

# REDE NEURAL ARTIFICIAL EM SIMULAÇÃO DA OCORRÊNCIA DE FOGO NO CERRADO

Letícia Meirelles; Manoel Cardoso

Instituto Nacional de Pesquisas Espaciais / Centro de Ciência do Sistema Terrestre – INPE/CCST

leticia.meirelles@inpe.br; manoel.cardoso@inpe.br

Trabalho apresentado à disciplina Análise de Algoritmo e Estrutura de Dados da UNIFESP, como requisito parcial de avaliação.

Prof Lilian Berton.

Orientador: Manoel Cardoso

## I. INTRODUÇÃO

Existe grande interesse na preservação do Cerrado devido principalmente à sua alta biodiversidade (Myers et al 2000) e potencial de recursos naturais (Nóbrega et al. 2017). Apesar disso, o bioma deverá sofrer com pressões devido à mudanças no clima e em seu uso e cobertura, como a produção de soja e milho, e pastagens (Miranda et al. 2014). Assim, a capacidade de simular a dinâmica do Cerrado pode contribuir em tomadas de decisões sobre sua preservação. Isto inclui o desenvolvimento de métodos para simular a ocorrência do fogo, o que têm uma forte influência na dinâmica da vegetação desta região (Arruda et al. 2018).

Neste contexto, esse estudo desenvolve uma aplicação de redes neurais artificiais (Werbos 1975) para simular a ocorrência do fogo a partir de variáveis ambientais calculadas pelo Integrated Land Surface Model (INLAND, baseado em Foley et al. 1998). O INLAND oferece a capacidade de simular a dinâmica do Cerrado sob diferentes condições climáticas e de uso na terra, e explorar mecanismos de interação entre a atmosfera e a superfície da região.

A presente metodologia e objetivos se concentram em validar a possibilidade de uso de redes neurais artificiais como suporte de cálculo da ocorrência natural de fogo a partir de saídas do INLAND em simulações da dinâmica de vegetação, pelas informações que o modelo provém sobre biomassa vegetal e umidade do solo. Além destas informações, são também utilizados dados meteorológicos de precipitação e velocidade do vento, e detecções de fogo pelo sensoriamento remoto com o Moderate Resolution Imaging Spectroradiometer (MODIS).

## II. METODOLOGIA

Como ferramentas principais para implementação da rede neural foram utilizados resultados do modelo INLAND e dados de focos de queimadas do produto MODIS MCD14ML (University of Maryland, EUA, <http://modis-fire.umd.edu>). O INLAND é desenvolvido no CCST/INPE com base em Foley et al. (1996), e simula as características dominantes da superfície vegetada continental, com resolução espacial de 50km, e em escala temporal de dias-anos. O modelo integra

processos superficiais que representam os efeitos do clima na superfície (como crescimento ou mortalidade da vegetação), e da superfície no clima (como evapotranspiração e reflexão da energia solar), e permite considerar cenários de desmatamento e agricultura.

Para obtenção dos resultados com o INLAND, realizou-se uma simulação no domínio de território do Brasil para um período total de 110 anos. A fim de se evitar efeitos de condições iniciais, a extração de valores simulados se concentrou nos últimos 10 anos apenas. A configuração desta simulação contempla a dinâmica de vegetação, concentração de CO<sub>2</sub> estática, e dinâmica de uso da terra desativada. Os dados de entrada sobre condições atmosféricas correspondem a uma climatologia de trinta anos (1961 a 1990) gerada a partir de dados do CRU (Climate Research Unit, Global Climate Dataset).

### A. Algoritmo Implementado

Para o desenvolvimento da Rede Neural Artificial (RNA) multicamadas no cálculo de ocorrência de fogo foi utilizado a linguagem de programação Python, treinadas com o algoritmo backpropagation (Werbos, 1975). A topologia da rede a ser utilizada consistem em 3 camadas (camada de entrada, camada oculta e camada de saída). Foram considerados 11 neurônios para as camadas de entrada; 2,3,5,10 e 100 neurônios na camada oculta; e 1 neurônio para a camada de saída.

