Database of 2019 Brazilian oil disaster: an overview of a dataset and its application on an Artificial Intelligence monitoring system.

[1]Maria Paula Graziotto; [2]Laércio Massaru Namikawa
[1] -Bolsista DTI no Instituto Nacional de Pesquisas Espaciais - INPE
[2] -Tecnologista Sênior III no Instituto Nacional de Pesquisas Espaciais - INPE
{maria.graziotto@inpe.br; laercio.namikawa@inpe.br}

Abstract. In September 2019 small tar balls and oil slicks showed up affecting eleven states, mostly in the Northeast region, which depends mostly on tourism and fishery activities. The Brazilian National Contingency Plan (PNC) failed since the oil spread subsurface making impossible to see thorough satellite images. To help out to discover what happened, the Federal Government activated the Monitoring and Evaluation Group (GAA) to study different aspects of this disaster and learn how to prevent and mitigate a possible new event. This work is part of the Work Group 1, which deals with the Monitoring and Modeling, and describes the dataset composed of SAR images used to train the Artificial Intelligence (AI) and the future data that will be incorporated on the database of the project.

1 - Introduction

In September 2019 small tar balls and oil slicks showed up at over 1000 locations along the Brazilian's coastline till December of 2020 (IBAMA 2020). Altogether, it affected 11 states, mostly in the Northeast region (>80%) and 20% in the Southeast region (Soares et al. 2020). It represented the worst oil spill disaster not only in Brazilian history, but also in any tropical coastal region worldwide (Nasri Sissini et al. 2020).

The socioeconomic and environmental impacts of this event are not only a short term problem, but also a long term situation, making it difficult to estimate the price of the damage (Câmara et al. 2021). In terms of socioeconomic matters, those states depend on tourism and fishery activities. In the Northeast region approximately 724 fishery and mussels extraction locations were spoiled, affecting 144 thousand artisanal fishers (IBAMA 2020), jeopardizing their incoming for a long period. In less than 2 months of the disaster 40% of the tourism movement decayed (FUNDAJ 2019). Studies show the public health issues which those places are passing through, such as mental health, breathing problems and skin diseases (Carmo and Teixeira 2020); (Pena et al. 2020); (CCI/ENSP 2021).

Despite the Brazilian National Contingency Plan (PNC) existing, it didn't apply correctly in this case. Firstly because it should be triggered with an emergency, but it took weeks for the government to act. Secondly, when specialists analyzed the satellite images to discover where and who caused the spill, they couldn't conclude what happened. This was because the oil spill spread subsurface, so any satellite or radar could not identify the exact location of the incident and inquire the indicted. With that scenario, the Federal Government activated the Monitoring and Evaluation Group (GAA), jointly coordinated by the Brazilian Institute of Environmental and Renewable Natural Resources (IBAMA), the Brazilian Navy and the National Petroleum Agency (ANP). To help out to discover where and who committed the spill, the GAA created a scientific coordination body composed of seven working groups (WGs) that collaborated with scientists working from all over the country to study different aspects of this disaster and learn how to prevent and mitigate a possible new event. To succeed in that, a combination of technologies will be used.

Remote Sensing (RS) has been developed in the past decade, becoming a powerful tool to monitor every spot in the planet with a high resolution and frequency of coverage. Multisensory data has been collected for many years, making the lack of data no longer a limit factor for Earth Observation (EO). The challenge nowadays is to analyze all that data. For that, Machine Learning and Artificial Intelligence came to promise maximization of the work and to make possible the monitoring of large areas 24/7.

This work is part of the WG 1, which deals with the Monitoring and Modeling, which aims to develop situational analysis and Artificial Intelligence algorithm for detection of oil slicks on the sea surface from orbital images, meteo-oceanographic fields of surface generated by coupled ocean-atmosphere models and ship positions. To do that, all data must be on an integrating platform that must be able to store, select and make these data available for analysis in a geographic information system. Analyzes must be carried out semi-automatically, using images processing algorithms associated with artificial intelligence techniques (Nobre et al. 2022). The aim of this paper is to describe the creation and dataset gathering for the Artificial Intelligence (AI) training and the future data that will be incorporated on the database of the project. This primary database set will nourish the ship and oil detection algorithms developed by WG that will monitor the Brazilian coast and help to prevent and mitigate any spill like the 2019 event.

