando a infraestrutura já estabelecida e otimizando o uso dos recursos. Nesse trabalho, é comparado o uso de abordagem MapReduce e matrizes multidimensionais. Apesar dos resultados de desempenho serem equivalentes, quando considerados o custo de organização dos dados, a abordagem baseada em SGBD-M é, segundo os autores, superior para o problema avaliado.

É importante observar que pesquisadores preferem testar novas ideias em conjuntos pequenos de dados, utilizando ferramentas conhecidas, antes de moverem seu trabalho para ambientes de computação massiva e de grandes volumes de dados [Camara et al. 2016, Vinhas et al. 2016]. No contexto desse trabalho, grandes volumes de dados são considerados os conjuntos de dados que excedem a capacidade de armazenamento ou processamento de hardware e software convencionais [Guo et al. 2017].

Na fase em que cientistas realizam a Análise Exploratória de Dados (EDA, do inglês *Exploratory Data Analysis*), por vezes, nem todo volume de dados disponível é necessário para que se possam entender melhor os dados ou obter relações preliminares entre variáveis. Além disso, há demanda que as soluções desta natureza forneçam suporte para o ciclo completo de pesquisa, permitindo que algoritmos desenvolvidos em computadores pessoais possam ser aplicados em grandes volumes de dados com poucas alterações [Camara et al. 2016, Vinhas et al. 2016].

Neste contexto e considerando a diversidade de tecnologias disponíveis, este trabalho visa apresentar um ambiente para a análise exploratória de grandes volumes de dados geoespaciais. O objetivo desse trabalho é avançar o ambiente computacional descrito em [Camara et al. 2016], integrando dados representados por matrizes multidimensionais (séries temporais) com dados vetoriais para análise exploratória do lado cliente, voltado para a linguagem Python. Nas demais seções desse trabalho, são apresen-

tados a estruturação do ambiente (seção 2), uma API desenvolvida para acesso aos dados (seção 2.1) e uma análise preliminar realizada com dados de queimadas (seção 3). Além disso, são apresentados os resultados iniciais (seção 3.1) e as considerações finais (seção 4).

2. Ambiente para Análise Exploratória de Dados

O ambiente apresentado neste trabalho considera a combinação de tecnologias para: i) o armazenamento; ii) o acesso; e iii) a análise exploratória de dados geoespaciais. Para a estruturação desse ambiente e a seleção das tecnologias, é importante considerar a natureza dos dados e as necessidades dos cientistas.

De maneira geral, dados geoespaciais são representados nas formas matricial e vetorial. A representação matricial, também conhecida como *raster*, é caracterizada por uma grade regular onde cada célula está associada a um ou mais valores. Dados como imagens de satélites e variáveis meteorológicas são tradicionalmente representados por matrizes multidimensionais. Essas matrizes são chamadas de *coverages* quando representam fenômenos no espaço-tempo. A representação vetorial, por sua vez, utiliza pontos, linhas e polígonos para representar feições (*features*) únicas localizadas no espaço e que possuem atributos não espaciais. Eventos pontuais, rios e limites políticos são exemplos de elementos frequentemente representados em formato vetorial.

Além da intenção de permitir acesso a essas duas representações, partimos da premissa que após a fase EDA, o pesquisador tem interesse em realizar análises completas sobre todo o volume de dados disponível. Desta forma, tomamos como referência a arquitetura proposta por [Camara et al. 2016] para o armazenamento de dados *raster* utilizando SGBD-M. O objetivo, nesse caso, é que as demais fases do ciclo de pesquisa possam ser realizadas usando a abordagem proposta pelos autores, sem a necessidade de reorganizar os dados em outra infraestrutura. O SGBD-M selecionado para esta tarefa foi o SciDB. Esse é um sistema no estado da arte, considerado um dos mais promissores SGBD-M da atualidade [Rusu and Cheng 2013]. Para o armazenamento de dados vetoriais, optou-se pelo uso do SGBD PostgreSQL com a extensão PostGIS [Ramsey et al. 2016], pois é a solução que provê a implementação mais completa da especificação OGC *Simple Feature* entre as opções *open source* disponíveis [Steiniger and Hunter 2012].

