

**AS CONTRIBUIÇÕES PREDITIVAS DO DANO AMBIENTAL PARA GESTÃO PÚBLICA
DE UNIDADES DE CONSERVAÇÃO: ESTUDO DE CASO DA APA DO ALTO DO MUCURI
EM MINAS GERAIS**

DIEGO DA SILVA PASSOS

UNIVERSIDADE FEDERAL DOS VALES DO JEQUITINHONHA E MUCURI (UFVJM)

JOÃO PAULO CALEMO BATISTA MENEZES

UNIVERSIDADE FEDERAL DOS VALES DO JEQUITINHONHA E MUCURI (UFVJM)

BRUNO DE ALMEIDA VILELA

UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO (UFES)

JORGE FULGÊNCIO SILVA CHAVES

UNIVERSIDADE FEDERAL DOS VALES DO JEQUITINHONHA E MUCURI (UFVJM)

AS CONTRIBUIÇÕES PREDITIVAS DO DANO AMBIENTAL PARA GESTÃO PÚBLICA DE UNIDADES DE CONSERVAÇÃO: ESTUDO DE CASO DA APA DO ALTO DO MUCURI EM MINAS GERAIS

1 INTRODUÇÃO

A crise ambiental no mundo, decorrente principalmente da ideologia errônea de recursos naturais inesgotáveis e exploração de maneira deliberada, começou a ser debatida a partir do surgimento de movimentos ambientalistas. A nova visão ambiental evidenciou a dificuldade da estabilidade da natureza em face a extração desordenada do ser humano e a necessidade do desenvolvimento sustentável para o bem-estar da sociedade (PRAES, 2012). Como desdobramento, percebeu-se o crescimento de políticas ambientais voltadas para a temática da proteção ambiental.

No Brasil, a intensificação de políticas ambientais ocorreu no século XX, com normatizações que promoviam a interação sustentável entre o ser humano e a natureza, e que perduram nos dias atuais (PECCATIELLO, 2011). Dentre elas, pode ser citada a Lei 9.985/00, conhecida como Sistema Nacional de Unidades de Conservação (SNUC). O SNUC regulamenta a criação e gestão de unidades de conservação no Brasil delimitando características para cada tipologia a ser preservada. Os artigos 8 e 14 da Lei 9.985/00 parametriza 12 categorias de áreas para proteção ambiental, podendo ser destacado a Área de Preservação Ambiental (APA).

A tipologia Área de Preservação Ambiental está elencada no SNUC como uma unidade de conservação de Uso Sustentável, dotada de riquezas naturais e responsável pela integração sustentável do homem e natureza (BRASIL, 2000). A partir da relevância da APA para a sociedade, o artigo realizou o estudo de caso da APA do Alto do Mucuri, área situada no nordeste de Minas Gerais e de grande importância no desenvolvimento socioambiental e econômico da região.

A APA do Alto do Mucuri, assim como as demais unidades de conservação, possui desafios que colocam em risco a sua sobrevivência no meio ambiente. O incêndio florestal é uma dessas problemáticas, responsáveis por gerar anualmente prejuízos inestimáveis para os cofres públicos e a biodiversidade (Anderson *et al.*, 2019). De acordo com Santos, Soares e Batista (2006), estima-se a perda de 85.735,02 hectares de vegetação ocorridas em 19.377 registros de incêndios em unidades de preservação ambiental no Brasil, entre os anos de 1998 a 2002. Nos últimos anos, os incêndios florestais estão cada vez mais evidenciados nos meios de comunicação, normalmente provocados pelo próprio ser humano, dando a sensação de impunidade e retrocesso nas políticas ambientais.

Uma das formas de combate ao incêndio florestal é conhecer o perfil das ocorrências de incêndios para a prevenção, controle e auxílio na tomada de decisões da gestão pública (IBAMA, 2009). Uma das ferramentas para alcance desses resultados é a utilização do Registro de Ocorrência de Incêndio (ROI). Os ROI's são formulários preenchidos por brigadistas ou gestor de unidade de conservação, contendo informações que vão desde a identificação do incêndio florestal até a sua debelação (BONTEMPO *et al.*, 2011).

A partir dos fatos elucidados, o presente estudo propõe analisar as contribuições preditivas que o dano ambiental pode gerar para a gestão pública no combate a incêndios florestais, utilizando os ROI's auferidos pela APA do Alto do Mucuri, no período de 2014 a 2018. Para isso, a pesquisa se pauta na utilização do algoritmo de aprendizado de máquina *Random Forest* como técnica e ferramenta potencial para a administração de unidades de conservação. Em complemento aos resultados, são abordados ainda a estatística descritiva dos

elementos obtidos dos ROI's e os fatores determinantes para a ocorrência dos incêndios florestais da APA do Alto do Mucuri.

2 O SNUC E SUA IMPORTÂNCIA NO CONTEXTO AMBIENTAL

No decorrer da história humana, pode ser percebida a perda de espécies da fauna e flora à medida que a urbanização e industrialização foram se ampliando mundialmente. De acordo com Godoy e Leuzinger (2015), muitos cientistas defendem e relatam a necessidade da criação de áreas protegidas de maneira legal para contenção dessa depredação ambiental, como é o caso das unidades de conservação no Brasil.

O Brasil é um país singular no quesito riquezas e belezas naturais, possuindo ainda uma das maiores biodiversidades do planeta:

O país possui 56 mil espécies de plantas com 22% do total de 250 mil plantas existentes em todo o planeta. A riqueza animal é representada por 524 espécies de mamíferos, 70 espécies de pássaros da ordem psittaciformes (araras, papagaios e periquitos) mais de três mil espécies de peixes de água doce e algo entre 10 e 15 milhões de espécies de insetos (a grande maioria ainda não foi descrita pela ciência) (CUNHA, OLIVEIRA, BARBALHO, 2017, p. 2).

