

Análise de Dados de Focos de Calor no Brasil Através de Técnicas de Visualização

Jéssica da Silva Costa¹, Felipe da Rocha Henriques^{1,2}, Kele Teixeira Belloze¹

¹Programa de Pós-graduação em Ciência da Computação (PPCIC) – CEFET/RJ
Rio de Janeiro/RJ – Brasil

²Coordenação de Telecomunicações – CEFET/RJ (Campus Petrópolis)
Petrópolis/RJ – Brasil

jessica.costa.1@aluno.cefet-rj.br
{felipe.henriques, kele.belloze}@cefet-rj.br

Abstract. *This work analyses fire outbreak data in brazilian states between 1998 and 2020. Data visualization techniques are used to perform a data exploratory analysis, in order to generate usefull information, such as months and years with larger numbers of fires. Furthermore, a prediction model is used to evaluate the tendency of amount of focus, for three specific states. Results showed that the visualization techniques, along with the prediction employed, form a set of valid tools for the data analythics and decision making process.*

Resumo. *Este trabalho analisa dados de focos de incêndio nos Estados brasileiros entre os anos de 1998 e 2020. Técnicas de visualização de dados são usadas para realizar uma análise exploratória dos dados, de modo a gerar informações úteis, tais como os meses e anos com maior quantidade de focos. Além disso, um modelo de predição foi usado, de modo a avaliar a tendência da quantidade de focos, para três Estados específicos. Os resultados mostraram que as técnicas de visualização, em conjunto com a predição, formam um conjunto válido de ferramentas para a análise de dados e tomada de decisões.*

1. Introdução

Incêndios florestais constituem uma grande ameaça que coloca em risco o meio ambiente e a população. O acompanhamento de dados de focos de calor é realizado mundo afora, e o Brasil não fica de fora desse movimento ambiental. Muitos destes incêndios podem se iniciar em áreas remotas, o que dificulta sua percepção, fazendo com que ele se espalhe e se torne perigoso antes de ser detectado.

De modo a resolver este problema, imagens de satélites permitem a detecção e monitoramento de focos de calor espalhados nos territórios. Tanto as imagens, quanto os dados gerados nesse processo podem ser analisados para, por exemplo, detectar padrões e realizar previsões sobre focos de calor em um determinado local, considerando distribuições espaço-tempo [Pinto et al. 2020]. A detecção antecipada de focos de calor se justifica na redução do tempo de resposta para a resolução do problema, e no apoio a tomadores de decisão para otimização de recursos no caso de vários incêndios ao mesmo tempo [Laneve et al. 2019]. Neste contexto, o uso de técnicas de visualização de dados pode trazer informações que permitam uma melhor caracterização de cenários, identificação de *outliers* e a descoberta de padrões entre séries temporais [Telea 2014].

Incidentes de focos de calor ocorridos entre 2011 e 2017 na cidade de Changsha (China) foram analisados em [Liu et al. 2019], com o intuito de se avaliar o risco de incêndios nessa região. A partir dos dados, um modelo de risco generalizado (GRM) foi proposto, e visualizações foram utilizadas de modo a mostrar graficamente tais riscos. Em [Hilary et al. 2020], uma Rede de Sensores Sem Fio é usada como infraestrutura para o monitoramento de incêndios em mercados na Tanzânia. Os dados coletados são transmitidos para a nuvem, onde são processados e visualizados, de modo a mitigar possíveis incêndios. Ainda na mesma linha, [Sowah et al. 2014] propõe um sistema *web* que se utiliza de uma infraestrutura de sensores para realizar a detecção e notificação de incêndios em tempo real.

De modo a auxiliar o combate a focos de incêndio por meio de aviões que sobrevoam tais locais lançando água na região, [Becker et al. 2019] se utilizam de técnicas de exploração de dados para analisar séries temporais correspondentes a voos onde se observou um comportamento anormal. Através de diversas visualizações de dados, *outliers* puderam ser detectados em dados replicados de séries temporais multivariadas.

Diferentemente dos artigos supracitados, neste trabalho, uma série temporal de dados de focos de calor no Brasil é analisada, através de técnicas de visualização de dados. Além disso, uma predição da tendência da quantidade de focos para três Estados brasileiros é avaliada, por meio do método de regressão SARIMAX [Hao et al. 2019].