2 - Assembling Data for the Artificial Intelligence

Oil spill is one of the biggest problems that affect the environment since the industrial revolution. Petroleum products have a wide range of short/ long-term impacts on the ecosystem. Animals exposed to oil can have health problems and changes to their physiology and behavior (Ober 2019). Fortunately, the amount of large spills declined over the decades. The International Tanker Owners Pollution Federation (ITOPF) maintains a database of oil spill from tank vessels since 1970, and shows the reduction of spills over the years. Despites that downward trend, due to improvement on the standards of operations in sea-borne transportation, most of those incidents occurred while the vessel was underway in open water (ITOPF 2022), where inspection is a problem.

Monitoring and detecting an oil slick as soon as possible is the key to a successful contingency plan. Satellite-based remote sensing became the monitoring into remote areas, as the middle of the ocean, possible. Radar images, like the Synthetic Aperture Radar (SAR) are the best instrument to do that since it provides data regardless of the hour of the day and the weather condition. Oil slick shows up as a dark area above a light gray color which represents water. However, dark areas in SAR images do not always mean petroleum slicks. These oil spill “look-alike” may represent: algae blooms; low wind track; cold upwelling water; divergent flow, e.g., wave or tidal flow; turbulent water; and other kinds of oil, e.g. plant oils (Alpers, Holt, and Zeng 2017).

One of the many techniques adopted to identify oil in SAR images is Machine Learn (ML). It has shown an effective way to extract information from RS data in a semi-automatic manner, turning the analyses of a massive amount of data faster, helping the right decision making on the time. Al-Ruzouq et al. (2020) made a great review of this technique showing the advantages and shortcomings, the preprocessing chain needed and the challenges of oil spill detection.

As previously said, to differ what is oil from oil’s looks-alike it is the big issue. To minimize that error, images from verified oily incidents were chosen within meteo-oceanographic data from those episodes.

For now, the dataset is composed of twenty-six RADARSAT-1 images from Cantarell (Gulf of Mexico) of 2020 (Figure 1). Cantarell is a heavy oilfield located 100km off the Yucatán coast, in Mexico. It was once the second largest oilfield in the world, but currently occupies the sixth position (Offshore Technology 2000; Wikipedia contributors 2022). All images were pre-processed with an image segmentation technique that analyzed 409 features of oil, classified in oil spill (349 features) or seepage slick (60 features) (Beisl et al. 2004; de Miranda et al. 2004).

Sixty Sentinel-1 images (Figure 2), being sixteen from the Suez Canal and four images from the Red Sea contained verified oil spill events, related to ship incidents, within the period 2015-2019 (Abou El-Magd et al. 2020). Five images from the oil spill accident with the Wakashio Japanese vessel in 2020 at Mauritius, an Indian Ocean island (Rajendran et al. 2021).

Four scenes from the period of 2017 to 2018, one close to Kamarajar Harbor, south India, where two ships collided. One at Sharjah, United Arab Emirates, due to an oil well drilling. Others from Kuwait coast, with an uncertain origin. And an image from Mubarak, Pakistan, related to an illegal discard (Naz et al. 2021). Two scenes from the Black Sea, where a tanker dropped oil, spreading for almost 80 km² on the sea in August 2021 (Voytenko 2021).

Thirteen images from Mediterranean Sea, seven from the Syrian oil refinery in Baniyas at August 2021, tree images of Emerald tanker oil spill that roamed from Israel to Lebanon coast in February 2021, and tree images from Cap Corse due to the collision between two vessels in 2018 (VisioTerra 2021; Surkes 2021; Bitelli 2019).

Eight scenes from the North Sea, near to the Brage oil platform, show water discharges containing oil in it (Skrunes, Johansson, and Brekke 2019). And four images from the Caspian

Sea, on the Azerbaijan coast. Two of them are from the fire incident on the gas pipeline in 2015, and two are from irregular discharges in 2019 (Mityagina and Lavrova 2019).

