

**UNIVERSIDADE ESTADUAL PAULISTA – UNESP
CÂMPUS DE JABOTICABAL**

**MODELAGEM AGROMETEOROLÓGICA POR INTELIGÊNCIA ARTIFICIAL PARA
PREVISÃO DA PRODUTIVIDADE DE PALMEIRAS NA AMAZÔNIA ORIENTAL**

José Reinaldo da Silva Cabral de Moraes
Engenheiro Agrônomo

**UNIVERSIDADE ESTADUAL PAULISTA – UNESP
CÂMPUS DE JABOTICABAL**

**MODELAGEM AGROMETEOROLÓGICA POR INTELIGÊNCIA ARTIFICIAL PARA
PREVISÃO DA PRODUTIVIDADE DE PALMEIRAS NA AMAZÔNIA ORIENTAL**

**José Reinaldo da Silva Cabral de Moraes
Orientador: Prof. Dr. Glauco de Souza Rolim**

Tese apresentada à Faculdade de
Ciências Agrárias e Veterinárias – UNESP,
Câmpus de Jaboticabal, como parte das
exigências para a obtenção do título de
Doutor em Agronomia (Produção Vegetal).

M827m Moraes, José Reinaldo da Silva Cabral de
Modelagem agrometeorológica por inteligência artificial para
previsão da produtividade de palmeiras na Amazônia Oriental /
José Reinaldo da Silva Cabral de Moraes. -- Jaboticabal, 2021
139 p.

Tese (doutorado) - Universidade Estadual Paulista (Unesp),
Faculdade de Ciências Agrárias e Veterinárias, Jaboticabal
Orientador: Glauco de Souza Rolim

1. Agrometeorologia. 2. Amazônia brasileira. 3. Arecaceae.
4. Machine Learning. 5. Python. I. Título.

Sistema de geração automática de fichas catalográficas da Unesp.
Biblioteca da Faculdade de Ciências Agrárias e Veterinárias,
Jaboticabal. Dados fornecidos pelo autor(a).

CERTIFICADO DE APROVAÇÃO

TÍTULO DA TESE: **MODELAGEM AGROMETEOROLÓGICA POR INTELIGÊNCIA ARTIFICIAL PARA PREVISÃO DA PRODUTIVIDADE DE PALMEIRAS NA AMAZÔNIA ORIENTAL**

AUTOR: JOSÉ REINALDO DA SILVA CABRAL DE MORAES

ORIENTADOR: GLAUCO DE SOUZA ROLIM

Aprovado como parte das exigências para obtenção do Título de Doutor em AGRONOMIA (PRODUÇÃO VEGETAL), pela Comissão Examinadora:



Assinado de forma digital por
Glauco de Souza Rolim
Dados: 2021.08.04 06:45:05
+03'00'

Prof. Dr. GLAUCO DE SOUZA ROLIM (Participação Virtual)
Engenharia e Ciências Exatas / FCAV / UNESP - Jaboticabal

Pesquisadora Dra. LUCIETA GUERREIRO MARTORANO (Participação Virtual)
NAPT-Embrapa Amazônia Oriental / Belém/PA



Assinado de forma digital por Glauco de Souza Rolim
Dados: 2021.08.04
06:45:36 -03'00'

Prof. Dr. ROGÉRIO TEIXEIRA DE FARIA (Participação Virtual)
Departamento de Engenharia e Ciências Exatas (DECEX) / FCAV / UNESP - Jaboticabal



Assinado de forma digital por Glauco de Souza Rolim
Dados: 2021.08.04
06:45:54 -03'00'

Prof. Dr. LUCAS EDUARDO DE OLIVEIRA APARECIDO (Participação Virtual)
Instituto Federal de Educação, Ciência e Tecnologia do Sul de Minas Gerais - Campus Muzambinho / Muzambinho/MG



Assinado de forma digital por Glauco de Souza Rolim
Dados: 2021.08.04
06:46:13 -03'00'

Prof. Dr. ALEXANDRE DAL PAI (Participação Virtual)
Bioprocessos e Biotecnologia / Faculdade de Ciências Agrônomicas de Botucatu - UNESP



Assinado de forma digital por Glauco de Souza Rolim
Dados: 2021.08.04
06:46:36 -03'00'

Jaboticabal, 02 de junho de 2021

DADOS CURRICULARES DO AUTOR

JOSÉ REINALDO DA SILVA CABRAL DE MORAES - Nascido em Castanhal, Pará, no dia 27 de abril de 1992, filho de José da Silva Moraes e Lilia da Silva Cabral de Moraes, natural de Castanhal, Pará. Coursou o ensino fundamental no colégio estadual Aristedes Santa Rosa, no município de Inhangapi, Pará e o ensino médio no Instituto de Educação Darwin, no município de Castanhal, Pará, tendo finalizado no ano de 2009. Ingressou no ensino superior no ano de 2010 no curso de Engenharia Agrônômica pela Universidade Federal Rural da Amazônia (UFRA), Campus Belém, obtendo o título de Engenheiro Agrônomo em janeiro de 2015. Durante a graduação foi bolsista do Programa de Educação Tutorial (PET) Agronomia, sob a tutoria do Prof. Dr. Carlos Augusto Cordeiro Costa por quatro anos e realizando iniciação científica na Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA) Amazônia Oriental, também durante quatro anos, na área de Agrometeorologia, sob orientação da pesquisadora Dra. Lucieta Guerreiro Martorano. Realizou durante um ano estágio curricular na área de agrometeorologia na UFRA, sob a orientação do Prof. Dr. Paulo Jorge de Oliveira Ponte de Souza, desenvolvendo seu trabalho de conclusão de curso com a calibração do modelo DSSAT para o feijão caupi. Em agosto de 2015, iniciou o curso de Mestrado em Agronomia, no Programa de Produção Vegetal, na área de Modelagem Agrometeorológica, pela Universidade Estadual Paulista “Júlio de Mesquita Filho” – Campus de Jaboticabal, São Paulo, sob a orientação do Prof. Dr. Glauco de Souza Rolim, concluindo no ano de 2017. No ano de 2017 iniciou o doutorado na mesma instituição sob a orientação do professor Glauco Rolim, submetendo sua tese em abril de 2021 à banca examinadora para obtenção do título de doutor em Agronomia.

O mundo é um livro, e quem fica sentado em
casa lê somente uma página.

Santo Agostinho

Aos meus pais, José Moraes e Lilia Moraes por
toda dedicação, amor e carinho que me foi dado,
e com seus sacrifícios, me proporcionaram
caminhos de muitas conquistas e realizações.
Minhas vitórias são graças a vocês.

DEDICO

Ao meu Avô Sebastião de Oliveira Moraes (*In memorian*). Que
apesar do pouco tempo que passamos juntos,
foi o suficiente para formar homem que sou hoje.
A minha noiva Aline Michelle da Silva Barbosa,
pelo seu carinho, amor, dedicação e apoio
incondicional nas batalhas da vida.

OFEREÇO

AGRADECIMENTOS

A Deus, por ter me guiado e protegido, mostrando sempre o melhor caminho a seguir.

Ao professor Glauco Rolim, orientador e amigo. Seus ensinamentos científicos e de vida me moldaram como profissional e pessoa, obrigado por toda paciência e parceria durante esses anos de mestrado e doutorado.

À professora Lucieta Guerreiro Martorano, pela amizade, confiança e por ter acreditado no meu potencial no início da minha graduação. Nossos caminhos terem se cruzado logo no começo da minha vida acadêmica e científica, foi fundamental para eu enxergar outras oportunidades e desafios.

Agradeço a Universidade Federal Rural da Amazônia (UFRA), instituição que me formei como engenheiro agrônomo e conquistei grandes amigos. Aos professores dessa universidade, em especial ao professor Paulo Jorge de Oliveira Ponte de Souza, pelas excelentes aulas de agrometeorologia na graduação, e posteriormente pela orientação no meu trabalho de conclusão de curso.

Ao Programa de Educação Tutorial (PET) Agronomia, em especial ao professor Carlos Augusto Cordeiro Costa, pelos conselhos, anos de amizade, ensinamentos. Este programa o qual fui bolsista durante quatro anos foi fundamental para minha formação como profissional. Também agradeço aos amigos bolsistas do PET, por toda convivência, viagens, experimentos de campo e amizade.

Registro meus agradecimentos a minha família, em especial as minhas irmãs, Liliane Moraes, Josi Moraes e Viviane Feio, por todo carinho, apoio e pelos sobrinhos maravilhosos que já me deram. Aos meus tios Reinado Moraes e Ieda Terra, por todo acolhimento, confiança e carinho.

A minha madrinha Ambrósia Marinho (*In memorian*) que sempre acreditou e torceu pelo meu sucesso. À minha avó paterna, Raimunda Monteiro da Silva Moraes e materna, Nazaré Cabral, e meus avós de coração, Maria das Dores Terra e Antônio Terra da Trindade (*In memorian*), pelos ensinamentos e carinhos frequentes.

Meus agradecimentos aos funcionários do departamento de ciências exatas da UNESP, em especial aos amigos Carlão, Vanessa, Zezé, Shirlei e Adriana. E aos amigos Bruna, Gustavo, Daniel e Mara.

Aos amigos do grupo de pesquisa GAS da UNESP, Lucas Aparecido, Taynara Valeriano, Victor Moreto, Kamila Meneses, João Trevizoli, Valter e Aline.

Aos irmãos que também passaram por Jaboticabal e foram trilhar novos caminhos, Daniel Pinheiro, Francisco Carlos, Flávio José e Raphael Leone. Amizades construídas durante a pós-graduação e que levo para vida. Registro também meus agradecimentos aos companheiros do Galáticos futebol, amigos que conquistei durante o mestrado e doutorado.

Aos amigos do Instituto Federal do Mato Grosso do Sul, em especial ao Cícero Teixeira, Gustavo Valente, Daniel Zimmermann, Matheus Bornelli, pela amizade e apoio no início da minha carreira docente.

À Embrapa Amazônia Oriental, em especial aos pesquisadores Dra. Socorro Padilha e Dr. João Tomé, pela concessão dos dados de campo para a realização dessa pesquisa. Aos amigos do Laboratório de Agrometeorologia da Embrapa Amazônia Oriental, por toda parceria científica, carinho e amizade.

À empresa Sococo em nome do Dr. Paulo Lins, assim como a empresa Marboges em nome do Dr. Abreu Pina, pelo fornecimento dos dados de campo para realização da minha tese.

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

A todos que contribuíram para minha formação de vida e profissional, meus sinceros agradecimentos.

SUMÁRIO

RESUMO.....	iii
ABSTRACT	iv
CAPÍTULO 1 – CONSIDERAÇÕES GERAIS.....	1
1.1 Palma de óleo.....	1
1.2 Coqueiro	4
1.3 Açaizeiro	5
1.4 Inteligência artificial.....	6
REFERÊNCIAS.....	10
CAPÍTULO 2 – Modelagem Agrometeorológica a partir de inteligência artificial para prever a produtividade da Palma de óleo no estado do Pará, Amazônia. 19	
2.1 INTRODUÇÃO	21
2.2 MATERIAL E MÉTODOS	24
2.2.1 Base de dados	24
2.2.2 Componentes de água no solo	28
2.2.3 Caracterização fenológicas da palma de óleo	30
2.2.4 Modelos de Machine Learning	32
2.2.5 Análise de dados	35
2.3 RESULTADOS	36
2.4 DISCUSSÃO	54
2.5 CONCLUSÃO.....	59
2.6 REFERÊNCIAS.....	59
CAPÍTULO 3 – Previsão de produtividade de frutos de açaí (<i>Euterpe oleracea</i> Mart.) na região Nordeste paraense na Amazônia Oriental modelado por inteligência artificial	68
3.1 INTRODUÇÃO	70
3.2 MATERIAL E MÉTODOS	72
3.2.1 Caracterização fenológica do açaizeiro	76
3.2.2 Modelos de Machine learning	79

3.3 RESULTADOS E DISCUSSÃO.....	82
3.4 CONCLUSÃO.....	94
3.5 REFERÊNCIAS.....	94
CAPÍTULO 4 – Previsão agrometeorológica da produtividade do coqueiro modelado por Machine Learning	104
4.1 INTRODUÇÃO	106
4.2 MATERIAL E MÉTODOS.....	108
4.2.1 Área de estudo e origem dos dados	108
4.2.1 Caracterização fenológica do coqueiro.....	114
4.2.2 Modelos de Machine Learning.....	115
4.2.3 Processamento gráfico e estatísticos	116
4.3 RESULTADOS	117
4.4 DISCUSSÃO	128
4.5 CONCLUSÃO.....	131
4.6 REFERÊNCIAS.....	132

MODELAGEM AGROMETEOROLÓGICA POR INTELIGÊNCIA ARTIFICIAL PARA PREVISÃO DA PRODUTIVIDADE DE PALMEIRAS NA AMAZÔNIA ORIENTAL

RESUMO

A expansão de palmeiras de alto valor comercial e industrial como a palma de óleo, o açaizeiro e o coqueiro, necessitam de informações e estratégias sustentáveis para seus sistemas de cultivos. Modelos agrometeorológicos atuam como uma ferramenta exploratória para os tomadores de decisão e aliado a inteligência artificial (IA) auxiliam no planejamento do uso da terra para cultivos agrícolas anuais e/ou perenes. Tais métodos desempenham um papel importante na previsão de produtividade, no entanto, poucas pesquisas têm sido aplicadas visando a modelagem de palmeiras, que sejam suficientemente simples e ao mesmo tempo incorporem o conhecimento fenológico e climático suficientes para ser estudada em locais com diferentes condições de crescimento e práticas de manejo. Neste sentido, objetiva-se avaliar o desempenho de modelos de IA na modelagem agrometeorológica para previsão da produtividade da palma de óleo, açaí e do coco na Amazônia Oriental. O estudo foi realizado em várias regiões do estado do Pará. Os dados de produtividade de palma de óleo e coco foram disponibilizados por empresas produtoras situadas na Amazônia, e de açaí de áreas experimentais da Embrapa Amazônia oriental. Foram utilizados diferentes modelos de IA para prever a produtividade das culturas. Para todos os modelos, foram separados 70% dos dados para treinamento e 30% para teste. Os modelos ajustados foram: Regressão linear múltipla (RLM), Random Forest (RF), Redes neurais artificiais Multilayer Perceptron (MLP) e Support Vector Machine de base radial RBF (SVM_RBF), Linear (SVM_Linear) e Polinomial (SVM_Poly), usando a linguagem de programação Python. Os modelos foram avaliados por meio da acurácia, precisão e tendência, comparando dados observados e previstos. Os dados observados em campo e obtidos pelos modelos de IA foram comparados pelo R^2 , MAPE, RMSE e ME. Modelos de IA com 4 meses antes da colheita apresentaram acurácia média (MAPE) de 22% para previsão de produtividade da palma de óleo. Para o açaí, variaram de 4 a 7 meses entre os diferentes tipos de manejo e épocas do ano com MAPE médio igual a 19%. Para o coqueiro, os resultados apontaram previsões com até 11 meses antes da colheita com MAPE médio igual a 15%. Houve uma perda de precisão dos modelos ao prever dados extremos, com subestimação na previsão em situações de altas produtividades e superestimação em baixas produtividades. No geral, os modelos de IA demonstraram desempenho importante para prever a produtividade das palmeiras estudadas, subsidiando produtores, agroindústrias e outros planejadores a melhor tomada de decisão.

Palavras-chave: Agrometeorologia, Amazônia brasileira, Arecaceae, Machine Learning, Python

AGROMETEOROLOGICAL MODELING WITH ARTIFICIAL INTELLIGENCE IN FORECASTING PALM YIELD IN THE EASTERN AMAZON

ABSTRACT

The expansion of high commercial and industrial value palm trees such as the oil palm, the açaizeiro and the coconut palm, require information and sustainable strategies for their cropping systems. Agrometeorological models consist of an exploratory tool for decision makers and allied with artificial intelligence (AI) assist in land use planning for annual and/or perennial agricultural crops. This methods play an important role in yield prediction, however, few researches have been applied aiming at palm tree modeling that are simple enough and at the same time incorporate enough phenological and climatic knowledge to be studied in locations with different growing conditions and management practices. This, we aimed to evaluate the performance of AI models in agrometeorological modeling for predicting the yield of oil palm, açaí and coconut in the Eastern Amazon. The study was conducted in several regions of the state of Pará, Brazil. Yield data for oil palm and coconut were provided by production companies located in the Amazon, and for açaí from experimental areas of Embrapa Amazônia oriental. Several AI models were used to predict crop yield. We use 70% of the data to calibrate and 30% to test the AI models. The fitted AI models were: Multiple Linear Regression (MLR), Random Forest (RF), Multilayer Perceptron (MLP) artificial neural networks, and Support Vector Machine of radial RBF (SVM_RBF), Linear (SVM_Linear) and Polynomial (SVM_Poly) bases. We using the Python programming language to perform the analysis. The models were evaluated using accuracy, precision, and trend, comparing observed and predicted data. Field-observed data and data obtained by the AI models were compared by R^2 , MAPE, RMSE, and ME. AI models with 4 months before harvest showed average accuracy (MAPE) of 22% for predicting oil palm yields. We observed that it ranged from 4 to 7 months between the different types of management and times of the year with an average MAPE equal to 19% for açaí. As for the coconut palm, the results showed predictions up to 11 months before harvest with an average MAPE equal to 15%. There was a loss of accuracy of the models when predicting extreme data, with underestimation in the prediction in situations of high yields and overestimation in low yields. Overall, the AI models demonstrated an important performance in predicting the yield of the studied palms, supporting producers, agro-industries and other planners to make better decisions.

Keywords: Agrometeorology, Brazilian Amazon, Arecaceae, Machine Learning, Python

CAPÍTULO 1 – CONSIDERAÇÕES GERAIS

A maioria das palmeiras se desenvolve em ecossistemas tropicais (Eiserhardt et al., 2011) sendo possível a expansão desses cultivos no Brasil, principalmente em áreas antropizadas. Os climas tropicais são caracterizados por padrões de temperatura previsíveis e com pequena amplitude térmica (Moron; Robertson; Wang, 2019). Contudo, os regimes de precipitação pluvial não tendem a ser simples e previsíveis, existindo grandes variações no ano e entre os anos (Moraes et al., 2020).

Os produtores agrícolas, agroindústrias e outros, necessitam ter alta produtividade e sustentabilidade em seus negócios. Neste sentido, conhecer a variabilidade do clima por meio de modelos agrometeorológicos de forma prévia na cadeia produtiva de qualquer cultura, é fundamental para estabelecer estratégias e melhores tomada de decisão (Klompenburg et al., 2020).

Dentre as palmeiras de alto valor comercial e de produção, destaca-se a palma de óleo o coqueiro e o açaizeiro, do gênero *Elaeis*, *Cocos* e *Euterpe*, respectivamente, que tem na Amazônia brasileira suas principais áreas de produção e comercialização. Diversos trabalhos vêm estudando a influência do clima no desenvolvimento e efeitos na produção dessas culturas, entre esses destaca-se, Benezoli et al. (2021), Liu et al. (2021) e Oettli et al. (2018) para palma de óleo, Santos et al. (2020), Samarasinghe et al. (2018) e Pathmeswaran et al. (2018) para o coqueiro e Moraes et al. (2020) e Viana et al. (2020) para o açaizeiro.

1.1 Palma de óleo

A palma de óleo (*Elaeis guineensis* Jacq.) uma das culturas oleaginosas mais importantes do mundo, originária da África Tropical. Trazida para o Brasil pelos escravos, se adaptou ao clima tropical úmido da Bahia, mas foi na região amazônica que o cultivo obteve as maiores produtividades (Vijay et al., 2016). A produção de óleo de palma é cinco vezes maior por unidade de terra do que outras culturas oleaginosas, como a soja, que, juntamente com a crescente demanda global de óleo vegetal e biocombustíveis, impulsiona sua lucratividade (Lamade et al., 2016).

Entre as culturas com maior expansão de área colhida nos trópicos úmido, a palma de óleo se destaca entre as três maiores, com crescimento de 176% entre os anos de 2000 a 2019 (Figura 1), sendo que 87% da produção é concentrada principalmente no Sudeste Asiático (FAO, 2021). No Brasil a região do nordeste paraense concentra aproximadamente 98% da produção, com mais de 160 mil hectares colhidos (Figura 2) (IBGE 2021).

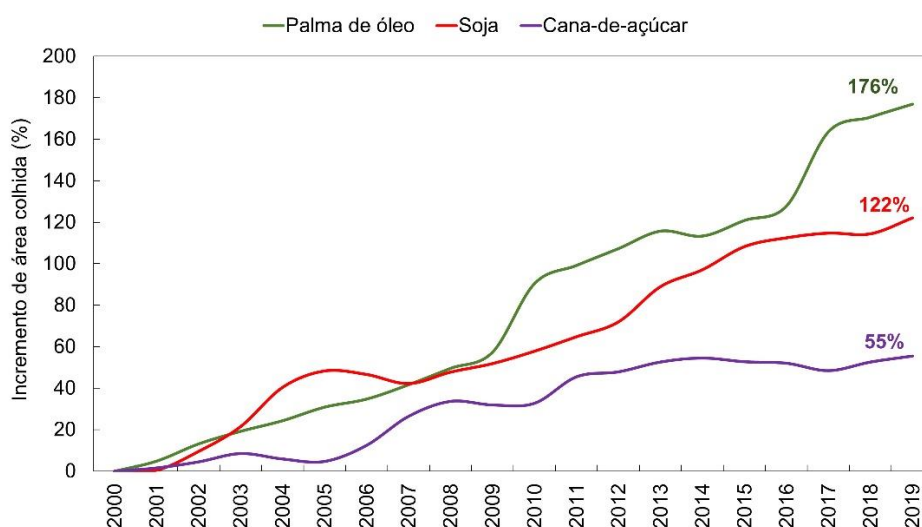


Figura 1. Principais culturas agrícolas com expansão de área colhida em porcentagem nos países tropicais úmidos.

Fonte. (FAO, 2021).

Sua inserção na cadeia do Biodiesel no Brasil foi incentivada pelo Programa Palma de Óleo, como cultura viável e rentável na recuperação de áreas em vias de degradação (Brandão et al., 2021, Garrett et al., 2019). A palma de óleo é extremamente versátil, sendo aproveitado óleos da semente, do mesocarpo e o óleo de palmiste (Awalludin et al., 2015; Mushtaq; Abdullah; Ani, 2015). Aproveitam-se ainda cachos, resíduos do processo de extração de óleo (glicerina), entre outros usos, além do seu potencial na produção de combustíveis alternativos e bioeletrecidade (Pirker et al., 2016).

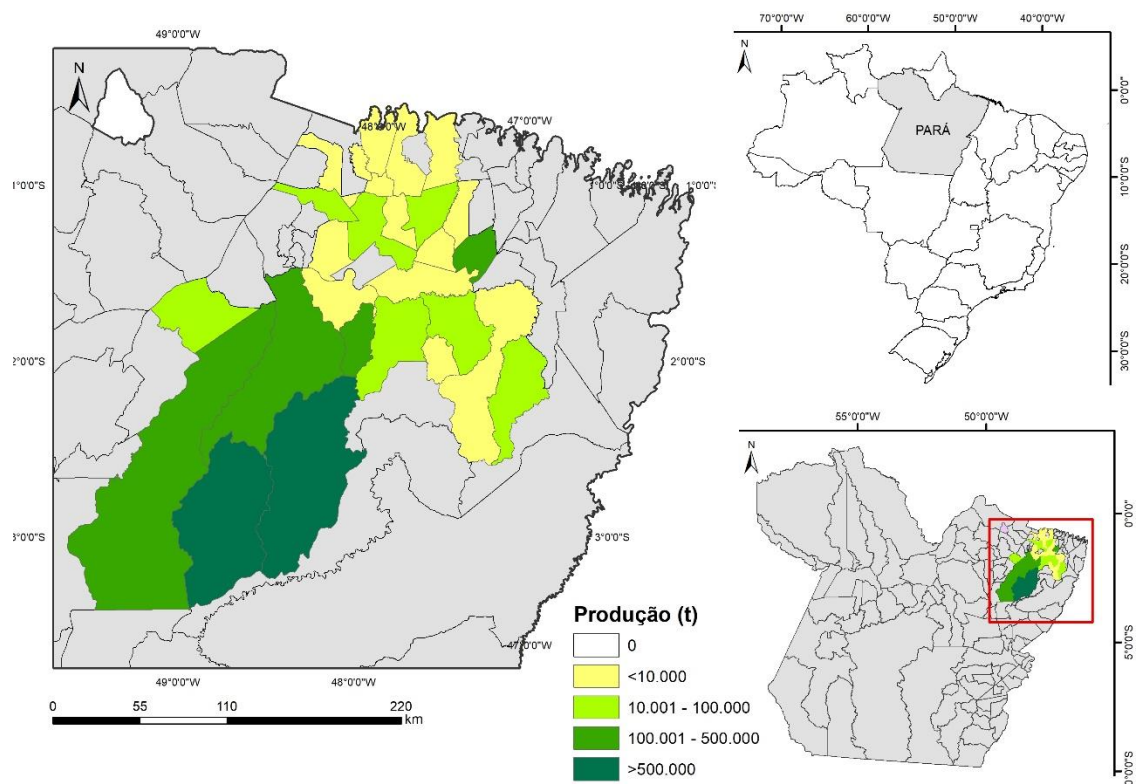


Figura 2. Mapa das principais áreas produtoras de palma de óleo no estado do Pará.

Fonte. (IBGE, 2021).

As evidências de aquecimento global, oscilações nos preços do petróleo e mudanças para uma matriz energética renovável incluem a palma de óleo como planta com alto potencial para produção de biocombustíveis, impulsionando sua lucratividade (Sheil et al., 2009).

O potencial dessa espécie na produção de óleo e conversão em biocombustível ganha importância em áreas com alto grau de degradação, inserindo-as ao sistema produtivo, minimizando a pressão do desmatamento, capturando carbono e mitigando emissões de gases de efeito estufa (GEE), além de promover oferta de emprego ao setor agrícola e industrial na Amazônia (Zarin et al., 2016).

1.2 Coqueiro

Outra palmeira de destaque, o coqueiro (*Cocos nucifera* L), é uma das mais importantes do mundo, cultivada em mais de 12 milhões de hectares de terras nas planícies costeiras tropicais e subtropicais (FAOSTAT, 2021). Cerca de 10 milhões de agricultores familiares são altamente dependentes do produto desta palmeira e muitos outros em áreas rurais e semi-urbanas que possuem reduzido número de áreas plantadas, contribuem para a subsistência dessas populações (Rethinam, 2006).

O coqueiro é uma palmeira de grande interesse econômico, uma vez que são utilizados todos os seus componentes: frutos, folhas, polpa, água, casca e fibras. Cada parte da planta pode produzir itens que têm valor comunitário, bem como fornecer uma gama de produtos comerciais e industriais (Nguyen et al., 2015). Esses produtos incluem aqueles com propriedades nutricionais e medicinais (Foale, 2003; Perera et al., 2009), com um núcleo maduro (endosperma sólido) que contém fibras comestíveis, proteínas, lipídios e minerais inorgânicos.

No Brasil, a principal atividade econômica envolvendo esta espécie destinam-se à produção de coco seco in natura, coco ralado, leite de coco, óleo de coco e outros derivados do coco seco e água de coco, a partir do coco verde (Benassi; Fanton; Santana, 2013).

Outros produtos derivados da fruta incluem bebidas, amêndoas frescas e leite (Lim, 2012), além de produtos refinados, como, o óleo virgem e a fibra da casca. O óleo virgem (extraído a baixa temperatura) possui propriedades antioxidantes potentes (Marina et al., 2009) e antimicrobianas (Chakraborty e Mitra, 2008) e possui potenciais ações anticancerígenas (Koschek et al., 2007).

No Brasil, o avanço do cultivo ocorre não só pela evolução em patamares produtivos, que condicionam ao país lugar de destaque entre os maiores produtores mundiais, mas também, pela expansão da área plantada, principalmente em regiões não tradicionais (Ferreira; Warwick; Siqueira, 2018).

O cultivo de coqueiro no Brasil tradicionalmente acontece na região Nordeste, mas nos últimos trinta anos as áreas estão se expandindo para outras regiões do Brasil, principalmente no Sudeste, Centro Oeste e Norte (Figura 3).

Esta expansão ocorre também em agroecossistemas frágeis como em regiões de Mata Atlântica e zonas litorâneas (Martins; Jesus Júnior, 2014).

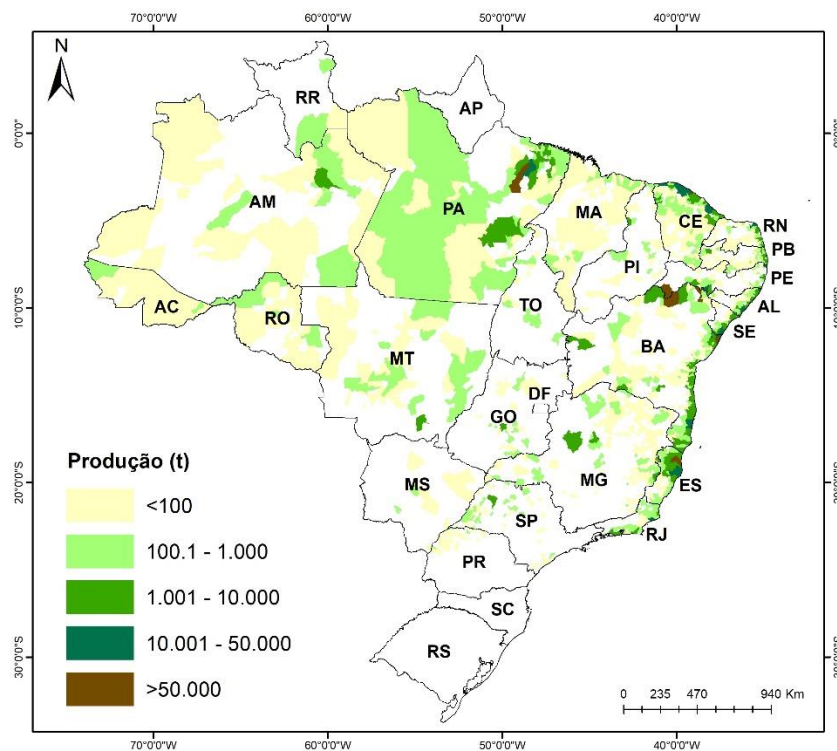


Figura 3. Principais áreas produtoras de coco no Brasil.

Fonte. (IBGE, 2021).

1.3 Açaizeiro

O açaizeiro (*Euterpe oleracea* Mart) que é uma palmeira de hábito perene possuindo frutos bastante valorizados para o preparo de bebidas, pertence à família Arecaceae, englobando aproximadamente 200 gêneros e cerca de 2600 espécies, cuja distribuição é predominantemente tropical e subtropical (Goldel et al., 2015). Na Amazônia esta família é representada por 39 gêneros e um número entre 150 a 180 espécies (Kahn, 1997).

O açaizeiro desempenha importante papel socioeconômico e ambiental para as regiões produtoras (Vallejo et al., 2014). Silva et al. (2006), verificaram aumentos dos retornos sociais do cultivo, após a inserção tecnológica na produção na ordem de 238 milhões de reais no ano de 2005, por meio de linhas de crédito de financiamento pelos bancos.

O novo contexto do mercado do açaí, tanto ao nível nacional como internacional, tem se caracterizado por uma demanda crescente e superior à oferta, o que pressiona os preços, sobretudo com o aumento das exportações (Homma; Santana, 2009). Nesse mercado a bebida é comercializada como *commodity*, onde o fator determinante é o preço, que oscila devido a sazonalidade da safra que se concentra no Pará, Amapá e Maranhão (Nogueira et al., 2005). No entanto, apresenta uma produção que ainda demanda de pesquisas que visem o melhor manejo do cultivo, assim como o conhecimento de áreas com potencial climático para sua expansão.

A distribuição espacial das espécies de palmeiras está predominantemente condicionada aos fatores ambientais limitantes (Bazzaz, 1998), tornando a biodiversidade do planeta vulnerável aos fenômenos naturais e processos resultantes das ações antrópicas sobre o ambiente.

Com a busca da expansão desses cultivos e melhor planejamento de safra, o uso de modelos de prognósticos agrometeorológicos auxilia no planejamento agrícola, identificando as condicionantes de variáveis atmosféricas que possam causar perdas de produção e buscar soluções estratégicas para tomada de decisão. No entanto, esse monitoramento prévio requer o uso de vários conjuntos de dados, não sendo uma tarefa trivial, em vez disso, consiste em várias etapas complicadas (Filippi et al., 2019). Neste sentido, o uso da inteligência artificial tem sido uma importante ferramenta de apoio a decisão no meio agrícola, podendo determinar diferentes padrões e correlações a partir de conjuntos de dados, ajudando a melhorar a eficiência no setor agrícola, como a produtividade da safra, irrigação, monitoramento de safra, entre outros (Talaviya et al., 2020).

1.4 Inteligência artificial

Na produção agrícola, existem várias metodologias e processos diferentes que requerem um consumo de energia bastante elevado (Jha et al. 2019). Ao mesmo tempo, o mercado exige produtos de saída de alta qualidade (Singh; Singh; Kaur, 2021).

Com o aumento da informação no campo, se amplia o horizonte de aplicabilidade e uso de modelos matemáticos. Dados históricos de clima e dados meteorológicos durante a estação de crescimento, bem como a previsão de curto, médio e longo prazo, desempenham papel importante nas aplicações de modelagem agrícola (Hogenboom, 2000).

A demanda por informações rápidas e consistentes impõe que pesquisadores, consultores, agricultores e gestores do setor agrícola ampliem as aquisições e conhecimento em modelos de simulação (Martorano, 2008). Dessa forma, as relações existentes entre os elementos climáticos e a produtividade de uma cultura podem ser simulados de forma previa antes da colheita, através de modelos agrometeorológicos (Rolim et al., 2008), usando algoritmos de Inteligência Artificial (IA) (Klompenburg; Kassahun; Catal, 2020).

A IA inicia com o estudo sobre a capacidade dos computadores de pensar sem a interferência humana (Russell; Norvig, 2011), definindo-se basicamente como uma tecnologia que funcione como o cérebro humano, que incluem aprendizagem, raciocínio e autocorreção (Parekh; Shah; Shah, 2020) para automatizar tarefas complexas (Tada, 2016). Neste conceito, softwares e sistemas inteligentes são desenvolvidos, baseados em como o cérebro humano pensa, aprende, toma decisões e trabalham enquanto resolvem um problema (Talaviya et al., 2020). Esses softwares são alimentados com dados de treinamento e, além disso, esses dispositivos inteligentes nos fornecem a saída desejada para cada entrada válida, assim como o cérebro humano (Jani et al., 2020).

A IA envolve duas grandes subáreas classificadas como, Deep learning (DL) e Machine Learning (ML) (Patel et al., 2020; Sukhadia et al., 2020). A DL é uma subárea de ML, que por sua vez é uma subárea da IA (Sharma; Sharma; Jindal, 2021) (Figura 4). Enquanto IA é a ciência de fazer máquinas e programas inteligentes, ML é a capacidade de aprender algo sem ser explicitamente programado e DL é o aprendizado por redes neurais profundas (Kodali; Sahu, 2016).

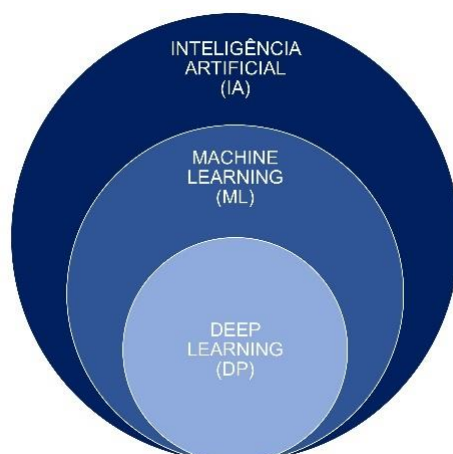


Figura 4: Inteligência artificial, Aprendizado de máquina e Aprendizagem profunda. Adaptado de Goodfellow et al. (2016).

Conforme discutido em Goodfellow et al. (2016), inicialmente o foco da inteligência artificial era enfrentar e solucionar problemas intelectualmente difíceis para os seres humanos, mas relativamente diretos para os computadores, problemas que podem ser descritos por uma lista de regras formais e matemáticas. Entretanto, observa-se que o verdadeiro desafio é resolver as tarefas fáceis para as pessoas, mas difíceis de descrever formalmente, problemas que resolvemos intuitivamente, que parecem automáticos, como reconhecer palavras ou faces nas imagens.

Para prover uma solução a esses problemas mais intuitivos é necessário que os computadores aprendam com a experiência e compreendam as informações em termos de hierarquia de conceitos, com cada conceito definido por sua relação com conceitos mais simples (Goodfellow et al., 2016), assim como o ser humano faz.

Os algoritmos de ML agrupam um grande número de métodos que permitem aos computadores generalizar a partir da experiência, aprendendo com um conjunto de dados de treinamento a reconhecer padrões em dados de alta dimensão (Bonetto; Latzko, 2020; Subasi, 2020). Em geral o ML requer um grande conjunto de dados para se treinar e prever o resultado. Entretanto, estudos recentes têm apontado o uso de ML como modelos preditivos usando pequenos conjuntos de dados (Shaikhina et al., 2015).

Patrício e Rieder (2018) definem que os processos de ML consistem basicamente nas seguintes etapas:

- Preparação e aquisição dos dados: etapa inicial que consiste em obter os dados adequados que contenham características a serem consideradas na aprendizagem.
- Seleção das características de Interesse: identificar as características mais significativas para o problema que será abordado.
- Seleção de algoritmos: consiste em selecionar o algoritmo mais adequado para tratar o problema em estudo. Como exemplo na agricultura temos a Regressão Linear Multipla, Random Forest, Redes Neurais Artificiais, Support Vector Machine e outros.
- Seleção de parâmetros: alguns algoritmos precisam ser ajustados por parâmetros que requerem experimentação para serem definidos, dessa forma os ajustes de seus parâmetros são fundamentais para um bom desempenho do modelo.
- Treinamento: Selecionado o algoritmo mais adequado e seus parâmetros, o treinamento consiste na construção do modelo computacional que será utilizado na predição das respostas aos novos dados.
- Teste: Validação dos resultados quanto a precisão e acurácia dos dados de treinamento.

Os algoritmos de ML permitem que se analise grandes volumes de dados independentemente da complexidade, com rapidez e precisão. Vastos são as aplicações de ML na agricultura, que passam deste a modelos preditivos de safras, como observado nos trabalhos de (Pant et al., 2021; Luciano et al., 2021) na detecção de doenças (Shah et al., 2016; Chung et al., 2016), qualidade de grãos (Zareiforush et al., 2015; Vithu; Moses, 2016), otimização de irrigação e aplicação de herbicidas (Talaviya et al., 2020) e tantas outras aplicações.

Poucos trabalhos têm investigado aplicações de modelos de IA em cultivos perenes como as palmeiras de forma geral e suas respostas produtivas em função do clima. O limitado número de trabalhos se deve principalmente ao longo período de monitoramento dessas culturas, aumentando os custos operacionais

de campo para aquisição de dados. Neste sentido, este estudo busca avaliar modelos agrometeorológicos para previsão de produtividade de palmeiras comerciais como a palma de óleo, coco e açaí na Amazônia Oriental, usando inteligência artificial.

REFERÊNCIAS

Awalludin MF, Sulaiman O, Hashim R, Aidawati WN, Nadhari W (2015) An overview of the oil palm industry in Malaysia and its waste utilization through thermochemical conversion, specifically via liquefaction. *Renewable and Sustainable Energy Reviews*, 50: 1469 – 1484. <http://dx.doi.org/10.1016/j.rser.2015.05.085>.

Bazzaz FA (1998) Elevated CO₂ and plant production in the 21st century: can we feed billions and preserve biological diversity. In : Garab G (ed) *Photosynthesis: mechanisms and effects*, V. Kluwer Academic, Dordrecht.

Benassi AC, Fanton CJ, Santana EM (2013) O cultivo do coqueiro-anão-verde: Tecnologias de produção. Vitória, ES: Incaper, 120 p. il. (Incaper. Documentos, 227).

Benezoli VH, Imbuzeiro HMA, Cuadra SV, Colmanetti, MAA, Araújo, AC, Stiegler C, Motoike, SY (2021) Modeling oil palm crop for Brazilian climate conditions. *Agricultural Systems*. 190: 103130. DOI: <https://doi.org/10.1016/j.agsy.2021.103130>.

Bonetto R, Latzko V (2020) Machine learning. *Computing in Communication Networks*. <https://doi.org/10.1016/B978-0-12-820488-7.00021-9>.

Brandão F, Schoneveld G, Pacheco P, Vieira I, Piraux M, Mota D (2021) The challenge of reconciling conservation and development in the tropics: Lessons from Brazil's oil palm governance model. *World Development* 139:105268. <https://doi.org/10.1016/j.worlddev.2020.105268>.

Chakraborty M, Mitra A (2008) The antioxidant and antimicrobial properties of the methanolic extract from *Cocos nucifera* mesocarp. *Food Chem.* v. 107, p. 994–999. doi:10.1016/j.foodchem. 2007.08.083.

Chung CL, Huang KJ, Chen SY, Lai MH, Chen YC, Kuo YF (2016) Detecting Bakanae disease in rice seedlings by machine vision. *Comput. Electron. Agric.* 121, 404–411. <https://doi.org/10.1016/j.compag.2016.01.008>.

Eiserhardt WL, Svenning JC, Kissling WD, Balslev H (2011) Geographical ecology of the palms (Arecaceae): determinants of diversity and distributions across spatial scales. *Annals of Botany*, 108: 1391–1416.

FAO (2021). FAOSTAT Data: Production – Crops. Food and Agriculture Organization of the United Nations (FAO) (Accessed 07 February 2021). <http://www.fao.org/faostat/en/#data/QC>.

Ferreira JMS, Warwick DRN, Siqueira LA (2018). A cultura do coqueiro no Brasil. Embrapa, Brasília, DF, p. 508.

Filippi P, Jones EJ, Wimalathunge NS, Somarathna PDSN, Pozza LE, Ugbaje SU, Bishop TFA (2019) An approach to forecast grain crop yield using multilayered, multi-farm data sets and machine learning. *Precis. Agric.* 1–15. <https://doi.org/10.1007/s11119-018-09628-4>.

Foale M (2003) The coconut odyssey: the bounteous possibilities of the tree of life. ACIAR Monography. Canberra, v. 101.

Garrett RD, Levy S, Carlson KM, Gardner TA, Godar J, Clapp J, Villoria N (2019) Criteria for effective zero-deforestation commitments. *Global Environmental Change*, 54, 135–147. <https://doi.org/10.1016/j.gloenvcha.2018.11.003>.

Goldel B, Kissling WD, Svenning JC (2015) Geographical variation and environmental correlates of functional trait distributions in palms (Arecaceae) across the New World. *Botanical Journal of the Linnean Society*, v. 179, p. 602–617.

Goodfellow I, Bengio Y, Courville A. *Deep learning*. MIT press, 2016.

Homma AKO, Santana ACA (2009) agroindústria na Região Norte, In: Zibetti DW & Barroso LA (Eds.) *Agroindústria: uma análise no contexto socioeconômico e jurídico brasileiro*. v.1. São Paulo, LEUD. p.19-43.

Hoogenboom G (2000) Contribution of agrometeorology to the simulation of crop production and its applications. *Agricultural and Forest Meteorology*, v. 103, n. 1, p. 137–157.

IBGE (2021). Instituto Brasileiro de Geografia e Estatística. Sistema IBGE de Recuperação Automática – SIDRA. (Accessed 16 March 2021). <https://sidra.ibge.gov.br/home/pimpfbr/brasil>.

Jani K, Chaudhuri M, Patel H, Shah M (2020) Machine learning in films: an approach towards automation in film censoring. *J. of Data, Inf. and Manag.* 2, 55–64. <https://doi.org/10.1007/s42488-019-00016-9>

Jha K, Doshi A, Patel P, Shah M (2019) A comprehensive review on automation in agriculture using artificial intelligence. *Artificial Intelligence in Agriculture*. 2, 1–12.

