

Universidade de São Paulo
Escola Superior de Agricultura “Luiz de Queiroz”

Mapeamento da probabilidade de incêndio e de cicatrizes de dano como suporte
ao manejo florestal

Gabriel Atticciati Prata

Tese apresentada para obtenção do título de Doutor em
Ciências, Programa: Recursos Florestais. Opção em:
Silvicultura e Manejo Florestal

Piracicaba
2019

Gabriel Atticciati Prata
Engenheiro Florestal

Mapeamento da probabilidade de incêndio e de cicatrizes de dano como suporte ao manejo florestal

versão revisada de acordo com a resolução CoPGr 6018 de 2011

Orientador:
Prof. Dr. **LUIZ CARLOS ESTRAVIZ RODRIGUEZ**

Tese apresentada para obtenção do título de Doutor em Ciências, Programa: Recursos Florestais. Opção em Silvicultura e Manejo Florestal

Piracicaba
2019

**Dados Internacionais de Catalogação na Publicação
DIVISÃO DE BIBLIOTECA – DIBD/ESALQ/USP**

Prata, Gabriel Atticciati

Mapeamento da probabilidade de incêndio e de cicatrizes de dano como suporte ao manejo florestal / Gabriel Atticciati Prata. - - versão revisada de acordo com a resolução CoPGr 6018 de 2011. - - Piracicaba, 2019.

97 p.

Tese (Doutorado) - - USP / Escola Superior de Agricultura "Luiz de Queiroz".

1. Incêndio 2. Probabilidade de incêndio 3. Cicatriz de dano 4. Aprendizado de máquinas I. Título

**À MINHA ESPOSA, SOFIA
E AO MEU FILHO, CAIO,
DEDICO.**

AGRADECIMENTOS

À toda minha família, por ter me proporcionado os meios possíveis para que eu chegasse até aqui, especialmente à minha esposa Sofia e ao meu filho Caio, que estiveram ao meu lado todos os dias desta jornada.

Ao Professor Dr. Luiz Carlos Estraviz Rodriguez, meu orientador e amigo, que me apresentou caminhos, idéias, contatos e conhecimentos para o êxito deste trabalho.

À pesquisadora Brigitte Botequim, inspiradora desta tese, amiga, que dedicou bastante tempo, atenção e proporcionou excelente contribuição a este trabalho.

À Empresa Fibria Papel e Celulose por ceder os dados do estudo.

Ao Jefferson Polizel, pela amizade, enorme disposição e ensinamentos.

Ao pesquisador Andrew Hudak por ter viabilizado meu intercâmbio ao Serviço Florestal Americano (USDA FS) em Moscow/Idaho e proporcionado todo o suporte da estadia nos EUA, além das contribuições com idéias e conhecimento para o desenvolvimento desta tese.

Aos pesquisadores do USDA FS Ben Bright e Kas Dumroese, pela contribuição no aprendizado, amizade e todo apoio proporcionado e ao Lucca Schonhaler pela parceria e amizade.

Ao Anderson Santiago, Cristiano Reis, Daniel Papa, Danilo Pulga, Eduardo Ruas, Silvio Gomes e Tiago de Conto, pela contribuição com idéias, disponibilidade e conhecimento.

Aos colegas do Centro de Métodos Quantitativos - CMQ e do Grupo de Estudos da Tecnologia LiDAR - GETLiDAR.

Aos meus amigos.

À CAPES pela bolsa de doutorado concedida.

SUMÁRIO

RESUMO.....	7
ABSTRACT.....	8
1. INTRODUÇÃO.....	9
2. REVISÃO BIBLIOGRÁFICA.....	13
2.1. INCÊNDIO FLORESTAL.....	13
2.1.1. Causas dos Incêndios Florestais	13
2.1.2. Propagação de Incêndios Florestais	14
2.1.3. Risco e Perigo de Incêndios Florestais.....	16
2.1.4. Mapeamento de risco de incêndios florestais com SIG	17
2.2. SENSORIAMENTO REMOTO NO MAPEAMENTO DE INCÊNDIOS FLORESTAIS	18
2.2.1. Sensores Satelitais	19
2.2.2. Uso de Sensores Remotos no Monitoramento de Incêndios Florestais	22
2.2.3. Sensores LiDAR e aplicações florestais	23
2.2.4. Fusão de Sensores Remotos em aplicações florestais.....	26
2.2.5. Índices de Vegetação.....	27
2.3. APRENDIZADO DE MÁQUINAS E SUA APLICAÇÃO NA AVALIAÇÃO DE SINISTROS FLORESTAIS	27
2.4. MODELOS PREDITIVOS PARA MANEJO DE INCÊNDIOS FLORESTAIS.....	29
2.4.1. Regressão Logística.....	29
2.4.2. Random Forest.....	30
2.4.3. Support Vector Machine	31
3. MATERIAIS E MÉTODOS	33
3.1. ÁREA DE ESTUDO.....	33
3.2. SOBREVOO ALS	34
3.3. DADOS DE OCORRÊNCIA DE INCÊNDIO	34
3.4. BASE PARA MAPEAMENTO DA PROBABILIDADE DE INCÊNDIO	35
3.4.1. Preparo dos dados LiDAR – Modelagem 1	35
3.4.2. Estrutura da base de dados para modelagem.....	37
3.4.3. Ocorrências de incêndios	38
3.4.4. Parcelas de inventário.....	38
3.4.5. Modelagem de dados biométricos	38
3.4.6. Dados climáticos	39

3.4.7.	Dados sociais.....	40
3.4.8.	Modelos testados	42
3.5.	BASE PARA MAPEAMENTO DAS CICATRIZES DE INCÊNDIO	46
3.5.1.	Realidade de campo – Dados de ocorrência de incêndios.....	47
3.5.2.	Preparo dos dados LiDAR – Modelagem 2	47
3.5.3.	Dados do Satélite RapidEye.....	49
3.5.4.	Fusão das imagens entre sensores remotos.....	51
3.5.5.	Classes de treinamento.....	52
3.5.6.	Classificação de imagens	53
3.5.7.	Algoritmos para classificação.....	54
4.	RESULTADOS E DISCUSSÃO	57
4.1.	PROBABILIDADE DE INCÊNDIO.....	57
4.2.	MAPEAMENTO DE CICATRIZES DE INCÊNDIO.....	66
5.	CONCLUSÃO.....	75
	REFERÊNCIAS	77
	ANEXOS.....	93

RESUMO

Mapeamento da probabilidade de incêndio e de cicatrizes de dano como suporte ao manejo florestal

O histórico de incêndios florestais pode ajudar o gestor na localização de áreas de maior risco e, conseqüentemente, alocar de maneira mais eficaz os recursos de produção. Este trabalho mostra como o histórico de incêndios em florestas de eucalipto pode ser usado para gerar dois modelos preditivos, um de probabilidade de incêndio em nível de talhão e outro de cicatrizes que identificam áreas com cobertura florestal danificada por incêndio. O ajuste do modelo preditivo de probabilidade anual de incêndio, teve como variáveis preditivas, uma combinação de variáveis biométricas (volume comercial com casca), climáticas (face de exposição do terreno, precipitação anual, precipitação total anualizada, temperatura média anual e média de umidade relativa do ar), sociais (distância para área urbana, para estradas e para assentamento rural, população municipal, densidade demográfica, e população da zona rural) e de dados processados de levantamentos a laser aerotransportados (ALS): volume estimado por métricas ALS; índice de área foliar para altura total das árvores (LAI), para frações de altura de 1 a 5 metros (LAI_1_5m) e 1 a 10 metros (LAI_1_10m), e estimativa de sub-bosque (proporção entre LAI_1_5m e LAI). Foram utilizadas como técnicas de ajuste, a regressão logística (LOGIT) e o algoritmo Random Forest (RF), que se mostrou superior após o processo de validação-cruzada (tipo “k-fold”, com k=10). Dados ALS não se mostraram significativos, e o método RF com as variáveis volume comercial com casca, precipitação total anualizada, distância para áreas urbanas e para assentamentos e população da zona rural foi o de melhor eficácia. Esse resultado se expressou nas medidas de especificidade (classificação correta de áreas com registro de incêndio) e performance (classificação correta de áreas preditas como incendiadas). O melhor resultado revela especificidade e performance de 77%. Dentre as variáveis preditoras, a de maior importância foi a precipitação total anualizada. O modelo preditivo de cicatrizes de áreas com cobertura danificada, teve sua classificação baseada em três classes: Incêndio, Colheita/Terra Nua e Plantação. O ajuste utilizou como variáveis preditivas 16 métricas multiespectrais, derivadas do sensor RapidEye, e 29 métricas ALS. A resolução espacial das predições é de 5 m. Os algoritmos Support Vector Machine (SVM) e Random Forest foram usados como técnicas de classificação, que após a validação-cruzada (“k-fold” com k=10), identificou o RF como superior. Neste caso, a inclusão das métricas ALS ao cenário em que se usam apenas dados multiespectrais, aumentaram a sensibilidade para aspectos estruturais da vegetação, verificado para as classes “Incêndio” e “Plantação” e melhorou a acurácia das predições de 94%, para 97%, e o índice kappa de 90% para 95%. Por importância de capacidade preditiva de cicatrizes de dano, destacam-se as variáveis banda vermelho e NDVI para o RapidEye e, as variáveis relacionadas à cobertura e densidade do dossel, para os dados ALS. Os modelos gerados são úteis para gestores florestais, pois permitem melhor planejamento das operações de combate a incêndio, podendo, inclusive, reduzir custos na operação devido a melhor eficiência logística.

Palavras-chave: Incêndio; Probabilidade de incêndio; Cicatriz de dano; Aprendizado de máquinas

ABSTRACT

Fire risk and fire scars mapping as support for forest management

Historical forest fire data can help managers to locate risk areas and, consequently, allocate more efficiently production resources. This work shows how historical fire data from eucalyptus plantations can be used to generate two predictive models, one for fire probability at stand level and another of scars generated from areas with forest cover damaged by fire. The adjustment of the predictive model for fire probability used, as predictive variables, a combination of biometric (volume), climatic (aspect, annual precipitation, annualized total precipitation, annual mean temperature and mean relative air humidity), social (distance to urban area, to roads and to rural settlement, municipal population, demographic density, and rural population), and LiDAR variables: predicted volume by ALS metrics, leaf area index for tree's total height (LAI), and for fractions of 1 to 5 meters height (LAI_1_5m) and 1 to 10 meters (LAI_1_10m), and a shrub estimation (fraction between LAI_1_5_m and LAI). Logistic regression (LOGIT) and Random Forest (RF) algorithms were compared and RF achieved better accuracy after the 10-fold cross-validation. Adding LiDAR data resulted non significance, and the best adjustment for RF method used wood volume, annualized total precipitation, distance to urban areas, distance to settlements and rural population. The model predictive performance was evaluated by computing the specificity (correct classification of areas with fire registry) and performance (correct classification of areas predicted as burned). The best model yields specificity and performance of 77%. Among the predictive variables, the one that presented the greatest importance was the annualized total precipitation. The predictive fire scars model had its classification based on three classes: Fire, Harvest / BareLand and Plantation. The adjustment used as predictive variables, 16 multispectral metrics, derived from the RapidEye sensor, and 29 ALS metrics. The spatial resolution of the predictions is 5 m. The algorithms Support Vector Machine (SVM) and Random Forest were used as classification techniques, and, after the 10-fold cross-validation RF reached the best tune. In this case, combining ALS metrics to the scenario that used only multispectral data, the sensitivity increased for vegetation structure, verified for the "Fire" and "Plantation" classes, and improved the prediction accuracy from 94% to 97%, and the kappa index from 90% to 95%. Red band and NDVI were the dominant factors from RapidEye to predict fire scars pixels, and variables related to canopy cover and canopy density were the most important variables from the ALS data. The generated models are useful for forest managers, as they allow better planning of fire-fighting operations, and may even reduce operating costs due to better logistics efficiency.

Keywords: Fire; Fire probability; Fire scar; Machine learning

1. INTRODUÇÃO

Desconsiderando os ecossistemas que necessitam do fogo para sua estabilidade, os incêndios são um grande problema para as florestas em todo o mundo (HANEWINKEL; HUMMEL; ALBRECHT, 2010), queimando anualmente grandes extensões florestais (SOARES; BATISTA, 2007). O fogo é considerado uma Variável Climática Essencial (Essential Climate Variables - ECVs) pelo programa Sistema Climático de Observação Global (*Global Climate Observing System* – GCOS, 2016) e, de acordo com a Classificação e Codificação Brasileira de Desastres (COBRADE), é um desastre natural climatológico (BRASIL, 2012). Predição, prevenção e medidas de manejo referentes aos incêndios florestais tem se tornado cada vez mais importantes (SAKR; ELHAJJ; MITRI, 2011).

Desde as primeiras publicações sobre plantio de eucalipto (*Eucalyptus* spp.) no Brasil, os incêndios florestais já eram tratados como uma ameaça (ANDRADE, 1911), mas somente a partir de 1963, quando ocorreu no Paraná um dos maiores incêndios já registrados no mundo, onde aproximadamente 2 milhões de hectares foram atingidos e 110 pessoas morreram (WASILEWSKI, 1965), esse assunto passou a ser mais estudado no país. Após essa catástrofe uma empresa florestal iniciou o registro sistemático das ocorrências de fogo, tornando o estado do Paraná pioneiro no país nos estudos de incêndio (SOARES, 1992; TETTO et al, 2012). Atualmente o registro das ocorrências de incêndio faz parte da rotina das grandes empresas de base florestal do Brasil, que se mostram preparadas para monitorar e combater sinistros dessa natureza (RODRIGUEZ et al., 2011).

A probabilidade da ocorrência de um incêndio e sua propagação, é em função da probabilidade de haver uma fonte de fogo e condições favoráveis para se propagar (SOARES; BATISTA, 2007). Segundo os autores, uma análise criteriosa da fonte de ignição e condições favoráveis para propagação do fogo, possibilita avaliar o potencial de risco de incêndios de uma região e determinar seu zoneamento, utilizando mapas de risco. A partir desses mapas, se pode elaborar medidas preventivas e planificar situações de emergência, com o objetivo de promover a defesa permanente contra os desastres naturais (INTERNATIONAL STRATEGY FOR DISASTER REDUCTION - ISDR, 2004).

Mapas de risco são gerados por meio do uso de Sistema de Informações Geográficas (SIG) e são criados pela sobreposição de mapas temáticos de riscos

específicos a partir da integração dos fatores que podem causar a ignição e que influenciam na propagação dos incêndios (SALAS; CHUVIECO, 1994; FERRAZ; VETTORAZZI, 1998; CHUVIECO et al., 2010; YOU et al., 2017). A elaboração do mapa final de zoneamento de risco envolve a criação de um algoritmo ou modelo matemático que expresse numericamente a interação das variáveis analisadas, na influência sobre o início e propagação do fogo (SOARES; BATISTA, 2007). Esses modelos são desenvolvidos com base em informações bibliográficas e levantamentos de campo, no qual os pesos associados à importância de cada mapa temático, muitas vezes, é definido a partir de um palpite do avaliador.

O desenvolvimento de bons modelos preditivos que expressem o risco da ocorrência de fogo ainda é uma tarefa difícil, pois incêndios florestais são tipicamente processos complexos e não-lineares, motivados por diversos fatores (NGUYEN et al., 2018), como a influência antropogênica, uma das principais causas dos incêndios florestais (MANN et al., 2016; SOARES; BATISTA; SANTOS, 2006) e também o clima, a topografia e a vegetação (SOARES; BATISTA, 2007; BIRCH et al., 2015). Exemplos de trabalhos que desenvolveram modelos de probabilidade de incêndio integrando as principais variáveis responsáveis pela ignição e propagação do incêndio, foram desenvolvidos na Espanha (GONZÁLEZ et al., 2006) e Portugal (GARCIA-GONZALO et al., 2012; MARQUES et al., 2012; BOTEQUIM et al., 2013), e obtiveram boas acurácias. Esses trabalhos apresentam diretamente um valor da probabilidade da ocorrência de incêndio para determinado local.

Em ciências ambientais, observa-se crescente aumento no uso de “Inteligência Artificial” (IA) em estudos de modelagem (CHEN et al., 2008), fato também verificado na área florestal (LAGERQUIST et al., 2017) pelo uso de técnicas de Aprendizado de Máquinas (em inglês, *Machine Learning* - ML). Estudos recentes mostraram que a performance de algoritmos de ML foram eficientes para predição da probabilidade da ocorrência de eventos climáticos indesejáveis às florestas (HART et al., 2019; KIM et al., 2019; LEUENBERGER et al., 2018; NGUYEN et al., 2018; TIEN BUI; LE; HOANG, 2018; YU et al., 2017). Esses algoritmos são capazes de obter bons desempenhos ao relacionar fenômenos não-lineares e complexos, assim como as variáveis que influem as causas de ignição e propagação dos incêndios florestais.

O uso de sensores remotos ativos, como o escaneamento a laser aeroembarcado (ALS) proporcionam acurada medição estrutural do dossel florestal em três dimensões (3D) (COOPS et al., 2007) e, com isso, permite boa estimativa das

características de combustível para grandes áreas (ANDERSEN; MCGAUGHEY; REUTEBUCH, 2005), podendo ser úteis na caracterização da vegetação, que contribui para a propagação do fogo, pela composição de espécies e sua estrutura, variáveis controláveis no manejo florestal (GONZALES et al., 2006).

Além da importância de se conhecer a probabilidade da ocorrência de incêndios, que podem vir a causar danos, a detecção de áreas florestais já danificadas é um processo importante nas pesquisas relacionadas ao ciclo de carbono e mudanças climáticas (CHEN et al., 2014; SCHROEDER et al., 2012), na descrição de impactos do incêndio ao ecossistema terrestre e à distribuição da vegetação (BOND, 2005). Afora, indicar áreas que exigem atenção especial ao manejo florestal, quando se trata de uma floresta para produção de madeira.

Considerando grandes dimensões espaciais e frequente dificuldade de acesso aos locais onde ocorreram incêndios, técnicas de sensoriamento remoto apresentam-se como um importante meio para juntar informações sobre áreas queimadas, em intervalos de tempo específicos e de maneira consistente (ROGAN; YOOL, 2001). Nos últimos anos foi observado progresso significativo no mapeamento de áreas queimadas, com o aprimoramento de observações via satélite (GCOS, 2016).

Embora a identificação de cicatrizes de incêndios possa parecer óbvia para um interpretador de imagens, o mapeamento desses locais pode ser difícil, pois técnicas automáticas de classificação não possuem a capacidade de reconhecimento de padrões assim como o cérebro humano (HUDAK; BROCKETT, 2004). O uso de algoritmos de aprendizado de máquinas podem suprir esse problema, pois possuem habilidade de generalização para tratar fenômenos complexos, não-lineares e de grande dimensão (KANEVSKI; POZDNOUKHOV; TIMONIN, 2010).

Nesse contexto, o objetivo deste estudo foi desenvolver metodologias para gerar dois modelos preditivos relacionados ao manejo de incêndios florestais em plantios de eucalipto:

- 1) Probabilidade de incêndio em nível de talhão;
- 2) Mapas de cicatrizes que identificam áreas com cobertura florestal danificada por incêndio.

Como objetivo específico, esse estudo procurou identificar as variáveis mais importantes utilizadas em cada um dos modelos.

Por fim, buscou-se responder as seguintes perguntas:

- i) A adição de dados ALS às variáveis biométricas, climáticas e sociais aprimora os modelos de cálculo de probabilidade de incêndio, gerando mapas de maior acurácia?
- ii) A adição de dados ALS aos dados multiespectrais aprimora a eficácia do mapeamento de cicatrizes de incêndio?

Esse trabalho foi executado utilizando dados de uma empresa de papel e celulose situada no sul da Bahia, onde há muitos casos de ocorrências de incêndio, principalmente associadas a interferência antrópica. As metodologias aqui propostas são úteis para produtores florestais, pois auxiliam na preservação do estoque de madeira. O conhecimento da probabilidade de incêndio proporciona ao gestor florestal melhor entendimento sobre o risco iminente à cada talhão, possibilitando a adoção de medidas estratégicas de prevenção e combate às chamas. Já o mapeamento de áreas danificadas, permite identificar locais onde houve perda na produção de madeira, ou que exige maiores cuidados no manejo.

A metodologia e resultados desta tese estão estruturados da seguinte maneira:

- a) Desenvolvimento de um modelo de cálculo da probabilidade de ocorrência de incêndio em nível de talhão para plantios de eucalipto, onde o algoritmo Random Forest foi comparado com a Regressão Logística;
- b) Desenvolvimento de um modelo para classificação de imagens provenientes do satélite RapidEye e da sua fusão com dados ALS, que determine a probabilidade de um local em pertencer a cicatriz de incêndio. Para isso, o RF foi comparado ao algoritmo Support Vector Machine utilizando a função kernel polinomial.

2. REVISÃO BIBLIOGRÁFICA

2.1. Incêndio Florestal

“Fogo é o termo aplicado ao fenômeno físico resultante da rápida combinação entre o oxigênio e uma substância qualquer (madeira, por exemplo), com produção de calor, luz e, geralmente, chamas. Fogo, ou mais precisamente combustão é, portanto, uma reação química de combustão” Soares e Batista (2007).

Para que ocorra um incêndio, é preciso que haja combustível para queimar, oxigênio para manter as chamas e calor para iniciar ou continuar o processo de queima. A relação entre os três elementos é conhecida como “triângulo do fogo”, e a ausência (ou diminuição significativa) de qualquer um desses componentes, inviabiliza o processo de combustão (SOARES; BATISTA, 2007).

Incêndios são influenciados pelo clima, pois estão associados a eventos climáticos extremos, como secas, no entanto, também afetam o clima devido ao impacto que exercem sobre o fluxo de carbono e emissões de gases do efeito estufa (VAN DER WERF et al., 2017). Podem causar grandes impactos, pois seus efeitos estão relacionados à modificação do ambiente e impacto nos âmbitos econômico, social e ambiental. Incêndios florestais são aqueles relacionados com queima de vegetação.

2.1.1. Causas dos Incêndios Florestais

É importante a obtenção de conhecimento sobre os agentes causadores do fogo (*Food and Agriculture Organization of the United Nations - FAO, 1953*). Segundo a organização, a classificação dos números de focos de incêndio e área queimada atribuída a uma específica causa, é a forma mais confiável de se obter estatísticas comparáveis entre si. O Serviço Florestal Americano definiu, então, oito grupos de causas: raios, incendiários, queimas para limpeza, fumantes, fogos de recreação, estradas de ferro, operações florestais e diversos; e essa metodologia tem sido bastante utilizada e apresenta bons resultados há muitos anos (FAO, 1953; SOARES; BATISTA, 2007).

Analisando um banco de dados de mais de 23 mil focos de incêndio no Brasil, para florestas nativas e plantadas, entre 1983 e 2002, Soares, Batista e Santos (2006) identificaram que nos primeiros anos da análise, as principais causas eram referentes às “queimas para limpeza” e, posteriormente, passaram a ser do tipo “incendiário”. Os resultados deste trabalho caracterizaram, portanto, a influência antropogênica como sendo a principal causa de incêndio, corroborando resultados verificados em outros países. No estado da Califórnia, Estados Unidos, 95% das causas são relacionadas ao homem (MANN et al., 2016). Na Espanha e países mediterrâneos, 90% dos fogos florestais são causados por pessoas, assim como na África, países do Caribe e também da América do Sul, entre outros (FAO, 2007).

A mudança dos agentes causadores de incêndio verificado por Soares; Batista; Santos, (2006), pode ser explicada por uma maior conscientização no uso da queima como limpeza, contrastando com conflitos agrários e problemas sociais, que intensificaram o índice de incêndios criminosos.

A ocorrência do fogo é influenciada não só pelos fatores de ignição, mas também pelos fatores que agravam as condições para propagação dos focos de incêndios.

2.1.2. Propagação de Incêndios Florestais

A propagação dos incêndios é um fenômeno complexo e os mecanismos que o definem, principalmente transferência ou fluxo de calor, são bastantes distintos conforme a intensidade do fogo (SOARES; BATISTA, 2007). Segundo os autores, os fatores que influenciam na propagação dos incêndios florestais são os seguintes:

- Material combustível (quantidade, tipo e arranjo);
- Umidade do material combustível;
- Condições climáticas;
- Topografia;
- Tipo de floresta.

O conhecimento das características dos combustíveis é essencial para o manejo da terra e do fogo, pois pode ser utilizado para computar ocorrências de incêndio, risco comportamento e efeitos do fogo (ARROYO; PASCUAL; MANZANERA, 2008). Material combustível representa a matéria orgânica disponível

para ignição e combustão e representam o componente do triângulo do fogo em que o homem é capaz de controlar (KOUTSIAS; KARTERIS, 2003; ROTHERMEL, 1972; SALAS; CHUVIECO, 1994). Os combustíveis florestais podem ser aéreos (galhos, copas e troncos das árvores acima de 1,8m) e superficiais (vegetação rasteira, troncos e galhos) (SOARES; BATISTA, 2007). Combustíveis superficiais são os que mais afetam a intensidade e poder de propagação dos incêndios (MUTLU; POPESCU; ZHAO, 2008).

A umidade do combustível influencia diretamente na velocidade da propagação e grau de severidade (PARKS et al., 2014), por isso é utilizado como *input* em modelos de comportamento do fogo (FINNEY, 2006). Para haver combustão, primeiramente deve haver calor suficiente para secagem do material, por isso, quanto mais úmido for o combustível, menores as chances de ignição e propagação.

As condições climáticas que mais influenciam na dispersão das chamas, são: temperatura, umidade relativa do ar, vento e precipitação. Maiores temperaturas, menores umidades, menor precipitação e maior velocidade de vento caracterizam maiores chances de propagação de incêndio (SOARES; BATISTA, 2007). Até mesmo em florestas como a Amazônia, que são caracterizadas por combustíveis úmidos, em períodos de seca, onde há acúmulo de material combustível, aumento da temperatura e baixa umidade relativa do ar, foi observado aumento da morte de plantas devido a incêndios (BRANDO et al., 2014).

A topografia pode influenciar o fogo de diferentes maneiras, como por exemplo na elevação, inclinação e exposição. Altas elevações apresentam ar rarefeito e temperatura mais baixa, por isso, locais situados em menores altitudes possuem maior risco de incêndio (SOARES; BATISTA, 2007). Além do fator climático, florestas situadas em maiores elevações são menos acessadas pela atividade humana e, conseqüentemente, possuem menor risco de incêndio (KIM et al., 2019). A exposição do terreno está relacionada à quantidade de calor recebida pela superfície, que provocam reações em cadeia e originam condições distintas ao risco de fogo (SOARES; BATISTA, 2007). A inclinação também afeta a severidade dos incêndios, que se propagam de forma mais intensa em áreas mais íngremes e causam maiores danos às árvores nestes locais (GONZÁLEZ et al., 2007; MARQUES et al., 2011). Áreas inclinadas também dificultam o acesso de equipes de combate à incêndio.

O tipo de vegetação influencia o potencial de propagação de incêndio, pois está relacionada com o microclima local (SOARES; BATISTA, 2007). Os autores

explicam que florestas densas e fechadas são menos susceptíveis à condução do fogo, pois interceptam a radiação solar, reduzindo a temperatura do ar, acumulando menos material combustível em seu interior. Além disso, mais biomassa acarreta em maior transpiração e maior umidade relativa do ar.

A integração dos diferentes fatores que exercem influência na propagação de incêndios varia conforme a região. Por exemplo, Birch et al., (2015), identificou que no noroeste dos Estados Unidos, vegetação e topografia exercem maior influência na severidade dos incêndios do que as condições climáticas, o que contrasta com o verificado por Brando et al. (2014), onde as condições climáticas foram determinantes para severidade de incêndios na Amazônia brasileira.

2.1.3. Risco e Perigo de Incêndios Florestais

Von Gadow, (2000) apresenta uma definição para o termo “risco” sugerida pelas Nações Unidas: “Risco é definido como a perda esperada devido a ocorrência de algum evento indesejado, sendo que a expectativa de perda por este evento é em função da probabilidade de sua ocorrência”. Segundo o autor, o manejo do risco inclui estratégias e ações para sua redução.

Para Brown e Davis (1973, apud SOARES; BATISTA, 2007) risco de incêndio (*fire risk*) é a probabilidade de um incêndio iniciar em função da presença e/ou atividade de agentes causadores e, para Philips e Nickey (1978), é a probabilidade de ocorrência de fogo em uma determinada área, durante um período específico de tempo. Perigo de incêndios (*fire hazards*) é composto pelo risco de incêndios, agregado às condições de combustível (BROWN; DAVIS, 1973). Chuvieco et al., (2010), apresenta a definição dada pela FAO, em que risco de incêndio é a probabilidade de um fogo se iniciar, em virtude da presença e atividade de agentes causadores e, perigo de incêndio, é a integração dos fatores fixos e variáveis do ambiente do fogo, que determina a facilidade da ignição, taxa de espalhamento, dificuldade de controle e impacto do fogo, geralmente expresso como um índice.

Por estas definições, entende-se que tendo o conhecimento da probabilidade da ocorrência de incêndio de uma determinada floresta, e havendo possibilidade de caracterizar as condições dos combustíveis envolvidos e seu entorno, é possível avaliar o potencial de risco da região e como se deve propagar o incêndio, possibilitando o mapeamento de risco.

2.1.4. Mapeamento de risco de incêndios florestais com SIG

O conhecimento dos locais de ocorrência de incêndios é um conceito prático e importante para o planejamento de prevenção e estratégias de combate ao fogo (FAO, 1953). A organização enfatiza que ao mapear os focos históricos de incêndio, são geradas importantes informações, e mesmo que não se saiba a causa exata, experientes gestores florestais podem identificar relações lógicas para essas concentrações e identificar padrões para as ocorrências, como por exemplo proximidade de estradas, centros urbanos, etc.

Segundo Soares e Batista (2007), a contínua coleta de dados estatísticos pode definir as regiões de maior ocorrência, fornecendo informações importantes para planos de prevenção e, assim, permitir o zoneamento das ocorrências de incêndio e a confecção de mapas de risco potencial de uma determinada região, utilizando Sistemas de Informação Geográfica (SIG).

