

ARTIGO ORIGINAL

Predição do potencial radioativo de queimadas: comparando duas técnicas mediante regressão espacial

Prediction of the radioactive potential of fires: comparing two techniques using spatial regression

João Batista Firmino Júnior¹ and Damires Souza¹

¹IFPB

*firminojunior83@gmail.com; damires@ifpb.edu.br

Recebido: 16/02/2022. Revisado: 19/07/2022. Aceito: 26/09/2022.

Resumo

Com as mudanças climáticas, estudos que considerem elementos para administrar prejuízos causados por queimadas nas proximidades ou no interior de municípios por meio do potencial radioativo do fogo podem ser úteis. Nesse contexto, este trabalho apresenta uma abordagem para prever, a partir de regressão ponderada geograficamente, a incidência do potencial radioativo de queimadas. Para isso, dados extraídos do projeto FIRMS da NASA e do INMET (Instituto Nacional de Meteorologia) foram obtidos, tratados, e algumas variáveis topográficas foram adicionalmente derivadas. O trabalho considera a área do município de João Pessoa (no nordeste do Brasil) como objeto de estudo e faz uso de métodos de aprendizado de máquina baseados em regressão linear espacial, considerando dois cenários, cujas diferenças se situam na extensão das amostras e na escolha das variáveis independentes. O primeiro cenário abrange amostras mais concentradas em João Pessoa, e inclui as variáveis climáticas, de satélite e topográficas; e o segundo cenário abrange toda a mesorregião da zona da mata paraibana, porém não considera as variáveis climáticas. Os resultados obtidos são promissores.

Palavras-Chave: Análise preditiva; queimadas; regressão espacial; R²; RMSE.

Abstract

With climate change, studies that consider elements to manage damages caused by fires in the vicinity or inside municipalities, through the radioactive potential of fire, may be useful. In this context, this work presents an approach to predict, from geographically weighted regression, the incidence of the radioactive potential of fires. For this, data extracted from the NASA's FIRMS project and from the INMET (National Institute of Meteorology) have been acquired, cleaned, and some topographic variables have been additionally derived. The work considers the city of João Pessoa (in the northeast of Brazil) as an object of study and makes use of machine learning methods based on spatial linear regression to predict the radioactive potential of fires. To this end, it considers two scenarios whose differences lie in the extension of the samples and in the choice of the independent variables. The first scenario covers samples which are more concentrated in the city of João Pessoa, and includes climatic, satellite and topographical variables; the second scenario covers the entire mesoregion of the forest zone in Paraíba close to João Pessoa, but does not consider the climatic variables. The work has achieved promising results.

Keywords: Predictive analysis; spatial regression; wildfire; R²; RMSE.

1 Introdução

As mudanças climáticas trazem impactos e repercussões em vários segmentos da vida humana, como, por exemplo, na Saúde, Infraestrutura, Meio Ambiente e Economia. Soluções de apoio à decisão e ao planejamento de medidas combativas podem ajudar gestores e responsáveis. Nesse contexto, técnicas computacionais que fazem uso de mineração de dados e aprendizado de máquina vêm sendo consideradas para avaliar ou prever valores relacionados a fenômenos subjacentes às queimadas, como nos trabalhos de [Pham et al. \(2020\)](#), em relação a uma província vietnamita, e o de [Ghorbanzadeh et al. \(2019\)](#), com respeito a uma província iraniana.

No tocante à região da Mata Atlântica Paraibana (uma das mesorregiões do Estado da Paraíba), não há ainda trabalhos de análises preditivas no que diz respeito a questões de incêndios florestais. Conforme o levantamento realizado para este trabalho, como descrito em [Xue et al. \(2021\)](#) e [Xu et al. \(2021\)](#), modelos de análise preditiva podem auxiliar na tomada de decisões e, no caso de dados geográficos, podem ser gerados a partir de métodos de Regressão Geograficamente Ponderada (RGP). RGP é um método de regressão linear que considera, para os valores de cada local de amostra, os valores do local de amostra da vizinhança, para o oferecimento de parâmetros, conforme [Dale and Fortin, 2014](#)), que podem ser utilizados como base para a elaboração de mapas temáticos úteis aos gestores e responsáveis. Dessa forma, investigar abordagens que possam prever tendências quanto ao potencial radioativo do fogo de queimadas é um tópico importante e pode fazer uso de métodos de RGP para utilizar os componentes geográficos como meios de obter análises mais assertivas.

A investigação neste trabalho busca, assim, prover um panorama mais claro, em termos de informações e análises que possam auxiliar processos de mitigação de queimadas (ou incêndios) em áreas como o bioma “Mata Atlântica”, que faz parte do estado da Paraíba, usando o município de João Pessoa como objeto de estudo. Isso é realizado a partir da definição de dois cenários, a saber: (i) com amostras que abrangem apenas o município e áreas próximas aos seus limites, e (ii) considerando a mesorregião da zona da mata paraibana e áreas próximas aos seus limites também para o município de João Pessoa.

Assim, o problema de pesquisa, objeto deste trabalho, consiste em como gerar um modelo preditivo que possa indicar o potencial radioativo do fogo na área citada e que sirva a estratégias de mitigação das consequências das queimadas. A contribuição deste trabalho está baseada em dados da década de 2010 e faz uso de métodos de regressão linear espacial.