Kahn F (1997) *Les palmiers de l'eldorado*. Paris: Éditions de l'Oprstom. *Journal of Nutrology*. p. 252.

Klompenburg TV, Kassahun A, Catal C (2020). Crop yield prediction using machine learning: A systematic literature review. *Computers and Electronics in Agriculture*, 177, 105709. <https://doi.org/10.1016/j.compag.2020.105709>.

Kodali RK, Sahu A (2016) An IoT based soil moisture monitoring on Losant platform. 2nd International Conference on Contemporary Computing and Informatics. IEEE. pp. 764–768.

Koschek PR, Alviano DS, Alviano CS, Gattass CR (2007) The husk fiber of *Cocos nucifera* L. (Palmae) is a source of anti-neoplastic activity. *Brazilian J Med Biol Res.* v.40, p. 1339–1343. doi:10.1590/ s0100-879x2006005000153.

Lamade E, Tcherkez G, Darlan NH, Rodrigues RL, Fresneau C, Mauve C, Lamothe-Sibold M, Sketriené D, Ghashghaie J (2016) Natural ¹³C distribution in oil palm (*Elaeis guineensis* Jacq.) and consequences for allocation pattern. *Plant, Cell and Environment.* v. 39, p. 199–212.

Lim TK (2012) *Cocos nucifera*. In: Lim TK (ed) *Edible medicinal and non-medicinal plants*. Springer-Verlag Berlin, Berlin, p. 301–334. doi:10.1007/978-90-481-8661-7_45.

Liu Y, Heuvelink GBM, Bai Z, He P, Xu Xinpeng, Ding W, Huang S (2021) Analysis of spatio-temporal variation of crop yield in China using stepwise multiple linear regression. *Field Crops Research*, 264:108098. <https://doi.org/10.1016/j.fcr.2021.108098>.

Luciano ACS, Picoli MCA, Duft DG, Rocha JV, Leal MRLV, Maire GL (2021) Empirical model for forecasting sugarcane yield on a local scale in Brazil using Landsat imagery and random forest algorithm. *Computers and Electronics in Agriculture* 184, 106063. DOI: <https://doi.org/10.1016/j.compag.2021.106063>.

Marina AM, Man YBC, Nazimah SAH, Amin I (2009) Antioxidant capacity and phenolic acids of virgin coconut oil. *Int J Food Sci Nutr.* v. 60, p.114–123. doi:10.1080/09637480802549127.

Martins CR, Jesus Júnior LA (2014) Produção e Comercialização de Coco no Brasil Frente ao Comércio Internacional: Panorama 2014. 51 p. (Documentos / Embrapa Tabuleiros Costeiros, 1517- 1329; 184.

Martorano LG, Faria RT, Bergamaschi H, Dalmago GA (2008) Evaluation of the COPGRO/DSSAT model performance for simulating plant growth and grain yield of soybeans, subjected to no-tillage and conventional systems in the subtropical southern Brazil. *Italian Journal of Agronomy*. 3, 795-796.

Moraes JRSC, Rolim GS, Martorano LG, Aparecido LEO, Oliveira MSP, Neto JTF (2019) Agrometeorological models to forecast açaí (*Euterpe oleracea* Mart.) yield in the Eastern Amazon. *J Sci Food Agric*, 100: 1558–1569. <http://dx.doi.org/10.1002/jsfa.10164>.

Moron V, Robertson AW, Wang L (2019) Weather Within Climate: Subseasonal Predictability of Tropical Daily Rainfall Characteristics. *Sub-Seasonal to Seasonal Prediction*, p. 47-64. <https://doi.org/10.1016/B978-0-12-811714-9.00003-6>.

Mushtaq F, Abdullah TAT, Mat R, Ani FN (2015) Optimization and characterization of bio-oil produced by microwave assisted pyrolysis of oil palm shell waste biomass with microwave absorber. *Bioresour Technol*.

Nguyen QT, Bandupriya HDD, Villalobos AL, Sisunandar S, Foale M, Adkins SW (2015) Tissue culture and associated biotechnological interventions for the improvement of coconut (*Cocos nucifera* L.): a review. *Planta* v. 242, p. 1059–1076.

Nogueira OL, Figueirêdo FJC, Muller AA (2005) Açaí. Embrapa Amazônia Oriental. Belém, Pará. Embrapa Amazônia Oriental. (Sistemas de Produção, 4). p.137.

Oettli P, Behera SK, Yamagata T (2018) Climate Based Predictability of Oil Palm Tree Yield in Malaysia. *Scientific Reports*, 8:2271. doi:10.1038/s41598-018-20298-0.

Pant J, Pant RP, Singh MK, Singh DP, Pant H (2021) Analysis of agricultural crop yield prediction using statistical techniques of machine learning. *Materials Today: Proceedings*. <https://doi.org/10.1016/j.matpr.2021.01.948>.

Parekh V, Shah D, Shah M (2020) Fatigue Detection Using Artificial Intelligence Framework. *Augmented Human Research*. 5:5. <https://doi.org/10.1007/s41133-019-0023-4>.

Patel D, Shah D, Shah M (2020) The Intertwine of Brain and Body: A Quantitative Analysis on How Big Data Influences the System of Sports. *Ann. Data. Sci.* 7, 1–16. <https://doi.org/10.1007/s40745-019-00239-y>.

Pathmeswaran C, Lokupitiya E, Waidyarathne KP, Lokupitiya RS (2018). Impact of extreme weather events on coconut productivity in three climatic zones of Sri Lanka. *European Journal of Agronomy* 96, 47–53. DOI: <https://doi.org/10.1016/j.eja.2018.03.001>.

Patrício DI, Rieder R (2018) Computer vision and artificial intelligence in precision agriculture for grain crops: A systematic review. *Computers and Electronics in Agriculture* 153, 69–81. DOI: <https://doi.org/10.1016/j.compag.2018.08.001>.

Perera PIP, Vidhanaarachchi VRM, Gunathilake TR, Yakandawala DMD, Hoche V, Verdeil JL, Weerakoon LK (2009) Effect of plant growth regulators on ovary culture of coconut (*Cocos nucifera* L.). *Plant Cell Tiss Org.* v. 99, p. 73–81. doi:10.1007/s11240-009-9577-z.

Pirker J, Mosnier A, Kraxner F, Havlík P, Obersteiner M (2016) What are the limits to oil palm expansion? *Global Environ. Change* 40, 73–81. <https://doi.org/10.1016/j.gloenvcha.2016.06.007>.

Rethinam P (2006) Asian and Pacific coconut community activities, achievements and future outlook. ACIAR Proceedings Series, v. 125, p.15–21.

Rolim GS, Ribeiro RV, Azevedo FA, Camargo MBP, Machado EC (2008) Previsão do número de frutos a partir da quantidade de estruturas reprodutivas em laranjeiras. *Revista Brasileira de Fruticultura*. 30, 48-53.

Russell SJ, Norvig P (2011) *Artificial Intelligence: A Modern Approach*, third ed. Pearson. Global Edition.

Samarasinghe CRK, Meegahakumbura MK, Dissanayaka HDMAC, Kumarathunge D, Perera L (2018) Variation in yield and yield components of different coconut cultivars in response to within year rainfall and temperature variation. *Scientia Horticulturae*, 238, 51–57. DOI: <https://doi.org/10.1016/j.scienta.2018.03.058>.

Santos MMS, Lacerda CF, Neves ALR, Sousa CHCS, Ribeiro AA, Bezerra MA, Araújo ICS, Gheyi HR (2020) Ecophysiology of the tall coconut growing under different coastal areas of northeastern Brazil. *Agricultural Water Management* 232, 106047. <https://doi.org/10.1016/j.agwat.2020.106047>.

Shah JP, Prajapati HB, Dabhi VK (2016) A survey on detection and classification of rice plant diseases. In: 2016 IEEE International Conference on Current Trends in Advanced Computing (ICCTAC), 1–8. doi:<https://doi.org/10.1109/ICCTAC.2016.7567333>.

Shaikhina T, Lowe D, Daga S, Briggs D, Hggins R, Khovanova N (2015) Machine learning for predictive modelling based on small data in biomedical engineering. *IFAC Pap Online*, 48–20:469–474. <https://doi.org/10.1016/j.ifacol.2015.10.185>.

Sharma N, Sharma R, Jindal N (2021) Machine Learning and Deep Learning Applications-A Vision. Global Transitions Proceedings. 2, 24-28. <https://doi.org/10.1016/j.gltp.2021.01.004>.

Sheil D, Casson A, Meijaard E, Van Noordwijk M, Gaskell J, Sunderland-Groves J, Wertz K, Kanninen M (2009) Palm Oil Impacts and Opportunities in Southeast Asia: What do we know and what do we need to know? Occasional Paper. CIFOR, Bogor, Indonésia. v. 51 p. 80.

Silva IM, Santana AC, Reis MS (2006) Análise dos retornos sociais oriundos de adoção tecnológica na cultura do açaí no Estado do Pará. Amazônia: Ciência & Desenvolvimento, v. 2, p. 25-37.

Singh G, Singh A, Kaur G (2021) Role of Artificial Intelligence and the Internet of Things in Agriculture. Artificial Intelligence to Solve Pervasive Internet of Things Issues, 317-330. <https://doi.org/10.1016/B978-0-12-818576-6.00016-2>.

Subasi A (2020) Machine learning techniques. Practical Machine Learning for Data Analysis Using Python, p. 91-202. <https://doi.org/10.1016/B978-0-12-821379-7.00003-5>.

Sukhadia A, Upadhyay K, Gundeti M, Shah S, Shah M (2020) Optimization of smart traffic governance system using artificial intelligence. Augment Hum Res 5, 13. <https://doi.org/10.1007/s41133-020-00035-x>.

Talaviya T, Shah D, Patel N, Yagnik H, Shah M (2020) Implementation of artificial intelligence in agriculture for optimisation of irrigation and application of pesticides and herbicides. Artificial Intelligence in Agriculture, 4, 58–73. <https://doi.org/10.1016/j.aiia.2020.04.002>.

Vallejo MI, Galeano G, Bernal R, Zuidem PA (2014) The fate of populations of Euterpe oleracea harvested for palm heart in Colombia. Forest Ecology and Management. v. 318, p. 274–284.

Viana LF, Homma AKO, Menezes AJEA, Santos JC, Farias Neto JT (2020) Viabilidade econômica do cultivo de açaizeiro (*euterpe oleracea* mart.) irrigado no nordeste paraense. International Journal of Development Research. 10, 39177-39182. <https://doi.org/10.37118/ijdr.19655.08.2020>.

Vijay V, Pimm SL, Jenkins CN, Smith SJ (2016) The Impacts of Oil Palm on Recent Deforestation and Biodiversity Loss. **PLoS ONE**. v. 11: e0159668. doi:10.1371/journal.pone. 0159668.

Vithu P, Moses JÁ (2016) Machine vision system for food grain quality evaluation: A review. Trends Food Sci. Technol. 56, 13–20. <https://doi.org/10.1016/j.tifs.2016.07. 011>.

Zareiforush H, Minaei S, Alizadeh MR, Banakar A (2015) Potential applications of computer vision in quality inspection of rice: a review. Food Eng. Rev. 7 (3), 321–345. <https://doi.org/10.1007/s12393-014-9101-z>.

Zarin DJ, Harris NL, Baccini A, Aksenov D, Hansen MC, Azevedo-Ramos C, Azevedo, T, Margono BA, Alencar AC, Gabris C, Allegretti A, Potapov P, Farina M, Walker WS, Shevade VS, Loboda TV, Turubanova S, Tyukavina A (2016) Can carbon emissions from tropical deforestation drop by 50% in 5 years?. Glob Chang Biol. v. 22, p.1336-47.

CAPÍTULO 2 – Modelagem agrometeorológica a partir de inteligência artificial para prever a produtividade da Palma de óleo no estado do Pará, Amazônia

Resumo: A modelagem agrometeorológica tem avançado a partir de análises integradas entre diferentes variáveis respostas biofísicas, apoiadas em diferentes áreas do conhecimento científico, como a matemática, física, estatística, geomática e lógica em tecnologia da informação. Os prognósticos utilizando soluções de modelagem com rotinas de inteligência artificial (IA), têm possibilitado combinar, por exemplo, variáveis climáticas e fisiológicas de uma determinada cultura melhorando o desempenho dos modelos preditivos. Os prognósticos agrometeorológicos visam subsidiar à tomada de decisão quanto a adoção de ajustes ou novas técnicas de manejo para atingir as produtividades estimadas pelos modelos. Objetiva-se com este trabalho avaliar o desempenho de diferentes modelos de IA para a previsão de produtividade da palma de óleo no estado do Pará. Foram utilizados dados de variáveis climáticas e fenológicas da cultura obtidos de levantamentos cedidos para a pesquisa. Os dados foram coletados em uma área comercial de 583 hectares com palma de óleo, localizada no município de Moju, Pará. Foram utilizados dados mensais de produtividade entre os anos de 2003 a 2018, sendo excluídos os anos entre 2008 a 2012, devido alta incidência da doença ‘Amarelecimento Fatal’, na área de estudo. Dados diários de precipitação pluvial (P em mm), temperatura máxima (Tmax em °C) radiação solar (RAD em MJ m⁻² dia⁻¹) e umidade relativa (UR em %) foram obtidos usando dados de estação meteorológica de superfície próxima ao local de estudo. O balanço hídrico foi calculado pela metodologia de Thornthwaite e Mather, sendo posteriormente estimada a porcentagem de armazenagem de água no solo. Os modelos de IA ajustados foram: Regressão linear múltipla (RLM), Random Forests (RF), Redes neurais artificiais – Multilayer Perceptron (MLP) e Support Vector Machines de base radial RBF (SVM_RBF), Linear (SVM_Linear) e Polinomial (SVM_Poly), sendo testados pelo R², MAPE, RMSE e ME. O desempenho dos modelos de IA com 4 meses antes da colheita representaram os melhores resultados para previsão de produtividade, com R² > 0.70, RMSE < 0.50 t h⁻¹, MAPE < 26% e ME (≤ -0.02 e ≥ -0.15 t h⁻¹). O modelo SVM_Poly apresentou os melhores ajustes, com R² = 0.86, RMSE = 0.39, MAPE = 16% e ME = -0.06. O modelo SVM_Linear foi o que apresentou os menores desempenhos, com R² = 0.71, RMSE = 0.48 t h⁻¹, MAPE = 24.64% e ME = -0.12 t h⁻¹. Os modelos de ML são indicados para prever a produtividade da palma de óleo em áreas cultivadas no Pará.

Palavras-chaves: Amazonia, *Elaeis guineensis* Jacq., Machine Learning, Modelagem agrícola, Python

Agrometeorological modeling using artificial intelligence to forecast oil palm yield in Pará state, Amazon

ABSTRACT

Agrometeorological modeling has advanced, from integrated analyses between different variables biophysical responses, supported in different areas of scientific knowledge, such as mathematics, physics, statistics, geomatics and logic in information technology. Predictions using modeling solutions with artificial intelligence (AI) routines have made it possible to combine, for example, climatic and physiological variables of a given crop, improving the performance of predictive models. The agrometeorological predictions aim to support decision making regarding the adoption of adjustments or new management techniques to achieve the yields estimated by the models. The objective of this work was to evaluate the performance of different AI models to predict oil palm productivity in the state of Pará. Data of climatic variables and crop phenology obtained from surveys provided for the research were used. The data were collected in a commercial area of 583 hectares with oil palm, located in the municipality of Moju, Pará. Monthly yield data from the years 2003 to 2018 were used, excluding the years between 2008 and 2012 due to the high incidence of the disease 'Fatal Yellowing' in the study area. Daily data of rainfall (P in mm), maximum temperature (Tmax in °C) solar radiation (RAD in MJ m⁻² day⁻¹) and relative humidity (RH in %) were obtained using data from surface weather station near the study site. The water balance was calculated by Thornthwaite and Mather methodology, and subsequently the percentage of water storage in the soil was estimated. The fitted AI models were: Multiple Linear Regression (MLR), Random Forests (RF), Artificial Neural Networks - Multilayer Perceptron (MLP) and Support Vector Machines of radial basis RBF (SVM_RBF), Linear (SVM_Linear) and Polynomial (SVM_Poly), being tested by R², MAPE, RMSE and ME. The performance of the AI models at 4 months before harvest represented the best results for yield prediction, with R² > 0.70, RMSE < 0.50 t h⁻¹, MAPE < 26% and ME (≤ -0.02 and ≥ -0.15 t h⁻¹). The SVM_Poly model showed the best fits, with R² = 0.86, RMSE = 0.39, MAPE = 16% and ME = -0.06. The SVM_Linear model was the lowest performer, with R² = 0.71, RMSE = 0.48 t h⁻¹, MAPE = 24.64% and ME = -0.12 t h⁻¹. The ML models are the most suitable for predicting oil palm productivity in cultivated areas in Pará.

Keywords: Amazon, *Elaeis guineensis* Jacq., Machine Learning, Agricultural modeling, Python

2.1 INTRODUÇÃO

A palma de óleo (*Elaeis guineensis* Jacq.) ou dendezeiro é uma planta perene que faz parte de um setor industrial globalizado, complexo e multissetorial, envolvendo pequenos, médio e grandes produtores, agroindústria, consumidores, cientistas e governo (Godswill et al., 2016). A palma de óleo se destaca pela produção e comercialização de óleo vegetal mais expressiva no mundo, apresentando no ano de 2019 uma produção de 74.60 milhões de toneladas de óleo beneficiado, superando a produção de óleo de soja que contabilizou 56.85 milhões de toneladas (Benezoli et al., 2021), além de render industrialmente de 3,5 a 10 vezes mais óleo vegetal do que a soja (Tan et al., 2009).

A tendência mundial do consumo de óleo de palma é expandir-se (Tapia; Doliente; Samsatli, 2021), devido à crescente demanda na produção de alimentos, cosméticos, químicos e biodiesel (European Sustainable Palm Oil, 2019; Khatun et al., 2017; Pirker et al., 2016). Nos países tropicais úmido, entre os anos de 2000 a 2019 a palma de óleo foi um dos cultivos com maior expansão agrícola, tendo um crescimento de 176,9% de área colhida, seguida da soja e cana de açúcar com 122.1% e 55.5%, respectivamente (FAO, 2021).

Com origem na África Ocidental (Corley; Tinker, 2015), a palma de óleo é produzida principalmente no Sudeste Asiático, compreendida principalmente pela Indonésia, Malásia e Tailândia (FAO, 2021). Em 2019, esses três países somaram 87,9% da produção mundial de cachos de frutas frescas (FFB) de palma de óleo (FFB), com 59,7% de FFB concentrados na Indonésia, 24.1% na Malásia e 4.1% na Tailândia (FAO, 2021).

O Brasil participa com 0,6% da produção de FFB mundial, ocupando a nona posição. Na América em geral, esses valores são mais relevantes, contribuindo com uma produção de 2,5 milhões de toneladas e respondendo por 11,2% da produção, atrás somente da Colômbia (36,4%) e Guatemala (14,2%) (FAO, 2021).

Impulsionado por benefícios socioeconômicos, a palma de óleo veio crescendo no Brasil apoiada pelo Projeto de Lei Nº 7.326, de 2010 que culminou com a criação do 'Programa de Produção Sustentável da Palma de Óleo'. A

expansão do cultivo no país ganhou saltos relevantes, concentrado no Pará seu principal polo de produção com 164.170 mil hectares colhidos, representando 98,5% da produção nacional, sendo que entre os anos de 2010 a 2019, houve um crescimento de 140,3% de produção de FFB e 214,2% de área plantada na região (IBGE, 2021).

O potencial da palma de óleo na produção de óleo e conversão em biocombustível no Pará vêm ganhando importância em áreas com alto grau de degradação, inserindo-as ao sistema produtivo, minimizando a pressão do desmatamento, capturando carbono e mitigando emissões de gases de efeito estufa (GEE), além de promover oferta de emprego ao setor agrícola e industrial na Amazônia. Logo, o aumento da produtividade combinado com a expansão direcionada para áreas degradadas (Rhebergen et al., 2016) foram as estratégias preferidas para atender à crescente demanda pelo cultivo (Woittiez et al., 2017).

Conseguir prever como as condições ambientais afetam os processos fisiológicos da planta e, consequentemente a eficiência de sua produção, é importante para determinar a sustentabilidade do cultivo no Brasil (Liakos et al., 2018). Investigações de Brum et al. (2021), destacaram a sensibilidade da palma de óleo às condições ambientais, principalmente precipitação pluvial, radiação solar e déficit de pressão de vapor, as quais apontaram variações expressivas na produção com base nessas variáveis climáticas.

Hoffmann et al. (2017), destacaram a importância do ambiente e fatores internos no ciclo reprodutivo da palma de óleo, comprometendo sua produção. Estudos de Caliman e Southworth (1998) apontaram que a variável mais determinante para produção da cultura foi o déficit hídrico, pois o estresse hídrico desencadeou distúrbios vegetativos, como acúmulo de folhas fechadas, senescência prematura de folhas mais velhas e abortamento de frutos nos cachos.

Portanto, as quebras de produtividade do cultivo devem ser estudadas e analisadas com base em variáveis agrometeorológicas em áreas de interesse produtivo. No entanto, essas relações envolvem sistemas de alta complexidade, que não se limitam apenas a modelos lineares, mas sobretudo, o uso de métodos

mais avançados de modelagem para apoiar os produtores na tomada de decisões de forma mais rápida e eficiente.

Neste sentido, os sistemas de gestão agrícola estão evoluindo para soluções de modelagem baseadas em machine learning (ML). Os algoritmos de ML usam uma abordagem de modelagem empírica para aprender padrões e relações úteis, a partir de dados de entrada (Willcock et al., 2018), fornecendo importantes caminhos para previsões de produtividade. Chlingaryan et al. (2018), ressaltam a vantagem do algoritmo de ML na modelagem nas relações não lineares entre várias fontes de dados. Goodfellow et al. (2016), apontaram o ganho de desempenho nos modelos de ML, quando se aumentam os dados de treinamento disponível, assim como os algoritmos podem ser mais robustos para dados com ruídos, diminuindo a variância e o erro das previsões.

Liakos et al., (2018) realizaram uma revisão sobre as aplicações do ML na agricultura, incluindo aplicações em previsão de produtividade, detecção de doenças, detecção de ervas daninhas, qualidade dos cultivos, reconhecimento de espécies, gestão animal, incluindo aplicações em bem-estar animal, gestão da água e manejo do solo, reforçando como a agricultura se beneficiará das tecnologias de ML.

Diferentes modelos de ML têm sido usados para a previsão de safra dos cultivos, entre esses se destaca as Redes Neurais Artificiais (MLP), Support Vector Machine (SVM), Random Forest (RF) e Regressão linear múltipla (RLM). Haghverdi, Allen e Leib (2018), aplicaram o algoritmo de MLP para estimar a produtividade da fibra de algodão nos Estados Unidos. Pant et al., (2021), avaliaram a previsão de safra usando RF e SVM para os cultivos do milho, batata, arroz e trigo na Índia.

Luciano et al. (2021), utilizaram o modelo RF para prever a produtividade da cana-de-açúcar no Brasil com até nove meses antes da colheita. Moraes et al. (2020), avaliaram a previsão de produtividade mensal do Açaí usando modelos de RLM no Brasil com bom desempenho até 6 meses antes da colheita. No entanto, observa-se que há reduzida disponibilidade de publicações que aplicaram os modelos de ML na previsão de Palma de óleo.

Portanto, objetivou-se avaliar o desempenho de diferentes modelos de ML para a previsão de produtividade da palma de óleo usando variáveis climáticas e fenológicas do cultivo.

2.2 MATERIAL E MÉTODOS

2.2.1 Base de dados

O estudo foi realizado em uma área comercial de 583 hectares de plantio de palma de óleo, localizada no município de Moju, (Figura 1). As amostras utilizadas foram das variedades Tenera (*Elaeis guineenses* Jacq. var. tenera) coletadas em vários anos de colheita. Foram utilizados dados mensais de produtividade de cachos de frutas frescas (FFB) entre os anos de 2003 a 2018. Foram excluídos os anos entre 2008 a 2012, devido alta incidência da doença ‘Amarelecimento Fatal’ na área de estudo.

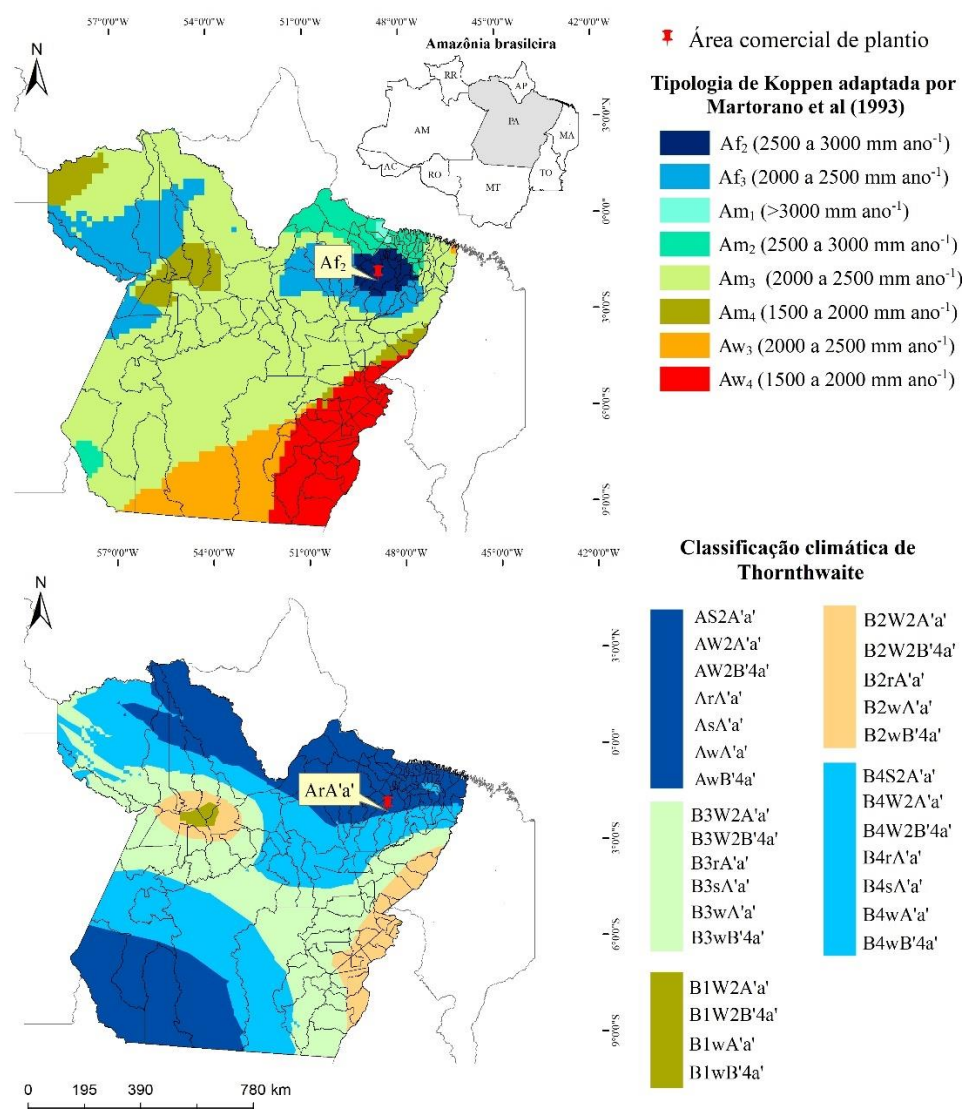


Figura 1. Mapa de localização da área comercial de cultivo de palma de óleo e classificação climática de Koppen adaptado por Martorano et al. (1993) e Thornthwaite (1948) para o estado do Pará.

O clima da região segundo a classificação de Köppen é definido com Af₂ (Martorano et al., 1993), apresentando o mês mais seco precipitação acima de 60 mm e chuvas anuais entre 2.500 a 3.000 mm ano⁻¹. Pela classificação de Thornthwaite (1948) o clima enquadra-se no tipo ArA'a' que é caracterizado como muito úmido, com pequena ou nenhuma deficiência hídrica, evapotranspiração potencial superior a 1.140 mm ano⁻¹ e evapotranspiração de

verão menor que 48% da evapotranspiração anual (Moraes et al., 2020; Torres; Machado, 2008) (Figura 1).

O conjunto de dados climáticos e de cultivo da palma de óleo, assim como as análises exploratórias para aplicação dos modelos de machine learning na previsão de produtividade da cultura aplicados neste estudo, é representado no fluxo grama de trabalho da figura 2.

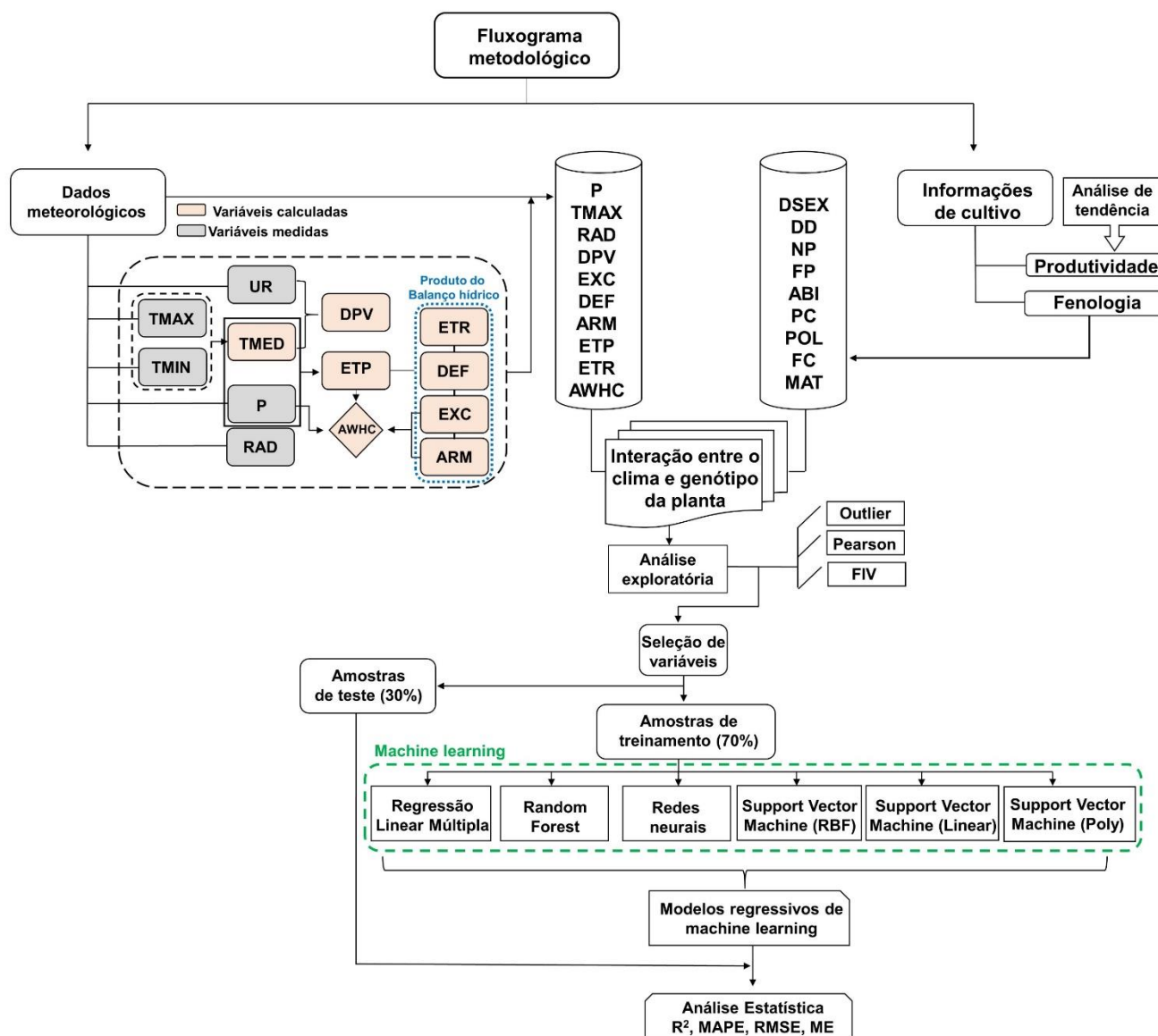


Figura 2. Etapas metodológicas conduzidos para este estudo. Variáveis meteorológicas: Precipitação pluvial (P mm), Deficiência hídrica (DEF mm), Excedente hídrico (EXC mm), Armazenamento de água no solo (ARM mm), Evapotranspiração potencial (ETP mm), Evapotranspiração real (ETR mm), Temperatura média (TMED °C), Temperatura mínima (TMIN °C), Temperatura

máxima (TMAX °C), Percentage Available Water holding Capacity (AWHC %), h) Radiação solar (RAD MJ m⁻² dia⁻¹), déficit de pressão de vapor (DPV kPa), umidade relativa (UR %). Variáveis de cultivo: Diferenciação do sexo (DSEX), Desenvolvimento (DD), Número de espiga (NP), Flor por espiga (FP), Aborto de inflorescência (ABI), Peso de carcaça (PC), Polinização (POL), Falha de cacho (FC), Maturação (MAT).

A base de dados agrometeorológicas utilizada corresponde ao período de 1993 a 2018. Foram analisados dados diários de precipitação pluvial (P em mm), com coletas pluviais localizadas na área de cultivo de palma de óleo, temperatura máxima (Tmax em °C), radiação solar (RAD em MJ m⁻² dia⁻¹) e umidade relativa (UR em %) de estações meteorológicas próximas a região de estudo, especificamente no raio de 36 km da área de coleta de dados da cultura. Com base nos dados de UR, foi calculado o déficit de pressão de vapor (DPV) por meio das seguintes equações (1 a 3):

$$DPV = Es - Ea \quad (1)$$

$$Es = 0.6108 \times 10^{\left(\frac{7.5 \times TMED}{237.3 + TMED}\right)} \quad (2)$$

$$Ea = \frac{UR \times Es}{100} \quad (3)$$

em que, DPV= Déficit de pressão de vapor (kPa); Es = Pressão de saturação de vapor (kPa); Ea = Pressão parcial de vapor (kPa); UR= Umidade relativa (%); TMED= Temperatura média do ar (°C).

Foi realizada a retirada de tendência temporal das séries por ajuste linear, conforme proposto por Gujarati e Porter (2011), para a visualização de outros componentes capazes de influenciar a produção em função de efeitos do clima na produção de FFB por hectare.

2.2.2 Componentes de água no solo

Para determinação do balanço hídrico (BH), seguindo a metodologia de Thornthwaite e Mather (1955), foi utilizado uma capacidade de água disponível no solo (CAD) igual a 125 mm, sendo mais indicada para fins de caracterização dos cultivos de palma de óleo (Gonçalves et al., 2010). A evapotranspiração foi calculada seguindo o método de Camargo (1971), equação 4, utilizando dados de temperatura média do ar para estimar a evapotranspiração potencial (ETP).

$$ETP = F Q_o T ND \quad (4)$$

sendo Q_o (mm dia^{-1}) a radiação solar extraterrestre diária expressa em equivalente de evaporação, no período considerado, T ($^{\circ}\text{C}$) é a temperatura média do ar durante o período; F é o fator de ajuste que varia com a temperatura média anual do local (para T_m até 23°C , $F = 0,01$; $T_m = 24^{\circ}\text{C}$, $F = 0,0105$; $T_m = 25^{\circ}\text{C}$, $F = 0,011$; $T_m = 26^{\circ}\text{C}$, $F = 0,0115$; $T_m > 26^{\circ}\text{C}$, $F = 0,012$); e ND é o número de dias do período.

Com a evapotranspiração potencial obtida foram estimados os seguintes componentes de BH: armazenamento de água no solo (STO), deficiência hídrica (DEF) e excedente hídrico (EXC) (Figura 3).

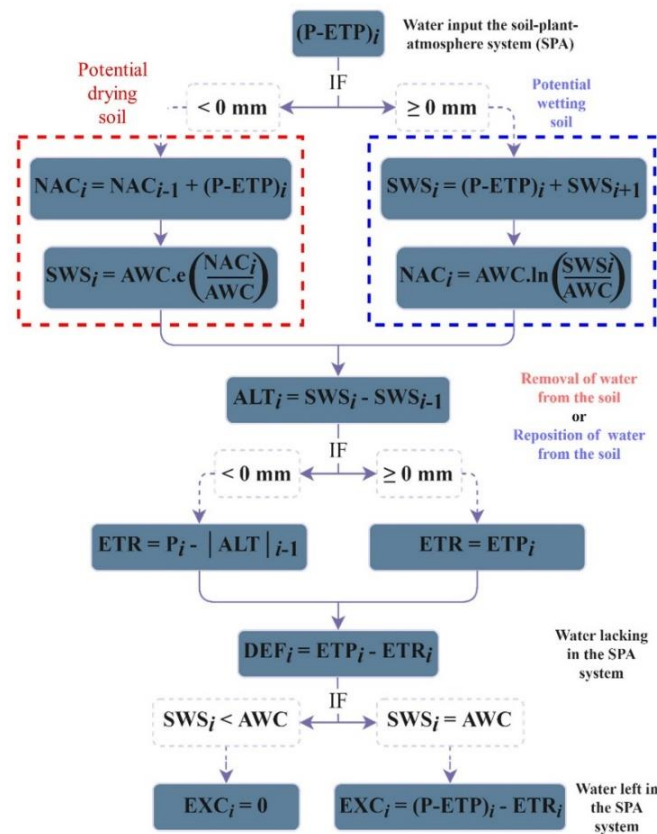


Figura 3. Diagrama do modelo de balanço hídrico de Thornthwaite e Mather (1955): Fonte Rolim et al. (2020).

em que ETP é a evapotranspiração potencial (mm), AWC é a capacidade de água disponível no solo (mm), SWS é o armazenamento de água no solo (mm), NAC é o negativo acumulado, ou seja precipitação acumulada menos a evapotranspiração potencial, P é a precipitação (mm), DEF é o déficit de água no sistema solo-planta-atmosfera (mm), ETR é a evapotranspiração real (mm), EXC é o excedente de água do sistema solo-planta-atmosfera (mm), ALT é o armazenamento de água do solo para o mês corrente menos o armazenamento de água do solo para o mês anterior (mm), e i é o período mensal (Adaptado de Rolim et al. 2020).

Com base nos componentes de BH, foi calculado a porcentagem de capacidade de armazenamento de água no solo disponível no final do mês (AWHC %) (Surre, 1968), utilizando-se a seguinte equação (5):

$$\% \text{AWHC} = ((\text{SWS} + \text{P} - \text{ETP} - \text{EXC}) / \text{AWC}) \times 100\% \quad (5)$$

sendo %AWHC é a porcentagem da capacidade de retenção de água disponível no final do mês, SWS é o armazenamento de água no solo (mm), e AWC é a capacidade de água disponível (mm).

2.2.3 Caracterização fenológicas da palma de óleo

Foram definidos 9 estádios fenológicos para palma de óleo (Hormaza; Fuquen EM; Romero HM, 2012; Forero DC, Hormaza P, Romero HM, 2012), considerando-se principalmente a linha do tempo de desenvolvimento da inflorescência até a maturação, com base na seguinte caracterização: Diferenciação do sexo (DSEX), estágio esse que leva a divisão de inflorescência masculina e feminina, ocorrendo entre 23 a 25 meses antes da colheita. Desenvolvimento (DD), com antecipação de (19 a 22 meses), Número de espiga (NP, 16 a 18 meses), Flor por espiga (FP, 12 a 15 meses), Aborto de inflorescência (ABI, 10 a 11 meses), Peso de carcaça (PC, 7 a 9 meses), Polinização/antese (POL, 5 a 6 meses), Falha de cacho (FC, 2 a 5 meses) e Maturação (MAT) com ocorrência 1 mês antes da colheita de FFB (Figura 4).

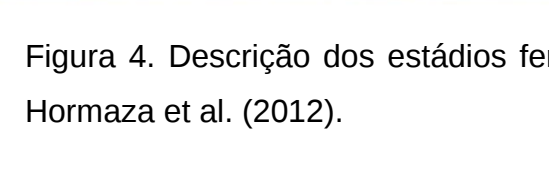
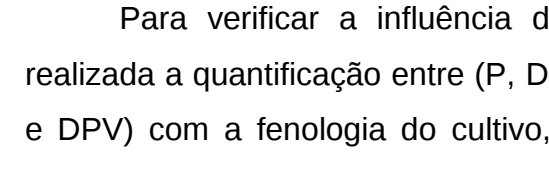
Diferenciação interna da inflorescência e desenvolvimento				Meses antes da colheita	Estádio fenológico	Índice
				-25	Diferenciação do sexo (Masculino/Feminino)	DSEX
				-24		
				-23		
Desenvolvimento de inflorescência masculina				-22	Desenvolvimento	DD
				-21		
				-20		
				-19		
Desenvolvimento de inflorescência feminina				-18	Número de espiga	NP
				-17		
				-16		
				-15	Flor por espiga	FP
Desenvolvimento do fruto				-14		
				-13	Aborto de inflorescência	ABI
				-12		
				-11	Peso de carcaça	PC
				-10		
				-9		
				-8	Polinização	POL
				-7		
				-6		
				-5	Falha de cacho	FC
				-4		
				-3		
				-2		
				-1	Maturação	MAT

Figura 4. Descrição dos estádios fenológicos da palma de óleo. Adaptado de Hormaza et al. (2012).

Para verificar a influência do clima em cada estágio fenológico foi realizada a quantificação entre (P, DEF, EXC, ARM, ETR, AWHC, TMAX, RAD e DPV) com a fenologia do cultivo, levando a um conjunto de dados de 81 variáveis para seleção entre os modelos de previsão. No entanto, antes da implementação do modelo final, avaliações prévias foram realizadas no banco de dados, buscando tanto a seleção quanto na extração de variáveis.

Foi analisada a multicolinearidade dos dados, ou seja, condição em que duas variáveis são correlacionadas e, portanto, apresentam informações semelhantes sobre a variação em um determinado conjunto de informações. Para isso, o Fator de Inflação de Variância (FIV), usando a biblioteca StatsModels da linguagem de programação Python foi utilizada para diagnosticar a multicolinearidade entre as variáveis. Esse método é baseado no valor R^2 obtido através da regressão de um preditor sobre todos os outros preditores na análise (Forthofer; Lee; Hernandez, 2007).

Em regra geral, se a FIV for maior que 10, então a multicolinearidade é alta, sendo assim, exclui-se a variável explicativa com FIV mais alta e reformase o modelo (Miles, 2014). Uma das técnicas também utilizada foi análise da matriz de correlação de Pearson, para a avaliação do grau de associação entre as variáveis de estudo.

Por fim, objetivando resumir a maior parte da variância dos dados a um número mínimo de fatores para fins de previsão, foi aplicado análise de componente principal (Farrugia et al., 2021, Hotelling, 1933), por meio da biblioteca `sklearn.decomposition.PCA`.

2.2.4 Modelos de Machine Learning

Foram utilizados diferentes modelos de machine learning para prever a produtividade da palma de óleo. Os dados de produtividade por hectare corresponderam a variável dependente e os elementos meteorológicos as variáveis independentes. Para todos os modelos, foram separados 70% dos dados para treinamento e 30% para teste. Os modelos machine learning ajustados foram: Regressão Linear Múltipla (RLM), Random Forests (RF), Redes neurais artificiais – Multilayer Perceptron (MLP) e Support Vector Machines de base radial RBF (SVM_RBF), Linear (SVM_Linear) e Polinomial (SVM_Poly), usando a linguagem de programação Python.

A análise de regressão linear múltipla (RLM) modela a relação entre uma variável dependente (produtividade da colheita) e variáveis independentes (climáticas) ajustando uma equação linear através de dados observados. Uma primeira etapa na regressão múltipla é examinar as relações de pares entre todas as variáveis, uma vez que isso é útil para entender os dados. Este método também evita o mau condicionamento da matriz das variáveis das independentes (Features), controlando a inflação e a instabilidade geral encontrados em estimadores de mínimos quadrados (Liu et al., 2021).