Buscando suprir a necessidade de conservação dos recursos naturais brasileiros é instituída no dia 18 de julho de 2000, a Lei nº 9.985, conhecida como o Sistema Nacional de Unidades de Conservação da Natureza (SNUC). A referida normatização tem como objetivos a manutenção da diversidade biológica, proteção de espécies ameaçadas de extinção, preservação e restauração da diversidade de ecossistemas naturais, promoção do desenvolvimento sustentável, recuperação dos recursos naturais e valorização socioeconômica da diversidade biológica (BRASIL, 2000).

Uma novidade que o SNUC trouxe consigo, foi a divisão das características de cada unidade de conservação em dois grandes blocos: Unidades de Proteção Integral – que permitem apenas o uso indireto dos recursos naturais, salvo os casos previstos em lei (5 tipos de unidades de conservação definidas); e Unidades de Uso Sustentável – que relaciona o equilíbrio entre o uso sustentável de parte dos recursos naturais com a conservação da biodiversidade do ambiente (7 tipos de unidades de conservação definidas) (BRASIL, 2000).

3 A APA DO ALTO DO MUCURI

A APA do Alto do Mucuri foi criada através do Decreto nº. 45.877, no dia 30 de dezembro de 2011 e está classificada como uma unidade de conservação do tipo Uso Sustentável. Localizada no Vale do Mucuri, região nordeste de Minas Gerais, ela abrange uma área correspondente a 325.148,8883ha (Figura 01). Seu território compreende oito municípios: Caraí, Catuji, Novo Cruzeiro, Malacacheta, Poté, Teófilo Otoni, Itaipé e Ladainha, sendo os dois últimos totalmente inseridos na circunscrição da unidade (MINAS GERAIS, 2011)

A região é caracterizada pela grande concentração do bioma Mata Atlântica, sobretudo, da fitofisionomia Floresta Estacional Semidecidual, vegetação conhecida pela perda de folhagens durante períodos de estiagem e grande quantidade de nutrientes depositadas no solo (PEREIRA, 2016). É uma das principais responsáveis pela recarga hídrica da bacia do rio Mucuri, com extensão de aproximadamente 14 mil km² no território mineiro e colabora no abastecimento de água regional (FELIPPE; BUENO; COSTA, 2009). Sob o aspecto socioeconômico, a APA do Alto do Mucuri possui como atividade principal a agropecuária em propriedades rurais, sendo relevante fonte para subsistência de pequenos agricultores familiares (PEREIRA, 2016). Segundo Fahel, Leite e Teles (2014), a região do Jequitinhonha/ Mucuri é

Os ROI's são preenchidos pelos indivíduos que atuam no combate de incêndios florestais, normalmente, o gestor da unidade de conservação ou o brigadista atuante na ocorrência. Para evitar perdas de informações, recomenda-se a confecção do ROI o mais breve possível, o que inclui a coleta dos dados logo ao início do incêndio florestal. Portanto, a qualidade do ROI depende do zelo, cuidado no preenchimento, a fim de evitar o registro incompleto de dados e a falta de clareza no documento (BONTEMPO *et al.*, 2011).

Segundo Bontempo *et al.* (2011), a utilização dos ROI's nas unidades de conservação encontrou respaldo em políticas públicas e foram aperfeiçoadas com o passar do tempo, sendo ajustadas conforme a realidade presenciada em eventos de incêndios. Para o IBAMA (2010), a adoção dos ROI's em unidades de conservação tem apresentado resultados positivos na redução dos incêndios florestais, apesar das queimadas serem ainda um grande problema ambiental no Brasil.

5 PROCEDIMENTOS METODOLÓGICOS

A coleta de dados dos ROI's da APA do Alto do Mucuri foi obtida por solicitação ao Instituto Estadual de Florestas, através do portal eletrônico de Acesso à Informação do Governo Federal, protocolo nº. 02100000001201970. Foram selecionados 138 ROI's entre os anos 2014 a 2018, com a seguinte distribuição de ocorrências: 45 (2014), 36 (2015), 13 (2016), 7 (2017) e 37 (2018). O período se estende desde o início da adoção desta ferramenta nas atividades da unidade de conservação até a data com dados disponíveis.

Como complemento, foram coletados dados climáticos e de poluentes atmosféricos do Sistema de Informações Ambientais Integrado a Saúde (SISAM) e cobertura florestal natural e agropecuária da região através do Projeto de Mapeamento Anual da Cobertura e Uso do Solo no Brasil (MapBiomias). Dados disponibilizados para amplo acesso da sociedade e obtidos nos respectivos portais eletrônicos.

A técnica de análise dos dados utilizada no estudo foi o algoritmo *Random Forest* ou *Random Decision Forest* (em português Florestas Aleatórias ou Florestas de Decisão Aleatórias). Considerada uma técnica de aprendizagem de máquina, utiliza Árvores de Decisão para realização de predições de eventos. Neste artigo, a predição de ocorrências de incêndios se deu com objetivo de identificar as variáveis determinantes do dano ambiental. Para aplicação da técnica foi utilizado o *software* R Studio, versão 1.3.1073. O R Studio utiliza a linguagem R para geração de cálculos e gráficos estatísticos.

5.1 Algoritmo *Random Forest*

Segundo Breiman (2001), responsável pela disseminação da técnica, o *Random Forest* possui a capacidade de gerar resultados robustos, de maneira rápida e com bom desempenho, em comparação a outros algoritmos preditivos. Para Oshiro (2013, p. 14-15), o *Random Forest* se configura em um método onde se:

[...] constrói muitas árvores de decisão que serão usadas para classificar um novo exemplo por meio do voto majoritário. Cada árvore de decisão usa um subconjunto de atributos selecionados aleatoriamente a partir do conjunto original, contendo todos os atributos.

Para aplicação do algoritmo foram selecionadas e tratadas 20 variáveis, sendo 1 variável dependente chamada de dano ambiental e 19 variáveis independentes (9 obtidas do SISAM, 8 obtidas do ROI e 2 obtidas do MapBiomias). Após tratamento dos dados aplicou-se o corte amostral em duas bases, na proporção de 70% para treino (97 observações) e 30% para teste (41 observações). Desta forma, aplica-se primeiramente o algoritmo nas observações de

treinagem e em seguida a testagem com intuito de identificar o nível de confiança e de confirmação dos resultados encontrados na fase do treino.