Images from other satellite sources like ICEYE, MODIS and CBERS, COSMO-SkyMed Constellation and TerraSAR-X with more oil spill cases will be incorporated, especially from the event of 2019. Also, wind, chlorophyll and sea surface data related to the images sites, from ERA5 and MODIS models are being added to the database along with real time data from PNBOIA/REMObs buoys to assist the AI to understand what is real oil spills from “look-alikes". All images will be stored at GeoTiff and the meteo-oceanography data in NetCDF formats.

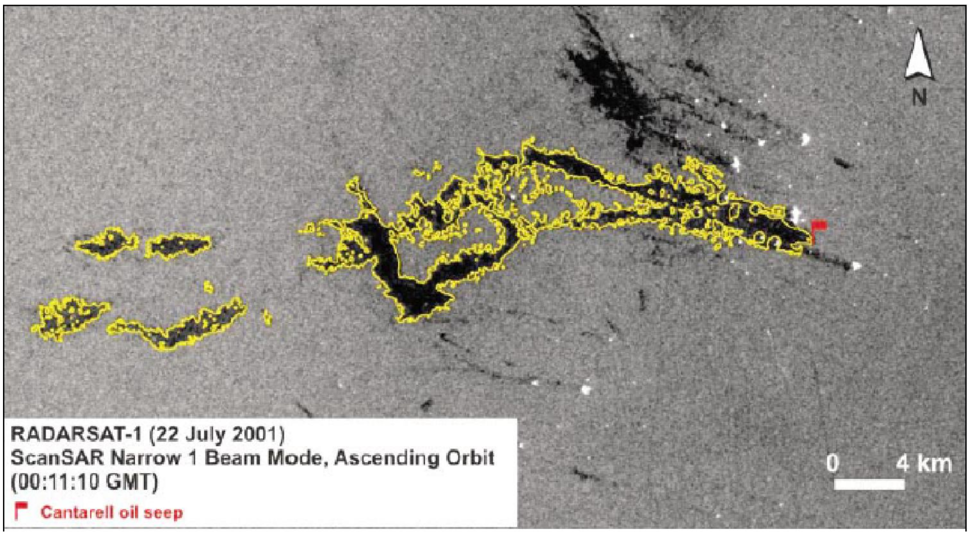

Figure 1. RADARSAT-1 image with seepage slick (highlighted in yellow) from Cantarell (2001). Source: de Miranda et al. 2004.