Para o acesso aos dados, são utilizados serviços web. As *coverages* armazenadas no SciDB são acessadas através de serviço WTSS (*Web Time Series Service*) [Queiroz et al. 2015] e para as *features* optou-se pelo OGC WFS (*Web Feature Service*) [OGC 2017]. O WTSS é um serviço web leve para manusear dados de séries temporais de imagens de sensoriamento remoto, possuindo implementação com suporte ao SciDB. O WTSS faz uso do formato JSON para o intercâmbio de dados e possui clientes em linguagens de programação R, C++, JavaScript e Python. A implementação do WTSS utilizada neste trabalho é a versão disponível na plataforma *open source* EOWS¹. O WFS é um padrão criado pelo OGC para criação, atualização e intercâmbio de informações geográficas em formato vetorial [OGC 2017]. Nesse padrão, os dados são intercambiados em formato XML, existindo implementações que também suportam o formato JSON em requisições específicas, como é caso do GeoServer, utilizado neste trabalho.

¹Repositório disponível em <https://github.com/e-sensing/eows>

Cientistas são conservadores na escolha das ferramentas para a análise de dados. Eles preferem trabalhar com ferramentas simples que permitem novos métodos analíticos com a adição de novos pacotes [Camara et al. 2016, Vinhas et al. 2016]. O ambiente de desenvolvimento R e a linguagem de alto nível Python são escolhas frequentes de analistas de dados por contarem com uma variedade de ferramentas estatísticas, gráficas e de análise numérica.

Para nosso ambiente, foi escolhida a linguagem Python. A escolha é motivada pela sua crescente popularidade na comunidade científica [Wagner et al. 2017] e a disponibilidade de pacotes para computação científica (NumPy² e SciPy library³), manipulação e análise de dados (Pandas⁴ e GeoPandas⁵ para dados geoespaciais) e visualização (Matplotlib⁶ e Seaborn⁷).

Para facilitar o acesso aos dados via WTSS e WFS, foi desenvolvida uma API, chamada `simple_geo.py`, para abstrair a construção de consultas e tratar os formatos utilizados pelos dois serviços. A subseção 2.1 apresenta os detalhes da API desenvolvida.

A Figura 1 apresenta um diagrama esquemático dos componentes do ambiente estruturado neste trabalho.

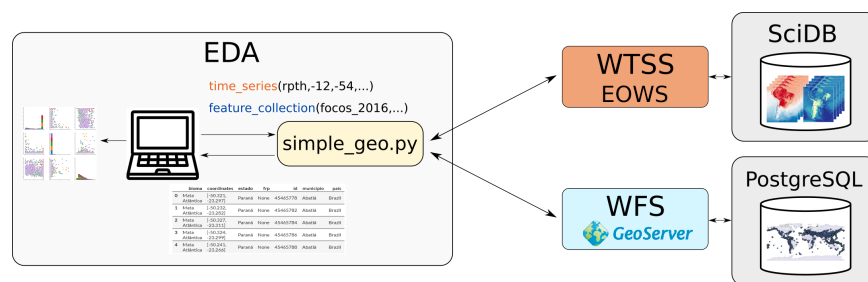


Figura 1. Diagrama esquemático do ambiente para análise exploratória de dados geoespaciais

2.1. API `simple_geo.py`

A API de acesso aos dados foi construída como um pacote Python que funciona como uma fachada (façade) para pacotes que realizam a interface com os serviços WTSS e WFS. Para o WTSS, foi utilizado o pacote WTSS.py⁸. Para o WFS, foi implementado um cliente para a construção de consultas e o tratamento dos retornos em XML e JSON. Para dados vetoriais, o `simple_geo.py` retorna dados na forma de `GeoDataFrame` e para `coverages` o retorno é um `DataFrame`, respectivamente das bibliotecas `GeoPandas` e `Pandas`. O pacote `simple_geo.py` fornece sete operações:

- `list_features`: lista as *features* disponíveis;

²NumPy: <http://www.numpy.org>

³SciPy library: <https://www.scipy.org/scipylib>

⁴Pandas: <http://pandas.pydata.org>

⁵GeoPandas: <http://geopandas.org>

⁶Matplotlib: <http://matplotlib.org>

⁷Seaborn: <https://seaborn.pydata.org>

⁸Repositório disponível em <https://github.com/e-sensing/wtss.py>

- `describe_feature`: obtém a descrição de uma *feature* selecionada;
- `feature_collection`: obtém os elementos de uma *feature* selecionada;
- `feature_collection.len`: obtém o número de elementos de uma *feature*;
- `list_coverage`: lista as *coverages* disponíveis;
- `describe_coverage`: obtém a descrição de uma *coverage* selecionada;
- `time_series`: obtém a série temporal de uma *coverage* selecionada.