A avaliação experimental foi desenvolvida de forma a gerar e avaliar modelos de regressão espacial. Os resultados foram avaliados por meio das métricas R2 ou coeficiente de determinação (que traz a confiabilidade de cada resultado, ou seja, se o modelo ajusta bem os dados) e RMSE (root mean squared error) que expressa os erros do modelo preditivo. A avaliação é realizada também de modo comparativo a um trabalho que serviu de base ([Xue et al., 2021](#)). Assim, foi verificado que os valores de R2 no presente trabalho, em geral, foram maiores que em [Xue et al. \(2021\)](#). Sendo que todos os valores de R2, incluindo os com

ajustamento desse R2, foram quase todos superiores aos do trabalho base em quase todos os cenários estudados.

O artigo está organizado da seguinte forma: a Seção 2 introduz alguns fundamentos e trabalhos relacionados. Na Seção 3, é abordada a metodologia experimental para os dois cenários definidos. A Seção 4 apresenta e discute os resultados obtidos com a avaliação experimental, e a Seção 5 tece algumas considerações e possibilidades de trabalhos futuros.

2 Referencial teórico e trabalhos relacionados

De acordo com [Kelleher et al. \(2020\)](#), a análise preditiva de dados é uma forma de construir e usar modelos que fazem previsões baseadas em padrões de dados históricos. Para isso, são construídos modelos com base em métodos de Aprendizado de Máquina (AM) capazes de extrair, a partir de uma variável dependente relacionada a variáveis independentes, valores (preditos) dessa mesma variável dependente.

A regressão linear é uma tarefa de AM supervisionado, tendo em vista que existe uma variável de destino a ser predita. Essa variável é numérica e contínua ([Harrington, 2012](#)). A regressão linear consiste, simplificada, na modelagem do AM a partir da Eq.(1):

$$y = ax + b \quad (1)$$

Nessa equação, os valores a e b são conhecidos como pesos da regressão. O processo de encontrar esses pesos é chamado de regressão. Uma vez encontrados, pode-se prever novos valores dado um outro conjunto de entradas ([Harrington, 2012](#)). Acrescentando-se o uso de componente espacial, através de uma função como o "Adaptive Bisquare", tem-se a regressão linear espacial. Ou seja, mantém-se a linearidade da equação, mas se acrescenta um fator a mais - o espacial. Inclusive, [Brunsdon et al. \(1996\)](#) descrevem Geographically Weighted Regression (GWR) como uma regressão ponderada geograficamente, que permite trazer os parâmetros para cada coordenada no espaço, a serem estimados e mapeados. Isso contribui para responder à seguinte questão: como compreender dados não-estacionários, mas que mudam conforme o lugar (a coordenada geográfica)?

Diferentemente da regressão linear básica, que parte do princípio de que todas as variáveis são completamente independentes e cujos valores estimados são universais, na regressão espacial têm-se variáveis que possuem uma característica em comum (a possibilidade de estarem situadas em um mesmo par de coordenadas), não sendo completamente independentes e cujos valores são alterados de acordo com o lugar a partir de uma calibração por mínimos quadrados a incluir nessas coordenadas.

Uma adaptação da regressão geograficamente ponderada (GWR) considerando múltiplas escalas vem sendo estudada e aplicada. Ela foi denominada de regressão geograficamente ponderada multiescala, do inglês Multiscale Geographically Weighted Regression (MGWR), de acordo com [Oshan et al. \(2019\)](#). Essa técnica é considerada menos

restritiva que a GWR pois nela são eliminadas as restrições de que todos os relacionamentos variam na mesma escala, minimizando o overfitting. A MGWR mitiga a colinearidade, dessa forma, permite que cada variável possa ser associada a uma diferente "bandwidth" (que equivale a células de uma matriz hipotética), generalizando o GWR (generalized additive model ou GAM), reformulando-o, através da seguinte equação presente no supracitado artigo:

$$y = \sum_{j=1}^k f_j + \varepsilon \quad (2)$$

Onde se tem uma função de suavização em f_j , multiplicada pelo somatório de uma matriz e tudo somado ao número de Euler. A matriz, no caso, é onde haverá os distintos valores de bandwidth, generalizados. Em síntese, termina por permitir a predição de valores da variável dependente a partir de todas as variáveis independentes.

Tanto para GWR como para MGWR é considerada a já citada função "Adaptive Bisquare", que atribui maior peso às variáveis geograficamente mais próximas. Essa função se baseia no conceito de autocorrelação espacial, que se trata da correlação entre a variável com ela mesma onde valores próximos tendem a ser familiares, conforme [Burrough et al. \(2015\)](#). Consequentemente, valores de amostras mais distantes tendem a ser diferentes.

É importante deixar evidente que não se tratam de técnicas de Aprendizado de Máquina Supervisionado. Mas de algoritmos que, mediante um método baseado no tratamento dos dados, permitem a geração direta de resultados. Não há a etapa de treinamento, mas o acréscimo da validação dos dados de entrada mediante as coordenadas. Por isso a necessidade de, ao realizar o fracionamento dos registros, preservar as coordenadas sem normalizá-las. Consequentemente, na avaliação dos resultados, as técnicas utilizadas foram sendo elaboradas por meio de um dado do relatório gerado (o R2) e do RMSE (ou Root Mean Squared Error).

O R2, ou coeficiente de determinação, indica a variância da variável dependente em relação às variáveis independentes, segundo [Fávero and Belfiore \(2020\)](#). Além desta métrica, para este trabalho, também é usada o RMSE, cujo valor deve ser inversamente proporcional ao R2. O RMSE é uma das métricas utilizadas para calcular erros entre valores observados (reais) e predições (hipóteses). Por meio da ponderação entre as duas medidas (R2 e RMSE), pode-se ter uma visão mais adequada do grau de confiabilidade dos resultados do modelo preditivo.

