REMUNOM
ISSN 2178-6925

Received: 22/08/2026 - **Accepted:** 22/09/2026
Vol: 13.20
DOI: 10.66104/x4hqfw39
Pages: 1-22

PREVISÃO DA QUANTIDADE COMERCIALIZADA DE LIMA ÁCIDA 'TAHITI' COM RANDOM FOREST E VALIDAÇÃO TEMPORAL

FORECASTING THE MARKETED QUANTITY OF 'TAHITI' ACID LIME USING RANDOM FOREST AND TEMPORAL VALIDATION

PREDICCIÓN DE LA CANTIDAD COMERCIALIZADA DE LIMA ÁCIDA 'TAHITI' MEDIANTE RANDOM FOREST Y VALIDACIÓN TEMPORAL

Letícia Moura da Silva
Ma. em Agronomia, Universidade Federal Rural da Amazônia (UFRA), Brasil
E-mail: leticiamouradasilva29@gmail.com

Antonia Benedita da Silva Bronze
Dra. em Ciências Agrárias, Universidade Federal Rural da Amazônia (UFRA), Brasil
E-mail: antonia.silva@ufra.edu.com

Glenda Rafaele da Silva
Graduanda em Agronomia, Universidade Federal Rural da Amazônia (UFRA), Brasil
E-mail: glendarafaelle@gmail.com

Larissa Pacheco Nogueira
Graduada em Agronomia, Universidade Federal Rural da Amazônia (UFRA), Brasil
E-mail: larissanogues@gmail.com

Marcos Paulo Souza de Almeida
Graduando em Agronomia, Universidade Federal Rural da Amazônia (UFRA), Brasil
E-mail: souzadealmeidamarcospaulo@gmail.com

Camila Galvão Viana
Graduanda em Agronomia, Universidade Federal Rural da Amazônia (UFRA), Brasil
E-mail: camilagalvao2026@hotmail.com

Resumo

A previsão da quantidade comercializada de produtos agrícolas pode contribuir para o planejamento das operações produtivas e comerciais. Este estudo teve como objetivo desenvolver e avaliar um modelo Random Forest para prever, com horizonte de um mês, a quantidade comercializada de lima ácida 'Tahiti'. Foram utilizados dados mensais de 2018 a 2023 de uma unidade produtiva localizada em Capitão Poço, Pará. A modelagem incorporou informações históricas da quantidade comercializada e componentes sazonais, enquanto as variáveis climáticas foram avaliadas adicionalmente por análise de sensibilidade. O desempenho foi analisado por validação temporal com

REMUNOM
ISSN 2178-6925

Received: 22/08/2026 - **Accepted:** 22/09/2026
Vol: 13.20
DOI: 10.66104/x4hqfw39
Pages: 1-22

janelas expansivas, utilizando 2021, 2022 e 2023 como períodos de teste. Nas 36 previsões fora da amostra, o Random Forest apresentou MAE de 279.100,1 kg, RMSE de 387.771,5 kg e R² preditivo de 0,530, superando os modelos de referência baseados na persistência, média histórica mensal e mesmo mês do ano anterior. Entretanto, a regressão linear apresentou melhor desempenho global (MAE = 239.165,4 kg; RMSE = 343.104,6 kg; R² = 0,632), indicando que maior complexidade algorítmica não resultou em maior capacidade preditiva neste conjunto de dados. As variáveis climáticas não proporcionaram ganho incremental de desempenho no horizonte avaliado. Os resultados evidenciam a importância da estrutura temporal e sazonal para a previsão da quantidade comercializada e reforçam a necessidade de comparar algoritmos de aprendizado de máquina com modelos mais parcimoniosos.

Palavras-chave: aprendizado de máquina; citricultura; previsão temporal.

Abstract

Forecasting the marketed quantity of agricultural products can contribute to the planning of production and commercial operations. This study aimed to develop and evaluate a Random Forest model to predict the marketed quantity of 'Tahiti' acid lime with a one-month forecast horizon. Monthly data from 2018 to 2023 were obtained from a production unit located in Capitão Poço, Pará, Brazil. The model incorporated historical information on marketed quantity and seasonal components, while climatic variables were additionally evaluated through sensitivity analysis. Predictive performance was assessed using expanding-window temporal validation, with 2021, 2022, and 2023 as test periods. Across the 36 out-of-sample predictions, the Random Forest achieved an MAE of 279,100.1 kg, an RMSE of 387,771.5 kg, and a predictive R² of 0.530, outperforming reference models based on persistence, historical monthly mean, and the same month of the previous year. However, linear regression showed better overall performance (MAE = 239,165.4 kg; RMSE = 343,104.6 kg; R² = 0.632), indicating that greater algorithmic complexity did not result in higher predictive performance for this dataset. Climatic variables did not provide incremental predictive gains over the evaluated forecast horizon. The results highlight the importance of temporal and seasonal structure in forecasting marketed quantity and reinforce the need to compare machine learning algorithms with more parsimonious models.**Keywords:** brazilian citrus growing; climate variables; agricultural management.

Keywords: machine learning; citrus production; time-series forecasting.

Resumen

La predicción de la cantidad comercializada de productos agrícolas puede contribuir a la planificación de las operaciones productivas y comerciales. Este estudio tuvo como objetivo desarrollar y evaluar un modelo Random Forest para predecir, con un horizonte de un mes, la cantidad comercializada de lima ácida 'Tahiti'. Se utilizaron datos mensuales de 2018 a 2023 procedentes de una unidad productiva ubicada en Capitão Poço, Pará, Brasil. El modelo incorporó información histórica de la cantidad comercializada y componentes estacionales, mientras que las variables climáticas fueron evaluadas adicionalmente mediante un análisis de sensibilidad. El desempeño predictivo se evaluó mediante validación temporal con ventanas expansivas, utilizando 2021, 2022 y 2023 como períodos de prueba. En las 36 predicciones fuera de muestra, el Random Forest presentó un MAE de 279.100,1 kg, un RMSE de 387.771,5 kg y un R² predictivo de 0,530, superando los modelos de referencia basados en la persistencia, el promedio histórico mensual y el mismo mes del año anterior. Sin embargo, la regresión lineal presentó un mejor desempeño global (MAE = 239.165,4 kg; RMSE = 343.104,6 kg; R² = 0,632), indicando que una mayor complejidad algorítmica no resultó en una mayor capacidad predictiva para este conjunto de datos. Las variables climáticas no proporcionaron ganancias predictivas incrementales en el horizonte evaluado. Los resultados destacan la importancia de la estructura temporal y estacional para la predicción de la cantidad comercializada y refuerzan la necesidad de comparar los algoritmos de aprendizaje automático con modelos más parsimoniosos.