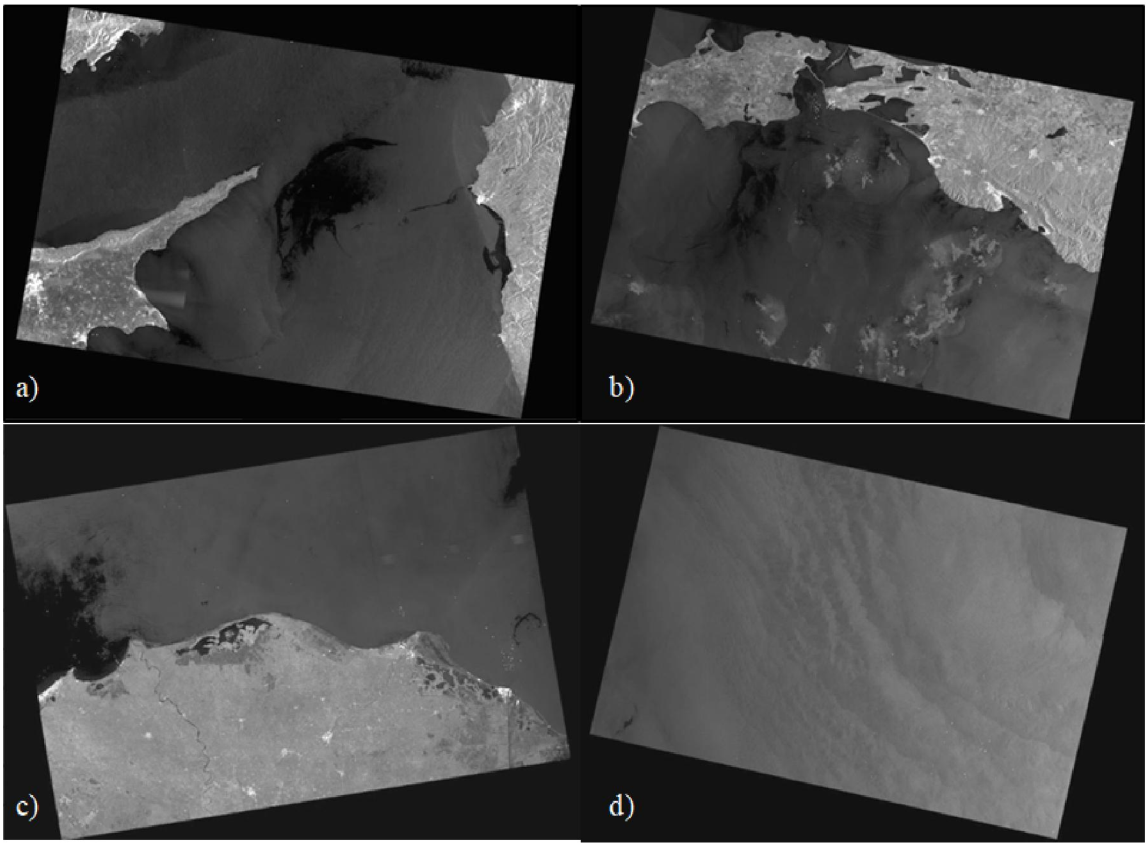

Figure 2. Sentinel-1 images with oil spills. a) Syrian coast 2021, b) Black Sea 2021 , c) Suez Canal 2019, d) North Sea 2019.

3 - Database Creation

To catalog the data, spreadsheets containing the necessary information to understand the environment of the image were made. One describes the metadata of the image and another describing the environment of the place at that time.

The information extracted from the images is:

-Location:

- ID_data - is the Mercator zone where the center of the image is. The name of the satellite and the date of acquisition. Used to connect with the others worksheets;
- ID_mercator: is the Mercator zone where the center of the image is.
- Lat_dec_central / Lon_dec_central - decimal degrees from the center of image;
- Sigla_uf - states there are in the image, if it is in Brazil territory.

-Images data:

- Satellite - the satellite name;
- Sensor - Instruments related to the acquisition of data or satellite mission. E.g.: MUX, WFI, SAR;
- Nome_arquivo_imagem - original name of the image;
- Bandas - refers to the spectral resolution at which the sensor operates. Usually they are: infrared, blue, green, red;
- Órbita - the orbit of a satellite depends on the function used for it. Some types of orbits are: circular elliptical, polar and geostationary;
- Data - the date of the acquisition;
- Process_Type - which level of processing it has undergoing;
- Status - if it has oil spill, seepages slicks or false oil;
- Features - how many oil features does the image contain.

The information extracted from the environment is:

-Locator:

- ID_data - worksheets connector;

-Input:

- Lat_dec / Lon_dec - decimal degrees from the center of site related to the image;
- Sigla_uf - if it is in Brazil territory, the states within;
- Data - date of the inputs data;
- Hora - hour of the data acquisition;
- Vento - minimum and maximum of wind on that quadrant;
- Mare - minimum and maximum of tide on that quadrant;
- Corrente - minimum and maximum of current speed on that quadrant;
- Batimetria - better source of bathymetry to run an oceanography model of the place;

To handle the capacity of storage and processing of all data that the project aims to deal with, a Spatial Data Infrastructure (SDI) was chosen. SDI is a platform that facilitates the management of EO data with an easy interface between the manager and the system. To improve the efficiency and performance to process the large amount of data, those platforms are using their own system of Array Database Management System (Array DBMS) such as TileDB (Papadopoulos et al. 2016) and RasDaMan (Baumann et al. 1998). They also developed their own query language for multidimensional arrays, like AQL (Array Query Language) and RasDaMan RasQL. The better fit to this project was the Open Data Cube (ODC), described next.