Nesse panorama, o trabalho de [Xu et al. \(2021\)](#) define a temperatura da superfície do solo como variável dependente. Seus valores são contínuos – ainda que não tratem do potencial radioativo do fogo. O trabalho traz resultados que se basearam tanto na regressão linear múltipla como na GWR. Seus resultados, obtidos e discutidos, mostram uma comparação entre três técnicas, a Multi-Factor Geographically Weighted Machine Learning (a técnica MFGWML), a Thermal Image Sharpening (TsHARP) e a High-Resolution Urban Thermal Sharpener (HUTS) – ou seja, três algoritmos ou técnicas –, sendo a primeira

técnica considerada a melhor para uma amostragem de cerca de 50 mil pontos. Tal trabalho aproxima-se deste por trazer resultados para uma variável que interfere na existência ou não das queimadas, e no uso de regressão espacial. Chegaram à conclusão que a técnica (a MFGWML) é a melhor, em relação aos outros dois algoritmos, ao utilizarem Mean Absolut Error (MAE), RMSE e R2 para avaliação.

O trabalho de [Xue et al. \(2021\)](#) foi considerado mais próximo a este e é usado como base inicial para comparação. As diferenças entre este trabalho e o artigo citado ocorrem principalmente com relação à variável dependente, que é a medida de concentração de uma substância no ar – além da inexistência, no artigo base, do uso da técnica com Multiescala ou "MGWR". Mesmo que a variável do potencial radioativo do fogo não tenha sido a variável dependente, mas uma das independentes – já que se objetivava achar um índice de qualidade do ar denominado PM2.5 –, encontrou-se alguma semelhança tanto quanto ao uso do método GWR como em relação aos uso das métricas R2 e RMSE para avaliação. Uma outra diferença diz respeito às grandezas físicas diferentes usadas nos trabalhos. Neste, foi usado o Megawatt normalizado e, no caso do artigo base, foi utilizada uma medida de concentração de substância no ar em metro cúbico não normalizado.

Há também os trabalhos que se utilizam de regressão espacial mediante outras ferramentas. Uma delas é o RSAGA, um pacote em R que, em 2007, já era utilizado para regressão espacial ([Brenning, 2008](#)). Outras ferramentas, além do software de geocomputação SAGA ou da linguagem R, é o QGIS, capaz de permitir operações dessa natureza. Além disso, há trabalhos que também se utilizam do próprio GWR e do MGWR, tais como em [Maiti et al. \(2021\)](#), que trabalharam com a problemática da COVID-19 nos Estados Unidos com base em regressão espacial global e local; e o trabalho de [Shabrina et al. \(2021\)](#), para a indústria das viagens e turismo.

Outros trabalhos, dessa vez se utilizando de Redes Neurais Artificiais (RNA), como os de [Elia et al. \(2020\)](#) e [Sayad et al. \(2019\)](#), ambos apresentando resultados positivos, porém focando-se em predição de ocorrência de queimadas propriamente ditas, existentes ou inexistentes, e não de valores contínuos. No primeiro caso, aliados também a uma regressão logística, utilizaram-se de polígonos de dados históricos georreferenciados, duas áreas de estudo na Itália, avaliando através de medidas de Acurácia e da área sob a curva (AUC). No segundo caso, utilizaram-se de conceitos e práticas de Big Data, Aprendizado de Máquina e Sensoriamento Remoto. No Aprendizado de Máquina supervisionado, as Redes Neurais Artificiais e Support Vector Machine (SVM). Nesses dois casos, utilizam-se valores categóricos, e não contínuos – o que significa que as técnicas e métodos em uso foram correspondentes a essa realidade.

3 Metodologia experimental

O problema deste trabalho, como já introduzido, consiste em como gerar um modelo preditivo que possa indicar o potencial radioativo do fogo na região da cidade de João Pessoa, no Brasil, – ainda que as amostras possam atingir uma extensão de toda a mesorregião da zona da mata paraibana. O objetivo é que o modelo preditivo gerado possa

auxiliar em estratégias de mitigação de consequências das queimadas.

Para a solução do problema em questão, dois cenários foram definidos e são descritos a seguir, bem como dados para cada um deles foram coletados e tratados. Da mesma forma, técnicas de regressão espacial foram estudadas e avaliadas.

3.1 Cenários e técnicas

Inicialmente, pretendia-se trabalhar com todas as amostras para toda a mesorregião da zona da mata paraibana, sem a diferenciação em dois cenários e sem o uso de técnicas para subamostragem.

Fundamental deixar evidente que as técnicas GWR e MGWR não determinam os cenários, mas sim a extensão geográfica das amostras. No primeiro cenário, foram consideradas amostras em coordenadas geográficas situadas em João Pessoa e suas proximidades, como mostra a [Fig. 1](#). Assim, buscou-se uma extensão geográfica das amostras (ou seja, de cada coordenada onde se situam os valores de radiação do fogo) concentradas no município. Quanto ao segundo cenário, foram consideradas amostras em coordenadas geográficas situadas por toda a mesorregião da zona da mata paraibana, e um pouco além dos limites, como universo de influência para as previsões dos valores de radiação do fogo também para os bairros desse município.

Quanto às técnicas GWR e MGWR, ambas pela biblioteca Python `mgwr`, a primeira é a mais simples e só considera a dimensão da variável dependente (a radiação do fogo), com apenas um valor de "bandwidth" ou quantidade de células de uma matriz hipotética; já na MGWR ou GWR com Multiescala, todas as variáveis são consideradas. É também importante explicar que, para ambas as técnicas, não há a convencional divisão clara entre dados de treino e dados de teste. Há o uso direto dos dados, após o particionamento através de técnica de separação como a validação cruzada.