Sistemas de Informação Geográfica são projetados para a captura, armazenamento, recuperação, análise e exibição de dados espaciais. Podem ser utilizados por meio de softwares específicos, onde usuários podem criar consultas interativas, analisar informações espaciais, editar dados em mapas e apresentar os resultados de todas essas operações (CLARKE, 1986). Para Tien Bui; Le; Hoang (2018), o SIG fornece várias ferramentas para incorporação de informações espaciais de vários fatores que influenciam os incêndios florestais e, por isso, inúmeros trabalhos sobre manejo do fogo utilizaram SIG em suas análises.

Alguns exemplos desses trabalhos, são: Salas e Chuvieco, (1994), onde foram combinados mapas de topografia, vegetação e influência antrópica para gerar mapas de risco de ignição e de comportamento do fogo que, quando combinados, integraram um mapa de perigo de incêndio para uma área de estudo na Espanha. Ferraz e Vetorazzi (1998) desenvolveram o mapeamento de risco para uma área em Piracicaba/SP pela sobreposição de mapas de uso da terra, declividade, exposição, vizinhança e vias de acesso. Jaiswal et al., (2002) aplicaram metodologia semelhante na Índia e validaram seus mapas utilizando imagens de satélite. Eugenio et al., (2016) aplicou a técnica para o estado do Espírito Santo. You et al. (2017) apresentou um conceito de mapeamento de risco que incluiu dados de inventário, sócio-econômicos e climáticos, entre as 12 variáveis utilizadas, em uma província do sudoeste da China. Chuvieco et al. (2010) desenvolveu um trabalho com princípio similar aos demais,

porém propôs uma estrutura integrada para avaliação de risco, que incluiu ações antropogênicas, umidade de combustível, taxas de propagação potencial de incêndios por modelos simuladores do fogo, entre outras variáveis, para o mapeamento do risco de 4 áreas distintas na Espanha.

O mapa final de risco envolve a sobreposição de mapas temáticos específicos, elaborados para cada variável de interesse, que possa representar risco para a área de estudo (exemplo: distância para estradas, centros urbanos, uso do solo, variáveis climáticas, biométricas, etc). Essa sobreposição é feita de forma ponderada, em função de um peso dado à cada mapa, normalmente definido por consulta à literatura (SOARES; BATISTA, 2007; YOU et al., 2017).

O peso atribuído ao mapa de risco, mesmo que apoiado por uma boa referência bibliográfica, é geralmente proveniente de um palpite de um profissional, o que significa que duas pessoas podem atribuir pesos diferentes ao mapa, gerando produtos distintos para uma mesma região. O uso de técnicas de aprendizado de máquinas pode transferir esta tarefa a um poderoso algoritmo, capaz de extrair informações a partir de relações complexas entre todas variáveis elencadas que possam contribuir ao perigo de incêndio, produzindo mapas que já incluem um valor da probabilidade da ocorrência de incêndios para determinado local. Mapas como esse podem ser utilizados em substituição ou complemento aos mapas de zoneamento de risco.

A cada ano, devido aos avanços tecnológicos e maior disponibilidade de recursos para utilização de dados de Sensoriamento Remoto (SR) e SIG, vai se aperfeiçoando a metodologia na elaboração dos mapas de risco (YOU et al., 2017). Ao integrar técnicas de aprendizado de máquinas, ao SR e SIG, os mapas de risco podem apresentar cada vez mais acurácia.

2.2. Sensoriamento Remoto no mapeamento de incêndios florestais

Sensoriamento Remoto é definido como uma técnica que realiza “a aquisição de informações sobre um objeto sem que se entre em contato físico com ele” (ELACHI; VAN ZYL, 2006). O termo “sensoriamento remoto” é comumente utilizado quando se trata de informações baseadas na captação dos sinais que cobrem todo o espectro das ondas eletromagnéticas (ELACHI; van ZIL, 2006; LIU, 2006).

A energia eletromagnética é o meio pelo qual a informação é transmitida de um objeto ao sensor (ELACHI; van ZYL, 2006) e é propagada pela radiação com a velocidade da luz (ELACHI; van ZYL, 2006; LIU, 2006). A interação das ondas eletromagnéticas com a atmosfera e superfícies naturais é fortemente dependente da frequência das ondas. Radiação solar é a principal fonte de energia incidente sobre o planeta terra e, conseqüentemente, tem papel fundamental na determinação da estrutura e composição da atmosfera (MEASURES, 1992; NOVO, 2006). Cerca de 20% da energia solar cai na banda espectral menor que 470 nm, enquanto 44% encontram-se sob bandas visíveis, na faixa entre 400 e 760 nm (MEASURES, 1992).

A intensidade da radiação solar recebida pela superfície depende do ângulo zenital solar, mas até chegar à superfície terrestre, passa por diversos tipos de interferências atmosféricas, que incluem espalhamento, absorções e reflectâncias múltiplas (LIU, 2016). Segundo o autor, a reflectância ocorre quando um feixe de luz que atinge um objeto não transparente e é redirecionado e influenciado pelas irregularidades da superfície em relação ao comprimento da onda da radiação considerada.

A atmosfera transmite quase totalmente a radiação em determinadas faixas de ondas eletromagnéticas, incluindo ultravioleta e visível (0,30 a 0,75 μm), infravermelha próxima (0,77 a 0,91 μm), infravermelha termal (8 a 9,2 μm e 10,2 a 12,4 μm) e micro-ondas (7,5 a 11,5 mm e > 20 mm), que são chamadas de janelas atmosféricas. Fora dessas bandas, a energia da radiação eletromagnética é atenuada severamente. Assim, os sensores satelitais são desenhados nas faixas de janelas atmosféricas que tentem evitar ao máximo o efeito das atenuações atmosféricas (LIU, 2006).

A maioria da radiação solar com comprimento de onda maior que 300 nm (0,3 μm) pode penetrar na atmosfera e chegar à superfície, portanto a maioria dos sensores de satélite é desenvolvido para registrar os sinais de energia eletromagnética refletida ou emitida no comprimento da onda localizado nas faixas do visível, infravermelho e micro-ondas (LIU, 2006).

2.2.1. Sensores Satelitais

Embora técnicas de sensoriamento remoto tenham sido utilizadas já por muitos anos, o início da era espacial estimulou o desenvolvimento da tecnologia de

sensores remotos, tanto espaciais quanto para sistemas aéreos (LANDGREBE, 2003). Conforme descrito pelo autor, a era espacial praticamente começou com o lançamento do primeiro satélite, o *Sputinik*, pela União Soviética em 1957. E esse acontecimento estimulou os Estados Unidos a criar, em 1958, sua agência espacial, a NASA (*National Aeronautics and Space Administration*), que em 1960 lançou com sucesso o satélite denominado *Television and Infrared Observational Satellite* (TIROS 1), o primeiro satélite designado para fins da observação do planeta, e que mostrou o globo terrestre e os sistemas de circulação atmosférica (LANDGREB, 2003; LIU, 2006). O TIROS 1 foi um satélite com fins climáticos e contribuiu para o aprimoramento das previsões meteorológicas e monitoramento de ameaças, como por exemplo, furacões.

A partir dos anos 60, uma série de satélites com diferentes finalidades foram desenvolvidos e testados. São destacados aqui os satélites mais utilizados em estudos ligados a incêndios florestais.

Lançado pela primeira vez em 1972, os satélites da série *Land Satellite* (Landsat), foram desenvolvidos com o propósito de monitorar os recursos naturais do planeta (LIU, 2006; NOVO, 2010). Atualmente estão em órbita o Landsat 7 ETM+ e Landsat 8 OLI. É previsto para 2020 o lançamento do Landsat 9, e o objetivo é que não haja lapso de dados disponibilizados pelo programa, que já fornece imagens há mais de 40 anos (NASA, 2019). A continuação na coleta de dados permite a elaboração de importantes estudos multitemporais, como por exemplo no monitoramento de distúrbios, estimativa e recuperação de biomassa.

Os satélites da série *National Oceanic and Atmospheric Administration* - NOAA também fazem parte de agências espaciais norte-americanas e surgiram da evolução do programa TIROS 1. Foram lançados pela primeira vez em 1970 pelo NOAA-1. São equipados com sensores que incluem radiômetros de alta resolução, com cinco canais espectrais: visível, infravermelho próximo, infravermelho médio e dois infravermelhos termais; denominado *Advanced Very High Resolution Radiometers* (AVHRR) (NOAA, 2019).

Em 1999 a NASA lançou o satélite TERRA, que inclui cinco instrumentos, dentre eles o MODIS (Moderate Resolution Spectroradiometer), capaz de promover observação de toda superfície terrestre à cada dois dias, obtendo dados para geração de produtos específicos, destinados à diversos fins, como no mapeamento da

vegetação, do oceano e detecção de focos de calor utilizados no monitoramento de incêndios (LIU, 2016; NASA, 2019).

No ano de 2008 foi lançado o sistema de cinco satélites multiespectrais denominado RapidEye, com capacidade de cobertura de extensas áreas, revisitadas em intervalos frequentes, produzindo imagens de alta resolução (RAPIDEYE, 2016). Este sensor trouxe como inovação a presença da banda “*Red Edge*” (traduzido como “borda do vermelho”) que ficou amplamente reconhecida pela eficiência na caracterização da vegetação (WEICHELDT et al., 2016).

Para avaliar a capacidade dos sensores a bordo dos diferentes tipos de satélite são aplicados dois critérios: resolução espectral e resolução espacial (LIU, 2006).

A resolução espectral representa a largura das faixas espectrais e da capacidade do sensor em distinguir entre dois níveis de intensidade do sinal de retorno (NOVO, 2010). A banda espectral mais estreita significa a resolução mais alta (LIU, 2006). A resolução espacial é definida como a capacidade que um sensor apresenta na identificação de um alvo-padrão na superfície terrestre (LIU, 2006), em outras palavras, representa a menor feição detectável pelo sensor (NOVO, 2010). De maneira geral, as imagens de sensores remotos são formadas por matrizes de elementos de imagem (*picture elements*), conhecido como “pixel”, que são geralmente um quadrado representando certa área do terreno (NOVO, 2010). Segundo o autor, a resolução espacial e o tamanho do pixel são termos equivalentes quando uma porção da imagem é visualizada em resolução plena, mas podem diferir quando há transformação da imagem.

Os sensores MODIS e da série NOAA, possuem resolução espacial mais grosseira, em média de 400 metros, podendo atingir até 6 km (Instituto Nacional de Pesquisas Espaciais, INPE, 2019a). São bastante utilizados para disponibilizar produtos relacionados a incêndios florestais pela identificação de focos de calor e são mais recomendados para utilização em estudos para grandes escalas, com por exemplo escala global, continental ou em nível de país. Para escalas menores, com objetivo de se obter maior acurácia, satélites de alta a média resolução são mais recomendados, como por exemplo os sensores RapidEye, que possui resolução espacial de 5 metros (RAPIDEYE, 2006), e os sensores landsat, com 30 metros.

2.2.2. Uso de Sensores Remotos no Monitoramento de Incêndios Florestais

O uso de sensoriamento remoto para mapeamento de incêndios florestais se iniciou nos anos 60 com uso de sensores infravermelhos aeroembarcados. A partir dos anos 70, após o desenvolvimento do programa Landsat, uma série de projetos foram conduzidos para testar a viabilidade do uso de imagens de satélite no mapeamento e estudos de severidade dos incêndios (CHUVIECO; CONGALTON, 1988; PATTERSON; YOOL, 1998; PEREIRA; SETZER, 1993; CHEN et al., 2014; HUDAK; BROCKETT, 2004; LI, 2003; MILLER; THODE, 2007).

Agências espaciais vêm contribuindo com o progresso no mapeamento dessas ocorrências, via satélite, (CHUVIECO et al., 2018) por meio de programas de mapeamento global de áreas queimadas (AQ). A NASA, por exemplo, lançou os produtos MCD45 (ROY et al., 2008) e C6 MCD64A1 (GIGLIO et al., 2018), obtidos do sensor MODIS. A Agência Espacial Européia lançou o programa “Fire_cci”, que mapeia também longas séries temporais de mapas globais de área queimada, com imagens Envisat-MERIS, com maior resolução espacial que o MODIS (CHUVIECO et al., 2016). No Brasil, o Instituto Nacional de Pesquisas Espaciais – INPE, mantém uma plataforma online denominada “INPE - Programa de Queimadas” que disponibiliza dados de focos de calor, detectado por diferentes satélites, a partir de 1998, para todo o país, permitindo ao usuário busca por sensores, regiões, biomas e municípios (INPE, 2019b).

Para menores escalas espaciais, uma série de sensores multiespectrais, de média a alta resolução (pixels de 1 a 30 metros) são utilizados para mapear cicatrizes de incêndio. Diferentes composições de bandas, índices de vegetação e tipos de sensores são utilizados (FORNACCA; REN; XIAO, 2018; HISLOP et al., 2018; HUDAK; BROCKETT, 2004; LU; HE; TONG, 2016; MALKINSON et al., 2016; PEREIRA; SETZER, 1993).

A acurácia de mapas de áreas queimadas depende não só do número de bandas, mas na sua proporção de informações significantes para o mapeamento das cicatrizes de incêndio (CHEN et al., 2014). Diferentes técnicas e índices já foram utilizados, mas os mais frequentes são o NDVI e o NBR - Índice Normalizado de Área Queimada (Normalized Burned Index), que podem ser utilizados individualmente ou em séries multi-temporais pelo dNDVI e dNBR, diferenças entre índices antes e depois do (ESCUIN; NAVARRO; FERNÁNDEZ, 2008; FORNACCA; REN; XIAO, 2018;

HISLOP et al., 2018; HOLDEN et al., 2005; KEY; BENSON, 2006; LEWIS et al., 2009; LU; HE; TONG, 2016).

O NDVI é um índice bastante utilizado para identificação do vigor vegetativo das plantas e na detecção de distúrbios na paisagem (BALDI et al., 2008; PICKELL et al., 2016). O NBR utiliza em seu cálculo a razão entre a banda do infravermelho e a banda termal SWIR 2 (2,08 a 2,35 μm) (KEY; BENSON, 2006). Em trabalho comparando os dois índices, Hudak et al. (2007) verificou que o NDVI não se apresentou significativamente inferior ao NBR no potencial para detecção de cicatrizes de incêndio, podendo ser utilizado quando o sensor utilizado para o estudo não possuir a banda termal, como no caso do RapidEye, por exemplo

O uso de técnicas passivas de sensoriamento remoto, como por exemplo imagens de satélite, permitem apenas a avaliação da floresta em duas dimensões (2D), mas são limitadas para representar sua estrutura (KEANE; BURGAN; VAN WAGTENDONK, 2001; LIM et al., 2003). Sensores passivos são aqueles que medem a energia solar refletida por um alvo na superfície terrestre, na faixa de 0,4 a 3 μm ou ainda aqueles que medem a energia emitida pela superfície na faixa de 10 a 12 μm . Já os sensores ativos, produzem fonte artificial em certa banda e medem suas intensidades emitidas e refletidas, simultaneamente (LIU, 2006). Um exemplo de sensor ativo é o *Light Detection and Ranging* (LiDAR), emissor de raio laser, que podem medir diretamente a distribuição dos componentes vegetativos da copa e também do sub-bosque em três dimensões e em alta resolução (COOPS et al., 2007; LEFSKY et al., 2002).

2.2.3. Sensores LiDAR e aplicações florestais

LiDAR se apresenta como uma solução rápida e precisa para mapeamento planialtimétrico para extensas áreas e em alta resolução. Utilizam um pulso do laser para medir a distância entre o sensor e um objeto (FLOOD; GUTELIUS, 1997) e podem ser acoplados a plataformas aéreas (ALS – *Airborne Laser Scanning*), terrestre (TLS – *Terrestrial Laser Scanning*), e orbitais.

Os equipamentos que compõem o sistema ALS são: unidade de emissão e recebimento a laser acoplado à aeronave (sensor LiDAR); unidades de posicionamento global (GNSS) no avião e no solo; sensor inercial acoplado à unidade laser para mensuração de movimentos do avião (IMU) e um computador para controlar

e armazenar os dados levantados (REUTEBUCH; ANDERSEN; MCGAUGHEY, 2005).

Os escaneamentos com LiDAR geram uma nuvem de pontos, que podem ser disponibilizadas em diferentes formatos, como por exemplo '.txt', '.xyz', '.las' e '.laz'. O formato mais comum é o LAS, o qual é de autoridade da Sociedade Americana de Fotogrametria e Sensoriamento Remoto (*The American Society for Photogrammetry & Remote Sensing* - ASPRS) (ASPRS, 2013), e em sua versão compactada, LAZ, que possui cerca de 20% do tamanho original (ISENBURG, 2012).

Os arquivos das nuvens de pontos são compostos, além da localização geográfica e altitude de cada ponto (XYZ), por outras informações, como por exemplo, número do retorno, quantidade de retornos, intensidade, entre outras (ASPRS, 2013).

Essa tecnologia apresentou-se como uma ferramenta de grande acurácia para medições topográficas, altura e cobertura da vegetação, assim como atributos mais complexos referentes à estrutura de dossel (LEFSKY et al., 2002). Os produtos topográficos mais comuns são o Modelo Digital de Terreno (MDT), Modelo Digital de Superfície (MDS) e Modelos de Altura de Dossel (CHM).

Sistemas LiDAR utilizados em aplicações florestais, podem ser classificadas em "retornos discretos" e sistemas "*full waveform*", que se diferenciam pela maneira em que amostram verticalmente e horizontalmente a estrutura 3D do dossel florestal (LIM et al., 2003). Segundo os autores, o sistema de retornos discretos possibilita múltiplos retornos de um único feixe laser. Esse é o sistema mais comum utilizado na área florestal. Em um caso típico (mas não padrão), os primeiros retornos atingem a copa das árvores e os demais são interceptados pelos estratos inferiores da floresta, até o último pulso chegar ao solo (Figura 1).

O processamento das nuvens de pontos LiDAR geram estatísticas denominadas métricas (NELSON; KRABILL; TONELLI, 1988). Podem ser utilizadas para diversas finalidades e classificadas como métricas de contagem; de elevação ou altura; de intensidade; topográficas ou ainda para representar algum fator ecológico ou de interesse para a pesquisa.

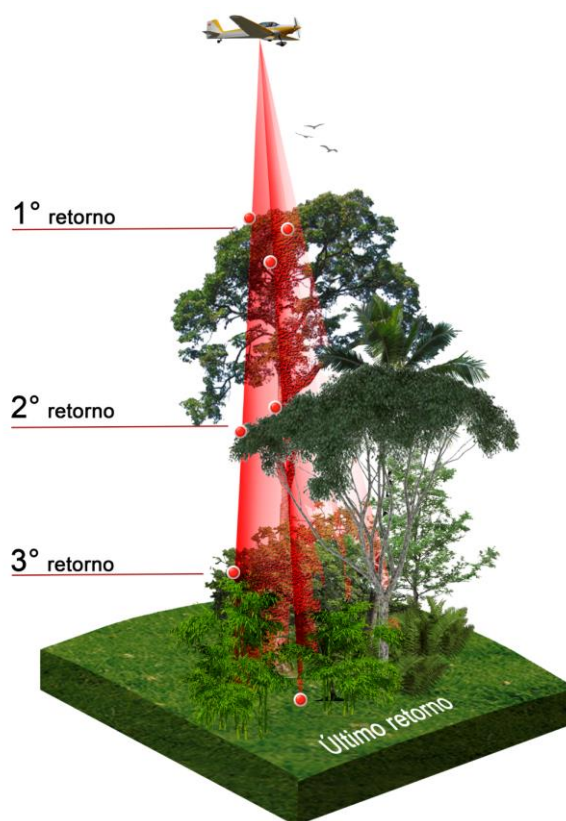


Figura 1. Múltiplos retornos de um feixe laser em um escaneamento a laser aerotransportado (ALS). Ilustração de Papa, 2018.

Métricas provenientes de ALS são capazes de compor modelos acurados para predição de biomassa florestal para diferentes espécies (NELSON; KRABILL; TONELLI, 1988; GUERRA-HERNANDEZ et al., 2016) e em diferentes escalas e estado de distúrbio (KNAPP; FISCHER; HUTH, 2018). O mesmo foi verificado para estimativa de volume de plantios clonais de eucalipto no Brasil (GÖRGENS et al., 2015; ZONETE; RODRIGUEZ; PACKALÉN, 2010), onde a modelagem final utilizando métricas ALS apresentou ótimas acurácias.

O conhecimento da estrutura do dossel pode explicar a variação no crescimento da biomassa sobre paisagens tropicais e melhorar a compreensão da função do ecossistema (STARK et al., 2012). Uma das variáveis mais usadas para caracterizar a estrutura vertical e horizontal do dossel é o índice de área foliar (IAF) (COOPS et al., 2007). Estimativas derivadas do IAF, fornecem acuradas estimativas para estrutura e função do dossel, incluindo a modelagem de ambientes foliares e estrutura demográfica florestal, importantes na abordagem de problemas em conservação e manejo florestal (ALMEIDA et al., 2019).

LiDAR demonstra-se ainda como uma importante ferramenta para linhar a informação espacial em nível de parcela para caracterização de grandes áreas com

objetivo de se detectar distúrbios por fogo em florestas heterogêneas (HOFFMAN et al., 2018).

2.2.4. Fusão de Sensores Remotos em aplicações florestais

Dadas as características específicas de sensores remotos passivos, como imagens satelitais, que oferecem alternativa no mapeamento no plano horizontal, em duas dimensões e sensores ativos como o LiDAR, que possibilitam obtenção de informação estrutural no plano vertical, em três dimensões, a fusão entre sensores pode apresentar benefícios e bons resultados na aplicação florestal.

Hudak et al., (2002) foram pioneiros ao integrar imagens de satélite e de sensores LiDAR para buscar soluções de problemas florestais. Os autores combinaram imagens Landsat ETM + para estimar altura de dossel em locais não sobrevoados com o sensor a laser, baseado em estatísticas e geoestatísticas entre dados provenientes do ALS e do satélite. Como resultado, observaram que a integração dos sensores apresentou o melhor método para o propósito do estudo.

A fusão de imagens multiespectrais, ou outros sensores remotos, à dados LiDAR foram utilizadas para diferentes fins nos últimos anos. Popescu; Wynne; Scrivani, (2004) observaram que o uso fusionado com sensores ópticos foi mais eficiente que o uso individual do LiDAR na estimativa de biomassa e volume para coníferas. A técnica se mostrou eficiente no mapeamento de material combustível de florestas para uso como input em modelos de comportamento do fogo (GARCÍA et al., 2011; MUTLU et al., 2008) e no mapeamento de alturas de dossel florestal em extensas áreas (WILKES et al., 2015). A técnica também foi eficiente no desenvolvimento de métodos para mapeamento de áreas florestais (MACHALA; ZEJDOVÁ, 2014). Já no trabalho de Xu, Manley e Morgenroth (2018), a fusão de LiDAR e dados RapidEye não mostrou grandes ganhos na predição de volume e idade de florestas, quando comparado somente com métricas LiDAR. Pflugmacher et al., (2013) utilizaram amostras de estimativas de biomassa por ALS para validar modelos de recuperação de biomassa em série histórica de Landsat, e obtiveram resultados favoráveis da utilização do LiDAR com esse propósito.

2.2.5. Índices de Vegetação

Foram criados nos anos 70 para monitorar e quantificar as condições e distribuições espaciais de vegetação e podem ser obtidos a partir de diversas combinações matemáticas das reflectâncias em várias faixas espectrais da radiação eletromagnética (JENSEN, 2000; LIU, 2006). São úteis para diferenciar a vegetação de outros objetos detectados por sensores remotos e avaliar, entre outras utilidades, ocorrências de eventos catastróficos meteorológicos.

Foram vários os índices criados (BANNARI et al., 1995) dentre eles, destaca-se o índice de vegetação da diferença normalizada (NDVI – *Normalized Difference Vegetation Index*), o mais utilizado por pesquisadores (LIU, 2006). Criado por Rouse et al. (1973), combinou a banda do vermelho (BVM), cuja energia é fortemente absorvida pela vegetação, à banda do infravermelho (BIV), cuja energia é fortemente refletida pela vegetação verde. O cálculo se dá pela diferença entre a reflectância das bandas infravermelho e vermelho, normalizada pela soma das mesmas. A normalização minimiza erros de uma simples combinação entre bandas e permite sua utilização como um índice (ROUSE et al., 1973):

$$NDVI = \frac{BIV - BVM}{BIV + BVM} \quad (1)$$

O valor de NDVI varia de -1 a +1. Os valores negativos representam as nuvens, próximos de zero representam solo nu ou sem vegetação e valores maiores que zero representam vegetação e, quanto mais próximo de um, maior seu vigor vegetativo (LIU, 2006).

2.3. Aprendizado de máquinas e sua aplicação na avaliação de sinistros florestais

O conceito de Inteligência Artificial (IA) foi utilizado pela primeira vez na conferência de Dartmouth em 1956, organizada por John McCarthy, que reuniu pesquisadores de diferentes áreas para discutir o tema (NILSSON, 2009). Kaplan e Haenlein, (2019) definem IA como “a habilidade de um sistema em interpretar corretamente dados externos, aprender com estes dados e usar esse aprendizado para atingir alguns objetivos e tarefas, por uma adaptação flexível”. Aprendizado de

Máquinas (em inglês, *Machine Learning - ML*) é uma parte essencial da IA, que descreve métodos que ajudam computadores a aprender, sem a necessidade específica de serem programados (KAPLAN; HAENLEIN, 2019). Podem ser subdivididos em três classes: aprendizado de máquinas supervisionado, não-supervisionado e de reforço. Neste trabalho será abordado apenas o ML do tipo supervisionado.

Aprendizado de máquinas é utilizado quando se acredita que há relação entre observações de interesse, mas não se sabe exatamente qual (ALPAYDIN, 2016). Se baseia em duas etapas: Treinamento e Predição. O treinamento envolve a aplicação de um determinado algoritmo em uma determinada base de dados para que a máquina aprenda padrões peculiares sobre os dados de treinamento, para enfim validar o aprendizado em uma nova base, até então desconhecida (ALPAYDIN, 2016). O objetivo é ajustar os parâmetros de um modelo, otimizando a performance para o critério pretendido.

Métodos de ML exercem papel cada vez mais importante em análise de dados, pois podem lidar com uma quantidade abundante de informação e, quanto mais informação for utilizada na fase de treinamento, melhor será sua performance (NILSSON, 2009). Os índices de acertos e erros na fase de validação, determinam a eficiência do algoritmo testado (ALPAYDIN, 2016).

São inúmeros os algoritmos utilizados em ML, uns mais sofisticados e outros mais simples, no entanto qualquer método que utiliza o entendimento de padrões de um banco de dados para predições em uma nova base, como por exemplo uma simples regressão linear, pode ser considerado uma técnica de ML.

Há centenas de algoritmos utilizados em ML, dos quais, destacam-se:

- Redes Neurais Artificiais (RNA), um dos primeiros algoritmos de inteligência artificial, utilizado desde antes de 1980 (NILSSON, 2009);
- Random Forest (RF) (BREIMAN, 2001);
- Support Vector Machine (SVM) (VAPNIK, 1988).

Há uma crescente tendência no uso de técnicas de aprendizado de máquinas para lidar com problemas de modelagem de susceptibilidade de florestas ao fogo (TIEN BUI; LE; HOANG, 2018) e também a outros fenômenos adversos. A seguir são citados alguns exemplos.

Um dos primeiros trabalhos na área florestal que utilizou inteligência artificial foi o de Hanewinkel; Zhou; Schill, (2004) que usou redes neurais artificiais para

identificação de florestas susceptíveis a danos por ventos. Rodrigues e de la Riva, (2014) compararam algoritmos de *machine learning*, entre eles o RF e SVM, juntamente com regressão logística, para avaliação da relação entre ocorrências de incêndios florestais causadas por ações antrópicas. Tien Bui; Le; Hoang, (2018) utilizaram novas técnicas de aprendizado de máquinas, derivadas de redes neurais, para modelagem espacial de risco de incêndio. Yu et al., (2017) compararam duas técnicas para previsão de chuva em tempo real, por radar. O RF e redes neurais também foram utilizados por Hart et al., (2019) na predição do risco de danos por vento para árvores individuais. Nguyen et al., (2018) utilizaram técnicas avançadas de machine learning, (variações do SVM e RNA), juntamente com RF para avaliar padrões de perigo de incêndio para floresta tropical. Kim et al., (2019) utilizaram RF e máxima entropia para predizer e analisar a distribuição espacial de risco de incêndio, considerando influência sócio-econômica. Nami et al., (2018) estimaram a probabilidade espacial de incêndios, empregando técnicas de ML e SIG. Bar Massada et al., (2012) compararam modelos lineares generalizados com RF e máxima entropia, para avaliar o potencial da probabilidade de incêndios florestais. Pham et al., (2018) utilizaram SVM e mais um algoritmo para predição de deslizamento de terra. Para mapear distúrbios em florestas, Hislop et al., (2019) testaram um conjunto de abordagens, entre elas, integrando o RF. Cortez e Morais, (2007) testaram RF e SVM, entre outros algoritmos, para avaliar a qualidade da predição de ocorrências de incêndios, baseando-se apenas em dados meteorológicos. Pereira et al., (2017) utilizaram SVM para mapear áreas queimadas na savanna brasileira, utilizando dados de produtos de incêndio de sensores satelitais.

2.4. Modelos preditivos para manejo de incêndios florestais

2.4.1. Regressão Logística

Na definição da probabilidade de incêndio, a variável resposta do modelo é uma variável binária quanto a ocorrência ou não ocorrência de fogo em determinado local assim como na regressão logística, onde a variável resposta é também uma variável binária (HOSMER; LEMESHOW, 2000).

Em análises de regressão cuja variável resposta é binária, as predições são restritas ao intervalo entre 0 e 1 e os erros são distribuídos por uma função binomial,

diferentemente das regressões lineares, nas quais os ajustes dos parâmetros são realizados pelo método dos mínimos quadrados e os erros seguem a distribuição normal (HOSMER; LEMESHOW, 2000). Conforme explicam os autores, o ajuste da regressão logística é realizado pelo método da máxima verossimilhança, que produz estimativas para os parâmetros que maximizam a probabilidade de se obter os dados observados.

Modelos logísticos foram utilizados em trabalhos de predição da probabilidade da ocorrência de incêndio e apresentaram boas acurácias (RICARDO, 2010; GARCIA-GONZALO et al., 2012; MARQUES et al., 2012; BOTEQUIM et al., 2013).