3.1 Open Data Cube

To choose the better platform for the project, some requirements and characteristics need to be considered. Gomes et al. (2020) listed the capacities of the platforms, which helped to select it. For that project, the capability of storage scalability should be easy; other interfaces and applications must have access to data of other services, like OGC Services; the extensibility for new tools; the capacity of sharing and reproducing the results among others researchers; and the processing capability is essential.

The Open Data Cube (ODC) is an open source project born from the need to manage a bunch of Satellite Data. It supports through command line tools and a Python API the cataloging, access and manipulation of a huge EO datasets, distributed on git repositories (https://github.com/opendatacube). Figure 4 shows the whole process of uploading the data, the data cube infrastructure and application platform. To visualize products, statistics and retrieve datasets on GeoTiff and NetCDF formats, a web portal is available, which can be managed through a REST API on Python Jupyter Notebooks which makes easy to access and manipulate the data (Gomes, Queiroz and Ferreira 2020; Open Data Cube 2022).

At this moment, this platform is in the process of installation and configuration on the project supercomputer, and soon will be available for access.

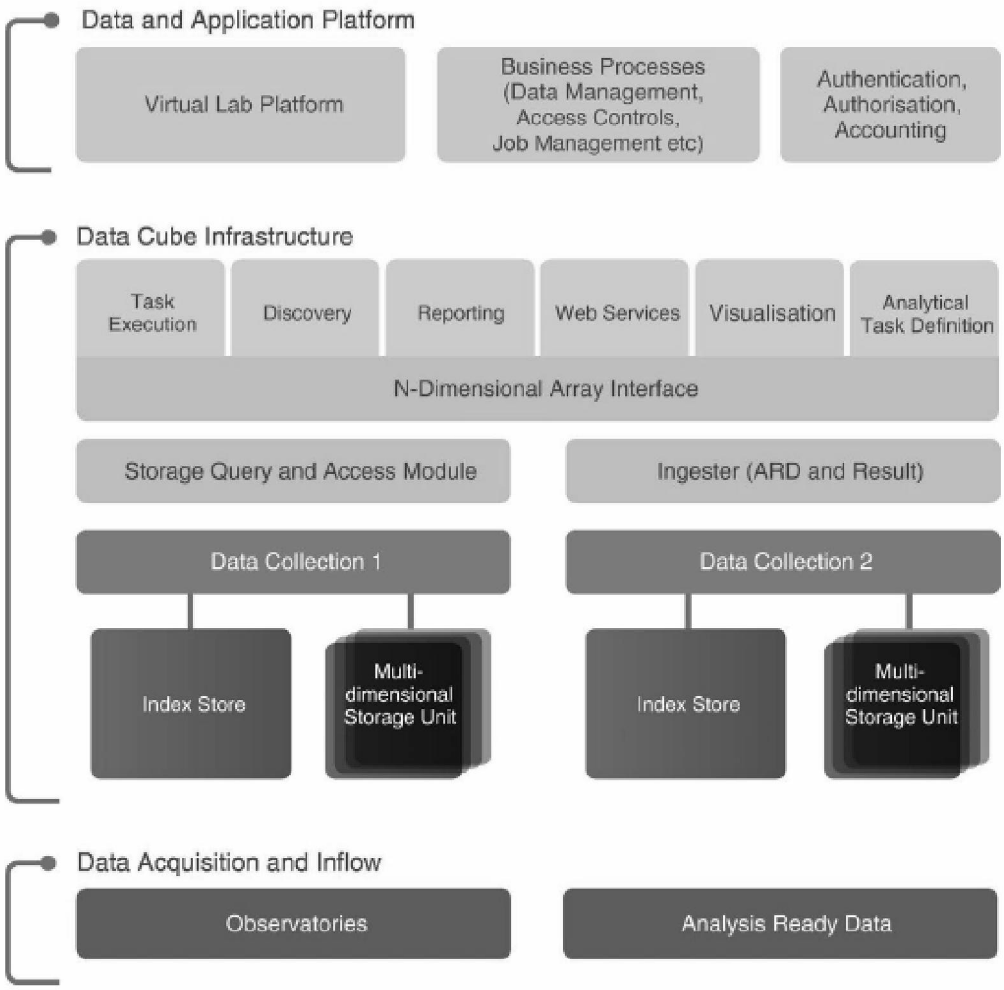

Figure 4. Open Data Cube (ODC) platform architecture diagram. Source: (Lewis et al. 2017)