2. A Questão Ambiental e Focos de Incêndio no Brasil

Incêndios florestais constituem impactos substanciais nos níveis sociais, econômico e ecológico [Chuvieco et al. 2019]. De acordo com o Sistema de Observação do Clima Global¹ (GCOS, em inglês), os incêndios contribuem para o aumento de CO₂ devido ao desmatamento gerado e ao empobrecimento das florestas, além de alterações nos regimes de incêndios (mais frequentes, maiores ou mais preocupantes). Embora um incêndio possa ser visto como uma condição natural, diversas vezes o fogo também é utilizado em atividades humanas para gerenciar e transformar terras para a agricultura. Por exemplo, há uma prática de se colocar fogo no solo após uma colheita de forma a limpá-lo e também a oferecer nutrientes por meio das cinzas que se formam; assim como para abrir pastagens ou áreas de cultivo [Hardesty et al. 2005]. Ambas as ocorrências podem prejudicar ecossistemas e colocar em risco populações e a saúde humana.

No Brasil, uma das áreas mais afetadas pelos incêndios florestais é o bioma Cerrado, também conhecido como a savana brasileira. O cerrado é uma área bastante susceptível aos focos de calor, devido a condições climáticas quentes e sazonalmente secas [Dantas et al. 2013]. Os incêndios florestais alteram historicamente a vegetação natural. Assim sendo, a frequência de incêndios vem mudando nos últimos anos, o que contribui para a perda da biodiversidade e afeta cada vez mais a vegetação nativa, os *habitats* naturais e o ecossistema nas regiões tropicais [Jr. et al. 2014, Arruda et al. 2021]. De acordo com [White 2019], o Brasil registrou a maior ocorrência de incêndios florestais entre os anos de 2003 a 2017.

O Instituto Nacional de Pesquisas Espaciais (INPE), por meio do Programa Queimadas, monitora e modela a ocorrência e a propagação de incêndios na vegetação, utilizando técnicas de sensoriamento remoto, geoprocessamento e modelagem numérica. Em

¹<https://gcos.wmo.int/>

seu Portal² são disponibilizados dados abertos sobre focos de calor, atualizados diariamente e relacionados aos continentes americano, africano e europeu. Embora os focos de calor sejam detectados por imagens de diferentes satélites, o INPE faz uso de um “satélite de referência”, que utiliza o mesmo método e a mesma hora para captar as imagens, possibilitando assim a formação e análises de séries temporais. Tais séries favorecem, portanto, a análise de tendências para os mesmos períodos em localidades de interesse. Cabe ressaltar que o INPE não tem a atribuição de fiscalizar ou combater os incêndios no país, e sim gerar o maior número de dados possíveis sobre o assunto de modo a favorecer o acesso à informação pelo governo e sociedade.

3. A Visualização de Dados como Ferramenta de Apoio à Análise de Dados

De acordo com [Telea 2014], a visualização de dados tem por objetivo gerar *insights*, de modo que se possa responder a questões acerca de algum fenômeno ou processo de interesse. Um tipo de informação útil que se pode extrair a partir de uma boa visualização de dados é responder a perguntas concretas sobre algum problema. Uma questão quantitativa, obtida através de visualizações, pode ser: “*Quais são os valores mínimos, máximos e outliers de uma base de dados?*” Por outro lado, visualizações também podem nos trazer respostas a questões qualitativas, tais como: “*Dado um exame de imagem de um paciente, existe alguma anomalia que pode indicar algum problema clínico?*”

Nesse sentido, pode-se compreender a visualização de dados como um processo de exploração interativa, que começa a partir de um conjunto (ou base) de dados, que pode ter sido gerado a partir da coleta de sensores em um ambiente de Internet das Coisas [Yasmin et al. 2018]. Esses dados são usados como entrada para uma aplicação de visualização, que gera como saída uma imagem a ser observada por um usuário final. O usuário pode interagir com a aplicação, de modo a refinar a imagem gerada, alterando-se parâmetros, cores, *zoom*, criando assim um processo retroalimentado de interação-visualização-observação.

O autor de [Runkler 2016] entende a visualização de dados como uma das fases de um projeto de análise de dados. Já em [Knaflic 2019], a autora compreende a visualização como uma forma de contar uma história acerca dos dados, e apresenta várias ferramentas para que o processo de *storytelling* atinja o público de maneira acurada e impactante. Em seu livro *Data Points: Visualization That Means Something* [Yau 2013], Nathan Yau utiliza a ideia do *storytelling* apresentando diversas fotos do seu casamento. A partir desses dados bastante comuns para a maioria das pessoas, ele é capaz de extrair outras informações, tais como a quantidade de fotos tiradas durante o tempo do casamento, como em um histograma. Ele ainda destaca que a conexão entre o dado e o que ele representa é um ponto chave para o processo de visualização e seu significado (nesse caso, um conjunto de imagens e a celebração de um casamento).

