

APRENDIZAGEM MÁQUINA NA PREVISÃO DA CERCOSPORIOSE DO CAFÉ

Ecologia Ambiental

Pedro Antonio Lorençone¹

João Antonio Lorençone²

Rafael Fausto de Lima³

Guilherme Botega Torsoni⁴

Lucas Eduardo de Oliveira Aparecido⁵

Resumo

O presente trabalho tem como objetivo prever a Cercosporiose do cafeeiro através do modelo Gradient Boosting Regression, alimentado com dados agrometeorológicos. Foram utilizados dados meteorológicos e da doença do café obtidos em 8 localidades do estado de São Paulo e Minas Gerais, compreendendo regiões do sul e do cerrado mineiro. Nesta etapa foram calibrados os algoritmos do modelo e topologias para previsão da cercosporiose, com uma antecipadamente possível para uma tomada de decisão. A segunda etapa do trabalho, foi a previsão e espacialização da severidade de cercosporiose para todos os municípios relevantes para a produção de café no Brasil. Os dados meteorológicos foram coletados por meio da plataforma NASA POWER. A evapotranspiração de referência foi estimada pelo método de Penman-Monteith, logo sendo gerado o balanço hídrico conforme Thonhwaite e Mather (1955). Os dados de doenças medidos em campo foram coletados entre 2010 e 2012, concentrando 191 epidemias de cercosporiose, em 7 localidades com alta produção de café arábica. Para carga alta o melhor modelo foi o GBR, apresentou precisão e acurácia ($R^2 = 0.63$) após o teste e acurácia, (RMSE = 15,76). Em carga baixa, a precisão foi de 0.61 (R^2) e acurácia 17,01 (RMSE). Utilizando o modelo Gradient Boosting Regression é possível estimar a incidência de cercosporiose em todas as localidades da região com maior produção de café arábica do país com 7 dias de antecedência.

Palavras-chave: Clima, Doença cafeeiro, Python, Gradient Boosting.

¹ Aluno do Curso de graduação em Agronomia, IFMS – Campus Naviraí, agrometeorologia, pedro.lorencone@estudante.ifms.edu.br.

² Aluno do Curso de graduação em Agronomia, IFMS – Campus Naviraí, agrometeorologia, joao.lorencone@estudante.ifms.edu.br.

³ Aluno do Curso de graduação em Agronomia, IFMS – Campus Naviraí, agrometeorologia, rafael.lima@estudante.ifms.edu.br.

⁴ Prof. Dr. Guilherme Botega Torsoni do IFMS – Campus Naviraí, Departamento agrometeorologia, guilherme.torsoni@ifms.edu.br.

⁵ Prof. Dr. Lucas Eduardo de Oliveira Aparecido, do IFSULDEMINAS – Campus Muzambinho, Departamento agrometeorologia, lucas.aparecido@muz.ifsuldeminas.edu.br.

INTRODUÇÃO

A Cercosporiose do cafeeiro, também conhecida como mancha-de-olho pardo é uma doença causada pelo fungo *Cercospora coffeicola*. O patógeno ataca as folhas, penetrando entre aberturas naturais em sua face superior ou em cutícula dos frutos do cafeeiro (SOUZA; MAFFIA; MIZUBUTI, 2012), atingindo todas as fases de desenvolvimento da lavoura, desde o viveiro até o campo (ACEVES-NAVARRO et al., 2020).

O clima interfere no desenvolvimento da doença, sendo esta favorecida pela luminosidade, temperatura do ar (10°C a 25°C), umidade relativa do ar (>90%) (BOTELHO et al., 2017) e período de água livre sobre o tecido vegetal (>9h dia⁻¹) (BOTELHO et al., 2017; SILVA et al., 2018). O fungo se dissemina principalmente pela água, por meio da precipitação pluviométrica ou irrigação por aspersão e vento (EDET et al., 2022).

Segundo (XU, 2003) sistemas de previsão de incidência de doenças com base nas condições ambientais ajudam na tomada de decisões para a aplicação de defensivos agrícolas. Esses sistemas mostram os períodos mais favoráveis para o seu desenvolvimento do patógeno.