2.4.2. Random Forest

O Random Forest é um classificador que utiliza árvores de decisão (BREIMAN, 2001) e tem sido amplamente utilizado para classificação, regressão e pelo seu potencial em apresentar a importância das variáveis do modelo (GENUER; POGGI; TULEAU-MALOT, 2010; YU et al., 2017).

O principal conceito do algoritmo RF é a criação de uma “floresta de decisão”, que consiste em um conjunto de combinações de árvores de decisões individuais, criadas por amostras compostas por observações e variáveis, extraídas aleatoriamente dos dados originais de treinamento, (esse processo, em inglês, é chamado de *bagging*, e consiste em seleções aleatórias com reposição, ou seja, as amostras podem ser extraídas mais de uma vez em cada seleção) (BREIMAN, 2001). Segundo o autor, para cada observação dos dados em que será feita a predição, é atribuída uma classificação, estimada pela combinação das classificações de árvore individuais. À cada classificação de uma árvore individual é contabilizado um voto, e a classificação final é atribuída à classe com maior votação.

Para a construção de um árvore de decisão individual, cerca de dois terços dos dados de treinamento selecionados pela técnica de *bagging* são utilizados pelo modelo, e o terço faltante é chamado de “*out-of-bag*” (OOB) (BREIMAN, 2001). A amostra OOB é utilizada para estimativa do erro de predição e, em seguida, para avaliação das importâncias das variáveis (GENUER; POGGI; TULEAU-MALOT, 2010).

2.4.3. Support Vector Machine

O SVM se baseia na utilização de otimização matemática, implementando um limite derivado da “Teoria de Aprendizagem Estatística” que busca encontrar um separador ótimo de hiperplanos em relação às amostras mais próximas de treinamento, conhecida como “vetores de suporte” (support vectors) (VAPNIK, 1988). O algoritmo é utilizado para reconhecimento de padrões, classificações, regressões. Em sua função mais simples, utiliza um hiperplano linear para criar a classificação, e em caso da impossibilidade da separação linear de dados, utiliza funções kernel (VAPNIK, 2000).

Em ML, kernel é normalmente utilizado para “transformar” classificadores lineares para solução de problemas não-lineares. A função kernel é aplicada a cada dado para mapear a observação originalmente não-linear em observações de um espaço com maior dimensão, na qual são separáveis (VAPNIK, 2000). Existem diferentes funções kernel incorporadas nas formulações de SVM que influenciam na performance do modelo, como por exemplo, polinomial, sigmoideal e função de base radial – RBF (TIEN BUI et al., 2016; VAPNIK, 2000). Após a transformação, o SVM encontra um hiperplano de separação linear nesse novo espaço (SAKR; ELHAJJ; MITRI, 2011). Neste trabalho foi testado o SVM com função kernel polinomial de graus 1, 2 e 3.

Em um exemplo de aplicação do SVM, quando utilizado para classificação de ocorrências e não ocorrência de incêndios, se mostrou mais eficiente que outro algoritmo de machine learning o qual foi comparado (SAKR; ELHAJJ; MITRI, 2011).

3. MATERIAIS E MÉTODOS

3.1. Área de estudo

A área de estudo compreende plantios de eucalipto (*Eucalyptus* spp.) de diferentes idades e materiais genéticos, pertencentes a uma empresa de papel e celulose, localizada no sul do estado da Bahia, que cedeu dados para realização deste trabalho. Abrange sete municípios: Alcobaça, Caravelas, Ibirapuã, Mucuri, Nova Viçosa, Prado e Teixeira de Freitas.

O clima predominante da região é o do tipo Af, segundo Koppen, clima tropical sem estação seca. A temperatura média anual é de aproximadamente 24° C, sendo julho o mês mais frio com média de 21°C e fevereiro o mês mais quente com média de 26°C. A precipitação média da região é de aproximadamente 1.350 mm por ano. O mês mais chuvoso é novembro, com média de 150 mm e o menos chuvoso é agosto, com média de 70 mm. (ALVARES et al., 2013).

A base cadastral completa disponibilizada pela empresa abrange um total de 407 fazendas, dispostas em 8.215 talhões, ocupando uma área de aproximadamente 100 mil hectares (Figura 2).

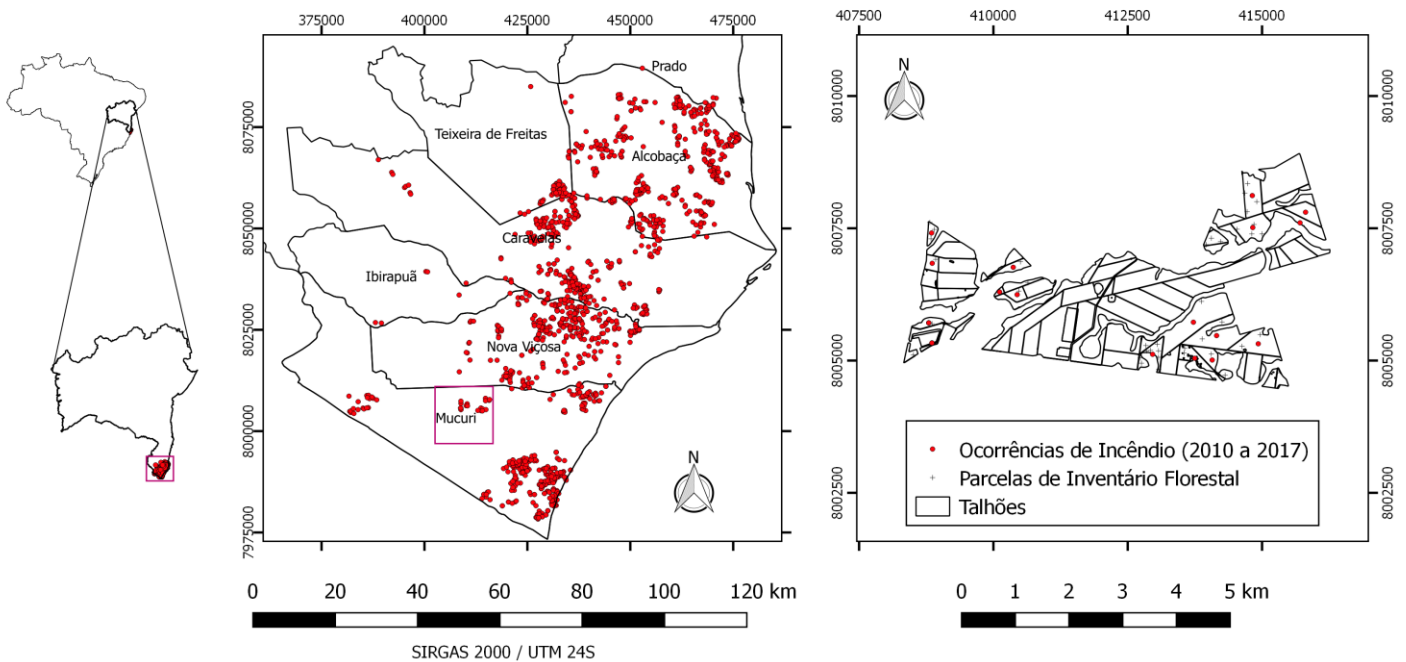


Figura 2. Localização da área de estudo. Destaque para localização das ocorrências de incêndio e parcelas de inventário florestal entre 2010 e 2017. À direita, uma fazenda localizada no município de Mucuri.

3.2. Sobrevoos ALS

Entre agosto e outubro de 2015 a região de estudo foi sobrevoada com ALS, utilizando o sensor LiDAR Harrier 68i da Trimble, Sistema Inercial Applanix (IMU+GPS), modelo POS AV 150 em uma aeronave CESSNA modelo 206. O sobrevoos contou com sobreposição de 30% e intensidade média de cinco pulsos por metro quadrado.

A área escaneada cobriu um total de 269 fazendas e 5.919 talhões, dispostos em aproximadamente 70 mil hectares. Esses talhões com sobrevoos LiDAR foi a base para a modelagem da probabilidade de incêndio.

3.3. Dados de ocorrência de incêndio

A empresa proprietária das fazendas localizadas na área de estudo compartilhou seu banco de dados de ocorrência de incêndio. Foram incluídos todos os registros com área queimada a partir de 0,01 hectares. Essas informações estão contidas na Tabela 1.

O número total de talhões com registros de incêndio para o intervalo de oito anos, foi de 1.422. O número de talhões não queimados, “acumulados” no período, foi de 34.443. O percentual de registros com incêndio para o período de oito anos foi de 4,1%. O ano de 2015 foi o de maior ocorrência, num total de 461, equivalente a 9,8% do total dos talhões avaliados no ano.

Tabela 1. Número de talhões queimados (Q) e não-queimados (NQ), por ano.

Ano	NQ	Q
2010	4.213	32
2011	4.091	142
2012	4.064	141
2013	4.198	196
2014	4.405	203
2015	4.235	461
2016	4.519	183
2017	4.718	64
Total	34.443	1.422

3.4. Base para mapeamento da probabilidade de incêndio

A modelagem da probabilidade de incêndio envolve variáveis que caracterizam fatores relacionados à ignição e propagação dos incêndios.

Quanto à ignição, é muito importante que sejam consideradas variáveis sociais, devido a conflitos existentes entre empresas plantadoras de eucalipto e a sociedade em seu entorno. A introdução do eucalipto no extremo sul do estado da Bahia aconteceu a partir dos anos 80 e foi intensificado após o ano 2000, com a chegada das empresas de papel e celulose. Embora tenha melhorado a economia da região, pouco impactou em mudanças sociais (ALMEIDA et al., 2008), mantendo os municípios com IDH abaixo da média nacional (ALMEIDA et al., 2008; IBGE, 2010). A expansão da silvicultura nessa região resultou em um processo intenso de êxodo rural e uma reorganização socioeconômica, ocasionando crescimento desordenado das áreas urbanas, provocando novas paisagens locais (ALMEIDA et al., 2008). Santos e Silva (2004) apontam que essa expansão contribuiu para inviabilidade da agricultura familiar na região, ocupou áreas que seriam destinadas à reforma agrária e terras indígenas. Os autores relatam atos de movimentos sociais em protestos contra as empresas de papel.

Quanto à propagação, variáveis climáticas e biométricas são os principais agentes relacionados à dispersão das chamas, embora também exerçam influência sobre a ignição.

3.4.1. Preparo dos dados LiDAR – Modelagem 1

A partir das nuvens LiDAR foi feita a classificação de pontos de solo e, a partir dos pontos de solo, foi extraído um Modelo Digital do Terreno (MDT) para grid de dez metros, onde, à cada pixel, foi atribuída a informação de elevação. Pela diferença entre a altura do retorno de cada pulso laser em relação ao MDT, foi feita a normalização das nuvens. Essas etapas foram realizadas utilizando o software LAsTools (ISENBURG, 2018).

A partir das nuvens de pontos normalizadas, foram calculadas as métricas LiDAR para modelagem de volume e caracterização da estrutura de sub-bosque. Essa etapa foi feita utilizando o software estatístico R (R CORE TEAM, 2019), por meio do pacote lidR (ROUSSEL; AUTY, 2019).

Para gerar o modelo de volume, foram geradas métricas de posição (elevação média, elevação máxima e percentis – 60, 70, 75, 80, 90, 95, 99) e de dispersão (assimetria e curtose da elevação e desvio padrão e variância). Os valores mínimos, máximos e médios de cada métrica estão apresentados na Tabela 2.

Tabela 2. Valores mínimos, médio e máximo para as métricas utilizadas para modelagem de volume dos dados LiDAR.

Métrica	Valor Mínimo	Valor Médio	Valor Máximo
Média	0,84	4,58	12,06
Máxima	9,47	20,09	30,51
Desvio Padrão	2,18	6,83	12,13
Variância	4,73	50,53	147,12
Assimetria	-0,55	1,14	2,86
Curtose	-1,94	-0,16	6,85
p60	0,00	1,82	20,89
p70	0,00	5,95	22,30
p75	0,00	9,19	24,76
p80	0,00	11,86	25,86
p90	4,52	16,19	27,37
p95	6,28	17,41	28,22
p99	7,79	18,85	29,14

A correlação entre métricas LiDAR e dados de inventário permitiu a confecção de um modelo de volume. O modelo gerado (equação 2) foi baseado em duas métricas, a primeira referente à altura do povoamento, representada pelo percentil 75 ($p75$) das alturas, e uma métrica referente a dispersão dos pontos, representada pela variância das alturas dos retornos (var). O valor do coeficiente de determinação ajustado (R^2 -ajust) do modelo foi de 0,8 e a Raiz do Erro Médio Quadrático (RMSE), de 28,4 m³/ha.

$$\text{Volume (m}^3\text{/ha)} = 12.5069 - 3.1148 * p75 + 2.56172 * var \quad (2)$$

As métricas geradas para caracterizar a estrutura de sub-bosque foram baseadas no índice de área foliar (LAI) (equação 4) do perfil vertical da copa, que é calculado em função da densidade de área foliar (LAD), que por sua vez é calculada pela equação de MacArthur-Horn (STARK et al., 2012; ALMEIDA et al., 2019) (equação 3).

$$LAD_i = \ln\left(\frac{pulses.in_i}{pulses.out_i}\right) * \frac{1}{Dz} * \frac{1}{K} \quad (3)$$

Para cada um dos 5.919 talhões foi gerada uma nuvem de pontos exclusiva, que posteriormente foi sub-dividida em voxels de 2x2x1 metros. Para essas etapas foi utilizada a ferramenta lasclip e lasgrid do pacote lidR (ROUSSEL; AUTY, 2019). O cálculo do LAD leva em consideração o número de pulsos que entram em cada voxel (*pulses.in*) e o número de pontos que passam pelo voxel (*pulses.out*). *Dz* é a resolução vertical, fixo em um metro e *K* é o coeficiente de calibração do LAD para um LAI conhecido. Foi fixado um valor de *K* igual 1, calculando-se assim um “LAD efetivo” e um “LAI efetivo” (ALMEIDA et al., 2019).

$$LAI = \sum_{i=1}^n LAD \quad (4)$$

Para este trabalho foi utilizado o valor de LAI correspondente à altura total, à camada de um a cinco metros e à camada de um a dez metros: LAI, LAI_1a5m, LAI_1a10m, respectivamente. Por último foi calculada a variável “SubBosque” para representar o sub-bosque das florestas, pela relação entre o LAI_1a5m e LAI total.

$$SubBosque (\%) = \frac{LAI_{1a5m}}{LAI_{total}} * 100 \quad (5)$$

3.4.2. Estrutura da base de dados para modelagem

O modelo anual de probabilidade de incêndios tem como variável dependente um dado binário referente à “ocorrência” ou “não-ocorrência” de incêndio em determinado talhão e em determinado ano, mais uma série de variáveis independentes, podendo ser biométricas, climáticas, sociais ou métricas relativa ao sobrevoo ALS.

Para cada talhão que faz parte do estudo, foi criada uma observação para cada ano do período de 2010 à 2017, desde que o mesmo estivesse cadastrado pela empresa como “plantado”, no respectivo ano. Ou seja, cada talhão pode ter até 8 observações na base final e, para cada uma, foi atribuído uma “data de referência” para as análises, obedecendo os seguintes critérios:

- a) Ano em que ocorreram um ou mais incêndios em um único talhão, utilizou-se a data de detecção do primeiro registro de fogo;
- b) Ano em que foi realizado inventário no respectivo talhão, utilizou-se a data de inventário;
- c) Ano em que o talhão não sofreu incêndio ou não teve parcela de inventário, foi utilizado o dia 31 de dezembro do ano em questão.

Foi necessário a definição da data de referência, com a finalidade de definir a idade da floresta.

Cada observação foi classificada entre queimada e não-queimada (Tabela 1).

3.4.3. Ocorrências de incêndios

As ocorrências de incêndio foram organizadas em duas escalas temporais: i) para o período de 2010 à 2017; ii) somente para 2015, ano do ALS.

Nos casos em que ocorreram mais de um incêndio em determinado ano e em um mesmo talhão, foi considerado apenas como uma ocorrência e a área afetada foi somada.

3.4.4. Parcelas de inventário

A empresa disponibilizou para este trabalho dados biométricos de 13.042 parcelas de inventário florestal, únicas ou multitemporais, coletadas no período de 2005 a 2017. Desta base de dados foi utilizado o Volume Comercial com Casca – VCCC, como variável para representar a estrutura do povoamento. As parcelas do ano de 2015 também foram utilizadas para modelagem dos dados ALS.

3.4.5. Modelagem de dados biométricos

Para cada observação da base para modelagem (talhão e ano), quando não existiram parcelas de inventário correspondente, o volume foi estimado a partir de modelos desenvolvidos utilizando a própria base de dados fornecidas pela empresa.

Para predição de volume foi utilizado um modelo não-linear adaptado de Richards (1959), bastante flexível quando utilizado para o ajuste de curvas de

crescimento (RODRIGUEZ, 2005; PRATA, 2012), determinado em função da idade do povoamento, neste caso, em anos. A modelagem se baseou em 7.285 parcelas.

$$\text{Volume} \left(\frac{\text{m}^3}{\text{ha}} \right) = [13,8173 * (1 - e^{-0,2670*idade})]^{2,2866} \quad (6)$$

3.4.6. Dados climáticos

Para compor o modelo foram criadas as seguintes variáveis climáticas:

Face de exposição do terreno (aspect) foi considerada como uma variável climática contínua. Essa informação foi obtida utilizando a função “terrain”, do pacote “raster” (HIJMANS; ET TEN, 2019), no software R (R CORE TEAM, 2019). A função “terrain” utiliza como input dados de elevação, por grid, que neste caso foram obtidos pelo DTM de cada local, criado em função dos dados LiDAR.

O resultado desta aplicação gera, para cada pixel, a direção em graus da face de exposição da declividade do terreno em relação aos pontos cardeais. É uma medição sentido horário de 0° (face norte) para 360° (também face norte), completando o círculo (GEORGIADOU et al., 2001). Caso essa informação, em graus, seja tratada diretamente como uma variável contínua pelos valores de 0 a 360, haverá um erro na interpretação dos dados, pois os valores extremos da escala possuem a mesma representação, no caso, face norte. Para contornar este problema, há a necessidade de transformação dos valores. A estratégia adotada foi utilizar o cosseno (cos) do ângulo, pois $\cos(0^\circ) = \cos(360^\circ) = 1$. Assim, a escala dos valores para a face do terreno pode variar de -1 a 1.

Foram utilizados dados diários de precipitação (mm), disponibilizados pelo Instituto Nacional de Meteorologia – INMET - do período de 2003 a 2017, para o município de Caravelas/BA, (Banco de Dados Meteorológicos para Ensino e Pesquisa – BDMEP, 2018) para criação da variável preditora “Precipitação Total Anualizada”, que representa a soma do total de precipitação recebida pela respectiva floresta, desde sua data de plantio à “data de referência” estabelecido para o ano em questão, dividido pela idade da floresta, conforme expressão:

$$\text{PRECIPITAÇÃO TOTAL ANUALIZADA} \left(\frac{\text{mm}}{\text{anos}} \right) = \frac{\text{PRECIPITAÇÃO TOTAL (mm)}}{\text{IDADE (anos)}} \quad (7)$$

Também fizeram parte da criação das variáveis climáticas dados de seis estações meteorológicas próprias da empresa. Para cada talhão, foi atribuído os valores da estação mais próxima, para compor as seguintes variáveis:

- Temperatura média anual (°C);
- Precipitação média anual (mm); e
- Umidade relativa do ar média.

Para encontrar a distância mais próxima entre o talhão e estação meteorológica, foi utilizado a ferramenta “Nearest Neighbour join”, NNJoin (TVEITE, 2018), disponível no software Quantum Gis (QGIS DEVELOPMENT TEAM, 2018). Essa mesma ferramenta foi utilizada para definir as variáveis sociais de distâncias.

3.4.7. Dados sociais

As seguintes “variáveis sociais” foram criadas:

- Distâncias para municípios: Mínima distância, em metros, do centróide de cada talhão em relação ao polígono de “Áreas Urbanizadas” mais próximo (IBGE, 2018) e vetores próprios;
- Distâncias para estradas: Mínima distância, em metros, do centróide de cada talhão em relação aos vetores de “Estradas” (Departamento Nacional de Infraestrutura e Transporte - DNIT, 2018) e vetores criados para as estradas da região;
- Distância para assentamentos: Mínima distância, em metros, do centróide de cada talhão em relação aos vetores de “Assentamentos” (INCRA, 2018);
- População Residente por município: número total de habitantes do município em determinado ano (Departamento de Informática do SUS - DATASUS, 2018);
- Densidade demográfica: habitantes por km² por município (SIDRA, 2018);
- População residente na zona rural: número de habitantes na Zona Rural, por município, no ano de 2010 (SIDRA, 2018).

Em resumo, as Tabelas 3 e 4 apresentam as variáveis contínuas utilizadas nos modelos, por seus valores mínimos, médios e máximos, encontrados para cada registro de talhões queimados ou não queimados.

Tabela 3. Valores mínimos, médio e máximo para variáveis contínuas, em relação a registros de talhões não queimados (NQ) e queimados (Q), no período de 2010 a 2017.

Variável	Descrição	Mínimo		Médio		Máximo	
		NQ	Q	NQ	Q	NQ	Q
Volume Comercial com Casca - VCCC	m ³ /ha	0	0	95	92	399	404
Face de Exposição do Terreno - Aspect	#	-1	-1	-1	-1	1	1
Precipitação do Ano - PA	mm/ano	672	672	1.112	1.018	1.480	1.480
Precipitação Total Anualizada- PTA	mm/ano	0	0	1.391	1.432	5.667	13.408
Temperatura Média- Temp	°C	22	22	23	23	25	25
Umidade Relativa do Ar Média – UR	%	59	59	79	79	84	83
Distância para Área Urbana – DistUrb	m	54	92	7.779	7.378	17.377	17.377
Distância para Estrada – DistEst	m	23	51	5.094	4.600	16.460	15.775
Distância Assentamento Rural – DistAssent	m	0	0	17.947	16.759	41.373	40.778
População Municipal – PopTot	nº hab	19.417	21.048	43.293	47.438	161.690	161.690
Densidade Demográfica – Dens	hab/km ²	9	9	29	31	119	119
População Zona Rural - PopZR	nº hab	5.030	5.030	8.590	8.543	12.153	12.153

Tabela 4. Valores mínimos, médio e máximo para variáveis contínuas, em relação a registros de talhões não queimados (NQ) e queimados (Q), no período de 2015.

Variável	Descrição	Mínimo		Médio		Máximo	
		NQ	Q	NQ	Q	NQ	Q
Volume Comercial com Casca - VCCC	m ³ /ha	0	0	84	66	395	404
Face de Exposição do Terreno - Aspect	#	-1	-1	-1	-1	1	1
Precipitação do Ano - PA	mm/ano	710	710	783	795	871	871
Precipitação Total Anualizada- PTA	mm/ano	0	0	1.085	1.358	1.905	3.709
Temperatura Média- Temp	°C	23	23	23	24	25	25
Umidade Relativa do Ar Média – UR	%	76	76	79	79	80	80
Distância para Área Urbana – DistUrb	m	92	531	7.912	6.873	17.377	16.508
Distância para Estrada – DistEst	m	23	51	5.154	4.509	16.460	14.653
Distância Assentamento Rural – DistAssent	m	0	0	17.974	18.317	41.373	40.778
População Municipal – PopTot	nº habs	22.548	22.548	44.344	55.122	157.804	157.804
Densidade Demográfica – Dens	hab/km ²	9	9	28	37	119	119
População Zona Rural - PopZR	nº hab	5.030	5.030	8.522	8.933	12.153	12.153
Volume - Vol_ALS	m ³ ha	10	12	50	69	346	305
LAI	índice	0	0	0,2936	0,3085	1,4755	1,1100
LAI 1 a 5m	índice	0	0	0,0948	0,0719	1,1890	0,5594
LAI 1 a 10m	índice	0	0	0,1623	0,1307	1,2135	0,7417
SubBosque	%	0	2	44	33	100	100

3.4.8. Modelos testados

Neste trabalho, a probabilidade da ocorrência de incêndios foi calculada comparando um método comumente utilizado neste tipo de estudo, a Regressão Logística (BOTEQUIM et al., 2013; MARQUES et al., 2012), com um algoritmo bastante utilizado em ML, Random Forest.

A probabilidade calculada pelo modelo logístico é definida da seguinte forma:

$$\text{Probabilidade (Incêndio)} = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \dots + \beta_n X_n)}} + \varepsilon \quad (8)$$

O Random Forest é um classificador que consiste em construir uma coleção de “árvores de decisão” independentes, onde os resultados são combinados para efetuar a predição em uma determinada base de dados (BREIMAN, 2001). O algoritmo é uma poderosa ferramenta para seleção de variáveis particularmente adequada para lidar com predições de problemas que inclui relações não-lineares e interações complexas entre variáveis (GENUER; POGGI; TULEAU-MALOT, 2010).

Para cada algoritmo foram feitas quatro abordagens, onde foi modificado o período dos dados modelados e as variáveis testadas (Tabela 5).

As duas primeiras abordagens utilizaram exatamente as mesmas variáveis para comparação entre a probabilidade anual de incêndio para o período de 2010 a 2017 e, para o ano de 2015, quando ocorreu o sobrevoo LiDAR. A abordagem 3 teve o incremento dos dados LiDAR ao modelo e, na abordagem 4, apenas métricas e provenientes do ALS como fonte de informações sobre a estrutura da floresta, junto aos dados climáticos e sociais.

Tabela 5. Tipos de variáveis testadas e período analisado em cada abordagem.

Abordagem	Período	Tipos de Variáveis Testadas
1	2010 a 2017	Biométricas, Climáticas, Sociais
2	2015	Biométricas, Climáticas, Sociais
3	2015	Biométricas, Climáticas, Sociais, Dados ALS
4	2015	Climáticas, Sociais, Dados ALS

Algoritmos de aprendizado de máquinas se mostraram eficientes para prever o risco de incêndio utilizando um número limitado de *inputs* (SAKR; ELHAJJ; MITRI, 2011). Para cada abordagem, foram testados três modelos, reduzindo o

número de variáveis em cada teste, conforme correlação entre as mesmas, avaliado por dendograma representando um *cluster* hierárquico.

Um cluster hierárquico é gerado a partir de uma matriz de similaridade, para construção de uma “árvore”, representando relações específicas entre “entidades” (variáveis preditoras, neste caso). Cada entidade gera um “ramo” específico, que vai se unindo em clusters, até a formação da “raiz”, que representa a coleção inteira de entidades. Quando dois clusters se unem, formam um “bloco” para associação com outros clusters (ANDERBERG, 1973). As alturas de conexão entre cluster pode ser definido por diferentes métodos. Neste trabalho, foi utilizado o método de Ward (ANDERBERG, 1973; MURTAGH; LEGENDRE, 2011; WARD, 1963).

Agrupamento Hierárquico Variáveis

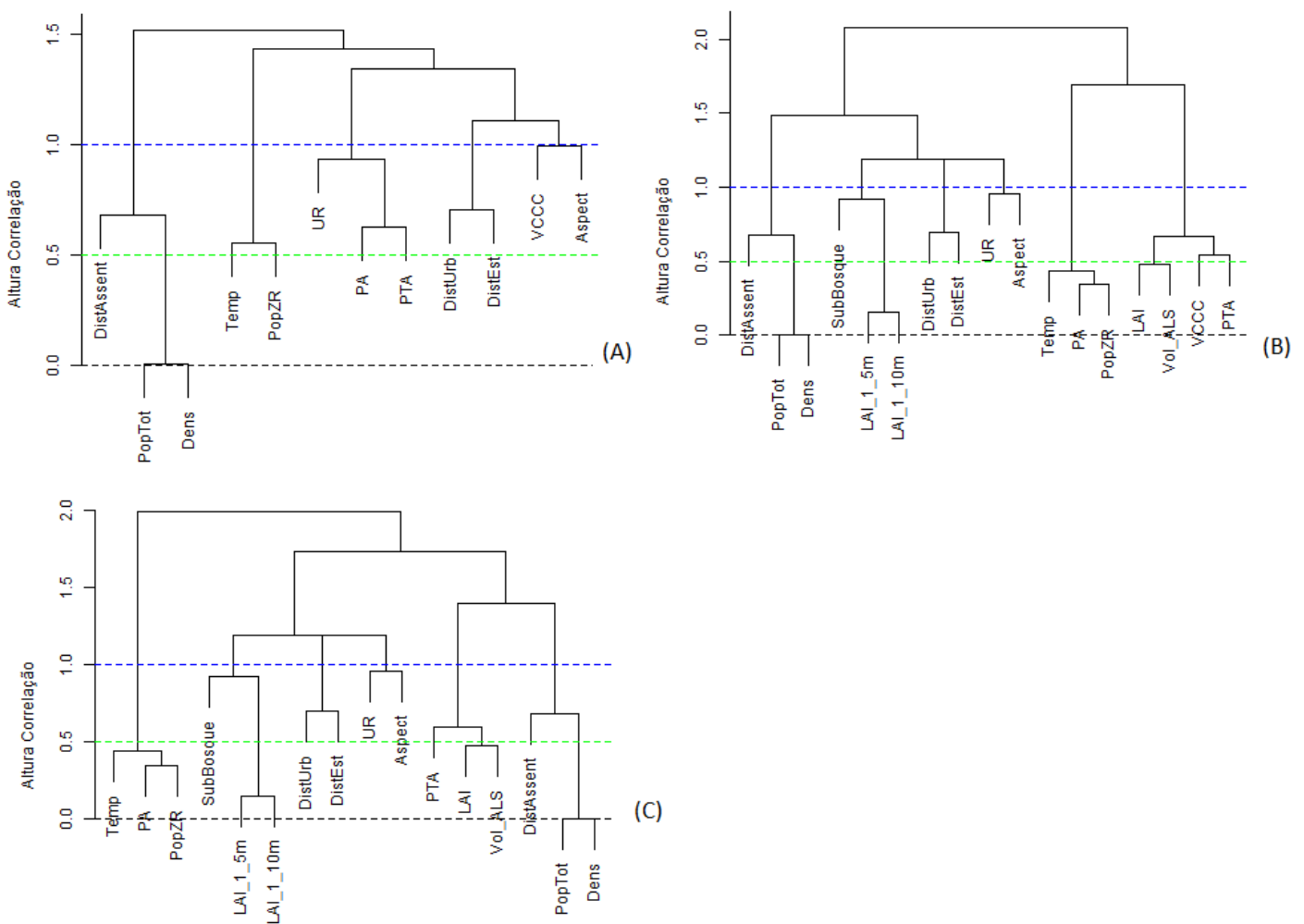


Figura 3. Cortes por altura de correlação para seleção de variáveis contínuas na composição dos modelos. Para cada abordagem, cortes nas alturas 0, 0,5 e 1. A) Variáveis para abordagens 1 e 2. B) Abordagem 3. C) Abordagem 4.