A Figura 2 mostra um exemplo de uso do `simple_geo.py` para a obtenção de um conjunto de *features* (linha 6). Além da opção de filtragem por atributo (`filter`), é possível limitar a região de busca (`within`), ordenar os dados (`order`), limitar o número de feições a serem retornadas (`max_features`) e selecionar quais atributos devem ser obtidos (`attributes`). Na linha 8, é recuperada uma série temporal com três atributos da *coverage climatologia*. Para esse tipo de dados, é possível especificar o período no qual se deseja obter a série temporal, usando os parâmetros `start_date` e `end_date`. Mais detalhes sobre o `simple_geo.py` e seu código fonte estão disponíveis online⁹.

```

1 from simple_geo import simple_geo as sgeo
2 # Conectando aos servidores
3 s = sgeo(wfs="http://server:8080/geoserver",
4         wtss="http://server:7654")
5 # Obtendo features
6 ft, ft_metadata = s.feature_collection("focos", filter=["regiao='N'"])
7 # Obtendo serie temporal de coverage
8 ts, ts_metadata = s.time_series("climatologia", ("prec", "temp", "umid"),
                                -12, -54)

```

Figura 2. Uso do `simple_geo.py` para obtenção de feições e série temporal

3. Análise

A necessidade de estruturar um ambiente para a análise exploratória de grandes volumes de dados geoespaciais surgiu do interesse em avaliar o índice risco de fogo [Setzer and Sismanoglu 2012] associado aos focos de queimadas detectados pelo Programa de Monitoramento de Queimadas do INPE. O risco de fogo é um índice produzido pelo INPE que indica o quão propícia a vegetação está para ser queimada. Os focos de queimadas são detectados a partir da análise de um conjunto de imagens de satélite. Cada foco detectado é registrado em um banco de dados com informações sobre sua posição, data e horário, sensor utilizado na detecção, entre outras variáveis de interesse dos especialistas.

Preliminarmente, nosso interesse é avaliar as condições de risco de fogo no momento da ocorrência dos focos, a fim de verificar a correlação entre o índice e os eventos, e verificar se essa relação tem variação espacial (biomas e regiões) e/ou variação temporal (estações climáticas). Esse tipo de análise pode contribuir para que futuras evoluções do cálculo do índice de risco de queimadas considerem, além das variáveis ambientais utilizadas, informações a respeito da região/bioma e estações do ano.

Para a realização dessa análise, o ambiente descrito na seção 2 foi implantado. Os SGBD SciDB e PostgreSQL e os serviços EOWS (WTSS) e GeoServer (WFS) foram

⁹Repositório disponível em https://github.com/e-sensing/simple_geo.py

instalados, por conveniência, em um único servidor. Destaca-se que esses sistemas poderiam ser distribuídos em um conjunto de máquinas, pois comunicam-se via rede, o que nos fornece capacidade futura de escalar o desempenho do ambiente. Em um computador pessoal, foi instalado um interpretador Python, o pacote `simple_geo.py` e os pacotes de análise e visualização de dados mencionados na seção 2. Além desses, configurou-se o ambiente colaborativo de análise Jupyter Notebook [Kluyver et al. 2016], o qual permite a elaboração de documentos que combinam programas, textos formatados e figuras.

Seguindo a abordagem de testar novas ideias em conjuntos pequenos de dados, foram preparados dados do ano de 2016 para esta análise inicial. A Tabela 1 apresenta um resumo dos dados carregados no ambiente. As *coverages* foram agrupadas em 4 *arrays* e carregadas no SciDB. Os dados vetoriais de focos de queimadas foram carregados no PostgreSQL. Uma coluna com valores aleatórios foi incluída através da construção de uma visão (*view*) para permitir a ordenação aleatória dos dados e que distintos subconjuntos possam ser recuperados. O conjunto de dados ocupa 25GB.