REMUNOM
ISSN 2178-6925

Received: 22/08/2026 - **Accepted:** 22/09/2026
Vol: 13.20
DOI: 10.66104/x4hqfw39
Pages: 1-22

Palabras clave: aprendizaje automático; citricultura; predicción temporal.

1. Introdução

A fruticultura constitui um dos principais pilares da agricultura brasileira, exercendo papel estratégico no desenvolvimento econômico de diversas regiões, além de contribuir para a geração de empregos e para a fixação da população no meio rural. Entre os segmentos frutícolas, a citricultura destaca-se pela sua expressiva relevância econômica, posicionando o Brasil entre os principais produtores mundiais de frutas cítricas, especialmente laranja, limão e lima ácida, com forte impacto na balança comercial e no setor agroindustrial (Passos, 1990; Lopes *et al*., 2013; CNA, 2019; Kist *et al*., 2018).

Dentre as espécies cítricas cultivadas no país, a lima ácida 'Tahiti' (*Citrus latifólia* Tanaka), popularmente conhecida como limão Tahiti ou limão verde (Mattos Jr *et al*., 2021) tem apresentado crescimento expressivo nos últimos anos, impulsionado pela elevada aceitação no mercado interno e externo, bem como por suas características agronômicas favoráveis, como ampla adaptação edafoclimática e elevada produtividade (Luchetti *et al*., 2003; Bastos *et al*., 2022).

No cenário nacional, o estado de São Paulo lidera a produção da lima ácida 'Tahiti', sendo responsável por aproximadamente 73,87% da produção brasileira, enquanto o estado do Pará ocupa a quarta posição no ranking nacional, com produção estimada em 74.442 toneladas, correspondendo a 4,33% da produção nacional em 2024, segundo dados do Instituto Brasileiro de Geografia e Estatística (IBGE, 2025).

Apesar do elevado potencial produtivo, uma parcela significativa da produção de lima ácida 'Tahiti' é perdida durante as etapas de colheita e pós-colheita. Essas perdas estão associadas a fatores como condições climáticas adversas, práticas de manejo inadequadas e danos mecânicos decorrentes do uso de equipamentos descalibrados ou de sistemas de colheita ineficientes. Tais injúrias comprometem a qualidade dos frutos e favorecem a incidência de distúrbios fisiológicos e

REMUNOM
ISSN 2178-6925

Received: 22/08/2026 - **Accepted:** 22/09/2026
Vol: 13.20
DOI: 10.66104/x4hqfw39
Pages: 1-22

patológicos, reduzindo o percentual de frutos aptos à comercialização e à exportação (Bassan, 2015; Bassan *et al*., 2016; Parisi *et al*., 2012).

Diante desse cenário, o uso de tecnologias digitais e métodos preditivos baseados em aprendizado de máquina tem se mostrado uma alternativa promissora para apoiar a tomada de decisão no setor agrícola. Essas técnicas permitem a análise integrada de grandes volumes de dados históricos, climáticos e produtivos, possibilitando a identificação de padrões complexos e a antecipação de cenários produtivos (Oliveira *et al*., 2023; Vasconcelos e Silva, 2024).

Entre os algoritmos de aprendizado de máquina, o Random Forest destaca-se por sua robustez, eficiência computacional e capacidade de modelar relações não lineares, sendo amplamente empregado em estudos de previsão de produtividade agrícola e rendimento de culturas frutíferas. Diversos trabalhos relatam desempenho superior desse algoritmo quando comparado a métodos tradicionais, especialmente em contextos agrícolas caracterizados por elevada variabilidade ambiental (Ferreira *et al*., 2023; Pereira *et al*., 2024; Andrade *et al*., 2020).

A integração de séries históricas de dados, variáveis climáticas e características específicas da cultura da lima ácida ‘Tahiti’ pode fornecer subsídios relevantes para os produtores, permitindo a otimização do uso de recursos, a redução de perdas e o planejamento estratégico da colheita e da comercialização. Além disso, a previsão da produção contribui para o ajuste das etapas da cadeia de suprimentos, favorecendo o equilíbrio entre oferta e demanda e aumentando a eficiência do sistema produtivo (Wolfart *et al*., 2021; Tajudeen *et al*., 2022).

O presente estudo teve como objetivo desenvolver um modelo preditivo para a comercialização de frutos de lima ácida ‘Tahiti’, utilizando o algoritmo Random Forest.

2. Revisão da Literatura

Produção de lima ácida 'Tahiti' e fatores que influenciam a produtividade

A lima ácida 'Tahiti' (*Citrus latifolia* Tanaka) é uma das principais espécies cítricas cultivadas no Brasil, destacando-se pela elevada demanda dos mercados

REMUNOM
ISSN 2178-6925

Received: 22/08/2026 - **Accepted:** 22/09/2026
Vol: 13.20
DOI: 10.66104/x4hqfw39
Pages: 1-22

interno e externo e pela sua contribuição para a geração de renda e diversificação da citricultura nacional. No estado do Pará, a cultura possui expressiva importância econômica, sendo o estado um dos principais produtores brasileiros de limão, com produtividade média anual acima de 22 t ha^{-1} (Pinto *et al*., 2023). O município de Capitão Poço concentra grande parte dessa produção, consolidando-se como um dos principais polos citrícolas da região amazônica (Pinto *et al*., 2023; Gurgel *et al*., 2025).

A competitividade da cultura está diretamente relacionada às condições climáticas e ao manejo adotado. Estudos demonstram que fatores como disponibilidade hídrica, temperatura e escolha adequada de porta-enxertos influenciam significativamente o número de frutos, a produtividade e a qualidade da produção. Dessa forma, o uso de tecnologias e ferramentas de apoio à decisão torna-se essencial para aumentar a eficiência produtiva e reduzir os impactos das variações climáticas sobre a cultura (Gurgel *et al*., 2024; Gurgel *et al*., 2025). Estudos recentes com *Citrus latifolia* demonstram que a disponibilidade de água no solo influencia o estado hídrico das plantas, as trocas gasosas, o crescimento e a produção de frutos, evidenciando a importância do manejo hídrico, principalmente durante períodos de menor disponibilidade de água (Rivera-Hernández *et al*., 2025).

A variabilidade climática constitui um importante fator de risco para a produção de lima ácida 'Tahiti', uma vez que alterações na distribuição das chuvas, períodos de déficit hídrico, temperaturas elevadas e ocorrência de eventos climáticos extremos podem afetar processos fisiológicos e comprometer a estabilidade da produção. Em escala mais ampla, os efeitos das mudanças climáticas sobre a citricultura estão associados à ocorrência de estresses como seca, calor, salinidade e alterações no regime de precipitação, com possíveis impactos sobre a produtividade e a qualidade dos frutos (Srivastava *et al*., 2026).