Foram selecionados os grupos de variáveis que compuseram cada modelo a partir de três “cortes” em determinadas alturas do gráfico de dendrograma, conforme ilustra a Figura 3, por meio de linhas tracejadas, coloridas à cada nível.

O primeiro passo foi executar para cada abordagem o modelo completo (corte em “zero”), com todas as variáveis preditoras, e avaliar a importância das variáveis identificada pelo algoritmo Random Forest. Em cada corte subsequente, foi decidido em permanecer no modelo com a variável com maior importância, por agrupamento.

Para as abordagens 1 e 2, partiu-se de 12 variáveis preditoras no primeiro teste, para 5 no terceiro (Tabela 6). Na abordagem 3, foram inseridas 5 variáveis provenientes do sobrevoo LiDAR, partindo de 17 variáveis, até chegar em 6 no teste 3 (Tabela 7). Já na abordagem número 4, o primeiro teste foi com 19 variáveis, finalizando com 6 no terceiro teste (Tabela 8).

Tabela 6. Conjunto de variáveis utilizadas em cada teste das abordagens 1 e 2.

Modelos	Variáveis
1	VCCC, Aspect, PA, PTA, Temp, UR, DistUrb, DistEst, DistAssent, PopTot, Dens, PopZR
2	VCCC, Aspect, PA, PTA, Temp, UR, DistUrb, DistEst, DistAssent, PopTot, PopZR
3	VCCC, PTA, DistUrb, DistAssent, PopZR

Tabela 7. Conjunto de variáveis utilizadas em cada teste da abordagem 3.

Modelos	Variáveis
1	VCCC, Aspect, PA, PTA, Temp, UR, DistUrb, DistEst, DistAssent, PopTot, Dens, PopZR, Vol_ALS, LAI, LAI_1_5m, LAI_1_10m, SubBosque
2	VCCC, Aspect, PTA, UR, DistUrb, DistEst, DistAssent, PopTot, PopZR, Vol_ALS, LAI_1_10m, SubBosque
3	Aspect, PTA, UR, DistUrb, DistAssent, PopZR, SubBosque

Tabela 8. Conjunto de variáveis utilizadas em cada teste da abordagem 4.

Modelos	Variáveis
1	Aspect, PA, PTA, Temp, UR, DistUrb, DistEst, DistAssent, PopTot, Dens, PopZR, Vol_ALS, LAI, LAI_1_5m, LAI_1_10m, SubBosque
2	Aspect, PTA, UR, DistUrb, DistEst, DistAssent, PopTot, PopZR, Vol_ALS, LAI_1_10m, SubBosque
3	Aspect, PTA, DistUrb, DistAssent, PopZR, SubBosque

Para avaliar a performance dos modelos, foi utilizado o método da validação cruzada, que é utilizada para obtenção de estimativa não-viesada da performance de cada modelo, em um banco de dados desconhecido. Neste trabalho, a validação dos modelos foi feita pela validação cruzada em pastas (*k-fold cross validation*), com *k* igual a 10.

Na avaliação do desempenho de modelos onde a proporção de um fenômeno é muito maior que a do fenômeno inverso (ex: proporção aproximada de 95% para não-ocorrência e 5% para ocorrência), é esperado que a acurácia da modelagem apresente sempre bons resultados (próximos de 100%), pois é natural que os modelos acertem os casos que não ocorreram incêndio e dêem uma falsa impressão de eficiência (METZ, 1978).

Neste estudo, a eficácia dos modelos foi avaliada utilizando os seguintes índices: Sensibilidade, Especificidade, Performance e AUC, também utilizados em trabalhos similares como de Nguyen et al., (2018) e Tien Bui; Le; Hoang, (2018). Os valores de AUC foram representados pela sigla ROC na comparação dos métodos.

A relação entre a Sensibilidade e a Fração Falso Positivo (1 - Especificidade), gera um gráfico denominado ROC (*Receiver Operator Characteristics*), e o valor da área de baixo da curva do gráfico (AUC) representa um parâmetro de eficiência do modelo (FAWCETT, 2006; METZ, 1978) e possui uma propriedade estatística importante: O AUC de um classificador vai ranquear aleatoriamente escolhas de instâncias positivas maiores que instâncias negativas (FAWCETT, 2006).

A matriz de confusão (Figura 4) sumariza a aplicação da modelagem do determinado algoritmo, no novo banco de dados teste. Um caso em que é bastante utilizada é o de duas classes, a primeira designada de “classe positiva” e a segunda de “classe negativa”, formando quatro células de uma matriz: verdadeiros positivos (VP), falsos positivos (FP), verdadeiros negativos (VN) e falsos negativos (FN) (TING, 2011).

		Dado Observado	
		Positivo	Negativo
Dado Predito	Positivo	VP	FP
	Negativo	FN	VN

Figura 4. Esquema da matriz de confusão para duas classes.

A Sensibilidade avalia como o modelo se comportou na predição dos valores positivos avaliados, e é calculado pela seguinte forma:

$$\text{Sensibilidade} = \frac{VP}{(VP+FN)} \quad (9)$$

A Especificidade indica a eficiência na predição dos valores negativos, sob a ótica dos dados observados:

$$\text{Especificidade} = \frac{VN}{(VN+FP)} \quad (10)$$

A Performance avalia o quanto o modelo foi eficiente na predição da classe de interesse da análise, neste caso, a ocorrência de incêndios. Sobre os valores preditos como tal, o índice indica o percentual de acerto:

$$\text{Performance} = \frac{VN}{(VN+FN)} \quad (11)$$

$$\text{Fração Falso Positivo (FFP)} = \frac{FP}{(FP+VN)} \quad (12)$$

Neste trabalho, a ocorrência de incêndio pertence à “classe negativa” da matriz de confusão, ou seja, para o cálculo da probabilidade de sua ocorrência, é importante que, o modelo execute boa performance para a classificação de registros como Verdadeiro Negativo (VN). Nos casos onde a eficácia de sensibilidade e ROC forem similares, as melhores performances nos índices de especificidade e performance definem o melhor modelo.

3.5. Base para mapeamento das cicatrizes de incêndio

Para este mapeamento, foram selecionadas ao acaso 34 fazendas que foram sobrevoadas com LiDAR, dentre as 269 pertencentes à área de estudo.

3.5.1. Realidade de campo – Dados de ocorrência de incêndios

O registro de ocorrências de incêndio para o ano de 2015 disponibilizado pela empresa, foram utilizadas como referência para realidade de campo e geração de dados de treinamento de áreas queimadas, para o período de janeiro a julho.

3.5.2. Preparo dos dados LiDAR – Modelagem 2

A partir das nuvens de pontos normalizadas foram extraídas as métricas (Tabela 9) para pixels com 5 metros de resolução, que compuseram 29 variáveis independentes do modelo. Essas etapas foram realizadas utilizando o software LAStools (ISENBURG, 2018). A Tabela 10 apresenta os valores mínimos, médios e máximos para cada uma das métricas utilizadas no modelo.

As seguintes métricas foram geradas.

Tabela 9. Relação de métricas ALS utilizadas para modelagem, baseada nas alturas dos retornos das nuvens de pontos normalizadas.

Métricas	Código
Altura Máxima dos Retornos	max
Altura média dos Retornos	media
Altura Mínima dos Retornos	min
Assimetria	assm
Bicentis de alturas: 10, 50 e 90 (eq 3.1 e 3.2)	b ₁₀ , b ₅₀ , b ₉₀
Cobertura de Dossel - ausência (eq 3.3)	doss_aus
Curtose das Alturas	curt
Densidade de Dossel - ausência (eq 3.4)	dns_aus
Desvio Padrão das Alturas	desvPad
Intensidade Assimetria	int_assm
Intensidade Curtose	int_curt
Intensidade Desvio Padrão	int_desvPad
Intensidade Máxima	int_max
Intensidade Média	int_media
Intensidade Média da Altura ao Quadrado	int_alt2
Intensidade Mínima	int_min
Intensidade Percentis: 5, 25, 50, 75, 90	int_p05, int_p25, int_p50, int_p75, int_t90
Média da Altura ao Quadrado	alt2
Percentis de altura: 5, 25, 50, 75, 90	p05, p25, p50, p75, p90

Os valores dos bicentis de altura representam o percentual de pontos existentes entre a altura de corte (AC) e a altura determinada (h_{b_x}) (ISENBURG, 2018). Por exemplo, no cálculo do bicentil “x” (b_x) de uma parcela com altura máxima (h_{max}), temos:

$$b_x = \frac{n^{\circ} \text{ total de pontos entre AC e } h_{b_x}}{n^{\circ} \text{ total de pontos } > AC} \quad (13)$$

Onde:

$$h_{b_x} = x\% * (h_{max} - AC) + AC \quad (14)$$

A métrica “Cobertura de Dossel - ausência” foi calculada pelo número total de primeiros retornos acima da altura de corte (AC) de 1.37 metros, dividido pelo número total de primeiros retornos acima do corte de cobertura (CC) de 5 metros:

$$doss_aus (\%) = \frac{n^{\circ} \text{ total de primeiros retornos } (>AC)}{n^{\circ} \text{ total de primeiros retornos } (>CC)} \quad (15)$$

A métrica denominada “Densidade de Dossel - ausência”, é similar à anterior, com a diferença de analisar todos os retornos:

$$dns_aus (\%) = \frac{n^{\circ} \text{ total de retornos } (>AC)}{n^{\circ} \text{ total de retornos } (>CC)} \quad (16)$$

Essas métricas indicam a ausência de cobertura e densidade de dossel, e é calculada pelo inverso do que seria a cobertura e densidade de dossel.

Tabela 10. Valores máximos, médios e mínimos obtidos por cada métrica, nas três classes: INC: áreas pertencentes a cicatriz de incêndio; C/TN: áreas consideradas como “colheita” ou “terra-nua”; e P: locais com plantio de eucalipto em diferentes idades.

Métrica ALS	Descrição	Mínimo			Médio			Máximo		
		INC	C/TN	P	INC	C/TN	P	INC	C/TN	P
max	m	0	0	0	11	0	11	29	17	30
media	m	0	0	0	8	0	9	28	11	25
min	m	0	0	0	4	0	3	28	5	24
assm	m	-5	-1	-6	0	0	-1	7	2	4
dns_aus	%	0	13	11	90	100	79	100	100	100
doss_aus	%	0	13	11	89	100	77	100	100	100
b10	%	0	0	0	6	1	2	96	57	88
b50	%	0	0	0	18	5	9	99	94	95
b90	%	0	0	0	41	8	45	100	100	99
curt	m	0	0	0	3	0	3	58	8	43
desvPad	m	0	0	0	2	0	2	17	4	14
int_assm	#	-1	-2	-1	0	0	1	5	2	6
int_curt	#	0	0	0	2	0	3	35	8	42
int_desvPad	#	0	0	0	39	4	30	279	212	166
int_max	#	0	0	0	209	25	188	1.039	517	786
int_media	#	0	0	0	116	18	102	583	322	299
int_alt2	#	-32.767	-32.757	-32.746	6.035	-32	13.210	32.765	32.762	32.766
int_min	#	0	0	0	64	12	59	583	302	299
int_p05	#	0	0	0	72	13	67	583	304	299
int_p25	#	0	0	0	90	16	82	583	312	299
int_p50	#	0	0	0	111	18	98	583	322	321
int_p75	#	0	0	0	136	21	117	583	363	416
int_p90	#	0	0	0	162	23	138	624	376	544
alt2	m ²	0	0	0	128	0	134	798	126	649
p05	m	0	0	0	5	0	5	28	6	24
p25	m	0	0	0	7	0	8	28	7	26
p50	m	0	0	0	9	0	10	28	13	27
p75	m	0	0	0	10	0	10	28	13	28
p90	m	0	0	0	10	0	11	29	15	28

3.5.3. Dados do Satélite RapidEye

As imagens foram obtidas a partir do portal GEO CATÁLOGO, do Ministério do Meio Ambiente (MMA) (MMA, 2018), que disponibiliza as imagens RapidEye como apoio de “informações geoespaciais temáticas” do Programa de Regularização Ambiental brasileiro. Por meio de convênio entre instituições de ensino e MMA, a disponibilidade das imagens é permitido para fins de pesquisa.

As cenas utilizadas da constelação RapidEye foram do nível 3A, ortoretificadas, com correções radiométricas, geométricas e de terreno e resolução de 5 metros para cada pixel (RAPIDEYE, 2016). As seguintes bandas espectrais compõem as imagens:

Tabela 11. Bandas espectrais dos satélites RapidEye e respectivos comprimentos de onda.

Banda	Comprimento de onda (mm)
Azul – BAZ	410 - 510
Verde – BVD	520 - 590
Vermelho – BVM	630 - 685
Red-Edge – BRE	690 - 730
Infravermelho Próximo - BIV	760 - 850

Além das cinco bandas individuais apresentadas na Tabela 11, foram utilizadas combinações entre elas, como por exemplo a subtração da Banda Vermelho da Banda Azul (BAZ – BVM), num total de 10 combinações, mais o Índice de Vegetação por Diferença Normalizada (NDVI), totalizando 16 variáveis derivadas de informações multiespectrais (Tabela 12).

Tabela 12. Variáveis derivadas dos dados multiespectrais RapidEye. Valores máximos, médio e mínimo obtidos nas três classes: INC: áreas pertencentes a cicatriz de incêndio; C/TN: áreas consideradas como “colheita” ou “terra-nua”; e P: locais com plantio de eucalipto em diferentes idades.

Variável	Mínimo			Médio			Máximo		
	INC	C/TN	P	INC	C/TN	P	INC	C/TN	P
Banda Azul - BAZ	3.362	3.591	3.136	4.059	5.233	3.754	6.795	7.980	5.102
Banda Verde - BVD	2.119	2.274	1.920	3.010	4.453	2.588	6.821	7.097	3.639
Banda Vermelho - BVM	1.057	1.575	1.018	2.066	4.226	1.410	6.553	6.785	2.349
Banda "Red Edge" - BRE	1.791	2.046	2.033	2.898	4.112	2.683	6.620	6.156	3.516
Banda InfraVermelho									
Próximo - BIV	1.753	2.014	4.046	5.444	5.056	7.449	10.093	8.012	9.858
Índice NDVI	-0,037	-0,062	0,438	0,446	0,090	0,678	0,736	0,589	0,790
Relação BAZ - BIV	-5.176	-3.482	-6.403	-1.385	177	-3.695	2.532	2.069	-295
Relação BAZ - BRE	-337	179	232	1.161	1.121	1.071	3.212	2.752	2.131
Relação BAZ - BVD	-631	-420	402	1.049	780	1.166	2.762	1.508	1.779
Relação BAZ - BVM	-520	-281	1.756	1.993	1.007	2.344	3.694	2.837	3.319
Relação BVD - BIV	-6.282	-3.998	-7.399	-2.434	-603	-4.861	972	1.012	-1.411
Relação BVD - BVM	-391	-681	353	944	227	1.178	1.921	1.893	1.950
Relação BVD - BRE	-624	-645	-891	112	341	-95	1.446	1.820	730
Relação BRE - BIV	-6.775	-3.984	-7.207	-2.546	-944	-4.766	187	174	-1.776
Relação BVM - BIV	-8.152	-5.532	-8.639	-3.378	-830	-6.039	415	759	-2.540
Relação BVM - BRE	-1.863	-1.848	-1.931	-832	114	-1.273	1.025	1.145	-641

3.5.4. Fusão das imagens entre sensores remotos

A fusão das imagens foi feita utilizando o pacote “raster” (HIJMAN; VAN ETTEN, 2019), disponível no software estatístico R (R CORE TEAM, 2019).

As métricas provenientes das nuvens LiDAR foram extraídas com a mesma resolução de pixel dos dados multiespectrais, de 5 metros, gerando rasters de banda única para cada métrica, em um total de 29 imagens. As áreas ocupadas pelos rasters com as métricas foram utilizadas para “recortar” as imagens de satélite, mantendo somente as áreas de interesse, correspondentes a cada fazenda. A fusão acontece por meio da adição das bandas geradas para cada sensor, em um raster único com 45 bandas. A Figura 5 ilustra o passo a da fusão das imagens.

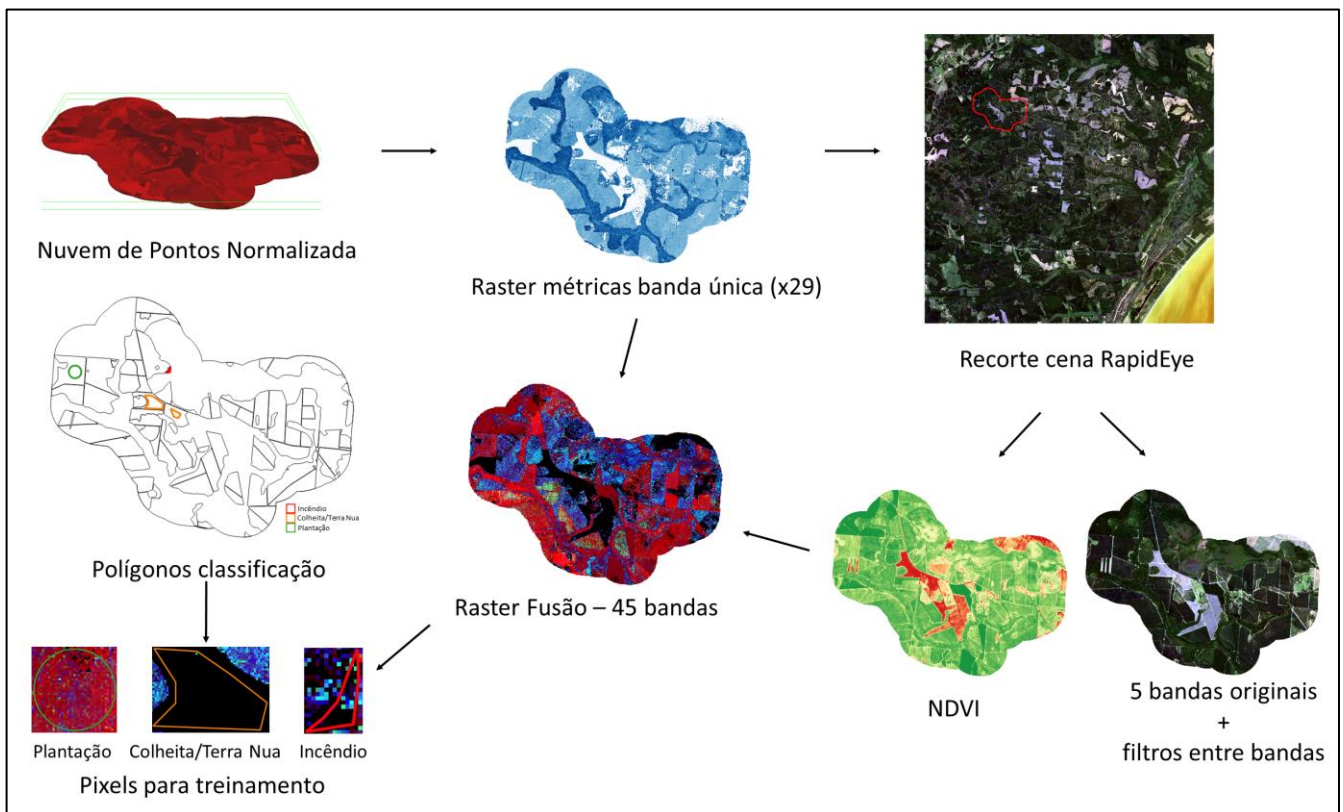


Figura 5. Sequência de procedimentos para fusão dos sensores em uma fazenda. Seleção de pixels de treinamento e geração do raster final com a fusão das imagens originadas pelas métricas LiDAR e das bandas derivadas do satélite RapidEye.

As imagens RapidEye são disponibilizadas em cenas de 25km x 25km, denominadas individualmente de “Tile”. Neste trabalho foram utilizadas 9 cenas, obtidas entre abril e julho de 2015. Para cada uma das cenas foram extraídos os demais filtros entre bandas e NDVI. Conforme pode ser observado pela Tabela 13, houve lapso temporal de até 5 meses entre os sensores utilizados.

3.5.5. Classes de treinamento

Foram definidas três classes para classificação das imagens:

- Incêndio – polígonos correspondentes às áreas de cicatrizes de incêndio;
- Colheita/Terra Nua – polígonos referentes às áreas de terra-nua (solo exposto) e/ou áreas colhidas;
- Plantação – polígonos sobre áreas de plantio, com diferentes idades.

Tabela 13. Lista de Imagens RapidEye utilizadas, datas das imagens e do sobrevoo ALS.

Tile RapidEye	Fazendas/Projetos	Data da Imagem	Data ALS
2430611	M401, M404	04/07/2015	07/09/2015
2430612	M404, M415	18/04/2015	07/09/2015
2430711	M153, M408, M413, M419, M423	04/07/2015	06/09/2015, 07/09/2015
2430712	M375, M419, M420, M423, M164, M667	04/07/2015	07/09/2015
2430812	M021, M045, M131, M375, M616, M667, T050, T665	15/06/2015	07/09/2015, 12/09/2015
2430813	M045, T706	15/06/2015	12/09/2015, 15/09/2015
2430911	M012, T604	15/06/2015	17/09/2015, 19/09/2015
2430912	T050, T604, T606, T607, T707, T665, T680, T716	15/06/2015	12/09/2015, 15/09/2015, 19/09/2015, 03/10/2015, 04/10/2015
2431012	T635, T636, T647, T649, T650, T651, T652, T655	15/06/2015	26/08/2015, 01/10/2015, 04/10/2015

A classificação das imagens é feita pela análise dos pixels extraídos de polígonos delimitados para áreas representativas de cada classe. Essa delimitação foi feita utilizando o software Quantum Gis (QGIS DEVELOPMENT TEAM, 2018). Os polígonos de cada classe foram vetorizados manualmente a partir das imagens RapidEye e locais correspondentes às fazendas.

Pelo fato de não ter sido feita uma delimitação real dos perímetros referentes às áreas queimadas, os locais classificados como incêndio foram vetorizados utilizando as imagens de satélite, a partir dos talhões com registro de fogo, sobrepostos à imagem do índice NDVI. A diferença da coloração do NDVI permitiu identificar, dentro de um talhão, a área danificada pelo incêndio.

A delimitação dos polígonos de colheita/terra nua e para plantação foi feita a partir das imagens com a composição cor verdadeira (RGB) das imagens RapidEye.

A extração dos valores dos pixels delimitados pelos polígonos classificatórios, formam o banco de dados final utilizado na modelagem. Cada observação desta base corresponde a um pixel e, cada coluna, às variáveis para a classificação. As variáveis de classificação foram separadas entre variáveis contínuas, preditoras e variável categórica, predita.

3.5.6. Classificação de imagens

A classificação das imagens para detecção de cicatrizes de incêndio é realizada em duas etapas: i) treinamento, onde os algoritmos de classificação são aplicados para a base de treino; e ii) predição, onde os resultados da modelagem são aplicados em uma outra base de dados.

Esses resultados do treinamento apontam a probabilidade do pixel em pertencer à determinada classe. Probabilidades acima de 50% determinam a classificação do pixel.

A base final foi construída com 156.690 observações, cada uma representando um pixel. Na abordagem 1, com apenas dados multiespectrais, foram 16 colunas classificadoras, e, para os dados da fusão entre sensores, 45. Para melhor eficiência computacional, foi gerada uma amostra aleatória de 30% dos dados para realização do treinamento. Nessa amostragem foi respeitada a proporção entre classes (Tabela 14).

Tabela 14. Número total de registros gerados para cada classe, número total utilizado para a modelagem, e sua proporção.

Classe	Número de Registros		Proporção por classe (%)
	Total	Amostra 30%	
Incêndio	46.940	14.082	30,0%
Colheita/Terra nua	55.503	16.651	35,4%
Plantação	54.147	16.245	34,6%
Total	156.590	46.978	100%

3.5.7. Algoritmos para classificação

Foi comparada a eficiência de dois algoritmos na classificação automática de cicatrizes de incêndio, o RF e o SVM, utilizando o pacote caret (KUHN, 2008), no software estatístico R (R CORE TEAM, 2018). A performance dos modelos foi definida após validação cruzada em pastas (*k-fold cross-validation*), com k igual a 10. Os critérios utilizados para avaliação foram: Acurácia, Coeficiente Kappa, Sensibilidade, Especificidade e Performance, calculados a partir da Matriz de Confusão.

A matriz de confusão para 3 classes possui 9 campos que representam as possíveis combinações entre classes AA, BA, CA, AB, BB, CB, AC, BC, CC, conforme pode ser observado no exemplo na Figura 6. Para cada combinação há uma interpretação distinta por classe, quanto às possíveis classificações: Verdadeiro Positivo (VP), Verdadeiro Negativo (VN), Falso Positivo (FP) e Falso Negativo (FN).

Cada classe possui 9 então classificações, sendo: 1 valor VP, 2 valores FP, 2 valores FN e 4 valores VN. Por exemplo, o valor de AA representa uma classificação VP para classe A, e VN para as classes B e C. Neste caso, $AA = VP_A = VN_{B1} = VN_{C1}$.

		Dado Observado			
		Classe A	Classe B	Classe C	Total
Dado Predito	Classe A	AA VP_A VN_{B1} VN_{C1}	AB FP_{A1} FN_{B1} VN_{C2}	AC FP_{A2} VN_{B2} FN_{C1}	P_A
	Classe B	BA FN_{A1} FP_{B1} VN_{C3}	BB VN_{A1} VP_B VN_{C4}	BC VN_{A2} FP_{B2} FN_{C2}	P_B
	Classe C	CA FN_{A2} VN_{B3} FP_{C2}	CB VN_{A3} FN_{B2} FP_{C1}	CC VN_{A4} VN_{B4} VP_C	P_C
	TOTAL	O_A	O_B	O_C	N

Figura 6. Matriz de confusão para 3 classes de treinamento, A, B e C. Nove combinações entre classes e possíveis classificações, como: Verdadeiro Positivo (VP), Verdadeiro Negativo (VN), Falso Positivo (FP) e Falso Negativo (FN), para cada classe. Dados totais preditos em P_A , P_B e P_C e totais observados em O_A , O_B e O_C . “N” se refere à soma dos nove campos da matriz.

Para interpretação das equações de cada critério de avaliação, foram utilizados os campos representados na matriz de confusão ilustrada na Figura 6.

Os valores totais foram calculados da seguinte forma:

$$P_A = AA + AB + AC \quad (17)$$

$$P_B = BA + BB + BC \quad (18)$$

$$P_C = CA + CB + CC \quad (19)$$

$$O_A = AA + BA + CA \quad (20)$$

$$O_B = AB + BB + CB \quad (21)$$

$$O_C = AC + BC + CC \quad (22)$$

$$N = P_A + P_B + P_C = O_A + O_B + O_C \quad (23)$$

A Acurácia da modelagem representa o quanto o modelo acertou em suas predições, considerando as as corretas classificações de cada classe (VP) em relação a todas obeserwações.

$$Acurácia = \frac{AA + BB + CC}{N} \quad (24)$$

O índice Kappa (COHEN, 1960) é um método estatístico que avalia o nível de concordância ou reprodutibilidade de um conjunto de dados.

$$Kappa = \frac{p_0 - p_e}{1 - p_e} \quad (25)$$

Onde:

$$p_0 = Acurácia = \frac{AA + BB + CC}{N} \quad (26)$$

$$p_e = \frac{(O_A * P_A) + (O_B * P_B) + (O_C * P_C)}{N^2} \quad (27)$$

Os índices Sensibilidade, Especificidade e Performance, são calculados individualmente para cada classe.

A Sensibilidade representa a eficiência do classificador em reconhecer as amostras de cada classe.

$$Sensibilidade_{classe} = \frac{VP_{classe}}{VP_{classe} + \sum FN_{classe}} \quad (28)$$

Já a Especificidade, representa a eficiência do classificador em reconhecer amostras que não pertence à determinada classe.

$$Especificidade_{classe} = \frac{\sum VN_{classe}}{\sum VN_{classe} + \sum FP_{classe}} \quad (29)$$

A Performance, por sua vez, avalia o percentual de acerto da classificação correta de uma determinada classe, em relação ao total predito para essa classe.

$$Performance_{classe} = \frac{VP_{classe}}{P_{classe}} \quad (30)$$

4. RESULTADOS E DISCUSSÃO

Os resultados para o cálculo de probabilidade de ocorrência de incêndio em nível de talhão são apresentados inicialmente. Logo em seguida são apresentados os resultados para o mapeamento de cicatrizes de incêndio e, ao final, o comportamento dos mapas de cicatrizes para predição de ocorrência de incêndios. Resumem-se nesta seção os aspectos que permitem identificar as melhores abordagens de modelagem para geração dos respectivos mapas.

4.1. Probabilidade de Incêndio

O ajuste do modelo preditor de probabilidade de incêndio foi feito por meio de dois métodos diferentes, Random Forest (RF) e regressão logística (LOGIT). Corroborando os dados obtidos por Hart et al. (2018), o método RF se mostrou melhor que o método LOGIT (Figura 7 e Figura 8), pois foi único capaz de identificar com eficiência as ocorrências de incêndio.

Na abordagem 1, onde foram utilizados dados de 2010 a 2017, os modelos RF 1 e 2 apresentaram o melhor desempenho. Apenas a variável referente à densidade demográfica os difere, sendo que a mesma foi retirada no segundo teste. No modelo 3, após a retirada de 6 variáveis (face de exposição do terreno, precipitação anual, temperatura média anual, umidade relativa do ar média, distância para estradas e população total dos municípios) foi identificada queda na eficiência da performance e especificidade, de 0,81 e 0,52, respectivamente, para 0,78 e 0,37.

Nesta primeira abordagem, para regressão logística, nenhum dos testes foi considerado eficiente, pois apresentaram especificidade próxima ou igual à zero, ou seja, não foram capazes de classificar corretamente as ocorrências de incêndio. Dada a similaridade das eficiências obtidas para os modelos 1 e 2 para o Random Forest e regressão logística, o segundo foi considerado o melhor, pois alcançou resultados similares utilizando uma variável a menos.