O corte amostral foi ainda utilizado para comparação de resultados, com base em dois modelos de aplicação do algoritmo *Random Forest*:

- a) Modelo I – análise aleatória, ou seja, tanto o treino quanto o teste podem possuir dados em qualquer data do período 2014 a 2018;
- b) Modelo II – análise cronológica, ou seja, o treino apresentou variáveis entre os anos 19/05/2014 a 07/09/2017 e o teste entre os anos 08/09/2017 a 15/10/2018.

A sustentação dos resultados dos Modelos I e II se pautaram pela avaliação da matriz de confusão, indicadores de desempenho preditivos e índices de avaliação da importância das variáveis.

5.2 Dano Ambiental – Variável Dependente

O dano ambiental é identificado como o quociente entre o atributo Área Queimada (AQ), quantidade de vegetação queimada em hectares e o atributo Tempo de Combate (TC), indicado pelo total de minutos empregados por brigadista para a debelação do incêndio. Os dados foram obtidos dos ROI's e nos informam que quanto maior for a razão entre AQ e TC, menor será a eficiência do combate aos incêndios e, conseqüentemente, maior o dano ambiental.

A junção dos dois atribuídos possui como objetivo, uma forma alternativa para a mensuração da eficiência do combate de incêndios florestais. Encontrou-se, portanto, respaldo nas obras de Lima e Soares (1992) e de Santos, Soares e Batista (2006) para formulação de classes de distribuição apresentadas na Tabela 1.

Tabela 1 - Classes do dano ambiental dos incêndios florestais da amostra.

Classe	Intensidade	AQ/TC (hectares/minuto)
I	Baixo	$Y \leq 0,0015$
II	Moderado	$0,0015 < Y \leq 0,0333$
III	Alto	$0,0333 < Y \leq 0,0833$
IV	Extremo	$0,0833 < Y$

Fonte: Dados da pesquisa.

As seguintes intensidades foram constatadas nos ROI's analisados, após identificação das classes do dano ambiental: 30,43% - Baixo (42), 48,55% - Moderado (67), 12,32% - Alto (17), 8,70% - Extremo (12).

5.3 Variáveis Independentes

Em relação as 19 variáveis independentes, são apresentados no Quadro 1 a origem de cada uma, juntamente da respectiva descrição elucidativa.

Quadro 1 - Variáveis Independentes da Amostra

(Continua)

Origem	Variáveis	Descrição
ROI	Município	Município atingido pelo incêndio florestal.
	Localidade	Tipo de localidade atingida pelo incêndio florestal.
	Causa	Causa identificada para a ocorrência do incêndio florestal.

(Conclusão)

	Identificação	Informa se houve identificação do proprietário do local atingido pelo incêndio florestal.
	Vegetação	Tipo de vegetação atingida pelo incêndio florestal.
	Mês	Mês da ocorrência do incêndio florestal.
	Detecção	Tipo de detecção do incêndio florestal.
	Horário da Detecção	Horário que foi realizado a detecção do incêndio florestal.
MapBiomias	Cobertura Agropecuária	Porção territorial identificada como agropecuária (em hectares).
	Cobertura Floresta Natural	Porção territorial identificada como formação florestal natural (em hectares).
SISAM	Concentração CO	Concentração de Monóxido de Carbono (Parte por bilhão - ppb).
	Concentração NO ₂	Concentração de Dióxido de Nitrogênio (Parte por bilhão - ppb).
	Concentração O ₃	Concentração de Ozônio (Parte por bilhão - ppb).
	Concentração PM _{2,5}	Concentração de Material Particulado Inalado Fino (micrograma por metro cúbico - µgm ³).
	Concentração SO ₂	Concentração de Dióxido de Enxofre (micrograma por metro cúbico - µgm ³).
	Temperatura	Temperatura do ar (em grau Celsius).
	Umidade	Umidade relativa do ar (em porcentagem).
	Direção do Vento	Direção do vento (em grau).
Velocidade do Vento	Velocidade do vento (em metro por segundo - m/s).	

Fonte: Dados da pesquisa.

Tanto as variáveis dependentes, quanto as variáveis independentes são elementos importantes para obtenção de resultados concisos e confiáveis. Estes elementos possuem alguma ligação com o incêndio florestal (direta ou indireta), seja das condições climáticas, topográficas ou características encontradas na ocorrência de incêndio.

6 ANÁLISE DE RESULTADOS

Nesta seção são apresentados os resultados obtidos após análise estatística descritiva dos ROI's e aplicação do algoritmo *Random Forest* nos modelos preditivos.

6.1 Estatística Descritiva Condensada dos ROI's

As estatísticas descritivas das variáveis utilizadas na aplicação metodológica do estudo, permite compreender as características da APA do Alto do Mucuri e, conseqüentemente, compartilhar o perfil de ocorrências de incêndios florestais da referida unidade de conservação. Este fato pode ser evidenciado inicialmente pelos atributos do dano ambiental. Foi verificado o total de 14.729,92 hectares de vegetação queimada, equivalentes a 20.630,14 campos de futebol (0,714 ha/campo). Essa área queimada representou 4,53% do território da APA do Alto do Mucuri. Em relação ao Tempo de Combate, foram empregadas 9.864,61 horas (591.877 minutos) no período de 2014 a 2018.

As maiores quantidades de ocorrências de incêndios ocorreram respectivamente nos anos 2014 (45), 2018 (37) e 2015 (36), sendo os municípios de Ladainha (56,52%) e Itaipé (23,19%) detentores dos maiores percentuais de ROI's registrados. Cerca de 90,58% (125) dos registros analisados ocorreram nas zonas rurais dos municípios. Segundo Medeiros e Fiedler (2004), Mistry e Bizerril (2011), os incêndios em áreas rurais podem estar relacionadas com atividades de renovação de pastagens, limpeza para agricultura, caça de animais, extrativismo ilegal, contrabando de madeira ou utilização na fabricação de carvão. Esta situação pode ser ressaltada na Figura 2, onde a atividade agropecuária obteve relevante resultado (28,26%), juntamente das causas desconhecidas (55,07%).