Aprendizagem de máquinas é um campo da ciência da computação que trabalha com reconhecimento de padrões utilizando a teoria da aprendizagem computacional em inteligência artificial (SAHOO et al., 2019). Sendo estes, muito eficazes para análises de grande escala em um curto período de tempo, com grande precisão e eficácia (LI et al., 2019).

Gradient Boosting Regression é uma técnica analítica projetada para explorar a relação entre duas ou mais variáveis (X e Y) (FRIEDMAN, 2001). Sua saída analítica identifica fatores importantes (X_i) que impactam a variável dependente (y) e a natureza da relação entre cada um desses fatores e a variável dependente. A regressão de aumento de gradiente é limitada a prever a saída numérica, de modo que a variável dependente deve

Realização

Apoio



ser de natureza numérica (ZHANG; HAGHANI, 2015). O tamanho mínimo da amostra é de 20 casos por variável independente. É uma técnica popular entre os cientistas de dados devido à sua precisão e velocidade, particularmente dados complexos e de grande dimensão (NATEKIN; KNOLL, 2013). Para evitar a manifestação e proliferação das doenças, ferramentas computacionais e modelos matemáticos vêm sendo adotados para prever a incidência delas, de maneira prática, precisa e rápida (CHEN et al., 2022).

Algoritmos de Inteligência Artificial se demonstram eficientes para simular as interações bióticas e abióticas nos cultivos agrícolas (LIANG et al., 2018). Modelos calibrados com dados climáticos também apresentam resultados satisfatórios para previsão de doenças (FENU; MALLOCI, 2021). Entretanto, não foram encontrados na literatura trabalhos que utilizem diferentes algoritmos de inteligência artificial para prever o desenvolvimento da Cercosporiose do Cafeeiro, no Brasil, maior produtor da cultura. Dessa forma, o presente trabalho tem como objetivo prever a Cercosporiose do cafeeiro através do uso de diferentes algoritmos, alimentados com dados agrometeorológicos.

METODOLOGIA

O trabalho foi executado em duas etapas. Primeiramente, foram utilizados dados meteorológicos e da doença do café obtidos em 8 localidades do estado de São Paulo e Minas Gerais, compreendendo regiões do sul e do cerrado mineiro. Nesta etapa foram calibrado o modelo Gradient Boosting Regression (GBR) para previsão da cercosporiose, com uma antecipadamente possível para uma tomada de decisão. A segunda etapa do trabalho, foi a previsão e especialização da severidade de cercosporiose para todos os municípios relevantes para a produção de café no Brasil.

Dados Meteorológicos e de cercosporiose do café

Os dados meteorológicos foram coletados do sistema Prediction Of Worldwide Energy Resources – NASA POWER do período de 2002 a 2022. O grid do sistema NASA-POWER é de $0,5^\circ \times 0,5^\circ$ (aproximadamente 56 km x 56 km) e os dados são provenientes de vários sistemas globais de medidas de superfície e de sensoriamento remoto, como também de modelagem meteorológica. Para mais detalhes, verificar em

Realização



Apoio



(STACKHOUSE et al., 2015).

Para cada ponto de coleta foi calculada a evapotranspiração de referência (ET_o) pelo método padrão da FAO, proposto por Penman-Monteith (ALLEN et al., 1998), considerando o fluxo de calor sensível no solo igual a 5% da radiação líquida. Foi calculado também o balanço hídrico (BH) conforme Thornthwaite and Mather, (1955).

Dos dados de campo foram analisadas 198 epidemias de cercosporiose de 2010 a 2022 de 8 cidades, Boa Esperança (21°03'59" S, 45°34'37" W, 830 m), Carmo de Minas (22°10'31" S, 45°09'03" W, 1080 m), Muzambinho (21°20'47" S, 46°32'04" W, 1033 m), Varginha (21°34'00" S, 45°24'22" W, 940 m), Araxá (19° 33' 21" S, 46° 58' 08" W, 960 m), Araguari (18° 59' 35" S, 46° 59' 01" W, 961 m), Patrocínio (18° 33' 21" S, 48° 12' 25" W, 933 m), e Franca (20o 28' 19''S, 47o 24' 33''W, 1025 m).