O algoritmo de rede neural artificial multiplas camadas com retroalimentação consiste na inicialização dos pesos de cada neurônio com valores aleatórios, repete o treinamento da série de amostras até que as saídas satisfaçam o critério de parada, a partir do erro quadráticos médio ou atinja o número limite de interações. Para cada elemento da amostra é calculada a saída, e comparada com a esperada, a diferença entre o erro e o valor simulado é utilizado para ajuste dos pesos.

Algoritmo BackPropagation:  
begin

```
  Inicia todas conexões com valores de pesos aleatórios
do if erro <  $\epsilon$   $\vee$   $nEpoca > maxEpoca$  then exit fi;
   $\epsilon$  é média do erro quadrático
for  $a : k$  to  $nAmostrasstep1do$ 
```

```

 $x := amostra[0 : k]$ 
 $y := amostra[1 : k]$ 
for  $i : 1$  to  $nCamada$  step 1 do
  for  $j : 1$  to  $nNeuronio$  step 1 do
     $fij(x)$  calcula saída
     $erro := y - fij(x)$ 
  end
end
for  $i : nCamada$  to 1 step 1 do
  for  $j : 1$  to  $nNeuronio$  step 1 do
     $atualizaPesos(erro)$ 
  end
end

```

A Figura 1 apresenta o modelo geral de um neurônio artificial.

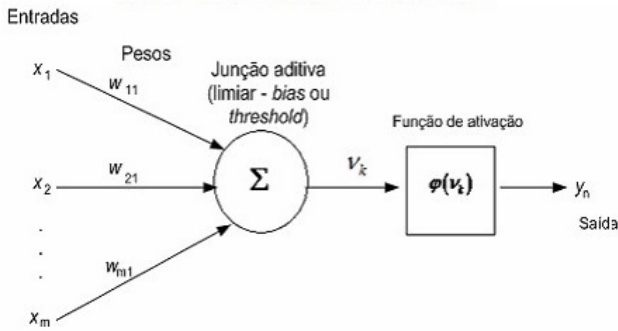


Figura 1. Modelo geral de um Neurônio Artificial (HAYKIN, 2001)

O primeiro da Figura 1 elemento é um conjunto de sinapses ou linha de conexão, cada qual caracterizada por um peso  $w_{kj}$ , com  $k = 1$  e  $j = 1, 2, \dots, m$ .

O segundo item da Figura 1 trata da soma dos sinais de entrada, ponderada pelas sinapses do neurônio (combinação linear), onde:

$$V_k = \sum_{j=0}^m W_{kj} X_j$$

A função de ativação, Figura 1, restringe a amplitude de saída de um neurônio, limitando o intervalo de saída para um valor finito.

$$Y_k = \varphi(v_k)$$

A equação da função de ativação, para este trabalho usada, tangente hiperbólica assumindo valores entre 1 e -1, onde  $\alpha$  é o parâmetro de inclinação da curva,  $\beta$  são os limites inferiores e superiores e  $\nu$  é o valor de ativação, deste modo:

$$f(v) = a \frac{e^{(bv)} - e^{(-bv)}}{e^{(bv)} + e^{(-bv)}}$$

Para a treinamento da RNA, o total dos dados de amostras foi dividido aleatoriamente em 70% para o treinamento, 15% para testes e 15% para validação. Em cada treinamento foram realizadas até 1440 interações da série de amostras de dados, até que o erro médio quadrático entre a saída desejada e a saída calculada fosse menor que 0.01 e tenha atingido o

limite de 1000 treinamentos da série de dados.

### B. Análise de Complexidade do Algoritmo

Esta seção apresenta uma análise computacional e de complexidade de mensagens para do algoritmo de rede neural perceptron multiplas camadas (MPL).

1) *Complexidade de Espaço*: A complexidade de espaço em memória de uma rede neural do tipo multiplas camadas é insignificante na maioria das vezes. Em um extremo, uma única interação pode abrigar um neurônio e, portanto, precisa armazenar as estruturas de dados, incluindo entradas, que são saídas de todos os outros neurônios, parâmetros e modelo computacional do neurônio, que coletivamente exigem alocação de espaço de memória.