4 - Consideration

This work belongs to a project that aims to create a lifetime monitoring system of Brazilian marine space using Artificial Intelligence and Satellite data. To do that many tests, trials and work will be necessary. The database's first steps are to collect information from the event of 2019 and other verified incidents to teach the AI to read oil in satellite images. Then, nourishing the DB with new oil spills data from our country and afford an integrated platform so users can have access to it, using the better SDI platform for that project.

5 - Acknowledges

The authors acknowledge the support of the National Institute for Space Research - INPE for support in the completion of this study. This work was supported by Conselho Nacional de

Desenvolvimento Científico e Tecnológico – CNPq Grant No. 440857/2020-1, CNPq/MCTI 06/2020 - Pesquisa e Desenvolvimento para Enfrentamento de Derramamento de Óleo na Costa Brasileira, Programa Ciência no Mar. The satellite data provided by Copernicus Marine Environment Monitoring Service - CMEMS and Open Data Cube are acknowledged. For our partner Carlos Beisl and his team to provide satellite data with segmentation of oil already done. And all the researchers and institutions who are part of this project.

6- References

Abou El-Magd, Islam, Mohamed Zakzouk, Abdulaziz M. Abdulaziz, and Elham M. Ali. 2020. "The Potentiality of Operational Mapping of Oil Pollution in the Mediterranean Sea near the Entrance of the Suez Canal Using Sentinel-1 SAR Data." *Remote Sensing* 12 (8): 1352. https://doi.org/10.3390/rs12081352.

Alpers, Werner, Benjamin Holt, and Kan Zeng. 2017. "Oil Spill Detection by Imaging Radars: Challenges and Pitfalls." *Remote Sensing of Environment* 201 (November): 133–47. https://doi.org/10.1016/j.rse.2017.09.002.

Al-Ruzouq, Rami, Mohamed Barakat A. Gibril, Abdallah Shanableh, Abubakir Kais, Osman Hamed, Saeed Al-Mansoori, and Mohamad Ali Khalil. 2020. "Sensors, Features, and Machine Learning for Oil Spill Detection and Monitoring: A Review." *Remote Sensing* 12 (20): 3338. https://doi.org/10.3390/rs12203338.

Baumann, P., A. Dehmel, P. Furtado, R. Ritsch, and N. Widmann. 1998. "The Multidimensional Database System RasDaMan." *SIGMOD Rec.* 27 (2): 575–77. https://doi.org/10.1145/276305.276386.

Beisl, C. H., E. C. Pedroso, L. S. Soler, A. G. Evsukoff, F. P. Miranda, A. Mendoza, A. Vera, and J. M. Macedo. 2004. "Use of Genetic Algorithm to Identify the Source Point of Seepage Slick Clusters Interpreted from Radarsat-1 Images in the Gulf of Mexico." *Institute of Electrical and Eletronics Engineers IEEE* 6: 4139–42. https://doi.org/10.1109/IGARSS.2004.1370044.

Bitelli, Francesca Franci Francesca Trevisiol. 2019. "SENTINEL-1 DATA FOR FUEL SPILL DETECTION. The Case of the Ships Collision near Corsica." https://www.enea.it/it/seguici/events/telerilevamento_25-26giu2019/Trevisiol.pdf.

Câmara, Samuel Façanha, Francisco Roberto Pinto, Felipe Roberto da Silva, Marcelo de Oliveira Soares, and Thiago Matheus De Paula. 2021. "Socioeconomic Vulnerability of Communities on the Brazilian Coast to the Largest Oil Spill (2019–2020) in Tropical Oceans." *Ocean & Coastal Management* 202 (March): 105506. https://doi.org/10.1016/j.ocecoaman.2020.105506.

