



Universidade de Brasília
Instituto de Ciências Humanas
Departamento de Geografia
Programa de Pós-graduação em Geografia

UNIVERSIDADE DE BRASÍLIA
PÓS-GRADUAÇÃO EM GEOGRAFIA

**PREVISÃO DE VULNERABILIDADE A INCÊNDIOS FLORESTAIS UTILIZANDO REGRESSÃO
LOGÍSTICA E REDES NEURAS ARTIFICIAIS: UM ESTUDO DE CASO NO DISTRITO
FEDERAL BRASILEIRO**

Pablo Pozzobon de Bem

Dissertação de Mestrado

Brasília-DF: dezembro de 2017

UNIVERSIDADE DE BRASÍLIA
PÓS-GRADUAÇÃO EM GEOGRAFIA

**PREVISÃO DE VULNERABILIDADE A INCÊNDIOS FLORESTAIS UTILIZANDO REGRESSÃO
LOGÍSTICA E REDES NEURAS ARTIFICIAIS: UM ESTUDO DE CASO NO DISTRITO
FEDERAL BRASILEIRO**

Pablo Pozzobon de Bem

Orientador: Osmar Abílio de Carvalho Junior

Dissertação de Mestrado

Brasília-DF: dezembro de 2017

UNIVERSIDADE DE BRASÍLIA
PÓS-GRADUAÇÃO EM GEOGRAFIA

**PREVISÃO DE VULNERABILIDADE A INCÊNDIOS FLORESTAIS UTILIZANDO REGRESSÃO
LOGÍSTICA E REDES NEURAS ARTIFICIAIS: UM ESTUDO DE CASO NO DISTRITO
FEDERAL BRASILEIRO**

Pablo Pozzobon de Bem

Dissertação de Mestrado submetida ao Departamento de Geografia da Universidade de Brasília, como parte dos requisitos necessários para a obtenção do Grau de Mestre em Geografia, área de concentração Gestão Ambiental e Territorial, opção Acadêmica.

Aprovado por:

Osmar Abílio de Carvalho Júnior, Doutor (GEA-UNB)
(Orientador)

Renato Fontes Guimarães, Doutor (GEA-UNB)
(Examinador interno)

Eraldo Aparecido Trondoli Matricardi, Doutor (EFL-UNB)
(Examinador externo)

Brasília-DF, 18 de dezembro de 2017

Ficha catalográfica

Bem, Pablo Pozzobon de

Previsão de vulnerabilidade a incêndios florestais utilizando regressão logística e Redes Neurais Artificiais: um estudo de caso no distrito federal brasileiro / Pablo Pozzobon de Bem; orientador Osmar Abílio de Carvalho Júnior -- Brasília, 2017.

39 p.

Dissertação de Mestrado -- Programa de Pós-graduação em Geografia -- Universidade de Brasília, 2017

- | | |
|---------------------------|--------------------|
| 1. Incêndios florestais | 2. Risco ambiental |
| 3. Aprendizado de máquina | 4. Regressão |

É concedida à Universidade de Brasília permissão para reproduzir cópias desta dissertação e emprestar ou vender tais cópias somente para propósitos acadêmicos e científicos. O autor reserva outros direitos de publicação e nenhuma parte desta dissertação de mestrado pode ser reproduzida sem a autorização por escrito do autor.

Pablo Pozzobon de Bem

Agradecimentos

Ao meu orientador, Dr. Osmar Abílio de Carvalho Júnior e aos demais docentes do Departamento de Geografia da Universidade de Brasília pela excelente postura para com os alunos de pós-graduação, motivando-os a buscarem um alto padrão científico.

Aos técnicos administrativos do Departamento de Geografia pelo apoio aos corpos docente e discente do curso.

À Coordenação de Aperfeiçoamento de Nível Superior (CAPES) pela bolsa concedida durante o curso de mestrado.

Ao professor Dr. Eraldo Aparecido Matricardi do Departamento de Engenharia Florestal por me atender e responder minhas dúvidas e fornecer sugestões para o desenvolvimento do trabalho.

Aos colegas de curso pelo compartilhamento de informações.

À minha família pelo apoio desde sempre.

Resumo: Incêndios florestais são um problema global e queimam milhões de hectares de vegetação nativa todos os anos. O Cerrado brasileiro é a savana neotropical mais rica em biodiversidade do mundo, e uma das regiões mais afetadas por incêndios, sendo considerado um ecossistema tolerante ao fogo. Apesar da adaptação do bioma ao fogo, a alta frequência de incêndios trazida pela ocupação humana tem danificado o ecossistema mais rápido do que ele é capaz de se recuperar. O combate a incêndios é custoso e, portanto, medidas de prevenção de incêndios são a melhor maneira de evitar seus danos em longo prazo. Prever a distribuição espacial da ocorrência de incêndios florestais é um passo importante para a realização do manejo do fogo. Para isto, podem ser utilizados modelos que relacionem a ocorrência do fogo às variáveis que o influenciam. Neste estudo, dois modelos distintos de previsão — Regressão Logística (RL) e uma Rede Neural Artificial (RNA) — foram aplicados à região do Distrito Federal brasileiro, que se encontra inserida dentro do bioma Cerrado. Produtos de área queimada baseados em imagens LANDSAT foram utilizados para gerar a variável dependente, e nove outras variáveis espacialmente explícitas e de origem antropogênica ou ambiental foram utilizadas como variáveis independentes. Os modelos foram otimizados em função da melhor medida de *Area under Receiver Operating Characteristic (AUROC, ou simplesmente AUC)* a partir da seleção de atributos, e posteriormente validados utilizando dados reais de áreas queimadas. Os modelos mostraram performances similares, mas o modelo utilizando a RNA demonstrou melhor AUC (0.7755), e melhor acurácia ao classificar áreas não queimadas (73.39%), porém pior acurácia média (66.55%) e ao classificar áreas queimadas, para as quais o modelo LR apresentou o melhor resultado (65.24%). Adicionalmente, foi comparada a importância de cada variável aos modelos, contribuindo para o conhecimento das causas principais de incêndios na região. As variáveis demonstraram importâncias similares em ambos os modelos utilizados, e as variáveis de maior importância foram a elevação do terreno e o tipo de uso do solo. Os resultados demonstraram bons desempenhos de todos os modelos testados, mas recomenda-se a execução de mais estudos similares mais detalhados em outras áreas de na savana Brasileira, dado que ainda são poucos os estudos deste tipo.

Palavras chave: Incêndios florestais, vulnerabilidade ambiental, regressão logística, redes neurais artificiais

Abstract: Wildfires are a global problem, burning millions of hectares of natural forests every year. The Brazilian Cerrado is the richest neotropical savanna of the world in regards to biodiversity, and one of the regions most affected by fires, and also considered a fire-dependent ecosystem. Despite being adapted to the occurrence of fire, the high frequency of wildfires in the region due to human occupation is damaging the ecosystem faster than it can recover. Fighting fires is costly, and therefore the best way to avoid damages in the long-term is through prevention techniques. Predicting the spatial distribution of wildfires is an important step towards proper wildfire management. For that purpose, models that relate the occurrence of fire to certain variables can be used. In this work, we applied two distinct prediction models — Logistic Regression (LR) and an Artificial Neural Network (ANN) — to the region of Brazil's Federal District, located inside the Brazilian Cerrado, the largest savanna in South America and the world's richest Neotropical Savanna. We used LANDSAT based burned area products to generate the dependent variable, and nine different anthropogenic and environmental factors were used as the explanatory variables. The models were optimized via feature selection for best Area Under Receiver Operating Characteristic Curve (AUROC) and then validated with real burn area data. The models had similar performance, but the ANN model showed a better AUC value (0.7755) and better accuracy when evaluating exclusively non-burned areas (73.39%), while it had worse accuracy overall (66.55%) and when classifying burned areas, in which LR performed better (65.24%). Moreover, we compared the contribution of each variable to the models, adding some insight into the main causes of wildfires in the region. Variables had similar contributions to the models, and the main driving aspects of the distribution of burned areas in the region were the land use type and elevation. The results showed good performance for both models tested, but further research regarding wildfire in the Brazilian savanna is recommended, as such studies are still scarce.

Keywords: Wildfires, environmental vulnerability, logistic regression, artificial neural networks

Sumário:

CAPÍTULO I – APRESENTAÇÃO GERAL	1
1.1. Introdução	1
1.2. Revisão metodológica	2
1.2.1. <i>Fatores que influenciam o fogo</i>	2
1.2.2. <i>Métodos de previsão de risco de incêndio</i>	5
1.3. Objetivos	6
1.4. Estruturação da dissertação	6
Referências	8
CAPÍTULO II – PREDICTING WILDFIRE VULNERABILITY USING LOGISTIC REGRESSION AND ARTIFICIAL NEURAL NETWORKS: A CASE STUDY IN BRAZIL'S FEDERAL DISTRICT	11
Abstract	11
2.1. Introduction	12
2.2. Material and methods	13
2.2.1. Study area	13
2.2.2. Dependent variable	14
2.2.3. Explanatory variables	15
2.2.3.1. <i>Topographic variables</i>	15
2.2.3.2. <i>Normalized Difference Senescent Vegetation Index (NDSVI)</i>	15
2.2.3.3. <i>Land use</i>	16
2.2.3.4. <i>Distance from features</i>	16
2.2.3.5. <i>Population density</i>	16
2.2.3.6. <i>Superficial water supply</i>	16
2.2.4. Prediction models	17
2.2.4.1. <i>Logistic Regression</i>	18
2.2.4.2. <i>Artificial Neural Network</i>	18
2.2.5. Feature Selection	19
2.3. Results and discussion	19
2.3.1. <i>Model performance</i>	19
2.3.2. <i>Variable importance</i>	21
2.3.3. <i>Burn area prediction</i>	22
2.4. Conclusions	22
References	23
CAPÍTULO III – CONCLUSÕES E CONSIDERAÇÕES FINAIS	27

Lista de figuras:

Figura 1. Enquadramento das principais variáveis responsáveis pelo comportamento do fogo em incêndios florestais. Modificado de CHUVIECO et al. (2003) 3

Fig 1. Spatial location of the study area in Brazil. 14

Fig 2. Burned and non-burned areas observed between 2010 and 2015..... 15

Fig 3. Spatial distribution of the variables used in this study..... 17

Fig 4. Architecture of a Multilayer Perceptron artificial neural network..... 19

Fig 5. Model ROC curves and the best classification thresholds. 20

Fig 6. Validation of classification values when compared to the historical burn area data..... 21

Fig 7. Distribution of validation cases in relation to the total area. 21

Fig 8. Models with only and without specific variables. All values obtained after 10-fold cross-validation. 22

Fig 9. Class membership probability map for each model..... 22

Lista de tabelas:

Table 1. Recent studies on fire risk prediction and their prediction methods: Logistic Regression (LR), Multi-criteria Decision Analysis (MCDA), Artificial Neural Networks (ANN), Multiple Linear Regression (MLR), Random Forest (RF), Multivariate Adaptive Regression Splines (MARS), Maximum Entropy (MAXENT), Generalized Linear Models (GLM), Boosted Regression Trees (BRT), Support Vector Machines (SVM), Geographically Weighted Regression (GWR), and Poisson Regression (PR). 12

Table 2. Model rankings by AUC and the explanatory variables used for each of them. 20

Table 3. Model accuracies after validation. 20

Lista de abreviaturas:

ANN	Artificial Neural Network
ASTER	Advanced Spaceborne Thermal Emission and Reflection Radiometer
AUC	Area Under Curve
AUROC	Area Under Receiver Operating Characteristic
BRT	Boosted Regression Trees
DEM	Digital Elevation Model
GDEM	Global Digital Elevation Model
GIS	Geographic Information System
GLM	Generalized Linear Model
GWR	Geographically Weighted Regression
IBRAM	Instituto Brasília Ambiental
IBGE	Instituto Brasileiro de Geografia e Estatística
INPE	Instituto Nacional de Pesquisas Espaciais
LR	Logistic Regression
MARS	Multivariate Adaptive Regression Splines
MAXENT	Maximum Entropy
MCDA	Multi-criteria Decision Analysis
MLA	Machine Learning Algorithm
MLR	Multiple Linear Regression
NDSVI	Normalized Difference Senescent Vegetation Index
PR	Poisson Regression
RF	Random Forest
SVM	Support Vector Machine
SWIR	Shortwave Infrared
USGS	United States Geological Survey
WEKA	Waikato Environment for Knowledge Analysis

CAPÍTULO I – APRESENTAÇÃO GERAL

1.1. Introdução

Incêndios florestais são um problema mundial, queimando áreas na ordem de milhões de hectares todos os anos. Na América do Sul a maior parte das áreas queimadas estão presentes dentro das formações savânicas como o Cerrado brasileiro (GIGLIO; RANDERSON; VAN DER WERF, 2013).

O Cerrado é o segundo maior bioma brasileiro e um dos “hotspots” de conservação da biodiversidade mundial. Porém, o avanço da fronteira agrícola e a crescente urbanização vêm há tempos devastando o bioma, restando hoje menos de um terço de sua cobertura vegetal original. Erroneamente, o Cerrado é considerado por muitos um bioma biologicamente pobre, sendo na realidade uma região de alto grau de endemismo de espécies e de alta biodiversidade, com uma diversidade estimada de 320 mil espécies animais e 20 mil espécies vegetais (AGUIAR et al. 2004, MENDONÇA et al. 2008).

O clima tropical sazonal que cobre grande parte da região do Cerrado é caracterizado pelas duas estações climáticas claramente definidas. A estação chuvosa ocorre entre os meses de outubro e março, enquanto a estação da seca ocorre entre os meses de maio e setembro. Durante a seca a umidade atmosférica chega a níveis extremamente baixos e a contagem de dias sem chuva pode ultrapassar 100 dias. Em condições de extrema seca, agrava-se o risco de ocorrência de incêndios florestais, ocorrência relativamente comum no Cerrado.

O Cerrado é considerado um ecossistema dependente do fogo. Além da vegetação ser morfológicamente adaptada para sobreviver incêndios, o fogo pode auxiliar os processos de ciclagem de nutrientes do solo e de dispersão e germinação de sementes de espécies nativas (COUTINHO, 1990; KUHLMANN; RIBEIRO, 2016). Originalmente a ocorrência de incêndios no Cerrado se dava de maneira natural, sendo iniciados por quedas de raios sobre a vegetação propensa ao fogo. Posteriormente o homem tomou o papel de principal causador com sua chegada na região e estima-se que hoje mais de 90% das ocorrências tenham origem antrópica (FAO, 2006).

Apesar da adaptação do bioma ao fogo, a alta frequência de incêndios traz diversos problemas ao ecossistema e à sociedade. Além de danificar e matar a

vegetação, o fogo atinge também a fauna que habita no ambiente. A ação do fogo traz ao solo modificações de caráter físico, químico e biológico, sendo muitas dessas permanentes (REDIN et al, 2011). Ademais, incêndios frequentes pioram a qualidade do ar em suas proximidades graças a presença de fumaça e fuligem.

Devido aos altos custos e perigo envolvidos em combates a incêndios florestais, a melhor maneira de evitar seus danos é através da adoção de técnicas de prevenção. Típicas medidas de prevenção envolvem a eliminação ou redução de fontes de propagação, a educação da população local, a aplicação da legislação e a criação de planos de prevenção. A elaboração de mapas de risco de incêndios é uma das principais ferramentas de prevenção por fornecer às autoridades informações sobre áreas prioritárias às quais deve ser dedicada maior atenção para implementação de outras técnicas preventivas, além de fornecer informações que auxiliam estratégias de combate ao fogo (SOARES; BATISTA, 2007)

1.2. Revisão metodológica

1.2.1. Fatores que influenciam o risco de incêndios

As técnicas e análises utilizadas para definição de áreas de risco são diversas, porém de modo geral baseiam-se na relação entre certas variáveis e sua influência sobre o comportamento do fogo. Os principais fatores que definem o comportamento do fogo são a condição do material combustível, as condições climáticas, o tipo de cobertura vegetal e fatores associados à presença humana. As variáveis podem ser divididas entre aquelas que afetam o risco de ocorrência (ignição) do fogo e as que influenciam a maneira como o fogo se propaga (CHUVIECO et al., 2003) (Figura 1).

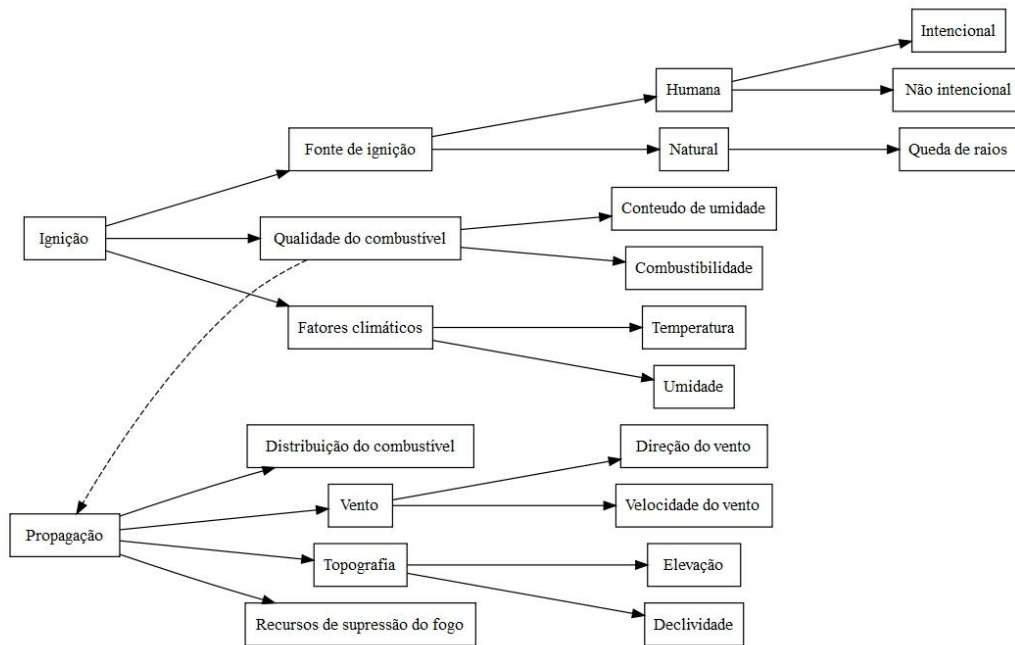


Figura 1. Enquadramento das principais variáveis responsáveis pelo comportamento do fogo em incêndios florestais. Modificado de CHUVIECO et al. (2003)

Quanto ao material combustível, deve-se levar em consideração a quantidade, a umidade, o tamanho e o arranjo do material e, no caso de espécies vegetais, a presença ou ausência de extrativos em seu interior. Espécies como as coníferas encontradas em regiões de clima temperado comumente possuem resinas de caráter inflamável, enquanto as espécies nativas do Cerrado brasileiro possuem lenhos adaptados para resistir ao fogo (MIRANDA, 2002; SOARES, 1985)

Segundo Oliveira (2002) condições climáticas como a temperatura, a umidade relativa do ar, a direção e velocidade do vento, e a precipitação possuem efeito notável tanto na ignição quanto propagação dos incêndios. Estes fatores atuam diretamente na modificação do conteúdo de umidade do material combustível, facilitando o início do fogo. A velocidade e direção do vento são fatores críticos para a tomada de decisões em combates aos incêndios, dado que o primeiro indica a taxa de propagação e o segundo sua possível direção.

Por ser um fator determinante do clima da região, a topografia é considerada também como um dos fatores que atuam na propagação do fogo. Apesar de ser um fator estático, o relevo define o comportamento dos ventos da região, além de definir o clima ou microclima local. As principais características a serem consideradas em relação à topografia são a elevação, responsável pelo perfil climático da área em questão; a declividade, que modifica a taxa e direção de propagação do fogo; e

orientação das encostas, que determina a quantidade de luz solar incidente sobre o solo e subsequentemente a temperatura de sua superfície. Além dos fatores principais mencionados, a presença ou não de barreiras como rios, lagos ou penhascos também é um fator a ser considerado (OLIVEIRA, 2002).

A cobertura vegetal, ou o tipo de vegetação também é uma característica importante a ser considerada. Vegetações mais densas tendem a manter um teor de umidade estável por criarem seu próprio microclima, enquanto vegetações mais esparsas estão mais sujeitas a perdas de umidade e possibilitam uma maior circulação de ar, facilitando a ocorrência do fogo (OLIVEIRA 2002).

A presença humana é hoje o principal agente causador de incêndios (MARTÍNEZ-FERNÁNDEZ et al, 2009). Áreas próximas a regiões densamente povoadas ou próximas a construções como rodovias, ferrovias e áreas de recreação devem ser tomadas como prioritárias para receber medidas de prevenção de incêndios devido à maior probabilidade de ignição por via de negligência humana. Este é um fator muitas vezes desconsiderado em metodologias de zoneamento por ser de difícil quantificação (MARTÍNEZ-FERNÁNDEZ et al, 2009; OLIVEIRA, 2002).

San-Miguel-Ayanz et al. (2003) dividem os modelos de previsão em três tipos em função da escala temporal em que as variáveis utilizadas para elaborá-los se encaixam:

- Modelos estáticos ou estruturais: compostos por fatores com pouca ou nenhuma variação ao longo do tempo. Topografia, tipo de cobertura e uso do solo, e certos fatores antropogênicos como a proximidade de rodovias são algumas das variáveis que comumente compõem estes modelos (PURI et al., 2011; VERMA; KUMAR, 2015). Estes modelos geram previsões para escalas temporais mais amplas, porém desconsideram o efeito da sazonalidade sobre o risco.