As incidências foram medidas através de um procedimento não destrutivo. As plantas foram escolhidas aleatoriamente num padrão de marcha em zigue zague na área, tal como recomendado por Chalfoun (1997). As pontuações de incidência foram determinadas em folhas a partir do terceiro ou quarto nó de ramos no terço médio das plantas. A metodologia para avaliar o desenvolvimento do café, doenças e pragas utilizada foi de acordo com o Tabela 1.

Tabela 01. Metodologia de avaliação do desenvolvimento do café, doenças e pragas utilizada pela Fundação Procafé do Brasil.

Problema fitossanitário	Metodologia
	Doenças
Índice de Cercosporiose (Cercospora coffeicola)	<ul style="list-style-type: none"> - Amostra de 20 plantas por campo - Recolher as folhas no terço médio da planta escolhida - Escolher ao acaso cinco ramos laterais de cada lado da planta - Retirar uma folha completamente desenvolvida, do 3º ou 4º par de folhas, de cada ramo - Total de 10 folhas/planta (cinco de cada lado) - Total de 200 folhas/campo - Quantificar a percentagem de doença nas amostras

A quantificação da taxa de manchas oculares castanhas no café em % foi obtida através da equação 1.

$$\text{Índice Cercospora (\%)} = \left(\frac{\text{Número de folhas com Cercospora}}{\text{Número de folhas total}} \right) \times 100 \quad (1)$$

A taxa de manchas castanhas nos olhos foi avaliada em duas situações de "alta" e "baixa" produtividade, que ocorrem devido à natureza bienal natural da planta do café. A "alta" representa mais de 30 sacos de 60 kg ha⁻¹, enquanto que a "baixa" é inferior a 10 sacos de 60 kg ha⁻¹.

1º Etapa – Calibração Gradient Boosting

Foram utilizados dados meteorológicos diários de Temperatura mínima - minT, Temperatura média - avgT, Temperatura máxima - maxT, temperatura do ponto de orvalho – dewT, adiação solar global diária – QG, Umidade relativa mínima – minRH, Umidade relativa média - avgRH, Umidade relativa máxima - maxRH, Duração do período de molhamento das folhas – LLW, Chuva – R, Evapotranspiração potencial – Eto, Evapotranspiração atual – ETR, Armazenamento da água do solo – SWS, Excedente hídrico - Sur, Déficit hídrico - Def, Velocidade mínima do vento – minWS, Velocidade média do vento – avgWS e Velocidade máxima do vento – maxWS

2º Etapa- Aplicação do modelo calibrado no Brasil

O mesmo procedimento de coleta de dados meteorológicos pelo sistema da NASA- POWER e cálculo de ETo e BH foi feito nesta etapa do trabalho, entretanto os dados foram obtidos para as maiores localidades produtoras de coffeea arábica (Figura 1), referente ao período de 1981 a 2022. Esse processo foi feito em linguagem de programação Python (<https://www.python.org/>, versão 3.7.3 ou superior) que permite o trabalho com grandes conjuntos de dados.

2.1. Modelagem dos algoritmos

A taxa de infecção foi a variável dependente e os elementos meteorológicas (Tabela 2) as variáveis independentes usados na calibração dos algoritmos. Para todas as metodologias foram separados de maneira aleatória 30% dos dados para o treinamento e os demais 70% para a calibração utilizando a biblioteca (sklearn.model_selection.train_test_split) do python. Os parâmetros calibrados, os valores testados em cada parâmetro e os valores definidos para o algoritmo buscando a maior

Realização

Apoio

acurácia pode ser observado na Tabela 2.

Tabela 2. Parâmetros calibrados, os valores testados em cada parâmetro e os valores definidos na previsão da cercosporiose do cafeeiro para todo Brasil.