Aprendizado de máquina aplicado à agricultura

O aprendizado de máquina, Machine Learning, tem se consolidado como uma das principais ferramentas da agricultura de precisão, permitindo analisar grandes volumes de dados e identificar padrões que auxiliam na tomada de decisão. A

REMUNOM
ISSN 2178-6925

Received: 22/08/2026 - **Accepted:** 22/09/2026
Vol: 13.20
DOI: 10.66104/x4hqfw39
Pages: 1-22

integração de informações provenientes de sensores, Internet das Coisas (IoT), drones, imagens de satélite e estações meteorológicas tem ampliado a capacidade de monitoramento das lavouras e de previsão do desempenho das culturas, tornando o manejo agrícola mais eficiente e sustentável (Liakos *et al*., 2018; Shawon *et al*., 2025).

Na agricultura, o Machine Learning é amplamente empregado para previsão de produtividade, monitoramento de culturas, detecção de doenças, manejo da irrigação e otimização do uso de insumos. Em estudos de previsão de rendimento agrícola, variáveis como temperatura, precipitação, características do solo e índices de vegetação são frequentemente utilizadas, enquanto algoritmos como Random Forest, Regressão Linear e Gradient Boosting Trees se destacam pela elevada capacidade de modelar relações complexas e não lineares entre os fatores que influenciam a produção (Liakos *et al*., 2018; Kamilaris *et al*., 2019; Istiak, 2023).

Apesar dos avanços, a aplicação dessas técnicas ainda enfrenta desafios relacionados à disponibilidade e à qualidade dos dados, à capacidade de generalização dos modelos para diferentes regiões e condições de cultivo e à necessidade de validação em campo. Assim, a construção de modelos baseados em dados locais e representativos torna-se fundamental para aumentar a confiabilidade das previsões e ampliar o potencial do aprendizado de máquina como ferramenta de apoio à gestão agrícola, especialmente em cenários de mudanças climáticas e crescente demanda por sistemas produtivos mais eficientes (Yao *et al*., 2026; Frimpong *et al*., 2025).

3. Metodologia

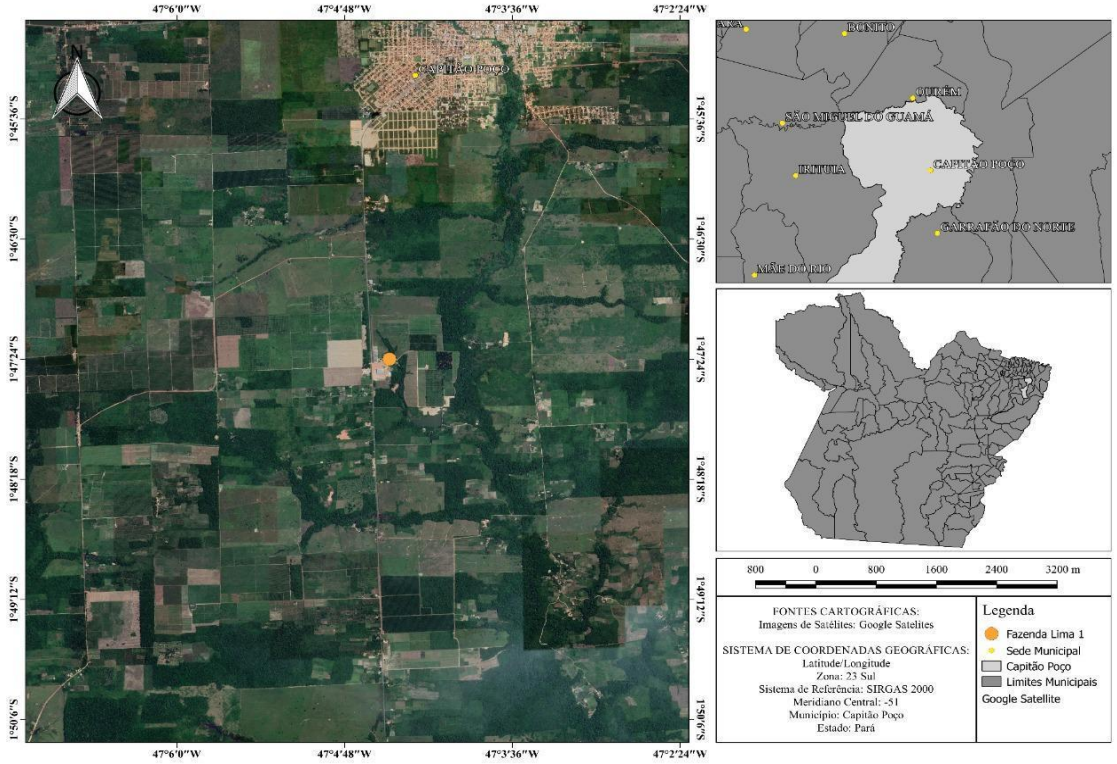
Área estudada

A área de estudo compreende o município de Capitão Poço, localizado no estado do Pará, na região leste da Amazônia brasileira. A pesquisa foi conduzida na Casa de Embalagem da empresa Lima Citros, situada na Rodovia PA-124, km 04, nº 17.304, a aproximadamente 6 km da sede do município (Figura 1).

REMUNOM
ISSN 2178-6925

Received: 22/08/2026 - **Accepted:** 22/09/2026
Vol: 13.20
DOI: 10.66104/x4hqfw39
Pages: 1-22

Figura 1. Mapa de localização da Casa de Embalagem em Capitão Poço - PA.



Fonte: Os autores (2026).

De acordo com a classificação climática de Köppen, adaptada por Martorano et al. (1993), o município apresenta clima do tipo Af_2, caracterizado por condições tropicais úmidas. A precipitação pluviométrica média anual varia entre 2.500 e 3.000 mm, com distribuição relativamente homogênea ao longo do ano. O período menos chuvoso ocorre no trimestre compreendido entre os meses de setembro e novembro, conforme descrito por Sousa (2021).

Essas condições climáticas favorecem o cultivo de espécies cítricas, especialmente a lima ácida ‘Tahiti’, cuja produção é fortemente influenciada por variáveis meteorológicas, como precipitação, temperatura e umidade relativa do ar.

Coleta de dados

Os dados utilizados para o desenvolvimento do modelo de previsão da quantidade de frutos comercializados de lima ácida ‘Tahiti’ foram obtidos a partir de uma base composta por informações climáticas e produtivas. A base de dados incluiu variáveis climáticas de precipitação, temperatura e umidade relativa do ar, bem como

REMUNOM
ISSN 2178-6925

Received: 22/08/2026 - **Accepted:** 22/09/2026
Vol: 13.20
DOI: 10.66104/x4hqfw39
Pages: 1-22

dados referentes à quantidade de frutos perdidos e comercializados, conforme apresentado no Quadro 1.