- Modelos dinâmicos: constituído por fatores com moderada ou grande variação ao longo do tempo. São compostos principalmente por variáveis associadas à sazonalidade, tais como temperatura, umidade, velocidade e direção dos ventos e a qualidade do material combustível (ALONSO-BETANZOS et al., 2002; HUESCA et al., 2009). As previsões geradas por estes modelos permanecem válidas para uma curta escala temporal.

- Modelos combinados ou integrados: formados pela combinação de fatores estáticos e fatores dinâmicos. Apresentam os melhores resultados por levarem em consideração os diferentes tipos de variáveis envolvidas na ocorrência do fogo, porém, assim como os modelos dinâmicos, são limitados a previsões em pequenas escalas temporais.

1.2.2. Métodos de previsão de risco de incêndio

Os métodos de previsão de risco de incêndio encontrados na literatura recente podem ser divididos em três grupos: modelos empíricos, modelos estatísticos e modelos baseados em algoritmos de aprendizado de máquina. Exemplos dos três métodos serão apresentados no capítulo II.

No Brasil, estudos recentes sobre o risco de incêndio têm se limitado ao uso de modelos empíricos como os elaborados em análises multicritérios (EUGENIO et al., 2016; CIPRIANI et al., 2011). Estes são métodos ponderativos nos quais escores e pesos são atribuídos às variáveis em função da opinião e experiência de especialistas, gerando um índice final de risco que se enquadra em faixas de risco, também determinadas por especialistas. A principal desvantagem deste tipo de modelo é a natureza subjetiva em sua elaboração.

Modelos estatísticos e algoritmos de aprendizado de máquina fundamentam-se na relação entre as variáveis independentes (clima, topografia, tipo de vegetação, etc.) e uma variável de resposta (quantidade de área queimada, número de focos, etc.), ou dependente. Os modelos estatísticos utilizados são em maior parte regressões. Dentre elas destaca-se a Regressão Logística, modelo comumente utilizado em estudos de risco de incêndio por calcular a probabilidade de ocorrência de um dado evento (WALKER; DUNCAN, 1967).

Algoritmos de aprendizado de máquina referem-se a modelos computacionais com a capacidade de aprender e realizar inferências sobre um conjunto de dados. Eles vêm sendo amplamente utilizados em diversos campos científicos devido à capacidade excepcional de aprender relações em problemas complexos e têm se tornado mais acessíveis recentemente graças à redução do custo de computação através de avanços no campo da informática (Jordan & Mitchell, 2015). Estudos sobre o risco de fogo têm feito uso destes algoritmos recentemente e têm demonstrado bons

resultados (BAR MASSADA et al., 2013; OLIVEIRA et al., 2012; AMATULLI et al., 2013)

1.3. Objetivos

O objetivo principal deste trabalho foi avaliar e comparar dois modelos diferentes quando utilizados para prever áreas de risco de incêndio a partir de um conjunto de variáveis independentes em uma região do Cerrado. O primeiro modelo, regressão logística, é um dos mais encontrados em estudos sobre o risco de fogo e consistentemente apresenta boa performance. O segundo modelo aplicado foi um algoritmo de Rede Neural Artificial, especificamente o Perceptron Multicamada

A região escolhida para realização do estudo foi a do Distrito Federal. O DF está inserido no Cerrado e é uma das unidades da federação com maior número de unidades de conservação da natureza. Segundo o Instituto Brasília Ambiental (IBRAM), o DF possui um total de 106 unidades, sendo 12 delas unidades federais. A região possui três tipos de cobertura terrestre conflitantes (áreas naturais, urbanas e agrícolas), fator importante para a avaliação de risco. Em adição, a dimensão da área de estudo permitiu que fosse observado o efeito das variáveis de modo mais localizado.

Os objetivos específicos desta pesquisa incluíram:

- Buscar a melhor combinação de variáveis para os modelos utilizando seleção de atributos;
- Qualificar os modelos em função de seus desempenhos;
- Avaliar a importância de cada variável e determinar os maiores responsáveis pelo fogo na região;
- Definir áreas de maior risco de incêndio;

1.4. Estruturação da dissertação

Esta dissertação foi elaborada no formato de apresentação de artigos científicos segundo as diretrizes estabelecidas para o Programa de Pós-graduação em Geografia da Universidade de Brasília.

Até então no primeiro capítulo foi apresentada uma breve descrição do tema, contendo o problema em questão, a justificativa e os objetivos gerais e específicos do trabalho.

No capítulo seguinte será apresentado o artigo intitulado “*Predicting wildfire vulnerability using logistic regression and artificial neural networks: a case study in Brazil’s Federal District*”. O Artigo visou cumprir os objetivos supracitados, e apresenta uma avaliação e comparação dos modelos utilizados, juntamente de uma avaliação das variáveis utilizadas, apresentado por fim os mapas de risco de incêndio gerados por cada modelo.

No último capítulo serão discutidas algumas conclusões e considerações sobre o desenvolvimento do trabalho e sobre o tema de modo geral, juntamente de algumas recomendações finais.

Referências

AGUIAR, L. M. S.; MACHADO, R. B.; MARINHO-FILHO, J. A diversidade biológica do Cerrado. In: Aguiar, L. M. S. & Camargo, A. J. A. **Cerrado: ecologia e caracterização**. Planaltina: Embrapa-CPAC, 2004.

ALONSO-BETANZOS, A. et al. **A Neural Network Approach for Forestal Fire Risk Estimation**. (Van Harmelen, Ed.) 15th European Conference on Artificial Intelligence. **Anais...**IOS Press, 2002

AMATULLI, G.; CAMIA, A.; SAN-MIGUEL-AYANZ, J. **Estimating future burned areas under changing climate in the EU-Mediterranean countries**. *Science of the Total Environment*, v. 450–451, p. 209–222, 2013.

BATISTA, A.C. Detecção de incêndios florestais por satélite. **Floresta**, Curitiba, v.34, p. 237-241, 2004.

BAR MASSADA, A. et al. **Wildfire ignition-distribution modelling: a comparative study in the Huron-Manistee National Forest, Michigan, USA**. *International Journal of Wildland Fire*, v. 22, n. 2, p. 174–183, 2013.

CIPRIANI, H. N. et al. **Fire Risk Map for the Serra de São Domingos Municipal Park, Poços de Caldas, MG**. *Cerne*, v. 17, n. 1, p. 77–83, 2011.

COUTINHO, L. M. Fire in the ecology of the Brazilian Cerrado. In J. G. Goldammer (Ed.), **Fire in the Tropical Biota**. v84, p. 82–105. Springer Berlin Heidelberg. 1990

CHUVIECO, E.; ALLGOWER, B.; SALAS, F. J. Integration of Physical and Human Factors in Fire Danger Assessment. In: CHUVIECO, E. (Ed.). **Wildland Fire Danger Estimation and Mapping: The Role of Remote Sensing Data**. 4. ed. Singapore: World Scientific Publishing, 2003. p. 197–218.

EUGENIO, F. C. et al. **Applying GIS to develop a model for forest fire risk: A case study in Espírito Santo, Brazil**. *Journal of Environmental Management*, v. 173, p. 65–71, 2016.

FAO Fire management – global assessment 2006. **A thematic study prepared in the framework of the Global Forest Resources Assessment 2005**. UN Food and Agriculture Organization, Forest Management Division, Forest Paper 151. (Rome, Italy) Disponível em: <<http://www.fao.org/docrep/009/a0969e/a0969e00.htm>> Acesso em 21 de julho 2017

GIGLIO, L.; RANDERSON, J. T.; VAN DER WERF, G. R. Analysis of daily, monthly, and annual burned area using the fourth-generation global fire emissions database (GFED4). **Journal of Geophysical Research: Biogeosciences**, v. 118, n. 1, p. 317–328, 2013.

HUESCA, M. et al. Assessment of forest fire seasonality using MODIS fire potential: A time series approach. **Agricultural and Forest Meteorology**, v. 149, n. 11, p. 1946–1955, 2009.

JORDAN, M. I.; MITCHELL, T. M. Machine learning: Trends, perspectives, and prospects. **Science**, v. 349, n. 6245, p. 255–260, 2015.

KUHLMANN, M., & RIBEIRO, J. F. Evolution of seed dispersal in the Cerrado biome:

ecological and phylogenetic considerations. **Acta Botanica Brasilica**, v30(2), p. 271–282. 2016

MARTÍNEZ-FERNÁNDEZ, J.; VEGA-GARCIA, C.; CHUVIECO, E. **Human-caused wildfire risk rating for prevention planning in Spain**. *Journal of Environmental Management*, v. 90, n. 2, p. 1241–1252, 2009.

MENDONÇA, R.C.; FELFILI, J.M.; WALTER, B.M.T.; SILVA JUNIOR, M.C.; FILGUEIRAS, T.S.; NOGUEIRA, P.E. & FAGG, C.W. Flora vascular do bioma Cerrado: checklist com 12.356 espécies. Pp. 423-1279. In: S.M. Sano; Almeida, S.P. & J.F. Ribeiro (eds.). **Cerrado: ecologia e flora**. v. 2. Brasília, Embrapa Informação e Tecnologia. 2008

MIRANDA, H. S.; BUSTAMANTE, M. M. C.; MIRANDA, A. C. **The Fire Factor**. p. 51-68 In: OLIVEIRA, P. S.; MARQUIS, R. J. *The cerrados of Brazil: ecology and natural history of a neotropical savanna*. Columbia University Press, New York. 368p 2002.

OLIVEIRA, D. dos S. **Zoneamento de risco de incêndios em povoados florestais no norte de Santa Catarina**. Dissertação, Universidade Federal do Paraná. Curitiba. 124p. 2002

OLIVEIRA, S. et al. **Modeling spatial patterns of fire occurrence in Mediterranean Europe using Multiple Regression and Random Forest**. *Forest Ecology and Management*, v. 275, p. 117–129, 2012.

PURI, K. et al. Forest fire risk assessment in parts of Northeast India using geospatial tools. **Journal of Forestry Research**, v. 22, n. 4, p. 641–647, 2011.

REDIN, M. et al. Impactos da queima sobre os atributos químicos, físicos e biológicos do solo. **Ciência Florestal**. Santa Maria, v. 21, n. 2, p. 381–392, 2011.

SAN-MIGUEL-AYANZ, J. et al. Current methods to assess fire danger potential. In: CHUVIECO, E. (Ed.). **Wildland Fire Danger Estimation and Mapping: The Role of Remote Sensing Data**. 4. ed. Singapore: World Scientific Publishing, 2003. p. 20–61.