A estrutura de dados mais dispendiosa é a matriz de entrada de neurônios unidimensional, que crescerá linearmente no número de neurônios, que enviam suas saídas como entradas para um determinado neurônio. Portanto, a complexidade espacial do caso de um neurônio por interação é linear no número de neurônios conectados ao único neurônio em consideração. Se houver múltiplos neurônios alojados por uma única entrada, a complexidade do espaço ainda é linear, pois a complexidade do espaço requisitado a um único neurônio é multiplicada pelo número de neurônios alojados naquela única interação para determinar a complexidade do espaço resultante.

2) *Complexidade de Tempo*: A complexidade de tempo de uma única interação depende da estrutura da rede neural. Para o algoritmo MLP, o tempo é dado pelas multiplicações de matrizes de entradas das amostras da série e neurônios de cada camada. Considerando a complexidade de um algoritmo para a multiplicação de matrizes, e D e seja o tamanho da lista de entradas. Em uma única camada com dimensão de entrada N dimensão de saída M, as propagações diretas e reversa serão sempre  $O(N \cdot M \cdot D)$  assumindo um algoritmo de produto de matriz. Soma-se isto para cada camadas, então pode-se obter o tempo para o calculo de uma interação de um algoritmo de backpropagation. Como a diferenciação em modo reverso é, no máximo, um fator constante mais lento do que o cálculo direto da função de saída da rede neural, todas as outras operações necessárias podem ser calculadas em tempo linear para a série de interações.

### C. Dados de Input

Depois da execução da simulação do INLAND, foi realizado o recorte para a região da Serra do Espinhaço-MG, para longitude -44.75 até -43.25 (3 pontos de grade) e latitude -18.25 até -17.25 (4 pontos de grade), escolhido como uma região representativa das condições do Cerrado. A sub-região de estudo totaliza uma área de 12 pontos de grade, onde foram extraídas saídas com frequência mensal. O número total de amostras é equivalente à  $10 \cdot 12 \cdot 12 = 1440$ . Como entradas para a rede neural, foram consideradas para cada amostra as seguintes variáveis: umidade do solo, biomassa total acima do solo, velocidade do vento na superfície, temperatura próxima à superfície, e precipitação média.

Após o recorte da região, e do período e variáveis das saídas do INLAND foi necessário gerar a matriz de entradas da rede e saídas esperadas. Para cada amostra foi verificado a saída desejada considerando os dados do MODIS para o mês equivalente, colocados em ordem aleatória para prevenção de tendências associadas à ordem de apresentação dos dados. Os dados foram ainda pré-processados através de normatizações e conversões de formato para torná-los mais apropriados à sua utilização na construção da rede.

As Figuras 2,3,4,5,6,7 e 9 apresentam as medias dos valores de dados de saídas do INLAND que foram utilizados como entrada para o treinamento da RNA.

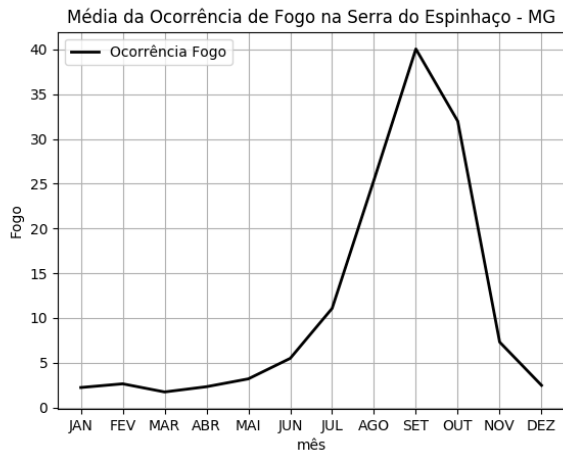


Figura 2. Climatologia da ocorrência de fogo entre 2002 a 2017 na região da Serra do Espinhaço - MG