O ponto comum entre as obras supracitadas, que dá sentido a todo o processo de visualização, é o dado a ser analisado. Neste trabalho, também partimos desse pressuposto, mas diferentemente das imagens de [Yau 2013], dados de focos de incêndio em Estados brasileiros são analisados, e a visualização é utilizada como ferramenta de apoio para a sua análise.

²<https://queimadas.dgi.inpe.br/queimadas/portal>

4. Metodologia

A metodologia utilizada para a análise dos dados de focos de incêndio dos Estados brasileiro seguiu as atividades de coleta de dados, pré-processamento e predição dos dados. Essas atividades são descritas a seguir.

Os dados foram extraídos da base do Programa Queimadas do INPE³. Nessa base pública, é possível coletar dados sobre os números de focos de calor ativos por biomas, por região e por estado, separados por anos e meses. Além disso, o portal apresenta alguns gráficos de séries históricas dos dados disponíveis. Para a coleta dos dados, foi feita a extração de 27 arquivos no formato *csv* (*Comma Separated Values*) compreendendo dados dos anos de 1998 a 2021 de cada Estado brasileiro, incluindo o Distrito Federal. Devido ao ano de 2021 estar em andamento, foi estipulado que a coleta e análise contemplaria até o ano de 2020.

Após a fase de coleta, foi necessário realizar uma preparação dos dados para adequá-los às análises posteriores. Os arquivos foram unificados de modo a gerar uma base única. Essa base contém o ano, o estado, o mês e a quantidade de focos de calor. Alguns registros não possuíam o número de focos e, por isso, foram substituídos por zero. Nessa etapa foram utilizadas a linguagem *Python* e suas bibliotecas. As bibliotecas *Numpy* e *Pandas* foram utilizadas para manipulação dos dados, e a *Plotly* e *Matplotlib* para a geração das visualizações.

Na sequência, foi aplicado um estudo preditivo com o modelo ARIMA (*Auto Regressive Integrated Moving Average*) sobre os dados de três Estados brasileiros, que representam três regiões, biomas e regimes de chuvas diferentes no Brasil, quais sejam: Bahia, Minas Gerais e Pará. O modelo ARIMA visa encontrar a autocorrelação na modelagem de séries temporais e dessa forma realizar previsões. Ele pode ser resumido em três parâmetros: p é o número de termos auto-regressivos; d é o número de diferenciações; e q é o número de termos móveis [Chintalapudi et al. 2020]. O modelo ARIMA possui uma variação que trabalha com sazonalidade chamada SARIMA. Este modelo, além das componentes p , d e q , também possui componentes dessa sazonalidade, definidas pelos parâmetros P , D e Q [Hao et al. 2019].

Para a realização da predição com o modelo ARIMA, utilizou-se a biblioteca *Pmdarima*. Ela possui uma função chamada `auto_arima` que opera como uma pesquisa em grade, em que dentre vários conjuntos dos parâmetros p e q , seleciona-se o modelo que minimiza o AIC (*Akaike information criterion*, em inglês) ou BIC (*Bayesian information criterion*). Para a diferenciação, ela usa um teste de estacionariedade (como um teste *Dickey-Fuller* aumentado) e sazonalidade (como o teste *Canova-Hansen*) para modelos sazonais [Smith et al. 2017].

Os dados das séries temporais para os três Estados brasileiros no período de 1998 a 2019 foram selecionados para constituírem o conjunto de dados de treinamento e os dados do ano de 2020 foram selecionados para constituírem o conjunto de dados de teste. Os dados têm uma característica sazonal e por isso o modelo SARIMA foi mais adequado. Todos os modelos treinados foram do tipo SARIMAX, uma variação do modelo SARIMA que amplia a sua capacidade de predição, integrando informações externas [Hao et al. 2019].

³Disponível em http://queimadas.dgi.inpe.br/queimadas/portal-static/estatisticas_estados/

5. Resultados

Nesta seção, são apresentadas as análises dos dados de focos de calor coletados, a partir de visualizações, e os resultados gerados por meio do estudo realizado com o modelo SARIMAX. As Figuras 1 e 2 apresentam visualizações da série temporal por mês dos focos de calor no Brasil e o somatório dos focos de calor por mês, respectivamente. A Figura 1 mostra que de 2002 até 2007, o Brasil teve muitos focos de calor, que diminuíram nos dois anos seguintes, para em 2010 ter outro pico. A linha vermelha representa a média móvel semestral dos focos de calor e mostra um comportamento sazonal no país, com uma subida em Agosto, um pico em Setembro, seguida por uma tendência de queda.