Quadro 1. Banco de dados para análise em Random Forest.

Dimensão	Variável	Unidade	Fonte	Utilização na modelagem
Climática	Precipitação	mm	Acervo da propriedade	Análise de sensibilidade
Climática	Temperatura do ar	°C	INMET	Análise de sensibilidade
Climática	Umidade relativa do ar	%	INMET	Análise de sensibilidade
Produtiva	Frutos perdidos	kg	Acervo da propriedade	Análise de sensibilidade
Produtiva	Quantidade comercializada	kg	Acervo da propriedade	Variável-alvo e construção das defasagens temporais

Fonte: Os autores (2026).

As informações foram coletadas a partir de fontes oficiais e registros da propriedade, abrangendo o período de 2018 a 2023, com frequência mensal para todas as variáveis analisadas.

Durante a etapa de verificação da qualidade dos dados, foram identificados valores iguais a zero nas séries de temperatura e umidade relativa do ar em períodos nos quais esses registros não representavam valores meteorológicos plausíveis. Esses casos foram tratados como dados ausentes. Para evitar o uso de informações futuras no processo de previsão, a imputação foi realizada separadamente em cada janela de treinamento, utilizando a média mensal histórica calculada exclusivamente com dados anteriores ao período de teste.

A base original foi constituída por 72 observações mensais e sete colunas: ano, mês, precipitação, temperatura, umidade relativa do ar, quantidade de frutos perdidos e quantidade de frutos comercializados. A variável de interesse do estudo foi a quantidade de frutos comercializados, expressa em quilogramas, enquanto as demais informações foram avaliadas como potenciais preditores.

Análise de dados

Inicialmente, os dados foram tabulados e organizados em planilha eletrônica utilizando o software Microsoft Office Excel 2016, estruturados em sete colunas,

REMUNOM
ISSN 2178-6925
Received: 22/08/2026 - **Accepted:** 22/09/2026
Vol: 13.20
DOI: 10.66104/x4hqfw39
Pages: 1-22

correspondentes às seguintes variáveis: ano, mês, precipitação, umidade relativa do ar, temperatura, quantidade de frutos perdidos e quantidade de frutos comercializados.

Posteriormente, o conjunto de dados foi importado para o software RStudio, onde foram realizadas as etapas de tratamento, estruturação temporal, modelagem e avaliação do desempenho preditivo. As análises foram conduzidas preservando as unidades originais das variáveis, sem aplicação de normalização, pois algoritmo Random Forest não requer padronização das escalas para a construção das árvores de decisão.

A modelagem foi estruturada como um problema de previsão temporal com horizonte de um mês, no qual as informações disponíveis no mês *t* foram utilizadas para prever a quantidade comercializada no mês subsequente (*t*+1). Dessa forma, a variável resposta foi definida como a quantidade comercializada, em quilogramas, no mês *t*+1, evitando o emprego de informações que somente estariam disponíveis após o momento da previsão.

Para representar a dependência temporal da série, foram construídas variáveis defasadas a partir do histórico da quantidade comercializada. Foram considerados o valor observado no mês imediatamente anterior ao mês-alvo e o valor correspondente ao mesmo mês do ano anterior. Essas variáveis permitiram incorporar ao modelo, respectivamente, a persistência de curto prazo e o comportamento sazonal anual da quantidade comercializada.

A sazonalidade também foi representada por meio da transformação cíclica do mês-alvo em componentes seno e cosseno. Essa representação preserva a natureza circular dos meses do ano, considerando, por exemplo, a proximidade temporal entre dezembro e janeiro, sem impor uma relação linear entre os valores numéricos atribuídos aos meses.

A avaliação preditiva foi realizada respeitando a ordem cronológica das observações, por meio de janelas temporais expansivas. Foram estabelecidos três períodos de teste independentes: 2021, 2022 e 2023. Para a previsão de 2021, o modelo foi treinado exclusivamente com as observações anteriores; para 2022, a janela de treinamento foi ampliada com os dados disponíveis até 2021; e, para 2023,

REMUNOM
ISSN 2178-6925
Received: 22/08/2026 - **Accepted:** 22/09/2026
Vol: 13.20
DOI: 10.66104/x4hqfw39
Pages: 1-22

foram utilizados apenas os dados disponíveis até 2022. Esse procedimento assegurou que nenhuma informação posterior ao período previsto fosse empregada no treinamento do modelo.

Nos casos de ausência de registros de temperatura e umidade relativa do ar, os valores foram imputados a partir da média histórica correspondente ao mesmo mês do ano, calculada exclusivamente com as observações pertencentes à respectiva janela de treinamento. Esse procedimento foi repetido em cada janela temporal, evitando a utilização de informações futuras durante o pré-processamento.

O desempenho preditivo foi avaliado por meio do Erro Absoluto Médio (MAE), da Raiz do Erro Quadrático Médio (RMSE) e do coeficiente de determinação preditivo (R^2). O MAE e o RMSE foram calculados na unidade original da variável resposta, em quilogramas. O R^2 preditivo foi obtido a partir da relação entre a soma dos quadrados dos erros de previsão e a variabilidade dos valores observados, podendo assumir valores negativos quando o desempenho do modelo é inferior à previsão baseada na média do conjunto avaliado.

Para contextualizar o desempenho do Random Forest, suas previsões foram comparadas a métodos de referência (benchmarks), incluindo a persistência do mês anterior, o valor observado no mesmo mês do ano anterior, a média histórica mensal e um modelo de regressão linear. A comparação com modelos de referência é uma etapa importante da avaliação preditiva, permitindo verificar se métodos de maior complexidade proporcionam ganhos efetivos em relação a abordagens tradicionais (Herrera *et al*., 2019; Karbassi Yazdi *et al*., 2026).

Também foram conduzidas análises de sensibilidade para avaliar a contribuição incremental das variáveis climáticas e de outras informações disponíveis na base.

Modelagem com Random Forest

O Random Forest é um método de Aprendizado de Máquina pertencente à classe de *ensemble learning*, baseado na combinação de múltiplas árvores de decisão construídas a partir de amostras e subconjuntos aleatórios de variáveis. Em problemas de regressão, as previsões individuais das árvores são agregadas,

REMUNOM
ISSN 2178-6925

Received: 22/08/2026 - **Accepted:** 22/09/2026
Vol: 13.20
DOI: 10.66104/x4hqfw39
Pages: 1-22

contribuindo para a redução da variância e para maior estabilidade do modelo. Sua capacidade de representar relações não lineares e interações entre variáveis tem favorecido sua aplicação em sistemas agrícolas, especialmente em problemas que envolvem informações ambientais e produtivas (Galera *et al*., 2023; Nieto *et al*., 2024; Meher *et al*., 2024).