Table 1. Resumo dos dados carregados no ambiente

Dado	Res. Temporal	Res. Espacial	Dimensões	Tipo	Serviço
Cobertura da Terra IGBP 2012	1 cena	5km	1200x1400	Coverage	WTSS
Temperatura	diário (366 cenas)	5km	1200x1400	Coverage	WTSS
Umidade	diário (366 cenas)	5km	1200x1400	Coverage	WTSS
Risco de Fogo	diário (366 cenas)	5km	1200x1400	Coverage	WTSS
Precipitação	diário (366 cenas)	5km	1200x1400	Coverage	WTSS
Ocorrência mensal de focos	mensal (12 cenas)	5km	1200x1400	Coverage	WTSS
Temperatura média normal	mensal (12 cenas)	20km	226x196	Coverage	WTSS
Precipitação média normal	mensal (12 cenas)	20km	226x196	Coverage	WTSS
Umidade relativa média normal	mensal (12 cenas)	20km	226x196	Coverage	WTSS
Focos de Queimadas	2.039.394 registros			Feature	WFS

Para essa fase de EDA, foi construído um Jupyter Notebook onde foram avaliadas correlações entre variáveis, ocorrência e distribuição do índice de risco associado a focos de queimadas por região, bioma e estações climáticas. Foi utilizado o pacote Seaborn para a elaboração de visualizações gráficas, como histogramas, *boxplot*, mapas de calor e gráficos de dispersão. A estruturação dos dados para a análise iniciou pela recuperação de 500 focos aleatórios para cada uma das 5 regiões do país. Na sequência, para cada foco, foram obtidos, através do método `time_series`, valores de risco, temperatura, umidade e precipitação para a localização e dia de ocorrência do foco. Todos esses dados foram consolidados em um único *DataFrame*, para facilitar a realização da análise estatística e a produção de gráficos. O documento com a análise preliminar encontra-se disponível online¹⁰.

3.1. Resultados Preliminares

As figuras numeradas de 3 a 8 apresentam alguns dos resultados produzidos. Além desses gráficos, também foram gerados matrizes de correlação, *boxplots* e matrizes de dispersão agrupando os dados por região, bioma e estação climática.

Observando os histogramas apresentados na Figura 3, verificamos que a maior parte dos focos acontece com risco igual a 1, como esperado. Entretanto, há focos

¹⁰Disponível em: https://github.com/e-sensing/simple_geo.py/blob/master/docs/eda.ipynb

que ocorrem em situações onde a previsão de risco é inferior a esse valor, possuindo ocorrências com valores próximos a zero. Observando os histogramas para cada região (Figura 4) detectamos que isso é mais destacado para a região Sul. As Figuras 5, 7 e 8 reforçam essa observação. No verão, por exemplo, a maior parte dos focos da região Sul ocorreram quando o índice de risco era inferior a 0,4.

Outra observação que podemos extrair quando comparamos as Figuras 7 e 8 é quanto a diferença entre a ocorrência dos focos e o índice associado a eles quando comparadas as duas estações. Os focos das regiões Nordeste, Centro-Oeste e Sudeste estão associados a riscos mais altos no inverno em relação aos observados no verão.

Além disso, observamos que nos biomas Amazônia, Mata Atlântica e Pampa (Figura 6) os focos ocorrem em uma maior amplitude do índice de risco quando comparados aos demais biomas.