A Figura 2 comprova que Setembro, Agosto e Outubro são os meses com mais focos de calor durante todos os anos com registros. No Brasil, os meses de Agosto e Setembro marcam o fim do inverno e o começo da primavera. Estes meses costumam ser mais secos em boa parte do país, o que pode justificar este comportamento.

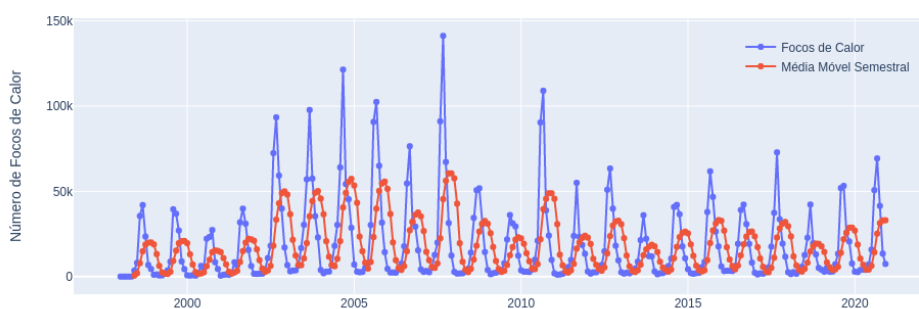


Figura 1. Série temporal de focos de calor por mês no Brasil.

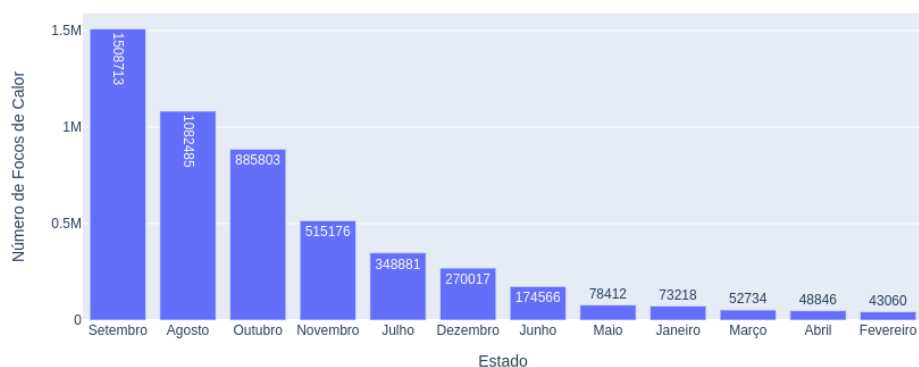


Figura 2. Somatório de focos de calor por mês, para cada Estado brasileiro.

Considerando o ano de 2020 como o último com dados completos de números de focos de calor, a Figura 3 dá um enfoque para todos os Estados brasileiros para esse

ano. Pode-se observar que o ano de 2020 teve uma tendência de crescimento em relação aos anos anteriores. Porém, um aspecto importante é que esse número subiu devido ao comportamento de alguns Estados brasileiros que tiveram mais problemas com queimadas em 2020. O estado do Mato Grosso foi o maior ofensor, tendo apresentado mais de 47 mil focos de calor, seguido pelo Pará com mais 38 mil focos no país. Os outros Estados brasileiros não ultrapassaram mais de 20 mil focos no ano de 2020. Algumas características importantes desses Estados com maior quantidade de focos são os períodos de seca, a agropecuária e a presença da Floresta Amazônica no Estado do Pará, que podem dar indícios de possíveis causas para estes números.

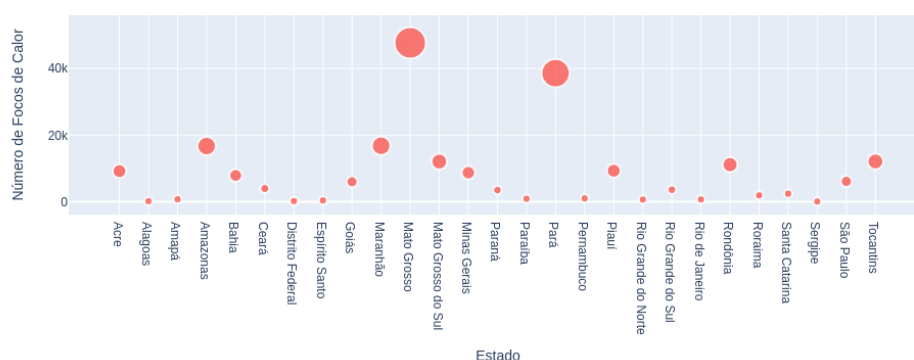


Figura 3. Focos de calor, em 2020, no Brasil.