O Random Forest (RF) são variantes de árvores de decisão que são procedimentos para classificação e/ou regressão que utilizam grande quantidade de dados (Adetiloye, Awasthi, 2017). A grande vantagem do RF é dada a sua adequação para lidar com ruído nos dados, sendo, portanto, muito robusto. Esse

algoritmo é amplamente usado no aprendizado de máquina, devido ao seu bom equilíbrio entre robustez e alta qualidade do modelo (Hariharan, 2021).

As redes neurais são um conjunto de unidades individuais interconectadas chamadas neurônios (Du e Swamy, 2014). O neurônio individual tem um número finito de entradas e uma saída. Cada entrada está associada a um valor de ponderação. Multilayer Perceptron (MLP) é um tipo de rede neural artificial que foi proposta por Frank Rosentblatt em 1957 (Shi et al., 2020), sendo mais comumente usadas devido à sua operação rápida, facilidade de implementação e requisitos de dados para o conjunto de treinamento menores (Kocuyigit et al., 2008).

O MLP consiste em camadas sequenciais: camadas de entrada, camadas ocultas e de saída (Alkhasawneh e Tay, 2017). A camada oculta processa e transmite as informações de entrada para a camada de saída. Um modelo MLP com número insuficiente ou excessivo de neurônios na camada oculta provavelmente causa problemas de má generalização e superajuste (Antanasijević; Antanasijević; Pocajt, 2018). Não existe um método analítico para determinar o número de neurônios na camada oculta. Portanto, ele só é encontrado por tentativa e erro (Wang; Hafshejani; Wang, 2021).

O Support Vector Machine (SVM) é um algoritmo de ML baseado na teoria de aprendizado estatístico. O SVM mostrou muitas vantagens, devido ao seu bom desempenho com um número limitado de amostras (Shaharum et al., 2018). Outras aplicações também são no reconhecimento de padrões não lineares e de alta dimensão, podendo ser aplicado a outros problemas de ML, como ajustes de funções (Cortes e Vapnik, 1995).

Os algoritmos de SVM são selecionados por parâmetros da função Kernel, dividindo-se em função de Base Radial (RBF), Linear e polinomial. No entanto, atualmente não existe um critério padrão para seleção da função kernel adequada para um problema específico no algoritmo SVM. Sendo assim, o método atual para seleção do melhor kernel é aplicar diferentes funções e selecionar a de menor erro (Manjunathan et al., 2020).

Os parâmetros de treinamento para cada modelo e seus respectivos valores padrão, foram estimados pelo método gridsearchCV uma biblioteca para

python. Esse método funciona da seguinte forma: a partir de valores pré-estabelecidos de cada parâmetro, o método faz uma combinação de todas as possibilidades formando um 'grid'. Será selecionado como resposta a combinação que apresentar menor erro em relação aos dados observados. Os valores dos hiperparâmetros e cada modelo selecionado são apresentados na tabela 1.

Tabela 1. Valores dos hiperparâmetros utilizados pelos modelos de Machine Learning entre as diferentes épocas de previsão. 4, 6 e 9 meses antes da colheita.

Modelo	Hiperparâmetros	4 meses	6 meses	9 meses
RF	max_depth	10	8	10
	min_impurity_decreas	0	0	0
	n_estimators	80	40	40
	random_state	0	0	0
MLP	hidden_layer_sizes	9,14,18	11,15,17	2,3,5
	learning_rate_init	0.1	0.1	0.1
	learning_rate	constant	constant	constant
	activation	logistic	logistic	logistic
	solver	lbfgs	lbfgs	lbfgs
	alpha	0.1	0.1	0.1
	random_state	50	50	50
	max_iter	1000	1000	1000
SVM_RBF	kernel	rbf	rbf	rbf
	C	3	8	8
	gamma	auto	auto	auto
	epsilon	1	1	1
SVM_LINEAR	kernel	linear	linear	linear
	C	4	2	2
	gamma	auto	auto	auto
	epsilon	1	2	3
SVM_POLY	kernel	poly	poly	poly
	C	1	2	2
	gamma	auto	auto	auto
	epsilon	0.1	0.1	0.1
	coef0	2	2	2

2.2.5 Análise de dados

Os dados observados em campo e obtidos pelos modelos de machine learning foram comparados através de indicadores estatísticos: precisão, acurácia e tendência (Tabela 2). A precisão que indica o grau de dispersão entre os valores estimados e observados, foi estimada pelo coeficiente de determinação (R^2) (Equação 12). A acurácia determina o afastamento entre valores estimados e observados, sendo medido pela Raiz do erro quadrático médio (RMSE) e do Erro absoluto médio (MAPE) (Equação 13, 14), respectivamente. O RMSE reflete o quão distantes os valores previstos em média estão dos valores observados e o MAPE qual a porcentagem de erro entre o valor previsto e observado. O erro médio (ME) (Equação 15), representa o erro sistemático do modelo, ou seja, quanto é o erro de subestimação ou superestimação em relação ao valor observado.

Tabela 2. Métricas de precisão, acurácia e tendência dos índices estatísticos utilizados.

Índice estatístico	Equação
Precisão	
R^2	$R^2 = \frac{\sum_{i=1}^n (Y_{est} - \overline{Y_{est}})^2}{\sum_{i=1}^n (Y_{obs} - \overline{Y_{est}})^2} \quad (6)$
Acurácia	
RMSE	$RMSE = \sqrt{\frac{\sum_{i=1}^n (Y_{obs} - Y_{est})^2}{n}} \quad (7)$
MAPE	$MAPE = \frac{\sum_{i=1}^n \left(\frac{ Y_{obs} - Y_{est} }{Y_{obs}} \right)}{n} * 100 \quad (8)$
Tendência	
ME	$ME = \frac{1}{n} \sum_{i=1}^n (Y_{est} - Y_{obs}) \quad (9)$

Y_{est} é o valor estimado de y ; Y_{obs} é o valor observado de y ; $\overline{Y_{est}}$ é o valor médio estimado de y ; n é número de dados.

2.3 RESULTADOS

Nas áreas de cultivo de palma de óleo foi observado grande variabilidade nos estoques hídricos e médias térmicas de variáveis agrometeorológicas entre janeiro e dezembro (Figura 5). A precipitação pluvial (P) (Figura 5.a) entre janeiro e maio apresentou os maiores valores em oferta de chuva, representando neste período quase 70% dos acumulados de P anual. As chuvas máximas ocorreram entre fevereiro e abril, com 650 mm cada, no entanto, as maiores médias entre os 24 anos analisados foram de, aproximadamente 455 mm para março e abril. No segundo semestre, principalmente entre julho e outubro, ocorreram as menores cotas pluviais, mas mantendo-se no mês menos chuvoso valores acima de 60 mm, confirmando-se a inclusão da área na tipologia climática Af₂, de acordo com Martorano et al. (1993) e Martorano et al. (2017), ao analisar as tipologias climáticas no estado do Pará.

A deficiência hídrica (DEF) (Figura 5.b), devido ao grande acumulado de chuva durante o primeiro semestre, não apresenta limitações evapotranspiratórias para plantas, principalmente entre janeiro e junho. Entre setembro e novembro, observa-se maior DEF, chegando a valores médios acima de 20 mm, mas como a CAD foi de 125 mm, reforça que esse fator não apresentou limitações a cultura.

O excedente hídrico (EXC) (Figura 5.c), de forma inversa ao DEF e direta a P, armazenamento de água no solo (ARM) (Figura 5.d) e capacidade de retenção de água disponível (AWHC) (Figura 5.e) apresentaram no primeiro semestre seus maiores valores acumulados e no segundo semestre os menores volumes. A Evapotranspiração real (ETR), obteve no segundo semestre a menor demanda evapotranspiratória, destacando-se entre setembro e novembro (Figura 5.e).

A temperatura máxima (Tmax) (Figura 5.f) e a radiação solar (Rs) (Figura 5.h), fatores que regulam diretamente a fotossíntese e o acúmulo de biomassa, apresentaram amplitudes de 1,8 °C e 5,3 MJ m² dia⁻¹, respectivamente. Entre agosto a novembro são os meses com os maiores índices dessas variáveis, chegando a médias de Tmax acima de 34,5 °C e médias de Rs acima de 22 MJ m² dia⁻¹.

O déficit de pressão de vapor (DPV) variou entre 0,7 a 1,7 kPa. De forma geral no segundo semestre, principalmente entre agosto e novembro ocorre maior demanda evaporativa da atmosfera, com valores acima de 1,5 kPa e médias acima de 1,2 kPa. Entre janeiro a abril são observados os menores valores de DPV, com variações médias abaixo de 0.7 kPa (Figura 5.i).

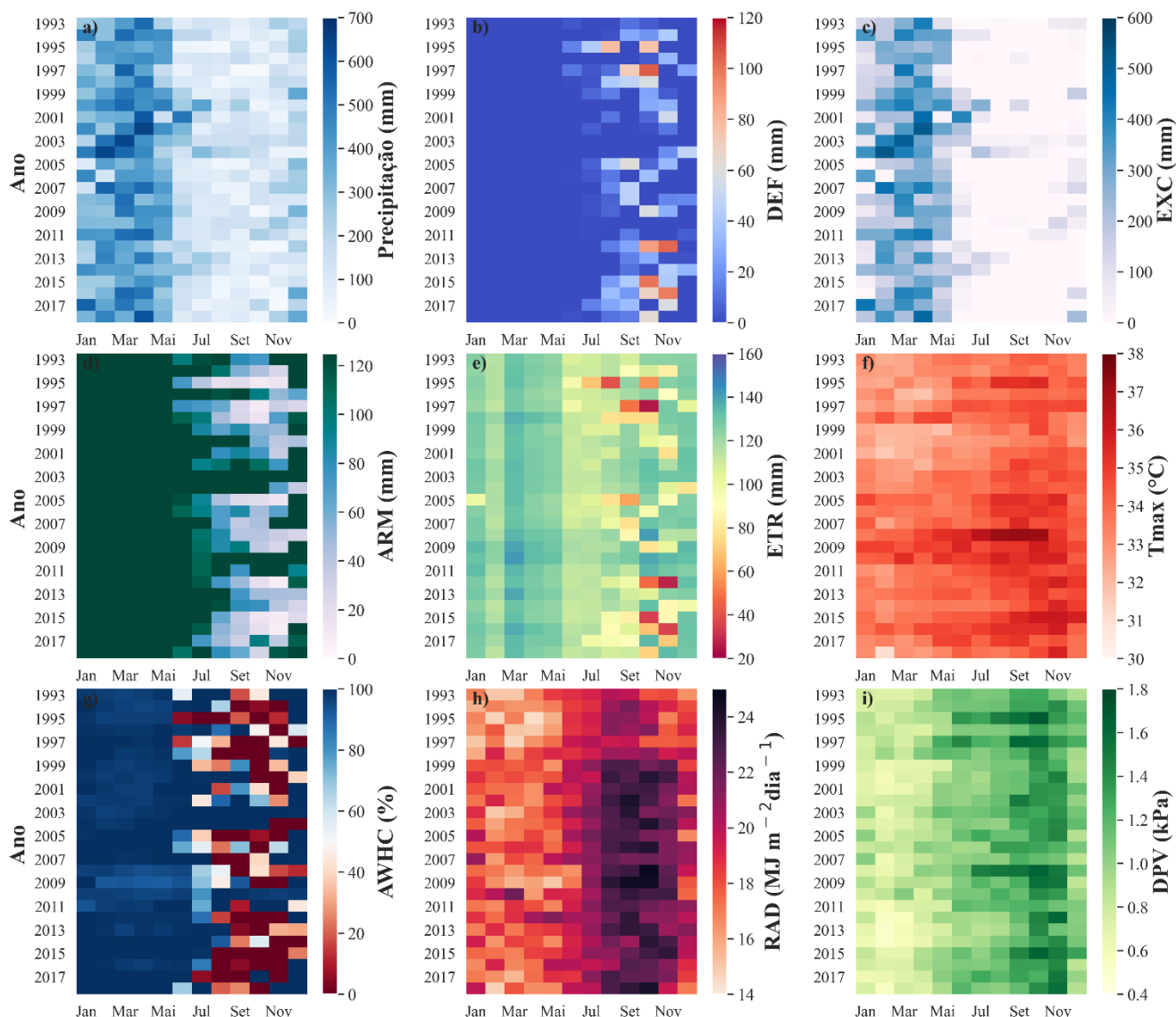


Figura 5. Mapa de calor das variáveis meteorológicas, a) Precipitação pluvial (P mm), b) Deficiência hídrica (DEF mm), c) Excedente hídrico (EXC mm), d) Armazenamento de água no solo (ARM mm), e) Evapotranspiração real (ETR mm), f) Temperatura máxima (TMAX °C), g) Percentage Available Water Holding Capacity (AWHC %), h) Radiação solar (RAD MJ m⁻² dia⁻¹) e déficit de pressão

de vapor (DPV kPa), em regiões de cultivo de palma de óleo, no município de Moju-Pará.

Todas as fases fenológicas da palma de óleo sofrem grande influência do clima (Figura 6). As variáveis hídricas para a fase DSEX (23 a 25 meses antes da colheita de FFB) como P, EXC, ARM e AWHC apresentam em média maiores valores entre os meses de colheita de fevereiro a abril. Portanto, durante esses meses as condições meteorológicas que acontecem 23 a 25 meses antes da colheita é condicionada por maiores ofertas hídricas. Por outro lado, os meses de colheita de setembro a novembro, apresentam os menores acumulados hídricos (Figura 6.a).

Quando avaliada as variáveis energéticas como radiação solar (RAD) e temperatura máxima (TMAX), os maiores valores médios obtidos em DSEX estão concentrados entre setembro e novembro e os menores valores entre fevereiro e março. Vale destacar que a amplitude obtida entre os meses de colheita para TMAX foi de 1,54 °C e RAD de 4,90 MJ m⁻² dia⁻¹, apontando as pequenas variações que ocorrem durante os meses de colheita durante o DSEX (Figura 6.g).

As maiores produtividades de FFB estão concentradas entre outubro e novembro (Figura 6.a), e as menores entre março a maio. Esses dados corroboram com os estudos de Benezoli et al. (2021) em investigações com produtividade de palma de óleo realizadas no nordeste do Pará. É importante observar que durante os meses de maior produção, as fases fenológicas de Desenvolvimento diferencial (DD) que se inicia 19 a 22 mês antes da colheita de FFB, Numero de espiga (NP) 16 a 18 meses antes da colheita, Peso de Cacho de carcaça (PC) 7 a 9 meses antes da colheita e polinização (POL) 5 a 6 meses antes da colheita, apresentam acumulados significantes de EXC e P, ou seja, para esses meses de maiores produtividades de FFB, as fases fenológicas descritas acima acontecem em média em períodos com maior oferta hídrica.

Durante março e maio, períodos de menor produtividade de FFB, as fases de Falha de Cacho (FC) 2 a 4 meses antes da colheita, Polinização (POL) 5 a 6 meses antes da colheita e Número de Espiga (NP) 16 a 18 meses antes da

colheita, acontecem em média em períodos mais secos, representando maiores déficits de água no solo e consequentemente menor capacidade evapotranspiratória da planta.

O déficit de pressão de vapor (DPV) (Figura 6.f), variável de grande importância dada as suas contribuições na troca de carbono entre a superfície e atmosfera, apresentou nos meses de menor produtividade (março a maio) nas fases, POL 5 a 6 meses antes da colheita, PC 7 a 9 antes da colheita e NP 16 a 18 meses antes da colheita, valores elevados de DPV. Por outro lado, em meses de maior produtividade de FFB, foi observado menores variações de DPV nas fases DD, ABI e PC, sendo esses 19 a 22, 10 a 11 e 7 a 9 meses antes da colheita, respectivamente.

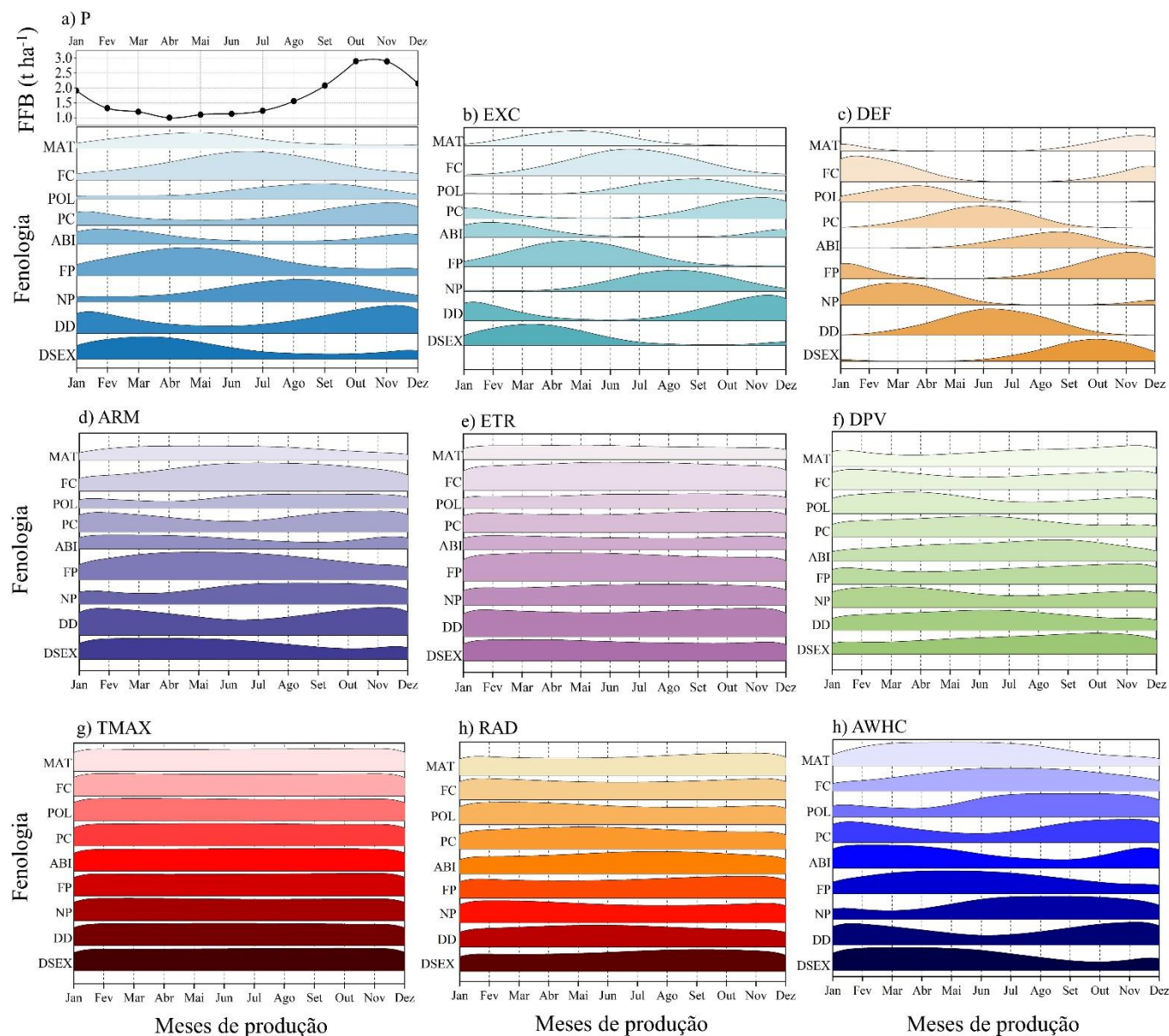


Figura 6. Acumulado das condições meteorológicas para diferentes meses de produção e fenológicos no cultivo de palma de óleo no município de Moju- Pará. a) Precipitação pluvial e Produtividade (FFB, t ha⁻¹), b) Excedente hídrico (EXC), c) Deficiência hídrica (DEF), d) Armazenamento de água no solo (ARM), e) Evapotranspiração real (ETR), f) déficit de pressão de vapor (DPV), g) Temperatura máxima (TMAX), h) Radiação solar (RAD), e) AWHC.

Do conjunto de dados obtidos pela interação entre as condições agrometeorológicas e a fenologia da palma de óleo foram selecionadas 9 variáveis de acordo com o critério estabelecido de correlação de Pearson, Fator

de Inflação de Variância, objetivando remover preditores que estão altamente correlacionados, e análise de fatores para redução de dimensionamento (Figura 7). Os valores correspondem a acumulados de EXC, e médias de TMAX, DPV, AWHC e RAD, durante os estádios fenológicos da planta.

O EXC durante a fase de DD (EXC_DD), 19 a 22 meses antes da colheita de FFB, apresentou entre os meses de novembro, dezembro e janeiro suas maiores médias, com 1078, 1086 e 934 mm, respectivamente. Enquanto maio a junho representaram as menores médias, com valores ≤ 70 mm (Figura 7.a). Na fase de polinização, os maiores EXC se concentram durante agosto a outubro, com acumulados ≥ 500 mm, e entre janeiro a maio se observou em média valores ≤ 50 mm (Figura 7.b).

Para TMAX durante a POL (TMAX_POL), maiores valores foram observados entre fevereiro a maio, com médias $\geq 34,5$ °C. Julho a setembro apresentaram as menores médias, com valores $\leq 34,0$ °C (Figura 7.c). O DPV durante a fase de PC (DPV_PC) (Figura 7.d), variou entre o período de maio a junho com valores $\geq 1,2$ kPa e entre outubro a novembro com valores $\leq 0,85$ kPa, sendo esses os períodos de maiores e menores valores, respectivamente. Durante a FP, o DPV, diferente da fase PC, obteve seus picos entre outubro a dezembro ($\geq 1,2$ kPa) e menores entre março a maio ($\leq 0,82$ kPa) (Figura 7.e).

O AWHC prevaleceu entre as variáveis hídricas e térmicas, sendo um importante elemento na variabilidade de três fases fenológicas para previsão de produtividade, sendo, NP, FP e ABI (Figura 7.f, h), respectivamente. AWHC durante os períodos de julho a outubro em NP, abril a junho em FP e de dezembro a abril em ABI não apresentam limitações hídricas, com valores médios ($\geq 95\%$). Por outro lado, fevereiro a abril, outubro a dezembro e julho a outubro, são os menores índices de AWHC ($\leq 50\%$), para NP, FP e ABI, respectivamente.

No conjunto de variáveis, a radiação solar contribui nos modelos de previsão durante a fase de FP (Figura 7.i). Esse estágio para os períodos de colheita de outubro e novembro apresentaram as maiores médias de radiação, com valores ≥ 22 MJ m⁻² dia⁻¹. Por outro lado, entre março a maio a radiação

variou entre 17.5 a 18 $\text{m}^{-2} \text{dia}^{-1}$, correspondendo, portanto, a períodos com menores valores.

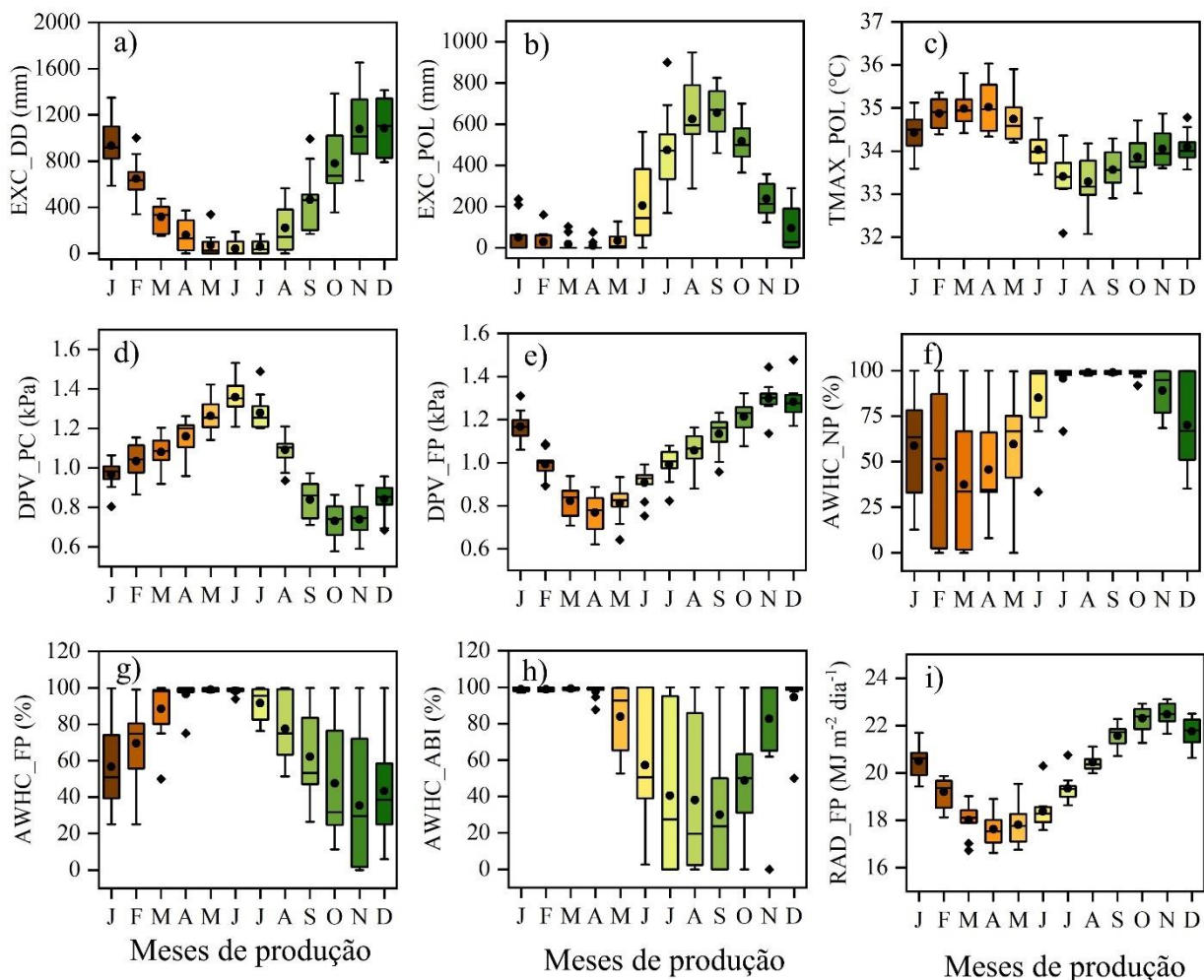


Figura 7. Boxplot da variabilidade mensal dos totais e médias climáticas para as principais fases fenológicas da palma de óleo, selecionadas pelo modelo de machine learning. EXC_DD (Excedente no desenvolvimento), EXC_POL (Excedente na polinização), TMAX_POL (Temperatura máxima na polinização), DPV_PC (Déficit de pressão de vapor no peso de cacho), DPV_FP (Déficit de pressão de vapor na flor por espiga), AWHC_NP (Porcentagem de armazenamento de água no número de espiga), AWHC_FP (Porcentagem de armazenamento de água em flor por espiga), AWHC_ABI (Porcentagem de armazenamento de água no aborto de inflorescência), RAD_FP (Radiação solar na flor por espiga).

Pela correlação de Pearson (Figura 8), as variáveis que respondem diretamente a produtividade, ou seja, quanto o valor de uma variável aumenta, maior será a produtividade, são (DPV_FP = 0.73), (RAD_FP = 0.65), (EXC_DD = 0.53), (AWHC_NP = 0.39), (EXC_POL = 0.33) e (AWHC_ABI = 0.1). De forma inversa, as variáveis que respondem de forma indireta a produtividade foram (DPV_PC = -0.67), (AWHC_FP = -0.58) e (TMAX_POL = -0.33).

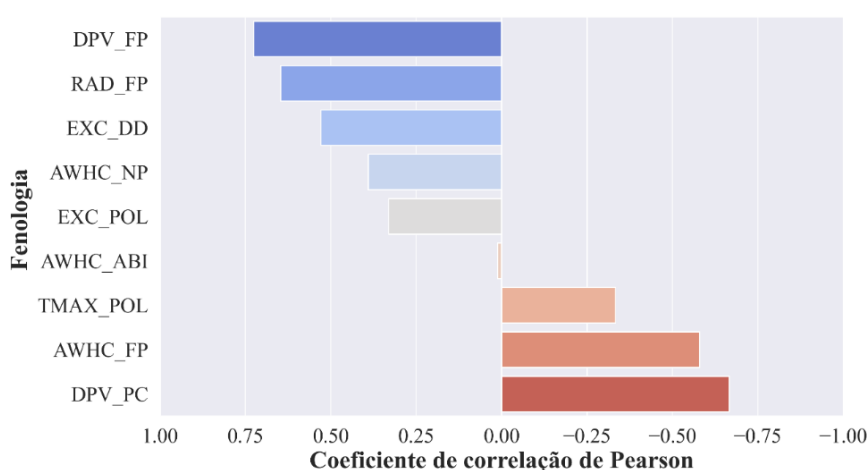


Figura 8. Coeficiente de correlação de Pearson das principais fases fenológicas e variáveis meteorológicas selecionadas pelo modelo de machine learning. DPV_FP (Déficit de pressão de vapor na flor por espiga), RAD_FP (Radiação solar na flor por espiga), EXC_DD (Excedente no desenvolvimento), AWHC_NP (Porcentagem de armazenamento de água no número de espiga), EXC_POL (Excedente na polinização), AWHC_ABI (Porcentagem de armazenamento de água no aborto de inflorescência), TMAX_POL (Temperatura máxima na polinização), AWHC_FP (Porcentagem de armazenamento de água em flor por espiga), DPV_PC (Déficit de pressão de vapor no peso de cacho).

O desempenho de todos os modelos de Machine Learning para previsão de produtividade de FFB com 4 meses de antecedência foram satisfatórios, apresentando $R^2 > 0,70$, $RMSE < 0,50 \text{ t ha}^{-1}$, $MAPE < 26\%$ e $ME (\leq -0,02 \text{ e } \geq -0,15 \text{ t ha}^{-1})$ (Figura 9). O modelo Support Vector Machine de kernel polinomial (SVM_Poly) (Figura 9.f), apresentou os melhores ajustes, com medidas de precisão ($R^2 = 0,86$), acurácia ($RMSE = 0,39$ e $MAPE = 16\%$) e tendência ($ME =$

-0,06). Em contrapartida, o modelo Support Vector Machine de kernel linear (SVM_Linear) apresentou os menores desempenhos, com ($R^2 = 0,71$, RMSE = $0,48 \text{ t h}^{-1}$, MAPE = 24,64% e ME = $-0,12 \text{ t h}^{-1}$). (Figura 9.e).

Embora tenha sido observado desempenho satisfatório na previsão de FFB, os modelos apresentaram limitações para prever os picos e vales de produtividade (Figura 9.a.1 a f.1). No entanto, o comportamento crescente e decrescente da produtividade dos dados observados ao longo do tempo é acompanhado pelo modelo estimado, tendendo os modelos em média a subestimação.

No conjunto de dados, observa-se também que o erro entre os valores observados e estimados, tende a ser maior à medida que se tem menores valores de produtividade, principalmente em faixas menores que 1 t ha^{-1} . Para produtividades mais altas, acima de 2 t ha^{-1} , os erros tendem a ser mais conservador. Importante destacar que à medida que a produtividade aumenta, os dados estimados tendem a subestimar os dados observados em campo, por outro lado, produtividades mais baixas os modelos tendem a superestimar os valores reais (Figura 9).

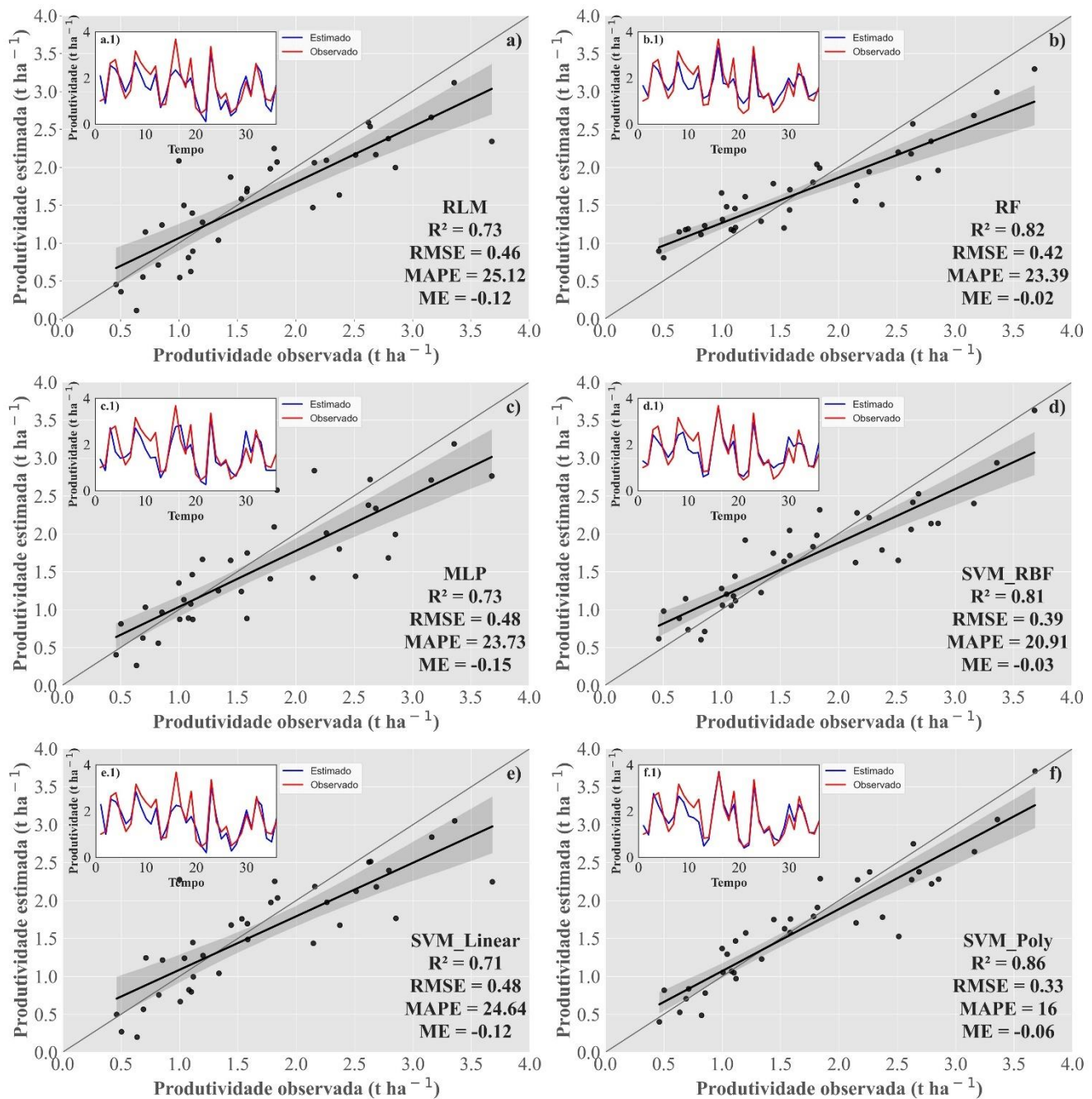


Figura 9. Comparação da produtividade observada obtida em campo e da produtividade prevista com 4 meses antes da colheita, simulada pelos diferentes modelos de machine learning. a) Regressão Linear Múltipla (RLM), b) Random Forest (RF), c) Multilayer Perceptron (MLP), d) Support Vector Machine RBF (SVM_RBF), e) Support Vector Machine Linear (SVM_Linear), f) Support Vector Machine Poly (SVM_Poly).

Para previsão de FFB com 6 meses de antecedência (Figura 10), os valores de precisão, tendência e acurácia apresentaram valores mais baixos. O nível de precisão variou entre ($R^2 \geq 0,70$ e $\leq 0,77$), acurácia ($\text{MAPE} \geq 24\%$ e $\leq 28\%$, $\text{RMSE} \geq 0,45 \text{ t ha}^{-1}$ e $\leq 0,52 \text{ t ha}^{-1}$) e tendência ($\text{ME} \geq -0,21 \text{ t ha}^{-1}$ e $\leq -0,02 \text{ t ha}^{-1}$). O RF (Figura 10.b) e RLM (Figura 10.a) apresentaram os melhores resultados de precisão ($R^2 = 0,77$ e $0,74$), acurácia ($\text{RMSE} = 0,45 \text{ t ha}^{-1}$ e $0,46 \text{ t ha}^{-1}$); ($\text{MAPE} = 24,41\%$ e $25,11\%$) e tendência ($\text{ME} = -0,02 \text{ t ha}^{-1}$ e $-0,13 \text{ t ha}^{-1}$), respectivamente. Desempenhos ligeiramente menores foram observados para os demais modelos previsores, com $R^2 \leq 0,73$, $\text{MAPE} \geq 25\%$ e $\text{RMSE} \geq 0,48 \text{ t ha}^{-1}$ e $\text{ME} \geq 0,14 \text{ t ha}^{-1}$.

Apesar do desempenho satisfatório para os níveis de precisão e acurácia na previsão com 4 meses de antecedência (Figura 9), o modelo SVM_Poly apresentou baixo desempenho para prever a produtividade com 6 meses antes da colheita, apresentando subestimação média de $-0,21 \text{ t ha}^{-1}$ (Figura 10.f).

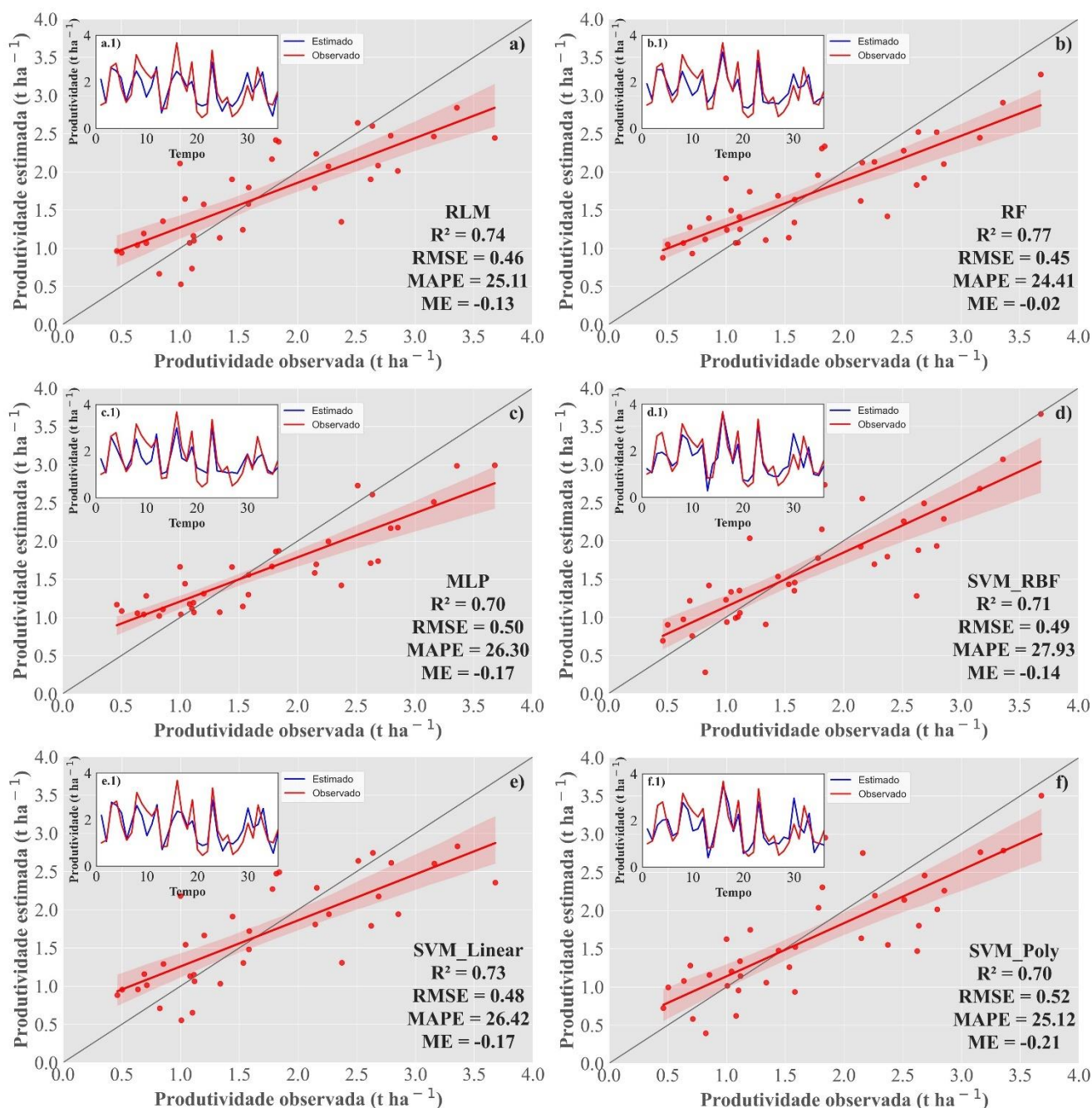


Figura 10. Comparação da produtividade observada obtida em campo e da produtividade prevista com 6 meses antes da colheita, simulada pelos diferentes modelos de machine learning. a) Regressão Linear Múltipla (RLM), b) Random Forest (RF), c) Multilayer Perceptron (MLP), d) Support Vector Machine RBF (SVM_RBF), e) Support Vector Machine Linear (SVM_Linear), f) Support Vector Machine Poly (SVM_Poly).

No conjunto dos modelos de previsão, observou-se menores desempenhos ao expandir-se o tempo das estimativas em previsibilidade. Portanto, para níveis de previsão de produtividade com 9 meses de antecedência (Figura 11), os índices estatísticos de precisão variaram entre ($R^2 \geq 0,62$ e $\leq 0,77$), acurácia ($MAPE \geq 24\%$ e $\leq 32\%$, $RMSE \geq 0,46 \text{ t ha}^{-1}$ e $\leq 0,75 \text{ t ha}^{-1}$) e tendência ($ME \geq -0,09 \text{ t ha}^{-1}$ e $\leq 0,0003 \text{ t ha}^{-1}$).

Modelos de MLP (Figura 11.c) e RF (Figura 11.b), obtiveram desempenho satisfatório, com reduzida diferença entre suas métricas estatísticas, R^2 (0.77 e 0.74), RMSE (0.46 ha^{-1}), MAPE (28.51% e 24.71%) e ME (-0.09 t ha^{-1} , 0.0003 t ha^{-1}), respectivamente. Por outro lado, RLM (Figura 10.a), SVM_Linear (Figura 11.e) e SMV_Poly (Figura 11. f), obtiveram os menores desempenho de previsões de produtividade de FFB, com R^2 ($\leq 0,66$), RMSE ($\geq 0,52 \text{ t ha}^{-1}$), MAPE ($\geq 29\%$).

Os resultados também evidenciaram limitações dos modelos para as produtividades mais baixas ($\leq 1 \text{ t ha}^{-1}$), com erros que podem chegar a ($\geq 50\%$). Reforçou-se também a superestimação à medida que se diminui a produtividade e subestimação com produtividades mais altas (Figura 11).