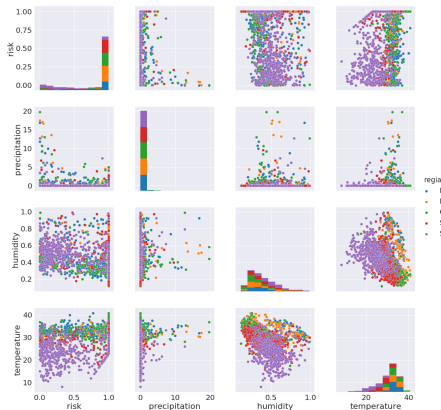


Figura 3. Matriz de dispersão por região

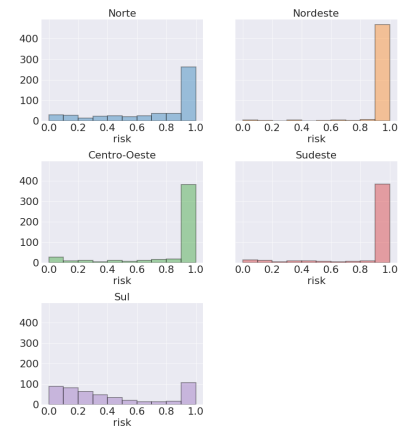


Figura 4. Histogramas de risco associados aos focos de queimadas por região

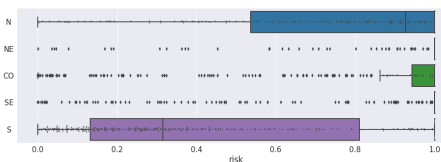


Figura 5. Risco por região

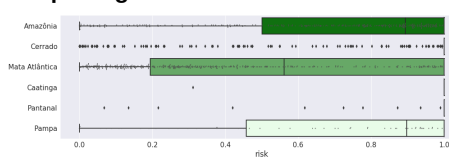


Figura 6. Risco por bioma

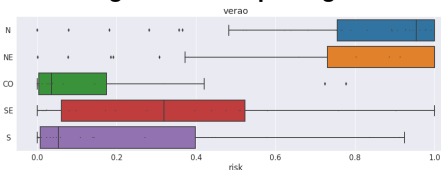


Figura 7. Risco por região no verão

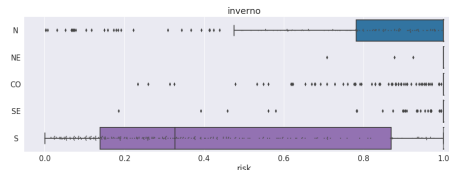


Figura 8. Risco por região no inverno

Essas primeiras observações sobre os dados nos permitem questionar se o índice representa de maneira igual a suscetibilidade à queima da vegetação nas diferentes regiões

e estações, ou se essa variação está associada a ocorrência de focos de queimada associados às práticas agrícolas, as quais variam entre as regiões e períodos do ano. Esse estudo está sob investigação e será tópico de discussão em trabalhos futuros.

4. Considerações Finais

Este trabalho apresentou a estruturação de um ambiente para a análise exploratória de grandes volumes de dados geoespaciais e as análises preliminares quanto ao índice de risco de fogo associado aos focos de queimadas. Uma ferramenta para a integração de serviços WTSS e WFS foi desenvolvida para facilitar o acesso a dados matriciais e vetoriais. Dados de meteorologia foram organizados na forma de matrizes espaço-temporais, permitindo o acesso aos dados na forma de séries temporais. Até onde temos conhecimento, não há trabalho com esse tipo de abordagem com este conjunto de dados.

Quanto ao ambiente estruturado, verifica-se que o mesmo é apropriado para o tipo de exploração que se deseja realizar, pois permite que o pesquisador recupere os dados utilizando diversas opções de seleção e subconjuntos amostrados aleatoriamente. Além disso, fazemos uso de tecnologias apropriadas para o armazenamento de cada tipo de dado geoespacial, integrando o acesso através do uso de única API. Observamos que o uso de uma API mais simples e especializada é mais efetivo, como também destacado por [Vinhas et al. 2016], pois permite que durante a análise a atenção seja dada a exploração dos dados e não a questões técnicas como protocolos e formatos de dados.

A análise realizada nos permitiu ter uma primeira visão sobre os dados e a variação do índice de risco de fogo associado aos focos quanto ao espaço (regiões e biomas) e tempo (estações). Além disso, permitiu a identificação de novas capacidades a serem incluídas no ambiente para que os dados possam ser melhor explorados. Uma das situações identificadas é a necessidade de obter todos os pontos da grade matricial onde o risco assume determinado valor ou faixa de valores. Isso permitiria verificar, por exemplo, as situações onde o risco é máximo (1) e não ocorrem focos de queimadas.

Os resultados obtidos neste trabalho nos motivam a estabelecer uma continuidade visando ampliar as capacidades do ambiente e a profundidade da análise exploratória sobre os dados de queimadas. Planejamos incluir a opção de obter as séries temporais considerando valores encontrados nas *coverages* e o suporte a realização de consultas por lote, visando aumento de desempenho. Além disso, acreditamos que incluir na API a possibilidade de recuperar dados das *coverages* associados as posições das *features* de maneira automática aumentaria a abstração quanto a estruturação interna dos dados e facilitaria o uso da API por cientistas. A inclusão do suporte a *Web Coverage Service* (WCS) e *Web Map Service* (WMS) ao ambiente facilitaria, respectivamente, a recuperação de regiões e a visualização dos dados matriciais. Por fim, esperamos mapear a probabilidade de ocorrências de focos associados ao índice de risco e as variações espacial (bioma ou região) e temporal (estação climática).