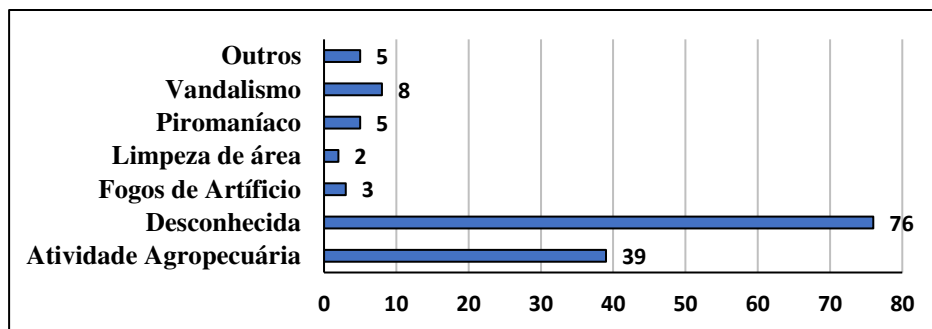


Figura 2 - Causas de incêndios na APA do Alto do Mucuri por ROI, no período de 2014 a 2018.
Fonte: Dados da pesquisa.

Em relação a vegetação desmatada, 94,93% (131) foram identificadas como Floresta Estacional Semidecidual, tipologia pertencente ao bioma Mata Atlântica e relatada por Pereira (2016) como espécie de vegetação abundante na região. Durante os cinco anos analisados foram encontrados uma taxa de 44,98% de identificação dos proprietários de terras onde ocorreram as queimadas, o que indica uma quantidade representativa de medidas preventivas aplicadas para evitar a reincidência de ocorrências (multas, conscientização, busca de infratores, etc).

O histórico mensal das ocorrências de incêndios florestais coincidiu com o período crítico relatado em literatura científica. Para Santos, Souto e Souto (2019), esse período crítico ocorre entre os meses de agosto a setembro, predominantemente de clima sazonal seco, o que favorece a incidência de queimadas. Na Figura 3 é possível perceber maiores índices de incêndios florestais entre os meses de agosto a novembro. Destaque para os meses de setembro e outubro que atingiram respectivamente os percentuais de 40,58% (56) e 35,51% (49), em relação as ocorrências de incêndios totais.

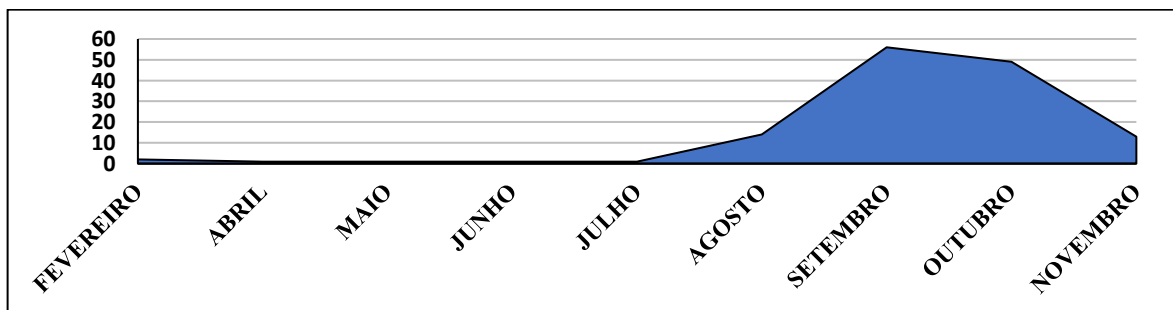


Figura 3 - Ocorrências mensais de incêndios na APA do Alto do Mucuri, no período de 2014 a 2018.
Fonte: Dados da pesquisa.

Ao longo do período foram observados evolução na participação da população local na identificação de incêndios florestais. O crescimento da participação representa aspecto positivo no engajamento da população residente nas questões ambientais da APA do Alto do Mucuri. O

crescimento é refletido nos índices anuais: 16,67% (2015), 15,38% (2016), 28,57% (2017) e 35,13% (2018).

6.2 Aplicação do algoritmo *Random Forest*

Para aplicação do algoritmo *Random Forest* foi necessário o tratamento dos dados, dividindo-os em dois cortes amostrais, na proporção de 70% para Treino (97 observações) e 30% para Teste (41 observações). Concluído esta etapa, foi necessário realizar ajustes para definição da quantidade de árvores aleatórias (ntree) e números de atributos utilizados em cada árvore de decisão (mtry) a ser utilizada no algoritmo. Por intermédio do R Studio, a melhor configuração obtida com os ajustes foram ntree com 300 árvores e mtry com 4 tipos de atributos. Com a compilação dos dados organizados, foi possível executar a aplicação e interpretação dos modelos preditivos demonstrados nos próximos tópicos.

6.2.1 Indicadores para interpretação dos Modelos I e II

Considerou-se três indicadores de classificação para interpretação dos Modelos I (aleatório) e II (cronológico) na aplicação do algoritmo *Random Forest* (Quadro 2).

Quadro 2 - Indicadores de desempenho preditivo.

Indicador	Equação	Descrição
Acurácia	$A = \frac{VP + VN}{VP + FN + FP + VN}$	Taxa de acertos gerados pelo modelo.
Sensibilidade	$S = \frac{VP}{VP + FN}$	Probabilidade de detecção de uma característica que realmente contém no objeto analisado.
Especificidade	$E = \frac{VN}{FP + VN}$	Probabilidade de detecção da ausência de uma característica em um objeto que realmente não possua determinada característica

Fonte: Adaptado de (BARBOSA, 2017), (BARROS, 2019).

Os indicadores são extraídos da matriz de confusão, esta “[...] representada pela tabulação cruzada de classes observadas e previstas para observações do conjunto de dados de teste [...]” (SANTOS, 2018, p.30). No Quadro 3 é retratado uma matriz de confusão binária.

Quadro 3 - Matriz de confusão para análise de problemas classificatórios binários.