Carmo, Eduardo Hage, and Maria Gloria Teixeira. 2020. "Technological Disasters and Public Health Emergencies: The Case of Oil Spill on the Brazilian Coast." *Cadernos de Saude Publica* 36 (2): e00234419. https://doi.org/10.1590/0102-311X00234419.

CCI/ENSP. 2021. "Derramamento de óleo no litoral brasileiro: pesquisadores concluem que desastre configurou Emergência em Saúde Pública." 2021. http://informe.ensp.fiocruz.br/noticias/48246.

FUNDAJ. 2019. “5 fatos que não te contaram sobre a poluição por óleo no Litoral do Nordeste.” FUNDAJ. October 21, 2019. https://www.gov.br/fundaj/pt-br/destaques/observa-fundaj-itens/observa-fundaj/revitalizacao-de-bacias/5-fatos-que-nao-te-contaram-sobre-a-poluicao-por-oleo-no-litoral-do-nordeste.

Gomes, Vitor C. F., Gilberto R. Queiroz, and Karine R. Ferreira. 2020. “An Overview of Platforms for Big Earth Observation Data Management and Analysis.” *Remote Sensing* 12 (8): 1253. https://doi.org/10.3390/rs12081253.

IBAMA. 2020. “Manchas de Óleo - Litoral Brasileiro.” IBAMA. 2020. http://www.ibama.gov.br/manchasdeoleo.

ITOPF. 2022. “Oil Tanker Spill Statistics 2021.” ITOPF. 2022. https://www.itopf.org/knowledge-resources/data-statistics/statistics/.

Lewis, Adam, Simon Oliver, Leo Lymburner, Ben Evans, Lesley Wyborn, Norman Mueller, Gregory Raevksi, et al. 2017. “The Australian Geoscience Data Cube — Foundations and Lessons Learned.” *Remote Sensing of Environment* 202 (December): 276–92. https://doi.org/10.1016/j.rse.2017.03.015.

Miranda, Fernando Pellon de, Arturo Mendoza Quintero Marmol, Enrico Campos Pedroso, Carlos Henrique Beisl, Pamela Welgan, and Luis Medrano Morales. 2004. “Analysis of RADARSAT-1 Data for Offshore Monitoring Activities in the Cantarell Complex, Gulf of Mexico, Using the Unsupervised Semivariogram Textural Classifier (USTC).” *Canadian Journal of Remote Sensing* 30 (3): 424–36. https://doi.org/10.5589/m04-019.

Mityagina, Marina I., and Olga Y. Lavrova. 2019. “Multi-Sensor Satellite Survey of Surface Oil Pollution in the Caspian Sea.” In *Remote Sensing of the Ocean, Sea Ice, Coastal Waters, and Large Water Regions 2015*, edited by Charles R. Bostater, Stelios P. Mertikas, and Xavier Neyt. SPIE. https://doi.org/10.1117/12.2194511.

Nasri Sissini, M., F. Berchez, J. Hall-Spencer, N. Ghilardi-Lopes, V. F. Carvalho, N. Schubert, G. Koerich, et al. 2020. “Brazil Oil Spill Response: Protect Rhodolith Beds.” *Science*. https://doi.org/10.1126/science.aba2582.

Naz, Saima, Muhammad Farooq Iqbal, Irfan Mahmood, and Mona Allam. 2021. “Marine Oil Spill Detection Using Synthetic Aperture Radar over Indian Ocean.” *Marine Pollution Bulletin* 162 (January): 111921. https://doi.org/10.1016/j.marpolbul.2020.111921.

Nobre, Paulo et al. The 2019 northeast Brazil oil spill: scenarios. Anais da Academia Brasileira de Ciências [online]. 2022, v. 94, suppl 2 [Accessed 13 September 2022] , e20210391. Available from: <https://doi.org/10.1590/0001-3765202220210391>. Epub 02 Sept 2022. ISSN 1678-2690. https://doi.org/10.1590/0001-3765202220210391.

Ober, Holly K. 2019. “Effects of Oil Spills on Marine and Coastal Wildlife.” *Department of Wildlife Ecology and Conservation*. https://edis.ifas.ufl.edu/publication/UW330.