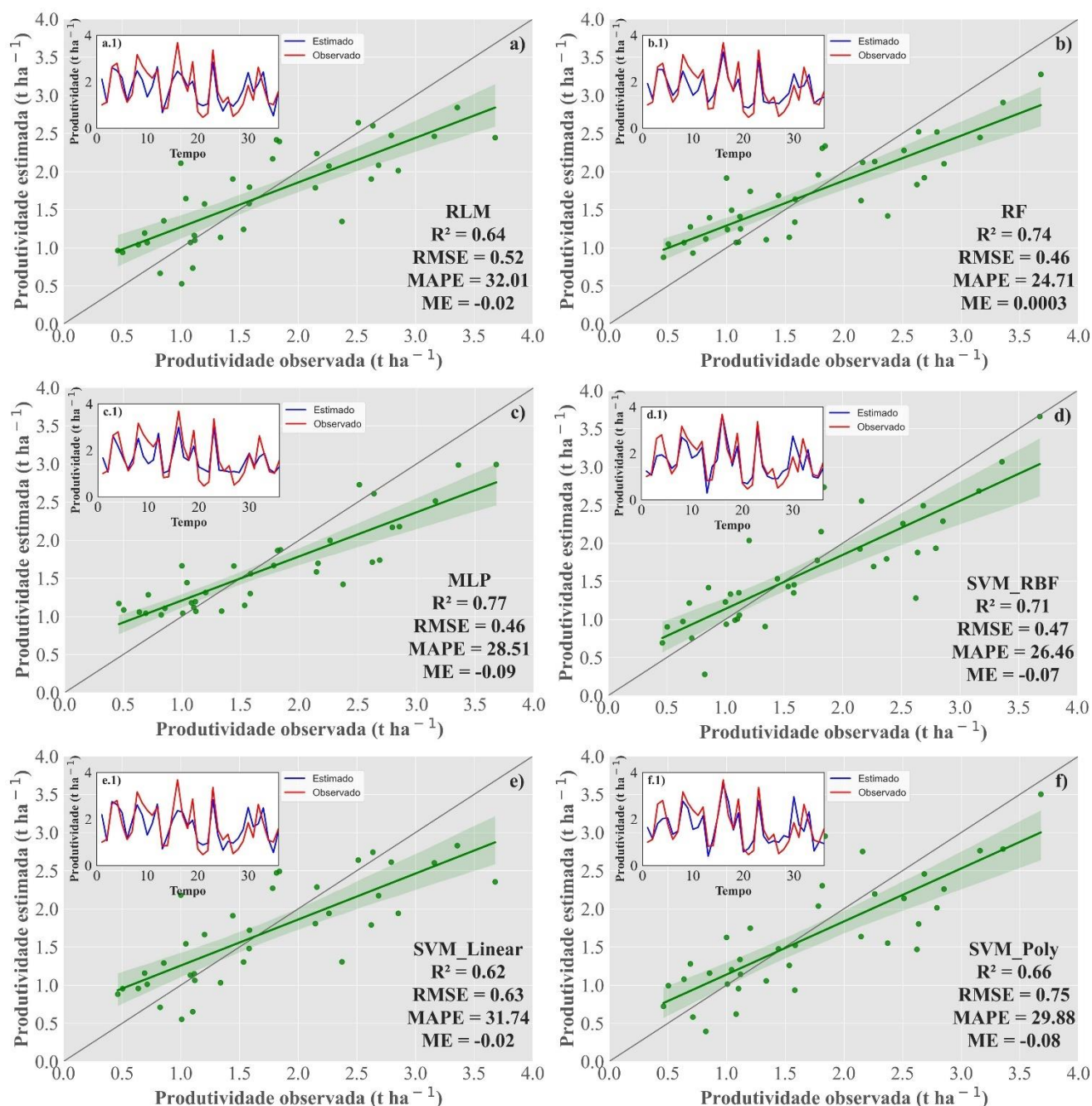


Figura 11. Comparação da produtividade observada obtida em campo e da produtividade prevista com 9 meses antes da colheita, simulada pelos diferentes modelos de machine learning. a) Regressão Linear Múltipla (RLM), b) Random Forest (RF), c) Multilayer Perceptron (MLP), d) Support Vector Machine RBF (SVM_RBF), e) Support Vector Machine Linear (SVM_Linear), f) Support Vector Machine Poly (SVM_Poly).

Avaliando o erro relativo (ER) em razão dos meses de produtividade (janeiro a dezembro) entre os modelos de ML, observa-se que em junho para todos os modelos e períodos diferentes de previsão (4 meses, 6 meses e 9 meses antes da colheita), apresenta em média os maiores erros, tanto para subestimação quanto para superestimação (Figura 12). Reforça-se também que os modelos de ML tendem subestimar os valores observados de FFB para produtividade mais elevadas.

Para as previsões com 4 meses de antecedência da colheita, os maiores ER foram observados para o modelo RF, apresentando em junho erros com valores que chegam em média a 75% (Figura 12.a). Em contrapartida, em janeiro, fevereiro, agosto, novembro e dezembro, apresentaram ER menores que 10%. O SVM_Poly, modelo de melhor desempenho para previsão de FFB (Figura 9), com exceção do mês de abril que apresentou ER de 30, os demais meses do ano obtiveram $ER < 10\%$. Em janeiro, março, maio, julho, agosto, outubro e dezembro destacaram-se entre o SVM_Poly com $ER < 4\%$ (Figura 12.a).

A superestimação observada em abril para o SVM_Poly, pode ser explicada pelo EXC na POL, TMAX na POL e DPV no PC, pois para as previsões com 9 meses antes da colheita, onde essas variáveis não são estimadas, a tendência do ER segue a mesma variação para todos os modelos, com maior ER em junho.

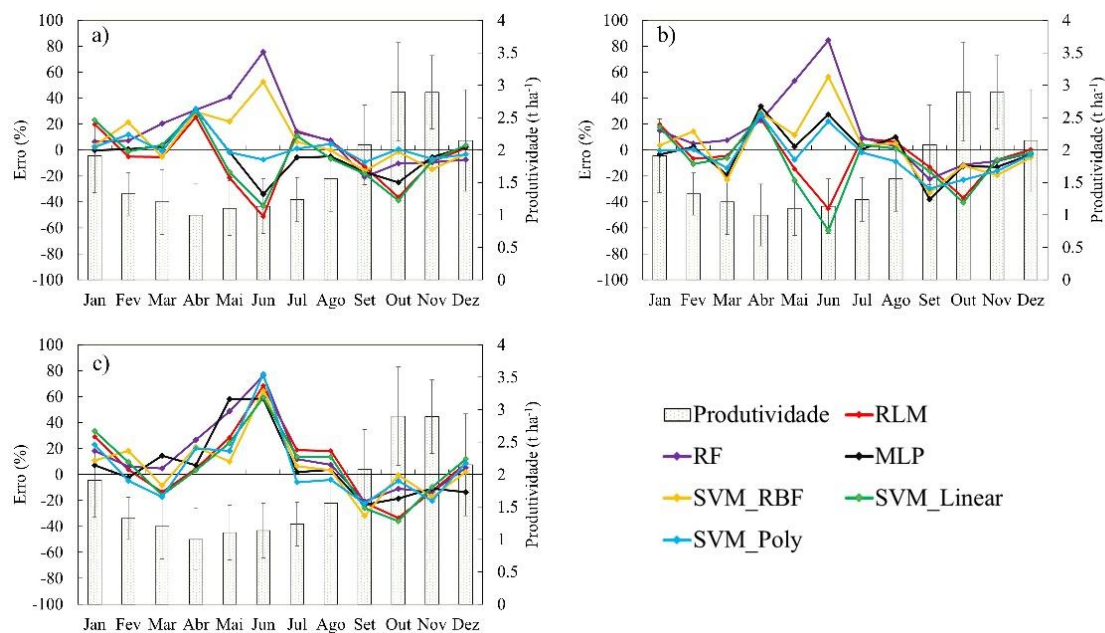


Figura 12. Erro relativo mensal de produtividade da palma de óleo pelos diferentes modelos de Machine Learning para os 3 períodos de previsão, a) 4 meses, b) 6 meses e c) 9 meses antes da colheita.

As variáveis independentes utilizadas, tiveram a mesma robustez pela matriz de correlação de Pearson entre os modelos de ML (Figura 13). O DPV durante a fase de FP (DPV_FP) e RAD_FP, apresentaram as maiores correlações positivas entre todos os modelos, com valores > 0.70 . Enquanto AWHC_FP e DPV_PC, apresentaram as maiores correlações negativas com valores > -0.65 .

O modelo SVM_Poly, que apresentou melhor desempenho para previsão com 4 meses de antecedência (Figura 9.f), obteve correlações positivas na ordem de (DPV_FP = 0.78), (RAD_FP = 0.74), (EXC_DD = 0.54), (AWHC_NP = 0.38), (EXC_POL = 0.38) e (AWHC_ABI = 0.03). Essas variáveis detêm a uma relação linear com a produtividade. De forma inversa, correlações negativas foram obtidas na ordem de -0.32, -0.67 e -0.77 para TMAX_POL, AWHC_FP e DPV_PC, respectivamente (Figura 13.a).

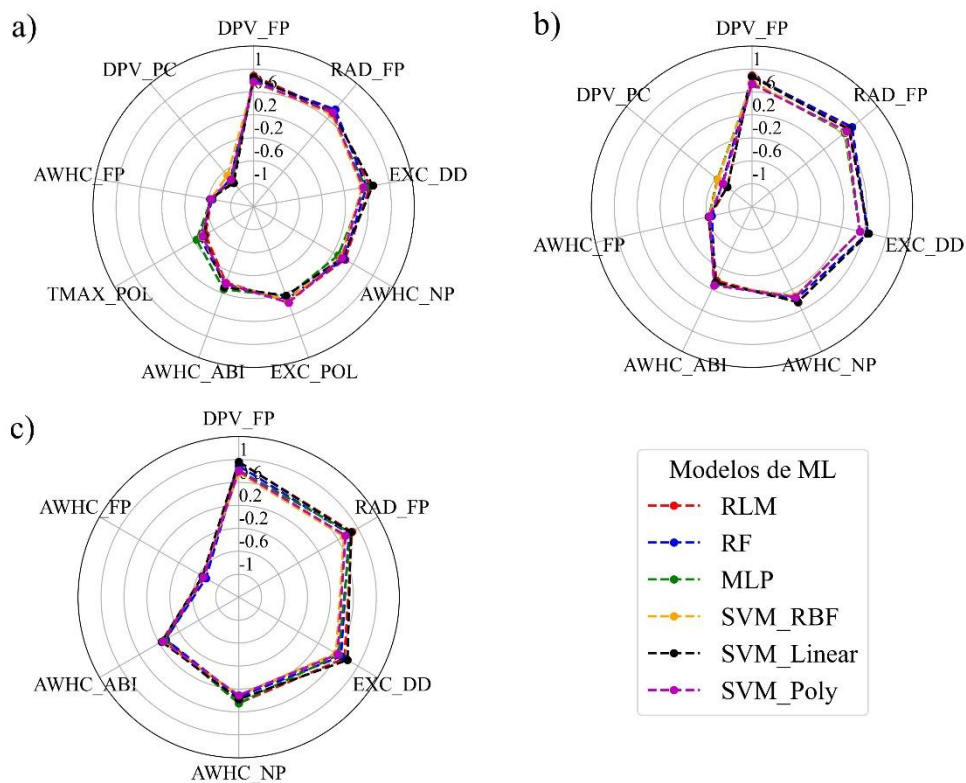


Figura 13. Correlação de Pearson entre a produtividade da palma de óleo e variáveis independentes selecionadas pelos modelos de Machine Learning para os 3 períodos de previsão, a) 4 meses, b) 6 meses e c) 9 meses antes da colheita.

No processo de seleção de variáveis/features de maior importância nos modelos de previsão utilizando o algoritmo de Random Forest (Figura 14), foi retornado 9 elementos para previsão com 4 meses, 7 elementos para previsão com 6 meses e 6 elementos para previsão com 9 meses. DPV_FP apresentou maior peso (> 0.4) no conjunto de variáveis para os três modelos de previsão (4, 6 e 9 meses de antecedência), acompanhada de AWHC_FP e RAD_FP. Por outro lado, as variáveis de menor peso foi EXC_POL para previsão com 4 meses e AWHC_ABI para previsões com 6 e 9 meses antes da colheita de FFB (Figura 14. a, b, c), respectivamente.

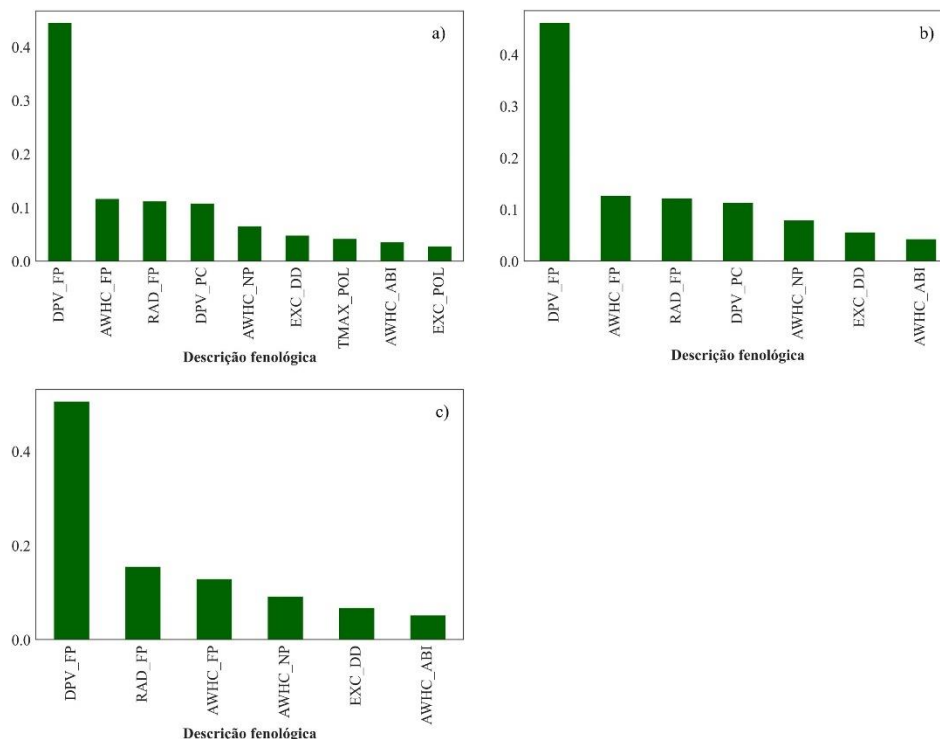


Figura 13. Análise de importância das variáveis selecionadas pelos modelos de machine learning. a) Previsão de produtividade com 4 (quatro) meses de antecedência, b) Previsão de produtividade com 6 (seis) meses de antecedência, c) Previsão de produtividade com 9 (nove) meses de antecedência.

Os modelos com melhores desempenho foram selecionados para estimar a produção de óleo de palma (Figura 15). A estimativa de óleo de palma em função dos dados observados de FFB, apresentou desempenho satisfatório com coeficiente de determinação $R^2 = 0,71$. O modelo de previsão SVM_Poly foi selecionado para estimativa de óleo de palma com 4 meses de antecedência, enquanto o RF para 6 e 9 meses, devido aos menores erros de produtividade de FFB obtidos nas suas validações (Figura 9,10,11), respectivamente.

Assim como para previsão de FFB, a produtividade de óleo de palma apresentou maior precisão na sua estimativa com o modelo de previsão SVM_Poly (Figura 15. b), com $R^2 = 0,61$. Mostrando que para cada aumento de 1 t ha^{-1} de FFB, ocorre um crescimento de produtividade de 0.0767 t ha^{-1} de óleo de palma. O modelo RF para estimativa com 6 e 9 meses de antecedência

apresentou precisões ligeiramente menores, com $R^2 = 0,52$ e $0,50$, respectivamente, (Figura 15. c,d).

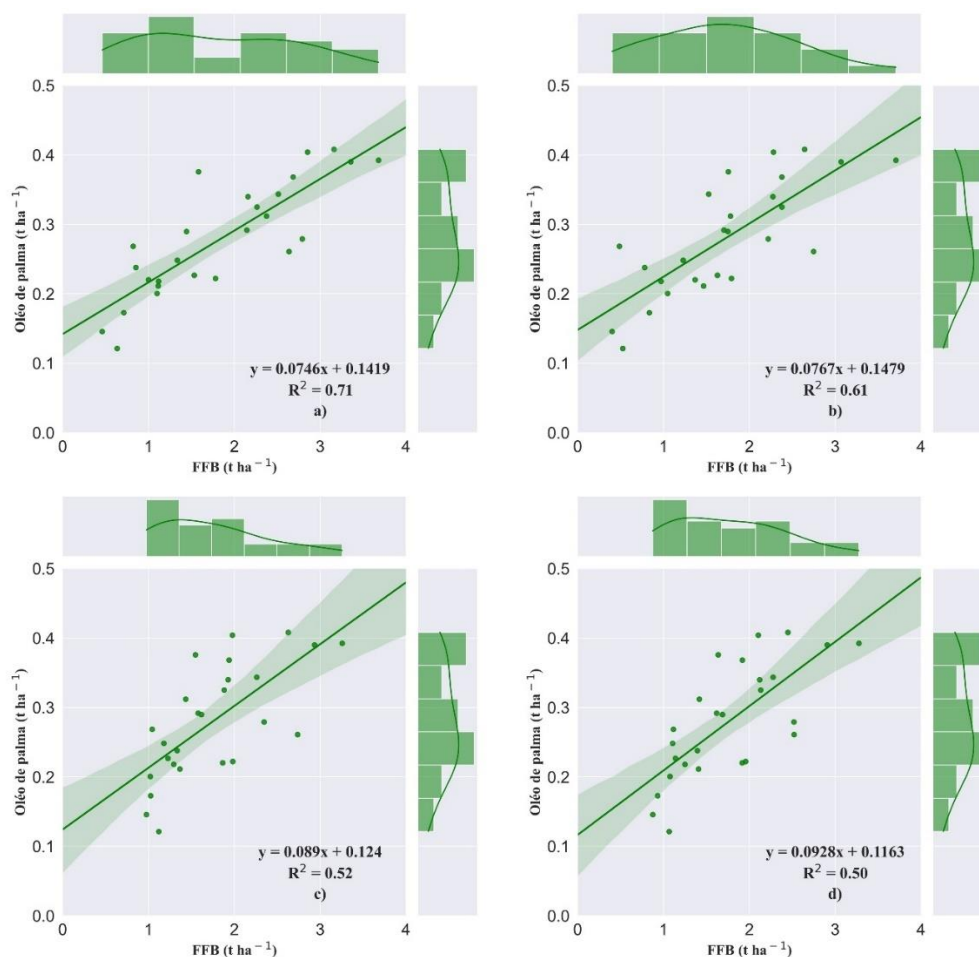


Figura 15. Análise de regressão para estimativa de óleo de palma em função da produção de frutos maduros de palma de óleo (FFB). a) observado, b) SVM_Poly_4meses, c) RF_6meses, d) RF_9meses.

2.4 DISCUSSÃO

A previsão de produtividade da palma de óleo é uma informação crucial para tomada de decisão em todo processo industrial, sendo uma contribuição agrometeorológica que permite subsidiar estratégias de decisão durante os estádios e crescimento, desenvolvimento e produtividade de uma cultura agrícola.

Os resultados das métricas estatísticas obtidos, corroboram com outros estudos para previsão de produtividade da palma de óleo com diferentes

modelos. Keong e Keng (2012) em estudos conduzido na Malásia, avaliaram um modelo de regressão linear múltipla (RLM) estabelecendo a produtividade mensal da palma como variável dependente e empregando variáveis agrometeorológicas em diferentes estádios fenológicos como variáveis independentes. O modelo de RLM apresentou desempenho aceitável, podendo prever com razoável precisão a produtividade da palma de óleo com 9 meses antes da colheita, já que o coeficiente de determinação R^2 foi igual a 0,68. Esses autores observaram que a porcentagem de capacidade de retenção de água no solo disponível (%AWHC) tem implicações significativas na produtividade mensal de cachos de frutas frescas por hectare do cultivo.

Mustakim, Buono e Hermadi (2016), avaliaram o desempenho dos métodos de ML Support Vector Machines (SVM) e Rede neural artificial (MLP) na predição de produtividade da palma de óleo em Riau, Indonésia. O modelo SVM de kernel RBF apresentou os melhores resultados ($R^2 = 0.95$) com erros de 6% de predição, enquanto o MLP apresentou menor coeficiente de determinação, com $R^2 = 0,74$. Os melhores desempenhos obtidos pelo modelo SVM_RBF é dado pela facilidade de modelo trabalhar com número limitado de informações. Contudo, deve-se ressaltar que os resultados obtidos não consideram a natureza ou outros fatores no campo que poderia afetar a produtividade da planta durante o ano, baseando-se apenas na temporalidade da produção.

Oettli, Behera e Yamagata (2018), avaliaram a predição de modelos estatístico de regressão linear múltipla na produtividade de cachos de frutas frescas de palma (FFB) na Malásia, combinando condições climáticas locais e modelos de variação climática em grande escala (ENSO). Foram ajustados dois modelos de previsão, um contento apenas valores reais de variáveis atmosféricas locais (Modelo 1) e outros adicionando os fenômenos climáticos de grande escala (Modelo 2). Para toda a Malásia, o modelo 1 apresentou superestimação de FFB, com $R^2 = 0,39$. Por outro lado, o modelo 2 apresentou melhor desempenho na simulação da produtividade observada com $R^2 = 0,78$, reforçando a importância de avaliações de fenômenos de grande escala como El Niño e La Niña na variabilidade produtiva da Palma de óleo. As médias espaciais

mensais para cada grupo de anos foram calculadas entre dezembro, do ano da colheita, e janeiro, três anos antes, ou seja, para um período de 48 meses. Este período cobre diferentes estágios do ciclo da palmeira.

Shanmuganathan et al. (2014) na Malásia, aplicaram análise de regressão linear múltipla na previsibilidade da FFB da palma de óleo, por meio de variações na temperatura em diferentes fases de desenvolvimento do cultivo e alterações climáticas. A partir das regressões realizadas em diversas regiões da Malásia, os períodos de abertura de flor a polinização foram importantes previsores de produtividade, apresentando alta sensibilidade a variações de temperatura. Os valores de R^2 para todas as regiões juntas no modelo de regressão, sendo um total de 10, foram baixas, com variações de 0,12 a 0,20 em oposição 0,27 a 0,84 para modelos regionais individuais para previsões com 8, 9 e 13 meses antes da colheita. Esses resultados indicam que os efeitos da mudança climática sobre a produtividade da palma de óleo não podem ser generalizados.

Khamis et al. (2006), avaliaram um modelo de rede neural (MLP) e de regressão linear múltipla (RLM) para avaliar a previsão de produtividade da palma de óleo. Foram consideradas como variáveis independente, porcentagem de nitrogênio, fósforo, potássio, cálcio e magnésio nas folhas e cachos de frutas frescas (FFB) como variável dependente. Os autores concluíram que o modelo de rede neural apresenta um desempenho melhor do que o modelo de regressão. Em comparação com o modelo de RLM, os valores de R^2 obtidos a partir do modelo MLP tiveram ganhos de até 62%, com valores partindo de 0,39 na RLM para 0,63 em MLP. Esses resultados podem ser explicados devido o MLP ser uma eficiente ferramenta de modelagem orientada a dados amplamente utilizadas para modelagem e identificação dinâmica de sistemas não lineares (Yang et al., 2017).

Firdawanti, Sumertajaya e Sartono (2019), objetivaram realizar a modelagem de produtividade da palma de óleo, com base no modelo de Random Forest (RF) por meio de variáveis climáticas, topográficas e de fertilizantes na região de Bornéu, sudeste asiático. Os resultados obtidos mostraram correlação de 0.93 e coeficiente de determinação ($R^2 = 0,82$). A previsão de dados de produtividade pode ser feita usando o atraso mínimo de 1 mês.

Oettli, Behera e Yamaga (2018) destacaram que as melhores produtividades possíveis de FFB são obtidos em condições climáticas ótimas com pelo menos 2.000 mm de precipitação distribuída homogeneamente ao longo do ano, correspondendo a cerca de 167 mm mês⁻¹. Além disso, a palma de óleo é sensível ao frio, apresentando temperaturas mínimas entre 22 e 24 °C e as temperaturas máximas entre 29 e 33 °C, enquanto a umidade relativa deve ser superior a 85%. O limite superior de temperatura para fotossíntese eficiente em folhas de palma (dendê) é >38 °C, desde que o DPV seja pequeno (Paramananthan, 2003)

Nos trópicos, a radiação solar é limitada principalmente pela nebulosidade, e por ser uma planta perene com folhagem permanente a palma de óleo é capaz de interceptar a radiação durante todo o ano, levando a produtividades de óleo a ser mais elevada que em comparação a outros cultivos. Contudo, faixas de radiação solar devem ser de pelo menos 16 ou 17 MJ m⁻² d⁻¹ para resultar em uma boa produção de biomassa de FFB (Woittiez et al., 2017).

As taxas de transpiração reais médias em plantações de palma de óleo são 4,0 a 6,5 mm dia⁻¹ na estação chuvosa e 1,0 a 2,5 mm dia⁻¹ em dias períodos secos (Carr, 2011), requerendo chuvas regulares ao longo do ano. A abertura estomática da planta é fortemente afetada pelo DPV, sendo observado que aumentos no DPV de 0,4 para 2,0 kPa resultou em um declínio na taxa fotossintética de 18-19 a 10–12 µmol CO₂ m⁻¹s⁻¹ na palma de óleo, mesmo sob condições de solo com suficiente disponibilidade de água (Carr, 2011).

Os cachos dos frutos de palma de óleo levam vários anos para se desenvolver, ocorrendo um lapso de tempo de 20 a 30 meses entre o início dos fatores de estresse e seu impacto na produtividade. Isso torna difícil separar e quantificar os efeitos de fatores de forma individuais (Adam et al., 2011). Portanto, os modelos de simulação para a previsão da produtividade da palma de óleo para melhor desempenho de suas previsões ou estimativas, devem levar em conta os ciclos de atividades endógenos resultantes dos efeitos das atividades fenológicas das plantas.

Correlações negativas entre produtividades anuais e temperaturas mostram o impacto negativo de temperaturas anormalmente altas na produtividade. E uma correlação positiva entre produtividade e precipitação, levando a ganhos produtivos da palma de óleo (Oettli; Behera; Yamagata, 2018).

Em regiões com P elevadas e regulares, as proporções sexuais da palma de óleo tendem a variar pouco ao longo do ano, em contraste com áreas que experimentam uma estação seca marcada, onde a proporção sexual (Inflorescência masculina e feminina) sofre grandes flutuações (Adam et al., 2011). Investigações de Corley e Tinker (2015), apontaram que a proporção sexual da palma de óleo é afetada por elevados déficit de água no sistema solo-planta-atmosfera, podendo reduzir a proporção de inflorescência femininas em relação a masculina.

Corley (1976) observaram que o período de menor proporção sexual (alta produção de inflorescências masculinas) ocorre durante a estação chuvosa e especularam que esse caráter é uma adaptação contra a redução da densidade de pólen no ar causada pela alta umidade atmosférica. Estes resultados reforçam que a determinação do sexo da palma de óleo é fortemente influenciada por fatores climáticos, com o lançamento da inflorescência masculina sendo promovida pelo DEF. No entanto, outros fatores podem participar da determinação de diferenciação do sexo, sendo esses, fatores metabólicos (reservas de carbono), estado hormonal, fatores genéticos e fatores abióticos (Adam et al. 2011).

Oettli, Behera e Yamaga (2018), observaram que o aborto de inflorescência também é um período sensível ao DEF. No entanto, respostas em abortamento floral ainda precisam ser identificados, devido à alta diversidade entre os materiais genéticos, locais de plantio e condições diversas de cultivo (Corley et al., 1995; Cros et al., 2013).

Durante a polinização chuvas excessivas levam a reduções de insetos polinizadores, portanto, nota-se que tem um limite para a relação entre maior chuva maior produtividade, pois chuvas contínuas muito altas durante a polinização reduzem a população de gorgulho polinizadores e consequentemente menores produções. Esses resultados são refletidos

principalmente na extração de óleo de palma, com taxas de extração caindo de 21.2% para 18.8% (Woittiez et al., 2017; Oettli, Behera; Yamaga, 2018)

2.5 CONCLUSÃO

Os modelos de Machine Learning (ML) quando bem treinados e testados apresentam desempenhos com RMSE médio de 0.42 t ha⁻¹ na previsão de produtividade da palma de óleo. Ao combinar o efeito em diferentes estádios fenológicos da planta com a produtividade, o modelo SVM Poly desenvolvido apresentou melhor precisão, tendência e acurácia, com antecipação de 4 meses antes da colheita.

A identificação de variáveis agrometeorológicas e suas interações com os processos fenológicos do cultivo é essencial para compreensão e determinação de práticas agrícolas para ganhos de produtividade e na formulação de modelos de previsão a níveis de produção agrícola. Os modelos de ML são ferramentas estratégicas para subsidiar a tomada de decisão nas indústrias e aos produtores de palma de óleo no estado do Pará.

2.6 REFERÊNCIAS

Adam H, Collin Myrian, Richaud R, Beulé T, Cros D, Omoré A, Nodichao L, Nouy B, Tregear JW (2011) Environmental regulation of sex determination in oil palm: current knowledge and insights from other species. *Annals of Botany* 108: 1529–1537. doi:10.1093/aob/mcr151.

Adetiloye T, Awasthi A (2017) Predicting Short-Term Congested Traffic Flow on Urban Motorway Networks. *Handbook of Neural Computation*, p.145-165. <https://doi.org/10.1016/B978-0-12-811318-9.00008-9>.

Alkhasawneh MS, Tay LT (2017) A hybrid intelligent system integrating the cascade forward neural network with Elman neural network, *Arabian J. Sci. Eng.* 1–13.

Antanasijević D, Antanasijević J, Pocajt V (2018) Prediction of the transition temperature of bent-core liquid crystals using fuzzy “digital thermometer” model based on artificial neural networks. *Engineering Applications of Artificial Intelligence*, 71, 251-258. DOI:<https://doi.org/10.1016/j.engappai.2018.03.009>.

Benezoli VH, Imbuzeiro HMA, Cuadra SV, Colmanetti, MAA, Araújo, AC, Stiegler C, Motoike, SY (2021) Modeling oil palm crop for Brazilian climate conditions. *Agricultural Systems*. 190: 103130. doi.org/10.1016/j.agry.2021.103130.

Brum M, Oliveira RS, López JG, Licata J, Pypker T, Chia GS, Tinôco RS, Asbjørnsen H (2021) Effects of irrigation on oil palm transpiration during ENSO-induced drought in the Brazilian Eastern Amazon. *Agricultural Water Management* 245:106569. doi.org/10.1016/j.agwat.2020.106569.

Caliman JP, Southworth A (1998) Effect of drought and haze on the performance of oil palm. *International Oil Palm Conference* (Bali).

Camargo AP (1971) *Balanço hídrico no Estado de São Paulo*. Campinas: Instituto Agrônômico, 24p. Boletim 116.

Carr MKV (2011) The water relations and irrigation requirements of oil palm (*elaeis guineensis*): a review. *Expl Agric.* 47:4, p. 629–652. doi:10.1017/S0014479711000494.

Chlingaryan A, Sukkariéh S, Whelan B (2018) Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: a review. *Comput. Electron. Agric.* 151, 61–69. <https://doi.org/10.1016/j.compag.2018.05.012>.

Corley RH, Tinker PB (2015) *The Oil Palm*, 5th ed. Blackwell Science Ltd., Oxford <https://doi.org/10.1002/9781118953297>.

Corley RHV (1976) Sex differentiation in oil palm: effects of growth regulators. *Journal of Experimental Botany* 27: 553–558.

Cortes C, Vapnik V (1995) Support-vector networks. *Machine learning*, 20, 273–297. DOI: [https://doi: 10.1023/A:1022627411411](https://doi.org/10.1023/A:1022627411411).

Cros D, Flori A, Nodichao L, Omoré A, Nouy B (2013) Differential Response to Water Balance and Bunch Load Generates Diversity of Bunch Production Profiles Among Oil Palm Crosses (*Elaeis guineensis*). *Trop. Plant Biol.* 6, 26–36.

Du KL, Swamy M (2014) *Neural Networks and Statistical Learning*. Springer-Verlag, London, 988p. DOI: <https://doi.org/10.1007/978-1-4471-5571-3>.

European Sustainable Palm Oil (2019) Choosing sustainable palm oil: Progress report on the import and use of sustainable palm oil in Europe. MVO – The Netherlands Oils and Fats Industry, Zoetermeer, The Netherlands (Accessed 09 march 2021).

FAO (2021). FAOSTAT Data: Production – Crops. Food and Agriculture Organization of the United Nations (FAO) (Accessed 7 January 2021). <http://www.fao.org/faostat/en/#data/QC>.

Farrugia J, Griffin S, Valdramidis VP, Camilleri K, Falzon O (2021) Principal component analysis of hyperspectral data for early detection of mould in cheeselets. *Current Research in Food Science*.

Firdawanti AR, Sumertajaya IM, Sartono B (2019) Random Forest Lag Distributed Regression for Forecasting on Palm Oil Production. *Proceedings of the 1st International Conference on Statistics and Analytics, ICSA*. <http://dx.doi.org/10.4108/eai.2-8-2019.2290493>.

Forero DC, Hormaza P, Romero HM (2012) Phenological growth stages of African oil palm (*Elaeis guineensis*). *Ann Appl Biol* 160, 56–65.

Forthofer RN, Lee ES, Hernandez M (2007) Linear Regression. Biostatistics (Second Edition), p. 349-386.

Gonçalves AO, Bastos TX, Barros AH, Filho AR, Motta EF (2010). Procedimento metodológico da Avaliação da Aptidão Climática para a Cultura da Palma de Óleo nas Áreas Desmatadas da Amazônia Legal. (Org), Filho AR, Motta PEF, Freitas PL, Teixeira WG. Zoneamento Agroecológico, Produção e Manejo para a Cultura da Palma de Óleo na Amazônia. Embrapa Solos, Rio de Janeiro, 222p.

Goodfellow I, Bengio Y, Courville A (2016). Deep Learning. MIT Press. <http://www.deeplearningbook.org> (Last accessed: Feb 15, 2021).

Gujarati DN and Porter DC (2011) Basics Econometrics, 5a edn. McGraw-Hill, New York.

Haghverdi A, Washington-Allen RA, Leib BG (2018) Prediction of cotton lint yield from phenology of crop indices using artificial neural networks. Computers and Electronics in Agriculture, 152:186–197. <https://doi.org/10.1016/j.compag.2018.07.021>.

Hariharan R (2021). Random forest regression analysis on combined role of meteorological indicators in disease dissemination in an Indian city: A case study of New Delhi. Urban Climate, 36, 100780. DOI: <https://doi.org/10.1016/j.uclim.2021.100780>.

Hoffmann MP, Donough CR, Cook SE, Fisher MJ, Lim CH, Lim YL, Cock J, Kam SP, Mohanaraj SN, Indrasuara K, Tittinutchanon P, Oberthür T (2017) Yield gap analysis in oil palm: framework development and application in commercial operations in Southeast Asia. Agric. Syst. 151, 12–19. <https://doi.org/10.1016/j.agsy.2016.11.005>.

Hormaza P, Fuquen EM, Romero HM (2012) Phenology of the oil palm interspecific hybrid *Elaeis oleifera* × *Elaeis guineenses*. *Sci. Agric.* 69, 275-280.

Hotelling H (1933) Analysis of a complex of statistical variables into principal components. *J. Educ. Psychol.* 24, 417-441.

IBGE (2021). Instituto Brasileiro de Geografia e Estatística. Sistema IBGE de Recuperação Automática – SIDRA. (Accessed 16 March 2021). <https://sidra.ibge.gov.br/home/pimpfbr/brasil>.

Keong YK, Keng WM (2012) Statistical Modeling of Weather-based Yield Forecasting for Young Mature Oil Palm. *APCBEE Procedia*, 4, p.58-65. <https://doi.org/10.1016/j.apcbee.2012.11.011>.

Khamis A, Ismail Z, Haron K, Mohammed AT (2016) Neural Network Model for Oil Palm Yield Modeling. *Journal of Applied Sciences* 6 (2): 391-399.

Khatun R, Reza MIH, Moniruzzaman M, Yaakob Z (2017) Sustainable oil palm industry: The possibilities. *Renew. Sustain. Energy Rev.* 76, 608–619. doi.org/10.1016/j.rser.2017.03.077.

Kocyigit Y, Alkan A, Erol A (2008) Classification of EEG recordings by using fast independent component analysis and artificial neural network, *J. Med. Syst.* 32, 17–20.

Liakos KG, Busato, P, Moshou D, Pearson S, Bochtis D (2018) Machine Learning in Agriculture: A Review. *Sensors*, 18, 2674; doi:10.3390/s18082674.

Liu Y, Heuvelink GBM, Bai Z, He P, Xu Xinpeng, Ding W, Huang S (2021) Analysis of spatio-temporal variation of crop yield in China using stepwise multiple linear regression. *Field Crops Research*, 264:108098. <https://doi.org/10.1016/j.fcr.2021.108098>.

Luciano ACS, Picoli, MCA, Duft DG, Rocha JV, Leal MRLV, Maire GL (2021) Empirical model for forecasting sugarcane yield on a local scale in Brazil using Landsat imagery and random forest algorithm. *Computers and Electronics in Agriculture*, 184: 106063. <https://doi.org/10.1016/j.compag.2021.106063>.

Manjunathan N, Rajesh P, Thangadurai E, Suresh A (2020) Crop Yield Prediction Using Linear Support Vector Machine. *European Journal of Molecular & Clinical Medicine*, 6, 1-7.

Martorano LG, Pereira LC, Cezar EGM, Pereira ICB (1993). Estudos climáticos do Estado do Pará, classificação climática (Köppen) e deficiência hídrica (Thornthwhite, Matter). Belém, SUDAM/EMBRAPA, SNLCS pp. 53.

Martorano LG, Vitorino MI, Silva BPPC, Moraes JRSC, Lisboa LS, Sotta ED, Reichardt K (2017) Climate conditions in the eastern amazon: Rainfall variability in Belem and indicative of soil water deficit. *African Journal Of Agricultural Research*. 12, 1801-1810.

Miles J (2014) Tolerance and Variance Inflation Factor. Wiley StatsRef: Statistics Reference Online.

Moraes JRSC, Rolim GS, Martorano LG, Aparecido LEO, Bispo RC, Valeriano TTB (2020) Performance of the ECMWF in air temperature and precipitation estimates in the Brazilian Amazon. *Theoretical and Applied Climatology*, 141:803–816. <http://dx.doi.org/10.1007/s00704-020-03231-2>.

Moraes JRSC, Rolim GS, Martorano LG, Aparecido LEO, Oliveira MSP, Neto JTF (2020) Agrometeorological models to forecast açaí (*Euterpe oleracea* Mart.) yield in the Eastern Amazon. *J Sci Food Agric*, 100: 1558–1569. <http://dx.doi.org/10.1002/jsfa.10164>.

Mustakim, Buono A, Hermadi I (2016) Performance comparison between support vector regression and artificial neural network for prediction of Oil Palm production. *Journal of Computer Science and Information*. 9:1, 1-8.

Oettli P, Behera SK, Yamagata T (2018) Climate Based Predictability of Oil Palm Tree Yield in Malaysia. *Scientific Reports*, 8:2271. doi:10.1038/s41598-018-20298-0.

Pant J, Pant RP, Singh MK, Singh DP, Pant H (2021) Analysis of agricultural crop yield prediction using statistical techniques of machine learning. *Materials Today: Proceedings*. <https://doi.org/10.1016/j.matpr.2021.01.948>.

Paramananthan S (2003) Land selection for oil palm. In: Fairhurst, T.H., Härdter, R. (Eds.), *Oil Palm: Management for Large and Sustainable Yields*. PPIC-PPI-IPI, Singapore, 27–57.

Pirker J, Mosnier A, Kraxner F, Havlík P, Obersteiner M (2016) What are the limits to oil palm expansion? *Global Environ. Change* 40, 73–81. <https://doi.org/10.1016/j.gloenvcha.2016.06.007>.

Rhebergen T, Fairhurst T, Zingore S, Fisher M, Oberthur T, Whitbread A (2016) Climate, soil and land-use based land suitability evaluation for oil palm production in Ghana. *Eur. J. Agron.* 81, 1–14. <https://doi.org/10.1016/j.eja.2016.08.004>.

Shaharum NSN, Shafri HZM, Ghani WAWAK, Samsatli S, Yusuf B, Al-Habshi M MA, Prince HM (2018). Image classification for mapping oil palm distribution via support vector machine using Scikit-learn module. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 42(4/W9).

Shanmuganathan S, Narayanan A, Mohamed M, Ibrahim R, Khalid H (2014) A Hybrid Approach to Modelling the Climate Change Effects on Malaysia's Oil Palm

Yield at the Regional Scale. *Recent Advances on Soft Computing and Data Mining*, 287, p. 335-345.

Shi K, Wang J, Tang Y, Zhong S (2020) Reliable asynchronous sampled-data filtering of T–S fuzzy uncertain delayed neural networks with stochastic switched topologies. *Fuzzy Sets Syst.* 381, 1–25.

Surre C (1968) Les besions en eau du palmier à huile. Calcul du bilan de l'eau et ses applications pratiques. *Oléagineux*. 23:165– 167.

Tan KT, Lee KT, Mohamed AR (2009) Production of FAME by palm oil transesterification via supercritical methanol technology. *Biomass and Bioenergy*. v.33, pp. 1096-1099.

Tápia DFJ, Doliente SS, Samsatli S (2021) How much land is available for sustainable palm oil?. *Land Use Policy* 102 (2021) 105187. doi.org/10.1016/j.landusepol.2020.105187.

Thornthwaite CW, Hare FK (1955) Climatic classification in forestry. *Unasylva* 9:51–59.

Torres FTP, Machado PJO (2008) *Introdução à Climatologia*. Ubá: Ed. Geographica. – (Série Textos Básicos de Geografia) 234p.

Wang P, Hafshejani BA, Wang D (2021) An improved multilayer perceptron approach for detecting sugarcane yield production in IoT based smart agriculture. *Microprocessors and Microsystems*. 82, 103822. <https://doi.org/10.1016/j.micpro.2021.103822>

Willcock S, Hooftman DA, Bagstad KJ, Balbi S, Marzo A, Prato C, Sciandrello S, Signorello G, Voigt B, Villa F, Bullock JM, Athanasiadis NL (2018) Machine learning for ecosystem services. *Ecosystem services* 33, 165–174. <https://doi.org/10.1016/j.ecoser.2018.04.004>.

Woittiez LS, Van Wijk MT, Slingerland M, Van Noordwijk M, Giller KE (2017) Yield gaps in oil palm: A quantitative review of contributing factors. *Eur. J. Agron.* 83, 57–77. <https://doi.org/10.1016/J.EJA.2016.11.002>.

Xavier AC, King CW, Scanlon BR (2016) Daily gridded meteorological variables in Brazil (1980-2013). *Int. J. Climatol.* 36, 2644–2659. DOI:<https://doi.org/10.1002/joc.4518>.

Yang T, Asanjan AA, Faridzad M, Hayatbini N, Gao X, Sorooshian S (2017). Na enhanced artificial neural network with a shuffled complex evolutionary global optimization with principal component analysis. *Inform. Sci. (Ny)* 418–419, 302–316. <https://doi.org/10.1016/j.ins.2017.08.003>.

CAPÍTULO 3 – Previsão de produtividade de frutos de açaí (*Euterpe oleracea* Mart.) na região Nordeste paraense na Amazônia Oriental modelado por inteligência artificial

RESUMO

O sucesso da produtividade agrícola depende de vários fatores biofísicos, qualidade das sementes, fator genético e manejo no campo. A previsão agrometeorológica visa subsidiar o planejamento agrícola tendo entre as variáveis respostas mais responsivas as condições de tempo e clima de uma determinada região. A modelagem utilizando a inteligência artificial tem apontado resultados promissores nas previsões de produtividade de várias culturas. O objetivo neste capítulo foi desenvolver modelos agrometeorológicos para previsão de produtividade de frutos de açaí, testando o desempenho de diferentes modelos usando inteligência artificial para condições irrigada e sequeiro. Os algoritmos de ML avaliados foram Regressão Linear Múltipla (MLP), Random Forest (RF), Redes Neurais Artificiais (MLP) e Support Vector Machine (SVM) de kernel de Função de Base Radial (SVM_RBF), Linear (SVM_Linear) e Polinomial (SVM_Poly). Para validar o desempenho dos modelos foram utilizados três métricas estatísticas, coeficiente de determinação (R^2), erro absoluto médio (MAPE) e o erro médio (ME). Dados de produção da cultura entre o período de 2003 a 2012 foram fornecidos por pesquisadores, oriundos de experimentos de campo da Embrapa Amazônia Oriental, em unidade no município de Tomé-Açu, no nordeste paraense. Foram analisados dados de dois sistemas de manejo: cultivo irrigado e sem irrigação (sequeiro). Informações climáticas de Precipitação pluvial (mm), Temperatura máxima (TMAX), Temperatura mínima (TMIN), Radiação Solar ($\text{MJ m}^2 \text{ dia}^{-1}$) e Umidade relativa (%) foram obtidos do banco de dados meteorológicos diários disponíveis em GRID com resolução espacial de $0,25^\circ$. Avaliou-se a existência ou não de multicolinearidade, bem como aplicou-se a análise de componente principal. Essas análises foram utilizadas como fatores importantes na seleção de variáveis independentes. Os resultados evidenciaram que os modelos baseados em algoritmos de RF apresentaram menor média de erro, ou seja, $\text{MAPE} < 19\%$, $R^2 > 0,80$ e menor tendência média dos dados, com ME variando para a condição de sequeiro entre -5,41 a -10,34 e para irrigado -162,81 e -40,59, entre o primeiro e segundo semestre, respectivamente. Sob condição irrigada as previsões agrometeorológicas apresentaram alto desempenho com 4 meses de antecipação, considerando o período mais chuvoso na região, ou seja, no primeiro semestre e com 7 meses antes da colheita, no segundo semestre. Para os cultivos de sequeiro os prognósticos são de até 6 meses de antecipação, em ambos os períodos. Modelos de ML são eficazes e versáteis com alto potencial para estimar a produtividade e subsidiar estratégias de decisão na cadeia produtiva de açaí no nordeste paraense.