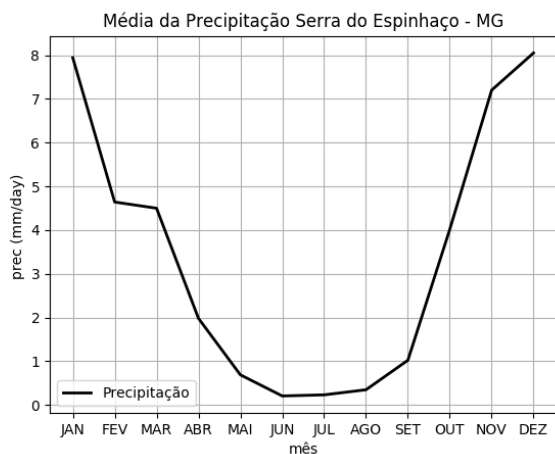


Figura 3. Climatologia 10 anos da precipitação (mm/dia) para região da Serra do Espinhaço - MG

O material combustível florestal pode ser definido como qualquer material orgânico vivo ou morto, no solo e acima dele, capaz de entrar em ignição e queimar, entretanto a matéria verde (madeira e folhas de vegetação viva) possui

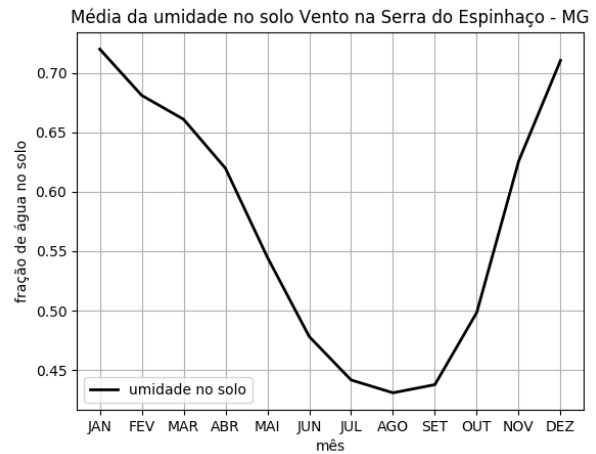


Figura 4. Climatologia 10 anos da umidade no solo (fração) para região da Serra do Espinhaço - MG

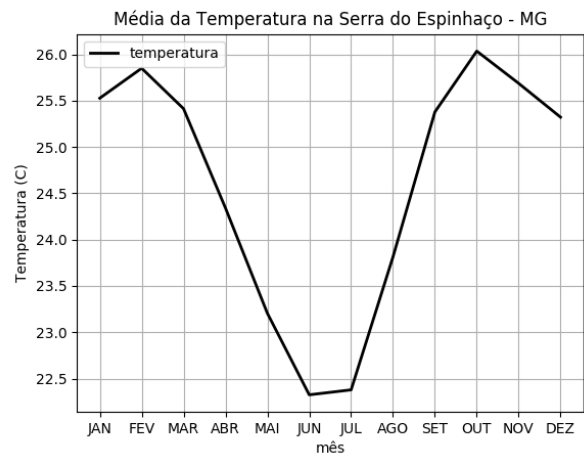


Figura 5. Climatologia 10 anos da temperatura na superfície (C) para região da Serra do Espinhaço - MG

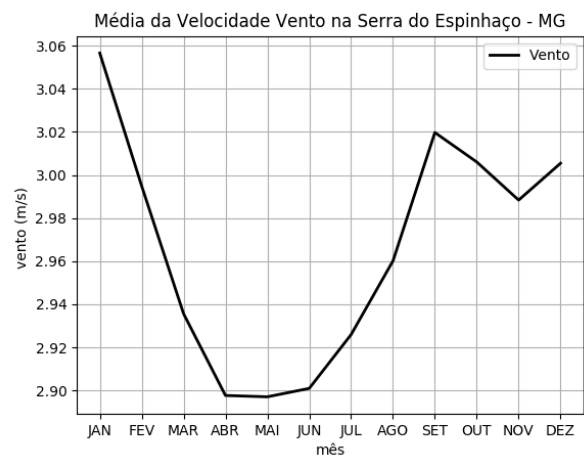


Figura 6. Climatologia 10 anos da velocidade no vento (m/s) para região da Serra do Espinhaço - MG

uma difícil ignição pois possuem alto teor de umidade. Deste modo, para modelagem da RNA foi considerada apenas as informações de biomassa de serrapilheira, para a análise de dados de entrada da RNA, figura 7.