Neste estudo, o Random Forest foi empregado para prever a quantidade comercializada de lima ácida 'Tahiti' com horizonte de um mês. A modelagem foi realizada no ambiente R, utilizando o pacote *RandomForest.* O ajuste dos hiperparâmetros foi conduzido exclusivamente com os dados disponíveis nas respectivas janelas de treinamento, evitando a utilização de informações dos períodos de teste. Foram avaliadas diferentes combinações do número de variáveis candidatas em cada divisão das árvores (mtry) e do tamanho mínimo dos nós terminais (nodesize), mantendo-se 500 árvores (ntree = 500) em cada ajuste.

A configuração foi selecionada com base no erro *out-of-bag* (OOB) médio obtido nas janelas de treinamento, resultando em mtry = 1, nodesize = 5 e ntree = 500. Para garantir a reprodutibilidade dos resultados, foi utilizada semente aleatória fixa (set.seed = 123).

O modelo principal utilizou como preditores os componentes cíclicos do mês-alvo, representados pelas transformações seno e cosseno, a quantidade comercializada disponível no mês imediatamente anterior à previsão e a quantidade correspondente ao mesmo mês do ano anterior. As variáveis climáticas (precipitação, temperatura e umidade relativa do ar) e a quantidade de frutos perdidos foram avaliadas separadamente em análises de sensibilidade, permitindo verificar sua contribuição incremental para o desempenho preditivo.

A importância relativa dos preditores foi examinada por meio da variação percentual do erro quadrático médio após a permutação de cada variável (%IncMSE). Essa medida foi utilizada exclusivamente para avaliar a contribuição preditiva das variáveis no modelo, não sendo interpretada como evidência de relação causal.

REMUNOM
ISSN 2178-6925

Received: 22/08/2026 - **Accepted:** 22/09/2026
Vol: 13.20
DOI: 10.66104/x4hqfw39
Pages: 1-22

4. Resultados e Discussão

A avaliação do modelo Random Forest foi realizada por validação temporal, considerando exclusivamente previsões fora da amostra. Embora a base compreenda o período de 2018 a 2023, os anos de 2018 a 2020 constituíram a janela inicial de treinamento, enquanto 2021, 2022 e 2023 foram utilizados sequencialmente como períodos de teste. Dessa forma, foram obtidas 36 previsões mensais fora da amostra, preservando-se a ordem temporal dos dados.

O desempenho do Random Forest apresentou variação entre os períodos de teste (Tabela 1). Em 2021, foram obtidos MAE de 405.573,0 kg, RMSE de 440.692,4 kg e R^2 preditivo de -0,231. O valor negativo de R^2 indica desempenho inferior ao de uma previsão baseada na média dos valores observados naquele período, evidenciando dificuldade de generalização do modelo no primeiro ano de teste.

Em 2022, observou-se melhora do desempenho, com MAE de 297.651,1 kg, RMSE de 467.077,0 kg e R^2 de 0,580. O melhor resultado foi obtido em 2023, quando o modelo apresentou MAE de 134.076,3 kg, RMSE de 196.798,2 kg e R^2 de 0,862. Considerando conjuntamente as 36 previsões fora da amostra, o MAE foi de 279.100,1 kg, o RMSE de 387.771,5 kg e o R^2 preditivo de 0,530.

Tabela 1. Desempenho preditivo do modelo Random Forest na validação temporal.

Período de teste	MAE (kg)	RMSE (kg)	R^2 preditivo
2021	405.573,0	440.692,4	-0,231
2022	297.651,1	467.077,0	0,580
2023	134.076,3	196.798,2	0,862
Global (2021–2023)	279.100,1	387.771,5	0,530

Fonte: Os autores (2026).

A variação das métricas entre os anos demonstra que a capacidade de generalização do Random Forest não foi uniforme ao longo da série. Embora o desempenho tenha sido superior nos períodos posteriores, esse comportamento não deve ser atribuído exclusivamente à ampliação progressiva da janela de treinamento, uma vez que características específicas de cada período também podem influenciar a previsibilidade da quantidade comercializada. Em bases temporais reduzidas, essa variabilidade reforça a necessidade de avaliar o desempenho em períodos

REMUNOM
ISSN 2178-6925

Received: 22/08/2026 - **Accepted:** 22/09/2026
Vol: 13.20
DOI: 10.66104/x4hqfw39
Pages: 1-22

efetivamente posteriores aos utilizados no ajuste, em vez de considerar apenas medidas de ajuste obtidas sobre a própria amostra.

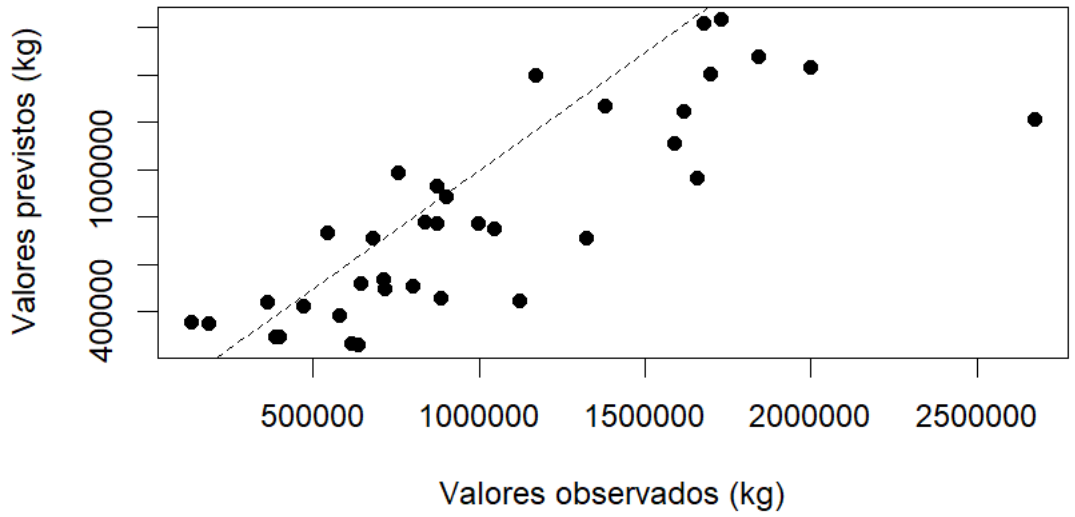
Estudos mostram que o modelo é eficaz na previsão da produção em quilogramas, com um elevado coeficiente de determinação (R^2), indicando que variáveis como dados climáticos e perdas são relevantes para explicar as variações na quantidade de frutos comercializados (Prasad *et al*., 2020; Pabitha; Benila; Suresh, 2023).