Modelo	Parameter	Values tested	Defined values
GradBOOSTING	n_estimators	10; 1000; 3000; 10000	1000

2.2. Definindo features de entrada

Foi considerado o período de latência da cercosporiose para a definição das variáveis de entrada. As análises foram feitas utilizando dados de 7º a 30º dia antes da ocorrência da doença, o que permite uma antecipação de 1 semana, tempo hábil para uma tomada de decisão na lavoura. Esses períodos foram escolhidos analisando o período de latência (tempo entre a infecção do patógeno na planta e a manifestação dos sintomas das doenças). Por exemplo, o período de latência da Cercosporiose é de 9 a 15 dias (KUSHALAPPA; AKUTSU; LUDWIG, 1983).

As FEATURES foram organizadas da seguinte forma: Tmin-30 correspondendo a temperatura mínima do ar no 30º dia antes do período de incubação da cercosporiose, seguindo da mesma forma Tmin-29, Tmin-28 ... Tmin7. O mesmo foi feito para os demais elementos climáticos, correspondendo inicialmente a camada de entrada dos modelos, totalizando inicialmente 378 FEATURES (18 variáveis x 21 dias antes do período de incubação da doença) para cada localidade.

2.3. Métricas de avaliação dos modelos

A comparação entre os dados reais (observados nas lavouras em campo) e os resultados de todas as metodologias foi realizada por meio de diversos índices estatísticos: 1) acurácia e 2) precisão. A acurácia ou exatidão ou bias, consiste no quanto a estimativa está próxima do valor observado, e foi avaliada pelo RMSE (raiz do erro ao quadrado médio), MSE (média quadrada do erro) e MAE (erro médio absoluto). A precisão é a

Realização

Apoio

capacidade do modelo repetir a estimativa e foi avaliada pelo coeficiente de determinação (R²) ajustado segundo Cornell and Berger (1987).

2.4. Espacialização dos dados

Com todos os algoritmos calibrados para a previsão da doença foi especializado as previsões do modelo mais acurado para as regiões com maior produção de café do Brasil. O método de interpolação utilizado foi a krigagem (KRIGE, 1951), com o modelo esférico, um vizinho e resolução de 0,25° (25 km).

RESULTADOS E DISCUSSÃO

Em relação ao modelo para previsão da cercosporiose Gradient Boosting Regression (GBR) houve uma variação entre anos de carga alta e anos de carga baixa. Para carga alta o modelo apresentou R² de 0,63, RMSE de 15,76, MSE de 248,45 e MAE de 9,98 (Tabela 2). Já para carga baixa o modelo demonstrou um R² de 0,61, RMSE de 17,01, MSE de 289,35 e MAE de 9,64, demonstrando que o modelo consegue prever com maior precisão em anos de alta produções dos cafeeiros (Tabela 2). A partir desses resultados pode se concluir que o modelo GBR é capaz de prever a Cercosporiose do cafeeiro (MORETO; ROLIM, 2015). Os gráficos de performances da previsão da cercosporiose para cafeeiro em carga alta e em carga baixa podem ser vistos na figura 1.

Table 02. Desempenho do modelo Gradient Boosting Regression para estimativa da incidência de cercospora para cafeeiros em carga alta e baixa

MODELO	Índice	Treino		Teste	
		Alta	Baixa	Alta	Baixa
GBR	R_correlação	0,9999	0,9999	0,6302	0,6141
	RMSE	0,0247	0,0002	15,7622	17,0101
	MSE	0,0006	4,2678	248,4477	289,3565
	MAE	0,0198	0,0001	9,9846	9,6461