SOARES, R. V. *Incendios Florestais: controle e uso do fogo*. Curitiba, Paraná: FUPEF. 213p. 1985

SOARES, R. V.; BATISTA, A. C.; **Incêndios Florestais: Controle, Efeitos e Uso do Fogo**. Universidade Federal do Paraná. 264p. 2007

VILAR DEL HOYO, L.; ISABEL, M. P. M.; VEGA, F. J. M. Logistic regression models for human-caused wildfire risk estimation: Analysing the effect of the spatial accuracy in fire occurrence data. **European Journal of Forest Research**, v. 130, n. 6, p. 983–996, 2011.

VERMA, A. K.; KUMAR, S. Mapping fire hazard in Rajaji National Park, future perspective of wildlife habitat conservation by using remote sensing and GIS. **International Journal of Technical Research and Applications**, v. 3, n. 5, p. 81–88, 2015.

WALKER, SH; DUNCAN, DB. "Estimation of the probability of an event as a function of several independent variables". **Biometrika**. v54: p. 167–178. 1967

CAPÍTULO II – PREDICTING WILDFIRE VULNERABILITY USING LOGISTIC REGRESSION AND ARTIFICIAL NEURAL NETWORKS: A CASE STUDY IN BRAZIL'S FEDERAL DISTRICT

Abstract: Predicting the spatial distribution of wildfires is an important step towards proper wildfire management. For that purpose, models that relate the occurrence of fire to certain variables can be used. In this work, we applied two distinct prediction models — Logistic Regression (LR) and an Artificial Neural Network (ANN) — to the region of Brazil's Federal District, located inside the Brazilian Cerrado, the largest savanna in South America and the world's richest Neotropical Savanna. We used LANDSAT based burned area products to generate the dependent variable, and 9 different anthropogenic and environmental factors were used as the explanatory variables. The models were optimized via feature selection for best Area Under Receiver Operating Characteristic Curve (AUROC) and then validated with real burn area data. The models had similar performance, but the ANN model showed a better AUC value (0.7755) and better accuracy when evaluating exclusively non-burned areas (73.39%), while it had worse accuracy overall (66.55%) and when classifying burned areas, in which LR performed better (65.24%). Moreover, we compared the contribution of each variable to the models, adding some insight into the main causes of wildfires in the region. Variables had similar contributions to the models, and the main driving aspects of the distribution of burned areas in the region were the land use type and elevation. The results showed good performance for both models tested, but further wildfire risk research in the Brazilian savanna is recommended, as such studies are still scarce despite the importance of the region.

Keywords: Wildfires, risk, vulnerability, logistic regression, artificial neural networks

2.1. Introduction

Wildfires are a recurring phenomenon that threaten native ecosystems around the globe by burning down millions of square kilometers of forests every year, damaging not only the native flora but also the fauna by destroying its natural habitat. The majority (71%) of burned areas worldwide between 1996 and 2012 took place inside savannas, with the bulk of them being in Africa and South America. Furthermore, in the portion of South America within the southern hemisphere, approximately 65% of the total burned areas during that same time span were distributed in such ecosystems (Giglio et al., 2013).

The largest savanna in South America is the ecoregion locally known as the Cerrado, Brazil's second largest biome, considered the world's richest Neotropical Savanna and one of the original 25 biodiversity hotspots for conservation proposed by Myers et al. (2000). Like in all savannas, fire typically plays a defining role in the Cerrado, and can impose benefic effects such as helping the flowering of plants and the dispersion of their seeds, and helping promote the cycling of nutrients in the soil (Conceição et al., 2013; Coutinho, 1990; Kuhlmann and Ribeiro, 2016). Despite that, the majority of wildfires in the region in the past were linked to human influence, and their increasing recurrence undermines the benefits of fire and damages the ecosystem faster than it can recover (Coutinho, 1990).

Controlling wildfires can be costly and dangerous. Therefore, prevention is essential for proper wildfire management. A series of precautionary measures can be employed to reduce the chances of ignition and to halt the spread of possible wildfires, but it is first necessary to define high-risk priority zones. One way of doing so is through the use of prediction methods, which can define vulnerable areas based on the correlation between wildfire occurrence and sets of environmental and anthropogenic variables. Recent studies have explored several different prediction models. Those models (Table 1) can be summarized into three main groups: Regressions, Multi-criteria Decision Analysis (MCDA) and Machine Learning Algorithms (MLA).

Table 1. Recent studies on fire risk prediction and their prediction methods: Logistic Regression (LR), Multi-criteria Decision Analysis (MCDA), Artificial Neural Networks (ANN), Multiple Linear Regression (MLR), Random Forest (RF), Multivariate Adaptive Regression Splines (MARS), Maximum Entropy (MAXENT), Generalized Linear Models (GLM), Boosted Regression Trees (BRT), Support Vector Machines (SVM), Geographically Weighted Regression (GWR), and Poisson Regression (PR).

Authors	Location of Area of Study	Methods
Bisquert et al. (2011)	Spain	LR
Vilar del Hoyo et al. (2011)	Spain	LR
Cipriani et al. (2011)	Brazil	MCDA
Bisquert et al. (2012)	Spain	LR, ANN
Mahdavi et al. (2012)	Iran	MCDA

Oliveira et al. (2012)	Mediterranean Europe	MLR, RF
Chang et al. (2013)	China	LR
Bar Massada et al. (2013)	United States of America	GLM, RF, MAXENT
Sitanggang et al. (2013)	Indonesia	LR, ID3, J48
Jung et al. (2013)	India	MCDA
Adab et al. (2013)	Iran	MCDA
Amatulli et al. (2013)	Mediterranean Europe	MLR, RF, MARS
Mohammadi et al. (2014)	Iran	LR
Oliveira et al. (2014)	Southern Europe	GWR
Rodrigues et al. (2014)	Spain	LR
Rodrigues and de la Riva (2014)	Spain	RF, BRT, SVM
Chuvienco et al. (2014)	Spain	GWR
Eskandari and Chuvienco (2015)	Iran	LR
Boubeta et al. (2015)	Spain	PR
Jafari Goldarag et al. (2016)	India	LR, ANN
Zhang et al. (2016)	Australia	LR
Guo et al. (2016)	China	LR
Eugenio et al. (2016)	Brazil	MCDA
Ajin et al. (2016)	India	MCDA

Fire modeling studies are still lacking in Latin America, despite being one of the regions most affected by wildfire (Costafreda-Aumedes et al., 2017). In recent years, the few studies of fire risk prediction in Brazil used MCDA as their main approach. Despite its ease of use since it does not need a dependent variable, this method is based on the collective opinion of experts, and is, therefore, susceptible to subjectivity in the form of personal bias (Chuvienco et al., 2003). Given the availability of countrywide spatiotemporal data, objective methods such as Regressions and MLAs can be used to achieve better results.

The aim of this study was to use two models (Logistic Regression and an Artificial Neural Network) to attempt to estimate possible fire risk areas within the region of Brazil's Federal District, located within the Brazilian Cerrado. In addition to evaluating the performance of the fire risk assessment models, we have attempted to determine the key factors driving the fire in the region.

2.2. Material and methods

2.2.1. Study area

The Federal District is one of Brazil's 27 federative units and houses the country's capital, Brasília. It is located between latitudes of 16°20'S and 15°35'S and longitudes of

48°31'W and 47°28'W, encompassing a total area of 5801.9km² (Fig 1). Agriculture and pastures cover approximately 63% of its territory, while natural vegetation covers about 22% of its area. Urban zones and water bodies make up for the remaining 15% of the territory.

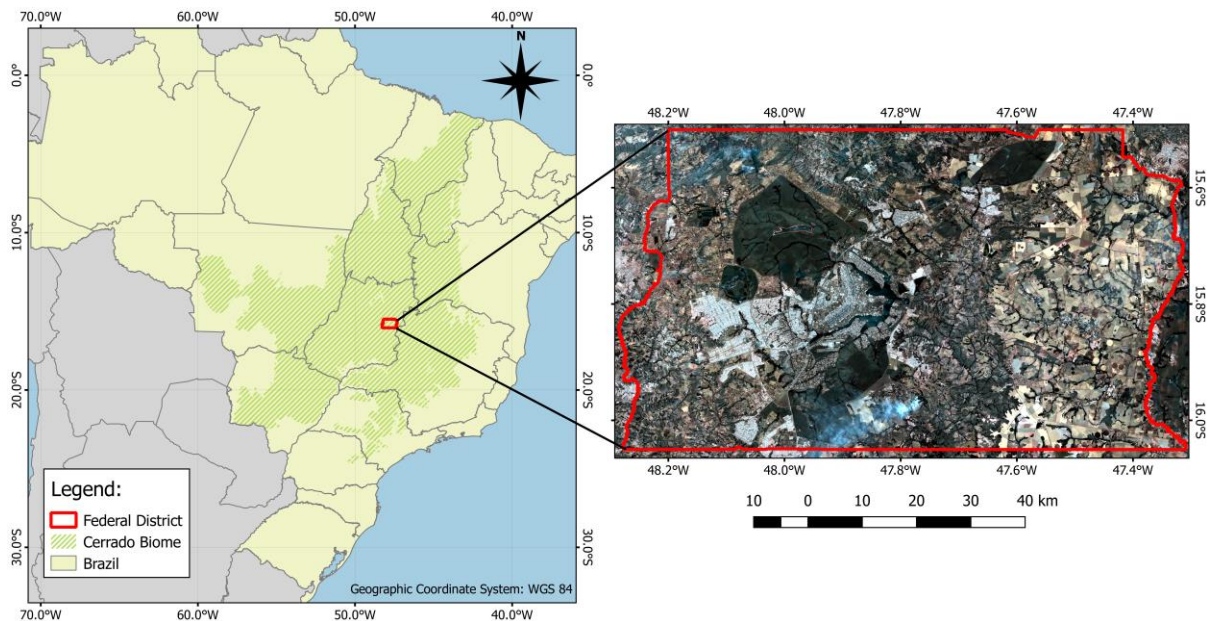


Fig 1. Spatial location of the study area in Brazil.

Based on the Köppen climatic classification system, the region is defined as a Tropical Savanna with distinct wet and dry seasons along the year. The natural vegetation in the study area comprises mainly a herbaceous stratum which consists of shrub savannas and grassland savannas, along with a woody stratum which consists of wooded savannas and savanna woodlands (Ribeiro and Walter, 2008). According to data from Brazil's National Institute for Space Research (*Instituto Nacional de Pesquisas Espaciais*, INPE), the number of wildfires in the region generally increases along with the start of the dry season, around May, and reaches its peak in August and September. With the beginning of the rainy season (by the end of September and early October) the number of fire occurrences drops sharply and stays at minimal numbers until the next dry season.

2.2.2. Dependent variable

To create our dependent variable, we used historical burned area data from the year of 2010 to 2015 (Fig 2). This data was produced by INPE (<https://prodwww-queimadas.dgi.inpe.br/aq30m/>) using the methodology proposed by Melchiori et al (2014). We randomly generated 5000 sampling points inside areas that had burned previously and 5000 points in areas that had never burned. The former were considered as positive cases (1), and the latter the negative ones (0). This process was done by using basic GIS tools in QGIS version 2.16.12 (QGIS Development Team, 2017).