Palavras-chaves: Arecaceae, Inteligência artificial, Manejo de cultivo, Variáveis climáticas, Python

Yield forecasting for açai fruit (*Euterpe oleracea* Mart.) in Eastern Amazonia modeled by artificial intelligence

ABSTRACT

The success of agricultural productivity depends on several biophysical factors, seed quality, genetic factors, and field management. The agrometeorological forecast aims to subsidize the agricultural planning having among the most responsive variables the weather and climate conditions of a given region. Modeling using the assumptions of artificial intelligence has shown promising results for crop yield forecasting. The objective of this chapter was to develop agrometeorological models for açai fruit yield prediction, testing the performance of different models using artificial intelligence for irrigated and rainfed conditions. The ML algorithms evaluated were Multiple Linear Regression (MLP), Random Forest (RF), Artificial Neural Networks (MLP) and Support Vector Machine (SVM) of Radial (SVM_RBF), Linear (SVM_Linear) and polynomial (SVM_Poly) basis function kernel. Three statistical metrics, coefficient of determination (R^2), Mean absolute error (MAPE) and the mean error (ME) were used to validate the performance of the models. Production data of the crop between the period 2003 to 2012 were provided by researchers, from field experiments at Embrapa Amazônia Oriental, in a unit in the municipality of Tomé-Açu, in northeastern Pará. Data from two management systems were analyzed: irrigated and non-irrigated (rainfed). Climate information of rainfall (mm), maximum temperature (TMAX), minimum temperature (TMIN), solar radiation (MJ m² day⁻¹) and relative humidity (%) were obtained from the daily meteorological database available in GRID with spatial resolution of 0.25°. The existence or not of multicollinearity was assessed, as well as principal component analysis was applied. These analyses were used as important factors in the selection of independent variables. The results showed that the models based on RF algorithms presented lower mean error, that is, MAPE < 19%, R^2 > 0.80 and lower mean trend of the data, with ME ranging for the rainfed condition between -5.41 and -10.34 and for irrigated -162.81 and -40.59, between the first and second semester, respectively. Under irrigated conditions the agrometeorological forecasts presented high performance with 4 months of anticipation, considering the rainiest period in the region, that is, in the first semester and with 7 months before harvest, in the second semester. For rainfed crops the predictions are up to 6 months in advance, in both periods. ML models are effective and versatile with high potential to estimate productivity and support decision strategies in the productive chain of açai in northeastern Pará.

Keywords: Arecaceae, Artificial Intelligence, Crop Management, Climatic Variables, Python

3.1 INTRODUÇÃO

Compreender a relação do clima com a resposta produtiva de um determinado cultivo, aumenta a capacidade dos agricultores e planejadores agrícolas na tomada de decisão de como alocar recursos e entender melhor as variabilidades existentes no campo (Culman et al., 2019).

O Açaí (*Euterpe oleracea* Mart.) a sua produtividade está fortemente relacionada às condições climáticas e à disponibilidade de água no solo (Moraes et al., 2020a). Essa cultura possui uma grande importância econômica para fruticultura da região amazônica, sendo uma fruta amplamente consumida no Brasil e no mundo (Barbosa et al., 2021), além dos aspectos sociais e ambientais prestadas pelo seu cultivo, contribuindo para a subsistência de centenas de comunidades tradicionais na Amazônia brasileira (Lopes et al., 2019).

Parte do interesse pelo açaí no mundo, se deve principalmente aos potenciais benefícios à saúde atribuídos à sua composição nutricional (Barbosa et al., 2021) e à presença de compostos bioativos (Insfran et al., 2004; Guerra et al., 2011).

O Brasil é o principal produtor de frutos de açaí, com produção em 2019 superior a 1.3 milhões de toneladas (IBGE, 2021). O estado do Pará concentra mais de 90% da produção nacional, apresentando um incremento de 38% de área colhida, entre 2015 a 2019, passando de 135.691 mil para 188.015 mil hectares, respectivamente, com uma produtividade média de 7 mil kg ha⁻¹.

Estima-se que do volume de açaí exportado para outros países, contabiliza-se 70% para os Estados Unidos da América (EUA), com um grande consumo atribuído aos valores energéticos da fruta (Tavares; Homma, 2015). Outro aspecto curioso é que os EUA aparecem como os principais exportadores de produtos processados de açaí no mundo (Conab, 2019), ou seja, a produção paraense está perdendo em valor agregado no mercado de polpas de açaí.

Possivelmente, apesar das grandes produções da cultura no Brasil e aumentos de áreas plantadas, a concentração da colheita da fruta de julho a dezembro, dificultam as exportações, pois os compradores exigem constância de volume e oferta permanente (Conab, 2019). Essa variação da produção é caracterizada principalmente pela sazonalidade do clima nas regiões produtoras,

que influencia diretamente na produção de plantas com ciclo contínuo e longo como o açaizeiro. Portanto, a aplicação de modelos que permitam previsões precisas de produtividade é essencial para o planejamento agrícola, comercial e alimentar, tanto em escala regional quanto global (Jeong et al., 2016).

A variabilidade da produção agrícola devido a eventos climáticos extremos, como secas e altas temperaturas, continuam sendo uma grande preocupação para os agricultores, governos e mercados, reforçando a necessidade de previsões precisas e oportunas da produção agrícola em um clima incerto (Jeong et al., 2016).

Dessa forma, métodos eficientes foram desenvolvidos para prever a respostas da produção agrícola à variabilidade climática, os quais podem ser classificados principalmente em modelos de cultivo e modelos estatísticos (Nayamuth, 1999; Dourado-Neto et al., 1998).

Os modelos de cultivo simulam processos fisiológicos de crescimento e desenvolvimento da planta em respostas as condições ambientais e práticas de manejo (Chipanshi et al., 2015). Esses modelos apresentam processos muito robustos e eficientes (Battisti et al., 2017), no entanto, devido sua calibração e validação serem baseados em processos das interações entre planta, meio ambiente e manejo, necessitam de informações detalhadas e variados parâmetros de ajustes (Patrício; Rieder, 2018).

Por outro lado, os modelos estatísticos estimam ou preveem as relações diretas entre as variáveis preditoras e a produtividade da cultura em um determinado conjunto de dados, não necessitando de um conhecimento sobre mecanismos físicos que produziram os dados (Jeong et al., 2016). Tais técnicas são baratas, relativamente fácil de aplicar, e não precisa de uma estrutura pré-definida de modelo (Sanchez; Solis; Bustamante, 2014).

Técnicas de modelagem estatística baseados em algoritmos de Machine Learning (ML) fornecem um caminho promissor para melhorar as previsões de produtividade de safra (Paudel et al., 2021). Modelos de ML é um ramo da Inteligência Artificial (IA) com foco no aprendizado, determinando padrões e correlações e retirar informações a partir de conjuntos de dados (Klompenburg; Kassahun; Catal, 2020). Os modelos precisam ser treinados usando conjuntos

de dados, onde os resultados são representados com base na experiência anterior, podendo modelar relacionamentos não lineares entre várias fontes de informações (Chlingaryan et al., 2018).

Muitos estudos aplicaram ML para prever as produtividades de certas safras em locais específicos (Manimekalai; Nandhini, 2018). Dentre esses se destaca os algoritmos de Random Forest (RF), Regressão Linear Multipla (RLM), Redes Neurais Artificiais e Supporte Vector Machine (SVM).

Diversos autores têm estudado a aplicação de ML na previsão da produtividade agrícola, como se verifica nos trabalhos de Wang et al. (2020) e Jiang et al. (2020) com Redes Neurais, Rolim et al. (2015) e Aparecido et al. (2020) aplicando modelos de RLM, Pant et al. (2021) e Prasad et al. (2020) utilizando RF e Kumar et al. (2019), Manjunathan et al. (2020) utilizando SVM na previsão de produtividade.

Contudo, observa-se que para o açaí é inovadora essa aplicação de modelos de ML, assim como para palmeiras agrícolas de alto valor comercial no Brasil. Neste sentido, objetiva-se com este trabalho testar modelos agrometeorológicos para prognosticar a produtividade de frutos de açaí na região Nordeste paraense na Amazônia Oriental usando inteligência artificial.

3.2 MATERIAL E MÉTODOS

Dados de produtividade de açaí referente ao período de 2003 a 2012 foram coletados na base experimental da Embrapa Amazônia Oriental, localizada no município de Tomé-Açu, Pará. Observa-se que as principais áreas produtoras de açaí no Pará estão localizadas na tipologia climática Af₂ e Af₃, com chuvas acima de 2000 mm ano⁻¹ (Figura 1). Foram analisados dados de dois sistemas de manejo: cultivo irrigado e sem irrigação (sequeiro). A região de estudo faz parte do nordeste paraense, e pertence à tipologia climática Af₃ (sem período seco definido com mês menos chuvoso superior a 60 mm), de acordo com o método de Koppen, adaptado por Martorano et al. (1993), assemelhando-se com o tipo climático do noroeste da Amazônia ocidental (Moraes et al., 2020b) (Figura 1).

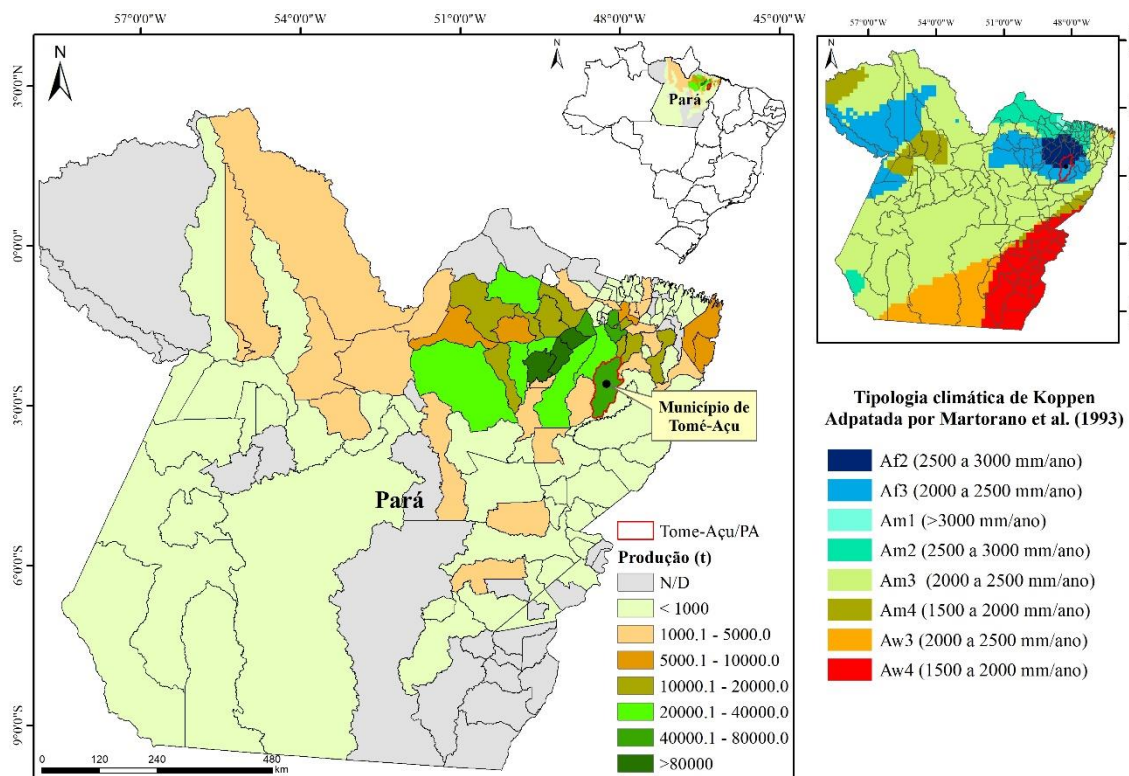


Figura 1. Mapa de produção de açaí e tipologia climática de Koppen adaptada por Martorano para o estado do Pará.

Fonte: IBGE, 2021. Mapa elaborado pelo autor.

O conjunto de dados climáticos e de cultivo do açaizeiro, assim como as análises exploratórias para aplicação dos modelos de machine learning na previsão de produtividade da cultura aplicados neste estudo, é representado no fluxograma de trabalho da figura 2.

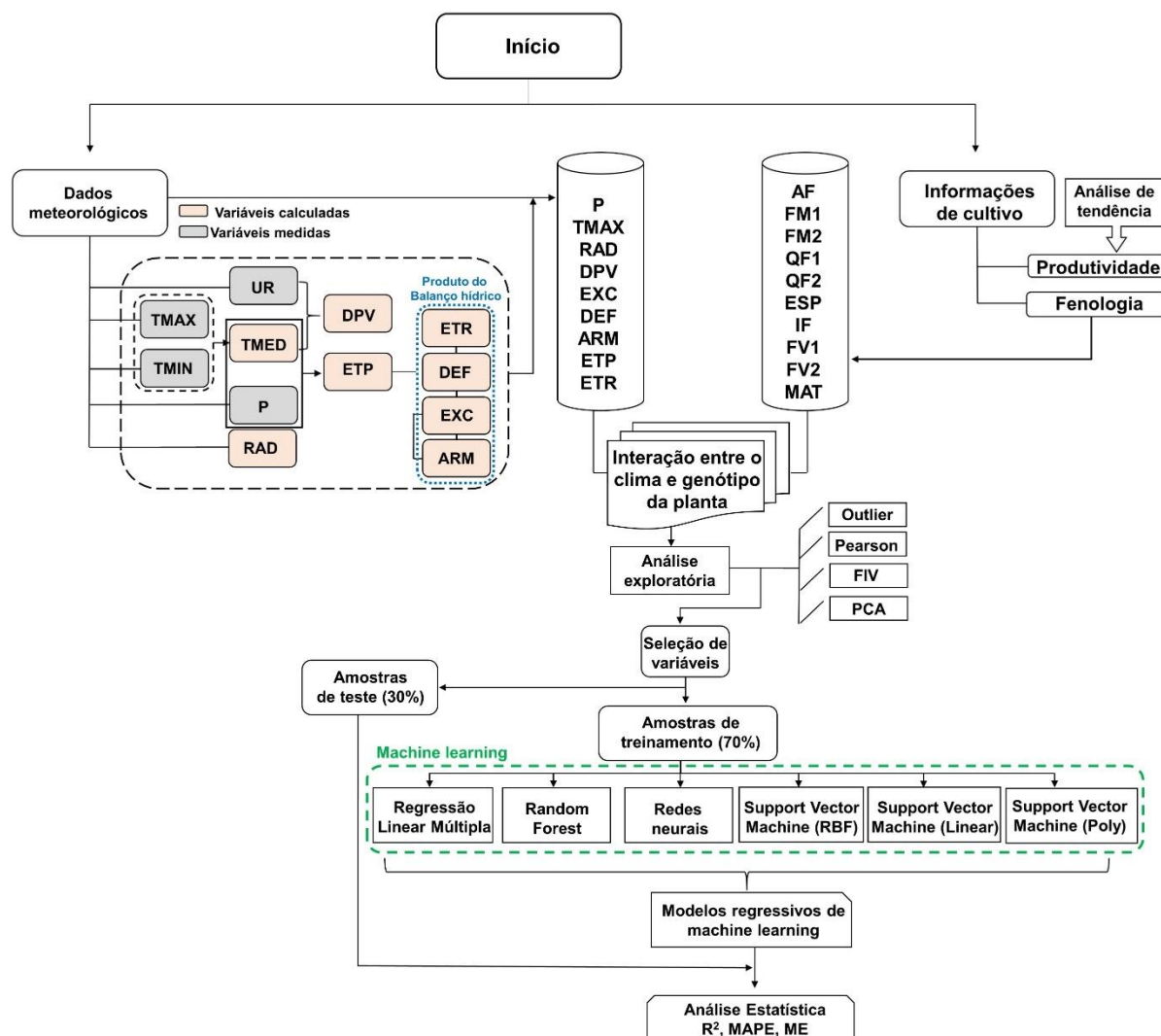


Figura 2. Etapas metodológicas conduzidos para este estudo. Variáveis meteorológicas: Precipitação pluvial (P mm), Deficiência hídrica (DEF mm), Excedente hídrico (EXC mm), Armazenamento de água no solo (ARM mm), Evapotranspiração potencial (ETP mm), Evapotranspiração real (ETR mm), Temperatura média (TMED °C), Temperatura mínima (TMIN °C), Temperatura máxima (TMAX °C), Radiação solar (RAD MJ m⁻² dia⁻¹), Déficit de pressão de vapor (DPV kPa), Umidade relativa (UR %). Variáveis de cultivo: AF (Emissão de flexa); FM1 (Folha madura 1); FM2 (Folha madura 2); QF1 (Folha senescente 1); QF2 (Folha senescente 2); ESP (Emissão de espata); IF (Inflorescência); FV1 (Fruto verde 1); FV2 (Fruto verde 2); MAT (Maturação).

Dados meteorológicos de precipitação pluvial (mm), Temperatura máxima (TMAX), Temperatura mínima (TMIN), Radiação Solar ($\text{MJ m}^2 \text{ dia}^{-1}$) e Umidade relativa (%) durante os anos de coleta de frutos de açaí, foram obtidos do banco de dados meteorológicos diários de Xavier et al. (2016), com GRID de resolução espacial de $0,25^\circ$. Battisti et al. (2018) e Battisti e Sentelhas (2019) avaliando o desempenho dos dados em GRID obtidos pela base de dados de Xavier, consideraram adequado para substituir os dados de superfície para simular a produtividade atingível de diferentes espécies em todo o Brasil.

Com base nos dados de UR, foi calculado o déficit de pressão de vapor (DPV) por meio das seguintes equações (1 a 3):

$$\text{DPV} = E_s - E_a \quad (1)$$

$$E_s = 0.6108 \times 10^{\left(\frac{7.5 \times \text{TMED}}{237.3 + \text{TMED}}\right)} \quad (2)$$

$$E_a = \frac{\text{UR} \times E_s}{100} \quad (3)$$

em que, DPV refere-se ao déficit de pressão de vapor (kPa); E_s corresponde pressão de saturação de vapor (kPa); E_a são os valores da pressão parcial de vapor (kPa); UR refere-se à Umidade relativa (%); TMED são os valores de Temperatura média do ar ($^\circ\text{C}$).

Seguindo o método de Camargo (1971), foi obtida a evapotranspiração potencial (equação 4), utilizando dados de temperatura média do ar e precipitação pluvial, realizado por meio de planilha eletrônica de Rolim et al., (1998).

$$\text{ETP} = F Q_0 T \text{ ND} \quad (4)$$

em que Q_0 (mm dia^{-1}) a radiação solar extraterrestre diária expressa em equivalente de evaporação, no período considerado, T ($^\circ\text{C}$) é a temperatura média do ar durante o período; F é o fator de ajuste que varia com a temperatura média anual do local (para T_m até 23°C , $F = 0,01$; $T_m = 24^\circ\text{C}$, $F = 0,0105$; T_m

= 25 °C, F = 0,011; Tm = 26 °C, F = 0,0115; Tm > 26 °C, F = 0,012); e ND é o número de dias do período.

Com os dados de ETP, foi realizado o balanço hídrico segundo Thornthwaite e Mather (1955), utilizando capacidades de água disponível (CAD) igual a 125 mm, equação (5 a10), obtendo posteriormente, os valores de Armazenamento de água (ARM) (mm), Deficiência hídrica (mm) (DEF) e Excedente hídrico (EXC) (mm) no solo.

$$\text{Se } (P - \text{ETP})_i < 0 = \begin{cases} \text{NAC}_i = \text{NAC}_{i-1} + (P - \text{ETP})_i \\ \text{ARM} = \text{CAD} e^{\frac{(\text{NAC})_i}{\text{CAD}}} \end{cases} \quad (5)$$

$$\text{Se } (P - \text{ETP})_i \geq 0 = \begin{cases} \text{NAC}_i = \text{CAD} \ln \frac{\text{ARM}_i}{\text{CAD}} \\ \text{ARM} = (P - \text{ETP})_i + \text{ARM}_{i-1} \end{cases} \quad (6)$$

$$\text{ALT}_i = \text{ARM}_i - \text{ARM}_{i-1} \quad (7)$$

$$\text{ETR}_i = \begin{cases} P + |\text{ALT}_i|, & \text{se } \text{ALT} < 0 \\ \text{ETP}_i, & \text{se } \text{ALT} \geq 0 \end{cases} \quad (8)$$

$$\text{DEF} = \text{ETP} - \text{ETR} \quad (9)$$

$$\text{EXC}_i = \begin{cases} 0, & \text{se } \text{CAD} < 0 \\ (P - \text{ETP})_i - \text{ALT}_i & \text{se } \text{CAD} = 0 \end{cases} \quad (10)$$

em que a ETP é a Evapotranspiração Potencial (mm); CAD refere-se a Capacidade de água disponível no solo (mm), ARM indica o armazenamento de água no solo (mm); NAC corresponde a Negativo acumulado que é resultado da subtração entre a precipitação acumulada e a evapotranspiração potencial; P é a precipitação pluvial (mm); DEF os valores de deficiência hídrica no sistema solo-planta-atmosfera (mm); ETR é a evapotranspiração real (mm), EXC é o excedente hídrico do sistema (mm); ALT refere-se ao armazenamento de água no solo no mês atual subtraído do armazenamento de água no solo do mês anterior (mm) e i indica o período mensal.

3.2.1 Caracterização fenológica do açaizeiro

O açaizeiro assim como outras palmeiras de interesse comercial, apresenta um ciclo reprodutivo indeterminando, com lançamento de inflorescência durante todo o ano. Entretanto, em períodos secos com manejo

do cultivo sem irrigação, o lançamento de novas inflorescências pode ser interrompido por deficiências hídricas acentuadas (Moraes et al. 2020a). Sendo assim, caracterizar o desenvolvimento reprodutivo do açaizeiro em sistemas irrigados e de sequeiro é essencial para avaliar a sazonalidade da produção de frutos da planta e promover melhor entendimento da relação solo-planta-atmosfera da cultura. Neste sentido, os modelos de previsão foram divididos em sistemas de sequeiro e irrigado, com previsões para o primeiro e segundo semestre em cada sistema, devido a variabilidade climática na região (Moraes et al., 2020b).

O ciclo reprodutivo do açaizeiro é de aproximadamente 24 meses, sendo a primeira fase a emissão de flexa ou lançamento foliar, dando início a formação dos primórdios florais, até a última fase de maturação dos frutos (Sousa, 2020; Oliveira, 2002) (Figura 3, 4).

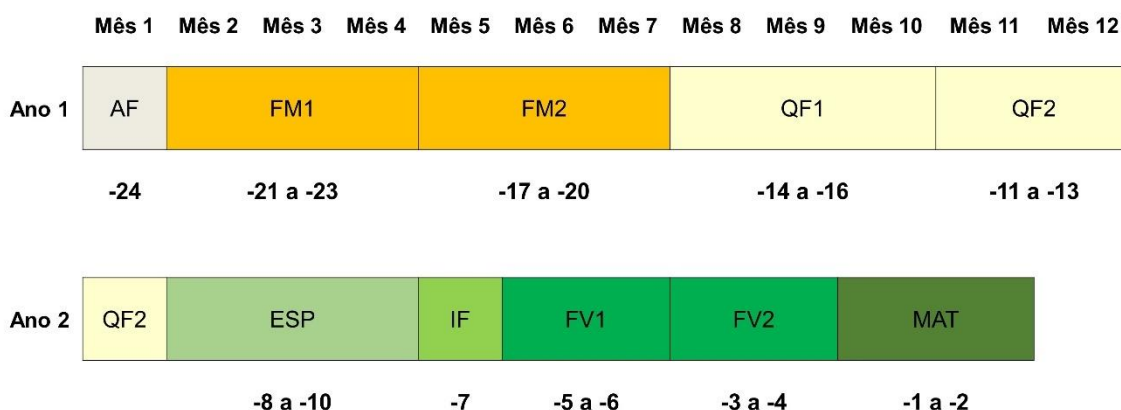


Figura 3. Desenvolvimento do ciclo reprodutivo do açaizeiro. AF (Emissão de flexa); FM1 (Folha madura 1); FM2 (Folha madura 2); QF1 (Folha senescente 1); QF2 (Folha senescente 2); ESP (Emissão de espata); IF (Inflorescência); FV1 (Fruto verde 1); FV2 (Fruto verde 2); MAT (Maturação).

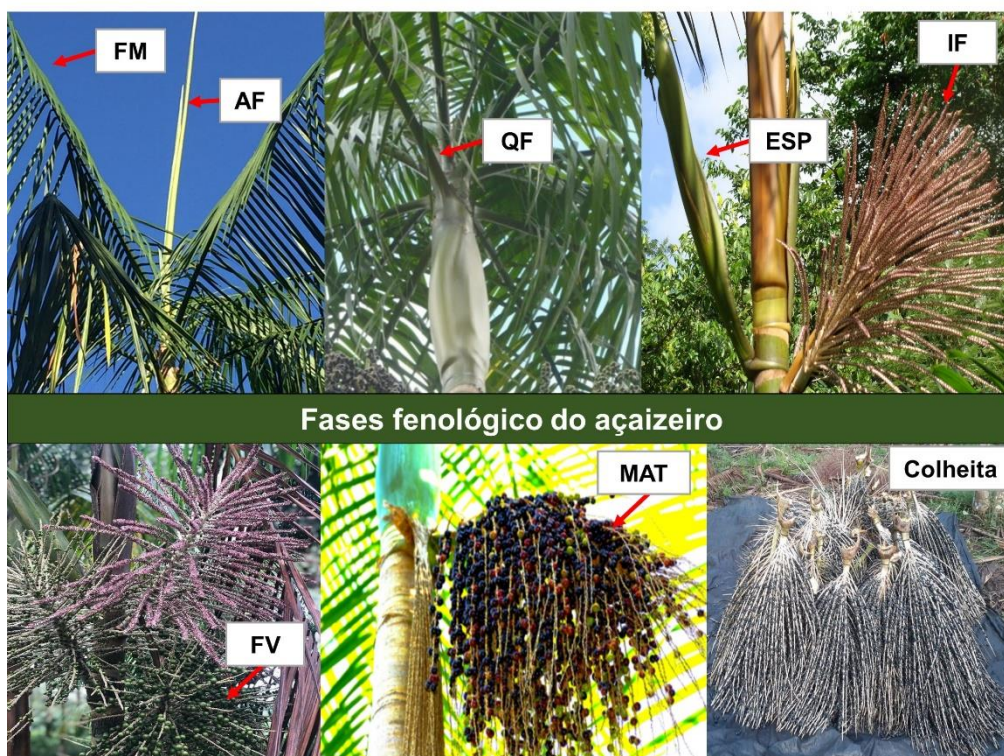


Figura 4. Fases de desenvolvimento reprodutivo do açaizeiro. AF (Emissão de flecha); FM (Folha madura); QF (Folha senescente); ESP (Emissão de espata); IF (Inflorescência); FV (Fruto verde); MAT (Maturação).

Do lançamento foliar até a emissão de espata são transcorridos em média ~13 meses. A inflorescência do açaizeiro abre após 75 dias da emissão da espata, colhendo frutos maduros em torno de 200 a 180 dias após a polinização. O período de lançamento foliar até a fase de emissão de espata é o ciclo mais longo da cultura. Esse período foi definido com base em experimentos de campo realizados nas dependências da Embrapa Amazônia Oriental, no município de Belém, Pará, entre os anos de 2017 a 2019. A coleta de dados da duração de uma folha de açaizeiro (indo de flecha até a senescência), foi acompanhado em 100 plantas.

Observou-se que do lançamento da flecha (AF) até sua abertura é descrito um ciclo de 25 dias, com desvio padrão (± 10 dias). Da abertura foliar a folha madura (FM) decorre em média 194 dias (± 62 dias) e da folha madura até a folha senescente (QF) 180 dias (± 58 dias) (Tabela 1). Logo, objetivando diminuir os intervalos de tempo para melhor caracterização do clima durante determinado

estádio fenológico, dividiu-se a fase de folha madura em dois períodos com 97 dias cada, e folha senescente também em dois intervalos com 90 dias cada (Figura 3). O mesmo processo foi realizado para formação de fruto verde (FV) que leva ao todo 130 dias, divididos em 2 intervalos de 65 dias.

Tabela 1. Desenvolvimento foliar do açaizeiro. AF (Emissão de flexa a abertura); FM (Folha madura), QF (Folha senescente).

Fase fenológica	Dias	Desvio Padrão
AF	25	10
FM	194	62
QF	180	58

O método de análise de componente principal (ACP) foi aplicado para a redução de dimensionamento e extração de variáveis entre o clima e a fenologia, criando assim uma matriz com um número menor de camadas, mas com a mesma variabilidade das informações originais (Salata; Grillenzoni, 2021). O primeiro fator extraído (ACP 1) é considerado a melhor combinação linear das variáveis explicativas da ACP, seguindo os demais fatores (ACP2...ACPN) a ordem de importância das combinações lineares.

3.2.2 Modelos de Machine learning

Modelos de Machine Learning (ML) são características muito úteis para modelar comportamentos complexos não lineares, tais como uma função para previsão de produtividade de safra (Sanchez; Solis; Bustamante, 2014). Entre os modelos de ML, foi aplicado para previsão de produtividade de frutos de açaí as técnicas de Regressão Linear Múltipla (RLM), Random Forest (RF), Redes Neurais Artificiais (MLP) e Support Vector Machine (SVM). Todos os modelos foram implementados no ambiente de linguagem de programação em Python.

A RLM é um dos algoritmos mais reconhecidos e usados em ML, pela sua fácil compreensão. A modelagem preditiva com RLM se preocupa principalmente em minimizar o erro de um modelo e fazer as previsões mais precisas possíveis (Kannangara et al., 2018), ajustando uma equação linear para minimizar a soma

residual dos quadrados entre os alvos observados no conjunto de dados e os alvos previstos pela aproximação linear (Kannangara et al., 2018).

O RF é um modelo de aprendizado supervisionado que ajusta várias árvores de decisão a partir de várias sub amostras do conjunto de dados, e usa uma árvore média para melhorar a precisão preditiva e controlar o ajuste excessivo (Hariharan, 2021; Breiman, 2001). Os múltiplos preditores das árvores de decisão são combinados para reduzir a variância da previsão e aumentar a precisão da mesma (Fouedjio, 2020).

A MLP são sistemas complexos de tratamento de dados amplamente conhecidas por resolver problemas multivariados por meio de classificação (Zhang et al., 2019) e regressão (Rahman; Zhang, 2018). Essa técnica aprende uma função de treinamento em um conjunto de dados para proporcionar uma ou mais saídas (Subasi, 2020).

O processo de configuração de qualquer rede neural é uma sequência de várias etapas que, em combinação, resultam em uma qualidade de predição específica da rede (Sack; Abom, 2020). Multilayer Perceptron (MLP) é um tipo de rede neural artificial que embora sejam as mais simples das arquiteturas de redes neurais, elas são aproximadores universais de funções (Cybenko; Cybenkot, 1989).

O SVM é outra técnica de ML, estabelecida por (Cortes e Vapnik, 1995), ativado como uma forma de classificador de margem máxima que tenta encontrar um hiperplano para o melhor ajuste entre os dados (Ayeleru et al., 2021). Vários núcleos estão disponíveis no SVM, como Função de Base Radial (RBF), modelo Linear, e regressão polinomial, sendo todos esses testados nas regressões preditivas.

Os parâmetros de treinamento para cada modelo e seus respectivos valores padrão, foram estimados pelo método gridsearchCV uma biblioteca para python. Foram separados 70% dos dados para construir os modelos e 30% para validação ou teste. Usando a biblioteca StatsModels da linguagem de programação Python, foi aplicado o Fator de Inflação de Variância para avaliar a multicolinearidade dos dados, evitando assim variáveis independentes

correlacionadas dentro dos modelos preditivos de ML (Forthofer; Lee; Hernandez, 2007; Miles, 2014).

Um dos principais fatores que governam o desempenho dos modelos de ML para capturar o máximo de informações nas variáveis preditivas brutas, bem como na relação entre as variáveis preditivas, são seus valores ajustados de hiperparâmetros. Cada modelo preditor nos diferentes períodos e manejo de cultivo são apresentados na tabela 2.

Tabela 2. Valores dos hiperparâmetros utilizados pelos modelos de Machine Learning entre as diferentes épocas de previsão.

Modelo	Hiperparâmetros	Sequeiro (1º Semestre)	Sequeiro (2º Semestre)	Irrigado (1º Semestre)	Irrigado (2º Semestre)
RF	max_depth	15	2	12	3
	min_impurity_decreas	1	1	1	1
	n_estimators	50	15	15	100
	random_state	0	0	0	0
MLP	hidden_layer_sizes	1,3,6	15,17,20	3,5,8	6,8,11
	learning_rate_init	0.1	0.1	0.1	0.1
	learning_rate	constant	constant	constant	constant
	activation	relu	relu	relu	relu
	solver	lbfgs	lbfgs	lbfgs	lbfgs
	alpha	0.01	0.3	0.01	0.3
	random_state	35	20	50	50
	max_iter	1000	1000	1000	1000
SVM_RBF	kernel	rbf	rbf	rbf	rbf
	C	100	100	200	800
	gamma	auto	auto	auto	auto
	epsilon	1	0.1	0.1	0.1
SVM_LINEAR	kernel	linear	linear	linear	linear
	C	200	7	9	200
	gamma	auto	auto	auto	auto
	epsilon	1	1	1	1
SVM_POLY	kernel	poly	poly	poly	poly
	C	10	8	1	200
	gamma	auto	auto	auto	auto
	epsilon	0.1	0.1	0.1	0.1
	coef0	2	2	2	2

As avaliações dos modelos de ML foram estimadas por meio do coeficiente de determinação (R^2), Erro absoluto médio (MAPE) e pelo erro médio (ME).

$$R^2 = \frac{\sum_{i=1}^n (\text{Yest} - \overline{\text{Yest}})^2}{\sum_{i=1}^n (\text{Yobs} - \overline{\text{Yest}})^2} \quad (11)$$

$$\text{MAPE} = \frac{\sum_{i=1}^n \left(\frac{|\text{Yobs} - \text{Yest}|}{\text{Yobs}} \right)}{n} * 100 \quad (12)$$

$$\text{ME} = \frac{1}{n} \sum_{i=1}^n (\text{Yest} - \text{Yobs}) \quad (13)$$

Yest é o valor estimado de y; Yobs é o valor observado de y; $\overline{\text{Yest}}$ é o valor médio estimado de y; n é número de dados.

3.3 RESULTADOS E DISCUSSÃO

A produção de açaí apresenta variações ao longo do ano, decorrente dos efeitos climáticos sazonais, principalmente quanto a oferta pluvial mensal (Figura 5). Observa-se que apesar da produção se estender durante todo o ano, no cultivo irrigado apresenta-se dois picos de maior produtividade, o primeiro de junho a julho e o segundo de novembro a dezembro. A produtividade média nesses meses de maior produtividade pode chegar a $\sim 2.319 \text{ kg ha}^{-1}$, com desvio padrão de $\sim 770 \text{ kg ha}^{-1}$. Em contrapartida, menores produtividades de frutos são observadas em dois intervalos, no primeiro semestre entre os meses de março a maio e no segundo semestre entre agosto a setembro, com valores de $\sim 664 \text{ kg ha}^{-1}$ no mês de setembro, e desvio padrão de $\sim 72 \text{ kg ha}^{-1}$. As produtividades médias anuais no cultivo irrigado foram de aproximadamente $13 \text{ mil kg ha}^{-1}$.

Esses resultados corroboram com as avaliações de produtividade realizadas por Farias Neto et al. (2020) na região do nordeste paraense, que visavam obter materiais genéticos que possibilitem diminuir a sazonalidade da produtividade de frutos de açaí, visando diminuir a segurança alimentar de toda a cadeia de produção da cultura maximizada pelos períodos de entressafra durante alguns meses do ano, gerando efeitos econômicos desfavoráveis (Nogueira et al., 2013; Nogueira; Santana, 2016).

Como esperado, os cultivos de açaí no sequeiro apresentam uma produtividade anual em torno de 400% a menos que nos cultivos irrigados. As médias anuais de produtividade no sequeiro foram de $\sim 2.880 \text{ kg ha}^{-1} \text{ ano}^{-1}$. Assim como no cultivo irrigado, maiores produtividades de frutos foram registradas nos meses de novembro e dezembro, variando de 542 a $770 \text{ kg ha}^{-1} \text{ mês}^{-1}$, respectivamente. Contudo, as menores médias foram principalmente entre os meses de fevereiro a abril e julho a setembro (Figura 5).

É importante destacar que os baixos valores de produção durante esses meses podem estar relacionados aos estágios de inflorescência e floração, uma vez que, durante os meses de fevereiro e março, a alta pluviosidade pode ter dificultado a visita dos insetos às flores, o que consequentemente impactaria a sua fertilização. Mukherjee et al. (2019) ressaltam que fatores ambientais abióticos, como as chuvas, afetam conjuntamente a produção e a disponibilidade de pólen e néctar e, portanto, influenciam o sucesso da polinização. Glenney et al. (2018) reforçam que Culturas dependentes de polinizadores, como o açaí, também podem ser comprometidas pela variabilidade climática devido a efeitos diretos na comunidade polinizadora.

Outro importante fator que pode influenciar nas quedas de produtividade são as temperaturas máximas na região. Em palmeiras como o coqueiro, aumentos de temperatura afetam negativamente a polinização por meio da capacidade de germinação do pólen (Hebbar et al., 2020) e na taxa de crescimento do tubo polínico (Hebbar et al., 2020; Snider et al., 2011). Na palma de óleo, outra importante palmeira na região amazônica, altas temperaturas presentes podem gerar a desidratação prematura das inflorescências masculinas, afetando significativamente a sobrevivência dos polinizadores, expressando em baixos pesos de cacho 5 meses depois (Corzo et al., 2021).

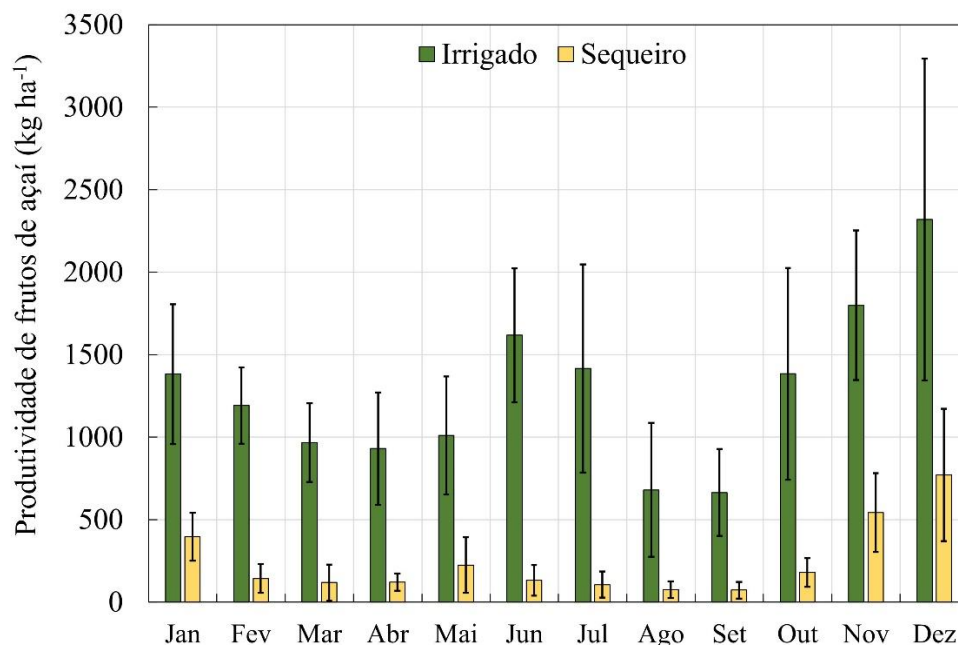


Figura 5. Produtividade de frutos de açaí irrigado e de sequeiro durante os meses do ano na área experimental da Embrapa Amazônia Oriental, na região de Tomé-Açu, Pará.

Entre os anos de 2004 a 2012, período de avaliação experimental com plantas de açaizeiro, observa-se que a variabilidade dos elementos climáticos na região de estudo (Figura 6), segue o padrão sazonal da Amazônia Oriental (Moraes et al., 2020b; Marengo et al. 2018). As maiores precipitações pluviiais são entre os meses de janeiro a maio, com máximas podendo chegar a valores acima de 600 mm mês⁻¹ e mínimas sempre acima de 100 mm mês⁻¹ (Figura 6.a). Esses resultados coincidem com os estudos de Wanzeler et al. (2018) para mesma região de trabalho.

Em média na região, os meses mais chuvosos estão concentrados entre março e abril, ambos com $P > 400$ mm, assim como também caracterizados como meses mais úmidos, com excedente hídrico que podem chegar a valores acima de 500 mm e médias superiores a 280 mm (Figura 6.c). As menores P são observadas entre agosto e setembro, com alguns anos apresentando mínimas abaixo de 10 mm e máximas que chegam a valores < 100 mm (Figura 6.a). As médias de P entre os meses menos chuvosos (agosto e setembro) estão entre

36 a 33 mm, respectivamente, com totais anuais de, aproximadamente 2.400 mm.

Agosto a outubro foram observados como períodos mais secos, com valores máximos que chegam a 94, 111 e 122 mm, respectivamente, e mínimas sempre acima de 10 mm (Figura 6.b). Essa caracterização é dada principalmente por menores P durante esses meses (Figura 6.a) e maiores temperaturas máximas (Tmax) que apresentam valores superiores a 34°C (Figura 6.f) e médias acima de 26°C. Essas variações também são observadas por Carvalho et al. (2020), que evidência temperaturas acima de 26°C mostradas na Amazônia oriental no estado do Pará, se espalhando para áreas na Amazônia central após o rio Amazonas.

Benezoli et al. (2021) também observaram as mesmas sazonalidades na região de estudo. Esses autores reforçaram que durante a estação chuvosa, a precipitação média é geralmente maior que 200 mm mês⁻¹ e na estação de menor pluviosidade (agosto e setembro) os acumulados de chuva geralmente são inferiores a 100 mm por mês⁻¹, com temperatura média anual de 26 °C, podendo ocorrer máximas superiores a 34°C entre setembro a novembro.