Já quanto à necessidade de comparação entre os dois cenários de amostras, isso se deve sobretudo à questão das variáveis climáticas, que se situam em apenas uma coordenada, e não em milhares distribuídas mais amplamente. Tal situação fez pensar na possibilidade de, em caso de uso das variáveis climáticas, haver uma deterioração dos resultados, uma perda de informação e de confiabilidade – algo que precisou ser testado. Além disso, uma análise inicial a respeito de possíveis influências ou ausência de influências, conforme a extensão espacial das amostras, foi realizada para os dois cenários.

Assim, para cada cenário, houve definição própria de variáveis independentes. Por isso o Cenário 2 possui outra abrangência espacial de amostras, como apresentado na [Fig. 2](#), e outra escolha de variáveis independentes.

3.2 Dataset

Todos os dados foram coletados a partir das fontes: [FIRMS-NASA](#) e [INMET](#), e agrupados conforme chaves (variáveis de data que coincidiam), formando um só dataset derivado, distribuídos em Variáveis Independentes de Clima (Precipitação, Pressão Atmosférica, Temperatura do Orvalho,

Temperatura Média, Temperatura Máxima, Temperatura Mínima, Umidade Média, Umidade Mínima, Velocidade Máxima do Vento e Velocidade Média do Vento) e em Variáveis Independentes de Satélite e Topografia (Brilho, Brilho do tipo t31, medidas do pixel como Scan e Track, DiaNoite, Erosão e Altitude).

Todas as variáveis independentes de clima foram consideradas no Cenário 1, enquanto a variável dependente de potencial radioativo do fogo foi considerada para os dois cenários. Já no Cenário 2, todas as variáveis independentes foram consideradas, exceto as de clima, cuja descrição das variáveis pode ser encontrada no [Glossário do INMET](#).

Todas essas variáveis possuem coordenadas por cada amostra, distribuídas somente em João Pessoa e arredores (Cenário 1) e por toda a mesorregião da zona da mata paraibana e arredores (Cenário 2). Lembrando que os valores de latitude e de longitude não foram considerados como variáveis nem dependentes nem independentes, pois apenas determinam a localização geográfica de cada amostra.

Quanto às variáveis de Satélite e de Topografia, estas foram consideradas em todos os cenários. Sendo que as variáveis de Topografia foram derivadas com o auxílio do software QGIS a partir dos valores dos pixels de um mosaico. As imagens para o mosaico de onde os valores dos pixels foram obtidas e calculadas foram obtidas em [USGS](#).

Deve-se salientar que não foram utilizados dados do INPE (Instituto Nacional de Pesquisas Espaciais), devido à relativa pouca riqueza dos atributos para resultados adequados; nem foram utilizados dados da IBM pois, indiretamente, baseiam-se no próprio FIRMS da NASA.

Quanto aos tratamentos dos dados, eles incluíram renomeação de colunas, escolha de variáveis numéricas ou transformações de valores em números além da retirada de valores nulos. Foi realizada adicionalmente a normalização de valores das variáveis por intervalo. Além disso, deve-se lembrar do fato de que, para a regressão espacial, as coordenadas geográficas não podem ser normalizadas (nem foram utilizadas como variáveis durante os processamentos), pois são elas que definem a localização das amostras e não poderiam ser configuradas de outra forma.

Importante ressaltar que foi utilizada, para ambos os cenários, uma função para remover os valores nulos porque optou-se por não prejudicar as amostras com a inserção de novos dados sem um conhecimento mais específico da área no que diz respeito às variáveis com maior presença de valores nulos: os de clima. Em testes com preenchimento desses valores nulos tanto com mediana como com média, os resultados foram mais desencorajadores após todos os processamentos, o que significaria a necessidade complementar da presença de um especialista – para análise dos dados climáticos, que apresentaram o maior número de valores nulos. Para este trabalho, não foi possível ter o acompanhamento de um especialista do domínio para essas questões.

Como os valores das variáveis escolhidas variam muito, optou-se por um processo de normalização utilizando-se técnica com normalização por intervalo min-max, conforme [Kelleher et al. \(2020\)](#).

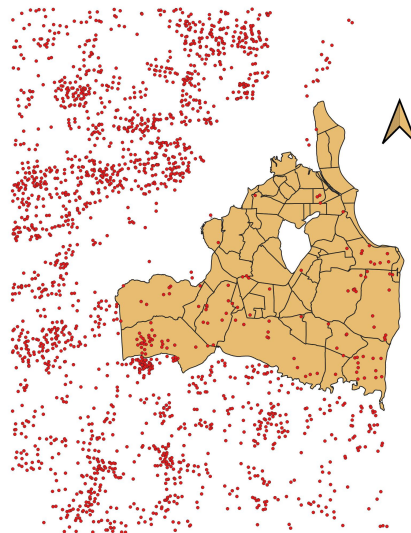


Figura 1: Representação espacial das amostras para o Cenário 1 sobre o município de João Pessoa, cujas geometrias dos bairros foram obtidas mediante o endereço eletrônico [Filipeia](#).

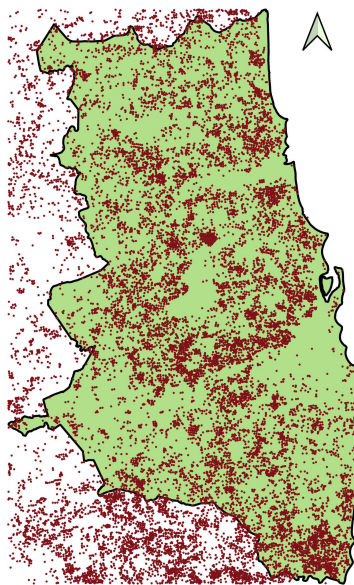


Figura 2: Representação espacial das amostras para o Cenário 2 sobre a mesorregião da Zona da Mata Paraibana, cuja geometria foi extraída das mesorregiões presentes no endereço eletrônico da [AESA](#).