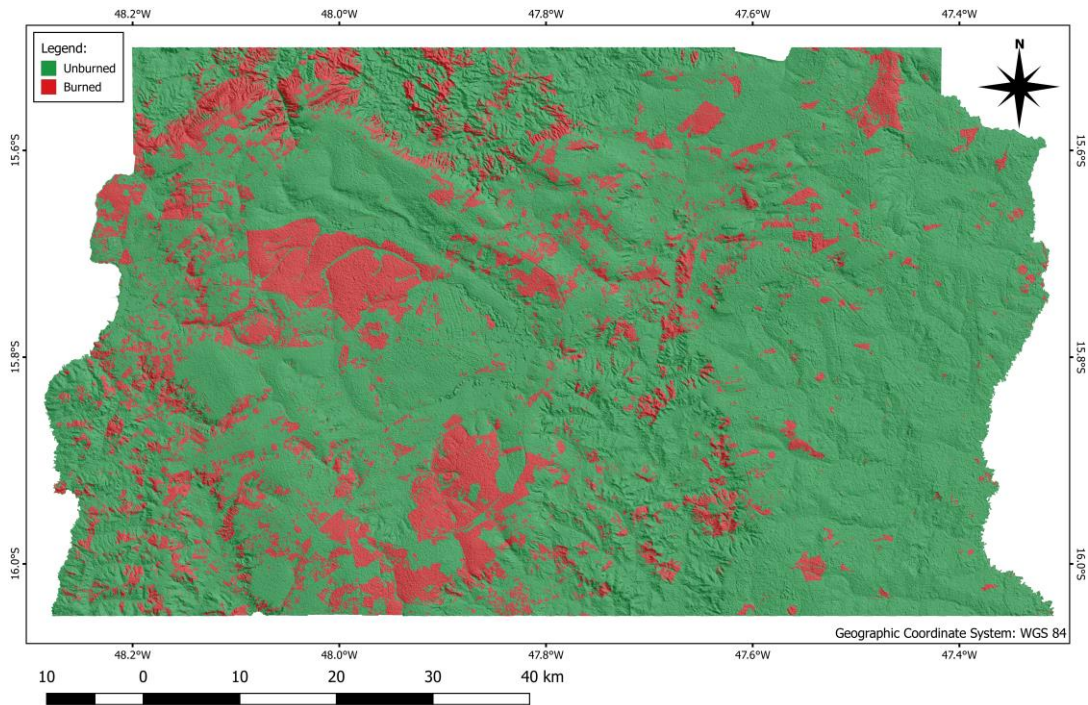


Fig 2. Burned and non-burned areas observed between 2010 and 2015.

2.2.3. Explanatory variables

A total of 9 different explanatory variables were chosen for this study, based on the knowledge of their relationship with fire and on their application in recent studies related to fire risk prediction (Bisquert et al., 2012; Chang et al., 2013; Mohammadi et al., 2014; Rodrigues et al., 2014; Sitanggang et al., 2013; Vilar del Hoyo et al., 2011; Zhang et al., 2016). Climatic variables such as temperature and humidity were ignored because of their homogeneous spatial distribution within the extent of the study area. Figure 3 shows the spatial distribution of the variables.

2.2.3.1. Topographic variables

The elevation layer consisted in a 30 meter spatial resolution raster extracted from the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) Global Digital Elevation Model (GDEM) product (Tachikawa et al., 2011), downloaded from the United States Geological Survey's (USGS) Earth Explorer platform (<https://earthexplorer.usgs.gov/>). Slope and aspect raster layers were produced from the GDEM, and the latter was converted into a categorical format by associating the slope orientation degree to the general directions North, East, South and West.

2.2.3.2. Normalized Difference Senescent Vegetation Index (NDSVI)

The NDSVI was developed and used by Marsett et al (2006) in order to assess senescent vegetation. That principle was the basis for the selection of this variable in this

study, considering that it is expected that dry, non-photosynthetically active vegetation is more vulnerable to fire. The index can be calculated following Eq. 4:

$$NDSVI = \frac{SWIR - RED}{SWIR + RED} \quad (4)$$

where SWIR and RED are the corresponding spectral bands, measured in atmospherically corrected reflectance. To generate this layer, we calculated the average NDSVI values for the month of September from 2011 to 2016, based on LANDSAT imagery obtained from the USGS.

2.2.3.3. *Land cover*

The land cover map for the area was downloaded from a database maintained by Brazil's Ministry of the Environment (*Ministro do Meio Ambiente*, MMA). The map accounted for urban areas, water bodies, exposed soil, agricultural lands, pastures, reforested areas, shrub savannas, grassland savannas, savanna woodlands, wooded savannas and gallery forests. The class vector was converted into a raster and a 30-meter spatial resolution was used.

2.2.3.4. *Distance from roads and urban zones*

These two variables were created by calculating the Euclidian distance to the nearest road or urban zone, resulting in raster layers with 30-meter spatial resolution in which each pixel contains the distance, in meters, to the corresponding feature (nearest road or urban area). Both were calculated based on data provided by Brazil's Institute of Geography and Statistics (*Instituto Brasileiro de Geografia e Estatística*, IBGE).

2.2.3.5. *Population density*

To generate this layer, we acquired tabular population data from IBGE distributed on a subdistrict basis, which was then linked to the respective vector polygons and converted into a raster file with 30m spatial resolution containing the number of habitants per square kilometer per grid cell.

2.2.3.6. *Superficial water supply*

This layer contained the values of total superficial water supply in cubic meters per second on a watershed basis. The Brazilian National Water Agency (*Agência Nacional de Águas*, ANA) provided the watershed dataset, which was converted from a vector to raster with spatial resolution of 30 meters.

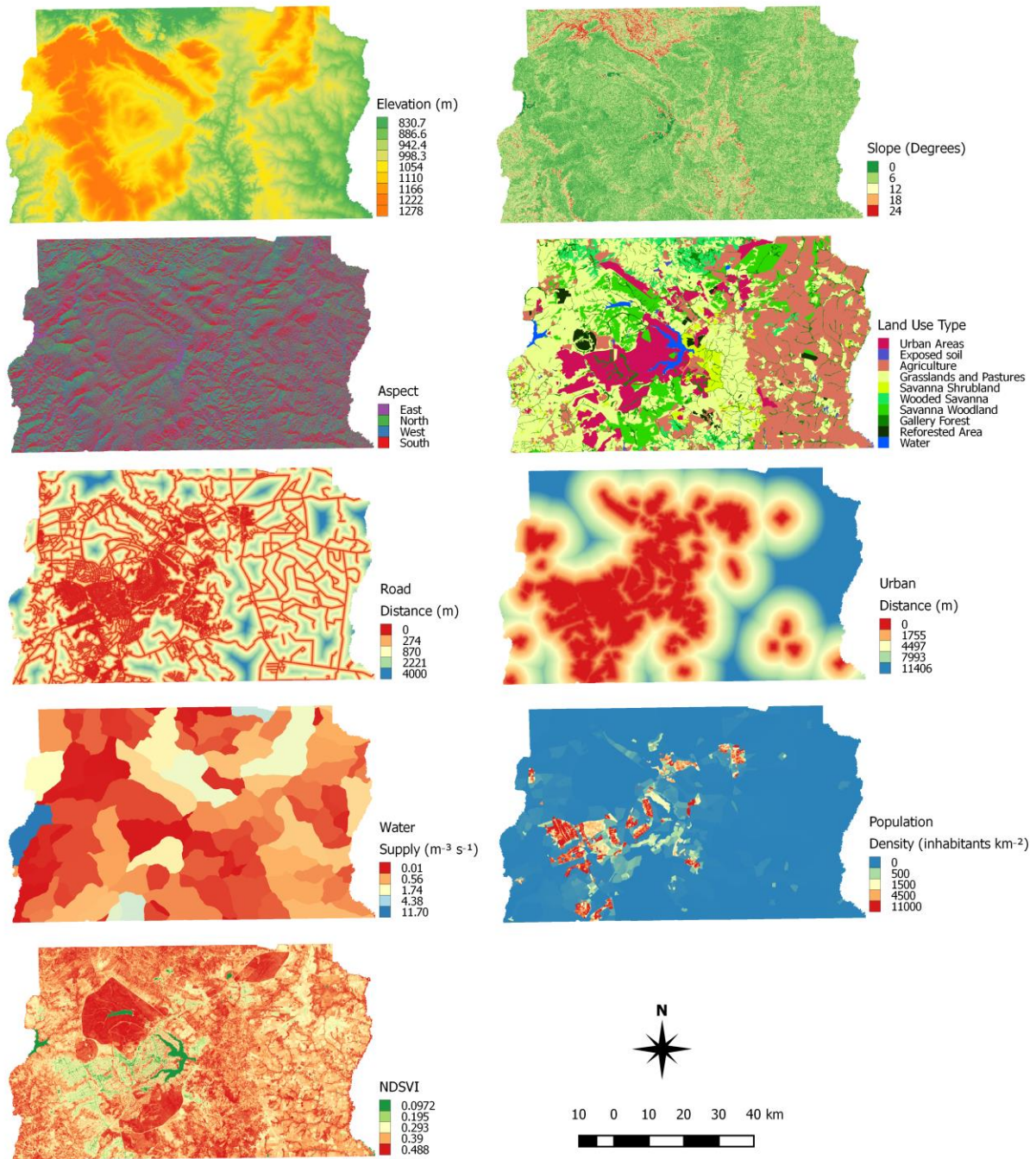


Fig 3. Spatial distribution of the variables used in this study.

2.2.4. Prediction models

We used two distinct models in this study to assess the probability of wildfire occurrence: Logistic Regression (LR) and an Artificial Neural Network (ANN), specifically the Multilayer Perceptron. The models were trained for best Area Under the Receiver Operating Characteristic Curve (AUROC, or simply AUC) and then compared and validated in function of their accuracy after setting an optimal decision threshold.

The training and validation of the models used R version 3.3.1 (R Development Core Team, 2017) in conjunction with the *RWeka* package (Hornik et al., 2009). Feature selection was performed by using the open-source software *Weka* (Waikato Environment for Knowledge Analysis) version 3.8.0 (Hall et al., 2009). Final models were applied to the raster dataset using the *raster* and *rgdal* R packages (Bivand et al., 2016; Hijmans, 2016).