		Previsão	
		Sim	Não
Situação Real	Sim	Verdadeiro Positivo (VP)	Falso Positivo (FP)
	Não	Falso Negativo (FN)	Verdadeiro Negativo (VN)

Fonte: Santos (2018), adaptado pelo autor.

Para a análise apresentada, o desejável são as situações consideradas como Verdadeiros Positivos (VP) e Verdadeiros Negativos (VN), ou seja, significa os acertos preditivos gerados no modelo. Nos próximos tópicos são desenvolvidos estes conceitos para compreensão dos resultados.

6.2.2 Modelo I – *Random Forest* Aleatório

A partir do corte amostral 70% para treino e 30% para teste, chegou-se nos resultados elencados na Tabela 2 para o Modelo I. Os valores sombreados da matriz de confusão indicam os acertos preditivos do modelo. Observou-se na treinamento da amostra o total de 49 (50,51%) classificações identificadas como corretas, sendo 43 (82,7%) para predição do dano ambiental do tipo moderado (em relação aos demais valores preditos como moderado).

		Referência							
		Treino				Teste			
		Baixo	Moderado	Alto	Extremo	Baixo	Moderado	Alto	Extremo
Predição	Baixo	6	21	0	0	4	0	0	1
	Moderado	5	43	3	1	11	15	7	2
	Alto	0	9	0	0	0	0	1	0
	Extremo	3	6	0	0	0	0	0	0
		Indicadores Preditivos							
		Baixo	Moderado			Alto			Extremo
Sensibilidade		0,2667	1,0000			0,1250			0,0000
Especificidade		0,9615	0,2308			1,0000			1,0000
Acurácia Balanceada		0,6141	0,6154			0,5625			0,5000

Tabela 2 - Matriz de confusão e Indicadores Preditivos – Modelo I.

Fonte: Dados da pesquisa.

Para validação dos resultados, foi feita a treinamento dos valores, resultando em 20 classificações corretas. Em vista a classificação correta, houve a acurácia de 48,78% dos valores, sendo a taxa de acerto de 80% para a predição de dano ambiental baixo e 44,12% para predição de dano ambiental moderado. A treinamento apresentou melhor predição para o dano ambiental baixo e não houve predição considerada como correta para o dano do tipo extremo.

Em relação aos indicadores preditivos, os resultados apontam serem satisfatórios para sensibilidade para dano do tipo moderado (100%), especificidade para dano do tipo baixo (96,15%), alto (100%) e extremo (100%). Os altos níveis de especificidade estão associados ao alto índice de assertividade quanto a variáveis consideradas ausentes no modelo. A acurácia balanceada (média dos resultados da sensibilidade e especificidade) indicam os danos do tipo baixo e moderado como melhores desempenhos preditivos.

6.2.3 Modelo II – *Random Forest* Cronológico

Os resultados do Modelo II podem ser considerados os que mais aproximam da realidade, uma vez que são temporais, ou seja, o algoritmo realiza o cálculo à medida que os eventos são registrados. Os resultados podem ser observados na Tabela 3.

		Treino				Teste			
		Baixo	Moderado	Alto	Extremo	Baixo	Moderado	Alto	Extremo
Predição	Baixo	9	15	0	1	11	9	0	1
	Moderado	9	34	5	1	6	9	3	2
	Alto	0	12	2	0	0	0	0	0
	Extremo	3	6	0	0	0	0	0	0

Indicadores Preditivos				
	Baixo	Moderado	Alto	Extremo
Sensibilidade	0,6471	0,5000	0,0000	0,0000
Especificidade	0,5833	0,5217	1,0000	1,0000
Acurácia Balanceada	0,6152	0,5109	0,5000	0,5000

Tabela 3 - Matriz de confusão e Indicadores Preditivos – Modelo II.

Fonte: Dados da pesquisa.

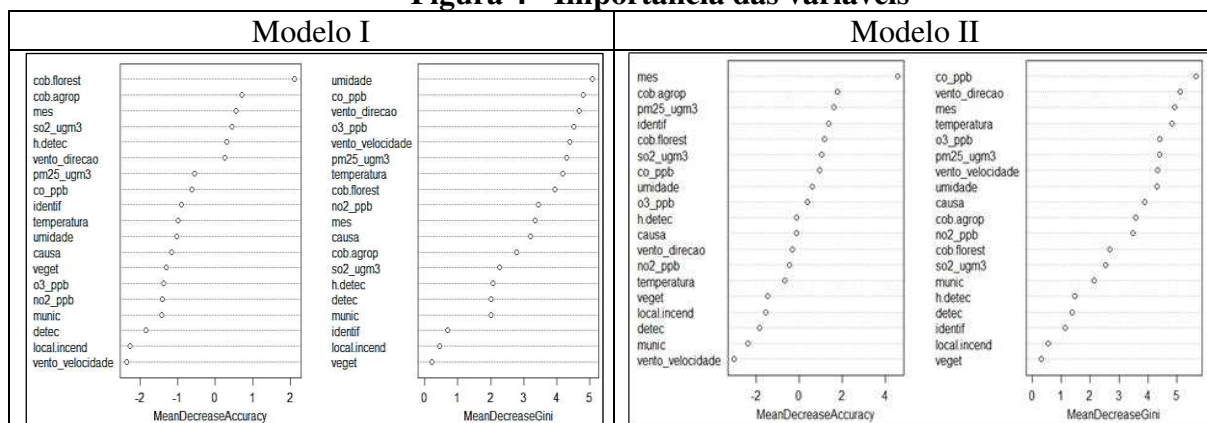
Na fase do treino, a matriz de confusão obteve 45 (46,49%) predições consideradas como corretas. A taxa de acerto com maior percentual foi destinada ao dano ambiental do tipo moderado (69,39%). Já na fase da testagem, o modelo gerou 20 predições consideradas como corretas, o que resultou em uma taxa de acurácia de 48,78%, igualmente encontrada no modelo I. Destes acertos, 52,38% foram preditas para o dano do tipo baixo e 45% para o dano do tipo moderado (percentuais com base nos valores totais preditos de cada classe).