Na segunda abordagem, onde foram utilizados apenas os registros de 2015 e os mesmos testes de variáveis da abordagem anterior, os três modelos testados com o RF mostraram boa acurácia para representar a probabilidade de incêndios. Dentre eles, o modelo 3 pode ser escolhido como o melhor, pois atingiu as mesmas performances com apenas 5 variáveis: VCCC, PTA, DistUrb, DistAssent e PopZR.

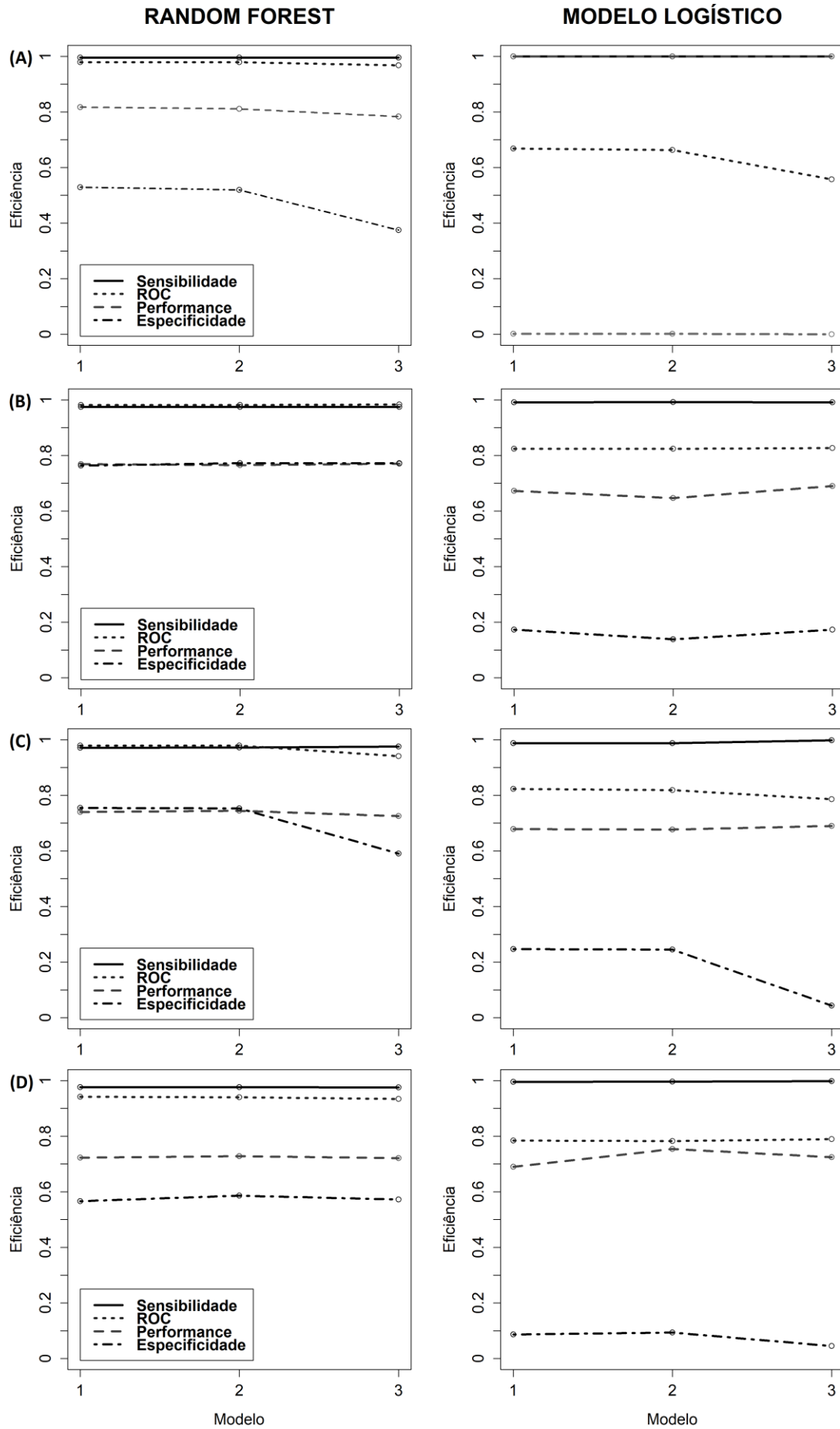


Figura 7. Eficiência do algoritmo Random Forest e do Modelo Logístico em relação aos índices Sensibilidade, ROC, Performance e Especificidade. Gráficos de A à D referem-se à abordagens 1 a 4, respectivamente.

Ainda na segunda abordagem, em relação ao modelo logístico, os índices não apresentaram “zero” respostas verdadeiras para a classificação de incêndio, mostrando evolução em relação à primeira abordagem.

Ambos os algoritmos apresentaram melhor desempenho na segunda abordagem. Esse fato pode ser explicado pela maior proporção entre as duas classes para a variável resposta, que aumentou de cerca de 4 para 10%. Considerando os melhores modelos de cada abordagem, os índices de especificidade subiram 48%, aumentando de 0,52 para 0,77 e a performance reduziu de 0,81 para 0,77 para o RF. O aumento da eficácia do modelo logístico para especificidade foi ainda maior, passando de 0,02, para 0,17. Esse algoritmo também obteve aumento na performance, para 0,69.

A abordagem 3 marcou a incorporação das métricas ALS aos dados de 2015. Tanto para RF quanto para LOGIT, os dois primeiros modelos apresentaram desempenho semelhante, mesmo após a retirada de 5 variáveis (precipitação anual, temperatura média anual, densidade demográfica, LAI e LAI_1_5m). Se mantiveram com sensibilidade e ROC próximos 0,97, especificidade em 0,75 e performance em 0,74. O terceiro modelo teve queda acentuada de especificidade para 0,59, ao retirar do modelo mais cinco variáveis, entre elas as referentes à estrutura da floresta (VCCC obtido pelo inventário e Vol_ALS, modelado por métricas LiDAR), variáveis climáticas (umidade relativa do ar) e social (população total e distância para estradas). Para a regressão logística, o mesmo comportamento entre os índices foi observado, mantendo a mesma eficiência entre os modelos 1 e 2, caindo no terceiro. Para ambos os métodos, o modelo 2 foi escolhido como o melhor nesta abordagem.

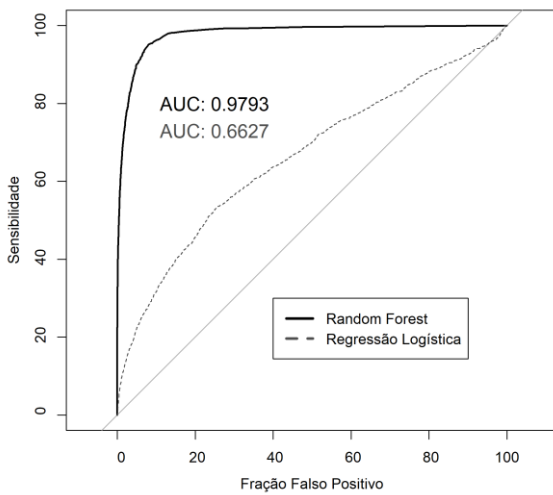
Comparando os modelos 1 dos gráficos B e C da Figura 7, quando as variáveis LiDAR foram adicionadas, os índices permaneceram estáveis para ambos os métodos, com leve oscilação. No Random Forest, a sensibilidade sem as métricas manteve-se em 0,97, a especificidade caiu de 0,76 para 0,75 e a performance passou de 0,77 para 0,74. Para o modelo logístico, a inclusão das métricas manteve a sensibilidade e performance em 0,99 e 0,7, respectivamente, porém, melhorou a especificidade em 42%, passando de 0,1735 para 0,2473, sendo este o maior valor entre todas as abordagens e modelos logísticos testados.

Na abordagem 4, foi retirado dos modelos a variável biométrica em relação à abordagem anterior, e a ausência da variável VCCC impactou ao modelo 1 queda da especificidade em 24% para o RF e 64% para LOGIT, porém, nos 3 testes os valores

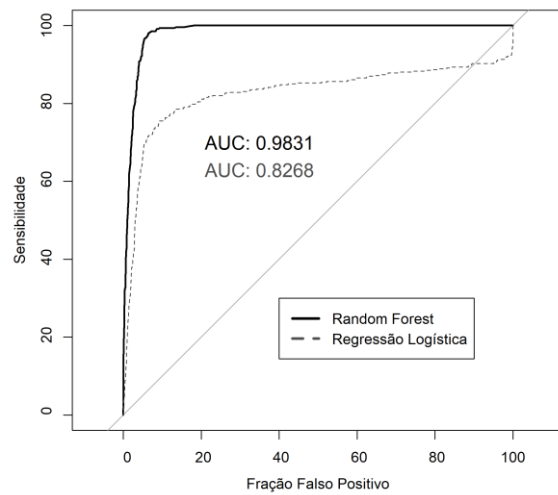
ficaram estáveis, sendo o terceiro o considerado melhor, por atingir performances semelhantes com menor número de variáveis, sendo o modelo final composto pela face de exposição do terreno, chuva total anualizada, distância para centros urbanos, distância para assentamento rural, população de zona rural e a variável estrutural do LiDAR, SubBosque.

A partir da escolha dos melhores modelos em cada abordagem e para cada método (RF e LOGIT), foram gerados os gráficos ROC (Figura 8), com o valor de AUC calculado em cada caso.

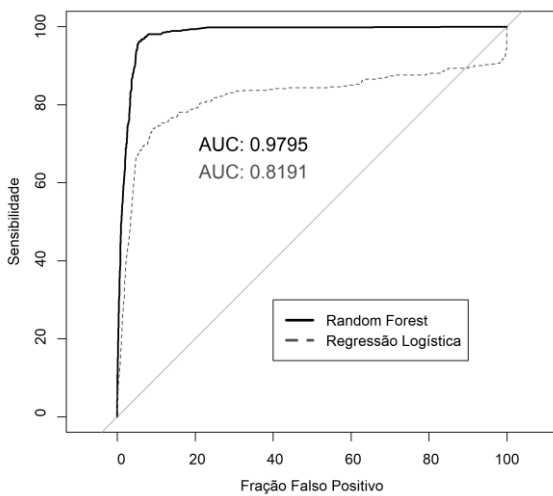
(1)



(2)



(3)



(4)

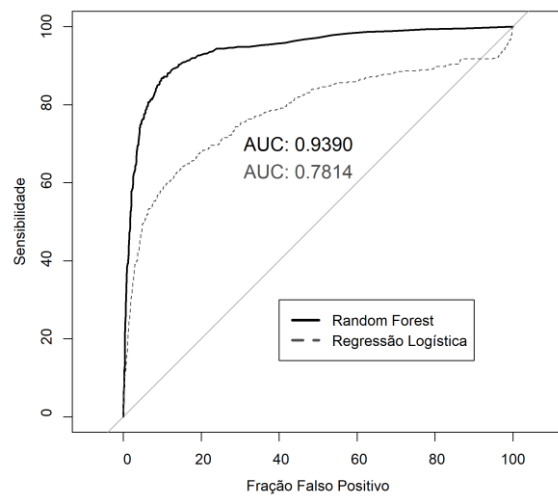


Figura 8. Curva ROC para os melhores modelos classificados nas abordagens 1 a 4.

Pelo gráfico, é possível observar limiares para a predição de incêndio bem distintos entre os modelos RF e LOGIT. Para valores maiores que 10% da fração falso positivo (FFP), já há estabilidade da sensibilidade pelo RF. Nas abordagens 2 e 3 para o modelo logístico, o valor de FFP que passa a reduzir a taxa de crescimento da sensibilidade ocorre a partir de 20% e, na abordagem 4, entre 20 e 40%. Na abordagem 1, o crescimento da curva é praticamente constante.

É possível que um modelo com baixa capacidade preditiva possua alto valor de AUC (HOSMER; LEMESHOW, 2000). Isso acontece pois o índice atribui pesos iguais para erros de classificação, enquanto, em muitas aplicações, o peso entre classes seja diferente (LOBO; JIMÉNEZ-VALVERDE; REAL, 2007), assim como observado neste trabalho, onde observou-se grande discrepância, 90% e 10%, entre valores pertencentes às classes positivo e negativo, respectivamente.

Os gráficos das Figura 7 e Figura 8 indicam que o índice sensibilidade também não foi capaz de selecionar os melhores modelos ou algoritmos, pois explicam o potencial da classificação dos dados positivos, neste caso, à não ocorrência de incêndio. Como a maior parte dos dados de treinamento pertencem a classe positivo, o índice apresentou, em todos os casos, eficiência próxima de 100%.

Desta forma, para os dados deste trabalho, os melhores índices para detectar os melhores modelos foram especificidade e performance.

Os resultados da modelagem pelo algoritmo RF permite identificar a importância de cada variável (BREIMAN, 2001; GENUER; POGGI; TULEAU-MALOT, 2010).

A variável mais importante para cálculo da probabilidade anual de ocorrência de incêndios, em todas as abordagens e modelo, foi em relação à precipitação (PTA) recebida pela floresta, desde o plantio até o ano da análise, o que faz sentido, pois há uma forte relação entre chuva e focos de incêndio, conforme corrobora o gráfico da Figura 9, onde anos mais chuvosos apresentaram menos registros.

A segunda variável mais importante nas abordagens 1 a 3 foi referente a estrutura do povoamento (VCCC), o que também é perfeitamente interpretável, pois a estrutura do povoamento reflete na disponibilidade de material combustível, um dos principais fatores relacionado aos incêndios (ANDERSEN; MCGAUGHEY; REUTEBUCH, 2005; KOETZ et al., 2008; KOUTSIAS; KARTERIS, 2003). Botequim et al., (2013) também identificou variáveis estruturais como uma das mais importantes na predição da probabilidade de incêndios.

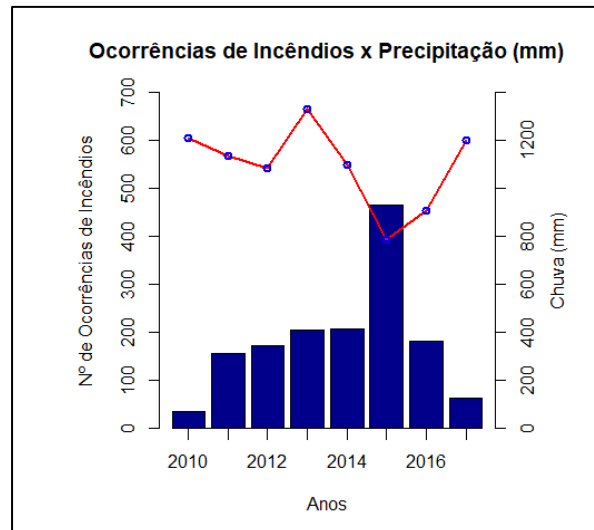


Figura 9. Número de ocorrências de incêndio por ano e precipitação média anual das estações meteorológicas onde estão inseridos os plantios. Dados de 2010 à 2017. Gráfico de barras referem-se ao número de ocorrências e, de linha, para a precipitação média anual.

As variáveis sociais mais importantes foram as de distância às áreas urbanas e aos assentamentos rurais. Resultado semelhante foi verificado por Tien Bui et al., (2018), que assim como neste trabalho, identificou esse fato como pertinente, pois a maior parte das ocorrências de incêndio no estudo foram ocasionadas por atividades humanas. Outro indício que corrobora esse resultado é a existência de estudo que demonstra conflitos sociais entre as empresas plantadoras de eucalipto, a população e movimentos sociais na região do extremo sul do estado da Bahia (SANTOS; SILVA, 2004).

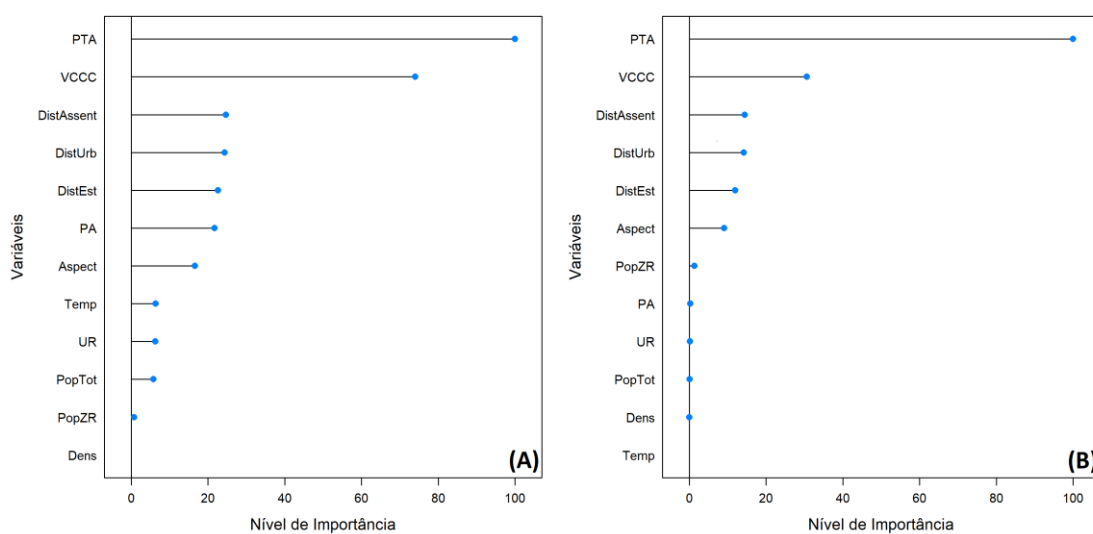


Figura 10. Nível de importância das variáveis na modelagem da probabilidade de incêndio. Modelo testado número 1 pelo algoritmo Random Forest. A) Dados de 2010 à 2017; Abordagem 1. B) Dados de 2015; Abordagem 2.

Comparando os resultados das abordagens 1 e 2, quando utilizadas as mesmas variáveis, modificando apenas o período da base de dados, a importância de variáveis verificada foi semelhante (Figura 10). A principal diferença observada está no fato de que as variáveis climáticas que representam média anual (Precipitação, Temperatura e Umidade Relativa) utilizam apenas os dados de 2015 na segunda abordagem e, por isso, não apresentam peso algum para as eficiências dos modelos. Verificou-se, ainda, que a variável Densidade Demográfica não apresenta peso para a modelagem 1.

Ao incorporar dados LiDAR ao modelo (Figura 11), foi verificado que a métrica referente ao volume foi a terceira variável mais importante, atrás apenas de PTA e VCCC. A métrica correspondente ao sub-bosque ficou atrás das variáveis sociais de distância e, na sequência apareceram as variáveis referente ao índice de área foliar (LAI).

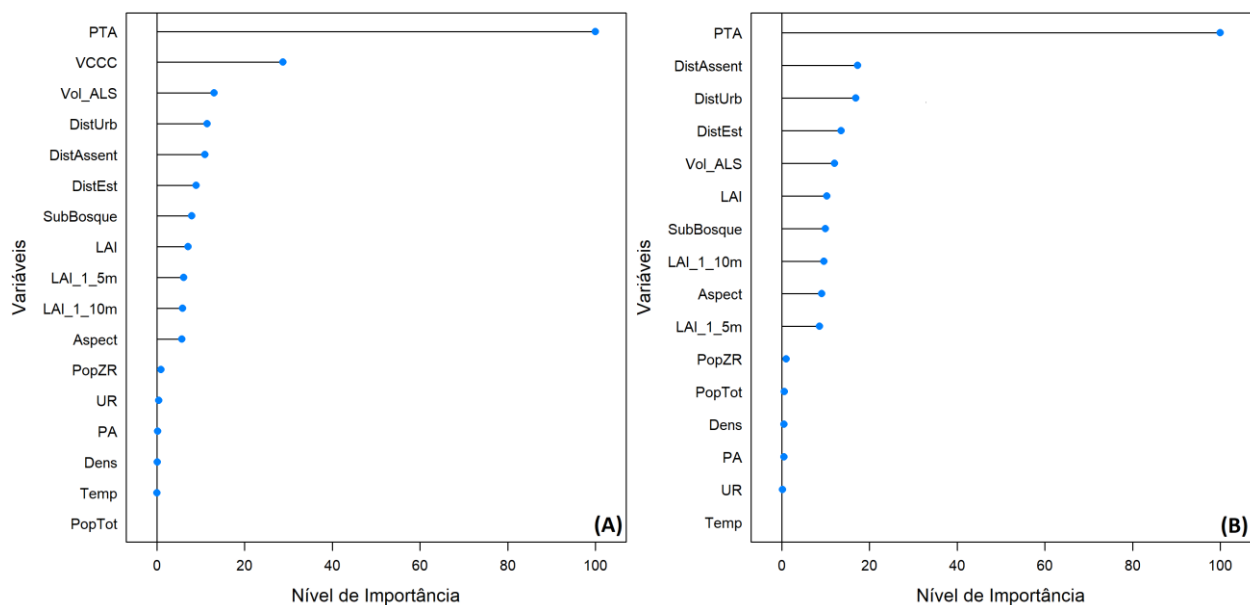


Figura 11. Nível de importância das variáveis na modelagem da probabilidade de incêndio. Modelo com todas as variáveis, algoritmo Random Forest. A) Abordagem 3. B) Abordagem 4.

No entanto, após a inclusão das métricas provenientes do ALS, não foi observada melhora na performance da modelagem pelo Random Forest. Essa informação é ratificada na abordagem 4, quando ao utilizar apenas métricas como fonte de informação de estrutura do povoamento florestal, o modelo perde eficiência e as variáveis sociais de distância passam a ser mais importantes que as variáveis de estrutura da vegetação. A variável Vol_ALS perde importância quando utilizada sozinha, comparado à abordagem onde foi utilizada junto da variável VCCC.

Considerando os conjuntos de dados e variáveis deste trabalho, a regressão logística não foi capaz de exercer um bom papel para predição de incêndio, diferente daquilo observado por Botequim, et al. (2013); Garcia-Gonzalo et al. (2012) e Marques et al. (2012), entre outros autores. O melhor modelo logístico observado, entre 12 testes, foi o segundo da abordagem 3, que integrou as métricas LiDAR. O modelo é definido pela equação 8, com as variáveis e parâmetros expressos na Tabela 15.

Pode-se observar pelos resultados do ajuste, que muitas variáveis possuem parâmetros próximos de zero, alguns com até cinco casas decimais zeradas, indicando baixa influência da respectiva variável ao modelo.

Tabela 15. Conjunto de variáveis e parâmetros para o melhor modelo logístico ajustado.

Variável	Parâmetro
Intercepto	β_0 : 3,08841182769
VCCC	β_1 : -0,01284047682
DistUrb	β_2 : -0,00007420022
DistEst	β_3 : -0,00001073199
DistAssent	β_4 : 0,00000393597
PopTot	β_5 : 0,00000414654
PopZR	β_6 : -0,00001155526
UR	β_7 : -0,11012221047
PTA	β_8 : 0,00362013847
Aspect	β_9 : 0,04808858820
LAI_1_10m	β_{10} : -2,85821984063
SubBosque	β_{11} : 0,01082673672
Vol_ALS	β_{12} : 0,00775012186

Em estudos de probabilidade de incêndio na europa, autores identificaram que a biomassa arbostiva no sub-bosque representa uma variável extremamente importante (CASTRO et al., 2003; GODINHO-FERREIRA et al., 2006; RICARDO 2010; BOTEQUIM et al., 2013). Neste trabalho essa variável foi representada pelas métricas SubBosque e pelo “índice de área foliar efetivo” LAI de 1 a 5 metros, porém, ambas apresentaram baixa importância para o modelo. Essa diferença entre a realidade europeia e brasileira pode ser explicada pela ausência de sub-bosque em florestas comerciais de eucalipto no Brasil, onde as empresas mantêm manejo constante e controle de mato-competição no sub-bosque.

A partir da probabilidade identificada para cada talhão e ano, foi possível construir mapas anuais de probabilidade de ocorrência de incêndio, que podem ser

extremamente úteis para gestão florestal, na identificação dos locais onde devem alocar recursos de monitoramento e combate a incêndios.

A Figura 12 ilustra, para 4 anos diferentes, as probabilidades de incêndio de uma determinada região da área de estudo. As probabilidades foram divididas em cinco categorias: Muito Baixa, Baixa, Moderada, Alta e Muito Alta.

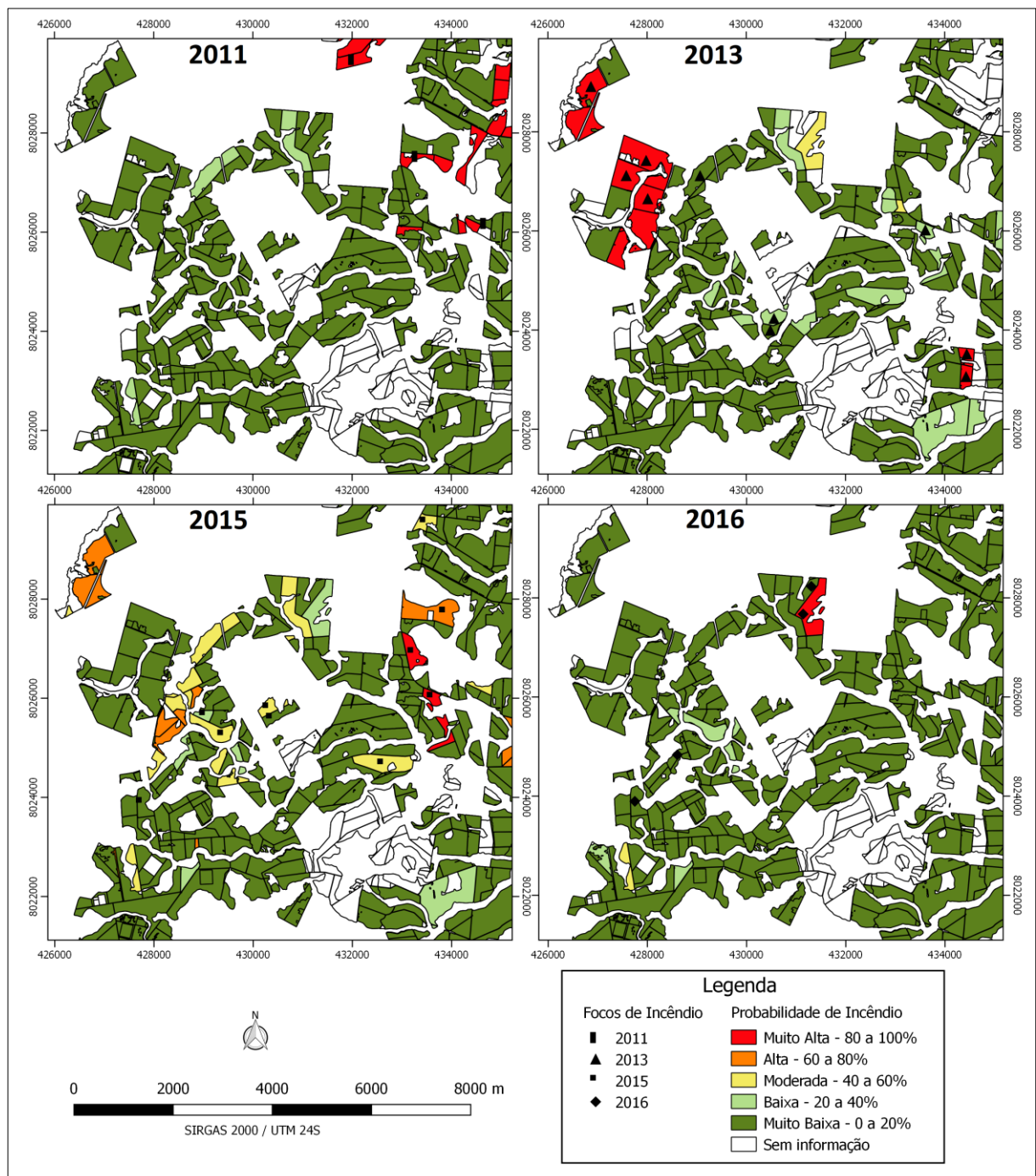


Figura 12. Mapa de probabilidade de incêndio em nível de talhão para os anos de 2011, 2013, 2015 e 2016 e identificação dos locais onde houve registro de incêndio para cada ano.

4.2. Mapeamento de cicatrizes de incêndio

A geração de mapas raster de cicatrizes que identificam áreas com cobertura florestal danificada foi feita pela classificação dos algoritmos RF e SVM. Foi comparada a eficiência da classificação utilizando apenas dados multiespectrais com a fusão entre sensores multiespectrais e LiDAR (Figura 13).