2.2.4.1. Logistic Regression

Recent studies have shown that LR Models are an effective way to predict fire occurrence (Chang et al., 2013; Chuvieco et al., 2014; Eskandari and Chuvieco, 2015; Guo et al., 2016; Vilar del Hoyo et al., 2011; Zhang et al., 2016). This model estimates the probability that a given event may occur based on a set of independent explanatory variables and it can be expressed by the following equations:

$$P = \frac{1}{(1 + e^{-\beta})} \quad (1)$$

$$\beta = a_0 + a_1x_1 + a_2x_2 + a_3x_3 + \dots + a_nx_n \quad (2)$$

where P = probability of a point belonging to either class 1 or 0, respectively fire or non-fire in this study; a_n = the coefficient associated to independent variable n ; and x_n = value of independent variable n .

2.2.4.2. Artificial Neural Network

ANNs are functions that attempt to mimic the way a human brain makes decisions. They are particularly useful when dealing with complex non-linear classification problems (Alonso-Betanzos et al., 2002). The method is not used quite as often as Logistic Regression models, but recent fire studies have evidenced its worth as a predictive tool (Bisquert et al., 2012; Goldarag et al., 2016). In this research, the ANN architecture used was that of the Multilayer Perceptron (Fig 4). In this specific architecture, an input layer receives the values of a set of independent variables (x_i to x_p), which are fed into a hidden layer node j , multiplying each value by a weight $W_{i,j}^h$ and then adding them together to produce a value u_j . This value passes through nonlinear transfer function $f(x)$, which produces a value h_j that goes through another weighting process before passing on to the output layer. Finally, the weighted v_j values are added together and transferred into another transfer function $f(x)$, which outputs the final y_k values of the model (Haykin, 2008).

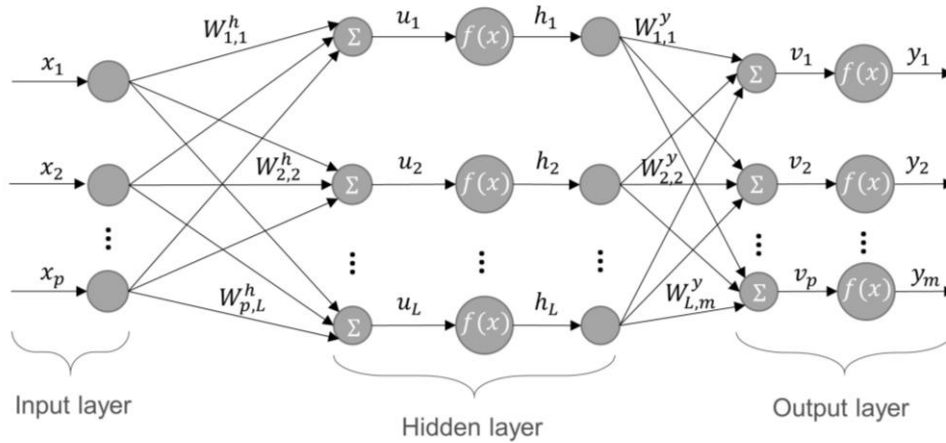


Fig 4. Architecture of a Multilayer Perceptron artificial neural network.

2.2.5. Feature selection

Feature selection is an optimization process used to extensively search for the best subset of input variables for a model based on a performance measure. In this study, we used the Genetic Search algorithm (Goldberg, 1989), implemented through *Weka*, version 3.8.0 (Hall et al., 2009). This algorithm attempts to imitate the process in which genetic information passes through generations of living beings. It generates a base population with a set number of random subsets and those with the best performance measure can share information, passing on their genetic information to the next generation and so on. The process repeats itself until it reaches a maximum number of generations, resulting in a final generation in which only the best models remain. In this paper, we selected the best model for each method and ranked them according to their AUC values after 10-fold cross-validation.

2.3. Results and discussion

2.3.1. Model performance

Table 2 shows the best instance of each model selected by the Genetic Search algorithm, ranked by AUC, and the variables used in each of them. Both model types showed good AUC values, ANN model achieved the best result among them, while LR performed only slightly worse. Both models had optimized thresholds close to the default value of 0.5 (Fig 5). The best subset of variables for both models was the same and excluded the aspect variable, leading to the conclusion that it had next to no effect on the occurrence of fire in the region. The aspect variable is related to the occurrence of fire by helping determine the amount of sunlight the ground surface receives throughout the day. Given that most of the region has very low slope values, most of its extent gets the same amount of sunlight despite the different aspects, so it shouldn't affect the spatial distribution of fires.

Table 2. Model rankings by AUC and the explanatory variables used for each of them.

Model	AUC	Variable								
		slope	water sup	aspect	elev	pop dens	urb dist	road dist	ndsvi	land use
ANN	0.7755	X	X		X	X	X	X	X	X
LR	0.7531	X	X		X	X	X	X	X	X

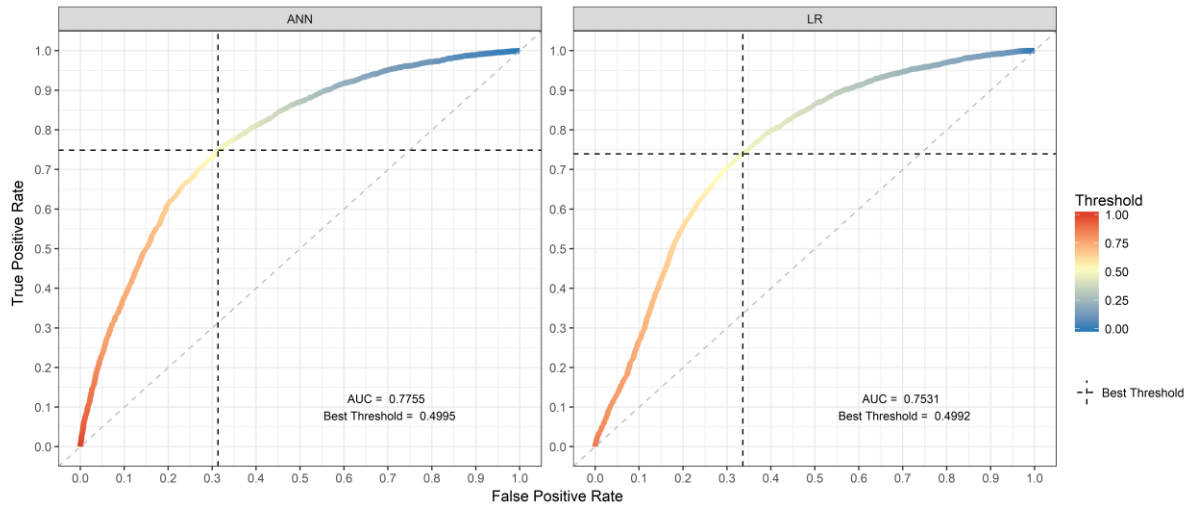


Fig 5. Model ROC curves and the best classification thresholds.

After classifying the land according to the probability values and the classification threshold we evaluated the models for their accuracy (Table 3). Both models had similar average accuracy measurements, but LR was more accurate when it came to burned areas, while the ANN had better accuracy with the negative class (unburned).

Table 3. Model accuracies after validation.

Model	Validation Accuracy (%)		
	Unburned	Burned	Average
ANN	73.39	59.70	66.55
LR	70.36	65.24	67.80

Fig 6 shows the spatial validation of the models when compared to the remainder of our burned area dataset. Considering the purpose of this study, special attention should be given to the false positive cases. Despite being treated as an error in traditional classification problems, they are highly significant in this study, as we assume that those cases share similar characteristics to the areas that have burned previously and therefore could be susceptible to burning in the future.

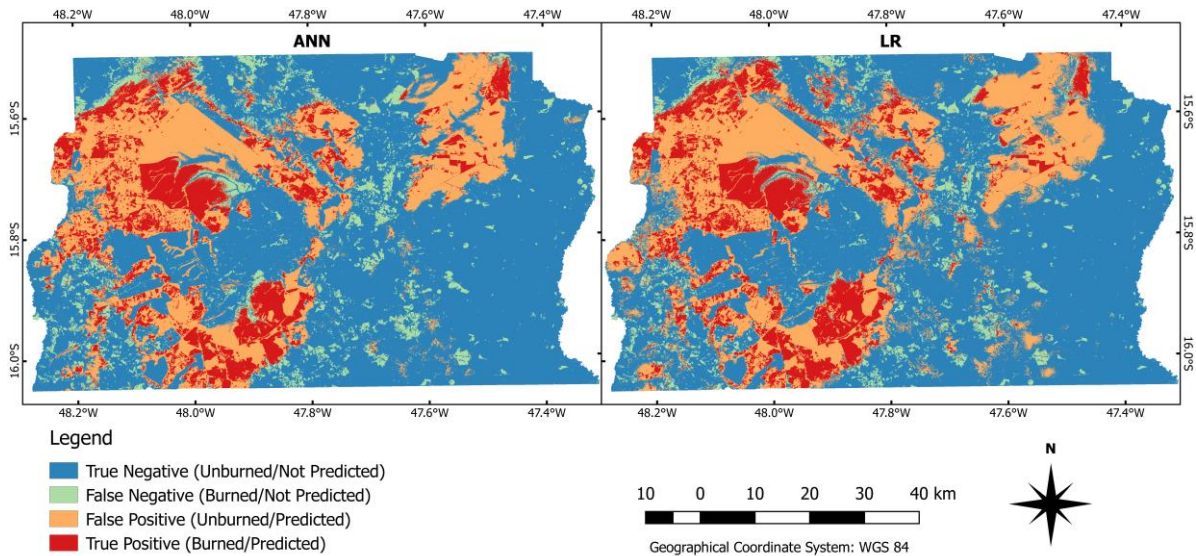


Fig 6. Validation of classification values when compared to the historical burn area data.

The number of false negative cases was also an important factor to consider as it is the main type of error in this study. It represents areas that have burned but were not classified as such by the models. In that regard, LR showed the least amount of false negative cases, although its values were very close to the percentage of false negatives from the ANN model (Fig 7).

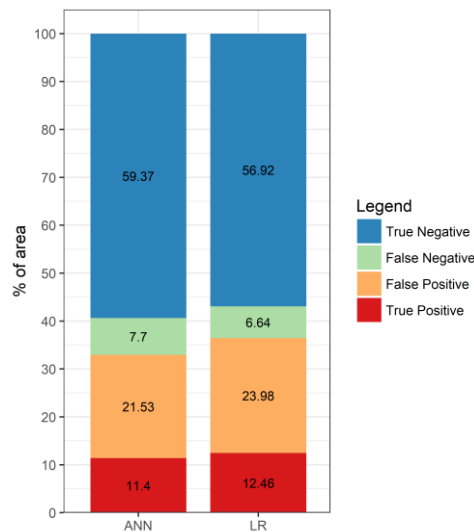


Fig 7. Distribution of validation cases in relation to the total area.

2.3.2. Variable importance

We attempted to describe the explanatory power of the variables by isolating them in models with only and without them (Fig 8). The variables showed similar explanatory power in both models. With the exception of aspect and water supply, every variable showed some explanatory power. Models using only land use or elevation had considerably higher AUC values than the rest of the variables. Water supply and aspect seemingly had no contribution

to the models, and in the case of the ANN the latter had a negative impact and removing it improved its AUC.