References

- Baumann, P., Dehmel, A., Furtado, P., Ritsch, R., and Widmann, N. (1998). The multidimensional database system rasdaman. *SIGMOD Rec.*, 27(2):575–577.
- Camara, G., Assis, L. F., Ribeiro, G., Ferreira, K. R., Llapa, E., and Vinhas, L. (2016). Big earth observation data analytics: Matching requirements to system architectures.

In *Proceedings of the 5th ACM SIGSPATIAL International Workshop on Analytics for Big Geospatial Data*, BigSpatial '16, pages 1–6, New York, NY, USA. ACM.

- Cudre-Mauroux, P., Kimura, H., Lim, K.-T., Rogers, J., Simakov, R., Soroush, E., Velikhov, P., Wang, D. L., Balazinska, M., Becla, J., DeWitt, D., Heath, B., Maier, D., Madden, S., Patel, J., Stonebraker, M., and Zdonik, S. (2009). A demonstration of scidb: A science-oriented dbms. *Proc. VLDB Endow.*, 2(2):1534–1537.
- Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., and Moore, R. (2017). Google earth engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*.
- Guo, H., Liu, Z., Jiang, H., Wang, C., Liu, J., and Liang, D. (2017). Big Earth Data: a new challenge and opportunity for Digital Earth's development. *International Journal of Digital Earth*, 10(1):1–12.
- Kluyver, T., Ragan-Kelley, B., Pérez, F., Granger, B., Bussonnier, M., Frederic, J., Kelley, K., Hamrick, J., Grout, J., Corlay, S., Ivanov, P., Avila, D., Abdalla, S., Willing, C., and development team [Unknown], J. (2016). Jupyter notebooks - a publishing format for reproducible computational workflows.
- OGC (2017). OGC Standards and Supporting Documents. <http://www.opengeospatial.org/standards>. Accessed: 29 sep 2017.
- Queiroz, G. R., Ferreira, K. R., Vinhas, L., Câmara, G., Costa, R. W., Souza, R. C. M., Maus, V. W., and Sanchez, A. (2015). WTSS: um serviço web para extração de séries temporais de imagens de sensoriamento remoto. In *Anais...*, pages 7553–7560, São José dos Campos. Simpósio Brasileiro de Sensoriamento Remoto, 17. (SBSR).
- Ramsey, P. (2016). Pointcloud: a PostgreSQL extension for storing point cloud (LIDAR) data. <https://github.com/pgpointcloud/pointcloud>. Accessed: 04 sep 2016.
- Ramsey, P., Santilli, S., Obe, R., Cave-Ayland, M., and Park, B. (2016). PostGIS - Spatial and Geographic objects for PostgreSQL. <http://postgis.net>. Accessed: 04 sep 2016.
- Rusu, F. and Cheng, Y. (2013). A survey on array storage, query languages, and systems. *CoRR*, abs/1302.0103.
- Setzer, A. W. and Sismanoglu, R. A. (2012). Risco de fogo: Metodologia do cálculo – descrição sucinta da versão 9.
- Steiniger, S. and Hunter, A. J. S. (2012). *Free and Open Source GIS Software for Building a Spatial Data Infrastructure*, pages 247–261. Springer Berlin Heidelberg, Berlin, Heidelberg.
- Vinhas, L., Queiroz, G. R., Ferreira, K. R., and Câmara, G. (2016). Web services for big earth observation data. In *Proceedings of Brazilian Symposium on Geoinformatics -(GEOINFO)*, Campos do Jordão, SP.
- Wagner, M., Llort, G., Mercadal, E., Giménez, J., and Labarta, J. (2017). Performance analysis of parallel python applications. *Procedia Computer Science*, 108(Supplement C):2171 – 2179. International Conference on Computational Science, ICCS 2017, 12-14 June 2017, Zurich, Switzerland.