A Figura 2 apresenta a relação entre os valores observados e previstos pelo Random Forest nas 36 observações fora da amostra. Observa-se aproximação dos pontos à linha de concordância, embora ocorram desvios ao longo da faixa de valores analisada, particularmente em algumas observações de maior magnitude. Esse comportamento é compatível com o desempenho preditivo global do modelo (R^2 = 0,530) e evidencia capacidade de generalização moderada no período avaliado.

REMUNOM
ISSN 2178-6925

Received: 22/08/2026 - **Accepted:** 22/09/2026
Vol: 13.20
DOI: 10.66104/x4hqfw39
Pages: 1-22

Figura 1. Relação entre os valores observados e previstos da quantidade comercializada de lima ácida 'Tahiti' pelo modelo Random Forest na validação temporal (2021–2023).



Fonte: Os autores (2026).

Comparação com modelos de referência

A capacidade preditiva do Random Forest foi comparada com estratégias de referência de diferentes níveis de complexidade (Tabela 2). Entre os modelos simples, a previsão baseada no valor observado no mesmo mês do ano anterior apresentou o melhor desempenho, com MAE de 301.377,7 kg, RMSE de 408.620,1 kg e R^2 de 0,478. O resultado evidencia a relevância da recorrência anual na série de quantidade comercializada.

O Random Forest apresentou desempenho superior aos modelos baseados no mesmo mês do ano anterior, na persistência do mês anterior e na média histórica mensal. Em relação ao *baseline* sazonal, o Random Forest reduziu o MAE em aproximadamente 7,4% e o RMSE em 5,1%, enquanto o R^2 aumentou de 0,478 para 0,530. Esses resultados indicam que o algoritmo conseguiu incorporar informação adicional à simples repetição do padrão anual.

Entretanto, a regressão linear apresentou o melhor desempenho global entre os modelos avaliados, com MAE de 239.165,4 kg, RMSE de 343.104,6 kg e R^2 preditivo de 0,632. Portanto, embora o Random Forest tenha superado os

REMUNOM
ISSN 2178-6925

Received: 22/08/2026 - **Accepted:** 22/09/2026
Vol: 13.20
DOI: 10.66104/x4hqfw39
Pages: 1-22

benchmarks simples, não apresentou vantagem preditiva em relação ao modelo linear.

Tabela 2. Desempenho global dos modelos na previsão da quantidade comercializada, 2021 a 2023.

Modelo	MAE (kg)	RMSE (kg)	R^2 preditivo
Regressão linear	239.165,4	343.104,6	0,632
Random Forest	279.100,1	387.771,5	0,530
Mesmo mês do ano anterior	301.377,7	408.620,1	0,478
Persistência - mês anterior	345.950,9	503.138,8	0,209
Média histórica mensal	366.231,0	480.879,0	0,277

Fonte: Os autores (2026).

O Random Forest apresentou capacidade preditiva e superou estratégias simples de referência; contudo, nesta base, a regressão linear apresentou melhor desempenho. Isso indica que, neste conjunto de dados, maior complexidade não resultou em maior capacidade preditiva. A comparação entre diferentes classes de modelos é importante para verificar se o aumento da complexidade algorítmica resulta efetivamente em ganhos de desempenho preditivo, especialmente em aplicações agrícolas (Karbassi Yazdi *et al.,* 2026).

Sazonalidade e dependência temporal

A análise da função de autocorrelação da quantidade comercializada indicou dependência temporal expressiva. Foram observados coeficientes de autocorrelação de aproximadamente 0,677 na defasagem de um mês (*lag* 1) e 0,657 na defasagem de 12 meses (*lag* 12). Esses resultados indicam, respectivamente, persistência de curto prazo e recorrência anual na série.

A autocorrelação observada no *lag* 12 sustenta a utilização do valor correspondente ao mesmo mês do ano anterior como informação preditiva. Da mesma forma, a autocorrelação no *lag* 1 demonstra que a quantidade comercializada recente contém informação relevante para a previsão do período subsequente. A estrutura temporal observada também é consistente com o desempenho do *baseline* sazonal, que apresentou R^2 preditivo de 0,478 e foi o melhor entre os métodos simples avaliados.

No Random Forest, a importância preditiva das variáveis apresentou distribuição relativamente equilibrada. Considerando o modelo ajustado com a maior

REMUNOM
ISSN 2178-6925
Received: 22/08/2026 - **Accepted:** 22/09/2026
Vol: 13.20
DOI: 10.66104/x4hqfw39
Pages: 1-22

janela de treinamento, os valores de %IncMSE foram de 14,73% para a quantidade comercializada correspondente ao mesmo mês do ano anterior, 14,73% para o componente seno do mês-alvo, 13,66% para a quantidade comercializada do mês anterior e 12,47% para o componente cosseno do mês-alvo.4

Essas medidas indicam que tanto o histórico recente quanto a recorrência anual e a representação cíclica do período contribuíram para as previsões. A consideração explícita de componentes temporais e sazonais também tem sido empregada na aplicação do Random Forest a séries temporais com padrões de tendência e sazonalidade (Marzak *et al*., 2025).

Contribuição das variáveis climáticas

As variáveis precipitação, temperatura e umidade relativa do ar foram avaliadas adicionalmente por meio de análise de sensibilidade. Embora essas condições sejam agronomicamente relevantes para o desenvolvimento e a produção dos citros, sua importância biológica não implica necessariamente ganho de desempenho quando incorporadas a um modelo de previsão mensal.

A inclusão conjunta das três variáveis climáticas no Random Forest resultou, no período global de avaliação, em MAE de 334.969,0 kg, RMSE de 466.631,0 kg e R^2 de 0,320. Esses valores foram inferiores aos obtidos pelo modelo temporal sem as variáveis climáticas, que apresentou MAE de 279.100,1 kg, RMSE de 387.771,5 kg e R^2 de 0,530.

A redução do desempenho também foi observada individualmente nos três anos de teste. Dessa forma, nas condições avaliadas, as variáveis climáticas mensais utilizadas não acrescentaram capacidade preditiva ao modelo que já incorporava informações históricas e sazonais da quantidade comercializada.

Esse resultado não deve ser interpretado como ausência de influência das condições meteorológicas sobre a cultura. Ele indica apenas que, para esta base, neste horizonte de um mês e com as variáveis climáticas disponíveis, sua incorporação não proporcionou ganho incremental de previsão. A resposta produtiva às condições ambientais pode envolver defasagens temporais, interações com

REMUNOM
ISSN 2178-6925

Received: 22/08/2026 - **Accepted:** 22/09/2026
Vol: 13.20
DOI: 10.66104/x4hqfw39
Pages: 1-22

práticas de manejo e efeitos acumulados que não são integralmente representados por valores climáticos mensais contemporâneos.