Realização

Apoio

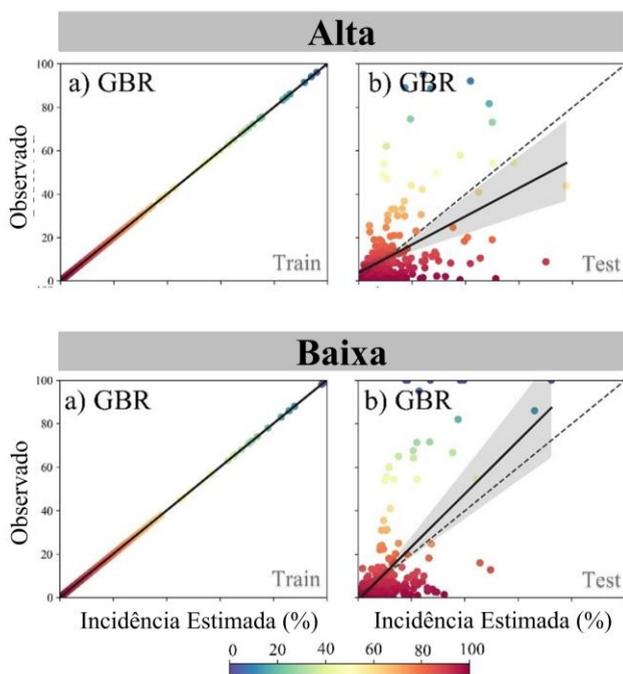


Figura 01: Performance dos algoritmos na calibração e teste para a previsão da cercosporiose do cafeeiro em carga alta

Foi perceptível diferença entre os cafeeiros em carga alta e carga baixa, sendo durante a primeira maior a severidade de cercosporiose na cultura (Figure 2 e 3). Quando o cafeeiro se encontra em carga alta ele direciona grande parte de seus foto assimilados para produção de frutos e enchimento de grãos, dificultando a manutenção de mecanismo de defesa, tornando-o mais suscetível (CAMARGO; CAMARGO, 2001).

Em carga alta para a região as classes mais predominantes foram de 0-10, 20-30 e 10-20%, médias anuais de 15.55, 36.11 e 52.63%, respectivamente. A classe mais grave de infestação (40-50%) concentrou em apenas 8.59% nos meses de maio, junho e julho. Durante junho 55.70% da área apresentou as duas classes mais graves, 33.48% para 30-40 e 13.58% em 40-50% (Figura 2 F G). Por outro lado, outubro e novembro demonstraram maior concentração de áreas com severidade abaixo de 20%, média de 24.51 e 55.39% para 0-10 e 10-20% (Figure 2 J K).

A ocorrência de cercosporiose variou conforme o estado dentro da região de estudo, com menores ocorrências nas localidades com maiores produções de café. Os

Realização

Apoio



únicos estados que apresentaram incidência dentre 40-50% foram BA (12.36%) e GO (12.52%), apresentaram a maior área com 20-30 também, 19.99 e 25.08%. Tanto na BA como em GO, junho e julho são os períodos mais críticos, em junho a média foi de 19.10% da área dos estados com cerca de 40 a 50% da lavoura contaminada. Ambos os estados demonstraram suscetibilidade ao ataque de *Cercospora coffeicola*, sendo necessário adotar medidas preventivas ao ataque do patógeno, métodos culturais como consórcio com árvores frutíferas, no qual o cafeeiro cresce sombreado, impossibilitando a ativação da toxina do fungo, não prejudicando as folhas (MOUEN BEDIMO et al., 2007).

Para carga baixa as classes de incidência de cercosporiose as que apresentaram valores na região estuda foram, 0-10 com 36.13 %, 10-20 com 50.22 % e 20-30 com 14.07 % (Figura 3). Para a classe 0-10 o estado de Minas Gerais é o que apresentou a maior porcentagem 44.11 % de seu território. Em relação a classe de 10-20 o estado que apresentou maior área foi a Bahia com 84.13 % de seu território durante todo ano, já o estado com a menor porcentagem foi Goiás com 43.67 %. Já para a classe 20-30 apenas alguns estados apresentaram porcentagem durante, sendo o Goiás o maior, com 47.23 % (Figura 3). Os meses com maior probabilidade de ocorrência de cercosporiose em anos de baixa produção, forma de julho a outubro (Figura 3).