Offshore Technology. 2000. “Cantarell Oilfield, Gulf of Mexico.” Offshore Technology. May 16, 2000. https://www.offshore-technology.com/projects/cantarell/.

Open Data Cube. 2022. Available online: https://www.opendatacube.org/overview [Accessed: 20 June 2022].

Papadopoulos, Stavros, Kushal Datta, Samuel Madden, and Timothy Mattson. 2016. “The TileDB Array Data Storage Manager.” *Proceedings of the VLDB Endowment International Conference on Very Large Data Bases* 10 (4): 349–60. https://doi.org/10.14778/3025111.3025117.

Pena, Paulo Gilvane Lopes, Amanda Laura Northcross, Mônica Angelim Gomes de Lima, and Rita de Cássia Franco Rêgo. 2020. “The Crude Oil Spill on the Brazilian Coast in 2019: The

Question of Public Health Emergency." *Cadernos de Saude Publica* 36 (2): e00231019. https://doi.org/10.1590/0102-311X00231019.

Rajendran, Sankaran, Ponnumony Vethamony, Fadhil N. Sadooni, Hamad Al-Saad Al-Kuwari, Jassim A. Al-Khayat, Vashist O. Seegobin, Himanshu Govil, and Sobhi Nasir. 2021. "Detection of Wakashio Oil Spill off Mauritius Using Sentinel-1 and 2 Data: Capability of Sensors, Image Transformation Methods and Mapping." *Environmental Pollution* 274 (April): 116618. https://doi.org/10.1016/j.envpol.2021.116618.

Schramm, Matthias, Edzer Pebesma, Milutin Milenković, Luca Foresta, Jeroen Dries, Alexander Jacob, Wolfgang Wagner, et al. 2021. "The openEO API–Harmonising the Use of Earth Observation Cloud Services Using Virtual Data Cube Functionalities." *Remote Sensing* 13 (6): 1125. https://doi.org/10.3390/rs13061125.

Skrunes, Stine, A. Malin Johansson, and Camilla Brekke. 2019. "Synthetic Aperture Radar Remote Sensing of Operational Platform Produced Water Releases." *Remote Sensing* 11 (23): 2882. https://doi.org/10.3390/rs11232882.

Soares, M.O, Teixeira, C. E. P., Bezerra, L. E. A., Paiva, S. V., Tavares, T. C. L., Garcia, T. M., Araújo,j. T., Campos, C. C., Ferreira, S. M. C., Matthews-Cascon, H., Frota, A., Mont'Alverne, T. C. F., Silva, S.T., Rabelo, E. F., Barroso, C. X., Freitas, J. E. P., Júnior, M. M., Campelo, R. P. S., Santana, C. S., Carneiro, P. B. M., Meirelles, A. J., Santos, B. A., Oliveira, A. H .B., Horta, P., Cavalcante, R. M. 2020. "Oil Spill in South Atlantic (Brazil): Environmental and Governmental Disaster." *Marine Policy* 115 (May): 103879. https://doi.org/10.1016/j.marpol.2020.103879.

Surkes, Sue. 2021. "Marine Monitor Urges Israel to Step up Surveillance or Risk More Oil Spill Woes." The Times of Israel. March 17, 2021. https://www.timesofisrael.com/marine-monitor-urges-israel-to-step-up-surveillance-or-risk-more-oil-spill-woes/.

VisioTerra. 2021. "Baniyas Thermal Power Plant Causes a 1000 Km^2 Oil Spill, Syria." https://www.sentinelvision.eu/gallery/pdf/e16326bf3cd44b2cbf689cd1ad134af2.

Voytenko, Mikhail. 2021. "Greek Aframax cleared of spill accusation in Novorossiysk, Black sea." FleetMon.com. August 11, 2021. https://www.fleetmon.com/maritime-news/2021/34869/greek-aframax-cleared-spill-accusation-novorossiys/.

Wikipedia contributors. 2022. "List of Oil Fields." Wikipedia, The Free Encyclopedia. August 14, 2022. https://en.wikipedia.org/w/index.php?title=List_of_oil_fields&oldid=1104337769.