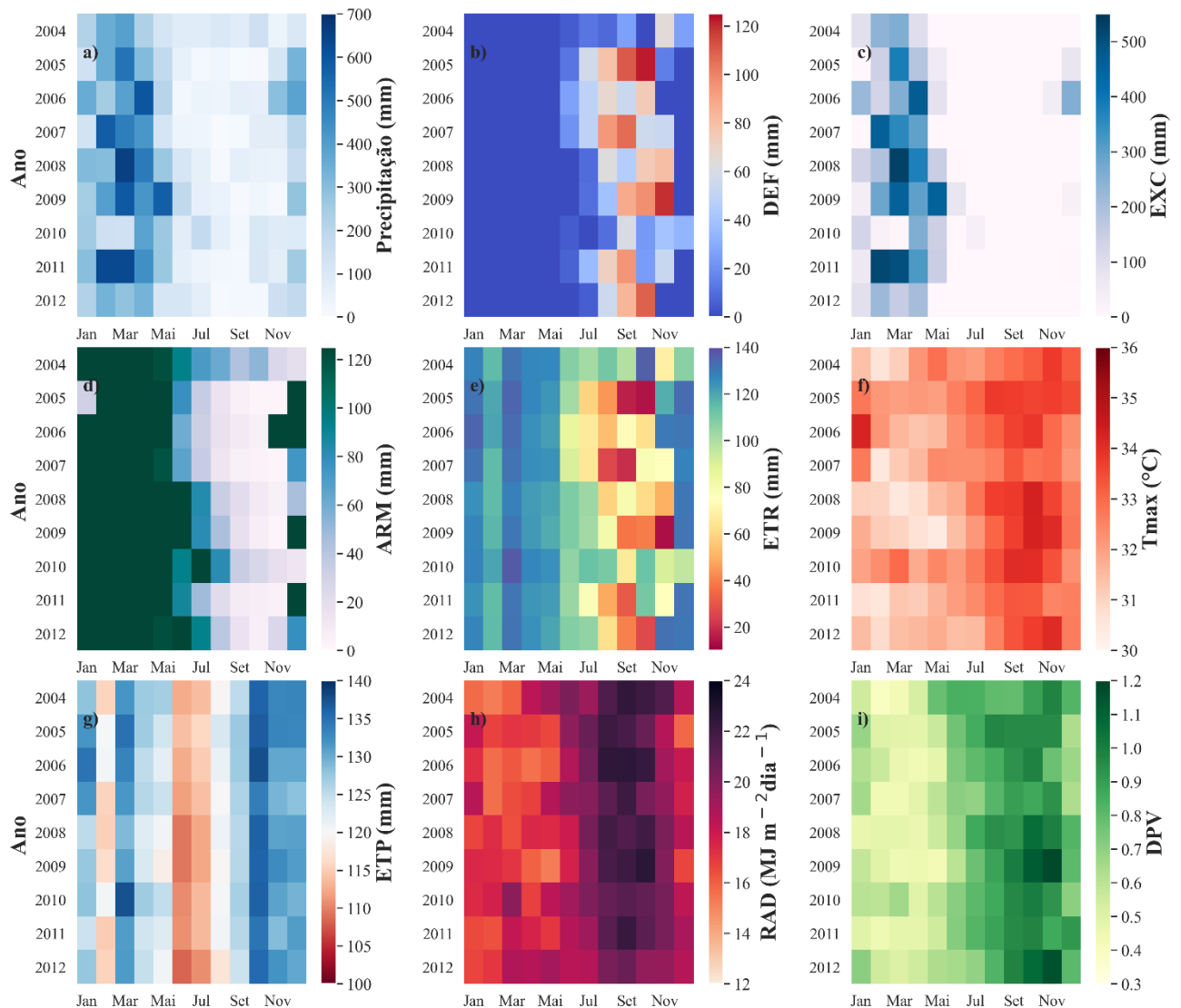


Figura 6. Mapa de calor das variáveis meteorológicas, a) Precipitação pluvial, b) Deficiência hídrica (DEF), c) Excedente hídrico (EXC), d) Armazenamento de água no solo (ARM), e) Evapotranspiração real (ETR), f) Temperatura máxima (TMAX), g) AWHC %, h) Radiação solar (RAD) e déficit de pressão de vapor (DPV), na área experimental de cultivo de açaizeiro em Tomé-Açu, Pará.

Com base na análise de componente principal, foram selecionadas 7 (sete) variáveis nas avaliações de açaí de sequeiro no primeiro semestre e 8 (oito) variáveis para açaí de sequeiro no segundo semestre, conforme critério de corte ($> 0,70$) considerado para interpretação do componente principal (Tabela 3).

Observa-se na figura 7.a, o gráfico biplot com os dois primeiros componentes principais (CP1 e CP2) para avaliações de sequeiro no primeiro semestre. O componente principal 1 (CP1) representa 71.38% da variância total dos dados e o segundo componente principal (CP2) 10.15% da variância.

A CP1 com exceção de DEF_FM2 é formada por todas as variáveis extraída da análise. Existe uma relação de dependência direta entre a produtividade (PROD) (-0.91), RAD_QF2 (-0.87) e EXC_IF (-0.79), com todos os valores projetados no mesmo plano bidimensional. Em contrapartida, TMAX_FM2 (0.81), DPV_IF (0.85), DPV_FV1 (0.89) e ETR_MAT (0.94), apresentam uma relação de dependência indireta com a PROD, localizada na direção oposta no plano bidimensional do gráfico (Figura 7.a, Tabela 3).

Nos cultivos de sequeiro no segundo semestre, a CP1 representa (69.79 %) da variância total dos dados, enquanto a CP2 contribui com (15.5%) (Figura 7.b, tabela 3). Assim como nas observações no primeiro semestre, a CP2 foi formada somente pela DEF_FM2 (0.77). Na CP1 a PROD (-0.83) apresenta relação de dependência direta com EXC_FM1 (-0.93), EXC_IF (-0.78) e RAD_QF2 (-0.75). No entanto, TMAX_FM1 (0.86), TMAX_IF (0.93), DPV_IF (0.96) e ETR_QF1 (0.88), apresentam relação de dependência indireta a PROD (Figura 7.b, tabela 3).

É importante observar que RAD_QF2 e EXC_IF apresentaram dependência direta com a produtividade nos dois períodos diferentes de análise, primeiro e segundo semestre, assim como, a CP2 foi formada somente pelo DEF_FM2.

Yokomizo et al., (2020) avaliando o comportamento das características fenológicas do açaí, visando determinar quais fases apresentam maior contribuição nas manifestações fenotípicas e aquelas com maior potencial a serem utilizados para seleção de genótipos, também observaram a formação de dois grupos distintos, um envolvendo somente folhas maduras e o outro com as demais características.

O DPV_IF e TMAX durante as fases de folha madura, representaram dependência indireta nos dois períodos de análise. Ou seja, quanto maior o valor

dessas variáveis durante esses estádios fenológicos da planta, menor será a produtividade e vice-versa.

Marenco et al., 2014 relatam que durante o dia a condutância estomática responde rapidamente à diferença de pressão de vapor entre a folha e o ar e, em geral, os estômatos se fecham à medida que a DPV aumenta (Dominguez et al., 2019), resultando um efeito direto na fotossíntese.

Outro fator relevante é que em plantas C3, como o açaí, as taxas de fotorrespiração podem ser tão dinâmicas quanto as da fotossíntese, sendo alteradas pela luz, concentração de dióxido de carbono (CO₂) e temperatura do ar (Wu et al., 2019). Sendo assim, as taxas de fotorrespiração aumentam com o aumento da intensidade da luz e da temperatura (Zhou et al., 2016). Nessa situação, a planta reduzirá a fotossíntese e aumentará a fotorrespiração, o que poderá ter um efeito direto e significativo no desenvolvimento da planta (Walker et al., 2016).

Tabela 3. Análise de componente principal das variáveis climáticas e fenológicas do açaizeiro em manejo de sequeiro, com meses de produção divididos entre o primeiro e segundo semestre do ano.

	Variáveis	CP1	CP2
1° Semestre Sequeiro	PROD	-0.91	-0.05
	RAD_QF2	-0.87	0.29
	TMAX_FM2	0.81	0.19
	DPV_IF	0.85	0.35
	DPV_FV1	0.89	-0.13
	DEF_FM2	0.59	-0.71
	EXC_IF	-0.79	-0.18
	ETR_MAT	0.94	0.21
	Variância explicada (%)	71.38	10.15
2° Semestre Sequeiro	PROD	-0.83	-0.34
	TMAX_FM1	0.86	-0.44
	TMAX_IF	0.93	-0.21
	DPV_IF	0.96	-0.18
	EXC_FM1	-0.93	0.25
	EXC_IF	-0.78	-0.25
	ETR_QF1	0.88	-0.27
	RAD_QF2	-0.75	-0.46
	DEF_FM2	0.52	0.77
	Variância explicada (%)	69.79	15.4

Produtividade (PROD); Radiação na fase de folha senescente 2 (RAD_QF2); Temperatura máxima na fase de folha madura 2 (TMAX_FM2); Déficit de pressão de vapor na inflorescência (DPV_IF); Déficit de pressão de vapor na fase de fruto verde 1 (DPV_FV1); Deficiência hídrica

na fase de folha madura 2 (DEF_FM2); Excedente hídrico na inflorescência (EXC_IF); Evapotranspiração Real na maturação (ETR_MAT); Temperatura máxima em folha madura 1 (TMAX_FM1); Excedente hídrico na fase de folha madura 1 (EXC_FM1).

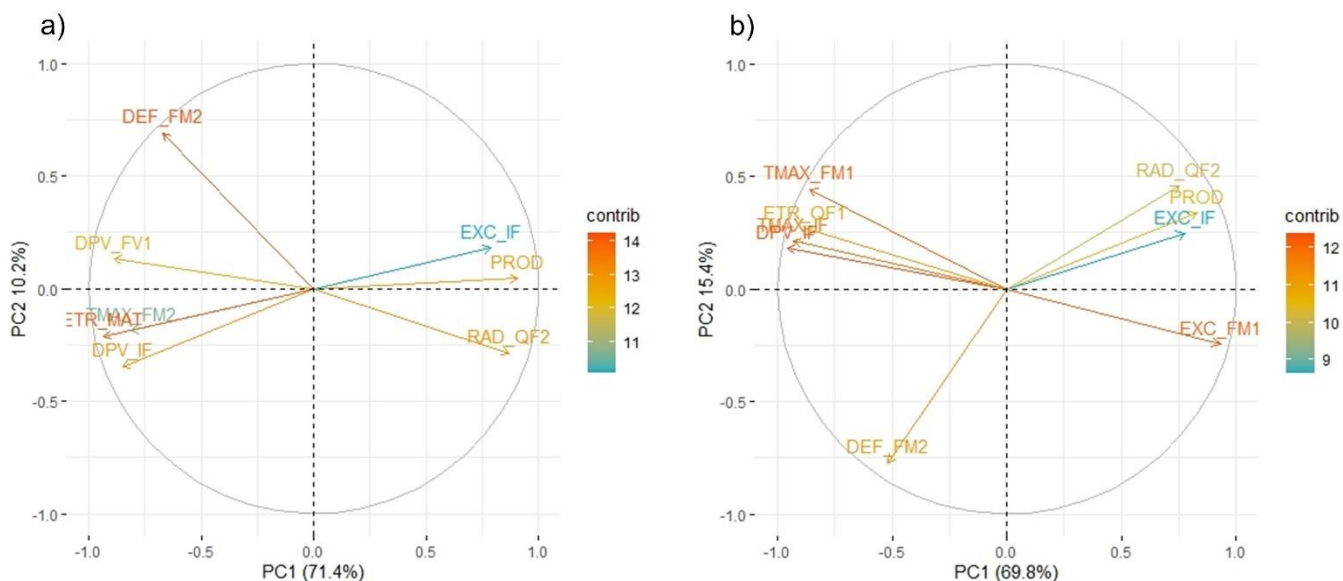


Figura 7. Representação biplot em contendo as variáveis climáticas acumuladas entre os estádios fenológicos do açaizeiro. a) Açaí de sequeiro com meses de produção no 1º semestre; b) Açaí de sequeiro com meses de produção no 2º semestre.

Produtividade (PROD); Radiação na fase de folha senescente 2 (RAD_QF2); Temperatura máxima na fase de folha madura 2 (TMAX_FM2); Déficit de pressão de vapor na inflorescência (DPV_IF); Déficit de pressão de vapor na fase de fruto verde 1 (DPV_FV1); Deficiência hídrica na fase de folha madura 2 (DEF_FM2); Excedente hídrico na inflorescência (EXC_IF); Evapotranspiração Real na maturação (ETR_MAT); Temperatura máxima em folha madura 1 (TMAX_FM1); Excedente hídrico na fase de folha madura 1 (EXC_FM1).

No cultivo irrigado nas análises de produção mensal do primeiro semestre, a CP1 representa (54.08 %) da variância total dos dados, enquanto a CP2 com (22.08%) (Figura 8.a, tabela 4). A CP1 é formada pela PROD (-0.84), ETP_ESP (-0.88), ETP_FM1 (-0.88), apresentando essas variáveis relação de dependência indireta com a RAD_FV1 (0.70), TMAX_FM2 (0.75), TMAX_IF (0.71) e TMAX_FV1 (0.90). Assim como na produção de açaí no sequeiro, a CP2 é formada somente pelo estágio fenológico FM2.

Nos meses de produção do segundo semestre, 73.87% da variância dos dados representa a CP1, enquanto a CP2 apresenta 8.46% da variância (Figura 8.b, tabela 4). Existe uma relação direta na CP1 entre PROD (0.89), PIR_AF (0.85), PIR_FM1 (0.82), PIR_QF2 (0.85) e DPV_QF1 (0.92), com projeções na

mesma direção no plano bidimensional (Figura 8.b). Entretanto, ETP_FM2 (-0.78) e ETP_ESP (-0.90) apresentam relação de dependência indireta a PROD, com direção oposta na CP1 do plano bidimensional (Figura 8.b).

Tabela 4. Análise de componente principal das variáveis climáticas e fenológicas do açaizeiro em manejo irrigado, com meses de produção divididos entre o primeiro e segundo semestre do ano.

	Variáveis	CP1	CP2
1° Semestre Irigado	PROD	-0.84	-0.24
	PIR_FM2	-0.19	-0.90
	ETP_ESP	-0.88	0.19
	RAD_FV1	0.70	0.65
	TMAX_FM2	0.75	-0.35
	TMAX_IF	0.71	-0.63
	TMAX_FV1	0.90	0.15
	PIR_IF	-0.61	-0.18
	ETP_FM1	-0.88	0.28
	Variância explicada (%)	54.08	22.08
2° Semestre Irigado	PROD	0.89	0.35
	PIR_AF	0.85	-0.26
	PIR_FM1	0.82	-0.33
	PIR_QF2	0.85	0.05
	ETP_FM2	-0.78	0.39
	ETP_ESP	-0.9	-0.32
	DPV_QF1	0.92	0.17
	Variância explicada (%)	73.87	8.46

Produtividade (PROD); Precipitação + irrigação na fase de folha madura 2 (PIR_FM2); Evapotranspiração potencial na emissão de espata (ETP_ESP); Radiação na fase de fruto verde 1 (RAD_FV1); Temperatura máxima na fase de folha madura 2 (TMAX_FM2); Temperatura máxima na inflorescência (TMAX_IF); Temperatura máxima na fase de fruto verde 1 (TMAX_FV1); Precipitação + irrigação na inflorescência (PIR_IF); Evapotranspiração potencial na fase de folha madura 1 (ETP_FM1); Precipitação + irrigação na emissão de flexa (PIR_AF); Precipitação + irrigação na fase de folha madura 1 (PIR_FM1); Precipitação + irrigação na fase de folha senescente (PIR_QF2); Evapotranspiração potencial na fase de folha madura 2 (ETP_FM2); Déficit de pressão de vapor na fase de folha senescente 1 (DPV_QF1).

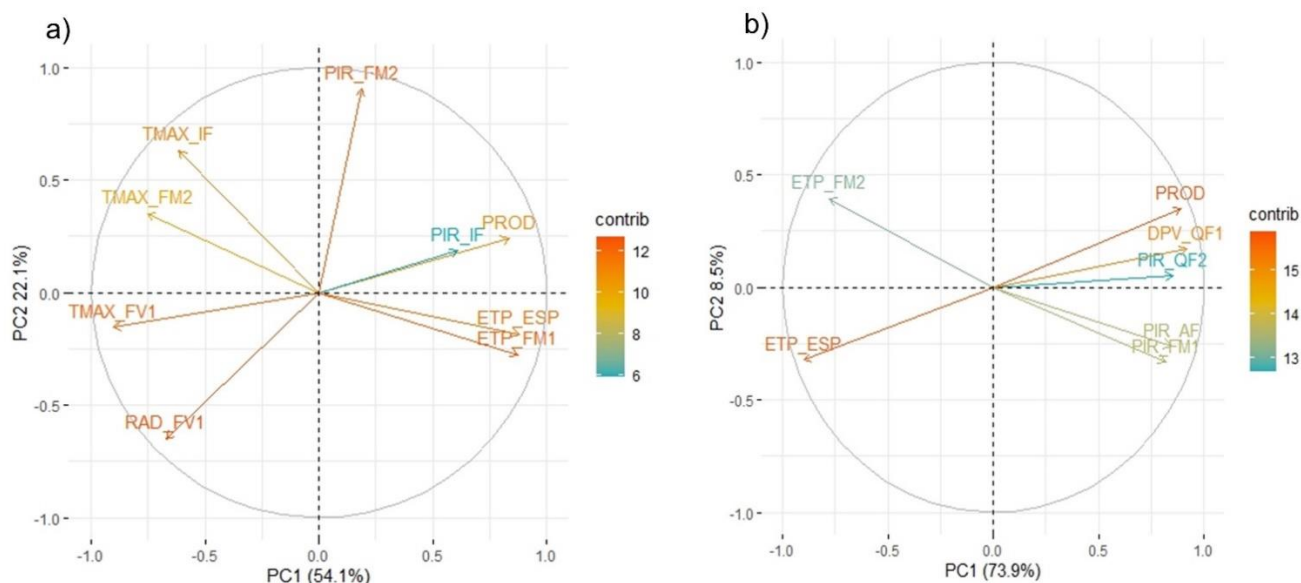


Figura 8. Representação em biplot contendo as variáveis climáticas acumuladas entre os estádios fenológicos do açaizeiro. a) Açaí irrigado com meses de produção no 1º semestre; b) Açaí irrigado com meses de produção no 2º semestre.

Produtividade (PROD); Precipitação + irrigação na fase de folha madura 2 (PIR_FM2); Evapotranspiração potencial na emissão de espata (ETP_ESP); Radiação na fase de fruto verde 1 (RAD_FV1); Temperatura máxima na fase de folha madura 2 (TMAX_FM2); Temperatura máxima na inflorescência (TMAX_IF); Temperatura máxima na fase de fruto verde 1 (TMAX_FV1); Precipitação + irrigação na inflorescência (PIR_IF); Evapotranspiração potencial na fase de folha madura 1 (ETP_FM1); Precipitação + irrigação na emissão de flexa (PIR_AF); Precipitação + irrigação na fase de folha madura 1 (PIR_FM1); Precipitação + irrigação na fase de folha senescente (PIR_QF2); Evapotranspiração potencial na fase de folha madura 2 (ETP_FM2); Déficit de pressão de vapor na fase de folha senescente 1 (DPV_QF1).

Entre os modelos de ML para previsão de produtividade para o açaí, em sistema de cultivo de sequeiro e irrigado com colheitas no primeiro e no segundo semestre (Tabela 5). Os resultados apontaram que o Random Forest (RF) e Support Vector Machine Linear (SVM_Linear), apresentam as menores porcentagem de erros e maiores coeficientes de determinação, com previsões de produtividade até 6 meses antes da colheita, para cultivo de sequeiro em colheitas no primeiro semestre. O RF e SVM_Linear obtiveram métricas de acurácia de (MAPE = 13.81%, 18.32%), Precisão ($R^2 = 0.84, 0.80$), e tendência (ME = -5.41, 11.86 kg ha⁻¹), respectivamente. Em contrapartida, o Support Vector

Machine Poly (SVM_Poly) apresentou os menores desempenhos, com $R^2 < 0.40$, MAPE de $\sim 30\%$ e $ME > 70 \text{ kg ha}^{-1}$.

Nas previsões para o cultivo de sequeiro com colheita no segundo semestre, também foram obtidas previsões com até 6 (seis) meses de antecedência. No entanto, com exceção da Regressão Linear Múltipla (RLM), todos os demais modelos de previsão apresentaram ($R^2 > 0.85$) e ($MAPE < 25\%$). No geral o RF obteve o melhor desempenho, com ($R^2 = 0.94$), ($MAPE = 18.35$) e ($ME = -10.3 \text{ kg ha}^{-1}$). É importante destacar que pela medida de tendência obtidas pelo ME, o RF tende em média a subestimar os valores observados.

No manejo irrigado foi possível obter previsões com até 4 meses antes da colheita no primeiro semestre e 7 meses no segundo semestre. No primeiro semestre os menores erros obtidos foram de ($MAPE = 14.14, 14.74$ e 15.46%) para SVM_Poly, RF e MLP, respectivamente, com $R^2 > 0.86$ e $ME > -135 \text{ kg ha}^{-1}$. No segundo semestre, todos os modelos apresentam precisões ($R^2 \geq 0.89$), com destaque para MLP, SVM_Linear, RF e SVM_Poly com $MAPE < 18\%$. No manejo irrigado, todos os modelos tendem também a subestimar a produtividade observada, com ME variando de -16 kg ha^{-1} (MLP) a -184 kg ha^{-1} (SVM_RBF).

No geral o RF ofereceu a melhor precisão nos prognósticos de produtividade do açaí entre os algoritmos testados. O RF obteve erros médios de 15% e R^2 de 0.90 , com valores em média subestimando os dados observado ($ME = -54.78$). Em contrapartida, a RLM obteve os menores desempenhos, com erros médios $> 100\%$ e $R^2 = 0.53$.

Pant et al. (2021), também obtiveram melhores desempenho para previsão de culturas anuais como milho, trigo e arroz na Índia, utilizando o algoritmo de RF. Os resultados também corroboram com os Li et al. (2020) que encontraram altos desempenhos na previsão da transpiração de plantas, com base em RF. As aplicações de RF foram promissoras, conforme Rakhra et al., (2021), com excelentes resultados para previsão de preço de safra agrícola, pois os modelos apontaram alto desempenho ao analisar situações complexas com reduzidas variáveis de entrada.

Estudos anteriores (Shahhosseini et al., 2019; Cai et al., 2019; Jeong et al., 2016) demonstraram que modelos de ML podem desempenhar um papel

importante na previsão de produtividade de safra, confirmando com os resultados obtidos para o açaí, nesta pesquisa. Da mesma forma, ML tem o potencial para prever a produção, com base em dados de campo, modelos de crescimento de culturas e sensoriamento remoto, reduzindo-se os erros nas estimativas.

Tabela 5. Valores de precisão estimada pelo coeficiente de determinação (R^2) e acurácia pelo Erro quadrado médio (ME) e MAPE (%), para previsão de produtividade de açaí sequeiro no primeiro semestre em diferentes fases de desenvolvimento da planta utilizando modelos de Machine Learning.

Períodos	Modelos	R^2	MAPE (%)	ME (kg ha ⁻¹)
1° Semestre Sequeiro (Previsão com 6 meses)	RLM	0.74	14.75	14.11
	RF	0.84	13.81	-5.41
	MLP	0.60	28.78	15.26
	SVM RBF	0.36	24.41	-60.66
	SVM Linear	0.8	18.32	11.86
	SVM Poly	0.38	29.73	-71.39
2° Semestre Sequeiro (Previsão com 6 meses)	RLM	0.29	476.62	349.46
	RF	0.94	18.35	-10.34
	MLP	0.81	20.17	-42.51
	SVM RBF	0.92	23.88	-29.94
	SVM Linear	0.92	19.34	-16.64
	SVM Poly	0.9	27.54	-29.42
1° Semestre irrigado (Previsão com 4 meses)	RLM	0.21	29.11	-208.7
	RF	0.86	14.74	-162.81
	MLP	0.91	15.46	-179.66
	SVM RBF	0.24	28.31	-197.99
	SVM Linear	0.93	27.67	-108.29
	SVM Poly	0.91	14.14	-135.1
2° Semestre irrigado (Previsão com 7 meses)	RLM	0.89	25.93	-179.94
	RF	0.97	15.68	-40.59
	MLP	0.93	15.64	-16.09
	SVM RBF	0.94	18.93	-184.28
	SVM Linear	0.89	15.54	-159.03
	SVM Poly	0.95	17.68	-72.32

Regressão Linear Múltipla (RLM); Random Forest (RF); Redes Neurais Artificiais (MLP); Support Vector Machine Kernel RBF (SVM_RBF); Support Vector Machine Kernel Linear (SVM_Linear); Support Vector Machine Kernel Poly (SVM_Poly).

3.4 CONCLUSÃO

O algoritmo RF é recomendado na modelagem estatística pelos altos desempenhos nas métricas de precisão e acurácia. Contudo, necessidade do aumento do número de amostras para melhorar a capacidade preditiva dos modelos.

Os prognósticos de produtividade do açaizeiro apresentam desempenho com antecedência de 6 meses em cultivo de sequeiro. Sendo que os modelos apontaram para os cultivos irrigados previsão de produtividade com 4 meses de antecipação no primeiro semestre e 7 meses no segundo semestre.

Modelos de ML são eficazes e versáteis com alto potencial de aplicação em prognósticos de produtividade, auxiliando agricultores e planejadores agrícolas na tomada de decisão.

3.5 REFERÊNCIAS

Aparecido LEO, Rolim GS, Moraes JRSC, Costa CT, Souza PS (2020) Machine learning algorithms for forecasting the incidence of *Coffea arabica* pests and diseases. *International Journal of Biometeorology*, 64:671–688. DOI:<https://doi.org/10.1007/s00484-019-01856-1>.

Ashburner GR (1995) Reproductive biology of coconut palms. C. Oropeza et al (eds.), *Lethal Yellowing: Research and Practical Aspects*, p. 111-121.

Ayeleru OO, Fajimi LI, Oboirien BO, Olubambi PA (2021) Forecasting municipal solid waste quantity using artificial neural network and supported vector machine techniques: A case study of Johannesburg, South Africa. *Journal of Cleaner Production* 289, 125671. DOI:<https://doi.org/10.1016/j.jclepro.2020.125671>.

Barbosa PO, Souza MO, Silva MPS, Santos GT, Silva ME, Bermano G, Freitas RN (2021) Açaí (*Euterpe oleracea* Martius) supplementation improves oxidative stress biomarkers in liver tissue of dams fed a high-fat diet and increases antioxidant enzymes' gene expression in offspring. *Biomedicine &*

Pharmacotherapy, 139, 111627. DOI: <https://doi.org/10.1016/j.biopha.2021.111627>.

Battisti R, Sentelhas PC, Boote KJ (2017) Inter-comparison of performance of soybean crop simulation models and their ensemble in southern Brazil. *Field Crops Research*, 200, 28-37. DOI: <https://doi.org/10.1016/j.fcr.2016.10.004>

Battisti R, Bender FD, Sentelhas PC (2018) Assessment of different gridded weather data for soybean yield simulations in Brazil. *Theor. Appl. Climatol.* 1–11. DOI: <https://doi.org/10.1007/s00704-018-2383-y>.

Battisti R, Sentelhas PC (2019) Characterizing Brazilian soybean-growing regions by water deficit patterns. *F. Crop. Res.* 240, 95–105. DOI: <https://doi.org/10.1016/J.FCR.2019.06.007>.

Benezoli VH, Imbuzeiro HMA, Cuadra SV, Colmanetti, MAA, Araújo, AC, Stiegler C, Motoike, SY (2021) Modeling oil palm crop for Brazilian climate conditions. *Agricultural Systems*. 190: 103130. DOI: <https://doi.org/10.1016/j.agsy.2021.103130>.

Breiman L (2001) Random forests. *Mach. Learn.* 45, 5–32.

Cai Y, Guan K, Lobell D, Potgieter AB, Wang S, Peng J, Xu T, Asseng S, Zhang Y, You L (2019) Integrating satellite and climate data to predict wheat yield in Australia using machine learning approaches. *Agric. For. Meteorol.* 274, 144–159.

Camargo AP (1971) Balanço hídrico no Estado de São Paulo. Campinas: Instituto Agronômico, 24p. Boletim 116.

Chipanshi A, Zhang Y, Kouadio L, Newlands N, Davidson A, Hill H, Warren R, Qian B, Daneshfar B, Bedard F (2015) Evaluation of the integrated Canadian

crop yield forecaster (ICCYF) model for in-season prediction of crop yield across the Canadian agricultural landscape. *Agri- cultural and Forest Meteorology* 206, 137–150 . DOI: <https://doi.org/10.1016/j.agrformet.2015.03.007>.

Chlingaryan A, Sukkarieh S, Whelan B (2018). Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review. *Computers and Electronics in Agriculture*, 151, 61–69.

CONAB (2019) Açaí (Fruto). (Org): Serra FR, Santos FD. Boletim técnico de análise mensal, Brasília, DF, 1-9.

Cortes C, Vapnik V (1995) Support-vector networks. *Machine learning*, 20, 273–297. DOI: <https://doi.org/10.1023/A:1022627411411>.

Corzo JM, Juracán R, Ical B, López LM (2021) Factores importantes en la formación y conformación del racimo en el cultivo de palma de aceite. *Boletín La Palma. Gremial de palmicultura de Guatemala*, v.18, p.1-12.

Culman M, Farias CM, Bayona C, Cruz JSC (2019) Using agrometeorological data to assist irrigation management in oil palm crops: A decision support method and results from crop model simulation. *Agricultural Water Management*, 213, 1047–1062. DOI: <https://doi.org/10.1016/j.agwat.2018.09.052>.

Cybenko G, Cybenkot G (1989) Approximation by superpositions of a sigmoidal function, *Math. Control Signals Syst.*, 2 (4), 303–314. DOI: <https://doi.org/10.1007/BF02551274>.

Dominguez CMR, Santana VH, Buckley TN and Fernández JE, Sensitivity of olive leaf turgor to air vapour pressure deficit correlates with diurnal maximum stomatal conductance. *Agric For Meteorol* 272–273:156–165 (2019).

Dourado-Neto D, Teruel DA, Reichardt K, Nielsen DR, Frizzzone JÁ, Bacchi OOS (1998) Principles of crop modeling and simulation: i. uses of mathematical models in agricultural Science. *Sci. agric.*, 55, 46-50.

Du KL, Swamy M (2014) *Neural Networks and Statistical Learning*. Springer-Verlag, London, 988p. DOI: <https://doi.org/10.1007/978-1-4471-5571-3>.

Farias Neto JT, Yokomizo GKI, Resende MDV (2020) Productive capacity and genetic variation behavior in progênies from irrigated açaí according to plant age. *Revista Agro@mbiente On-line*, 14, 1-14. DOI: <http://dx.doi.org/10.18227/1982-8470ragro.v14i0.6409>.

Ferreira JMS, Warwick DRN, Siqueira LA (2018). *A cultura do coqueiro no Brasil*. Embrapa, Brasília, DF, p. 508.

Forthofer RN, Lee ES, Hernandez M (2007) *Linear Regression*. Biostatistics (Second Edition), p. 349-386.

Fouedjio F (2020) Exact Conditioning of Regression Random Forest for Spatial Prediction. *Artificial Intelligence in Geosciences*, 1, 11–23. DOI: <https://doi.org/10.1016/j.aiig.2021.01.001>.

Glenny WR, Runyon JB and Burkle L (2018) Drought and increased CO₂ alter floral visual and olfactory traits with context-dependent effects on pollinator visitation . *New Phytol* 220:785–798.

Sanchez GA, Frausto-Solis J, Ojeda-Bustamante W (2014) Predictive ability of machine learning methods for massive crop yield prediction. *Spanish J. Agric. Res.* 12 (2), 313–328. DOI: <https://doi.org/10.5424/sjar/2014122-4439>.

Hariharan R (2021). Random forest regression analysis on combined role of meteorological indicators in disease dissemination in an Indian city: A case study

of New Delhi. Urban Climate, 36, 100780. DOI: <https://doi.org/10.1016/j.uclim.2021.100780>.

Hebbar KB, Neethu P, Sukumar PA, Sujithra M, Santhosh A, Ramesh SV, Niral V, Hareesh GS, Nameer PO, Prasad PVV (2020) Understanding Physiology and Impacts of High Temperature Stress on the Progametic Phase of Coconut (*Cocos nucifera* L.). *Plants*, 9, 1651. DOI: <https://doi.org/10.3390/plants9121651>.

IBGE (2021). Instituto Brasileiro de Geografia e estatística. Sistema IBGE de Recuperação Automática (Accessed 10 January 2021). <https://sidra.ibge.gov.br/home/pms/brasil>.

Insfran DDP, Brenes CH, Talcott ST (2004) Phytochemical composition and pigment stability of Acai (*Euterpe oleracea* Mart.), *J. Agric. Food Chem.* 52, 1539–1545.

Guerra JF, Magalhaes CL, Costa DC, Silva ME, Pedrosa ML (2011) Dietary acai modulates ROS production by neutrophils and gene expression of liver antioxidant enzymes in rats, *J. Clin. Biochem. Nutr.* 49, 188–194.

Jeong JH, Resop JP, Mueller ND, Fleisher DH, Yun K, Butler EE, Timlin DJ, Shim KM, Gerber JS, Reddy VR, Kim SH (2016) Random Forests for Global and Regional Crop Yield Predictions. *PLoS ONE* 11(6): e0156571. DOI: <https://doi.org/10.1371/journal.pone.0156571>.

Jiang H, Hu H, Zhong R, Xu J, Xu J, Huang J, Lin T (2020) A deep learning approach to conflating heterogeneous geospatial data for corn yield estimation: a case study of the US Corn Belt at the county level. *Glob. Change Biol.* 26 (3), 1754–1766.

Kannangara M, Dua R, Ahmadi L, Bensebaa F (2018) Modeling and prediction of regional municipal solid waste generation and diversion in Canada using machine learning approaches. *Waste Manag.* 74, 3e15.

Klompenburg TV, Kassahun A, Catal C (2020). Crop yield prediction using machine learning: A systematic literature review. *Computers and Electronics in Agriculture*, 177, 105709. DOI: <https://doi.org/10.1016/j.compag.2020.105709>.

Kumar S, Kumar V, Sharma RK (2019) Rice Yield Forecasting Using Support Vector Machine. *International Journal of Recent Technology and Engineering*, 8, 1-6. DOI:10.35940/ijrte.D7236.118419.

Li I, Chen S, Yang C, Meng F, Sigrimis N (2020) Prediction of plant transpiration from environmental parameters and relative leaf area index using the random forest regression algorithm. *Journal of Cleaner Production*, 261, 121136. DOI: <https://doi.org/10.1016/j.jclepro.2020.121136>.

Lopes E, Soares-Filho B, Souza F, Rajão R, Merryd F and Carvalho Ribeiro S (2019) Mapping the socio-ecology of non timber forest products (NTFP) extraction in the Brazilian Amazon: the case of açaí (Euterpe precatoria Mart) in acre. *Landscape and Urban Plann* 188:110–117. DOI: <https://doi.org/10.1016/j.landurbplan.2018.08.025>.

Manimekalai S, K. Nandhini (2018) Crop Yield Prediction from Soil Parameters through Neupper Rule Established Algorithm, *Int. J. Eng. Technol.* 7 (3.34), 908–912.

Manjunathan N, Rajesh P, Thangadurai E, Suresh A (2020) Crop Yield Prediction Using Linear Support Vector Machine. *European Journal of Molecular & Clinical Medicine*, 6, 1-7.

Martorano LG, Pereira LC, Cezar EGM, Pereira ICB (1993) Estudos climáticos do Estado do Pará, classificação climática (Köppen) e deficiência hídrica (Thorntwhite, Matter). Belém, SUDAM/EMBRAPA, SNLCS pp. 53.

Miles J (2014) Tolerance and Variance Inflation Factor. Wiley StatsRef: Statistics Reference Online.

Moraes JRSC, Rolim GS, Martorano LG, Aparecido LEO, Bispo RC, Valeriano TTB (2020b) Performance of the ECMWF in air temperature and precipitation estimates in the Brazilian Amazon. *Theoretical and Applied Climatology*, 141:803–816. DOI: <http://dx.doi.org/10.1007/s00704-020-03231-2>.

Moraes JRSC, Rolim GS, Martorano LG, Aparecido LEO, Oliveira MSP, Neto JTF (2020a) Agrometeorological models to forecast açaí (*Euterpe oleracea* Mart.) yield in the Eastern Amazon. *J Sci Food Agric*, 100: 1558–1569. DOI: <http://dx.doi.org/10.1002/jsfa.10164>.

Mukherjee R, Deb R, Devy SM (2019) Diversity matters: effects of density compensation in pollination service during rainfall shift. *Ecol Evol* 9:9701–9711.

Nayamuth BFC (199) Crop Modelling / Simulation : An Overview. AMAS: Food and Agricultural Research Council, Réduit, Mauritius, 1-16.

Nogueira AKM, Santana AC (2016) Benefícios socioeconômicos da adoção de novas tecnologias no cultivo do açaí no Estado do Pará. *Revista Ceres*, 63, 1–7.

Nogueira AKM, Santana AC, Garcia WS (2013) A dinâmica do mercado de açaí fruto no Estado do Pará: de 1994 a 2009. *Revista Ceres*, 60, 324– 331.

Oliveira MSP (2002) Biologia floral do açaizeiro em Belém, PA. Embrapa Amazônia Oriental. *Boletim de Pesquisa e Desenvolvimento*, 26p.

Pant J, Pant RP, Singh MK, Singh DP, Pant H (2021) Analysis of agricultural crop yield prediction using statistical techniques of machine learning. *Materials Today: Proceedings*. DOI: <https://doi.org/10.1016/j.matpr.2021.01.948>.

Patrício DI, Rieder R (2018) Computer vision and artificial intelligence in precision agriculture for grain crops: A systematic review. *Computers and Electronics in Agriculture* 153, 69–81. DOI: <https://doi.org/10.1016/j.compag.2018.08.001>.

Paudel D, Boogaard H, Wit A, Janssen S, Osinga S, Pylaniadis C, Athanasiadis NI (2021) *Agricultural Systems* 187, 103016. DOI: <https://doi.org/10.1016/j.agsy.2020.103016>.

Perera PIP, Hoher V, Weerakoon LK, Yakandawala DMD, Fernando SC, Verdeil JL (2010) Early inflorescence and floral development in *Cocos nucifera* L. (Arecaceae: Arecoideae). *South African Journal of Botany* 76, 482–492.

Prasad NR, Patel NR, Danodia A (2020) Crop yield prediction in cotton for regional level using random forest approach. *Spat. Inf. Res.* 1, 1-12. DOI: <https://doi.org/10.1007/s41324-020-00346-6>.

Rahman AA, Zhang X (2018) Prediction of acoustic wave parameters of thermoacoustic prime mover through artificial neural network technique: practical approach for thermoacoustics, *Therm. Sci. Eng. Prog.* 257–268. DOI: <https://doi.org/10.1016/j.tsep.2018.09.002>.

Rakhra M, Soniya P, Tanwar D, Singh P, Bordoloi D, Agarwal P, Takkar S, Jairath K, Verma N (2021). *Materials Today: Proceedings*. DOI: <https://doi.org/10.1016/j.matpr.2021.03.261>.

Rolim GS, Caporusso NB, Moreto VB, Pantano AP (2015) Influence of El Niño and La Niña on sugarcane yield and sucrose production in northern São Paulo, Brazil. *Australian Journal of Crop Science (Online)*, 9, 509-516.

Rolim GS, Sentelhas PC, Barbieri V (1998) Planilhas no ambiente Excel™ para os cálculos de balanços hídricos: normal, sequencial, de cultura e de

produtividade real e potencial. *Revista Brasileira de Agrometeorologia*, 6, 133-137.

Sack S, Abom M (2020) Acoustic plane-wave decomposition by means of multilayer perceptron neural networks. *Journal of Sound and Vibration* 486, 115518. DOI: <https://doi.org/10.1016/j.jsv.2020.115518>.

Salata S, Grillenzoni C (2021) A spatial evaluation of multifunctional Ecosystem Service networks using Principal Component Analysis: A case of study in Turin, Italy. *Ecological Indicators*, 127, 107758. DOI: <https://doi.org/10.1016/j.ecolind.2021.107758>.

Shahhosseini M, Martinez-Feria RA, Hu G, Archontoulis SV (2019) Maize yield and nitrate loss prediction with machine learning algorithms. *Environmental Research Letters* 14, 124026. DOI: <https://doi.org/10.1088/1748-9326/ab5268>.

Snider JL, Oosterhuis DM, Kawakami EM, Loka DA (2011) High temperature limits in vivo pollen tube growth rates by altering diurnal carbohydrate balance in field-grown *Gossypium hirsutum* pistils. *Journal of Plant Physiology*, 168, 1168-1175.

Sousa DP (2020) Estimativa do balanço de energia e das necessidades termohídricas do açaizeiro irrigado no leste da Amazônia. 107 f. Tese de doutorado – UFRA, Belém.

Subasi A (2020) Machine learning techniques. *Practical Machine Learning for Data Analysis Using Python*, p. 91-202. DOI: <https://doi.org/10.1016/B978-0-12-821379-7.00003-5>.

Tavares GS, Homma AKO (2015) Comercialização do açaí no Estado do Pará: Alguns cometários. *Revista Observatorio de la Economia Latino americana*, Brasil, n. 211. Disponível em: <http://www.eumet.net/coursecon/ecolat/br/15/acaipara.html>.

Thornthwaite CW, Mather Jr (1955) The water balance. Centerton: Drexel Institute of Technology. 104 p. (Publications in Climatology, v. 8, n. 1).

Walker BJ, South PF and Ort DR (2016) Physiological evidence for plasticity in glycolate/glycerate transport during photorespiration. *Photosynth Res* 129:93–103.

Wang X, Huang J, Feng Q, Yin D (2020) Winter wheat yield prediction at county level and uncertainty analysis in main wheat-producing regions of china with deep learning approaches. *Remote Sens.* 12 (11), 1744.

Wanzeler RTS, Mota MAS, Nogueira DS (2018) Temporal Variability of Rainfall and Convective Available Potential Energy (CAPE) in Eastern Amazon During Chuva Project. *Anuário do Instituto de Geociências*, 14, 305-317. DOI: http://dx.doi.org/10.11137/2018_2_305_317.

Wu J, Serbin SP, Ely KS, Wolfe BT, Dickman LT, Grossiord C (2019) The response of stomatal conductance to seasonal drought in tropical forests. *Glob Change Biol* 00:1–17.

Xavier AC, King CW, Scanlon BR (2016) Daily gridded meteorological variables in Brazil (1980-2013). *Int. J. Climatol.* 36, 2644–2659. DOI:<https://doi.org/10.1002/joc.4518>.

Zhang Y, Lin H, Yang Z, Wang J, Sun Y, Xu B, Zhao Z (2019) Neural network-based approaches for biomedical relation classification: a review, *J. Biomed. Inf.*, 1, 103294. DOI:<https://doi.org/10.1016/j.jbi.2019.103294>.

Zhou Y, Grice K, Stuart-Williams H, Hocart CH, Gessler A, Farquhar GD (2016) Hydrogen isotopic differences between C3 and C4 land plant lipids: consequences of compartmentation in C4 photosynthetic chemistry and C3 photorespiration. *Plant, Cell and Environ* 39:2676–2690.

CAPÍTULO 4 – Previsão agrometeorológica da produtividade do coqueiro modelado por Machine Learning

Resumo:

Modelos agrometeorológicos de previsão de produtividade baseados em algoritmos de Machine Learning (ML), uma subárea da inteligência artificial, são essenciais para o planejamento agrícola, dada a alta velocidade de processamento e a capacidade de fornecer resultados precisos. Objetivou-se neste trabalho avaliar diferentes algoritmos de ML para prever a produtividade do coqueiro em resposta ao clima e fases fenológicas do cultivo. Foram obtidos dados de frutos de coco entre os anos de 1994 a 2016, de plantio comercial da empresa sococo, localizada no município de Moju, Pará, assim como, informações climáticas de precipitação pluvial (P, mm), temperatura máxima (Tmax, °C), insolação (h), umidade relativa (%) e radiação solar (RAD, MJ m⁻² dia⁻¹) pela conversão da insolação, de estação automática pertencente a empresa. Os algoritmos de ML avaliados foram Regressão Linear Múltipla (MLP), Random Forest (RF), Redes Neurais Artificiais (MLP) e Support Vector Machine (SVM) de kernel de Função de Base Radial (SVM_RBF), Linear (SVM_Linear) e Polinomial (SVM_Poly), sendo validados pelo R², MAPE, RMSE e ME. Os resultados apontaram que o modelo RF é um método de ML eficaz para previsão de produtividade de frutos do coqueiro, apresentando melhores desempenhos em comparação aos demais métodos preditivos. Foi possível obter precisões pela determinação do R² para o RF variando de 0.67 a 0.69, enquanto esses valores variaram de 0.19 a 0.45 para MLP. A ETR_TFS, TMAX_DF(IV) e DPV_DF(II), foram as variáveis independentes mais importantes para previsão de produtividade. Previsões de produtividade para o coqueiro apresentam desempenho satisfatório com a até 11 meses antes da colheita, com R² de 0,68 e MAPE de aproximadamente 11%.