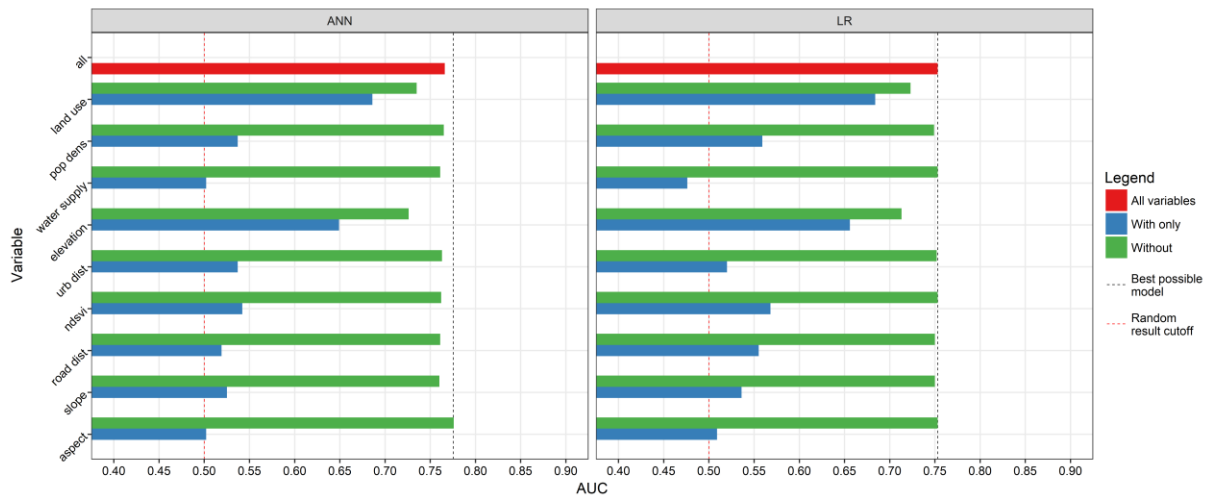


Fig 8. Models with only and without specific variables. All values obtained after 10-fold cross-validation.

2.3.3. Burn area prediction

Figure 9 shows the spatial distribution of class membership probabilities for each of the models. Both models produced maps with similar spatial distribution of risk, but the LR model concentrated its predictions within intermediate values, whereas the ANN model expected relatively more very high and very low risk areas.

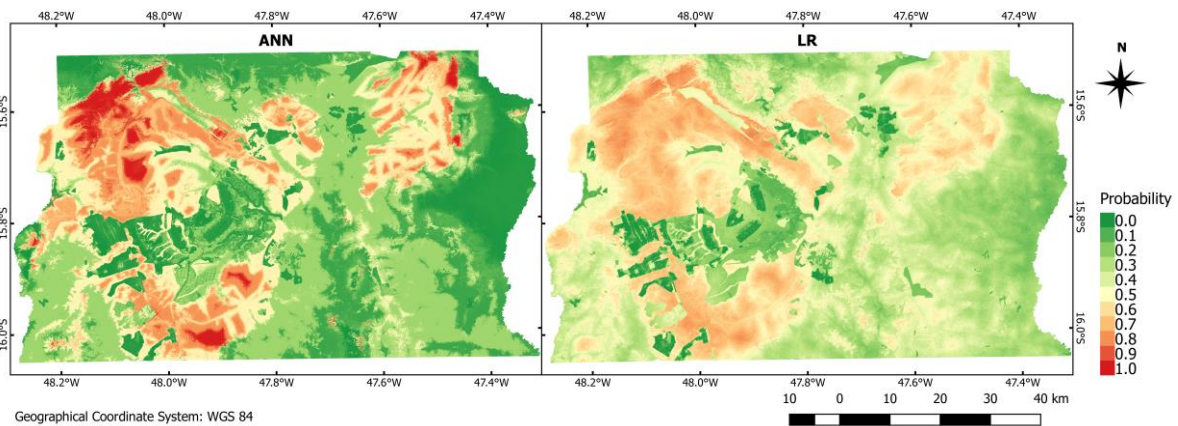


Fig 9. Class membership probability map for each model.

2.4. Conclusions

In this study, we successfully compared two different prediction models and obtained some insight into the relation between certain variables and the distribution of wildfires in a Brazilian savannah environment. Both models showed good results, and had similar values of accuracy and AUC. The ANN model showed a better AUC value and better accuracy when

evaluating exclusively non-burned areas, but it had worse accuracy regarding the positive class and overall.

Defining which model was the best when taking in mind the purpose of this study is a difficult task. Although the probability gradient produced by each model was different, the risk areas estimated by the models showed similar spatial distribution, especially considering true and false positive cases.

The studied variables had similar importance to both models, and the elevation and land use type variables had the biggest influence on their predictions. Given the size of our study area, climatic variables did not have significant spatial variation to justify changes in the distribution of fire. However, the spatial distribution of climatic variables in larger scales is less homogenous, and might have a more noticeable effect on the occurrence of fire.

Fire risk prediction studies in the Brazilian savannas are still scarce. We have concluded that it is possible to utilize methods such as LR and ANN with the data available in the country, and results reveal that they can offer reasonably good performances. Given the lack of comparable data, we recommend further investigation of the relationship between explanatory factors and the occurrence of wildfires in the Cerrado region, whether by exploration of new variables or further examination of the ones used in this work. Furthermore, we encourage studies in different scales, as our area of study area is only a small portion of the ecoregion and variables might work differently depending on the scale observed. Lastly, we recommend additional studies regarding the use of Machine Learning Algorithms for fire risk prediction, as the field is constantly growing and brings new and better algorithms every year.

References

- Adab, H., Kanniah, K. D., and Solaimani, K. (2013). Modeling forest fire risk in the northeast of Iran using remote sensing and GIS techniques. *Natural Hazards*, 65(3), 1723–1743. doi:10.1007/s11069-012-0450-8
- Ajin, R. S., Loghini, A., Vinod, P. G., and Jacob, M. K. (2016). RS and GIS Based Forest Fire Risk Zone Mapping in the Periyar Tiger Reserve, Kerala, India. *Journal of Wetlands Biodiversity*, 6, 139–148.
- Alonso-Betanzos, A., Fontenla-Romero, O., Guijarro-Berd nas, B., Hernández-Pereira, E., Canda, J., Jimenez, E., Luis Legido, J., niz, S. M., Paz-Andrade, C., and Inmaculada Paz-Andrade, M. (2002). A Neural Network Approach for Forestal Fire Risk Estimation. In Van Harmelen (Ed.), *15th European Conference on Artificial Intelligence* (pp. 643–647). IOS Press.
- Amatulli, G., Camia, A., and San-Miguel-Ayanz, J. (2013). Estimating future burned areas under changing climate in the EU-Mediterranean countries. *Science of the Total Environment*, 450–451, 209–222. doi:10.1016/j.scitotenv.2013.02.014
- Bar Massada, A., Syphard, A. D., Stewart, S. I., and Radeloff, V. C. (2013). Wildfire ignition-

- distribution modelling: a comparative study in the Huron-Manistee National Forest, Michigan, USA. *International Journal of Wildland Fire*, 22(2), 174–183. doi:10.1071/WF11178
- Bisquert, M., Caselles, E., Snchez, J. M., and Caselles, V. (2012). Application of artificial neural networks and logistic regression to the prediction of forest fire danger in Galicia using MODIS data. *International Journal of Wildland Fire*, 21(8), 1025–1029. doi:10.1071/WF11105
- Bisquert, M. M., Sánchez, J. M., and Caselles, V. (2011). Fire danger estimation from MODIS Enhanced Vegetation Index data: Application to Galicia region (north-west Spain). *International Journal of Wildland Fire*, 20(3), 465–473. doi:10.1071/WF10002
- Bivand, R., Keitt, T., and Rowlingson, B. (2016). rgdal: Bindings for the Geospatial Data Abstraction Library. R package version 1.1-10. Retrieved from <https://cran.r-project.org/package=rgdal>
- Boubeta, M., Lombardía, M. J., Marey-Pérez, M. F., and Morales, D. (2015). Prediction of forest fires occurrences with area-level Poisson mixed models. *Journal of Environmental Management*, 154, 151–158. doi:10.1016/j.jenvman.2015.02.009
- Brazil's Ministry of the Environment. (2016). i3GEO. Retrieved January 1, 2016, from <http://mapas.mma.gov.br/i3geo/>
- Brazil's National Institute of Space Research. (2016). Portal do Monitoramento de Queimadas e Incêndios. Retrieved September 30, 2016, from <http://www.inpe.br/queimadas/>
- Brazilian Institute of Geography and Statistics. (2016). Geoscience Downloads. Retrieved from http://downloads.ibge.gov.br/downloads_geociencias.htm
- Chang, Y., Zhu, Z., Bu, R., Chen, H., Feng, Y., Li, Y., Hu, Y., and Wang, Z. (2013). Predicting fire occurrence patterns with logistic regression in Heilongjiang Province, China. *Landscape Ecology*, 28(10), 1989–2004. doi:10.1007/s10980-013-9935-4
- Chuvieco, E., Aguado, I., Jurdao, S., Pettinari, M. L., Yebra, M., Salas, F. J., ... Martínez-Vega, F. J. (2014). Integrating geospatial information into fire risk assessment. *International Journal of Wildland Fire*, 23(5), 606–619. doi:10.1071/WF12052
- Chuvieco, E., Allgöwer, B., and Salas, F. J. (2003). Integration of Physical and Human Factors in Fire Danger Assessment. In E. Chuvieco (Ed.), *Wildland Fire Danger Estimation and Mapping* (4th ed., pp. 197–218). Singapore: World Scientific Publishing.
- Cipriani, H. N., Pereira, J. A. A., Silva, R. A., De Freitas, S. G., and De Oliveira, L. T. (2011). Fire Risk Map for the Serra de São Domingos Municipal Park, Poços de Caldas, MG. *Cerne*, 17(1), 77–83.
- Conceição, A. A., Alencar, T. G., Souza, J. M., Moura, A. D. C., and Silva, G. A. (2013). Massive post-fire flowering events in a tropical mountain region of Brazil: high episodic supply of floral resources. *Acta Botanica Brasilica*, 27(4), 847–850. doi:10.1590/S0102-33062013000400025
- Costafreda-Aumedes, S., Comas, C., and Vega-Garcia, C. (2017). Human-caused fire occurrence modelling in perspective: a review. *International Journal of Wildland Fire*, 26(12), 983. doi:10.1071/WF17026
- Coutinho, L. M. (1990). Fire in the ecology of the Brazilian Cerrado. In J. G. Goldammer (Ed.), *Fire in the Tropical Biota* (Vol. 84, pp. 82–105). Springer Berlin Heidelberg. doi:10.1007/978-3-642-75395-4_6