Realização



Apoio



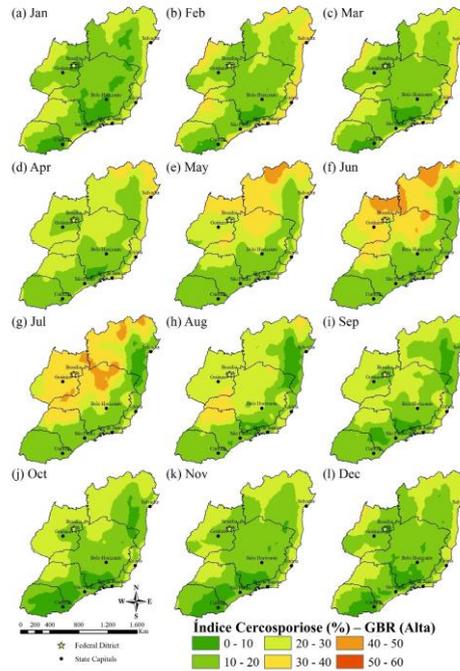


Figura 02: Incidência mensal de cercosporiose do café em carga alta previsto pelo algoritmo GBR para todas as regiões produtoras do Brasil.

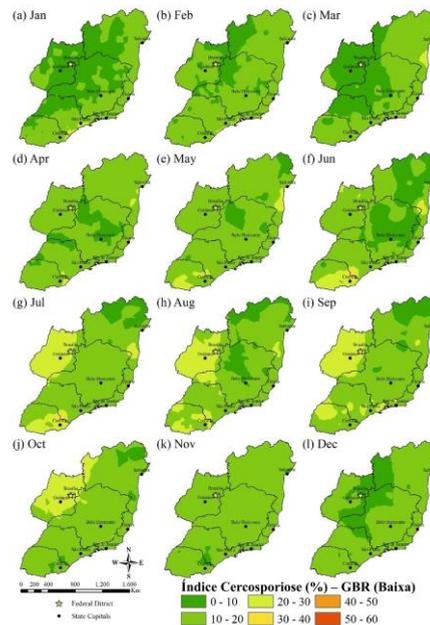


Figura 03: Incidência mensal de cercosporiose do café em carga baixa previsto pelo algoritmo GBR para todas as regiões produtoras do Brasil.

Realização

Apoio



CONCLUSÕES OU CONSIDERAÇÕES FINAIS

Com o auxílio de variáveis agrometeorológicas e o modelo Gradient Boosting Regression é possível estimar a incidência da cercosporiose do cafeeiro nas lavouras de café.

O desempenho do modelo para estimativa da severidade de cercosporiose é diferente em função da bionalidade do cafeeiro. Para carga alta o modelo demonstra precisão e acurácia ($R^2 = 0.63$) após o teste e acurácia, (RMSE = 15,76). Para estimar em carga baixa, a precisão foi de 0.61 (R2) e acurácia 17,01 (RMSE).

Utilizando o modelo Gradient Boosting Regression é possível estimar a incidência de cercosporiose em todas as localidades da região com maior produção de café arábica do país com 7 dias de antecedência.

As localidades com maiores produções de café apresentam baixa incidência de cercosporiose durante o ano todo, tanto para carga alta e baixa. Em carga alta o norte de Minas Gerais, da Bahia e Nordeste de Goiás demonstram alta ocorrência da doença durante maio a julho, alcançando 50% da lavoura contaminada em junho.

O clima é fator essencial na relação patógeno hospedeiros e ao mesmo tempo as doenças fúngicas reduzem expressivamente produção dos cafeeiros.

AGRADECIMENTOS

Agradecemos ao CNPq pelo apoio e financiamento dessa pesquisa

REFERÊNCIAS

- ACEVES-NAVARRO, L. A. et al. Impact of climatic change on the adaptation of coffee (*Coffea arabica* L.) crops in Tabasco, Mexico. **Agroproductividad**, v. 13, n. 4, p. 53–58, 2020.
- ALLEN, R. G. et al. Crop evapotranspiration-Guidelines for computing crop water requirements-FAO Irrigation and drainage paper 56. **Fao, Rome**, v. 300, n. 9, p. D05109, 1998.
- BOTELHO, D. M. DOS S. et al. Cercosporiosis resistance in coffee germplasm collection. **Euphytica**, v. 213, n. 6, p. 117, jun. 2017.
- CAMARGO, Â. P. D.; CAMARGO, M. B. P. D. Definição e esquematização das fases fenológicas do cafeeiro arábica nas condições tropicais do Brasil. **Bragantia**, v. 60, n. 1, p. 65–68, 2001.