3.3 Avaliação experimental

Para o particionamento das variáveis independentes e da variável dependente, no uso do algoritmo MGWR, foi utilizada a técnica K-FOLD (com 10 subconjuntos), com o intuito de utilização de apenas um décimo das amostras.

Em geral, os passos para ambos os métodos GWR e MGWR foram: configurar o seletor (com uso e configuração de função correspondente denominada "Sel_BW()", onde "sel" é de "selector" e "bw" ou "bandwidth" é referente às bandas ou células de uma matriz hipotética), realizar uma busca com esse seletor e, ao final, o ajuste – em ambos os casos também com as funções correspondentes, "gwr_selector.search()" e "GWR().fit()", por exemplo, no caso do GWR. Para a técnica MGWR, todas as palavras "gwr" são substituídas por "mgwr". Isso é feito através de adaptações do código presente em [Oshan et al. \(2019\)](#).

Quanto às variáveis topográficas, as de Índice de Rugosidade do Terreno (IRT) ou "erosão" e de Altitude, foram extraídas de um arquivo tif com coordenadas geográficas e, através da ferramenta de extração de valores de dados matriciais para pontos, do programa SAGA integrado ao QGIS, esses valores de quantidade de energia de cada pixel situado nas coordenadas dos pontos foi embutido. O arquivo tif em questão é um mosaico de quatro quadros que mostram o litoral nordestino (parte dele), que foi obtido de um da NASA.

Sobre as variáveis dos sensores dos satélites, estas podem ser encontradas no endereço eletrônico da [NASA](#).

4 Resultados e discussão

Os resultados, em relação aos modelos preditivos para cada um dos cenários, ainda que com diferenças próprias, apresentaram coerência entre R2 e RMSE. Essas diferenças correspondem às já citadas abrangências das amostras, geograficamente, e à ausência das variáveis de clima no Cenário 2.

Importante acrescentar sobre o RMSE que a comparação com o trabalho base teria que ser feita entre valores normalizados e, como o RMSE do base ocorreu sobre uma grandeza física diferente (um valor relacionado ao metro cúbico), isso não foi possível.

A [Tabela 1](#) traz os resultados obtidos nos experimentos de modo comparativo às métricas consideradas no trabalho base [Xue et al. \(2021\)](#).

Em suma, conforme salientado, de fato os valores de R2 não ajustado, em geral, foram maiores que os obtidos no artigo de [Xue et al. \(2021\)](#). Dessa forma, os valores mais altos de R2 foram quando do uso da técnica MGWR no Cenário 1, e, ao menos a partir de valores normalizados. O RMSE de 0.1650, na ??, demonstra que R2 foi um valor alto para acertos junto a valores baixos de erros – não havendo, por exemplo, valores altos de acertos com valores altos de erros, o que seria incoerente. Houve, nesse caso específico, mais acertos que erros.

Na [Fig. 3](#), o mapa mostra 6 classes e quebras naturais (técnica de divisão de intervalos para a determinação das classes), a partir de um arquivo geojson exportado dos parâmetros da variável "mgwr result.params" (separado por um underline). O mapa foi gerado usando o software QGIS

Tabela 1: Resultados e comparações

Trabalho:	R2	R2 Ajustado	RMSE
Base (baseado em Xue et al. (2021) , sem LO-OCV)	R2 = 0.834	R2 Ajustado = Não teve	RMSE = 3.46 µg por metro a -3
Base (baseado em Xue et al. (2021) , com LO-OCV)	R2 = 0.797	R2 Ajustado = Não teve	RMSE = 3.84 µg por metro a -3
Este Trabalho (GWR no Cenário 1)	R2 = 0.861	R2 Ajustado = 0.819	RMSE = 0.0759
Este Trabalho (MGWR no Cenário 1)	R2 = 0.885	R2 Ajustado = 0.831	RMSE = 0.1650
Este Trabalho (GWR no Cenário 2)	R2 = 0.705	R2 Ajustado = 0.661	RMSE = 0.1346
Este Trabalho (MGWR no Cenário 2)	R2 = 0.846	R2 Ajustado = 0.791	RMSE = 0.1283

(REF). Para isso, foi considerada a coluna "zero", com os parâmetros da variável dependente (potencial radioativo do fogo de incêndios florestais).

Nas Figuras [Fig. 4](#) e [Fig. 5](#), todos os mapas foram gerados no ambiente Python. Sendo que, no primeiro caso, tem-se o Cenário 1, com GWR e MGWR, respectivamente; e, no segundo caso, o Cenário 2, com GWR e MGWR, respectivamente.

É preciso deixar evidente se teve, em síntese, dois cenários, cujas diferenças se situam (i) nas amostras, geograficamente falando e (ii) na presença ou ausência das variáveis de clima. Para ambos os cenários, foram empregadas duas técnicas, a GWR, que se trata de uma regressão linear espacial mais tradicional; e a MGWR, que se trata de uma regressão linear espacial capaz de considerar não apenas todas as variáveis, mas, também, diferentes bandwidths ou, melhor dizendo, diferentes escalas.