Palavras-chaves: Amazônia oriental, Arecaceae, *Cocos nucifera* L., Inteligência artificial, Python

Agrometeorological forecast of coconut yield modeled by Machine Learning

ABSTRACT

Agrometeorological models for yield prediction based on Machine Learning (ML) algorithms, a subarea of artificial intelligence, are essential for agricultural planning, given the high processing speed and the ability to provide accurate results. It was aimed in this work to evaluate different ML algorithms to predict coconut fruit productivity in response to climate and crop phenological phases. Data of coconut fruits between the years 1994 to 2016 were obtained from commercial plantation of sococo company, located in the municipality of Moju, Pará, as well as, climate information of rainfall (P mm), maximum temperature (Tmax, °C), Insolation (h), relative humidity (%) and solar radiation (RAD, MJ m⁻² day⁻¹) by conversion of insolation, from automatic station belonging to the company. The ML algorithms evaluated were Multiple Linear Regression (MLP), Random Forest (RF), Artificial Neural Networks (MLP) and Support Vector Machine (SVM) of Radial (SVM_RBF), Linear (SVM_Linear) and Polynomial (SVM_Poly) basis function kernel, being validated by R², MAPE, RMSE and ME. The results pointed out that the RF model is an effective ML method for coconut fruit yield prediction, showing better performances compared to the other predictive methods. It was possible to obtain accuracies by R² determination for RF ranging from 0.67 to 0.69, while these values ranged from 0.19 to 0.45 for MLP. ETR_TFS, TMAX_DF(IV) and DPV_DF(II), were the most important independent variables for yield prediction. Yield predictions for coconut palm performed satisfactorily with the up to 11 months before harvest, with R² of 0.68 and MAPE of approximately 11%.

Keywords: Eastern Amazonia, Arecaceae, *Cocos nucifera* L., Artificial intelligence, Python

4.1 INTRODUÇÃO

Com exploração comercial em mais de 90 países e centro de origem no sudeste e o sul do continente asiático (Gunn et al., 2011), o coqueiro (*Cocos nucifera* L.) é cultivado em escala global em cerca de 12 milhões de hectares, com uma produção anual de aproximadamente 62 bilhões de toneladas (FAOSTAT, 2021).

Os países do continente asiático (Filipinas, Indonésia e Índia) concentram 72% da área plantada e 74% da produção global (FAOSTAT, 2021). A indonésia destaca-se como o maior produtor mundial de coco, com uma produção de 17 milhões de toneladas, seguida pelas Filipinas e Índia, ambas com cerca 14 milhões de toneladas. Contudo, em área colhida, as Filipinas destacam-se como maior área cultivada com 3.6 milhões de hectares (FAOSTAT, 2021).

O Brasil possui 188.113 mil hectares plantados, ocupando a oitava posição no cenário global com 1.58% de área plantada, no entanto, é o quinto maior produtor em toneladas da cultura, com mais de 2 milhões de toneladas, representando 3.7% da produção mundial (IBGE, 2021; FAOSTAT, 2021).

Os maiores produtores de coco no Brasil são localizados nos estados da Bahia, Ceará, Pará, Sergipe e Espírito Santo. Esses estados somam mais de 70% da produção nacional. A produção de coco no Brasil nos últimos 30 anos, a qual em 1989 era de 681 mil toneladas, passou para 1.5 milhões de toneladas em 2019, crescimento de aproximadamente 128% (IBGE, 2021).

Esse crescimento é destacado pelos ganhos econômico, sociais e ambientais, refletidos pela exploração sustentável da cultura (Martins; Cuenca, 2018). Além da gama de produtos gerados e comercializados mundialmente, sendo uma das principais fontes de proteínas e calorias para a população (Santos et al., 2019). Sendo assim, a matéria prima principalmente do coco seco por meio do seu albúmem sólido, é considerada como de maior valor agregado na indústria de alimentos, como na produção de óleo de coco, leite de coco, farinha de coco, além de seu uso largamente na culinária, doces e salgados (Martins; Cuenca, 2018).

O coqueiro encontra-se em condições climáticas favoráveis entre os paralelos 20°N e 20°S (Passos et al., 2009). Contudo, mesmo dentro desses

limites geográficos, os elementos climáticos interferem nos diferentes modos de desenvolvimento da planta. Uma boa produção de coco requer chuvas anuais de 1300 mm ou mais e bem distribuídas no ano (Abid; Sharmeen; Praveen, 2007). Temperatura média anual variando de 27 a 30 °C (com variação diurna de 5 a 7 °C) e radiação solar menor que 22 MJ m² dia⁻¹ (Samarasinghe et al., 2018).

Cada começo de inflorescência do coqueiro leva 44 meses de formação, estabelecendo-se desde o início do primórdio floral até a maturação completa do fruto (Perera et al., 2010; Ashburner, 1995). Desta forma, as condições climáticas de ocorrência afetam todas as fases do longo ciclo da cultura. Sendo assim, é possível que haja uma previsibilidade espaçada com base na variabilidade climática para produção de frutos de coco.

A previsão de produtividade de safra é essencial para o planejamento agrícola e melhor tomada de decisão no campo. No entanto, este problema requer o uso de vários conjuntos de dados, uma vez que a produtividade depende de muitos fatores diferentes, como clima, solo, fenologia e outros (Klompenburg; Kassahun; Catal, 2020; Xu et al., 2019).

Os modelos de Machine Learning (ML), que é um ramo da inteligência artificial, é uma abordagem prática que pode determinar padrões e correlações e construir os resultados a partir de conjuntos de dados, podendo assim fornecer uma melhor previsão com base em vários recursos (Klompenburg; Kassahun; Catal, 2020). Portanto, modelos de ML podem combinar os benefícios de outros métodos, como modelos de crescimento de cultura (Battisti et al., 2018), sensoriamento remoto (Lopez Lozano et al., 2015), com modelagem baseada em dados para fazer previsões confiáveis da produtividade agrícola (Paudel et al., 2021).

Os métodos de ML mais comumente usados como modelos preditivos são Random Forest (Luciano et al., 2021), Redes Neurais Artificiais (Oliveira et al., 2021), Support Vector Machine (Pant et al., 2021) e Regressão Linear Múltipla (Rolim et al., 2020; Aparecido et al., 2020). Revisões recentes de Klompenburg, Kassahun e Catal, (2020) e Liakos et al. (2018) analisaram as aplicações do aprendizado de máquina na agricultura.

Como exemplo, Crane Droesch (2018) aplicou redes neurais para prever a produção de milho nos Estados Unidos. Paudel et al. (2021) combinaram princípios agronômicos de modelagem de safra com ML para previsão de safra em grande escala, utilizando algoritmos de Support Vector Machine, KNN, Ridge Regression e Gradient-Boosted Decision Trees. Charoen-Ung e Mittrapiyanuruk (2019), apresenta um método baseado em Random Forest para prever o grau de produtividade da cana-de-açúcar. Pant et al. (2021), avaliaram a previsão de safra para os cultivos do milho, batata, arroz e trigo, usando Random Forest e Support Vector Machine. Rolim et al. (2020), avaliaram a qualidade natural da bebida de café por meio de modelos de regressão linear múltipla.

Poucas pesquisas a fim de avaliar a previsão da produtividade do coco vem sendo realizadas, tornando-se esses resultados ainda mais escassos quando aplicados em algoritmos de ML. Portanto, o presente estudo foi conduzido para investigar o desempenho dos modelos de ML na previsão de produtividade de frutos de coco na região do nordeste paraense, em resposta a variabilidade climática e fenológica da cultura.

4.2 MATERIAL E MÉTODOS

4.2.1 Área de estudo e origem dos dados

Foram obtidos dados de produção de frutos de coco da variedade híbrido, em plantios comerciais da empresa Sococo, localizada no município de Moju-Pará (02° 07' 00" S, 48° 38' 00" W) (Figura 1). Os dados coletados de produção são referentes aos anos de 1994 a 2016 em escala mensal. Informações climáticas entre os anos de produção de frutos, foram coletados por uma estação meteorológica automática localizada nas dependências da empresa próxima as áreas de plantio.

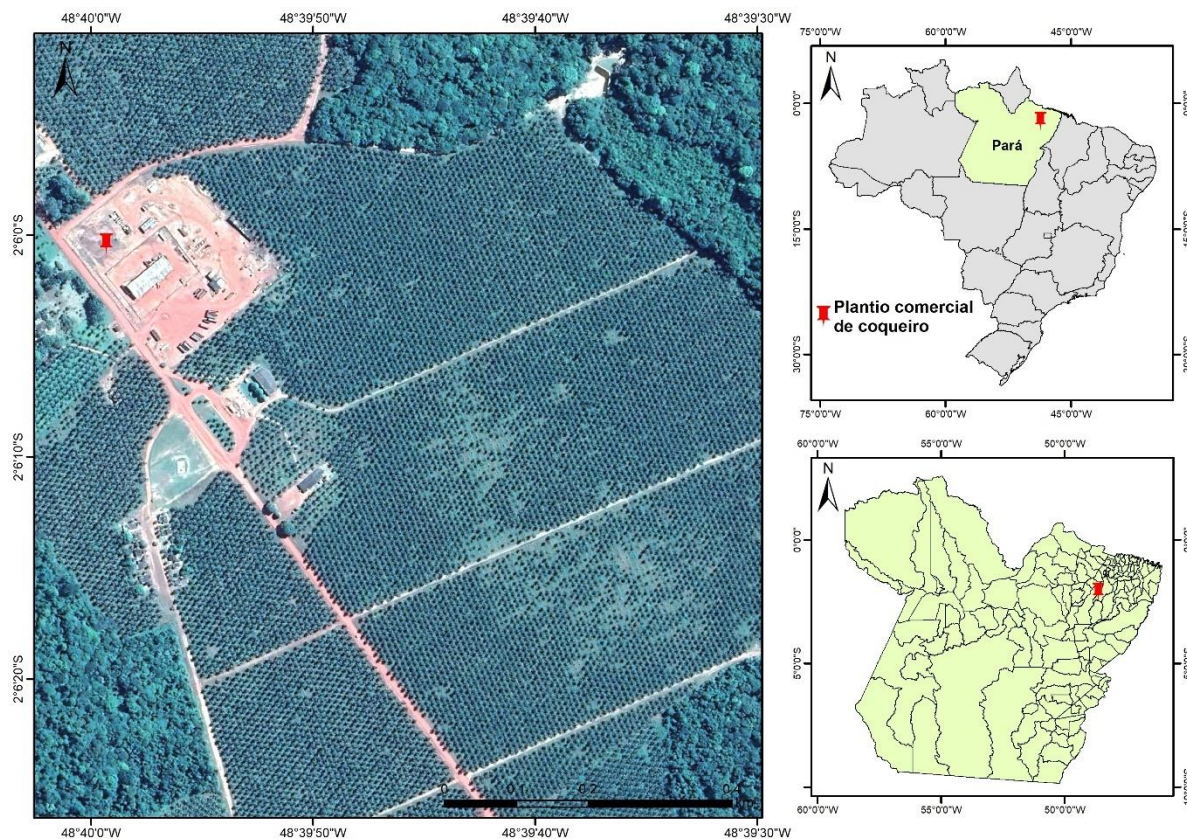


Figura 1. Mapa de localização da área comercial de cultivo da cultura do coqueiro na região do nordeste paraense.

Na figura 2, é possível observar a representação do fluxo grama de trabalho do conjunto de dados climáticos e de cultivo de coqueiro, assim como as análises exploratórias para aplicação dos modelos de machine learning na previsão de produtividade da cultura aplicados neste estudo.

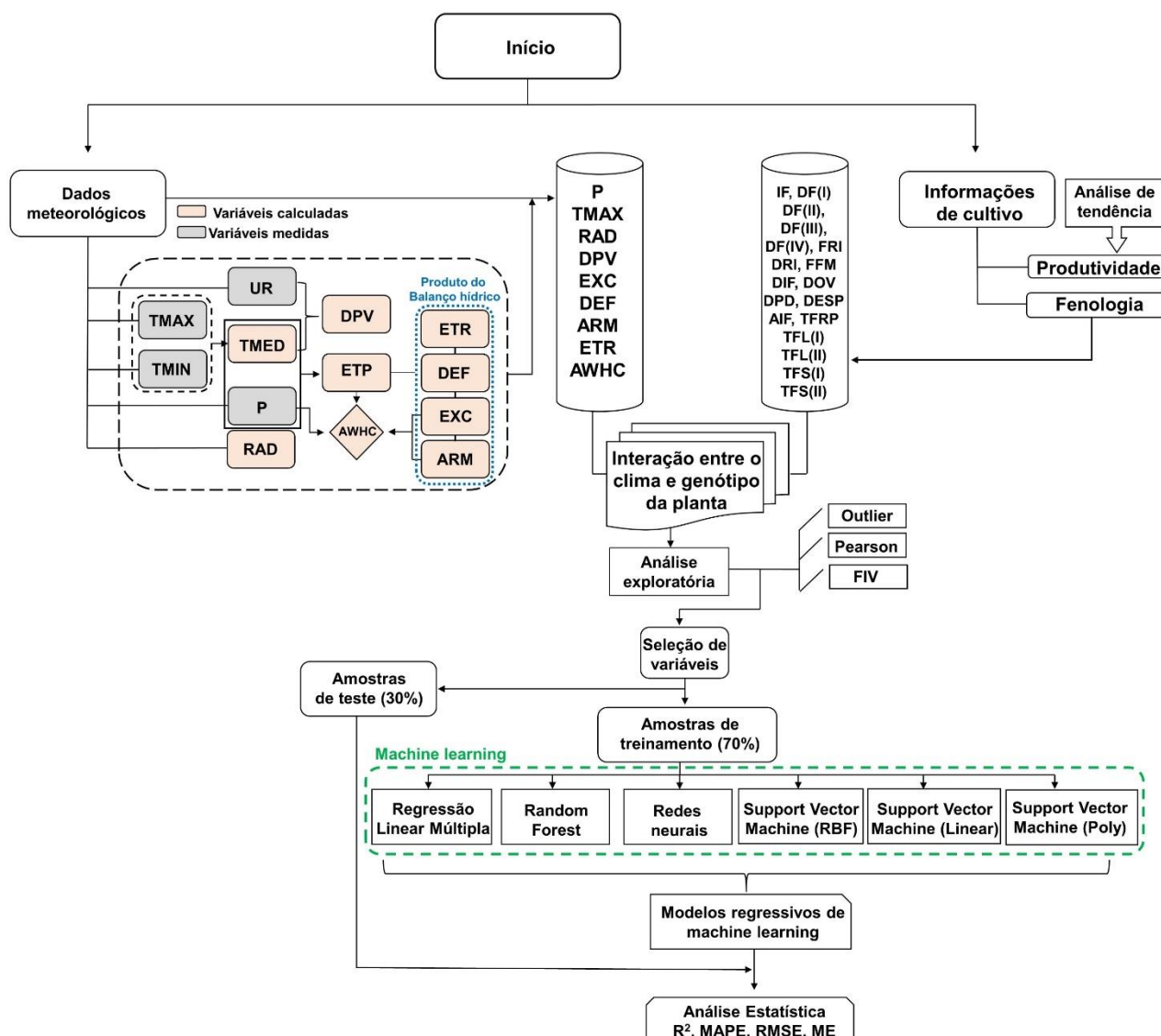


Figura 2. Etapas metodológicas conduzidos para este estudo. Variáveis meteorológicas: Precipitação pluvial (P mm), Deficiência hídrica (DEF mm), Excedente hídrico (EXC mm), Armazenamento de água no solo (ARM mm), Evapotranspiração potencial (ETP mm), Evapotranspiração real (ETR mm), Temperatura média (TMED °C), Temperatura mínima (TMIN °C), Temperatura máxima (TMAX °C), Radiação solar (RAD MJ m⁻² dia⁻¹), Déficit de pressão de vapor (DPV kPa), Umidade relativa (UR %), Percentage Available Water Holding Capacity (AWHC %). Variáveis de cultivo: Início do primórdio floral (IF), Desenvolvimento floral (I) (DF I), Desenvolvimento floral (II) (DF II), Desenvolvimento floral (III) (DF III), Desenvolvimento floral (IV) (DF IV), Desenvolvimento floral (V) (DF V), Formação do ramo da inflorescência (FRI),

Desenvolvimento do ramo da inflorescência (DRI), Formação de flores femininas e masculinas (FFM), Desenvolvimento de inflorescência (DIF), Diferenciação do ovário (DOV), Desenvolvimento pós diferenciação (DPD), Desenvolvimento de espata (DESP), Abertura de Inflorescência (AIF), Desenvolvimento do mesocarpo e endocarpo (TFRP), Desenvolvimento fruto (fase líquida) (TFL), Desenvolvimento fruto (fase sólida) (TFS).

Foram disponibilizados dados diários de precipitação pluvial (P mm), temperatura máxima (Tmax, °C), Insolação (h) que posteriormente foram convertidos em radiação solar (RAD, MJ m⁻² dia⁻¹) por meio do programa clima (Faria et al. 2002) e umidade relativa (%). Com base nos dados de UR, foi calculado o déficit de pressão de vapor (DPV) por meio das seguintes equações (1 a 3):

$$DPV = Es - Ea \quad (1)$$

$$Es = 0.6108 \times 10^{\left(\frac{7.5 \times TMED}{237.3 + TMED}\right)} \quad (2)$$

$$Ea = \frac{UR \times Es}{100} \quad (3)$$

em que, DPV= Déficit de pressão de vapor (kPa); Es = Pressão de saturação de vapor (kPa); Ea = Pressão parcial de vapor (kPa); UR= Umidade relativa (%); TMED= Temperatura média do ar (°C).

Seguindo o método de Camargo (1971), foi obtido a evapotranspiração potencial (equação 4), utilizando dados de temperatura média do ar e precipitação pluvial.

$$ETP = F Q_o T ND \quad (4)$$

sendo Qo (mm dia⁻¹) a radiação solar extraterrestre diária expressa em equivalente de evaporação, no período considerado, T (°C) é a temperatura média do ar durante o período; F é o fator de ajuste que varia com a temperatura

média anual do local (para T_m até 23 °C, $F = 0,01$; $T_m = 24$ °C, $F = 0,0105$; $T_m = 25$ °C, $F = 0,011$; $T_m = 26$ °C, $F = 0,0115$; $T_m > 26$ °C, $F = 0,012$); e ND é o número de dias do período.

Com os dados de ETP, foram realizados os balanços hídricos climatológicos segundo Thornthwaite e Mather (1955), utilizando capacidades de água disponível (CAD) igual a 125 mm (Ferreira, Warwick e Siqueira, 2018), equação (5 a10).

$$Se (P - ETP)_i < 0 = \begin{cases} NAC_i = NAC_{i-1} + (P - ETP)_i \\ ARM = CAD e^{\frac{(NAC)_i}{CAD}} \end{cases} \quad (5)$$

$$Se (P - ETP)_i \geq 0 = \begin{cases} NAC_i = CAD \ln \frac{ARM_i}{CAD} \\ ARM = (P - ETP)_i + ARM_{i-1} \end{cases} \quad (6)$$

$$ALT_i = ARM_i - ARM_{i-1} \quad (7)$$

$$ETR_i = \begin{cases} P + |ALT_i|, & se \ ALT < 0 \\ ETP_i, & se \ ALT \geq 0 \end{cases} \quad (8)$$

$$DEF = ETP - ETR \quad (9)$$

$$EXC_i = \begin{cases} 0, & se \ CAD < 0 \\ (P - ETP)_i - ALT_i, & se \ CAD = 0 \end{cases} \quad (10)$$

em que E é a ETP é a Evapotranspiração Potencial (mm); CAD é a Capacidade de Água Disponível no solo (mm), ARM é o armazenamento de água no solo (mm); NAC é (Negativo acumulado)= precipitação acumulada – evapotranspiração potencial; P é a precipitação (mm); DEF é a deficiência hídrica no sistema solo-planta-atmosfera (mm); ETR é a evapotranspiração real (mm), EXC é o excedente hídrico do sistema solo-planta-atmosfera (mm); ALT é o armazenamento de água no solo no mês atual – armazenamento de água no solo do mês anterior (mm) e i é o período mensal.

Com base nos componentes de BH, foi calculado a porcentagem de capacidade de armazenamento de água no solo disponível no final do mês (AWHC %) (Surre, 1968), por meio da seguinte equação (11):

$$\% AWHC = ((SWS + P - ETP - EXC) / AWC) \times 100\% \quad (11)$$

sendo %AWHC é a porcentagem da capacidade de retenção de água disponível no final do mês, SWS é o armazenamento de água no solo (mm), e AWC é a capacidade de água disponível (mm).

O clima da região segundo a classificação de Köppen é caracterizado como Af₂ (Martorano et al., 1993), apresentando o mês mais seco precipitação acima de 60 mm e chuvas anuais entre 2500 a 3000 mm ano⁻¹. A precipitação pluvial média anual da região é de 2.856 mm, com temperatura máxima anual média de 34°C. As variáveis climáticas utilizadas para a previsão da produtividade de frutos de coco foram, Precipitação pluvial (P mm), Deficiência hídrica (DEF mm), Excedente hídrico (EXC mm), Armazenamento de água no solo (ARM mm), Evapotranspiração real (ETR mm), Temperatura máxima (TMAX °C), Percentage Available Water Holding Capacity (AWHC %), Radiação solar (RAD MJ m⁻² dia⁻¹) e Déficit de pressão de vapor (DPV kPa) (Figura 3).

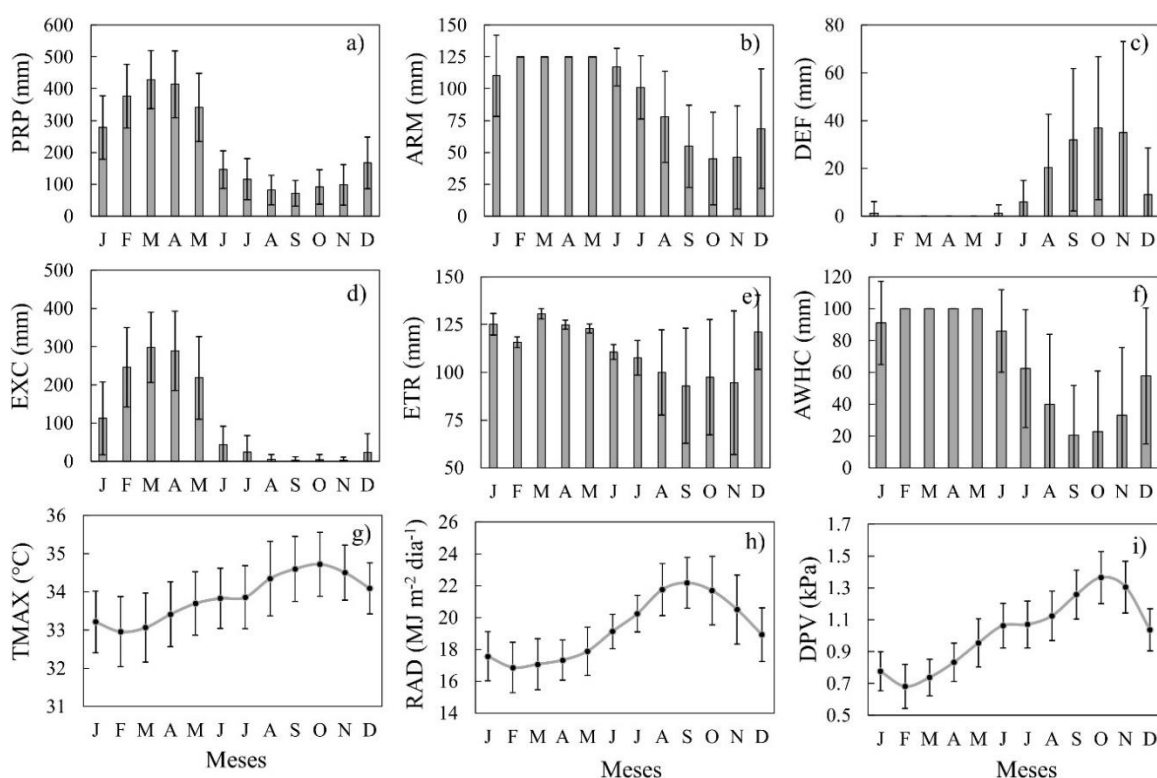


Figura 3. Variabilidade do clima na região de estudo, a) Precipitação pluvial (P mm), b) Deficiência hídrica (DEF mm), c) Excedente hídrico (EXC mm), d) Armazenamento de água no solo (ARM mm), e) Evapotranspiração real (ETR mm), f) Temperatura máxima (TMAX °C), g) Percentage Available Water Holding

Capacity (AWHC %), h) Radiação solar (RAD MJ m⁻² dia⁻¹) e Déficit de pressão de vapor (DPV kPa), na regiões de cultivo de coco no município de Moju-Pará.

4.2.1 Caracterização fenológica do coqueiro

O ciclo de crescimento de um cacho de coco dura em torno 44 meses, dando início nos primeiros primórdios florais até a maturação dos frutos (Perera et al. 2010; Ashburner, 1995) (Figura 4 e 5). Do período total da fenologia, a fase de desenvolvimento floral é mais longa, com ciclo médio de 14 meses. A inflorescência do coqueiro abre em média a cada mês do ano, produzindo frutos maduros para serem colhidos após 11 meses em média.

Vale destacar que neste trabalho a produção se destina a frutos de coco seco, fase de plena maturação, em que o albúmen sólido se encontra totalmente formado com médias de 11 meses após a polinização. No entanto, outro mercado de importância é a de coco verde, que devem ser colhidos para melhor produtividade de água e sabor entre o sexto e o sétimo mês após a polinização, fase que se inicia a formação do albúmen sólido, (Martins; Jesus Júnior, 2014).

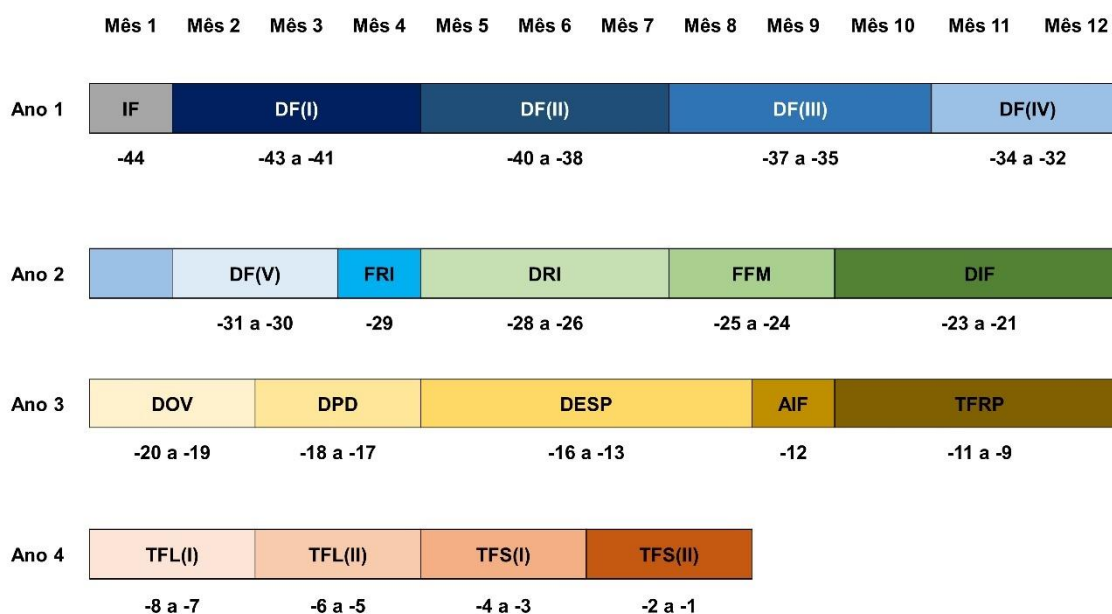


Figura 4. Desenvolvimento da inflorescência do coco. Início do primórdio floral (IF), Desenvolvimento floral (I) (DF I), Desenvolvimento floral (II) (DF II), Desenvolvimento floral (III) (DF III), Desenvolvimento floral (IV) (DF IV),

Desenvolvimento floral (V) (DF V), Formação do ramo da inflorescência (FRI), Desenvolvimento do ramo da inflorescência (DRI), Formação de flores femininas e masculinas (FFM), Desenvolvimento de inflorescência (DIF), Diferenciação do ovário (DOV), Desenvolvimento pós diferenciação (DPD), Desenvolvimento de espata (DESP), Abertura de Inflorescência (AIF), Desenvolvimento do mesocarpo e endocarpo (TFRP), Desenvolvimento fruto (fase líquida) (TFL), Desenvolvimento fruto (fase sólida) (TFS).

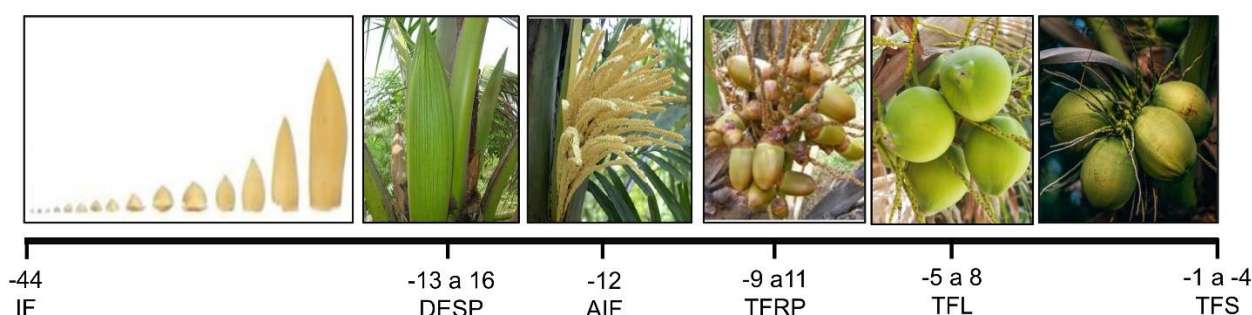


Figura 5. Fases de desenvolvimento de um cacho de coco.

4.2.2 Modelos de Machine Learning

Os modelos de previsão de produtividade para os frutos de coco foram simulados por seis modelos de Machine Learning: Regressão Linear Múltipla (RLM) (Oribe-Garcia et al., 2015), Random Forest (RF) (Hariharan, 2021; Breiman, 2001), Redes Neurais Artificiais (MLP) (Sack; Abom, 2020; Antanasijević; Antanasijević; Pocajt, 2018), Support Vector Machine (SVM) (Manjunathan et al., 2020) de Base Radial RBF (SVM_RBF), Linear (SVM_Linear) e Polinomial (SVM_Poly), usando a linguagem de programação Python.

Para todos os modelos, foram separados 70% dos dados para treinamento e 30% para teste. Sendo que para seleção dos melhores parâmetros de treinamento dos modelos foi utilizado o método gridsearchCV, uma biblioteca para python, selecionado como resposta a combinação que apresentar menor erro em relação aos dados observados.

A multicolinearidade dos dados foi avaliada por meio do Fator de Inflação de Variância (FIV), usando a biblioteca StatsModels da linguagem de

programação Python, evitando assim variáveis independentes correlacionadas dentro dos modelos de ML (Forthofer; Lee; Hernandez, 2007; Miles, 2014).

4.2.3 Processamento gráfico e estatísticos

Os gráficos foram construídos usando o Seaborn, uma biblioteca de visualização de dados Python baseada em matplotlib, fornecendo uma interface de alto nível para desenhar gráficos estatísticos.

Um mapa de calor com dendrograma foi desenvolvido para facilitar a apresentação do coeficiente de correlação de Pearson entre a produtividade e as variáveis independentes dos modelos de machine learning durante os meses do ano. Foi utilizado o agrupamento hierárquico no dendrograma, que envolve o cálculo das matrizes de distância de objetos de dados e, em seguida, a fusão de objetos que estão próximos uns dos outros para formar sub-agrupamentos. O método de agrupamento foi definido pela distância euclidiana.

A necessidade desse tipo de visualização surge da complexidade de grandes conjuntos de dados experimentais. Uma vantagem importante de usar um mapa de calor é a possibilidade da identificação de padrões de forma mais fácil.

As avaliações dos modelos de ML foram estimadas por meio do coeficiente de determinação (R^2), Raiz do erro quadrático médio (RMSE) e do Erro absoluto médio (MAPE) e pelo erro médio (ME).

$$R^2 = \frac{\sum_{i=1}^n (Y_{est} - \overline{Y_{est}})^2}{\sum_{i=1}^n (Y_{obs} - \overline{Y_{est}})^2} \quad (12)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (Y_{obs} - Y_{est})^2}{n}} \quad (13)$$

$$MAPE = \frac{\sum_{i=1}^n \left(\frac{|Y_{obs} - Y_{est}|}{Y_{obs}} \right)}{n} * 100 \quad (14)$$

$$ME = \frac{1}{n} \sum_{i=1}^n (Y_{est} - Y_{obs}) \quad (15)$$

Y_{est} é o valor estimado de y ; Y_{obs} é o valor observado de y ; $\overline{Y_{est}}$ é o valor médio estimado de y ; n é número de dados.

4.3 RESULTADOS

O coqueiro apesar de emitir frutos durante todos os meses do ano, a sua produtividade varia em escala de tempo mensal (Figura 6). Em média, a produção mensal por hectare observada foi de 2.310 cocos $\text{ha}^{-1} \text{mês}^{-1}$ com médias de 15 frutos por planta mês. Os valores máximo e mínimos de frutos por hectare mês observado foram de 3.662 e 1.313 $\text{ha}^{-1} \text{mês}^{-1}$, respectivamente (Figura 6.b).

Aproximadamente 75% da produção de frutos é concentrada com valores acima de 2.000 cocos $\text{ha}^{-1} \text{mês}^{-1}$. A partir do mês de setembro a produtividade média começa a subir, com maiores picos no mês de janeiro (2.975 cocos $\text{ha}^{-1} \text{mês}^{-1}$). Junho a agosto apresentam as menores produtividades, com valores abaixo de 2.000 cocos $\text{ha}^{-1} \text{mês}^{-1}$, sendo o mês de agosto o menos produtivo, com média de 1.811 cocos $\text{ha}^{-1} \text{mês}^{-1}$ (Figura 6.a).

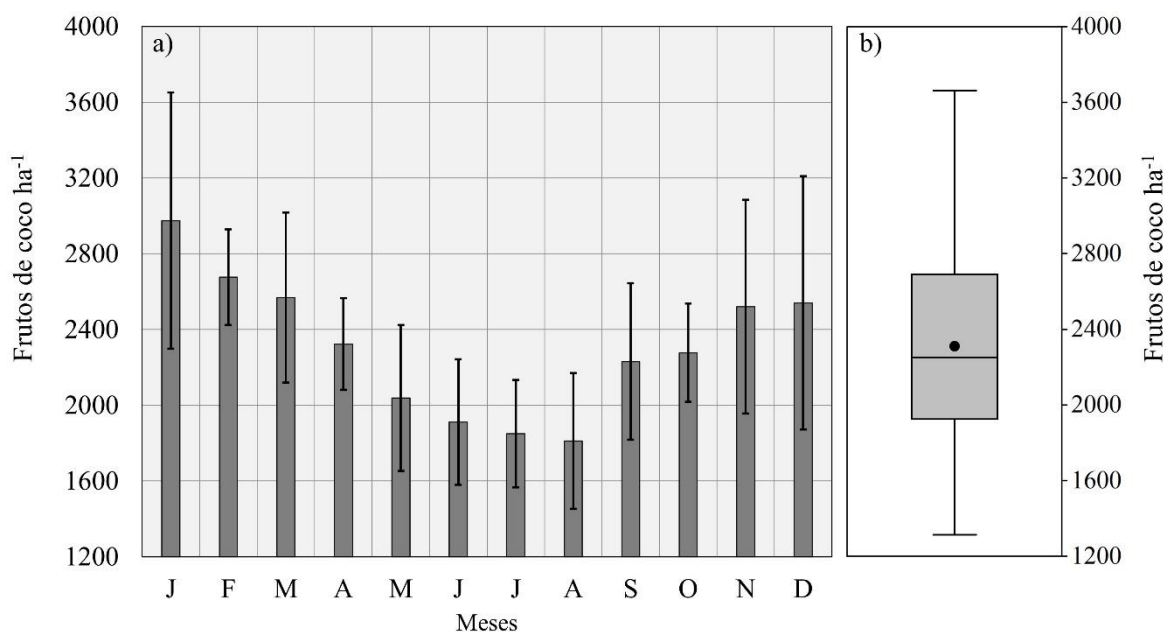


Figura 6. Produtividade média mensal do número de coco seco durante os meses do ano, em plantação comercial na região do nordeste paraense.

Do total de 153 variáveis avaliadas para compor os modelos de machine learning na previsão de produtividade do coqueiro, foram selecionadas 14 (quatorze) variáveis independentes (Figura 7). Uma grande variabilidade

climática é observada entre os meses de colheitas durante diferentes estádios fenológicos do coqueiro.

Para exemplo, o coqueiro na abertura de inflorescência (AIF), estágio que ocorre com antecipação de 12 Meses Antes da Colheita (MAC), a P apresenta diferentes acumulados hídricos no ano (Figura 7.a). Nas colheitas de fevereiro a maio, os acumulados hídricos durante AIF apresentam suas maiores médias e máximas com valores acima de 350 mm e 550 mm, respectivamente. De forma inversa, colheitas realizadas no segundo semestre principalmente entre os meses de agosto a novembro, são observadas as menores médias de P em AIF, com valores acumulados médios abaixo de 80 mm e mínimas que podem chegar a 11 mm entre os meses de agosto e setembro (Figura 7.a).

Observando as oscilações térmicas, a TMAX nas fases de desenvolvimento floral IV (DF IV) 32 a 34 MAC (Figura 7.f), e na formação de frutos pequenos (TFRP) 9 a 11 MAC (Figura 7.h), apresentam variações semelhantes com maiores médias entre maio a setembro, com valores superiores a 33.5°C e menores entre dezembro a março com valores menores que 32.5 °C. Contudo, durante a abertura de inflorescência (AIF), as maiores médias são observadas a partir de julho até novembro, com maiores valores em novembro 33.5 °C e começando a partir das colheitas de janeiro a diminuir as médias mensais das temperaturas máximas, com valores que chegam a 31.5 °C entre fevereiro e março (Figura 7.g)

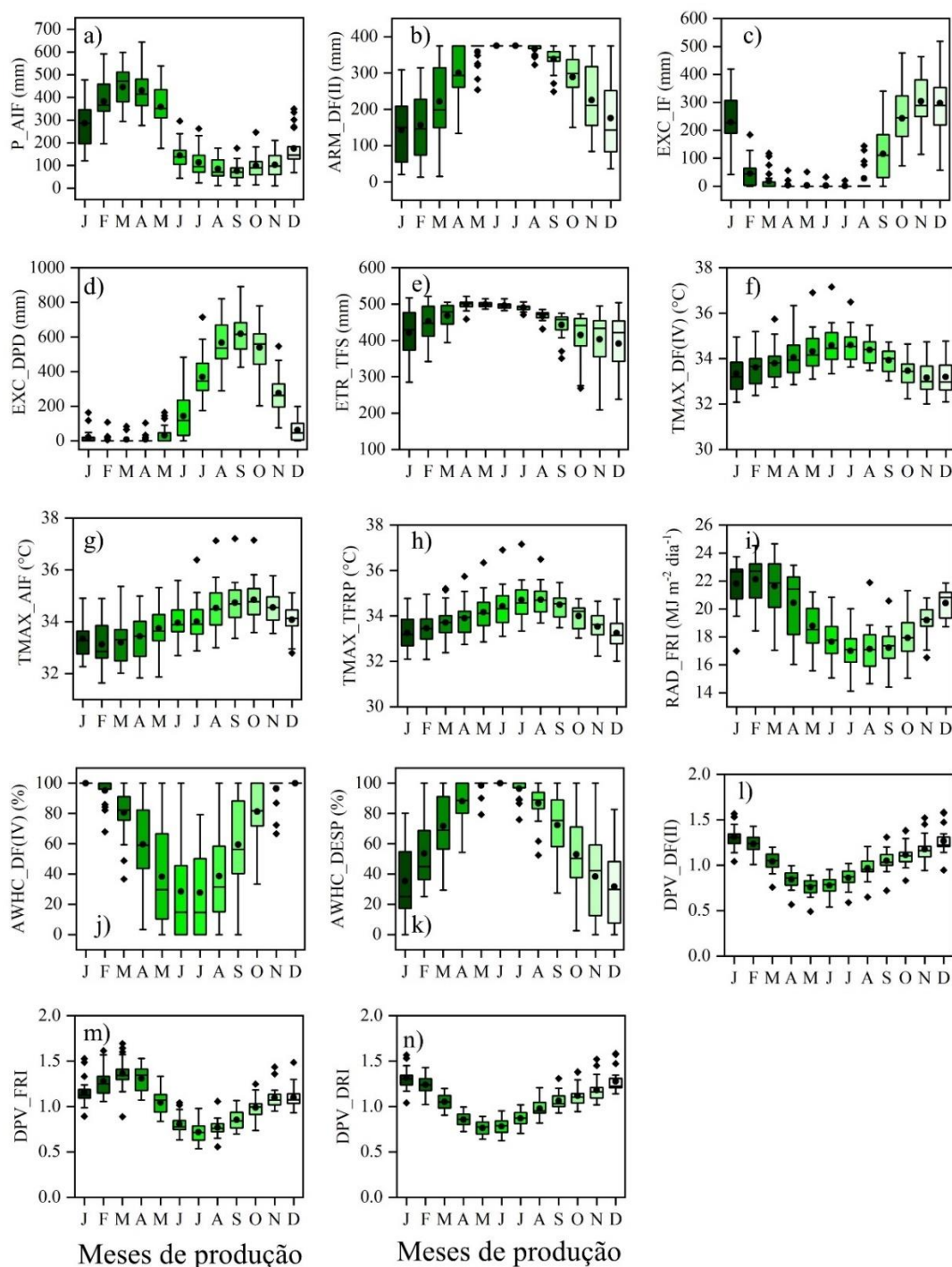


Figura 7. Boxplot da variabilidade mensal dos totais e médias climáticas para as principais fases fenológicas do coqueiro, selecionadas pelo modelo de machine learning.

O mapa de calor e agrupamento das variáveis independentes em função da produtividade de frutos de coqueiro em diferentes meses de produção, mostraram diferenças relevantes entre seus grupos formados (Figura 8). O dendrograma revelou a formação de dois grupos principais. O primeiro grupo é caracterizado por variáveis termodinâmicas, sendo (DPV_DF(II), DPV_DRI, TMAX_DF(IV), TMAX_TFRP, DPV_FRI, ETR_TFS, TMAX_AIF), apresentando maior correlação negativa principalmente no segundo semestre de produção.

O segundo grupo foi definido por variáveis hídricas, destacando-se (EXC_IF, P_AIF, AWHC_DESP, ARM_DF(II), EXC_DPD, AWHC_DF(IV)). Como observado, de forma inversa ao agrupamento termodinâmico, o agrupamento hídrico se caracteriza com maiores correlação positivas no segundo semestre. Esses resultados indicam no geral que quanto menor a disponibilidade de P, menor a produtividade de frutos de coco por hectare e vice-versa.

O mapa de calor evidencia prontamente as mudanças nos valores do coeficiente de correlação de Pearson entre os diferentes estádios fenológicos do coqueiro em diferentes meses de colheita (Figura 8). Como exemplo, observa-se que o estágio DPV_DF(II) tende nos primeiros meses do ano a níveis de correlações positivas, enquanto nos meses seguintes os valores de correlação tendem a aumentar negativamente. Variações semelhantes são observadas no ARM_DF(II), contudo, nos primeiros meses com correlações em sua maioria negativa e no segundo semestre positiva.