Realização

Apoio



- CHALFOUN, N. **Design and application of natural down-draft evaporative cooling devices.** AMERICAN SOLAR ENERGY SOCIETY INC, 1997.
- CHEN, Z. et al. Plant Disease Recognition Model Based on Improved YOLOv5. **Agronomy**, v. 12, n. 2, p. 365, 31 jan. 2022.
- CORNELL, J.; BERGER, R. Factors that influence the value of the coefficient of determination in simple linear and nonlinear regression models. **Phytopathology**, v. 77, n. 1, p. 63–70, 1987.
- EDET, I. A. et al. Identification and molecular characterisation of cercospora leaf spot disease pathogen on cowpea (*Vigna unguiculata* L. Walp). **Archives of Phytopathology and Plant Protection**, v. 55, n. 1, p. 109–120, 2 jan. 2022.
- FENU, G.; MALLOCI, F. M. Forecasting Plant and Crop Disease: An Explorative Study on Current Algorithms. **Big Data and Cognitive Computing**, v. 5, n. 1, p. 2, 12 jan. 2021.
- FRIEDMAN, J. H. Greedy function approximation: a gradient boosting machine. **Annals of statistics**, p. 1189–1232, 2001.
- KRIGE, D. G. A statistical approach to some basic mine valuation problems on the Witwatersrand. **Journal of the Southern African Institute of Mining and Metallurgy**, v. 52, n. 6, p. 119–139, 1951.
- KUSHALAPPA, A.; AKUTSU, M.; LUDWIG, A. Application of survival ratio for monocyclic process of *Hemileia vastatrix* in predicting coffee rust infection rates. **Phytopathology**, v. 73, n. 1, p. 96–103, 1983.
- LI, J. et al. Diversifying crop rotation improves system robustness. **Agronomy for Sustainable Development**, v. 39, n. 4, p. 1–13, 2019.
- LIANG, Y. et al. How disturbance, competition, and dispersal interact to prevent tree range boundaries from keeping pace with climate change. **Global Change Biology**, v. 24, n. 1, p. e335–e351, jan. 2018.
- MORETO, V. B. M.; ROLIM, G. DE S. Estimation of annual yield and quality of Valencia orange related to monthly water deficiencies. **African Journal of Agricultural Research**, v. 10, n. 6, p. 543–553, 5 fev. 2015.
- MOUEN BEDIMO, J. A. et al. Effect of cultural practices on the development of arabica coffee berry disease, caused by *Colletotrichum kahawae*. **European Journal of Plant Pathology**, v. 119, n. 4, p. 391–400, 30 out. 2007.
- NATEKIN, A.; KNOLL, A. Gradient boosting machines, a tutorial. **Frontiers in neurorobotics**, v. 7, p. 21, 2013.
- SAHOO, B. B. et al. Long short-term memory (LSTM) recurrent neural network for low-flow hydrological time series forecasting. **Acta Geophysica**, v. 67, n. 5, p. 1471–1481, out. 2019.
- SILVA, H. R. et al. Cercospora leaf spot in *Toona ciliata*: Epidemiology and infection process of *Cercospora* cf. *alchemillicola*. **Forest Pathology**, v. 48, n. 6, p. e12451, 2018.
- SOUZA, A. G. C.; MAFFIA, L. A.; MIZUBUTI, E. S. G. Cultural and Aggressiveness Variability of *Cercospora coffeicola*: Variability of *Cercospora coffeicola*? **Journal of Phytopathology**, v. 160, n. 10, p. 540–546, out. 2012.
- THORNTHWAITE, C. W.; MATHER, J. R. The water balance climatology. **Laboratory of Climatology**, 1955.
- XU, X. Effects of environmental conditions on the development of Fusarium ear blight. Em: XU, X.; BAILEY, J. A.; COOKE, B. M. (Eds.). **Epidemiology of Mycotoxin Producing Fungi**. Dordrecht: Springer Netherlands, 2003. p. 683–689.
- ZHANG, Y.; HAGHANI, A. A gradient boosting method to improve travel time prediction. **Transportation Research Part C: Emerging Technologies**, v. 58, p. 308–324, 2015.

Realização

Apoio