Os indicadores de robustez da predição revelaram melhores resultados de acurácia balanceada para os danos do tipo baixo (61,52%) e moderado (51,09%), ou seja, as referidas classes de dano ambiental apresentaram maiores forças preditivas em relação as demais. Ressalta-se que a sensibilidade gerou resultados em 0% para os danos do tipo alto e extremo e, portanto, a combinação predita pode ter influenciado nos altos índices de especificidade para as mesmas classes. Esta situação demonstra uma especificidade com predição correta de 100% para ausência de valores das classes.

6.3 Importância das Variáveis

A análise procurou relacionar as variáveis com maiores influências para a predição do dano ambiental. Foram utilizados dois parâmetros para avaliação de desempenho: *mean decrease accuracy* (MDA) e o *mean decrease Gini* (MDG) (Figura 4). O MDA ou redução média da acurácia é definido como a perda de acurácia quando os valores de determinada variável sofrem alterações de categorias e o MDG ou redução média Gini, utiliza o critério de impureza Gini na elaboração das árvores de decisão aleatórias presentes no *Random Forest* (PALCZEWSKA *et al.*, 2014).

Figura 4 - Importância das variáveis



Fonte: Elaborado e calculado no *software* R Studio versão 1.3.1073.

É possível perceber na Figura 4 que as variáveis com maiores influências nos modelos apresentados possuem marcadores posicionados para o lado direito do quadro. Verifica-se, ao fazer a confrontação das cinco primeiras variáveis de cada parâmetro entre os Modelos I e II, a seguinte apuração da importância:

- MDA – Cobertura Agropecuária (cob.agrop), Mês (mes) e Cobertura Floresta Natural (cob.florest);
- MDG – Monóxido de Carbono (co_ppb), Direção do Vento (vento_direcao) e Concentração de Ozônio (o3_ppb).

As variáveis relatadas possuem maior influência entre as demais, para predição das classes de dano ambiental. Desta forma, as oscilações de resultados das referidas variáveis terão maior impacto na definição de determinado tipo de dano ambiental, em relação as demais variáveis com níveis de importância menores.

6.4 Discussão dos Modelos I e II

Os Modelos I e II convergiram características que são necessárias para compreensão e adoção de medidas na gestão da APA do Alto do Mucuri. Ambos modelos alcançaram acurácia de predição de dano ambiental em 48,78% na fase de teste, refletindo em uma capacidade preditiva melhor para o dano do tipo baixo e moderado. Sendo assim, caso o algoritmo *Random Forest* venha classificar determinada ocorrência de incêndio como dano do tipo baixo ou moderado, há maiores chances de a predição estar correta, se comparado a classificação de dano ambiental do tipo alto ou extremo.

Na prática, significa dizer que ao tomar conhecimento das características de um incêndio florestal, o gestor da unidade de conservação poderá utilizar o algoritmo *Random Forest* para realizar a predição da possível dimensão do dano ambiental, e por conseguinte, estabelecer a alocação de recursos necessários para o combate do incêndio florestal. Desse modo, a previsão auxiliará na tomada de decisões de maneira eficiente, uma vez que o conhecimento da proporção do incêndio florestal propiciará a distribuição de recursos humanos, materiais, equipamentos, transporte com maior efetividade e menor gasto público.

Ressalta-se que o algoritmo *Random Forest*, por ser uma técnica de aprendizado de máquina, tende a gerar resultados cada vez mais assertivos em relação ao poder preditivo, à medida que a sua base de dados seja alimentada. A aprendizagem de máquina executa decisões baseadas no sucesso de experiências anteriores (MONARD; BARANAUSKAS, 2003). Isso significa que a acurácia dos Modelos I e II obterá evolução positiva de resultados com o acréscimo contínuo dos ROI's na compilação de dados a serem aplicados pelo algoritmo *Random Forest*.

Quanto a investigação da importância das variáveis, verificou-se que as variáveis de destaque são principalmente do grupo: condições climáticas, poluentes atmosféricos e dados de cobertura agropecuária e floresta natural. Importante mencionar que os referidos grupos de informações não são encontrados nos atuais formulários de preenchimento do ROI. Logo, sugere a reformulação do registro de ocorrência de incêndio, no intuito de incorporar esses elementos relevantes para o perfil de incêndio.

7 CONSIDERAÇÕES FINAIS

Com o objetivo de analisar as contribuições preditivas do dano ambiental para a gestão pública no combate a incêndios florestais, o artigo abordou a utilização do algoritmo *Random Forest* como potencial ferramenta para gestão da APA do Alto do Mucuri. Assim sendo, foi utilizado como base amostral, os ROI's apresentados pela unidade de conservação entre os anos 2014 a 2018.

Por intermédio dos ROI's, foi traçado o perfil dos incêndios florestais da APA do Alto do Mucuri. Foram identificados o registro de 14.729,92 hectares de vegetação queimada, sendo principalmente do tipo Floresta Estacional Semidecidual (bioma Mata Atlântica). Sobre os municípios com maiores incidências de queimadas, foram destacados Ladainha (56,52%) e Itaipé (23,19%). Estas, são localidades totalmente inseridas dentro da área da APA do Alto do Mucuri e que demandam maior atenção para medidas preventivas deste tipo de evento.

Durante os cinco anos estudados, observou-se um período crítico de incêndios florestais entre os meses de agosto a novembro. A principal causa relatada das ocorrências dos eventos foram desconhecidas (55,07%) e atividade agropecuária (28,26%). Segundo Ramos, Fonseca e Morello (2016), normalmente há a contratação de brigadistas para o alívio da demanda de combate de incêndios durante o período crítico.

Sobre a aplicação do *Random Forest* nos Modelos I (aleatório) e II (cronológico), verificou-se acurácia de 48,78% em ambos, com melhores capacidades preditivas para o dano ambiental do tipo baixo e moderado. Os fatores determinantes para ocorrência de incêndio foram medidos em dois parâmetros: o MDA, que resultou nas variáveis mais relevantes: Cobertura Agropecuária, Mês e Cobertura Floresta Natural e o MDG, com as variáveis Monóxido de Carbono, Direção do Vento e Concentração de Ozônio, como as mais importantes do modelo.