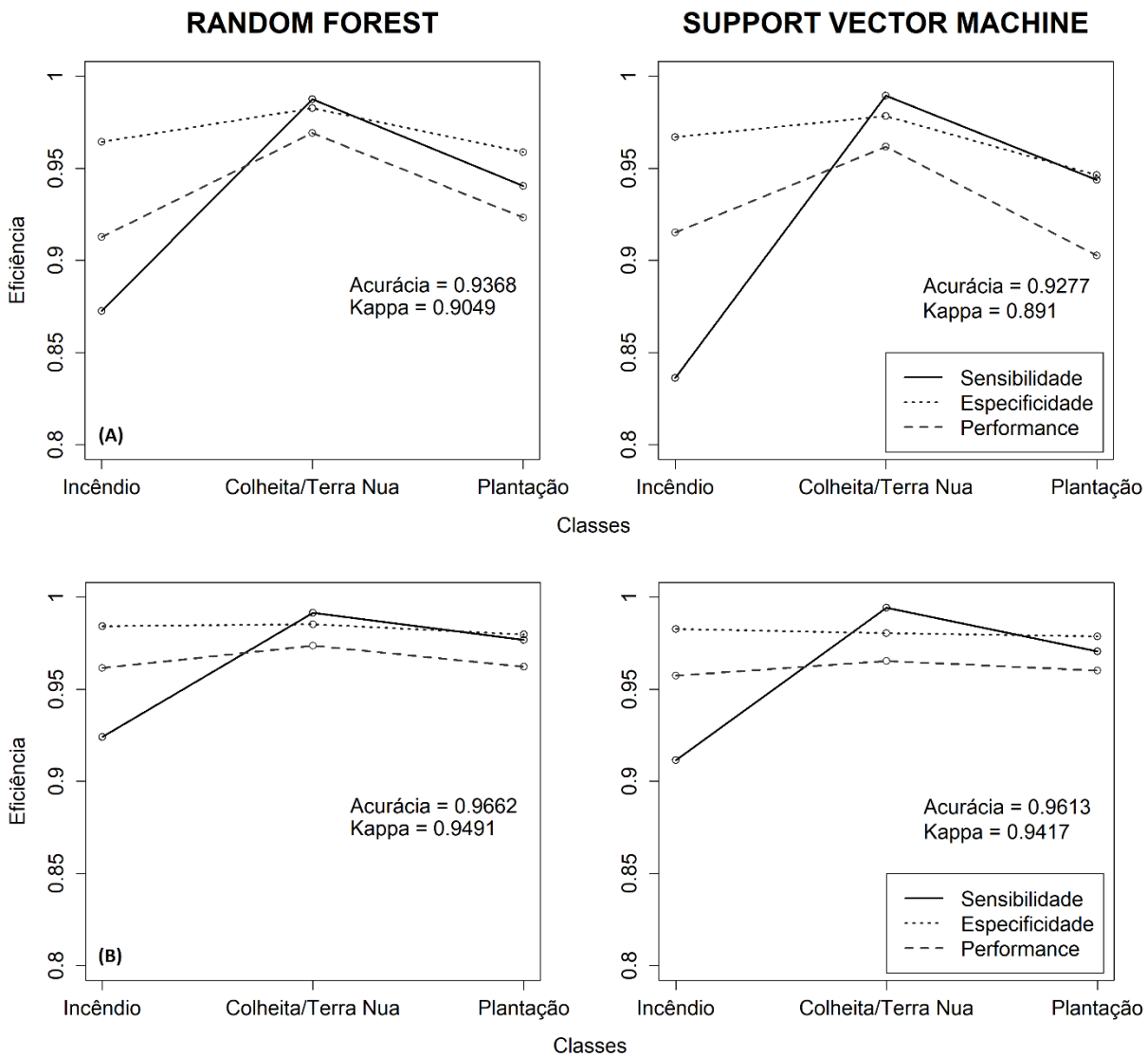


Figura 13. Modelagem a partir de dados multiespectrais (A) e com dados da fusão entre sensores multiespectrais e LiDAR (B). Eficiência dos modelos quanto à Acurácia e índice Kappa, e à Sensibilidade, Especificidade e Performance, por classe classificatória.

Utilizando apenas dados multiespectrais, a acurácia geral da classificação foi de 0,937 para o algoritmo Random Forest e 0,928 para o Support Vector Machine,

com Kappa de 0,904 e 0,891, respectivamente, para cada modelo. Após incorporar os dados LiDAR na modelagem, os índices melhoraram. A acurácia geral chegou a 0,966 para RF e 0,961 para SVM, com o coeficiente kappa de 0,949 e 0,942, respectivamente.

Analisando os gráficos da Figura 13, é possível observar que as linhas representando a eficiência de cada índice, apresentaram tendência semelhante entre as três classes e os dois algoritmos. Sensibilidade, especificidade e performance aumentam para a classe colheita/terra nua em relação à classe incêndio e depois apresentaram ligeira queda para a classe plantação.

Foi identificada maior eficiência dos índices na classe correspondente a áreas de colheita/terra nua segunda classe, onde observou sensibilidade média de 99% em todas abordagens e algoritmos. Esses resultados acontecem, pois solo exposto possui característica espectral peculiar (BELLINASSO; DEMATTÊ; ROMEIRO, 2010), que o diferenciam nitidamente em relação aos plantios e áreas incendiadas. Dependendo da estação do ano e idade da floresta, o vigor vegetativo se altera, impactando na resposta espectral para o plantio em determinado momento (JENSEN, 2000). Quanto às áreas queimadas, dependendo da intensidade do fogo e da situação da floresta, pode-se esperar diferentes respostas espectrais após o sinistro.

A sensibilidade para classificação de incêndio foi de 0,873 para RF e 0,836 para SVM quando modelados usando apenas os dados multiespectrais. Embora esses sejam os menores valores observados entre índices e classes, os mesmos apresentam eficiência superior a 83% para ambos algoritmos. Ao incorporar os dados LiDAR na análise, a sensibilidade aumentou para 0,924 (RF) e 0,911 (SVM).

Para ambas as abordagens (com e sem LiDAR) e ambos algoritmos, 30% dos dados de treinamento se referem a classe de incêndio. Utilizando apenas dados do sensor RapidEye, os algoritmos RF e SVM classificaram corretamente como incêndio 26,2% e 25,1%, respectivamente. Com a fusão dos dados ALS, passaram a classificar corretamente 27,7% e 27,3%.

Para todas as classes e algoritmos, foi observada especificidade entre 95% e 98%, indicando que independentemente das abordagens, os modelos foram eficientes em reconhecer quando uma amostra não pertence à determinada classe.

A fusão entre sensores também aprimorou o índice “performance” para as classes incêndio e plantação. Utilizando apenas o sensor espectral, coletando dados em duas dimensões (2D), os algoritmos obtiveram acuidade superior entre 90%

(classes 1 e 3) e 97% (classe 2). Para os dados fusionados, incorporando a análise na terceira dimensão (3D), o índice performance aumentou sua eficiência para 96% nas classes 1 e 3, identificando que os dados LiDAR permitiram um melhor reconhecimento de padrões, principalmente para as classes que envolvem elementos em três dimensões como no caso das áreas queimadas e de plantio, mantendo-se na média de 97% para a classe 2. No entanto, a classificação apenas com dados multiespectrais foi eficiente e pode ser utilizada no caso da impossibilidade de escanemanto com sensor a laser.

Esses resultados corroboram o potencial do uso de algoritmos de Machine Learning como ferramentas para mapeamento de áreas florestais perturbadas (HISLOP et al., 2019; LEUENBERGER et al., 2018; PEREIRA et al., 2017), e o potencial de fusão dos dados entre sensores satelitais e LiDAR para aprimorar as performances da tarefa pretendida (MACHALA; ZEJDOVÁ, 2014; MUTLU et al., 2008; XU; MANLEY; MORGENROTH, 2018).

Cada técnica utilizada para classificação dos mapas de cicatrizes, permite identificar a ordem de importância de variáveis para o modelo final. As variáveis mais importantes para o Random Forest são apresentadas por meio de um índice único para as três classes preditoras (Figura 14). Para o SVM há, para cada classe, um ranking específico para importância de variável (Figura 15).

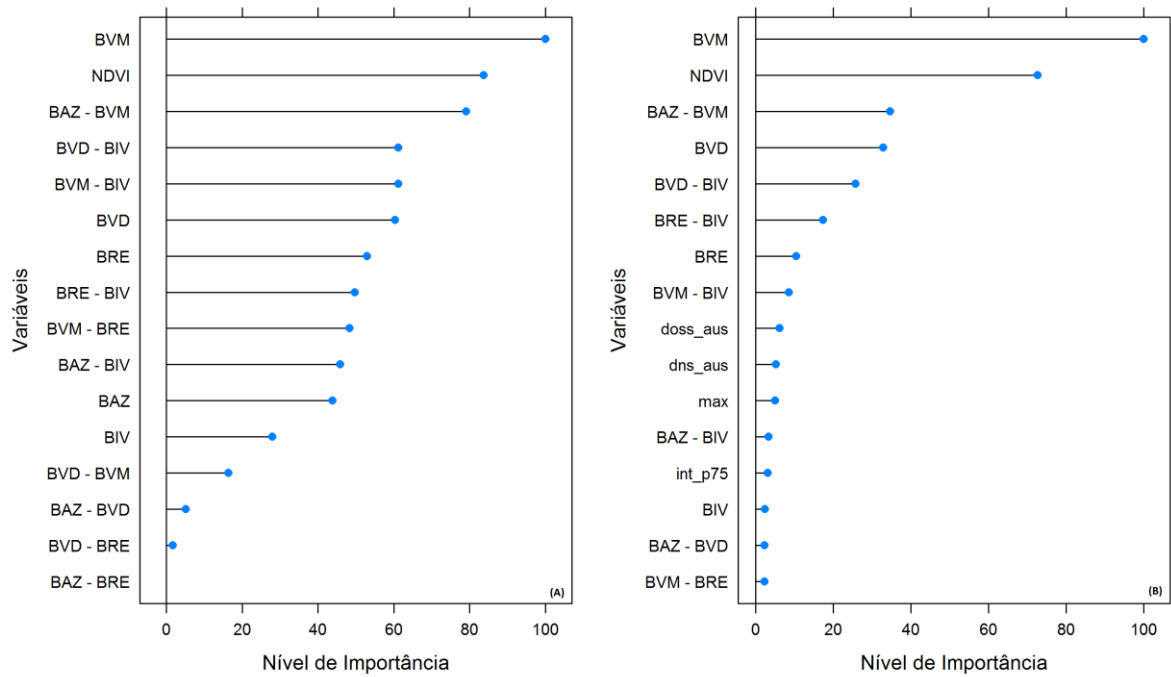


Figura 14. Importância de variáveis para a classificação de áreas danificadas, sugerida pelo algoritmo RF. À esquerda, utilizando apenas dados multiespectrais. À direita, dados da fusão entre sensores.

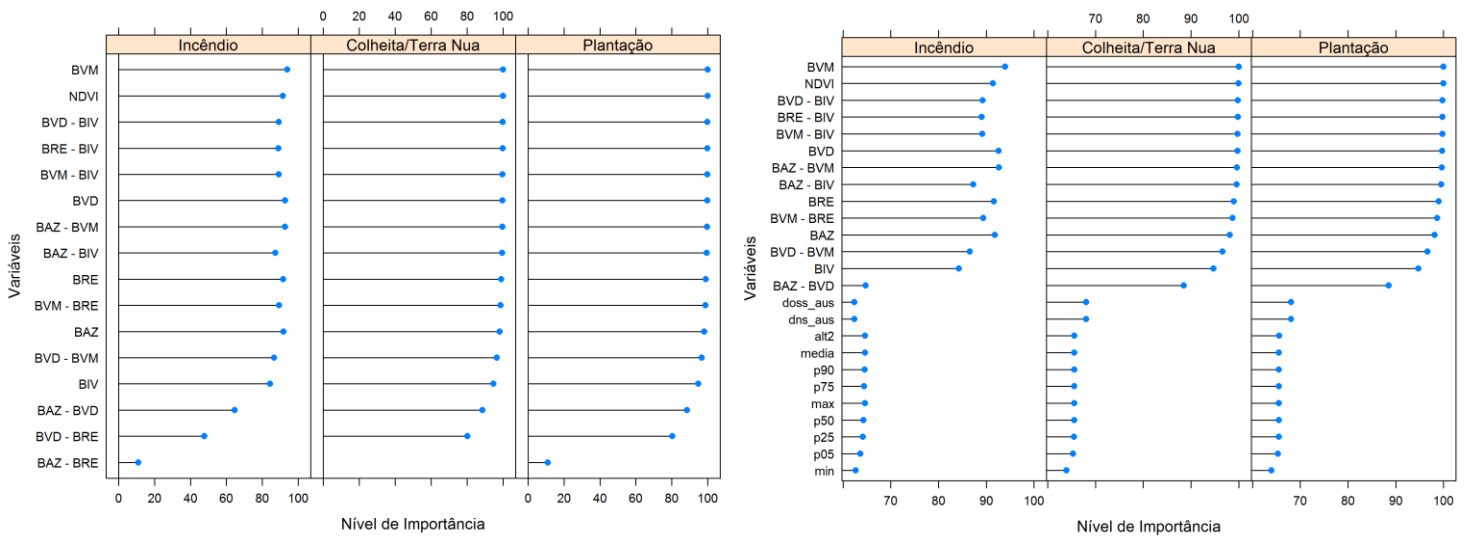


Figura 15. Importância de variáveis para a classificação de áreas danificadas, sugerida pelo algoritmo SVM, para cada classe. À esquerda, utilizando apenas dados multiespectrais. À direita, dados da fusão entre sensores.

A variável correspondente à banda Vermelho (BVM) foi a mais importante para todos os algoritmos e classes. O NDVI foi o segundo em importância para o algoritmo RF e às classes Colheita/Terra Nua e Plantação do SVM. Para a classe Incêndio no Support Vector Machine, o NDVI aparece na sexta colocação na ordem de importância, atrás das bandas individuais verde (BVD), azul (BAZ) e “red-edge” (BRE).

Resultado semelhante, onde uma banda única obteve maior importância em relação às combinações entre bandas e índices de vegetação foi observado nos trabalhos de Pereira e Setzer (1993) e Chen et al., (2014), porém, nestes casos, a Banda do Infravermelho próximo (BIV) que obteve melhor acurácia para o mapeamento de cicatrizes de incêndio, junto com uma banda termal.

Chen et al. (2014) encontrou ainda boa acurácia para as bandas vermelho (80%) e azul (70%). A combinação dessas bandas (BAZ – BVM) foi a terceira variável mais importante utilizada pelo Random Forest e a segunda para identificar pixels pertencentes às áreas queimadas utilizando o SVM. A relação entre as bandas azul e vermelho apresentou alta correlação com o índice de vegetação NDVI (Figura 24 e Figura 25, ANEXO C – MODELAGEM CICATRIZ DE INCÊNDIO). Este foi um resultado inédito, aliás, são poucos os trabalhos encontrados na literatura que compararam bandas únicas à índices de vegetação e/ou às combinações entre bandas para gerar mapas de cicatrizes

As Figura 14 e Figura 15 apresentam algumas das relações que foram importantes aos modelos, como por exemplo BVD e BRE utilizadas individualmente e as relações BVD – BIV e BRE – BIV. Por outro lado, o NDVI, identificado neste trabalho como uma variável importante no mapeamento de áreas danificadas, já foi amplamente testada e comprovada como eficaz na identificação de áreas queimadas (CHEN et al., 2014; ESCUIN; NAVARRO; FERNÁNDEZ, 2008; HISLOP et al., 2019; HUDAK et al., 2007; LEWIS et al., 2009).

As variáveis LiDAR mais importantes para a classificação de cicatrizes de incêndio, foram referentes à métrica cobertura de dossel, representado pelas variáveis “Cobertura de dossel - ausência”, “Densidade de Dossel – ausência”. Essa métrica também se demonstrou eficiente na diferenciação entre plantios queimados e não queimados, no trabalho de Hoffman et al. (2018). O número máximo de retornos foi outra métrica importante utilizada por ambos algoritmos.

A lista completa com a importância das variáveis pode ser encontrada no ANEXO C – MODELAGEM CICATRIZ DE INCÊNDIO.

O resultado final da modelagem dos mapas de cicatriz de danos atribui, para cada pixel, um valor que representa a probabilidade daquela célula em pertencer a uma determinada classe. Probabilidades acima de 50% definem a classificação e, baseado nesse limiar, é que são representadas as classificações apresentadas pela matriz de confusão e que geraram os índices de avaliação das eficiências dos

modelos. No entanto, o usuário final do modelo pode criar mapas com limiares de seu interesse, de maneira que seja garantindo a classificação pretendida para o determinado fim.

A Figura 16 traz três exemplos de mapas gerados com a classificação da probabilidade de cada pixel em fazer parte de uma cicatriz de incêndio, considerando um limiar de 95% para a classificação. É possível notar pelos mapas, que os locais onde houve registro de ocorrência de incêndio (representado por um ponto no mapa), os algoritmos foram capazes de classificar se houve ou não danos às florestas, pela coloração dos pixels. Em vermelho os locais onde houve danos e, em coloração clara, locais onde não houve danos e, provavelmente, o fogo tenha sido de baixa intensidade.

As informações de ocorrência de incêndio foram apresentadas pela empresa que forneceu os dados. Os registros apenas indicaram o talhão onde ocorreram o sinistro, mas não apresentaram o grau de intensidade ou o perímetro atingido, para que fosse possível uma análise mais segura de cada caso e a possibilidade de relacionar a classificação dos mapas à severidade dos incêndios.

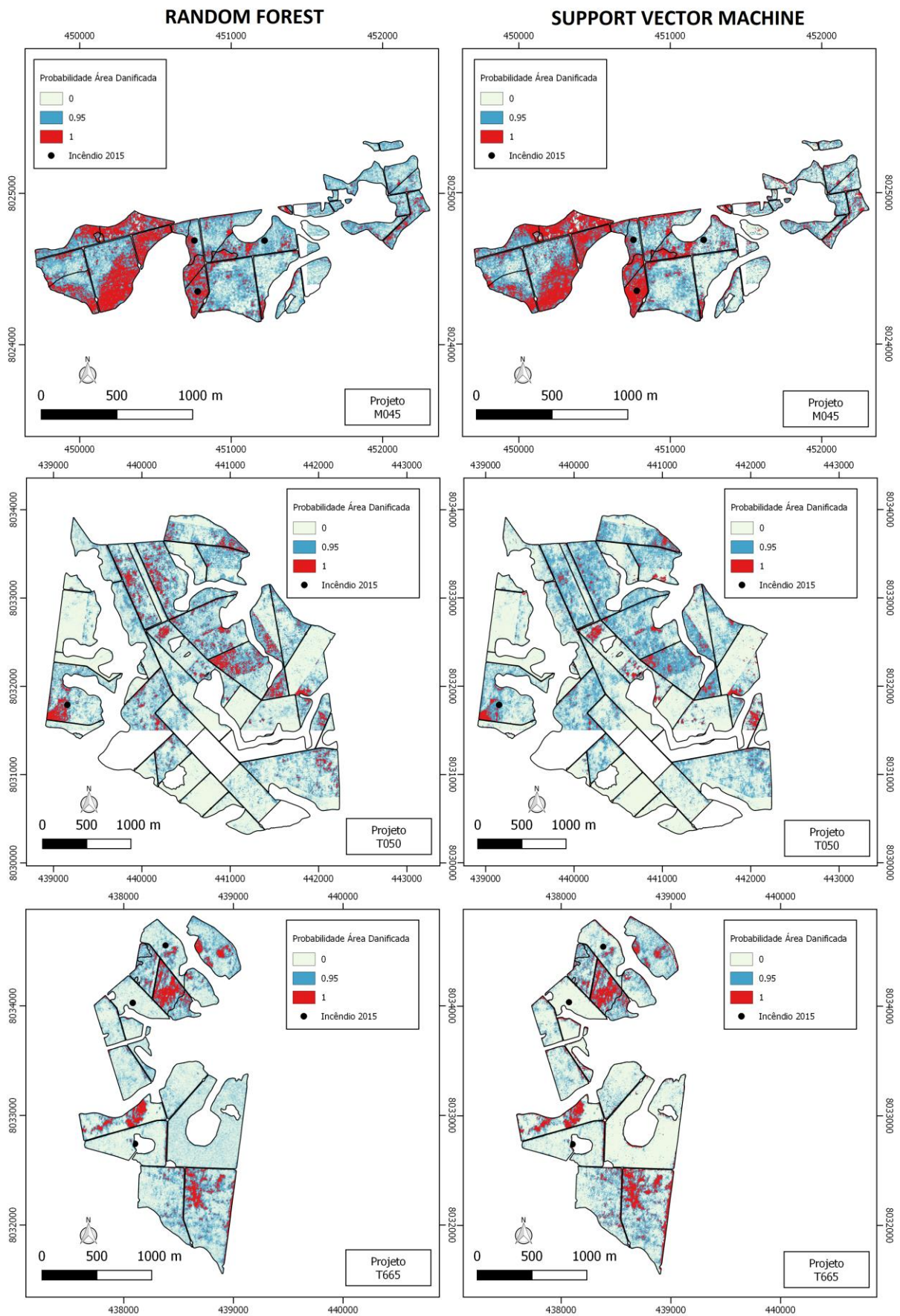


Figura 16. Mapas de Danos Florestal para os projetos (fazendas) M045, T050 e T665, gerados pelos algoritmos Random Forest e Support Vector Machine ao classificar dados provenientes da fusão dos sensores RapidEye e LiDAR. Pixels com probabilidade acima de 95% destacados em vermelho. Talhões com registro de incêndio no ano de 2015, anterior a data das imagens RapidEye, foram identificados por um ponto.

Os mapas permitem identificar que a presença de áreas danificadas não ocasionadas por incêndio que, em alguns casos, foram mais representativas em relação às cicatrizes que são de fato provenientes de áreas queimadas, conforme ilustra o gráfico de caixas da Figura 17.

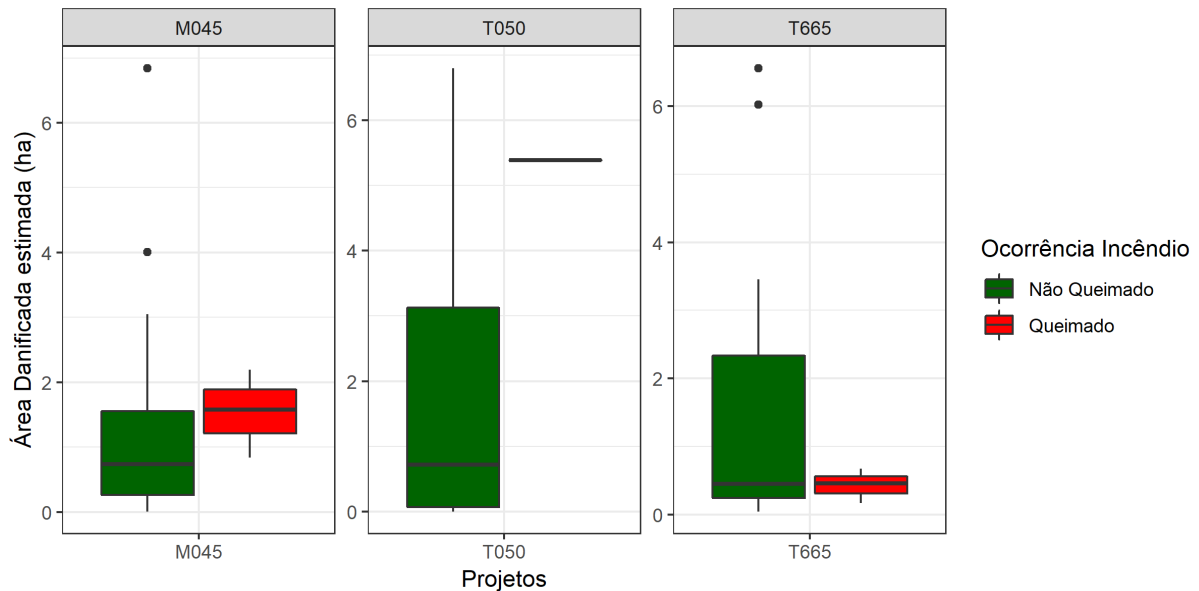


Figura 17. Proporção de áreas danificadas para os projetos (fazendas) M045, T050 e T665, em talhões não queimados e queimados.

Na Figura 17, as caixas em cor verde representam as áreas danificadas por talhão sem ocorrências de fogo, e são comparadas com as caixas em vermelho, sobre talhões onde houve registro de incêndio. É natural que a amplitude das caixas para talhões não queimados seja maior, pois, de maneira geral, são muito mais numerosos os casos de áreas não sinistradas, por fazenda.

Na fazenda M045 foram encontrados talhões com 4 e 6 hectares de áreas danificadas em locais sem registro de incêndio, além de um grupo de talhões com danos variando entre 0,5 e 3 hectares. Para este local, os talhões queimados apresentaram danos entre 0,8 e 2,2 hectares. No projeto T050, por exemplo, a própria empresa rotulou alguns talhões como “Plantio com Distúrbio” e os modelos gerados foram capazes de detectar esses locais. O gráfico aponta danos de até 6,8 hectares para um dos talhões não queimados e, em um único local com registro de fogo, área afetada de 5,4 hectares. Na fazenda T665 também foi observado maior quantidade de áreas danificadas para locais sem registro de incêndio.

Diante do fato em que os mapas gerados foram eficientes nas identificações das reais cicatrizes de incêndio, no entanto, também se mostraram bastante sensíveis ao classificar cicatrizes de florestas danificadas por outros agentes (pragas, vendaval, etc.), falhas de plantio ou qualquer outro tipo de distúrbio, uma nomenclatura mais adequada ao mapa gerado seria “Mapa de Perturbação Florestal”.

Foi identificado ainda casos com a classificação de pixels danificados para itens não ligados à plantios, como por exemplo para estradas e nuvens. Essas classificações são consideradas como “Falso Negativo” e podem ser corrigidas criando mais classes para treinamento dos modelos.

Neste trabalho, os polígonos utilizados para treinamento da classe “Incêndio”, foram vetorizados a partir de rasters de índice de vegetação referentes à talhões especificados pela empresa com ocorrência de incêndio. A utilização de polígonos representando áreas reais queimadas aferidas em campo, poderia também minimizar erros de classificação.

Outro fator que pode causar erros de classificação utilizando os dados fusionados é a diferença temporal entre a aquisição dos dados dos sensores. Neste trabalho houve até cinco meses de diferença entre a idade da imagem de satélite e do voo com LiDAR. Ou seja, pode haver casos de incêndios ocorridos entre a passagem dos sensores, além da possibilidade de recuperação da vegetação após a queimada, alterando a resolução espectral ou a estrutura da floresta após o evento.

5. CONCLUSÃO

Este trabalho propôs o desenvolvimento de metodologias para predição da probabilidade de incêndios florestais em nível de talhão para plantios de eucalipto e para a geração de mapas de cicatrizes com cobertura florestal danificada com incêndio. As metodologias desenvolvidas utilizando técnicas de aprendizado de máquinas apresentaram boa acurácia, tanto para a probabilidade anual da ocorrência de incêndio, quanto para os mapas com a localização das cicatrizes. O algoritmo Random Forest se mostrou a ferramenta mais eficiente em ambos os modelos.

Como objetivo específico, este estudo procurou identificar as variáveis mais importantes utilizadas em cada uma das modelagens. As variáveis relacionadas à quantidade de chuva recebida pelos plantios, ao volume comercial do talhão e às distâncias para áreas de assentamento rural e áreas urbanas foram as mais importantes para predição da probabilidade anual de incêndio. Na classificação dos mapas de cicatrizes de danos, as variáveis mais eficientes foram referentes ao uso individual da banda vermelho e ao índice de vegetação NDVI. A combinação das bandas RapidEye azul e vermelho também se destacou para este tipo de classificação, sendo este um resultado que não encontra similar na literatura.

Em relação as perguntadas levantadas, a adição de dados ALS às variáveis biométricas, climáticas e sociais, não aprimoraram os modelos de cálculo de probabilidade de incêndio utilizando essa base de dados, no entanto, a inclusão de métricas LiDAR no conjunto de variáveis que expressam sinais espectrais de imagens RapidEye melhoraram a capacidade preditiva do modelo de localização de cicatrizes de danos na cobertura florestal.

Os estudos e modelos gerados neste trabalho podem certamente apoiar a capacidade gestora e o processo de tomada de decisão de equipes responsáveis pelo manejo e monitoramento de plantios florestais. A identificação da probabilidade de ocorrência de incêndios permite melhor planejamento para a estruturação das equipes de combate a incêndio, priorizando locais de maior risco, podendo, inclusive, reduzir custos na operação devido a ganhos de eficiência logística.

As metodologias apresentadas podem ser replicadas a outros locais, inclusive para áreas de vegetação nativa, desde que se disponha de dados suficientes para compor as variáveis preditoras de cada modelo.

REFERÊNCIAS

- ALMEIDA, D. R. A.; STARK, S. C.; SHAO, G.; SCHIETTI, J.; NELSON, B. W.; SILVA, C. A.; GORGENS, E. B.; VALBUENA, R.; PAPA, D. DE A.; BRANCALION, P. H. S. Optimizing the remote detection of tropical rainforest structure with airborne lidar: Leaf area profile sensitivity to pulse density and spatial sampling. *Remote Sensing*, v. 11, n. 1, p. 15, 2019.
- ALMEIDA, T. M. DE; MOREAU, A. M. S. DOS S.; MOREAU, M. S.; PIRES, M. DE M.; FONTES, E. DE O.; GÓES, L. M. Reorganização socioeconômica no extremo sul da Bahia decorrente da introdução da cultura do eucalipto. *Sociedade & Natureza*, v. 20, n. 2, p. 5–18, 2008.
- ALPAYDIN, E. *Machine Learning The new AI*. MIT Press ed. MIT Press, Cambridge, Massachusetts. 2016.
- ALVARES, C. A.; STAPE, J. L.; SENTELHAS, P. C.; DE MORAES GONÇALVES, J. L.; SPAROVEK, G. Köppen's climate classification map for Brazil. *Meteorologische Zeitschrift*, v. 22, n. 6, p. 711–728, 1 dez. 2013.
- ANDERBERG, M. R. (MICHAEL R. Cluster analysis for applications. Academic Press. 1973.
- ANDERSEN, H. E.; MCGAUGHEY, R. J.; REUTEBUCH, S. E. Estimating forest canopy fuel parameters using LIDAR data. *Remote Sensing of Environment*, v. 94, n. 4, p. 441–449, 2005.
- ANDRADE, E. N. *Manual do plantador de eucalyptus*. São Paulo: Typographia Brazil, 1911. 338p.
- ARROYO, L. A.; PASCUAL, C.; MANZANERA, J. A. Fire models and methods to map fuel types: The role of remote sensing. 2008.
- ASPRS, A. S. FOR P. & R. S. American Society for Photogrammetry & Remote Sensing. Bethesda, Maryland: [s.n.]. Disponível em: <https://www.asprs.org/wp-content/uploads/2010/12/LAS_1_4_r13.pdf>. Acesso em: 2 abr. 2019.
- BALDI, G.; NOSETTO, M.; ARAGÓN, R.; AVERSA, F.; PARUELO, J.; JOBBÁGY, E. Long-term Satellite NDVI Data Sets: Evaluating Their Ability to Detect Ecosystem Functional Changes in South America. *Sensors*, v. 8, n. 9, p. 5397–5425, 3 set. 2008.
- BANNARI, A.; MORIN, D.; BONN, F.; HUETE, A. R. A review of vegetation indices. *Remote Sensing Reviews*, v. 13, n. 1–2, p. 95–120, 1995.

BAR MASSADA, A.; SYPHARD, A. D.; STEWART, S. I.; RADELOFF, V. C. 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, 2012.

BELLINASO, H.; DEMATTÊ, J. A. M.; ROMEIRO, S. A. Soil spectral library and its use in soil classification. *Revista Brasileira de Ciência do Solo*, v. 34, n. 3, p. 861–870, 2010.