A Figura 8, apresenta a comparação entre as variáveis de

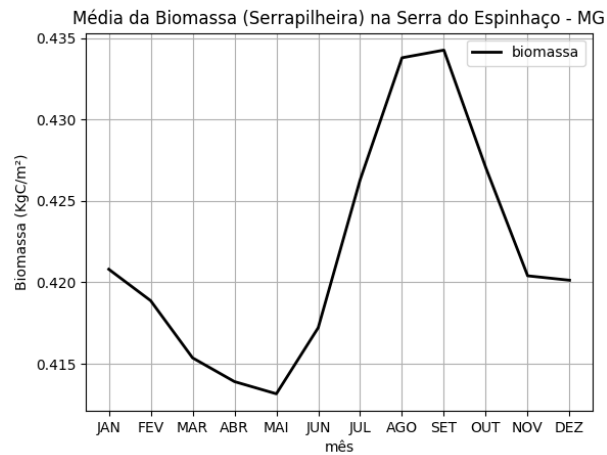


Figura 7. Climatologia 10 anos da Biomassa serrapilheira (Kg C/ m<sup>2</sup>) para região da Serra do Espinhaço - MG

entrada da RNA para as seguintes variáveis: precipitação e umidade no solo (que possuem correlação inversa com a ocorrência de fogo), temperatura e velocidade de vento (onde o crescimento de ocorrência de fogo coincidem com o aumento de valores de temperatura e velocidade de vento durante a seca), e quantidade de biomassa de serrapilheira.

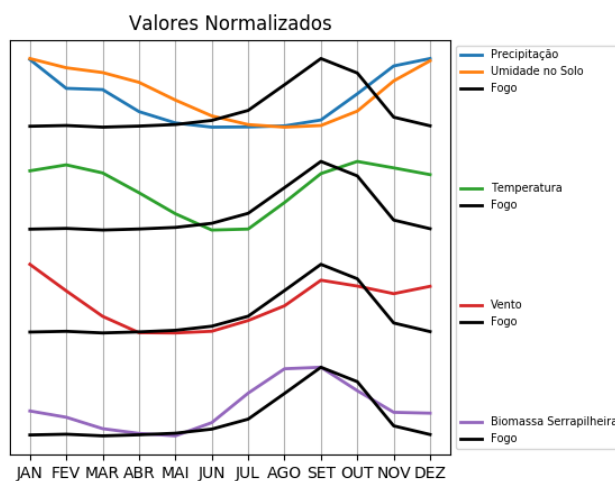


Figura 8. Valores normalizados da climatologia 10 anos para: precipitação, umidade, vento e biomassa para região da Serra do Espinhaço - MG

As Figuras 8 e 10 apresenta a comparação das entradas normalizadas com as saídas esperadas.

A precipitação de chuva irá influenciar no umidade do solo,

quanto mais seco for o ambiente maior será a periculosidade do material combustível florestal, as figuras 3 e 4 apresentam as médias dos valores e chuva e umidade no solo para a região da serra do espinhaço.

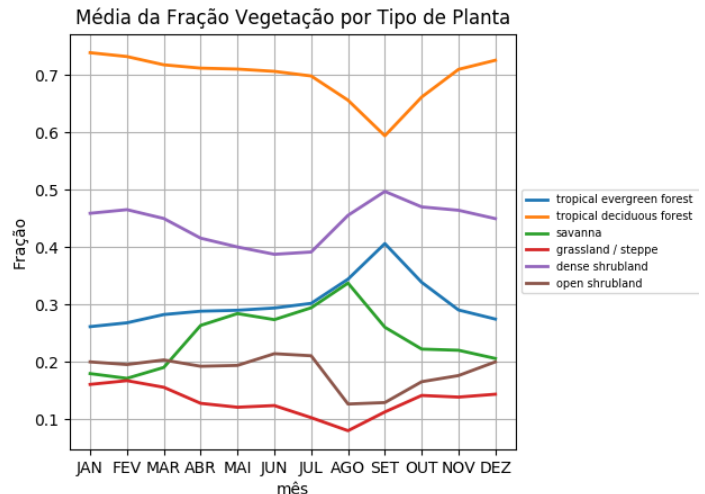


Figura 9. Climatologia 10 anos da fração do tipo de vegetação existente na para região da Serra do Espinhaço - MG