- Eskandari, S., and Chuvieco, E. (2015). Fire danger assessment in Iran based on geospatial information. *International Journal of Applied Earth Observation and Geoinformation*, 42, 57–64. doi:10.1016/j.jag.2015.05.006
- Eugenio, F. C., dos Santos, A. R., Fiedler, N. C., Ribeiro, G. A., da Silva, A. G., dos Santos, Á. B., Paneto, G. G., and Schettino, V. R. (2016). Applying GIS to develop a model for forest fire risk: A case study in Espírito Santo, Brazil. *Journal of Environmental Management*, 173, 65–71. doi:10.1016/j.jenvman.2016.02.021
- Giglio, L., Randerson, J. T., and Van Der Werf, G. R. (2013). Analysis of daily, monthly, and annual burned area using the fourth-generation global fire emissions database (GFED4). *Journal of Geophysical Research: Biogeosciences*, 118(1), 317–328. doi:10.1002/jgrg.20042
- Goldberg, D. E. (1989). *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison-Wesley. Addison-Wesley. doi:10.5860/CHOICE.27-0936
- Guo, F., Su, Z., Wang, G., Sun, L., Lin, F., and Liu, A. (2016). Wildfire ignition in the forests of southeast China: Identifying drivers and spatial distribution to predict wildfire likelihood. *Applied Geography*, 66, 12–21. doi:10.1016/j.apgeog.2015.11.014
- Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., and Witten, I. H. (2009). The WEKA data mining software. *SIGKDD Explorations*, 11(1), 10. doi:10.1145/1656274.1656278
- Haykin, S. O. (2008). *Neural Networks and Learning Machines* (3rd ed.). Upper Saddle River, NJ, USA: Pearson.
- Hijmans, R. J. (2016). raster: Geographic Data Analysis and Modeling. R package version 2.5-8. Retrieved from <https://cran.r-project.org/package=raster>
- Hornik, K., Buchta, C., and Zeileis, A. (2009). Open-source machine learning: R meets Weka. *Computational Statistics*, 24(2), 225–232. doi:10.1007/s00180-008-0119-7
- Jafari Goldarag, Y., Mohammadzadeh, A., and Ardakani, A. S. (2016). Fire Risk Assessment Using Neural Network and Logistic Regression. *Journal of the Indian Society of Remote Sensing*, 1–10. doi:10.1007/s12524-016-0557-6
- Jordan, M. I., and Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, 349(6245), 255–260. doi:10.1126/science.aaa8415
- Jung, J., Changjae, K., Jayakumar, S., Seongsam, K., Han, S., Dong, H. K., and Heo, J. (2013). Forest fire risk mapping of Kolli Hills, India, considering subjectivity and inconsistency issues. *Natural Hazards*, 65, 2129–2146. doi:10.1007/s11069-012-0465-1
- Kuhlmann, M., and Ribeiro, J. F. (2016). Evolution of seed dispersal in the Cerrado biome: ecological and phylogenetic considerations. *Acta Botanica Brasilica*, 30(2), 271–282. doi:10.1590/0102-33062015abb0331
- Mahdavi, A., Shamsi, S. R. F., and Nazari, R. (2012). Forests and rangelands' wildfire risk zoning using GIS and AHP techniques. *Caspian Journal of Environmental Sciences*, 10(1), 43–52.
- Marsett, R. R. C. R., Qi, J., Heilman, P., Biedenbender, S. H., Watson, M. C., Amer, S., Weltz, M., Goodrich, D., and Marsett, R. R. C. R. (2006). Remote Sensing for Grassland Management in the Arid Southwest. *Rangeland Ecology & Management*, 59(5), 530–540. doi:10.2111/05-201R.1
- Melchiori, A. E., W. Setzer, A., Morelli, F., Libonati, R., Cândido, P. de A., and Jesús, S. C.

- de. (2014). A Landsat-TM/OLI algorithm for burned areas in the Brazilian Cerrado: preliminary results. In *Advances in forest fire research* (Vol. 4, pp. 1302–1311). Imprensa da Universidade de Coimbra. doi:10.14195/978-989-26-0884-6_143
- Mohammadi, F., Bavaghar, M. P., and Shabanian, N. (2014). Forest Fire Risk Zone Modeling Using Logistic Regression and GIS: An Iranian Case Study. *Small-Scale Forestry*, 13(1), 117–125. doi:10.1007/s11842-013-9244-4
- Myers, N., Mittermeier, R. A., Mittermeier, C. G., da Fonseca, G. A. B., and Kent, J. (2000). Biodiversity hotspots for conservation priorities. *Nature*, 403(6772), 853–858. doi:10.1038/35002501
- Oliveira, S., Oehler, F., San-Miguel-Ayanz, J., Camia, A., and Pereira, J. M. C. (2012). Modeling spatial patterns of fire occurrence in Mediterranean Europe using Multiple Regression and Random Forest. *Forest Ecology and Management*, 275, 117–129. doi:10.1016/j.foreco.2012.03.003
- Oliveira, S., Pereira, J. M. C., San-Miguel-Ayanz, J., and Lourenço, L. (2014). Exploring the spatial patterns of fire density in Southern Europe using Geographically Weighted Regression. *Applied Geography*, 51, 143–157. doi:10.1016/j.apgeog.2014.04.002
- QGIS Development Team. (2017). QGIS Geographic Information System. Open Source Geospatial Foundation Project. Retrieved from <http://qgis.osgeo.org>
- R Development Core Team. (2017). R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <https://www.r-project.org/>
- Ribeiro, J. F., and Walter, B. M. (2008). As principais fitofisionomias do bioma Cerrado. In S. M. Sano & S. P. Almeida (Eds.), *Cerrado: Ambiente e Flora* (pp. 89–166). Embrapa Cerrados, Brasilia - Brazil.
- Rodrigues, M., and de la Riva, J. (2014). An insight into machine-learning algorithms to model human-caused wildfire occurrence. *Environmental Modelling and Software*, 57, 192–201. doi:10.1016/j.envsoft.2014.03.003
- Rodrigues, M., de la Riva, J., and Fotheringham, S. (2014). Modeling the spatial variation of the explanatory factors of human-caused wildfires in Spain using geographically weighted logistic regression. *Applied Geography*, 48, 52–63. doi:10.1016/j.apgeog.2014.01.011
- Sitanggang, I. S., Yaakob, R., Mustapha, N., and Ainuddin, A. N. (2013). Predictive models for hotspots occurrence using decision tree algorithms and logistic regression. *Journal of Applied Sciences*, 13(2), 252–261. doi:10.3923/jas.2013.252.261
- Tachikawa, T., Hato, M., Kaku, M., and Iwasaki, A. (2011). Characteristics of ASTER GDEM version 2. In *2011 IEEE International Geoscience and Remote Sensing Symposium* (pp. 3657–3660). IEEE. doi:10.1109/IGARSS.2011.6050017
- Vilar del Hoyo, L., Isabel, M. P. M., and Vega, F. J. M. (2011). Logistic regression models for human-caused wildfire risk estimation: Analysing the effect of the spatial accuracy in fire occurrence data. *European Journal of Forest Research*, 130(6), 983–996. doi:10.1007/s10342-011-0488-2
- Zhang, Y., Lim, S., and Sharples, J. J. (2016). Modelling spatial patterns of wildfire occurrence in South-Eastern Australia. *Geomatics, Natural Hazards and Risk*, 5705(March), 1–16. doi:10.1080/19475705.2016.1155501

CAPÍTULO III – CONCLUSÕES E CONSIDERAÇÕES FINAIS

Conclui-se que não só a Regressão Logística, mas também Redes Neurais Artificiais são métodos adequados para a elaboração de mapas de risco de incêndio no Cerrado.

No entanto, a área onde o estudo foi realizado é apenas uma pequena porção do bioma, e, portanto, os modelos elaborados neste trabalho podem ter dificuldades ao serem aplicados à totalidade da região não só devido à baixa representatividade dos dados em relação à escala do Cerrado, mas também devido às maiores amplitudes das variáveis em maiores escalas. Recomendam-se, portanto, estudos similares a este, porém em diferentes escalas e localidades.

Destaca-se que apenas dois métodos foram testados nesta pesquisa, dentre uma grande variedade de modelos e algoritmos disponíveis. O campo de aprendizado de máquina continua evoluindo, revelando algoritmos cada vez mais complexos e inteligentes. Portanto, é recomendável que sejam explorados novos algoritmos estudos similares no futuro.

Ainda existem lacunas no conhecimento de metodologias de análise de risco de incêndios florestais no Brasil. Para a elaboração de uma metodologia mais robusta é necessária uma avaliação holística do risco de incêndio, levando em consideração todas as variáveis envolvidas, ou ao menos as de maior poder estatístico explicativo.

A descentralização dos bancos de dados espaciais no Brasil representa um grande problema para o desenvolvimento de metodologias para análises de risco de incêndio. Seria desejável que fosse desenvolvido e estruturado um banco de dados de atualização contínua contendo diversos fatores correlatos ao fogo, juntamente de dados mais descritivos sobre ocorrências históricas do fogo. Não há, por exemplo, dados disponíveis online contendo as causas de ocorrência e local de ignição dos incêndios. O conhecimento dos principais fatores causadores de incêndios representa um grande passo no desenvolvimento de medidas preventivas de incêndios florestais.

No contexto da conservação da biodiversidade local, este estudo revelou que grande parte das áreas de risco se encontram dentro de áreas protegidas e delineadas segundo o Sistema Nacional de Unidades de Conservação segundo a lei

brasileira número 9.985 de julho de 2000. Dentre as principais áreas afetadas encontram-se:

- Área de Proteção Ambiental da bacia do rio Descoberto (Decreto nº 88.940 de 7 de novembro de 1983);
- Área de Proteção Ambiental dos ribeirões Gama e Cabeça de Veado (Decreto nº 9.417, de 21 de abril de 1986);
- Área de Proteção Ambiental da bacia do rio São Bartolomeu (Decreto nº 88.940 de 7 de novembro de 1983);
- Estação Ecológica de Águas Emendadas (Decreto nº 771 de 16 de agosto de 1968);
- Floresta Nacional de Brasília (Decreto s/nº de 10 de junho de 1999);
- Parque Nacional de Brasília (Decreto nº 241, de 29 de novembro de 1961 / Lei nº 11.285, de 8 de março de 2006);
- Reserva Biológica da Contagem (Decreto s/nº de 13 de dezembro de 2002).

Com o crescimento populacional contínuo no Distrito Federal, estas e outras Unidades de Conservação locais se encontram cada vez mais ameaçadas. Espera-se que este estudo possa vir a servir como fonte de auxílio para o manejo do fogo na região, e que os resultados aqui encontrados sirvam como fonte de motivação para a realização futuras pesquisas similares, em especial no Cerrado.