Essa interpretação substitui a afirmação do manuscrito anterior de que chuva, temperatura e umidade apresentavam ligação direta com a quantidade comercializada com base no desempenho do Random Forest, algo que a análise anterior não permitia estabelecer.

Análises de sensibilidade

A quantidade de frutos perdidos foi avaliada separadamente devido à sua disponibilidade no processo de classificação. Sua inclusão no Random Forest resultou, globalmente, em MAE de 282.890,4 kg, RMSE de 406.214,9 kg e R^2 de 0,484, desempenho inferior ao modelo principal. Além da ausência de ganho preditivo global, a utilização dessa variável no modelo prospectivo principal foi evitada por sua relação com uma etapa operacional contemporânea à classificação dos frutos, reduzindo o risco de utilização de informação que não estaria necessariamente disponível no momento da previsão.

A inclusão do ano como variável adicional apresentou efeito limitado. O modelo contendo essa informação alcançou MAE de 271.210,0 kg, RMSE de 387.766,3 kg e R^2 de 0,530. Embora tenha ocorrido pequena redução do MAE em relação ao modelo principal, o RMSE permaneceu praticamente inalterado e o R^2 foi equivalente. Além disso, o efeito da inclusão do ano não foi uniforme entre os períodos de teste. Assim, a variável não foi mantida no modelo principal, priorizando-se uma estrutura mais parcimoniosa.

Em problemas de previsão multivariada, a seleção dos preditores constitui uma etapa relevante da modelagem, uma vez que a incorporação de novas variáveis deve ser avaliada quanto à sua contribuição efetiva para o desempenho preditivo (Jiménez *et al*., 2020).

Diagnóstico dos erros e limitações

A análise dos resíduos das 36 previsões fora da amostra indicou autocorrelação de 0,482 na primeira defasagem e de 0,206 na defasagem de 12

REMUNOM
ISSN 2178-6925

Received: 22/08/2026 - **Accepted:** 22/09/2026
Vol: 13.20
DOI: 10.66104/x4hqfw39
Pages: 1-22

meses. O teste de Ljung–Box aplicado especificamente ao primeiro *lag* indicou autocorrelação residual significativa (Q = 9,085; p = 0,0026), evidenciando dependência de curto prazo ainda não integralmente capturada pelo modelo.

Por outro lado, considerando conjuntamente as defasagens até 12 meses, o teste de Ljung–Box não foi significativo ao nível de 5% (Q = 19,245; p = 0,0828). Esse resultado sugere que parte relevante da estrutura temporal foi representada pelo modelo, embora permaneça dependência residual de curto prazo. A presença dessa autocorrelação deve ser considerada uma limitação e indica possibilidade de aprimoramento por meio de abordagens que representem mais explicitamente a dinâmica temporal da série.

Outras limitações estão relacionadas à dimensão e à origem da base de dados. O estudo utilizou 72 observações mensais referentes a uma única unidade produtiva, o que restringe a generalização dos resultados para outras propriedades ou regiões. A validação temporal adotada permite avaliar o desempenho em períodos posteriores aos utilizados no treinamento, mas não constitui validação externa.

Além disso, informações relacionadas ao manejo da cultura, como irrigação, adubação, práticas fitossanitárias e outras intervenções realizadas durante o ciclo produtivo, não foram incorporadas à modelagem. Esses fatores podem contribuir para as variações na quantidade comercializada e representar fontes de heterogeneidade não capturadas pelos preditores disponíveis.

A presença de registros ausentes em parte das séries meteorológicas constitui outra limitação. Embora a imputação tenha sido realizada exclusivamente a partir das informações disponíveis nas respectivas janelas de treinamento, evitando a utilização de dados futuros, valores estimados não substituem integralmente observações meteorológicas efetivamente registradas.

Por fim, a variação do desempenho entre os anos de teste e a superioridade da regressão linear evidenciam a necessidade de cautela na escolha do algoritmo. A aplicação do Random Forest em séries mais extensas, diferentes unidades produtivas e conjuntos de dados que incluam informações de manejo poderá permitir

REMUNOM
ISSN 2178-6925

Received: 22/08/2026 - **Accepted:** 22/09/2026
Vol: 13.20
DOI: 10.66104/x4hqfw39
Pages: 1-22

avaliar com maior robustez em quais condições sua capacidade de representar relações não lineares proporciona vantagem efetiva sobre modelos mais simples.

5. Conclusão

O Random Forest demonstrou capacidade para a previsão mensal da quantidade comercializada de lima ácida ‘Tahiti’, apresentando, nas 36 previsões fora da amostra, MAE de 279.100,1 kg, RMSE de 387.771,5 kg e R^2 preditivo de 0,530. O modelo superou estratégias simples baseadas na persistência, na média histórica mensal e no comportamento sazonal anual, mas apresentou desempenho inferior à regressão linear, que alcançou R^2 preditivo de 0,632. Esse resultado evidencia que, nas condições avaliadas, maior complexidade algorítmica não implicou necessariamente maior capacidade preditiva e reforça a importância da validação temporal e da comparação com modelos de referência.

Mas, para assegurar a robustez e a aplicabilidade do modelo ao longo do tempo, ressalta-se a importância da atualização contínua da base de dados, bem como da validação periódica do modelo, especialmente diante das variações climáticas e das mudanças nos sistemas de manejo.

Agradecimentos

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001.

Referências

Akinbile, Christopher O., et al. "Trends and spatial analysis of temperature and rainfall patterns on rice yields in Nigeria." Atmospheric Science Letters 21.3 (2020): e944.

Andrade, Adalberto, et al. "Uso de ciência de dados para predição do consumo de fertilizantes no Brasil." Brazilian e-Science Workshop (BreSci). SBC, 2020.

REMUNOM
ISSN 2178-6925

Received: 22/08/2026 - **Accepted:** 22/09/2026
Vol: 13.20
DOI: 10.66104/x4hqfw39
Pages: 1-22

Arantes, Leticia Tondato. "Drone aplicado na agricultura digital." Ipê Agronomic Journal (2019).

Bassan, Meire Menezes, et al. "Beneficiamento pós-colheita de lima ácida'Tahiti'afeta sua qualidade e conservação." Ciência Rural 46.1 (2015): 184-190.

da Silva Carvalho, Roberto Luís, and Angel Ramon Sanchez Delgado. "ESTIMATIVAS DAS MODELAGENS POR REDES NEURAIS TIPO GMDH E BOX-JENKINS PARA AS SÉRIES DE TEMPERATURAS MÁXIMAS E MÍNIMAS DO MUNICÍPIO DE ARIQUEMES (RO)." Revista Brasileira de Climatologia 26 (2020)

Cavalcante, WS da S., et al. "Tecnologias e inovações no uso de drones na agricultura." Brazilian Journal of Development 8.1 (2022): 7108-7117.