Conforme Figura 10, os picos de ocorrência de fogo coincidem com os maior índices de fração maior concentração de fração de vegetação do tipo floresta tropical e gramíneas esparsas, e correlação inversa entre pastagem/estepes, e floresta topical decídua.

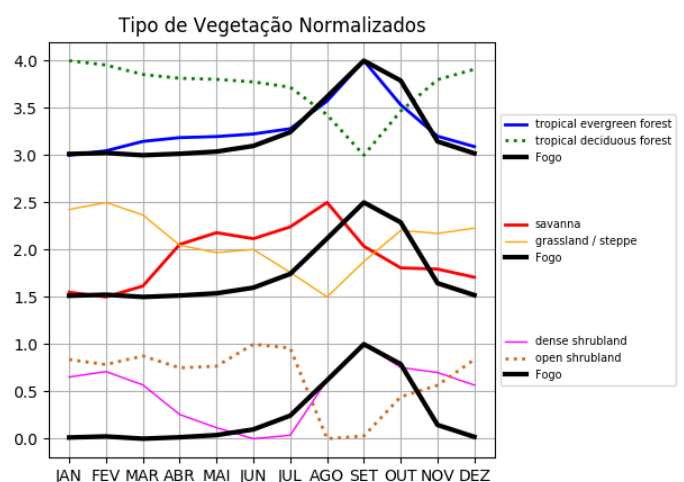


Figura 10. Valores normalizados da Climatologia 10 anos da fração do tipo de vegetação existente na para região da Serra do Espinhaço - MG

### III. RESULTADOS E DISCUSSÃO

Conforme tabela 1 apresenta resultado boas aproximação dos valores calculados com 5 neurônios na camada oculta, os resultados obtidos com treinamento com acima de 5 neurônios na camada oculta, geram maior custo de tempo para treinamento sem aumento significativo nos resultados encontrados.

Neurônio Camada Oculta	Erro médio quadrático	custo tempo
2	0.0078	23m18s
3	0.0087	25m36s
5	0.0069	27m54s
10	0.0070	28m16s
100	0.0067	32m31s

Tabela 1

CUSTO DE TEMPO PARA TREINAMENTO DA RNA COM DIFERENTES NÚMEROS DE NEURÔNIOS E ERRO MÉDIO QUADRÁTICO ENCONTRADO PARA 1000 ÉPOCAS

A Figura 11 abaixo apresenta os resultados dos valores de ocorrência de fogo calculados pela RNA após o treinamento (simulado, em vermelho), e os valores de ocorrência de fogo observados a partir das detecções com o sensor MODIS (observado, em azul). Os valores correspondem às distribuições das médias mensais para toda a sub-região e período estudado. Conforme mostra a figura, as distribuições dos valores observados e simulados são em geral bem próximos. Os valores simulados são capazes de reproduzir a variação intra-anual típica para a região de estudo, onde os meses entre Agosto e Outubro apresentam os valores máximos, e os meses entre Novembro e Julho apresentam os valores mínimos. Outro aspecto interessante é que o range dos valores simulados também é próximo do range dos valores observados na maioria dos meses. No geral, porém, há também uma pequena tendência de sobre-estimativa de valores máximos na maior parte dos meses.