No primeiro cenário, o GWR foi 136 de bandwidth, ou 136 amostras para a convergência. Enquanto que o MGWR foi de 156, porém apenas para a variável dependente (os valores de bandwidth para as variáveis independentes constam no relatório gerado durante o processamento). No segundo cenário, o GWR foi de 140 bandwidths e o MGWR convergiu em 86 bandwidths. Recordando que essas "larguras de banda" correspondem a cada célula ou a cada amostra. É como dizer que para a técnica MGWR aplicada ao Cenário 1, 156 registros, correspondentes a 156 bandas, foram suficientes para um resultado adequado.

Também foi visto que a ausência das variáveis de clima no segundo cenário não foi conclusivamente determinante

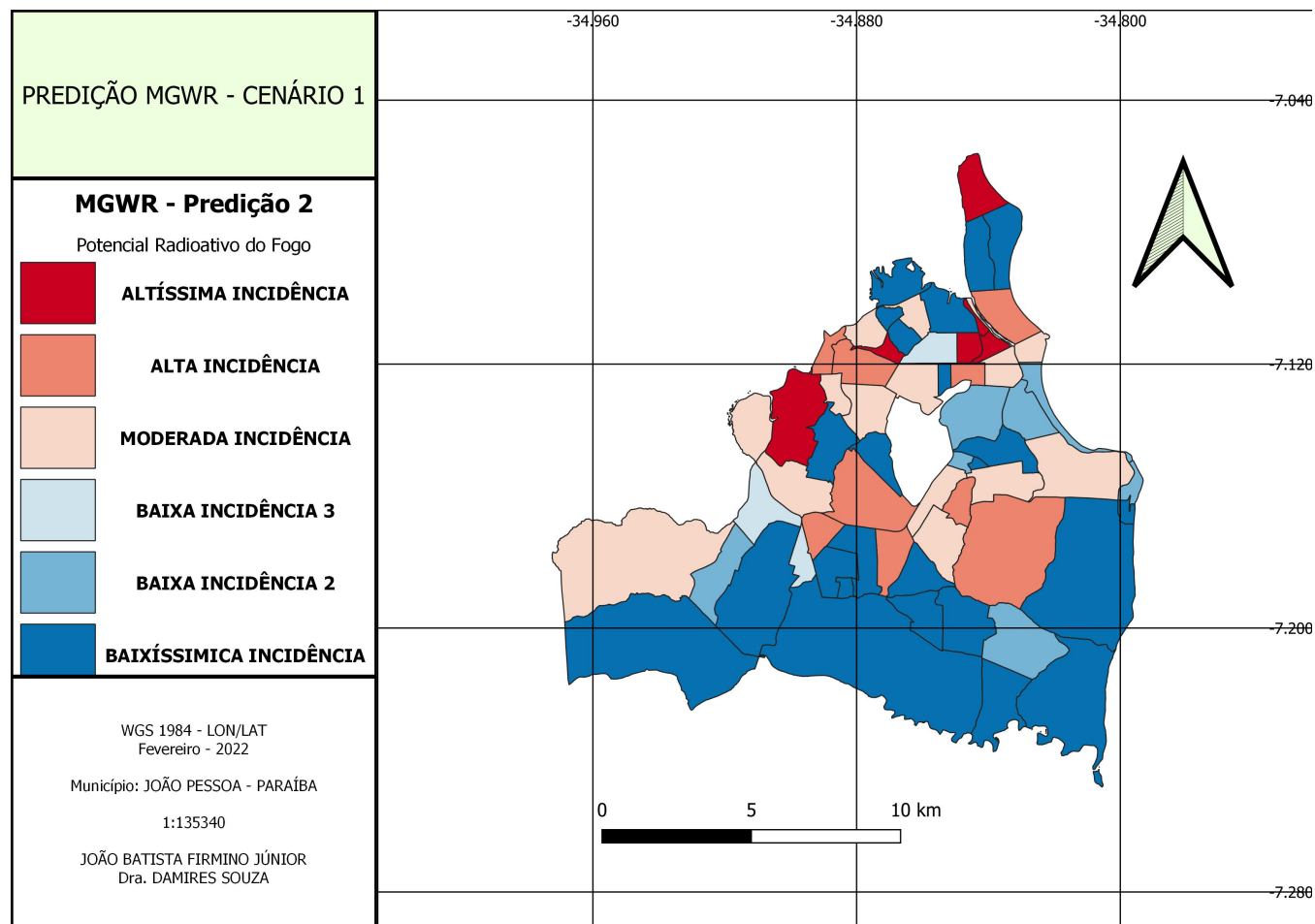


Figura 3: Mapa elaborado no QGIS com os valores preditos de potencial radioativo do fogo, em que quanto mais vermelho, maior a tendência de haver radiação dos focos de queimadas.