Coutinho, Eluã Ramos, Robson Mariano Silva, and Angel Ramon Sanchez Delgado. "Utilização de técnicas de inteligência computacional na predição de dados meteorológicos." Revista Brasileira de Meteorologia 31.1 (2016): 24-36.

Frimpong, Samuel Akwasi, et al. "Machine and deep learning in agricultural engineering: A comprehensive survey and meta-analysis of techniques, applications, and challenges." Computers 14.10 (2025): 438.

Furtado, Kenedy Daniel Calegari, et al. "O papel dos drones na agricultura 4.0 e 5.0: Auxílio tecnológico para uma agricultura eficiente, produtiva e sustentável." Tecnologia e Inovação Em Ciências Agrárias e Biológicas Avanços para a Sociedade Atual (2023).

Gatera, Antoine, et al. "Comparison of random forest and support vector machine regression models for forecasting road accidents." Scientific African 21 (2023): e01739.

Gurgel, Fábio de Lima, et al. "Early rootstock selection under ‘tahiti’acid lime crown in Capitão Poço, Pará State, Brazil." Acta Scientiarum. Agronomy 46 (2024): e63302.

Gurgel, Fábio de Lima, et al. "Rootstocks for citrus orchards in the state of Pará." Pesquisa Agropecuária Brasileira 60 (2025): e04134.

Jiménez, Fernando, et al. “Feature Selection Based Multivariate Time Series Forecasting: An Application to Antibiotic Resistance Outbreaks Prediction.” *Artificial Intelligence in Medicine*, vol. 104, 2020, article 101818.

REMUNOM
ISSN 2178-6925

Received: 22/08/2026 - **Accepted:** 22/09/2026
Vol: 13.20
DOI: 10.66104/x4hqfw39
Pages: 1-22

Kamilaris, Andreas, and Francesc X. Prenafeta-Boldú. "A review of the use of convolutional neural networks in agriculture." The Journal of Agricultural Science 156.3 (2018): 312-322.

Karbassi Yazdi, Amir, et al. “Forecasting Crop Yields in Rainfed India: A Comparative Assessment of Machine Learning Baselines and Implications for Precision Agribusiness.” *Agriculture*, vol. 16, no. 1, 2026, article 65.

Martorano, L. G., et al. "Estudos climáticos do estado do Pará, classificação climática (Köppen) e deficiência hídrica (Thornthwaite, Mather)." (1993).

Mattos Jr., Dirceu de, et al. "Lima ácida Tahiti no Centro de Citricultura: pesquisas e difusão de tecnologia." Informativo Centro de Citricultura Sylvio Moreira 308.309 (2021): 1-12.

Marzak, Zahira, et al. “Forecasting Multivariate Time Series with Trend and Seasonality: A Random Forest Approach.” *Innovative Intelligent Industrial Production and Logistics*, Springer Nature, 2025, pp. 128–144.

Meher, Bharat Kumar, et al. "Forecasting stock prices of fintech companies of India using random forest with high-frequency data." Journal of Open Innovation: Technology, Market, and Complexity 10.1 (2024): 100180.

Nieto, Paulino José García, et al. "Predicting the critical superconducting temperature using the random forest, MLP neural network, M5 model tree and multivariate linear regression." Alexandria Engineering Journal 86 (2024): 144-156.

Pabitha, C., S. Benila, and A. Suresh. "A digital footprint in enhancing agricultural practices with improved production using machine learning." (2023).

Parisi, M. C. M., C. M. Henrique, and P. Prati. "Perdas pós-colheita: um gargalo na produção de alimentos." Pesquisa & Tecnologia 9.2 (2012): 1-5.

Pinto, João V. de N., et al. "Fruit yield and gas exchange of Tahiti lime at different irrigation depths in the Amazon." Revista Brasileira de Engenharia Agrícola e Ambiental 27.6 (2023): 463-471.

Prasad, N. R., N. R. Patel, and Abhishek Danodia. "Crop yield prediction in cotton for regional level using random forest approach: NR Prasad et al." spatial information research 29.2 (2021): 195-206.

REMUNOM
ISSN 2178-6925

Received: 22/08/2026 - **Accepted:** 22/09/2026
Vol: 13.20
DOI: 10.66104/x4hqfw39
Pages: 1-22

Rivera-Hernández, Benigno, et al. "Seasonal changes in physiological responses and yield of Citrus latifolia under high-density planting and different soil moisture tensions." Horticulturae 11.12 (2025): 1472.

Roberts, Michael J., et al. "Comparing and combining process-based crop models and statistical models with some implications for climate change." *Environmental Research Letters* 12.9 (2017): 095010.

Rustick, Andressa, José Airton Azevedo dos Santos, and Cidmar Ortiz dos Santos. "Aplicação de modelos de redes neurais artificiais na previsão da produção, importação e consumo de fertilizantes no Brasil." *Peer Review* 5.16 (2023): 1-14.

Shawon, Sarowar Morshed, et al. "Crop yield prediction using machine learning: An extensive and systematic literature review." *Smart Agricultural Technology* 10 (2025): 100718.

Sim, Ha Seon, et al. "Prediction of strawberry growth and fruit yield based on environmental and growth data in a greenhouse for soil cultivation with applied autonomous facilities." *Horticultural Science and Technology* 38.6 (2020): 840-849.

Sousa, Denis de Pinho, et al. "Estimation of evapotranspiration and single and dual crop coefficients of acai palm in the Eastern Amazon (Brazil) using the Bowen ratio system." *Irrigation Science* 39.1 (2021): 5-22.

Srivastava, Anoop Kumar, et al. "Climate change and citrus nutrition: mechanisms, nutrient imbalances, and inclusive management." *Frontiers in Plant Science* 17 (2026): 1798826.

Tajudeen, Tawakalitu Titilayo, et al. "The effect of climate change on food crop production in Lagos State." *Foods* 11.24 (2022): 3987.

Wolfart, Simone Beatriz, et al. "ESTUDO DE MÉTODOS PARA PREVISÃO DA PRODUÇÃO DE GRÃOS NO BRASIL VIA SÉRIES TEMPORAIS." *Cadernos do IME-Série Estatística* 49 (2020): 17-17.

Yao, Zhaoyuan, et al. "Improving high-resolution soil moisture mapping in China via dense ground-based observations and multi-source data fusion." *Agricultural Water Management* 333 (2026): 110570.