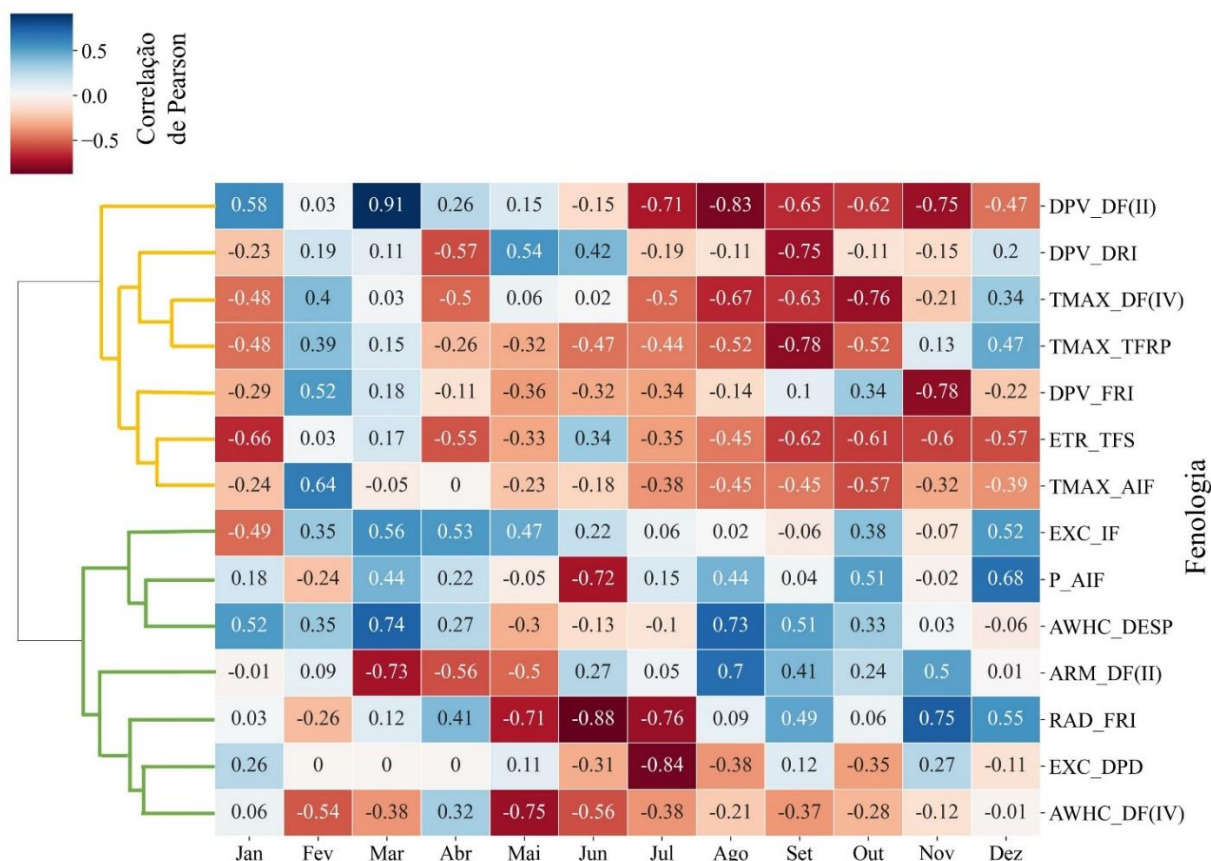


Figura 8. Mapa de calor com dendrograma ilustrando o coeficiente de correlação de Pearson da produtividade com as variáveis independentes de entrada dos modelos de machine learning entre os meses do ano no município de Moju-Pará.

A comparação dos dados de teste que não foram incluídos nos modelos de treinamento, indicaram estimativas satisfatórias para o RF (Figura 9.b). As métricas de desempenho do modelo apresentaram coeficiente de determinação ($R^2 = 0.69$). O RMSE do modelo foi de 328.15 frutos ha^{-1} com MAPE = 11.5%, ou seja, os resultados representam um erro de 11.5% na produtividade estimada em comparação ao observado em campo.

Diferente do RF o modelo MLP (Figura 9.c) foi menos satisfatório, com valores de $R^2 = 0.45$, RMSE = 427.53 frutos ha^{-1} e MAPE de 16.9%, apresentando assim um erro de quase 100 cocos a mais por hectare em relação ao RF.

É importante observar que todos os modelos tentem em média a superestimar a produtividade observada, com tendência estimada pelo ME

variando de 127 frutos ha^{-1} (Figura 9.a) a 41.1 frutos ha^{-1} (Figura 9.b). No entanto, os gráficos de dispersão na reta 1:1 indicam que à medida que a produtividade aumenta principalmente a partir de > 2.500 frutos ha^{-1} os modelos tendem a subestimar os dados observados, e em produtividade mais baixas < 2.500 frutos ha^{-1} os modelos tendem a superestimação (Figura 9). Esses resultados são observados tanto para os modelos de estimativas (Figura 9) quanto para previsão com 8 e 11 meses de antecedência da colheita (Figura 10 e 11) respectivamente.

Os resultados com subestimação e superestimação em produtividades altas e baixas respectivamente, são reforçados pelas limitações dos modelos de ML em estimar meses com pico e declínio de produção (Figura 9.b1 a e.1). A menor produtividade observada foi de 1.326 frutos de coco por hectare, no entanto, os modelos de ML são mais conservadores em suas variações, superestimando os valores com baixas produtividades. O RF obteve a menor diferença, com 1.983 frutos ha^{-1} . De forma distinta, os maiores valores observados de frutos de coco foram de 3.366 coco ha^{-1} , apresentando o modelo RF também o menor erro, com valores de 2.926 coco ha^{-1} . (Figura 9.b1).

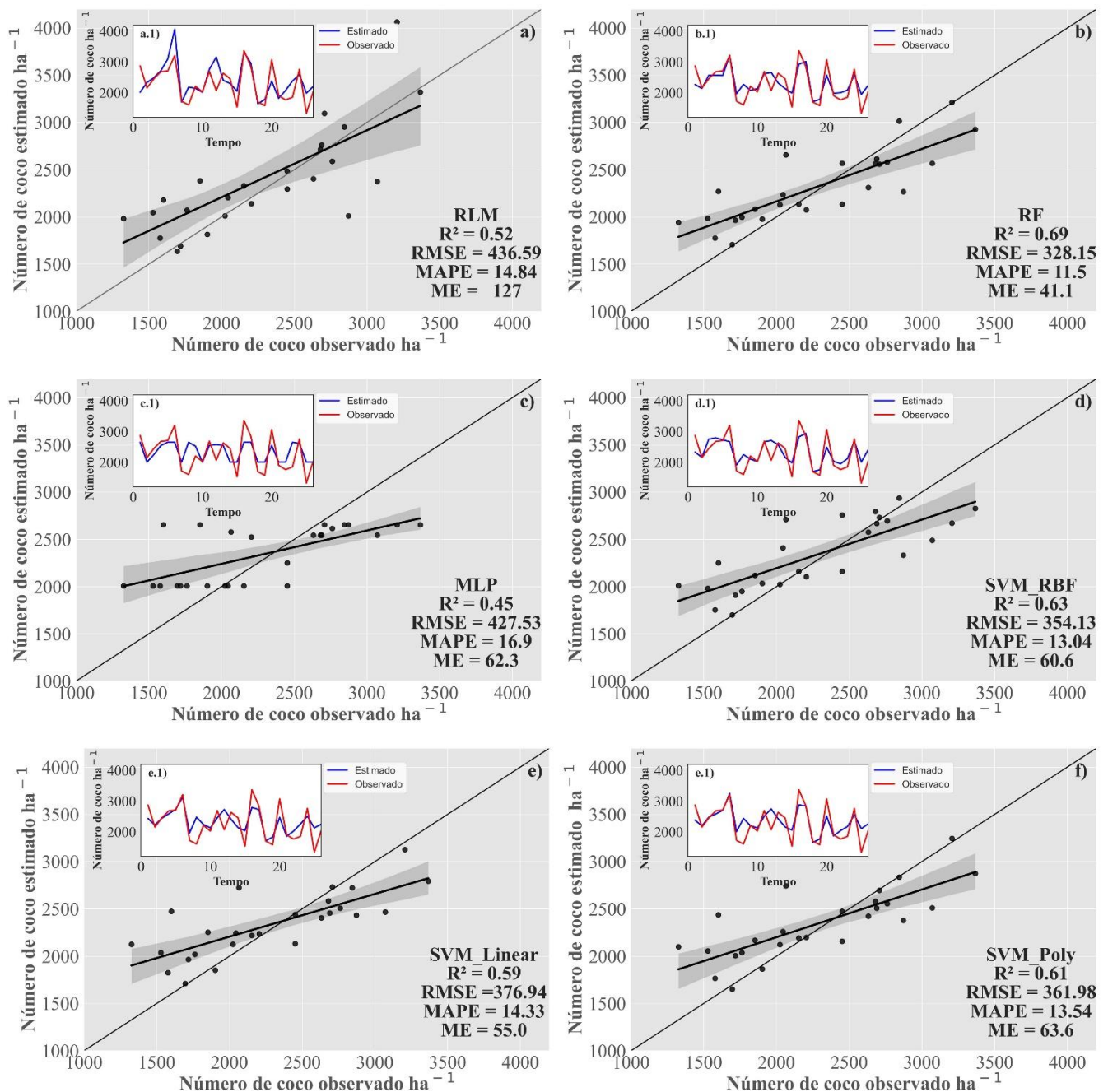


Figura 9. Comparação da produtividade observada obtida em campo em função da produtividade simulada pelos diferentes modelos de machine learning na estimativa de produtividade do coqueiro. a) Regressão Linear Múltipla (RLM), b) Random Forest (RF), c) Multilayer Perceptron (MLP), d) Support Vector Machine RBF (SVM_RBF), e) Support Vector Machine Linear (SVM_Linear), f) Support Vector Machine Poly (SVM_Poly).

Na previsão com 8 e 11 meses de antecedência da colheita (Figura 10 e 11), o modelo RF obteve métricas estatísticas próximas dos valores obtidos pelo modelo de estimativa de produtividade (Figura 9), obtendo ($R^2 = 0.67$ e 0.68), ($RMSE = 335.13$ e 331.7 frutos ha^{-1}), ($MAPE = 11.54\%$ e 11.40%) e ($ME = 55.2$ e 58.6 frutos ha^{-1}), respectivamente. Em contrapartida, o MLP apresentou os menores desempenhos, com valores de ($R^2 = 0.37$ e 0.19), ($RMSE = 446.63$ e 556.36 frutos ha^{-1}), ($MAPE = 17.29\%$ e 19.27%) e ($ME = 165.3$ e 41.6 frutos ha^{-1}), entre as previsões com 8 e 11 meses respectivamente.

Assim como para os modelos de estimativa (Figura 9), as previsões obtidas apresentam valores mais conservadores, com limitações em acompanhar oscilações de produtividade positivas e negativas, ou seja, maiores e menores produtividades ao longo dos meses (Figura 10 e 11, a1, e1). Jeong et al. (2016), ressaltam que o algoritmo de regressão pode resultar na subestimação ou superestimação das respostas marginais na faixa onde os pontos de dados são escassos em extremos. Esses resultados também são reforçados por Breiman (2001), evidenciando que os modelos de ML superestimam as respostas com valores mais altos.

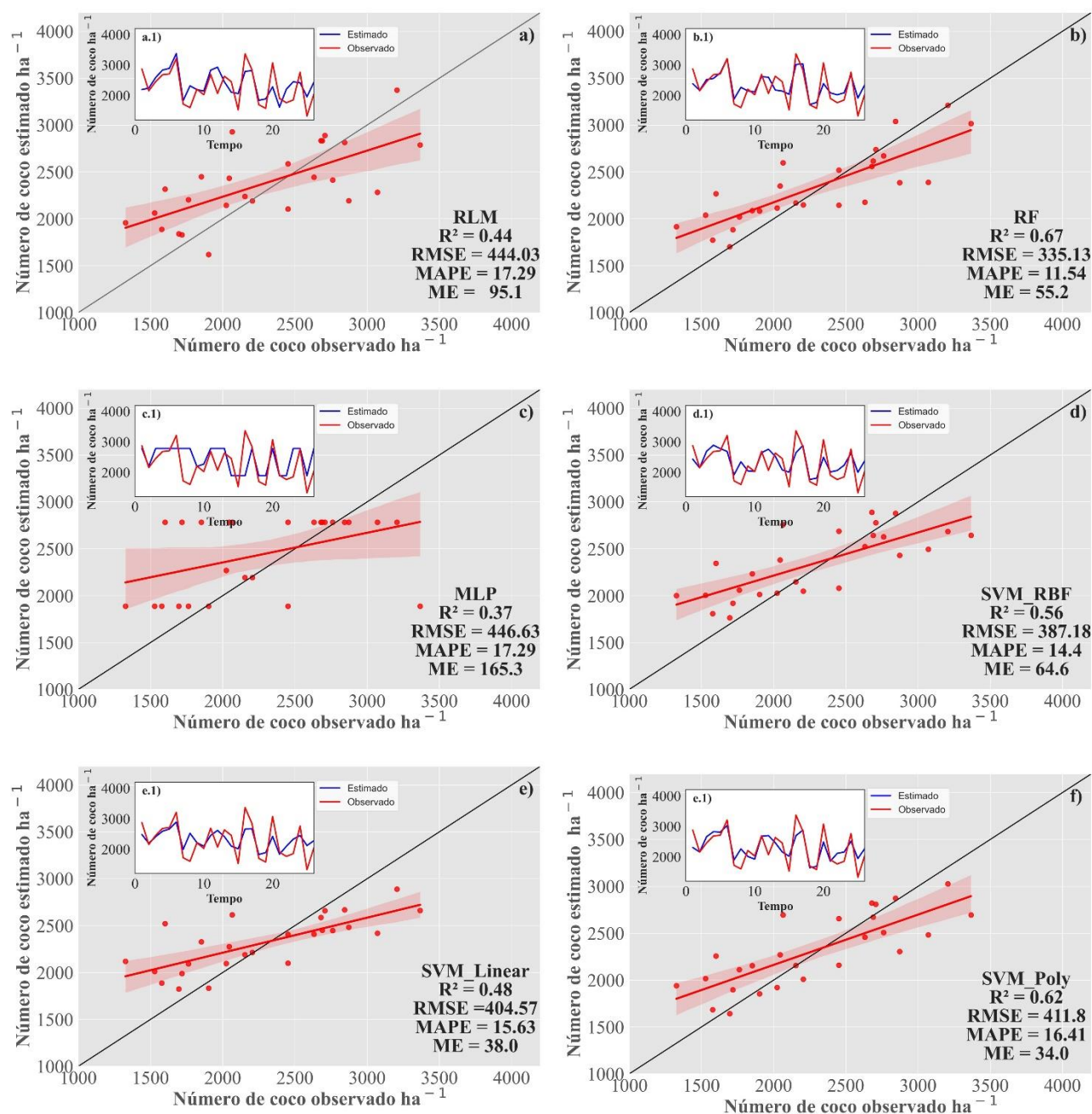


Figura 10. Comparação da produtividade observada obtida em campo em função da produtividade simulada pelos diferentes modelos de machine learning na previsão com 8 (oito) meses de antecipação da produtividade do coqueiro. a) Regressão linear múltipla (RLM), b) Random Forest (RF), c) Multilayer Perceptron (MLP), d) Support Vector Machine RBF (SVM_RBF), e) Support Vector Machine Linear (SVM_Linear), f) Support Vector Machine Poly (SVM_Poly).

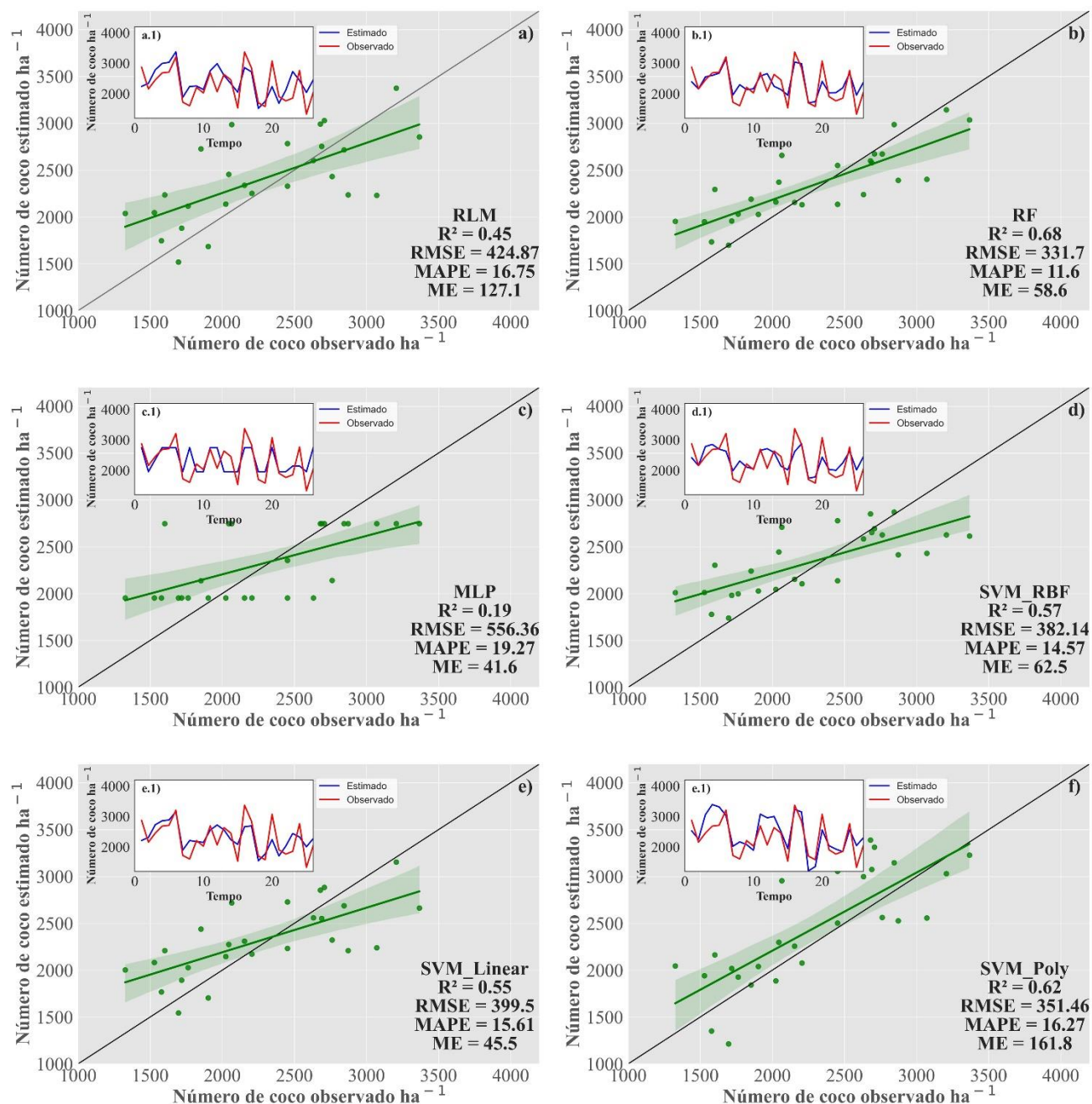


Figura 11. Comparação da produtividade observada obtida em campo em função da produtividade simulada pelos diferentes modelos de machine learning na previsão com 11 (onze) meses de antecipação da produtividade do coqueiro. a) Regressão Linear Múltipla (RLM), b) Random Forest (RF), c) Multilayer Perceptron (MLP), d) Support Vector Machine RBF (SVM_RBF), e) Support Vector Machine Linear (SVM_Linear), f) Support Vector Machine Poly (SVM_Poly).

Além da capacidade preditiva dos modelos, também foi possível fornecer informações úteis sobre a importância das variáveis. A classificação de importância da variável e o impacto parcial da variável na resposta dos modelos podem ser avaliados para fins de análise mais detalhada dos resultados. Sendo assim, a medida de importância das variáveis do modelo revelou que a ETR_TFS é uma das variáveis mais influente, seguida por TMAX_DF(IV) e DPV_DF(II). É importante destacar que Altas temperaturas foliares, altas DPV e baixo potencial de água na folha, levam a limitações estomáticas do coqueiro (Gomes et al., 2006), afetando as taxas fotossintéticas e resultando na redução do acúmulo de matéria seca e da produção da cultura (Kumar e Aggarwal, 2013). Outras variáveis como EXC_DPD, EXC_IF, AWHCDF(IV) foram classificadas em níveis mais baixos de importância (Figura 12).

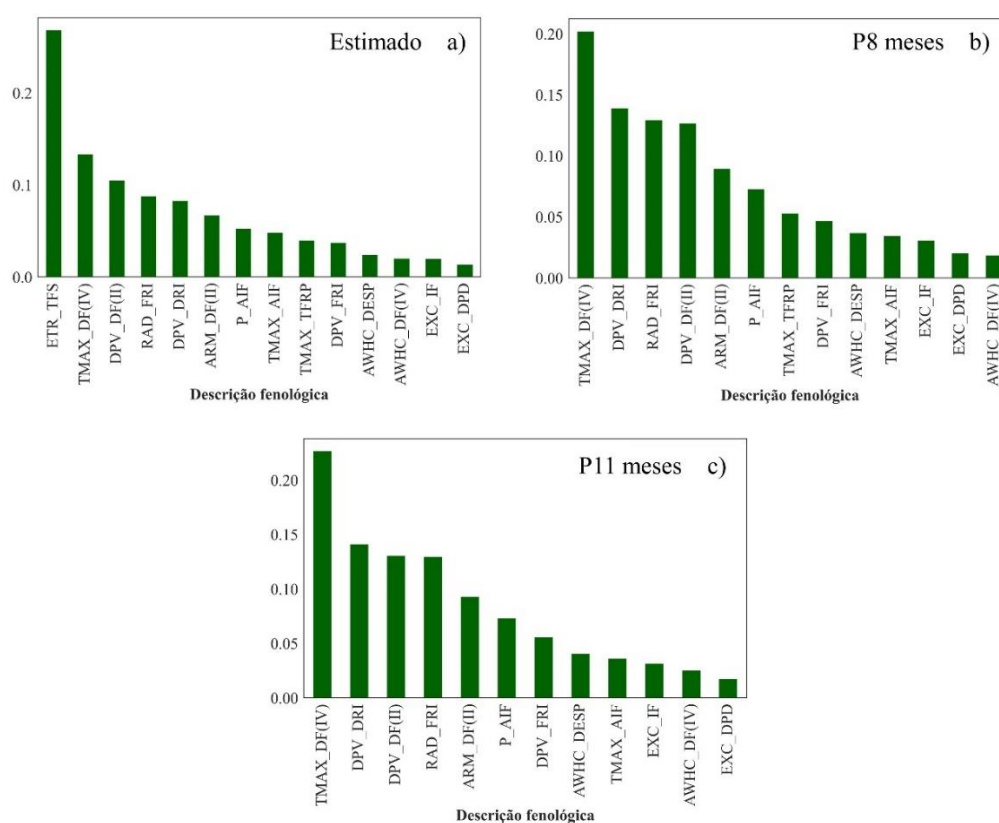


Figura 12. Análise de importância das variáveis selecionadas pelos modelos de machine learning. a) Estimativa de produtividade, b) Previsão de produtividade

com 8 (seis) meses de antecedência, c) Previsão de produtividade com 11 (doze) meses de antecedência.

4.4 DISCUSSÃO

No geral para estimativa e previsão de produtividade de frutos de coco, o RF superou claramente todos os modelos de machine learning testados. Em contrapartida, o MLP apresentou menor desempenho em comparação aos demais modelos.

Os resultados indicaram que o modelo de regressão de RF é capaz de realizar a previsão de produtividade de frutos de coqueiro por hectare em escala local para região de estudo. À medida que se aproxima da colheita, as previsões dos modelos tendem a melhorar ligeiramente e isso se deve ao fato de mais variáveis preditoras serem usadas pelos modelos de previsão, devido agruparem mais estádios fenológicos da planta. Contudo, o desempenho do modelo não apresentou diferença significativa entre suas estimativas e previsões. Sendo assim, previsões de produtividade com 11 meses de antecedência apresentaram desempenho satisfatório, sem efeito substancial com a exclusão de preditores.

Resultados semelhantes para previsão de produtividade por meio de modelos de ML também foram encontrados por Jeong et al. (2015). Os autores avaliaram métodos de ML para prever a produtividade do trigo, milho e batata nos Estados Unidos (EUA). Os resultados obtidos destacaram que o modelo RF previu com sucesso a produção de trigo, milho e batata nos EUA, com desempenho satisfatório na explicação do modelo, apresentando índice de concordância de Willmott's d igual a 0.99, 0.92 e 0.95, respectivamente. Os autores concluíram que RF é um método de ML eficaz e versátil para previsões de produtividade de safras em escalas regionais e globais por sua alta acurácia e precisão, facilidade de uso e utilidade na análise de dados.

Aparecido et al. (2020), avaliando diferentes modelos de ML para prever a incidência de doenças do café como, ferrugem do café, cercospora, bicho mineiro e broca do café, também observaram melhores desempenhos utilizando algoritmos de RF, com precisões médias de R^2 acima de 0.80.

É importante destacar que os modelos de ML separam de forma aleatória um subconjunto de dados para teste de desempenho dos dados de calibração, usando apenas o conjunto de dados restante para o teste do modelo. Portanto, a divisão de dados para fins de treinamento e teste é algo extremamente importante para modelagem de produtividade, e seu desempenho pode aumentar à medida que mais dados são incluídos nos modelos.

Desta forma, cenários de predição com pequenas amostras tornam-se grandes limitadores no desempenho de modelos de ML, como observado para MLP. Onde mesmo que os modelos de MLP tenham boa capacidade de previsão, seus desempenhos sofrem devido ao treinamento de sobreajuste de dados e baixa generalização (Abbasi; El Hanandeh, 2016). Contudo, algoritmos de RF são ferramenta comumente usada em vários modelos de predição, devido à sua alta precisão e capacidade em lidar com problemas complexos usando poucas amostras.

Explicar variações de produtividade na cultura do coqueiro não é algo tão simples e direto (Pham, 2016). Isso se deve principalmente a natureza perene dessa planta que apresenta uma fase reprodutiva prolongada de 44 meses, partindo do início do primórdio floral até a maturidade completa dos frutos, que acontece 11 a 12 meses após abertura de inflorescência (Perera et al., 2010; Ashburner, 1995). Portanto, a fenologia da planta está associada aos elementos meteorológicos como radiação solar, precipitação pluvial, temperatura, influenciando o desenvolvimento do fruto do coco e consequentemente sua produtividade (Jeong et al., 2016).

O aborto de flores femininas e frutos jovens é um fenômeno comum no coco em condições climáticas desfavoráveis e em menor grau mesmo em condições climáticas favoráveis (Ranasinghe et al., 2015).

Ranasinghe, Silva e Premasiri (2015) observaram que a temperatura máxima, intensidade da radiação solar e precipitação pluvial foram consideradas como um dos principais elementos climáticos que podem influenciar a frutificação e o aborto de frutos de coqueiro. Os autores identificaram que os três primeiros meses da abertura de inflorescência é o período mais sensível as variações climáticas, estando correlacionadas a produção de frutos de cocos

mensais. Assim como, temperaturas acima de 33 °C foram relatadas como crítica para o aborto de frutas (Wubs et al., 2009).

O estresse de altas temperaturas durante as fases de desenvolvimento reprodutivo da cultura, pode sinalizar quedas significativas nos componentes de produtividade (Samarasinghe et al., 2018). Ranasinghe, Silva e Premasiri (2015), observaram que o estresse térmico durante um período prolongado nos 3 primeiros meses após abertura de inflorescência do coqueiro, levou a um número muito baixo de frutas por plantas. Enquanto em outros anos, o número de dias reduzido com temperaturas máximas elevadas (>33 °C), o conjunto de frutos por planta foi mais expressivo. Estudos revelaram que aumentos de temperatura influenciam negativamente a polinização, por meio da capacidade de germinação do pólen (Ranasinghe; Silva; Premasiri, 2015), taxa de crescimento do tubo polínico (Snider et al, 2011) e sementes e frutos de plantas (Erickson et al., 2002).

Amarasinghe et al. (2016) em seu estudo, observaram que a porcentagem de germinação do pólen e o comprimento do tubo polínico, tiveram efeito significativos na redução de produção de frutos de coco em temperaturas máximas superiores a 33 °C em comparação a temperaturas de 28 °C. Esses resultados corroboram com as correlações negativas com temperatura máximas obtidas neste trabalho.

Outros estudos como de Nainanayake et al. (2008) revelaram que maioria dos abortos de flores e frutas foi observada em plantações de coco nas zonas secas e intermediárias do Sri Lanka. As possíveis razões para isso incluem a diminuição da qualidade do pólen e das flores femininas devido à exposição das flores a clima extremamente seco (Pathmeswaran et al., 2018).

O coqueiro apresenta um padrão de crescimento indeterminado e geralmente produz uma inflorescência em cada axila da folha em intervalos que variam de 25 a 30 dias, dependendo das condições ambientais e da idade da palma (Ranasinghe et al., 2015). Patel (1938) observou que a os primórdios florais da planta sofrem abortos quando submetidos a secas severas. Menon e Pandalai (1958) notou que um período de seca ocorrendo 15-16 meses antes da abertura da inflorescência leva ao aborto de espatas do coqueiro.

Passos (1998), observaram que estiagem prolongada pode destruir os primórdios florais do coqueiro, causando abortamento da inflorescência e, conseqüentemente, redução na produção de frutos, 28 a 30 meses após a estiagem severa.

Outros estudos avaliando o acúmulo de déficits de água durante os meses do ano, concluíram que as maiores correlações com baixas produtividades foram observadas no início do primórdio floral e diferenciação do ovário (Coomans, 1975). No entanto, deve-se destacar que eventos extremos de precipitação nem sempre reflete em ganhos de produtividade (Passos et al., 2009). Como exemplo, Vijayaraghavan et al. (1988) observaram que precipitação muito alta interfere na polinização, afetando assim a produtividade de frutos de coco.

A radiação solar é outro elemento climático que afeta a biologia floral do coqueiro. Kumar (2011), observou que a iniciação da espata e a produção de coco no Sri Lanka são mais expressivas, quando a duração média do dia é maior do que durante o resto do ano, ofertando mais radiação para planta. Assim como maiores correlações positiva entre a taxa de radiação 29 a 30 meses antes da colheita, com maiores flores femininas produzidas.

4.5 CONCLUSÃO

O modelo RF apresentou melhores desempenhos na capacidade de prever as respostas produtivas do coqueiro em função da fenologia da planta e variáveis agrometeorológicas.

Foi possível obter modelos preditivos com a até 11 meses antes da colheita, com R^2 de 0,68 e MAPE de aproximadamente 11%. A ETR_TFS, TMAX_DF(IV) e DPV_DF(II), foram as variáveis mais importantes para a previsão da produtividade de frutos de coco.

A disponibilidade de variáveis agrometeorológicas e a caracterização da fenologia da cultura, possibilita a calibração de modelos robustos com rotinas baseadas em Inteligência artificial.

4.6 REFERÊNCIAS

Abbasi M, El Hanandeh A (2016) Forecasting municipal solid waste Generation using artificial intelligence modelling approaches. *Waste Manag.* 56, 13e22.

Abid R, Sharmeen S, Praveen A (2007) Stomata type of monocots within the flora of Karachi, Pakistan. *Pak. Bot.* 39, 15–21.

Amarasinghe KGAPK, Ranasinghe CS, Abeysinghe DC, Perera AAFLK (2016) Early fruit set in dwarf x tall crosses of coconut (*Cocos nucifera* L.) under reciprocal pollination between heat and drought stressed and non stressed flowers. *International Journal of Agriculture, Forestry and Plantation*, 3, 1-9.

Antanasijević D, Antanasijević J, Pocajt V (2018) Prediction of the transition temperature of bent-core liquid crystals using fuzzy “digital thermometer” model based on artificial neural networks. *Engineering Applications of Artificial Intelligence*, 71, 251-258. DOI:<https://doi.org/10.1016/j.engappai.2018.03.009>.

Aparecido LEO, Rolim GS, Moraes JRSC, Costa CT, Souza PS (2020) Machine learning algorithms for forecasting the incidence of *Coffea arabica* pests and diseases. *International Journal of Biometeorology*, 64:671–688. DOI:<https://doi.org/10.1007/s00484-019-01856-1>.

Ashburner GR (1995) Reproductive biology of coconut palms. C. Oropeza et al (eds.), *Lethal Yellowing: Research and Practical Aspects*, p. 111-121.

Battisti R, Sentelhas PC, Boote KJ (2017) Inter-comparison of performance of soybean crop simulation models and their ensemble in southern Brazil. *Field Crops Research*, 200, 28-37.

Breiman L (2001) Random forests. *Mach. Learn.* 45, 5–32.

Camargo AP (1971) Balanço hídrico no Estado de São Paulo. Campinas: Instituto Agrônômico, 24p. Boletim 116.

Charoen-Ung P, Mittrapiyanuruk P (2019) Sugarcane Yield Grade Prediction Using Random Forest with Forward Feature Selection and Hyper-parameter Tuning. International Conference on Computing and Information Technology, 769, 33–42. DOI:https://doi.org/10.1007/978-3-319-93692-5_4.

Coomans P (1975) Influence des facteurs climatiques sur les fluctuations saisonnières et annuelles de la production du cocotier. Oléagineux, Paris, 30, 153-159.

Crane-Droesch A (2018) Machine learning methods for crop yield prediction and climate change impact assessment in agriculture. Environ. Res. Lett. 13, 114003. DOI:<https://doi.org/10.1088/1748-9326/aae159>.

Cuenca MAG, Ferreira JMS, Martins CR (2018) Importância e evolução da cocoicultura no Brasil. Embrapa, Brasília, 3. Ed. p. 25-56.

Sousa DP (2020) Estimativa do balanço de energia e das necessidades termohídricas do açaizeiro irrigado no leste da Amazônia. 107 f. Tese de doutorado – UFRA, Belém.

Erickson NA, Markhart AH (2002) Flower developmental stage and organ sensitivity of bell pepper (*Capsicum annum* L.) to elevated temperature. Plant Cell Environ, 25, 123-130.

FAO (2021). FAOSTAT Data: Production – Crops. Food and Agriculture Organization of the United Nations (FAO) (Accessed 07 February 2021). <http://www.fao.org/faostat/en/#data/QC>.

Faria RT, Caramon PH, Chibana EY, Brito LRS, Nakamura AK, Ferreira AR (2002) CLIMA – programa computacional para organização de análises de dados meteorológicos. IAPAR, Boletim.

Ferreira JMS, Warwick DRN, Siqueira LA (2018). A cultura do coqueiro no Brasil. Embrapa, Brasília, DF, p. 508.

Forthofer RN, Lee ES, Hernandez M (2007) Linear Regression. Biostatistics (Second Edition), p. 349-386.

Gomes FP, Oliva, MA, Mielke MS, Almeida AAF, Leite HG (2006) Photosynthetic irradiance-response in leaves of dwarf coconut palm (*Cocos nucifera* L. 'nana', Arecaceae): Comparison of three models. *Scientia Horticulturae*. 109, 101-105. DOI:<https://doi.org/10.1016/j.scienta.2006.02.030>.

Gunn BF, Baudouin L, Olsen KM (2011) Independent origins of cultivated coconut(*Cocos nucifera* L.) in the old world tropics. *PLoS One* 6, e21143.

Raveendran TS, Ramanathan T (1988) Influence of Weather factors on yield of rainfed coconut. *Indian Cocon. J.* 18:7-9.

Hariharan R (2021). Random forest regression analysis on combined role of meteorological indicators in disease dissemination in an Indian city: A case study of New Delhi. *Urban Climate*, 36, 100780. DOI: <https://doi.org/10.1016/j.uclim.2021.100780>.

IBGE (2021). Instituto Brasileiro de Geografia e estatística. Sistema IBGE de Recuperação Automática (Accessed 05 February 2021). <https://sidra.ibge.gov.br/home/pms/brasil>.

Jeong JH, Resop JP, Mueller ND, Fleisher DH, Yun K, Butler EE, Timlin DJ, Shim KM, Gerber JS, Reddy VR, Kim SH (2016) Random Forests for Global and

Regional Crop Yield Predictions. PLoS ONE 11(6): e0156571. <https://doi.org/10.1371/journal.pone.0156571>.

Klompenburg TV, Kassahun A, Catal C (2020). Crop yield prediction using machine learning: A systematic literature review. *Computers and Electronics in Agriculture*, 177, 105709. <https://doi.org/10.1016/j.compag.2020.105709>.

Kuma SN, Aggarwal PK (2013). Climate change and coconut plantations in India: Impacts and potential adaptation gains. *Agricultural Systems* 117, 45–54. <https://doi.org/10.1016/j.agry.2013.01.001>

Kumar KKN (2011) Coconut Phenology and Yield Response to Climate Variability and Change. Department of Atmospheric Sciences, Cochin University Of Science And Technology Kochi, India. p.241.

Liakos K, Busato P, Moshou D, Pearson S, Bochtis D (2018) Machine learning in agriculture: a review. *Sensors* 18, 2674. DOI: <https://doi.org/10.3390/s18082674>.

Lopez-Lozano R, Duveiller G, Seguini L, Meroni M, Garcia-Condado S, Hooker J, Leo O, Baruth B (2015) Towards regional grain yield forecasting with 1 km resolution EO biophysical products: strengths and limitations at pan-European level. In: *Agricultural and Forest Meteorology*, 206, 112–32. DOI: <https://doi.org/10.3390/s18082674>.

Luciano ACS, Picoli MCA, Duft DG, Rocha JV, Leal MRLV, Maire GL (2021) Empirical model for forecasting sugarcane yield on a local scale in Brazil using Landsat imagery and random forest algorithm. *Computers and Electronics in Agriculture* 184, 106063. DOI: <https://doi.org/10.1016/j.compag.2021.106063>.

Manjunathan N, Rajesh P, Thangadurai E, Suresh A (2020) Crop Yield Prediction Using Linear Support Vector Machine. *European Journal of Molecular & Clinical Medicine*, 6, 1-7.

Martins CR, Jesus Júnior La (2014) Produção e comercialização de coco no Brasil frente ao comércio internacional: panorama 2014. Aracaju. Embrapa Tabuleiros Costeiros, p.51.

Martorano LG, Pereira LC, Cezar EGM, Pereira ICB (1993). Estudos climáticos do Estado do Pará, classificação climática (Köppen) e deficiência hídrica (Thorntwhite, Matter). Belém, SUDAM/EMBRAPA, SNLCS pp. 53.

Menon KPV, Pandalai KM (1958) The coconut palm; A Monograph. Published by Dr. Gregory, P.J., Indian Central Coconut Committee, Times of India Press, Bombay, India.

Miles J (2014) Tolerance and Variance Inflation Factor. Wiley StatsRef: Statistics Reference Online.

Nainanayake NPAD, Ranasinghe CS, Tennakoon AT (2008) Effect of drip irrigation on canopy and soil temperature, leaf gas exchange, flowering and nut setting of mature coconut (*Cocos nucifera* L.). Journal of the National Science Foundation of Sri Lanka 36(1): 47 – 54. DOI: <https://doi.org/10.4038/jnsfsr.v36i1.131>.

Oliveira MF, Santos AF, Kazama EH, Rolim GS, Silva RP (2021) Determination of application volume for coffee plantations using artificial neural networks and remote sensing. Computers and Electronics in Agriculture 184, 106096. DOI: <https://doi.org/10.1016/j.compag.2021.106096>.

Oribe-Garcia I, Kamara-Esteban O, Martin C, Macarulla-Arenaza AM, Alonso-Vicario A (2015) Identification of influencing municipal characteristics regarding household waste generation and their forecasting ability in Biscay. Waste Manag. 39, 26e34.

Pant J, Pant RP, Singh MK, Singh DP, Pant H (2021). Analysis of agricultural crop yield prediction using statistical techniques of machine learning. Materials Today: Proceedings. DOI: <https://doi.org/10.1016/j.matpr.2021.01.948>.

Passos EEM, Castro CP, Silva AAG, Nogueira LC. Coqueiro. In: Monteiro JEBA (2009). (Org) Agrometeorologia dos cultivos: o fator meteorológico na produção agrícola. Brasília, DF: Inmet, p. 394-406.

Passos EEM (1998) Morfologia do coqueiro. In: FERREIRA, J. M. S.; WARWICK, D. R. N.; SIQUEIRA, L. A. A cultura do coqueiro no Brasil. Brasília, DF, Embrapa Serviço de Produção e Informação, p. 57-64.

Patel JS (1938) The Coconut: A Monograph. Government Press, Madras.

Pathmeswaran C, Lokupitiya E, Waidyarathne KP, Lokupitiya RS (2018). Impact of extreme weather events on coconut productivity in three climatic zones of Sri Lanka. European Journal of Agronomy 96, 47–53. DOI: <https://doi.org/10.1016/j.eja.2018.03.001>

Paudel D, Boogaard H, Wit A, Janssen S, Osinga S, Pylaniadis C, Athanasiadis NI (2021). Agricultural Systems 187, 103016. DOI: <https://doi.org/10.1016/j.agry.2020.103016>.

Perera PIP, Hoher V, Weerakoon LK, Yakandawala DMD, Fernando SC, Verdeil JL (2010) Early inflorescence and floral development in *Cocos nucifera* L. (Arecaceae: Arecoideae). South African Journal of Botany 76, 482–492.

Pham LJ (2016) Coconut (*Cocos nucifera*). Industrial Oil Crops, p. 231-242. DOI: <https://doi.org/10.1016/B978-1-893997-98-1.00009-9>.

Ranasinghe CS, Silva LRS, Premasiri RDN (2015) Major determinants of fruit set and yield fluctuation in coconut (*Cocos nucifera* L.). *J.Natn.Sci.Foundation Sri Lanka*, 43 (3): 253-264. DOI: <https://doi.org/10.4038/jnsfsr.v43i3.7955>.

Rolim GS, Aparecido LEO, Souza PS, Lamparelli RAC, Santos ER (2020). Climate and natural quality of *Coffea arabica* L. drink. *Theoretical and Applied Climatology*. 141, 87–98. DOI: <https://doi.org/10.1007/s00704-020-03117-3>141:87–98.

Sack S, Abom M (2020) Acoustic plane-wave decomposition by means of multilayer perceptron neural networks. *Journal of Sound and Vibration* 486, 115518. DOI: <https://doi.org/10.1016/j.jsv.2020.115518>.

Samarasinghe CRK, Meegahakumbura MK, Dissanayaka HDMAC, Kumarathunge D, Perera L (2018) Variation in yield and yield components of different coconut cultivars in response to within year rainfall and temperature variation. *Scientia Horticulturae*, 238, 51–57. DOI: <https://doi.org/10.1016/j.scienta.2018.03.058>.

Santos HO, Howell S, Earnest CP, Teixeira FJ (2019) Coconut oil intake and its effects on the cardiometabolic profile – A structured literature review. *Progress in Cardiovascular Diseases* 62, 436–443. DOI: <https://doi.org/10.1016/j.pcad.2019.11.001>.

Snider JL, Oosterhuis DM, Kawakami EM, Loka DA (2011) High temperature limits in vivo pollen tube growth rates by altering diurnal carbohydrate balance in field-grown *Gossypium hirsutum* pistils. *Journal of Plant Physiology*, 168, 1168-1175.

Surre C (1968) Les besions en eau du palmier à huile. Calcul du bilan de l'eau et ses applications pratiques. *Oléagineux*. 23, 165 – 167.

Thornthwaite CW, Mather Jr (1955) The water balance. Centerton: Drexel Institute of Technology. 104 p. (Publications in Climatology, v. 8, n. 1).

Wubs AM, Heuvelink E, Marcelis LFM (2009) Abortion of reproductive organs in sweet pepper (*Capsicum annum* L.): a review. *Journal of Horticultural Science and Biotechnology* 84(5): 467 – 475.

Xu X, Gao P, Zhu X, Guo W, Ding J, Li C, Wu X (2019) Design of an integrated climatic assessment indicator (ICAI) for wheat production: a case study in Jiangsu Province, China. *Ecol. Ind.* 101, 943–953. DOI: <https://doi.org/10.1016/j.ecolind.2019.01.059>.