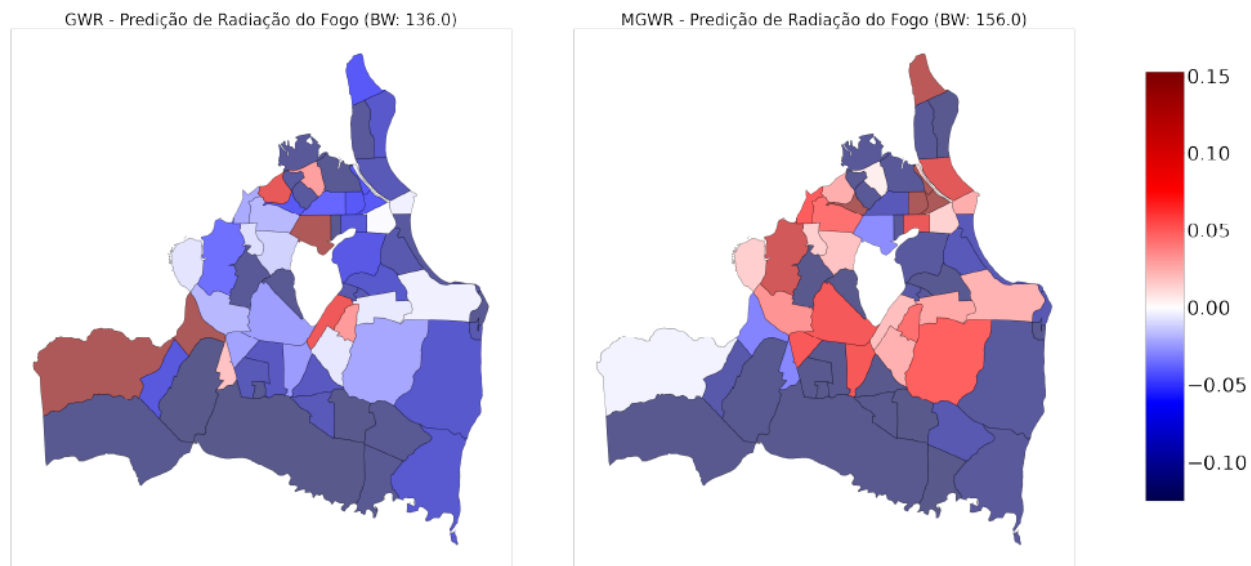


Figura 4: Mapas das predições do Cenário 1, para GWR e MGWR da esquerda para a direita, respectivamente.

de melhora ou piora de desempenho, sendo necessário mais cenários para avaliação.

5 Considerações finais e trabalhos futuros

Este trabalho foi realizado buscando comparar dois cenários de amostras geograficamente localizadas associadas ao município de João Pessoa. Os resultados obtidos na geração dos modelos preditivos e a avaliação por meio de métricas foram comparados a um trabalho base. Os resultados parecem confirmar a lógica de que dados mais próximos são mais parecidos que dados mais distantes geograficamente, sendo que a maioria deles foram com coordenadas de satélite e topográficos. Quanto aos dados de clima, estes vieram de uma estação localizada apenas em João Pessoa.

No primeiro cenário, os dados considerados foram apenas os relativamente mais próximos de João Pessoa e internos aos bairros do município, talvez por isso pôde-se elaborar um mapa mais confiável, sobretudo com a técnica MGWR, que considera "bandwidths" para todas as variáveis. Os resultados do segundo cenário foram também coerentes, mas não como no uso da técnica MGWR no Cenário 1, que se destacou tanto em relação a cada cenário como em relação aos experimentos do trabalho comparado.

É interessante acrescentar que tanto o trabalho relacionado quanto o levantamento realizado inicialmente nesta pesquisa realizada mostraram a inexistência de utilização de regressão linear espacial – até o momento – para prever informações sobre substâncias ou radiação relacionada a queimadas ou a incêndios na vegetação da Mata Atlântica. Isso se particulariza ainda mais no caso da mesorregião da zona da mata paraibana, com foco em municípios do litoral.

Como trabalhos futuros, pode ser útil pesquisar outras métricas de avaliação a serem usadas em regressão linear espacial, tal qual a Akaike e a Bayesiana (conforme melhor detalhado em [Oshan et al. \(2019\)](#)).

Há, sobretudo, a necessidade, em tempos de mudanças climáticas inesperadas, da utilização desses estudos e de sua maturação, até mesmo para prever ocorrências das queimadas em si – e não apenas, como faz este trabalho, o de uma medida relacionada a queimadas já ocorridas. Para tal, seria excelente a existência de acesso a dados de mais tipos de variáveis (demografia, proximidade das queimadas ocorridas no passado a determinados fatores de risco, mais variáveis topográficas etc.).

Assim, espera-se o melhoramento de três pilares: surgimento de mais bibliotecas como a mgwr, mais acesso a informações (incluindo as variáveis climáticas com mais estações, mais coordenadas disponíveis) e aprofundamento da pesquisa sobre quais as melhores métricas de avaliação para este tipo de estudo.

Enfim, pode-se dizer que seriam bem interessantes mais trabalhos futuros com GWR, sobretudo no Brasil, bem como com a presença de uma variável dependente como o "frp" ou potencial radioativo do fogo, para enriquecer a literatura da área com mais possibilidades de oferecer subsídios ao gerenciamento da gravidade das queimadas. Da mesma forma, trabalhos em que os valores de clima possam ser georreferenciados e medidos seguindo determinada metodologia de amostragem, por toda a área em questão, são importantes e promissores.

Referências

Brenning, A. (2008). Statistical geocomputing combining r and saga: The example of landslide susceptibility analysis with generalized additive models.