BIRCH, D. S.; MORGAN, P.; KOLDEN, C. A.; ABATZOGLOU, J. T.; DILLON, G. K.; HUDAK, A. T.; SMITH, A. M. S. Vegetation, topography and daily weather influenced burn severity in central Idaho and western Montana forests. *Ecosphere*, v. 6, n. 1, 2015.

BOND, WILLIAM, J. Large parts of the world are brown or black: A different view on the ' Green World ' hypothesis. *Journal of Vegetation Science*, v. 16, n. 3, p. 261–266, 2005.

BOTEQUIM, B.; GARCIA-GONZALO, J.; MARQUES, S.; RICARDO, A.; BORGES, J. G.; TOMÉ, M.; OLIVEIRA, M. M. Developing wildfire risk probability models for Eucalyptus globulus stands in Portugal. *IForest*, v. 6, n. 4, p. 217–227, 2013.

BRANDO, P. M.; BALCH, J. K.; NEPSTAD, D. C.; MORTON, D. C.; PUTZ, F. E.; COE, M. T.; SILVÉRIO, D.; MACEDO, M. N.; DAVIDSON, E. A.; NÓBREGA, C. C.; ALENCAR, A.; SOARES-FILHO, B. S. Abrupt increases in Amazonian tree mortality due to drought-fire interactions. *Proceedings of the National Academy of Sciences of the United States of America*, v. 111, n. 17, p. 6347–52, 2014.

BRASIL. Ministério da Integração Nacional. Instrução Normativa nº 1, de 24 de Agosto de 2012. *Diário Oficial da União*, Seção 1, nº169, p. 30, 30 ago. 2012. Disponível em: <<http://pesquisa.in.gov.br/imprensa/jsp/visualiza/index.jsp?data=30/08/2012&jornal=1&pagina=30&totalArquivos=120>>. Acesso em 31/03/2019.

BREIMAN, L. Random Forests. *Machine Learning*, v. 45, n. 1, p. 5–32, 2001.

BROWN, A.A.; DAVIS, K.P. **Forest Fire**: Control and use. 2.ed. New York: McGraw Hill, 1973. 686 p.

CHEN, W.; MORIYA, K.; SAKAI, T.; KOYAMA, L.; CAO, C. X. Mapping a burned forest area from Landsat TM data by multiple methods. *Geomatics, Natural Hazards and Risk*, v. 7, n. 1, p. 384–402, 2014.

CHUVIECO, E.; AGUADO, I.; YEBRA, M.; NIETO, H.; SALAS, J.; MARTÍN, M. P.; VILAR, L.; MARTÍNEZ, J.; MARTÍN, S.; IBARRA, P.; DE LA RIVA, J.; BAEZA, J.;

RODRÍGUEZ, F.; MOLINA, J. R.; HERRERA, M. A.; ZAMORA, R. Development of a framework for fire risk assessment using remote sensing and geographic information system technologies. *Ecological Modelling*, v. 221, n. 1, p. 46–58, 2010.

CHUVIECO, E.; CONGALTON, R. G. Mapping and inventory of forest fires from digital processing of tm data. *Geocarto International*, v. 3, n. 4, p. 41–53, 1988.

CHUVIECO, E.; LIZUNDIA-LOIOLA, J.; LUCRECIA PETTINARI, M.; RAMO, R.; PADILLA, M.; TANSEY, K.; MOUILLOT, F.; LAURENT, P.; STORM, T.; HEIL, A.; PLUMMER, S. Generation and analysis of a new global burned area product based on MODIS 250 m reflectance bands and thermal anomalies. *Earth System Science Data*, v. 10, n. 4, p. 2015–2031, 22 maio. 2018.

CHUVIECO, E.; YUE, C.; HEIL, A.; MOUILLOT, F.; ALONSO-CANAS, I.; PADILLA, M.; PEREIRA, J. M.; OOM, D.; TANSEY, K. A new global burned area product for climate assessment of fire impacts. *Global Ecology and Biogeography*, v. 25, n. 5, p. 619–629, maio. 2016.

CLARKE, K. C. *Advances in Geographic Information Systems. Computers, Environment and Urban Systems*, v. 10, n. 3–4, p. 175–184, 1986.

COOPS, N. C.; HILKER, T.; WULDER, M. A.; ST-ONGE, B.; NEWNHAM, G.; SIGGINS, A.; TROFYMOW, J. A. (TONY). Estimating canopy structure of Douglas-fir forest stands from discrete-return LiDAR. *Trees*, v. 21, n. 3, p. 295–310, 2 maio. 2007.

CORTEZ, P.; MORAIS, A. A Data Mining Approach to Predict Forest Fires using Meteorological Data. IN: PROCEEDINGS OF THE 13TH PORTUGUESE CONFERENCE ON ARTIFICIAL INTELLIGENCE. Anais... Guimarães, Portugal: 2007. p. 512–523. Disponível em: <<http://www.dsi.uminho.pt/~pcortez>>. Acesso em: 30/03/2019.

DATASUS – Departamento de Informática do SUS. Informação de Saúde (TABNET). Demográficas e Socioeconômicas. Disponível em: <<http://datasus.saude.gov.br/informacoes-de-saude/tabnet/demograficas-e-socioeconomicas>>. Acesso em: 27/10/2018. 2018.

DNIT – Departamento Nacional de Infraestrutura de Transportes. Shapefiles. Disponível em: <<http://www.dnit.gov.br/mapas-multimodais/shapefiles>>. Acesso em: 27/10/2018. 2018

ELACHI, C.; VAN ZYL, J. *Introduction to the physics and techniques of remote sensing*. Wiley-Interscience. 2006.

ESCUIN, S.; NAVARRO, R.; FERNÁNDEZ, P. Fire severity assessment by using NBR (Normalized Burn Ratio) and NDVI (Normalized Difference Vegetation Index) derived from LANDSAT TM/ETM images. *International Journal of Remote Sensing*, v. 29, n. 4, p. 1053–1073, fev. 2008.

EUGENIO, F. C.; DOS SANTOS, A. R.; FIEDLER, N. C.; RIBEIRO, G. A.; DA SILVA, A. G.; DOS SANTOS, Á. B.; PANETO, G. G.; SCHETTINO, V. R. 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, maio. 2016.

FAO. *Forest Fire Control*. Rome: 1953.

FAO. *Fire management - global assessment 2006*. Rome: 2007.

FAWCETT, T. An introduction to ROC analysis. *Pattern Recognition Letters*, v. 27, n. 8, p. 861–874, 2006.

FINNEY, M. A. An Overview of FlamMap Fire Modeling Capabilities. *Fuel Management-How to Measure Success: Conference Proceedings*. 28-30 March, p. 213–220, 2006.

FLOOD, M.; GUTELIUS, B. Commercial implications of topographic terrain mapping using scanning airborne laser radar.pdf. *Photogrammetric Engineering & Remote Sensing*, v. 63, p. 327–366, 1997.

FORNACCA, D.; REN, G.; XIAO, W. Evaluating the Best Spectral Indices for the Detection of Burn Scars at Several Post-Fire Dates in a Mountainous Region of Northwest Yunnan, China. *Remote Sensing*, v. 10, n. 8, p. 1196, 30 jul. 2018.

GARCÍA, M.; RIAÑO, D.; CHUVIECO, E.; SALAS, J.; DANSON, F. M. Multispectral and LiDAR data fusion for fuel type mapping using Support Vector Machine and decision rules. *Remote Sensing of Environment*, v. 115, n. 6, p. 1369–1379, 2011.

GARCIA-GONZALO, J.; ZUBIZARRETA-GERENDIAIN, A.; RICARDO, A.; MARQUES, S.; BOTEQUIM, B.; BORGES, J.G.; OLIVEIRA, M.M.; TOME, M.; PEREIRA, J.M.C. Modelling wildfire risk in pure and mixed forest stands in Portugal. *Allgemeine Forst und Jagdzeitung (AFJZ) - German Journal of Forest Research* 183 (11/12): 238-248 .2012.

GCOS. *The Global Observing System For Climate Implementation Needs*World Meteorological Organization. 2016. WMO GCOS. p. 316.

GENUER, R.; POGGI, J.-M.; TULEAU-MALOT, C. Variable Selection Using Random Forests Version. *Pattern Recognition Letters*, v. 31, n. 14, p. 2225–2236., 2010.

GEORGIADOU, Y.; KNIPPERS, R. A.; SIDES, E. J.; VAN WESTEN, C. J. Data entry and preparation. In: DE BY, R. A. (Ed.). Principles of Geographic Information Systems - An introductory textbook. The International Institute for Aerospace Survey and Earth Sciences (ITC), Enschede. 2001. p. 490.

GIGLIO, L.; BOSCHETTI, L.; ROY, D. P.; HUMBER, M. L.; JUSTICE, C. O. The Collection 6 MODIS burned area mapping algorithm and product. Remote Sensing of Environment, v. 217, p. 72–85, nov. 2018.

GONZÁLEZ, J. R.; PALAHÍ, M.; TRASOBARES, A.; PUKKALA, T. A fire probability model for forest stands in Catalonia (north-east Spain). Annals of Forest Science, v. 63, n. 2, p. 169–176, mar. 2006.

GONZALEZ, J.R.; PALAHI, M.; TRASOBARES, A.; PUKKALA, T. A fire probability model for forest stands in Catalonia (north-east Spain). Annals of Forest. Science 63: 169-176. 2006.

GONZÁLEZ, J. R.; TRASOBARES, A.; PALAHÍ, M.; PUKKALA, T. Predicting stand damage and tree survival in burned forests in Catalonia (North-East Spain). Annals of Forest Science, v. 64, n. 7, p. 733–742, 2007.

GÖRGENS, E. B.; DA SILVA, A. G. P.; RODRIGUEZ, L. C. E. LIDAR: APLICAÇÕES FLORESTAIS. 1. ed. Editora CRV, Curitiba, PR. 2014.

GÖRGENS, E. B.; PACKALEN, P.; DA SILVA, A. G. P.; ALVARES, C. A.; CAMPOE, O. C.; STAPE, J. L.; RODRIGUEZ, L. C. E. Stand volume models based on stable metrics as from multiple ALS acquisitions in Eucalyptus plantations. Annals of Forest Science, v. 72, n. 4, p. 489–498, 2015.

GUERRA-HERNANDEZ, J.; GÖRGENS, E. B.; GARCÍA-GUTIÉRREZ, J.; RODRIGUEZ, L. C. E.; TOMÉ, M.; GONZÁEZ-FERREIRO, E. Comparison of ALS based models for estimating aboveground biomass in three types of Mediterranean forest. European Journal of Remote Sensing, v. 49, n. February, p. 185–204, 2016.

HANEWINKEL, M.; HUMMEL, S.; ALBRECHT, A. Assessing natural hazards in forestry for risk management: A review. European Journal of Forest Research, v. 130, n. 3, p. 329–351, 2010.

HANEWINKEL, M.; ZHOU, W.; SCHILL, C. A neural network approach to identify forest stands susceptible to wind damage. Forest Ecology and Management, v. 196, n. 2–3, p. 227–243, 2004.

HART, E.; SIM, K.; KAMIMURA, K.; MEREDIEU, C.; GUYON, D.; GARDINER, B. Use of machine learning techniques to model wind damage to forests. *Agricultural and Forest Meteorology*, v. 265, p. 16–29, fev. 2019.

HIJMANS, R.J.; ETEN, J.V. raster: Geographic analysis and modeling with raster data. R package version 2.8-19. 2019.

HISLOP, S.; JONES, S.; SOTO-BERELOV, M.; SKIDMORE, A.; HAYWOOD, A.; NGUYEN, T. H. A fusion approach to forest disturbance mapping using time series ensemble techniques. *Remote Sensing of Environment*, v. 221, n. November 2018, p. 188–197, 2019.

HISLOP, S.; JONES, S.; SOTO-BERELOV, M.; SKIDMORE, A.; HAYWOOD, A.; NGUYEN, T.; HISLOP, S.; JONES, S.; SOTO-BERELOV, M.; SKIDMORE, A.; HAYWOOD, A.; NGUYEN, T. H. Using Landsat Spectral Indices in Time-Series to Assess Wildfire Disturbance and Recovery. *Remote Sensing*, v. 10, n. 3, p. 460, 15 mar. 2018.

HOFFMAN, K. M.; TRANT, A. J.; NIJLAND, W.; STARZOMSKI, B. M. Ecological legacies of fire detected using plot-level measurements and LiDAR in an old growth coastal temperate rainforest. *Forest Ecology and Management*, v. 424, n. January, p. 11–20, 2018.

HOLDEN, Z. A.; SMITH, A. M. S.; MORGAN, P.; ROLLINS, M. G.; GESSLER, P. E. Evaluation of novel thermally enhanced spectral indices for mapping fire perimeters and comparisons with fire atlas data. *International Journal of Remote Sensing*, v. 26, n. 21, p. 4801–4808, 10 nov. 2005.

HOSMER, D. W.; LEMESHOW, S. *Applied Logistic Regression*. 2. ed. John Wiley & Sons, Inc, New York. 2000.

HUDAK, A. T.; BROCKETT, B. H. Mapping fire scars in a southern African savannah using landsat imagery. *International Journal of Remote Sensing*, v. 25, n. 16, p. 3231–3243, 3 ago. 2004.

HUDAK, A. T.; LEFSKY, M. A.; COHEN, W. B.; BERTERRETICHE, M. Integration of Lidar and Landsat Etm+ Data. *Remote Sensing of Environment*, v. 82, p. 95–103, 2002.

HUDAK, A. T.; MORGAN, P.; BOBBITT, M. J.; SMITH, A. M. S.; LEWIS, S. A.; LENTILE, L. B.; ROBICHAUD, P. R.; CLARK, J. T.; MCKINLEY, R. A. the Relationship of Multispectral Satellite Imagery. *Fire Ecology*, v. 3, n. 1, p. 64–90, 2007.

INCRA – INSTITUTO NACIONAL DE COLONIZAÇÃO E REFORMA AGRÁRIA. Acervo Fundiário INCRA. Disponível em: <<http://acervofundiario.incra.gov.br/acervo/acv.php>>. Acesso em: 21/01/2019. 2019.

ISDR - INTERNATIONAL STRATEGY FOR DISASTER REDUCTION. Living with Risk - A global review of disaster reduction initiatives. United Nations publication, New York and Geneva. 2004. v. I

INPE – INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS. Programa Queimadas. Perguntas frequentes. Disponível em: <<http://www.inpe.br/queimadas/portal/informacoes/perguntas-frequentes>>. Acesso em 29/03/2019. 2019a.

_____ – INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS. Programa Queimadas – Banco de Dados. Disponível em: <<http://www.inpe.br/queimadas/bdqueimadas/>>. Acesso em 29/03/2019. 2019b

IPEA – INSTITUTO DE PESQUISA ECONÔMICA APLICADA. Ipeadata. Disponível em: < <http://www.ipeadata.gov.br/Default.aspx>> Acesso em: 27/10/2018. 2018

ISENBURG, M. 2012. LASzip: lossless compression of LiDAR data. Disponível em: <<https://www.cs.unc.edu/~isenburg/lastools/download/laszip.pdf>>. Acesso em: 14/02/2018.

_____. LAStools - efficient tools for LiDAR processing (version 181001, licensed). 2018.

JAISWAL, R. K.; MUKHERJEE, S.; RAJU, K. D.; SAXENA, R. Forest fire risk zone mapping from satellite imagery and GIS. International Journal of Applied Earth Observation and Geoinformation, v. 4, n. 1, p. 1–10, 2002.

JENSEN, J. R. Remote Sensing of the environment: an earth resource perspective. 1. ed. KEITH C. CLARK, Upper Sadle River, NJ. 2000.

KANEVSKI, M.; POZDNOUKHOV, A.; TIMONIN, V. Machine Learning for Spatial Environmental Data. 1. ed. EPFL press, Laussane, Switzerland. 2010.

KAPLAN, A.; HAENLEIN, M. Siri, Siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence. Business Horizons, v. 62, n. 1, p. 15–25, 1 jan. 2019.

KEANE, R. E.; BURGAN, R.; VAN WAGTENDONK, J. Mapping wildland fuels for fire management across multiple scales: Integrating remote sensing, GIS, and biophysical modeling. *International Journal of Wildland Fire*, v. 10, n. 4, p. 301–319, 2001.

KEY, C. H.; BENSON, N. C. Landscape assessment: ground measure of severity, the composite burn index, and remote sensing of severity, the normalized burn index. In: Lutes, D., Keane, R., Caratti, J., Key, C., Benson, N., Sutherland, S., Gangi, L. (Eds.). *FIREMON: Fire Effects Monitoring and Inventory System*, Rocky Mountains Research Station, USDA Forest Service: Fort Collins, CO, USA, pp. 1– 51., 2006.

KIM, S.; LEE, W.-K.; GEIGER, T.; RAHMATI, O.; KIM, G.; SON, Y.; LIM, C.-H.; LEE, J. Multi-Temporal Analysis of Forest Fire Probability Using Socio-Economic and Environmental Variables. *Remote Sensing*, v. 11, n. 1, p. 86, 2019.

KNAPP, N.; FISCHER, R.; HUTH, A. Linking lidar and forest modeling to assess biomass estimation across scales and disturbance states. *Remote Sensing of Environment*, v. 205, p. 199–209, 1 fev. 2018.

KOETZ, B.; MORSDORF, F.; VAN DER LINDEN, S.; CURT, T.; ALLGÖWER, B. Multi-source land cover classification for forest fire management based on imaging spectrometry and LiDAR data. *Forest Ecology and Management*, v. 256, n. 3, p. 263–271, 30 jul. 2008.

KOUTSIAS, N.; KARTERIS, M. Classification analyses of vegetation for delineating forest fire fuel complexes in a Mediterranean test site using satellite remote sensing and GIS. *International Journal of Remote Sensing*, v. 24, n. 15, p. 3093–3104, 2003.

KUHN, M. Building Predictive Models in R Using the caret Package. *Journal of Statistical Software*, v. 28, n. 5, p. 11–21, 2008.

LAGERQUIST, R.; FLANNIGAN, M. D.; WANG, X.; MARSHALL, G. A. Automated prediction of extreme fire weather from synoptic patterns in northern Alberta, Canada. *Canadian Journal of Forest Research*, v. 47, n. 9, p. 1175–1183, set. 2017.

LANDGREBE, D. A. *Signal Theory Methods in Multispectral Remote Sensing*. John Wiley & Sons, Inc., Hoboken, NJ, USA. 2003.

LEFSKY, M. A.; COHEN, W. B.; PARKER, G. G.; HARDING, D. J. Lidar Remote Sensing for Ecosystem Studies. *Bioscience*, v. 52, n. 1, p. 19–30, 2002.

LEUENBERGER, M.; PARENTE, J.; TONINI, M.; PEREIRA, M. G.; KANEVSKI, M. Wildfire susceptibility mapping: Deterministic vs. stochastic approaches. *Environmental Modelling and Software*, v. 101, p. 194–203, 2018.

LEWIS, S. A.; MCKINLEY, R. A.; BOBBITT, M. J.; ROBICHAUD, P. R.; LENTILE, L. B.; SMITH, A. M. S.; MORGAN, P.; CLARK, J. T.; HUDAK, A. T. The Relationship of Multispectral Satellite Imagery to Immediate Fire Effects. *Fire Ecology*, v. 3, n. 1, p. 64–90, nov. 2009.

LI, Z. Evaluation of algorithms for fire detection and mapping across North America from satellite. *Journal of Geophysical Research*, v. 108, n. D2, p. 4076, 2003.

LIM, K.; TREITZ, P.; WULDER, M.; ST-ONGÉ, B.; FLOOD, M. LiDAR remote sensing of forest structure. *Progress in Physical Geography*, v. 27, n. 1, p. 88–106, 2003.

LIU, W. T. H. *Aplicações de Sensoriamento Remoto*. Ed. UNIDERP, Campo Grande. 2006.

LOBO, J. M.; JIMÉNEZ-VALVERDE, A.; REAL, R. AUC: A misleading measure of the performance of predictive distribution models. *Global Ecology and Biogeography*, v. 17, n. 2, p. 145–151, 2007.

LU, B.; HE, Y.; TONG, A. Evaluation of spectral indices for estimating burn severity in semiarid grasslands. *International Journal of Wildland Fire*, v. 25, n. 2, p. 147, 25 fev. 2016.

MACHALA, M.; ZEJDOVÁ, L. Forest mapping through Object-based image analysis of multispectral and LiDAR Aerial data. *European Journal of Remote Sensing*, v. 47, n. 1, p. 117–131, 2014.

MALKINSON, L. D.; BOTTAI, L.; WITTENBERG, L.; MASELLI, F.; FOX, D. M.; LAAROUSSI, Y.; ANDRIEU, J. POSTFIRE: A model to map forest fire burn scar and estimate runoff and soil erosion risks. In: *Remote Sensing Applications: Society and Environment*. Elsevier, [s. L.]. 2016. v. 4p. 83–91.

MANN, M. L.; BATLLORI, E.; MORITZ, M. A.; WALLER, E. K.; BERCK, P.; FLINT, A. L.; FLINT, L. E.; DOLFI, E. Incorporating anthropogenic influences into fire probability models: Effects of human activity and climate change on fire activity in California. *PLoS ONE*, v. 11, n. 4, p. 1–21, 2016.

MARQUES, S.; GARCIA-GONZALO, J.; BORGES, J. G.; BOTEQUIM, B.; OLIVEIRA, M. M.; TOMÉ, J.; TOMÉ, M. Developing post-fire *Eucalyptus globulus* stand damage and tree mortality models for enhanced forest planning in Portugal. *Silva Fennica*, v. 45, n. 1, p. 69–83, 2011.

MARQUES, S.; GARCIA-GONZALO, J.; BOTEQUIM, B.; RICARDO, A.; BORGES, J. G.; TOME, M.; OLIVEIRA, M. M. Assessing wildfire occurrence probability in Pinus pinaster Ait. stands in Portugal. *Forest Systems*, v. 21, n. 1, p. 111, 2012.

MEASURES, R. M. *Laser Remote Sensing Fundamentals and Applications*. 2. ed. Krieger Publishing Company, Malabar, Florida. 1992.

METZ, C. E. Basic principles of ROC analysis. *Seminars in Nuclear Medicine*, v. 8, n. 4, p. 283–298, 1 out. 1978.

MILLER, J. D.; THODE, A. E. Quantifying burn severity in a heterogeneous landscape with a relative version of the delta Normalized Burn Ratio (dNBR). *Remote Sensing of Environment*, v. 109, n. 1, p. 66–80, 12 jul. 2007.

MMA. MINISTÉRIO DO MEIO AMBIENTE. Geocatálogo. Disponível em: <<http://www.geocatalogo.mma.gov.br>>. Acesso em: 01/04/2018. 2018.

MURTAGH, F.; LEGENDRE, P. Ward's Hierarchical Clustering Method: Clustering Criterion and Agglomerative Algorithm. [S.l.: s.n.]. Disponível em: <www.clustan.com>. Acesso em: 8 abr. 2019.

MUTLU, M.; POPESCU, S. C.; STRIPLING, C.; SPENCER, T. Mapping surface fuel models using lidar and multispectral data fusion for fire behavior. *Remote Sensing of Environment*, v. 112, n. 1, p. 274–285, 2008.

MUTLU, M.; POPESCU, S. C.; ZHAO, K. Sensitivity analysis of fire behavior modeling with LIDAR-derived surface fuel maps. *Forest Ecology and Management*, v. 256, n. 3, p. 289–294, 2008.

NAMI, M. H.; JAAFARI, A.; FALLAH, M.; NABIUNI, S. Spatial prediction of wildfire probability in the Hyrcanian ecoregion using evidential belief function model and GIS. *International Journal of Environmental Science and Technology*, v. 15, n. 2, p. 373–384, 2018.

NASA – MODIS. Moderate Resolution Imaging Spectroradiometer. About. Disponível em: <<https://modis.gsfc.nasa.gov/about/>>. Acesso em: 31/03/2019. 2019.

NELSON, R.; KRABILL, W.; TONELLI, J. Estimating forest biomass and volume using airborne laser data. *Remote Sensing of Environment*, v. 24, n. 2, p. 247–267, 1 mar. 1988.

NGUYEN, N. T.; DANG, B.-T. N.; PHAM, X.-C.; NGUYEN, H.-T.; BUI, H. T.; HOANG, N.-D.; TIEN BUI, D. Spatial pattern assessment of tropical forest fire danger at Thuan Chau area (Vietnam) using GIS-based advanced machine learning algorithms: A comparative study. *Ecological Informatics*, v. 46, p. 74–85, 1 jul. 2018.

NILSSON, N. J. *The Quest for Artificial Intelligence: A history of ideas and achievements*. Cambridge University Press, Cambridge, UK. 2009.

NOAA – NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION. Our Satellites. History of NOAA Satellites. Disponível em: <<https://www.nesdis.noaa.gov/content/history-noaa-satellites>>. Acesso em: 31/03/2019.

NOVO, E. M. L. DE M. *Sensoriamento remoto : princípios e aplicações*. Blucher, São Paulo. 2010.

PAPA, D. DE A. *Impacto do manejo de precisão em florestas tropicais*. 94 p. Dissertação (Mestrado em Ciências Florestais – ESALQ/USP), Piracicaba, 2018.

PARKS, S. A.; PARISIEN, M.-A.; MILLER, C.; DOBROWSKI, S. Z. Fire Activity and Severity in the Western US Vary along Proxy Gradients Representing Fuel Amount and Fuel Moisture. *PLoS ONE*, v. 9, n. 6, p. e99699, 18 jun. 2014.

PATTERSON, M. W.; YOOL, S. R. Mapping fire-induced vegetation mortality using landsat thematic mapper data: A comparison of linear transformation techniques. *Remote Sensing of Environment*, v. 65, n. 2, p. 132–142, 1998.

PEREIRA, A. A.; PEREIRA, J. M. C.; LIBONATI, R.; OOM, D.; SETZER, A. W.; MORELLI, F.; MACHADO-SILVA, F.; DE CARVALHO, L. M. T. Burned area mapping in the Brazilian Savanna using a one-class support vector machine trained by active fires. *Remote Sensing*, v. 9, n. 11, 2017.

PEREIRA, M. C.; SETZER, A. W. Spectral characteristics of fire scars in Landsat-5 TM images of amazonia. *International Journal of Remote Sensing*, v. 14, n. 11, p. 2061–2078, 1993.

PFLUGMACHER, D.; KENNEDY, W. B.; YANG, R. E.; PFLUGMACHER, D.; COHEN, W. B.; KENNEDY, R. E.; YANG, Z. Using Landsat-derived disturbance and recovery history and lidar to map forest biomass dynamics. *Remote Sensing of Environment*, v. 151, p. 124–137, 2013.

PHAM, B. T.; DIEU, .; BUI, T.; PRAKASH, I. Bagging based Support Vector Machines for spatial prediction of landslides. *Environmental Earth Sciences*, v. 77, p. 146, 2018.

PHILLIPS, C.; NICKEY, B. The concept of “spatial risk” and its application to fire prevention. *Fire management notes*, v. 39, p. 4, 7-8, 19, 1978.

PICKELL, P. D.; HERMOSILLA, T.; FRAZIER, R. J.; COOPS, N. C.; WULDER, M. A. Forest recovery trends derived from Landsat time series for North American boreal forests. *International Journal of Remote Sensing*, v. 37, n. 1, p. 138–149, 2 jan. 2016.

POPESCU, S. C.; WYNNE, R. H.; SCRIVANI, J. A. Fusion of Small-Footprint Lidar and Multispectral Data to Estimate Plot- Level Volume and Biomass in Deciduous and Pine Forests in Virginia, USA. *Forest Science*, v. 50, n. 4, p. 551–565, 1 ago. 2004.

PRATA, G.A. Estimação do risco e do valor das florestas para fins securitátios no Brasil. 126 p. Dissertação (Mestrado em Ciências Florestais – ESALQ/USP), Piracicaba, 2012.

QGIS DEVELOPMENT TEAM. QGIS Geographic Information System. Open Source Geospatial Foundation Project. <http://qgis.osgeo.org>. V.2.18.15. 2018

R CORE TEAM. **R: A language and environment for statistical computing**. 2008

RICHARDS, F.J. A flexible growth function for empirical use. **Journal of Experimental Botany**, Oxford, v. 10, n. 29, p. 290-300, 1959.

RAPIDEYE. **Imagery Product Especifications**. Version 6.1. 2016. 50 p. Disponível em: < <https://www.planet.com/products/satellite-imagery/files/160625-RapidEye%20Image-Product-Specifications.pdf>>. Acesso em 18/01/2019.

REUTEBUCH, S. E.; ANDERSEN, H.; MCGAUGHEY, R. J. Light Detection and Ranging (LIDAR): An Emerging Tool for Multiple Resource Inventory. *Journal of Forestry*, v. 103, n. 6, p. 286–292, 2005.

RODRIGUES, M.; DE LA RIVA, J. An insight into machine-learning algorithms to model human-caused wildfire occurrence. *Environmental Modelling and Software*, v. 57, p. 192–201, 2014.

RODRIGUEZ, L.C.E. Técnicas quantitativas para a gestão de florestas plantadas. Piracicaba: ESALQ, Depto. Ciências Florestais, 2005. 106 p.

RODRIGUEZ, L.C.E.; SEIXAS, F. PRATA, G.A.; NOBRE, S.R. Caracterização da história recente de incêndios em plantações industriais no Brasil. In: SIMPÓSIO SUL-AMERICANO SOBRE CONTROLE DE INCÊNDIOS FLORESTAIS, 5., 2011, Campinas. Anais,. Campinas: UFPR, 2011. 1 CD-ROM.

ROGAN, J.; YOOL, S. R. International Journal of Remote Sensing Mapping fire-induced vegetation depletion in the Peloncillo Mountains, Arizona and New Mexico Mapping re-induced vegetation depletion in the Peloncillo Mountains, Arizona and New Mexico. *International Journal of Remote Sensing*, v. 22, n. 16, p. 3101–3121, 2001.

ROTHERMEL, R. C. A mathematical model for predicting fire spread in wildland fuels. Disponível em: <https://www.fs.fed.us/rm/pubs_int/int_rp115.pdf>. Acesso em: 30 mar. 2019.