Foram observados que as variáveis mais importantes para a predição do dano ambiental não são encontradas nos atuais formulários ROI's da APA do Alto do Mucuri. Esses elementos são fatores necessários para melhor compreensão das queimadas e, portanto, sugere-se a reformulação do formulário ROI para inclusão dessas características nas ocorrências.

A utilização do algoritmo *Random Forest* na administração pública servirá para ajudar a definir melhores decisões a serem tomadas diante dos incêndios florestais. Em situação hipotética, entre duas ocorrências de incêndios florestais, o gestor da unidade de conservação poderá alocar um número maior de brigadistas, equipamentos, transportes no evento que demonstrar maior dimensão de dano ambiental, após classificação de dano ambiental pelo algoritmo *Random Forest*. Desta forma, a tendência é a melhor distribuição de recursos disponíveis e adoção de logísticas com maior eficiência de combate de queimadas.

Os resultados alcançados são positivos e tendem a ser aperfeiçoados, conforme a base de dados é alimentada. Ressalta-se que a análise de predição de dano ambiental pode ser estendida as demais unidades de conservação, demonstrando flexibilidade na utilização de mais ou menos variáveis. Como delimitação do estudo, não foram encontrados dados confiáveis para a utilização de outros fatores que poderiam colaborar no desempenho dos modelos preditos, como por exemplo, dados de precipitação e temperatura.

Por fim, o artigo propõe a adoção da técnica de aprendizado de máquina para administração pública de unidades de conservação, sobretudo, para o combate de incêndios florestais. Para isso, é recomendado o registro contínuo dos ROI's em uma base de dados sistemática e online com amplo acesso das partes interessadas, interligada a um *software* de fácil manuseio para utilização do algoritmo *Random Forest*.

REFERÊNCIAS

ANDERSON, L. O. *et al.* *Conceptual model of disaster risk management and warning system associated with wildfires and public policy challenges in Brazil*. **Territorium: Revista Portuguesa de riscos, prevenção e segurança**, n. 26 (I), p. 43–61, 2019. Disponível em: <<https://digitalis.uc.pt/handle/10316.2/46128>>. Acesso em: 28 abr. 2019.

BARBOSA, J. H. DE F. *Early Warning System para Distress Bancário no Brasil*. 2017. 210f. **Tese (Doutorado) – Faculdade de Economia, Administração e Contabilidade (FACE)**. – UnB, Brasília, 2017.

BARROS, C. B. Classificadores de Regressão Logística, *Naive Bayes* e *Random Forest* na análise do Tropismo do HIV-1 de subtipo B. **Dissertação de mestrado em Engenharia Biomédica**, COPPE, UFRJ: Rio de Janeiro, 2019. Disponível em: <<https://pantheon.ufrj.br/bitstream/11422/13240/1/CesarBorgesBarros.pdf>>. Acesso em: 19 dez. 2020.

BONTEMPO, G. C. *et al.* Registro de Ocorrência de Incêndio (ROI): evolução, desafios e recomendações. **Biodiversidade Brasileira**, n. 2, p. 247-263–263, 30 dez. 2011. Disponível em: <<http://www.icmbio.gov.br/revistaeletronica/index.php/BioBR/article/view/108>>. Acesso em: 20 abr. 2019.

BRASIL. Lei 9.985, de 18 de julho de 2000 (Sistema Nacional de Unidades de Conservação). 2000. Disponível em: <http://www.planalto.gov.br/ccivil_03/LEIS/L9985.htm>. Acesso em: 13 abr. 2019.

BREIMAN, L. *Random Forests*. **Machine Learning**, v. 45, n. 1, p. 5–32, 1 out. 2001. Disponível em: <<https://doi.org/10.1023/A:1010933404324>>. Acesso em: 15 abr. 2019.

CUNHA, A. L.; OLIVEIRA, A. P. C.; BARBALHO, M. G. S. UNIDADE DE CONSERVAÇÃO AMBIENTAL. **Anais SNCMA**, v. 8, n. 1, 14 nov. 2017. Disponível em: <<http://anais.unievangelica.edu.br/index.php/sncma/article/view/54>>. Acesso em: 12 abr. 2019.

FAHEL, M. C. X.; LEITE, G. P.; TELES, L. R. Pobreza Multidimensional no estado de Minas Gerais: uma mensuração para além da renda. **Revista Brasileira de Monitoramento e Avaliação**, v. 8, p.50-69, 2014. Disponível em: <<https://www.rbaval.org.br/article/10.4322/rbma201408004/pdf/1598015026-8-50.pdf>>. Acesso em: 21 nov. 2020.

FELIPPE, M. F.; BUENO, J.; COSTA, A. Desmatamento na Bacia do Rio Mucuri (MG), Brasil no período de 1989 a 2008: uso de imagens Cbers e Landsat na espacialização dos remanescentes florestais. **Simpósio Brasileiro De Sensoriamento Remoto (SBSR)**, v. 16, p. 2713–2720, 2009.

GODOY, L. R. DA C.; LEUZINGER, M. D. O financiamento do Sistema Nacional de Unidades de Conservação no Brasil Características e tendências. **Revista de Informação Legislativa**, v. 52, n. 206, p. 223, 2015. Disponível em:

<https://www12.senado.leg.br/ril/edicoes/52/206/ril_v52_n206_p223.pdf>. Acesso em: 14 abr. 2019.

IBAMA. Perfil dos Incêndios Florestais Acompanhados pelo Ibama. Prevfogo: 2010. Disponível em: <http://www.ibama.gov.br/phocadownload/prevfogo/relatorios/incendios_florestais/prevfogo-incendios-florestais-perfil_dos_incendios_2009.pdf>. Acesso em: 20 abr. 2019.

IBAMA. Relatório de ocorrências de incêndios em Unidades de Conservação Federais 2005-2008. Prevfogo: 2009. Disponível em: <http://queimadas.cptec.inpe.br/~rqueimadas/material3os/2009_Ibama_Relatorio_FogoUCs_2005_08_DE3os.pdf>. Acesso em: 20 abr. 2019.