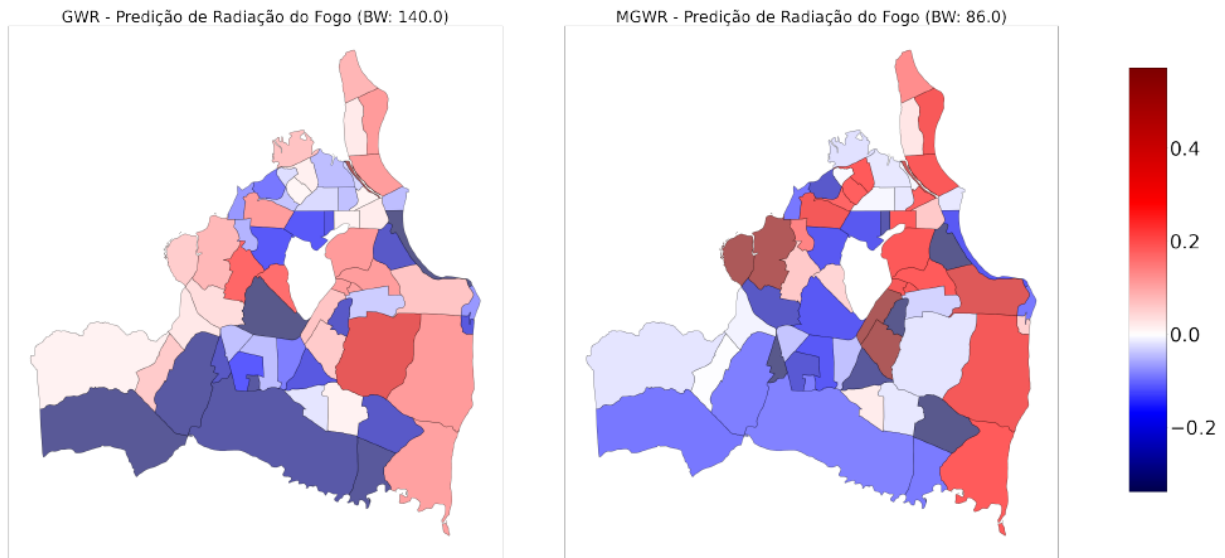


Figura 5: Mapas das predições do Cenário 2, para GWR e MGWR da esquerda para a direita, respectivamente.

- Brunsdon, C., Fortheringham, A. S. and Charlton, M. E. (1996). Geographically weighted regression: A method for exploring spatial nonstationarity, *Ohio State University Press* 28(4). <https://doi.org/10.1111/j.1538-4632.1996.tb00936.x>.
- Burrough, P., McDonnell, R. A. and D., L. C. (eds) (2015). *Principles of Geographical Information Systems*, Oxford United Press, United Kingdom.
- Dale, M. R. T. and Fortin, M.-J. (eds) (2014). *Spatial analysis: a guide for ecologists*, Cambridge University Press, United Kingdom.
- Elia, M., D'Este, M., Ascoli, D., Giannico, V., Spano, G., Ganga, A., Colangelo, G., Laforzezza, R. and Sanesi, G. (2020). Estimating the probability of wildfire occurrence in mediterranean landscapes using artificial neural networks, *Environmental Impact Assessment Review* 85. <https://doi.org/10.1016/j.eiar.2020.106474>.
- Fávero, L. P. and Belfiore, P. (eds) (2020). *Manual de análise de dados*, LTC, Rio de Janeiro.
- Ghorbanzadeh, O., Kamran, K. V., Blaschke, T., Aryal, J., Naboureh, A., Einali, J. and Bian, J. (2019). Spatial prediction of wildfire susceptibility using field survey gps data and machine learning approaches, *Fire* 2(43). <https://doi.org/10.3390/fire2030043>.
- Harrington, P. (ed.) (2012). *Machine Learning in Action*, Manning, Shelter Island, NY, USA.
- Kelleher, J. D., Namee, B. M. and D'Arcy, A. (eds) (2020). *Fundamentals of Machine Learning for Predictive Data Analytics: algorithms, worked examples, and case studies*, MIT Press, USA e England.
- Maiti, A., Zhang, Q., Sannigrahi, S., Pramanik, S., Chakraborti, S., Cerda, A. and Pilla, F. (2021). Exploring spatiotemporal effects of the driving factors on covid-19 incidences in the contiguous united states, *Sustainable Cities and Society* 68. <https://doi.org/10.1016/j.scs.2021.102784>.
- Oshan, T. M., Li, Z., Kang, W., Wolf, L. J. and Fotheringham, A. S. (2019). Mgw: A python implementation of multiscale geographically weighted regression for investigating process spatial heterogeneity and scale, *ISPRS Int. J. Geo-Inf.* <https://doi.org/10.3390/ijgi8060269>.
- Pham, B. T., Jaafari, A., Avand, M., Al-Ansari, N., Du, T. D., Yen, H. P. H., Van, P. T., Nguyen, D. H., Li, H. V., Mafi-Gholami, D., Prakash, I., Thuy, H. T. and Tuyen, T. T. (2020). Performance evaluation of machine learning methods for forest fire modeling and prediction, *Symmetry* 12(4022). <https://doi.org/10.3390/sym12061022>.
- Sayad, Y. O., Mousannif, H. and Moatassime, H. A. (2019). Predictive modeling of wildfires: A new dataset and machine learning approach, *Fire Safety Journal* 104: 130–146. <https://doi.org/10.1016/j.firesaf.2019.01.006>.
- Shabrina, Z., Buyuklieva, B. and Ng, M. K. M. (2021). Short-term rental platform in the urban tourism context: A geographically weighted regression (gwr) and a multiscale gwr (mgwr) approaches, *Geographical Analysis* 53: 686–707. <https://doi.org/10.1111/gean.12259>.
- Xu, S., Zhao, Q., Yin, K., He, G., Zhang, Z., Wang, G., Wen, M. and Zhang, N. (2021). Spatial downscaling of land surface temperature based on a multi-factor geographi-

cally weighted machine learning model, *Remote Sens.* **13**(1186). <https://doi.org/10.3390/rs13061186>.

Xue, Z., Gupta, P. and Christopher, S. (2021). Satellite-based estimation of the impacts of summertime wild-fires on pm_{2.5} concentration in the united states, *Atmospheric Chemistry and Physics* **21**. <https://doi.org/10.5194/acp-21-11243-2021>.