A Figura 4 apresenta o resultado obtido com as previsões para Bahia, Minas Gerais e Pará. A partir das visualizações, podemos perceber que nos estados que tiveram uma tendência acentuada comparada aos anos anteriores, o algoritmo não apresentou um desempenho tão bom. O modelo que obteve o melhor resultado foi o modelo de Minas Gerais, pois seu comportamento não foi tão discrepante em relação ao ano anterior, o que possibilitou uma melhor predição por parte do algoritmo. O modelo da Bahia teve números baixos nos anos anteriores, porém uma tendência mais acentuada em 2020, o que também prejudicou um pouco o desempenho do algoritmo. O modelo do Pará teve um crescimento mais acentuado em 2020, o que pode ter prejudicado a predição para esse Estado, não tendo conseguido prever tão bem essa tendência de alta.

Além da análise gráfica, foi realizada uma avaliação de uma métrica do modelo de predição. Para isso foi escolhida uma métrica bem utilizada em séries temporais, a RMSE (*Root Mean Square Error*), definida como a raiz quadrada do erro médio quadrático [Athiyarath et al. 2020]. Essa métrica compara a diferença entre os valores preditos e os valores reais do conjunto de teste para verificar o desempenho preditivo do modelo. Pode-se observar na Tabela 1 que o RMSE comprova os resultados da predição realizada nos dados de teste. O modelo do estado que apresentou menor RMSE foi o de Minas Gerais, enquanto que o modelo que apresentou o maior RMSE foi o do Pará.

6. Conclusões e Futuras Direções

Este trabalho apresentou uma análise por meio de técnicas de visualização de dados sobre os focos de calor no Brasil. Embora sejam análises exploratórias, foi possível perceber

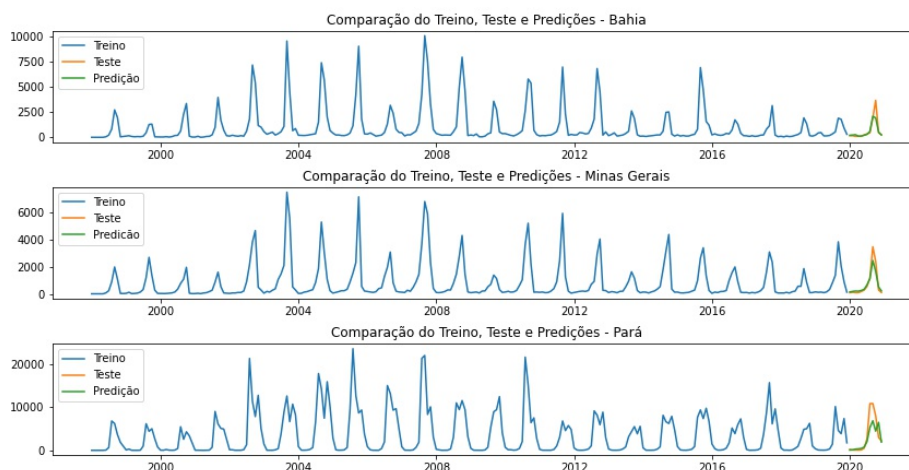


Figura 4. Previsões de tendência dos focos de calor para Bahia, MG e Pará.

Tabela 1. Resultados do RMSE das previsões, para Bahia, MG e Pará.

Estado	RMSE
Bahia	510,15
Minas Gerais	378,42
Pará	2.472,50

como os focos de calor são delineados pelo país. Pôde-se observar uma tendência de alta em 2020, principalmente nos estados do Mato Grosso e do Pará. Um aspecto notável é que Agosto e Setembro são os meses com mais focos em todos os anos de dados registrados. Essas observações são muito importantes para ações preventivas nos próximos anos.