LIMA, G. S. *et al.* Avaliação da eficiência de combate aos incêndios florestais em unidades de conservação brasileiras. **Floresta**, v. 48, n. 1, p. 113–122, 2018. Disponível em: <<https://revistas.ufpr.br/floresta/article/view/53550>>. Acesso em: 20 abr. 2019.

LIMA, G. S.; SOARES, R. V. Avaliação de eficiência de combate aos incêndios florestais no Brasil. **Floresta**, v. 22, n. 1/2, p. 25-38, 1992.

MEDEIROS, M. B. DE; FIEDLER, N. C. Incêndios florestais no parque nacional da Serra da Canastra: desafios para a conservação da biodiversidade. **Ciência Florestal**, v. 14, n. 2, p. 157–168, 2004. Disponível em: <<https://periodicos.ufsm.br/index.php/cienciaflorestal/article/viewFile/1815/1079>>. Acesso em: 20 abr. 2019.

MINAS GERAIS. Decreto nº 45 .877, de 30 de dezembro de 2011. 2011. Disponível em: <<https://www.almg.gov.br/consulte/legislacao/completa/completa-nova-min.html?tipo=DEC&num=45877&comp=&ano=2011&texto=original>>. Acesso em: 19 abr. 2019.

MISTRY, J.; BIZERRIL, M. Por que é importante entender as inter-relações entre pessoas, fogo e áreas protegidas? **Biodiversidade Brasileira**, v. 0, n. 2, p. 40-49–49, 30 dez. 2011. Disponível em: <<http://www.icmbio.gov.br/revistaeletronica/index.php/BioBR/article/view/137>>. Acesso em: 20 abr. 2019.

MONARD, M. C.; BARANAUSKAS, J. A. Conceitos sobre aprendizado de máquina. **Sistemas inteligentes-Fundamentos e aplicações**, v. 1, n. 1, p. 32, 2003.

OSHIRO, T. M. Uma abordagem para a construção de uma única árvore a partir de uma *Random Forest* para classificação de bases de expressão gênica. 2013. **Dissertação de Mestrado**, Universidade de São Paulo, 2013. Disponível em: <<http://www.teses.usp.br/teses/disponiveis/95/95131/tde-15102013-183234/>>. Acesso em: 14 abr. 2019.

PALCZEWSKA, A. *et al.* *Interpreting random forest classification models using a feature contribution method*. In: BOUABANA-TEBIBEL, T.; RUBIN, S. H. **Integration of reusable systems**. [S.l.]: Springer, 2014. p. 193–218.

PECCATIELLO, A. F. O. Políticas públicas ambientais no Brasil: da administração dos recursos naturais (1930) à criação do Sistema Nacional de Unidades de Conservação (2000). **Desenvolvimento e Meio Ambiente**, v. 24, n. 0, 28 dez. 2011. Disponível em: <<https://revistas.ufpr.br/made/article/view/21542>>. Acesso em: 7 abr. 2019.

PEREIRA, J. M. Políticas públicas florestal e de proteção à biodiversidade em prol da APA do Alto do Mucuri. 2016. **Dissertação de Mestrado em Tecnologia, Ambiente e Sociedade** – UFVJM, Teófilo Otoni - MG, 2016. Disponível em: <https://sucupira.capes.gov.br/sucupira/public/consultas/coleta/trabalhoConclusao/viewTrabalhoConclusao.jsf?popup=true&id_trabalho=5393296#>. Acesso em: 19 abr. 2019.

PEREIRA, R. S. Avanços e desafios em áreas protegidas: um estudo da APA do Alto do Mucuri no município de Ladainha -MG. 2018. **Dissertação de Mestrado em Administração Pública** – UFVJM, Teófilo Otoni - MG, 2018. Disponível em: <https://sucupira.capes.gov.br/sucupira/public/consultas/coleta/trabalhoConclusao/viewTrabalhoConclusao.jsf?popup=true&id_trabalho=7053229#>. Acesso em: 19 abr. 2019.

PRAES, E. O. Código florestal brasileiro: evolução histórica e discussões atuais sobre o novo código florestal. **VI Colóquio Internacional "Educação e Contemporaneidade"**, São Cristovão - SE, p. 14, 2012. Disponível em: <http://educonse.com.br/2012/eixo_19/PDF/20.pdf>. Acesso em: 08 abr. 2019.

RAMOS, R. M.; FONSECA, R. L.; MORELLO, T. F. Unidades de conservação e proteção contra incêndios florestais: relação entre focos de calor e ações articuladas pelas brigadas contratadas. **Revista Biodiversidade Brasileira**, n.2, 2016. Disponível em: <<https://revistaeletronica.icmbio.gov.br/index.php/BioBR/article/view/558> >. Acesso em: 01 out. 2020.

SANTOS, H. G. DOS. Comparação da performance de algoritmos de *machine learning* para a análise preditiva em saúde pública e medicina. 2018. **Doutorado em Epidemiologia** – Universidade de São Paulo, São Paulo, 2018. Disponível em: <<http://www.teses.usp.br/teses/disponiveis/6/6141/tde-09102018-132826/>>. Acesso em: 8 maio 2019.

SANTOS, J. F.; SOARES, R. V.; BATISTA, A. C. Perfil dos incêndios florestais no Brasil em áreas protegidas no período de 1998 a 2002. **Floresta**, v. 36, n.1, p. 93-100, 2006. Disponível em: <<https://revistas.ufpr.br/floresta/article/view/5510/4040>>. Acesso em: 26 set. 2020.

SANTOS, W. S. S.; SOUTO, P. C.; SOUTO, J. S. Riscos de incêndios florestais em Unidades de Conservação no bioma Mata Atlântica, Paraíba, Brasil. **Agropecuária Científica no Semiárido**, V. 14, N. 4, 2018. Disponível em: <<http://revistas.ufcg.edu.br/acsa/index.php/ACSA/article/view/1030/pdf>>. Acesso em: 01 out. 2020.