ROUSE, J. .; HAAS, R. H.; SCHELL, J. A.; DEERING, D. . MONITORING THE VERNAL ADVANCEMENT AND RETROGRADATION (GREEN WAVE EFFECT) OF NATURAL VEGETATION. NASA SP-351 I, p. 309–317, 1973.

ROUSSEL, J.R.; AUTY, D. lidR: Airborne LiDAR Data Manipulation and Visualization for Forestry Applications. 2019.

ROY, D. P.; BOSCHETTI, L.; JUSTICE, C. O.; JU, J. The collection 5 MODIS burned area product - Global evaluation by comparison with the MODIS active fire product. *Remote Sensing of Environment*, v. 112, n. 9, p. 3690–3707, 15 set. 2008.

SAKR, G. E.; ELHAJJ, I. H.; MITRI, G. Efficient forest fire occurrence prediction for developing countries using two weather parameters. *Engineering Applications of Artificial Intelligence*, v. 24, p. 888–894, 2011.

SALAS, J.; CHUVIECO, E. Geographic information systems for wildland fire risk mapping. *Wildfire*, v. 3, n. 2, p. 7–13, 1994.

SANTOS, C. S.; SILVA, J. L. C. Os impactos do plantio de eucalipto e da produção de celulose em comunidades tradicionais no extremo sul baiano. Disponível em <http://www.anppas.org.br/encontro/segundo/papers/GT17/gt17_jose_caetano.pdf> 2004. Acesso em: 22/11/2018.

SCHROEDER, T. A.; WULDER, M. A.; HEALEY, S. P.; MOISEN, G. G. Detecting post-fire salvage logging from Landsat change maps and national fire survey data. *Remote Sensing of Environment*, v. 122, p. 166–174, 2012.

SIDRA - Sistema IBGE de Recuperação Automática. Disponível em: <<https://sidra.ibge.gov.br>>. Acesso em: 27/10/2018. 2018

SOARES, R.V. Ocorrência de incêndios em povoamentos florestais. *Floresta*, Curitiba, v. 22, n. 1/2, p. 39-53, 1992.

SOARES, R. V.; BATISTA, A. C. Incêndios Florestais: controle, efeito e uso do fogo. 1. ed. Os autores, Curitiba/PR. 2007.

SOARES, R. V.; BATISTA, A. C.; SANTOS, J. F. Evolução do perfil dos incêndios florestais em áreas protegidas no Brasil, de 1983 a 2002. *Floresta*, v. 36, n. 1, p. 93–100, 2006.

STARK, S. C.; LEITOLD, V.; WU, J. L.; HUNTER, M. O.; DE CASTILHO, C. V.; COSTA, F. R. C.; MCMAHON, S. M.; PARKER, G. G.; SHIMABUKURO, M. T.; LEFSKY, M. A.; KELLER, M.; ALVES, L. F.; SCHIETTI, J.; SHIMABUKURO, Y. E.; BRANDÃO, D. O.; WOODCOCK, T. K.; HIGUCHI, N.; DE CAMARGO, P. B.; DE OLIVEIRA, R. C.; SALESKA, S. R. Amazon forest carbon dynamics predicted by profiles of canopy leaf area and light environment. *Ecology Letters*, v. 15, n. 12, p. 1406–1414, 2012.

TETTO, A. F.; SOARES, R. V. Eficiência do combate aos incêndios florestais, no período de 1965 a 2009 , na Fazenda Monte Alegre , Paraná Fire suppression efficiency in the Fazenda Monte Alegre , Paraná , from 1965 to 2009. p. 483–489, 2012.

TIEN BUI, D.; LE, H. VAN; HOANG, N. D. GIS-based spatial prediction of tropical forest fire danger using a new hybrid machine learning method. *Ecological Informatics*, v. 48, n. April, p. 104–116, 2018.

TIEN BUI, D.; TUAN, T. A.; KLEMPE, H.; REVHAUG, I.; PRADHAN, B. Spatial prediction models for shallow landslide hazards: a comparative assessment of the efficacy of support vector machines, artificial neural networks, kernel logistic regression, and logistic model tree. *Landslides*, v. 13, n. 2, p. 361–378, 2016.

TING K.M. Confusion Matrix. In: Sammut C., Webb G.I. (eds) *Encyclopedia of Machine Learning*. Springer, Boston, MA. 2011

TVEITE, N. R. NNJoin. QGIS Python Plugins Repository. V. 3.1.2. 2018

VETTORAZZI, C.A.; FERRAZ, S.F.B. Uso de sistemas de informações geográficas aplicados à prevenção e combate a incêndios em fragmentos florestais. *Série Técnica IPEF*, Piracicaba, v. 12, n. 32, p. 111-115, dez. 1998.

VAN DER WERF, G. R.; RANDERSON, J. T.; GIGLIO, L.; VAN LEEUWEN, T. T.; CHEN, Y.; ROGERS, B. M.; MU, M.; VAN MARLE, M. J. E.; MORTON, D. C.; COLLATZ, G. J.; YOKELSON, R. J.; KASIBHATLA, P. S. Global fire emissions estimates during 1997-2016. *Earth System Science Data*, v. 9, n. 2, p. 697–720, 2017.

- VAPNIK, V. N. *Statistical Learning Theory*. John Wiley & Sons, Inc, New York. 1988.
- VAPNIK, V. N. *The nature of statistical learning theory*. New York. Springer. 2nd ed. 2000.
- VON GADOW, K. Evaluating risk in forest planning models. *Silva Fennica*, v. 34, n. 2, p. 181–191, 2000.
- WARD, J. H. Hierarchical Grouping to Optimize an Objective Function. *Journal of the American Statistical Association*, v. 58, n. 301, p. 236–244, mar. 1963.
- WASILEWSKI, T. Nota prévia sobre os incêndios em plantações de *Araucaria angustifolia* na fazenda Monte Alegre. In: SIMPÓSIO DE REFLORESTAMENTO DA REGIÃO DA ARAUCÁRIA, 1., Curitiba, 1963. Anais,. Curitiba: FIEP, 1965. p. 75-77.
- WEICHEL, H.; ROSSO, P.; MARX, A.; REIGBER, K. D.; HEYNEN, M. White Paper: The RapidEye Red Edge Band. 2016.
- WILKES, P.; JONES, S.; SUAREZ, L.; MELLOR, A.; WOODGATE, W.; SOTO-BERELOV, M.; HAYWOOD, A.; SKIDMORE, A. Mapping Forest Canopy Height Across Large Areas by Upscaling ALS Estimates with Freely Available Satellite Data. *Remote Sensing*, v. 7, n. 9, p. 12563–12587, 23 set. 2015.
- XU, C.; MANLEY, B.; MORGENROTH, J. Evaluation of modelling approaches in predicting forest volume and stand age for small-scale plantation forests in New Zealand with RapidEye and LiDAR. *International Journal of Applied Earth Observation and Geoinformation*, v. 73, n. June, p. 386–396, 2018.
- YOU, W.; LIN, L.; WU, L.; JI, Z.; YU, J.; ZHU, J.; FAN, Y.; HE, D. Geographical information system-based forest fire risk assessment integrating national forest inventory data and analysis of its spatiotemporal variability. *Ecological Indicators*, v. 77, p. 176–184, 2017.
- YU, P. S.; YANG, T. C.; CHEN, S. Y.; KUO, C. M.; TSENG, H. W. Comparison of random forests and support vector machine for real-time radar-derived rainfall forecasting. *Journal of Hydrology*, v. 552, p. 92–104, 2017.
- ZONETE, M. F.; RODRIGUEZ, L. C. E.; PACKALÉN, P. Estimación de parâmetros biométricos de plantios clonais de eucalipto no sul da Bahia: Uma aplicação da tecnologia laser aerotransportada. *Scientia Forestalis/Forest Sciences*, v. 38, n. 86, p. 225–235, 2010.

ANEXOS

ANEXO A – GRÁFICOS MODELO NÃO-LINEAR

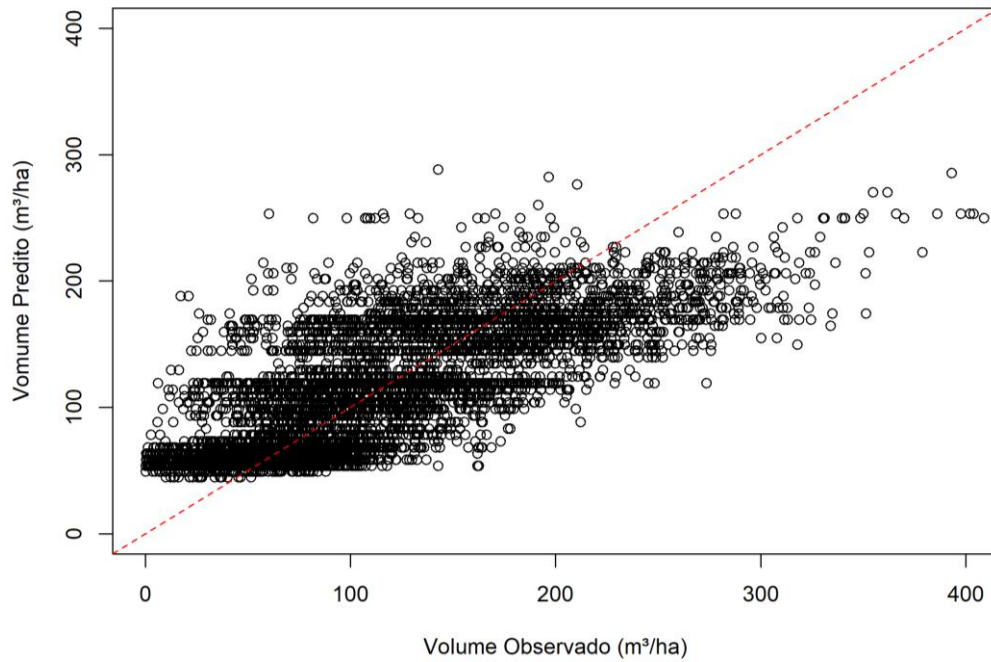


Figura 18. Volume Observado x Volume Predito. Modelo não-linear criado pela relação entre idade e volume.

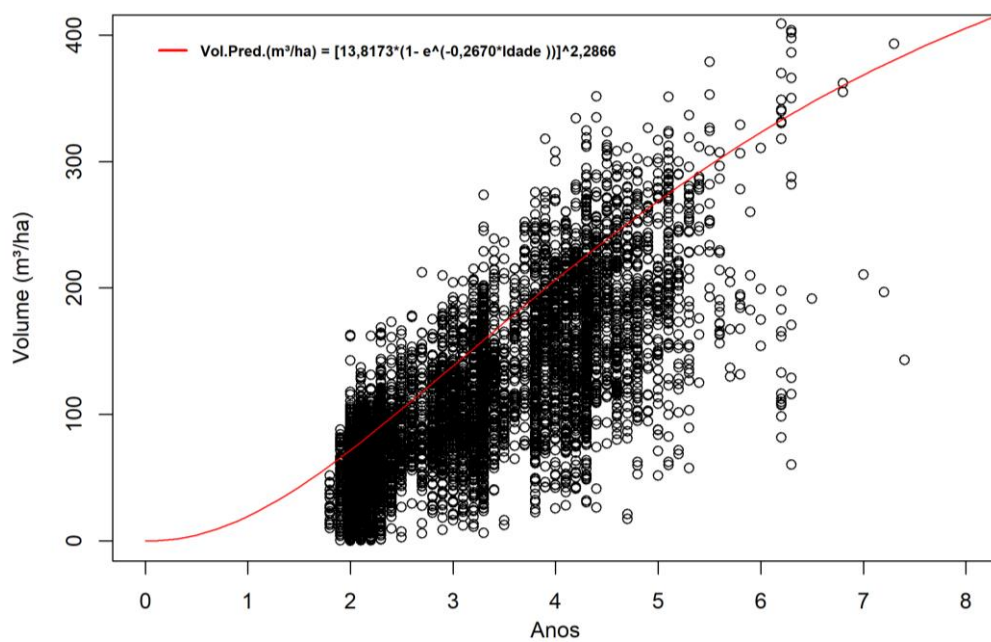


Figura 19. Ajuste não-linear entre os dados observados de volume em função da idade.

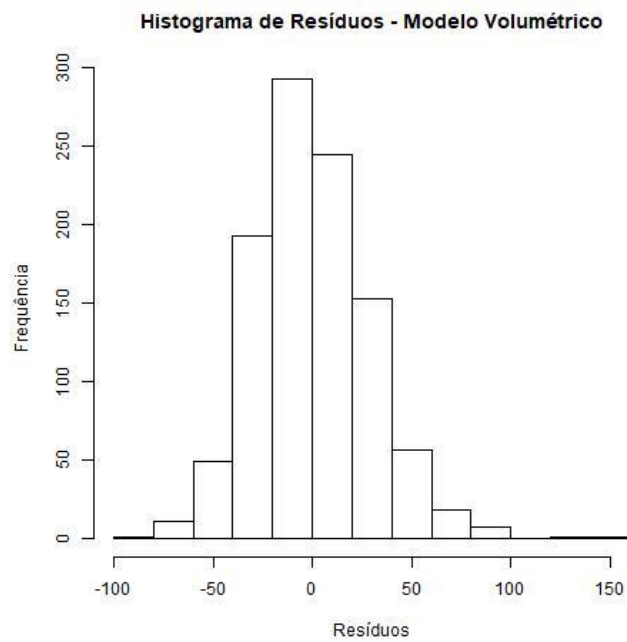


Figura 20. Histograma de resíduos para o modelo volumétrico não-linear.

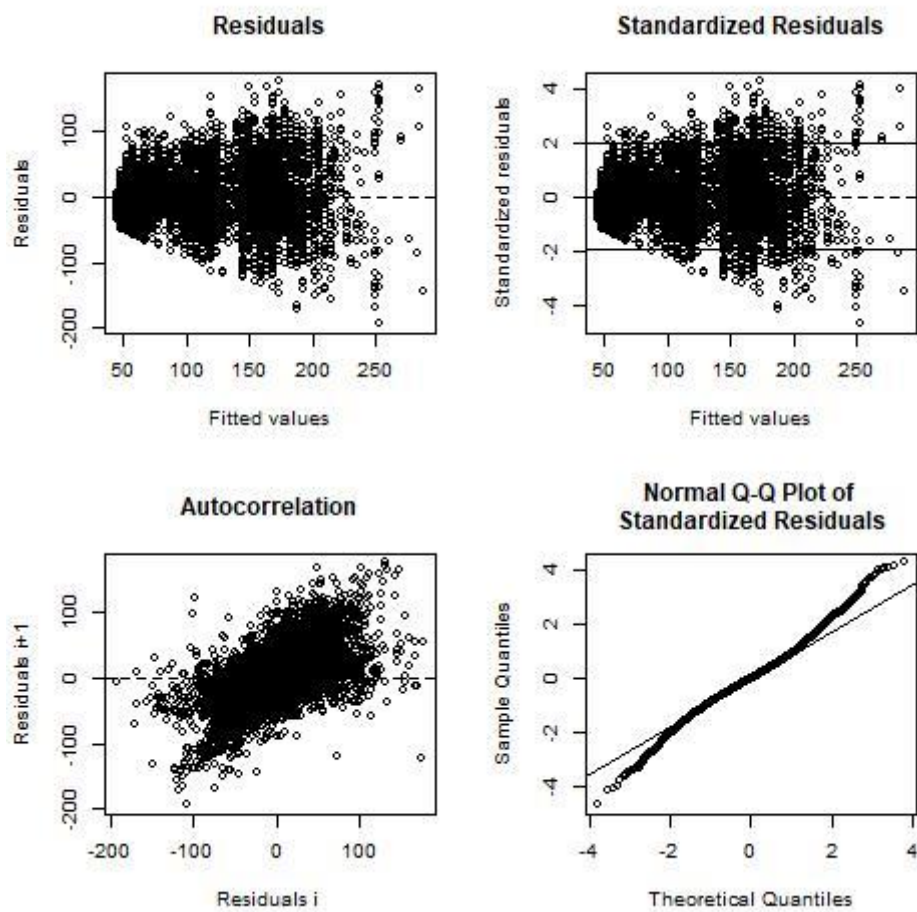


Figura 21. Gráficos de resíduos para o modelo volumétrico não-linear.

ANEXO B – GRÁFICOS MODELO DE VOLUME POR MÉTRICAS LIDAR

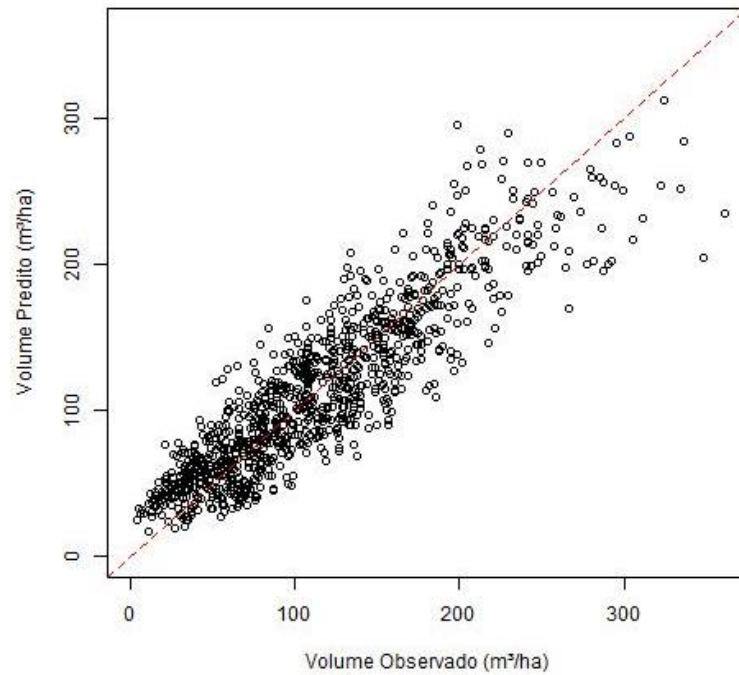


Figura 22. Volume Observados x Volume Predito. Modelagem utilizando métricas LiDAR e dados de parcelas de inventário para o ano de 2015

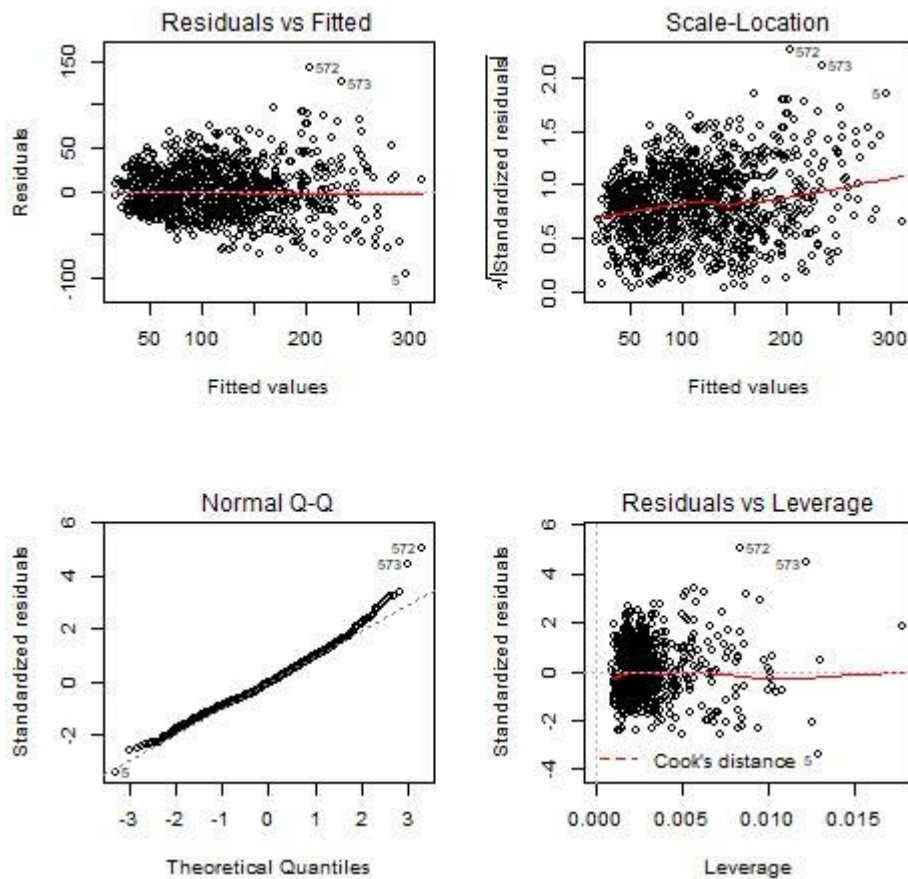


Figura 23. Gráficos de resíduos para o modelo volumétrico não-linear.

ANEXO C – MODELAGEM CICATRIZ DE INCÊNDIO

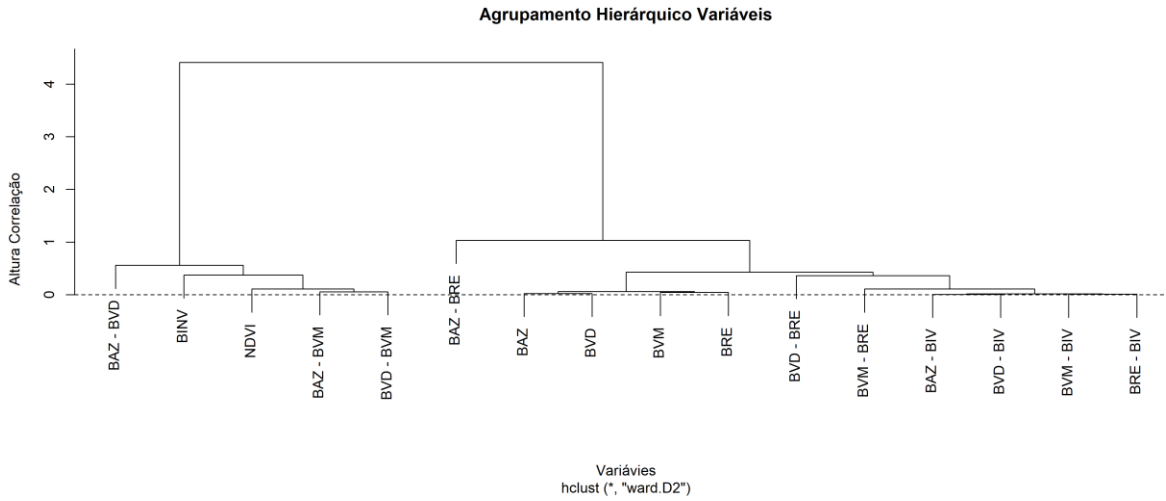


Figura 24. Agrupamento hierárquico de variáveis, apenas com dados multiespectrais.

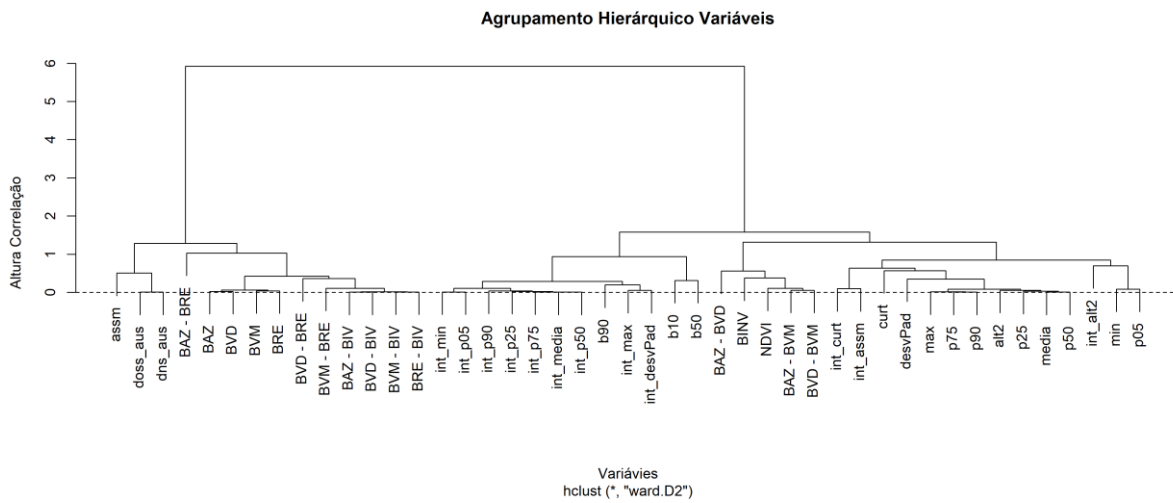


Figura 25. Agrupamento hierárquico de variáveis com os dados fusionados dos sensores multiespectrais e métricas LiDAR.

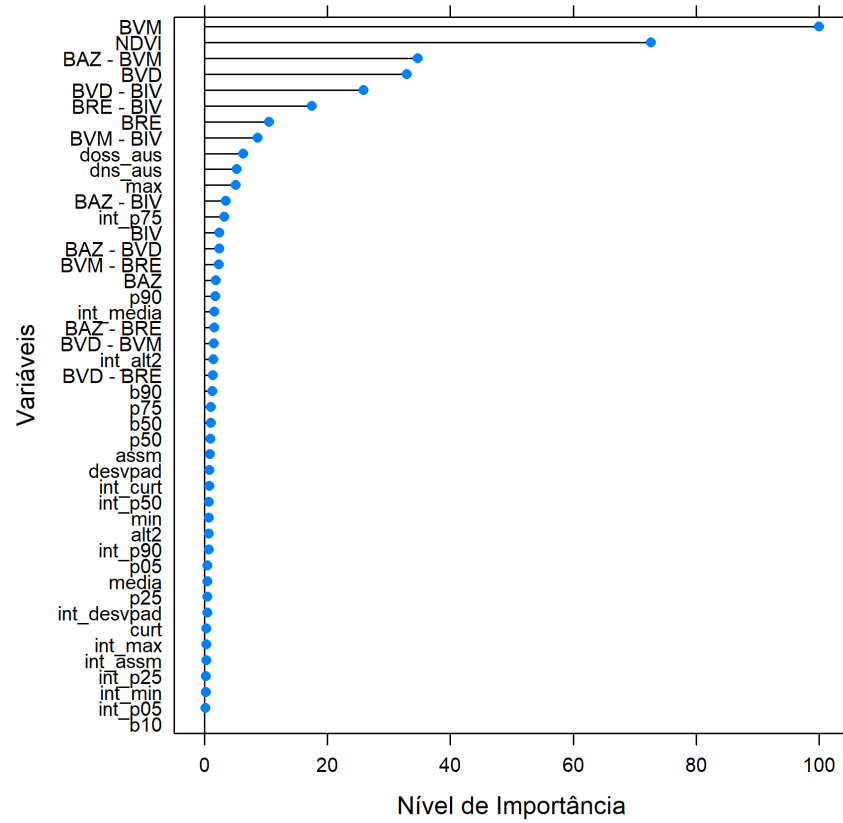


Figura 26. Importância de variáveis identificadas pelo algoritmo Random Forest utilizando dados da fusão entre sensores.

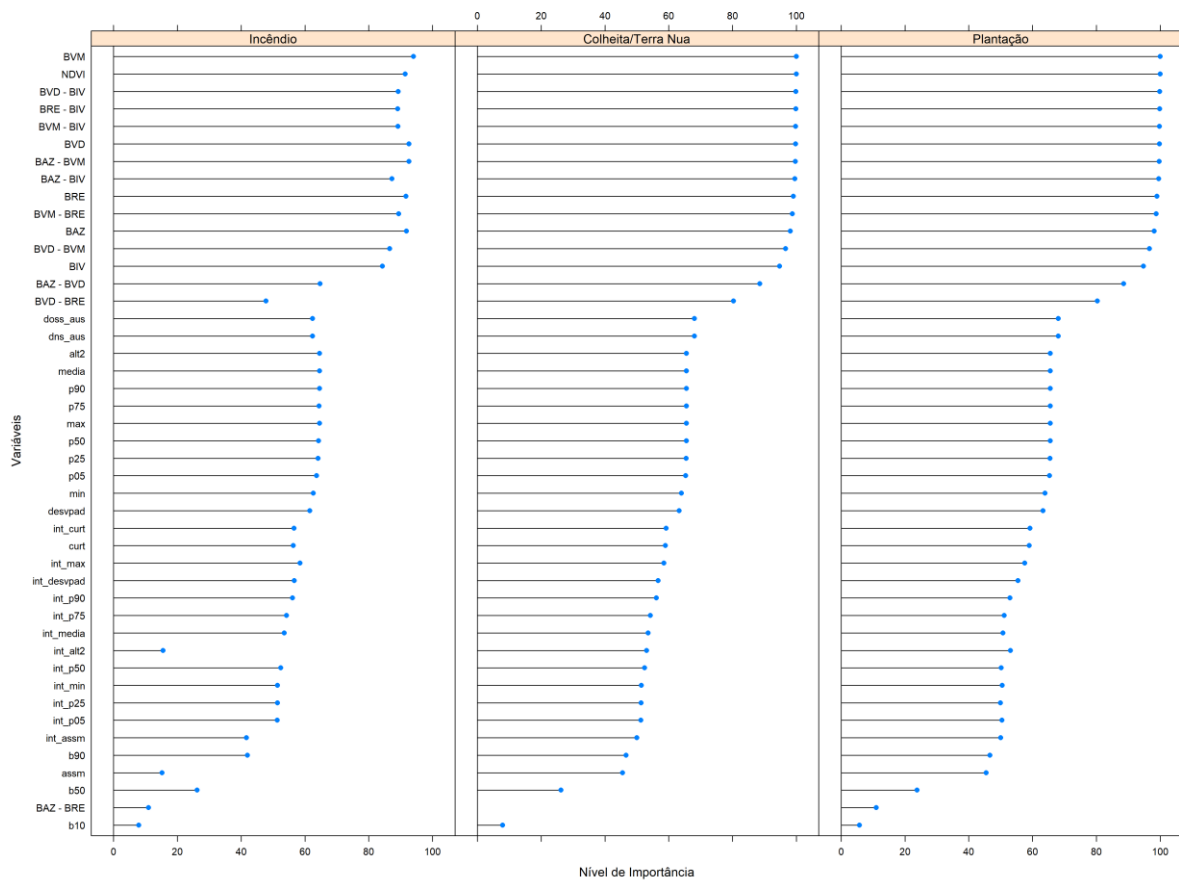


Figura 27. Importância de variáveis identificadas pelo algoritmo Support Vector Machine, para cada classe.