Na sequência deste trabalho, os dados dos demais Estados brasileiros e do Distrito Federal serão analisados por meio do modelo SARIMAX, com o propósito de gerar uma base para outras avaliações. De forma a obter previsões mais acuradas sobre o focos de calor no Brasil, utilizando imagens de satélite em conjunto com as séries temporais, pretende-se conduzir novas investigações com modelos de aprendizado de máquina.

Referências

- Arruda, V., Piontekowski, V., Alencar, A., Pereira, R., and Matricardi, E. (2021). An alternative approach for mapping burn scars using landsat imagery, google earth engine, and deep learning in the brazilian savanna. *Remote Sensing Applications: Society and Environment*, 22.
- Athiyarath, S., Paul, M., and Krishnaswamy, S. (2020). A comparative study and analysis of time series forecasting techniques. *SN Computer Science*, 1.
- Becker, D. G., Braun, W. J., Dean, C., and Woolford, D. G. (2019). Visualizing multivariate time series of aerial fire fighting data. *Journal of Environmental Statistics*, 9(1).
- Chintalapudi, N., Battineni, G., and Amenta, F. (2020). Covid-19 virus outbreak forecasting of registered and recovered cases after sixty day lockdown in italy: A data driven model approach. *Journal of Microbiology, Immunology and Infection*, 53(3):396–403.

- Chuvieco, E., Mouillot, F., van der Werf, G., Miguel, J. S., Tanasse, M., Koutsias, N., García, M., Yebra, M., Padilla, M., Gitas, I., Hawbaker, T., and Giglio, L. (2019). Historical background and current developments for mapping burned area from satellite earth observation. *Remote Sensing of Environment*, 225:45–64.
- Dantas, V. D. L., Batalha, M., and Pausas, J. (2013). Fire drives functional thresholds on the savanna-forest transition. *Ecology*, 94:2454–2463.
- Hao, Y., Wang, R.-R., Han, L., Wang, H., Zhang, X., Tang, Q.-L., Yan, L., and He, J. (2019). Time series analysis of mumps and meteorological factors in Beijing, China. *BMC Infectious Diseases*, 19(1).
- Hardesty, J., Myers, R., and Fulks, W. (2005). Fire, ecosystems, and people: A preliminary assessment of fire as a global conservation issue. *The George Wright Forum*, 22(4):78–87.
- Hilary, R., Rotich, P., Geoffrey, A., and Sam, A. (2020). Early fire detection system in tanzania markets. *Engineering Proceedings*, 2(1).
- Jr., A. P., Oliveira, S., Pereira, J., and Turkman, M. (2014). Modelling fire frequency in a cerrado savanna protected area. *PLoS ONE*, 9.
- Knaflic, C. N. (2019). *Storytelling com dados: um guia sobre visualização de dados para profissionais de negócios*. Alta Books, 2nd edition.
- Laneve, G., Santilli, G., and Luciani, R. (2019). Improving sevir-based hotspots detection by using multiple simultaneous observations. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 12:2349–2356.
- Liu, D., Xu, Z., and Fan, C. (2019). Generalized analysis of regional fire risk using data visualization of incidents. *Fire and Materials*, 43(4):413–421.
- Pinto, M., Libonati, R., Trigo, R., Trigo, I., and DaCamara, C. (2020). A deep learning approach for mapping and dating burned areas using temporal sequences of satellite images. *ISPRS Journal of Photogrammetry and Remote Sensing*, 160:260–274.
- Runkler, T. A. (2016). *Data Analytics: Models and Algorithms for Intelligent Data Analysis*. Springer, 2nd edition.
- Smith, T. G. et al. (2017). pmdarima: Arima estimators for Python. [Online; accessed 2021-04-11].
- Sowah, R., Ofoli, A. R., Krakani, S., and Fiawoo, S. (2014). A web-based communication module design of a real-time multi-sensor fire detection and notification system. In *2014 IEEE Industry Application Society Annual Meeting*, pages 1–6.
- Telea, A. C. (2014). *Data Visualization: Principles and Practice*. CRC Press, 2nd edition.
- White, B. (2019). Satellite detection of wildland fires in south america. *Floresta*, 49:851–858.
- Yasmin, R., Salminen, M., Gilman, E., Petäjäjärvi, J., Mikhaylov, K., Pakanen, M., Niemelä, A., Riekkö, J., Pirttikangas, S., and Pouttu, A. (2018). Combining iot deployment and data visualization: experiences within campus maintenance use-case. In *2018 9th International Conference on the Network of the Future (NOF)*, pages 101–105.
- Yau, N. (2013). *Data Points: Visualization That Means Something*. Willey, Indianapolis.