### IV. CONCLUSÃO

Os resultados mostram que a RNA aqui desenvolvida apresenta boa eficiência no cálculo de ocorrência do fogo na vegetação do Cerrado. As condições de ambientes geradas pelo INLAND para a sub-região da Serra do Espinhaço MG aliado a implementação da RNA em comparação aos dados do MODIS demonstram uma convergência satisfatória. Os valores simulados foram capazes de reproduzir o ciclo anual bem como o range dos valores medidos. Houve, porém, uma pequena tendência de sobre-estimativa de valores máximos na maior parte dos meses. Uma característica importante da metodologia e resultados é a possibilidade de aplicar RNAs para avaliar impactos de mudanças climáticas, e criar projetos ambientais que informem sobre possíveis alterações nos padrões do fogo. Para trabalhos futuros sugere-se também investir na utilização do presente método para uma área que contemple todo o Cerrado.

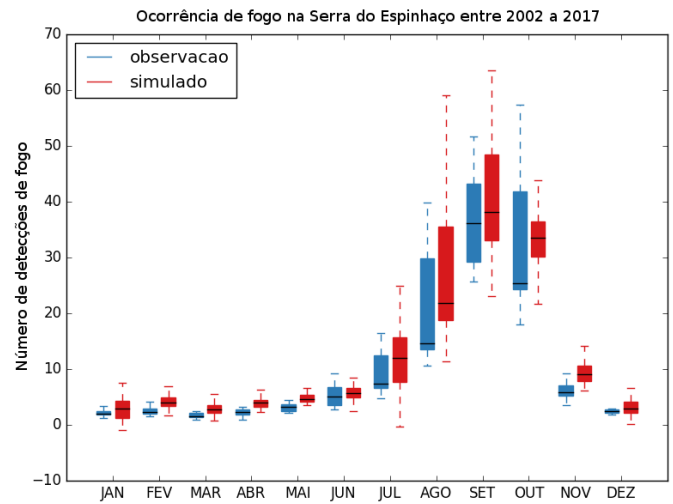


Figura 11. Comparação entre o número de detecções de fogo observado pelo sensor MODIS e simulado pela RNA entre 2002 e 2017 para a Serra do Espinhaço - MG.

### REFERÊNCIAS

- ARRUDA, F. V.; SOUSA, D. G.; TERESA, F. B.; PRADO, V. H. M.; CUNHA, H. F.; IZZO, T. J. *Trends and gaps of the scientific literature about the effects of fire on Brazilian Cerrado*. Biota Neotropica. 18(1): e20170426, 2018.
- FOLEY, J. A.; PRENTICE, I. C.; RAMANKUTTY, N.; LEVIS, S.; POLLARD, D.; SITCH, S.; HAXELTINE, A. *An integrated biosphere model of land surface processes, terrestrial carbon balance, and vegetation dynamics*. Global Biogeochemical Cycles, 1996.
- HAYKIN S. *Redes Neurais: princípios e práticas*, 2. ed. Porto Alegre, Bookman, 2001.
- KIMES, D. S.; NELSON, R. F.; MANRY, M. T.; FUNG, A. K. *Attributes of neural networks for extracting continuous vegetation variables from optical and radar measurements*. International Journal of Remote Sensing, v.19, n.14, p.2639-2663, 1998.
- MIRANDA, S. D.; BUSTAMANTE, M.; PALACE, M.; HAGEN, S.; KELLER, M.; FERREIRA, L. G. *Regional Variations in Biomass Distribution in Brazilian Savanna Woodland*. Biotropica, 46: 125-138. doi:10.1111/btp.12095, 2014.
- MYERS N.; MITTERMEIER R.A.; MITTERMEIER C.G.; DA FONSECA G.A.; Kent J. *Biodiversity hotspots for conservation priorities*. Nature, 403:853-858. doi: 10.1038/35002501, 2000.
- NOBREGA, R. L. B.; GUZHA, A. C.; TORRES, G. N.; KOVACS, K.; LAMPARTER, G.; AMORIM, R. S. S. et al. *Effects of conversion of native cerrado vegetation to pasture on soil hydro-physical properties, evapotranspiration and streamflow on the Amazonian agricultural frontier*. PLoS ONE 12(6): e0179414, 2017.
- WERBOS, P. J. *Beyond regression: new tools for prediction and analysis in the behavioral sciences*, 1975.