

UNIVERSIDADE ESTADUAL PAULISTA – UNESP

CÂMPUS DE JABOTICABAL

**ESTIMAÇÃO E PREVISÃO DE PRODUTIVIDADE DE
SOJA POR REDES NEURAIS NO MATOPIBA**

Valter Barbosa dos Santos

Engenheiro Agrônomo

2020

UNIVERSIDADE ESTADUAL PAULISTA – UNESP

CÂMPUS DE JABOTICABAL

**ESTIMAÇÃO E PREVISÃO DE PRODUTIVIDADE DE
SOJA POR REDES NEURAIS NO MATOPIBA**

Valter Barbosa dos Santos

Orientador: Prof. Dr. Glauco de Souza Rolim

Dissertação apresentada à Faculdade de Ciências Agrárias e Veterinárias – Unesp, Câmpus de Jaboticabal, como parte das exigências para a obtenção do título de mestre em Agronomia (Ciência do Solo).

2020

S237e

Santos, Valter Barbosa dos

Estimação e previsão de produtividade de soja por redes neurais no MATOPIBA / Valter Barbosa dos Santos. -- Jaboticabal, 2020

70 p. : tabs., mapas

Dissertação (mestrado) - Universidade Estadual Paulista (Unesp), Faculdade de Ciências Agrárias e Veterinárias, Jaboticabal

Orientador: Glauco de Souza Rolim

1. Inteligência artificial. 2. Soja. 3. Produtividade agrícola. 4. Python (Linguagem de programação de computador). 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).

Essa ficha não pode ser modificada.

CERTIFICADO DE APROVAÇÃO


TÍTULO DA DISSERTAÇÃO: ESTIMAÇÃO E PREVISÃO DE PRODUTIVIDADE DE SOJA POR REDES NEURAIS NO MATOPIBA

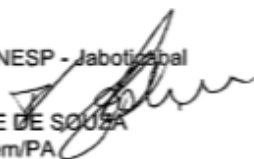
AUTOR: VALTER BARBOSA DOS SANTOS

ORIENTADOR: GLAUCO DE SOUZA ROLIM

Aprovado como parte das exigências para obtenção do Título de Mestre em AGRONOMIA (CIÊNCIA DO SOLO), pela Comissão Examinadora:


Prof. Dr. GLAUCO DE SOUZA ROLIM
Departamento de Ciências Exatas / FCAV / UNESP - Jaboticabal


Prof. Dr. ROGÉRIO TEIXEIRA DE FARIA
Departamento de Engenharia Rural / FCAV / UNESP - Jaboticabal


Prof. Dr. PAULO JORGE DE OLIVEIRA PONTE DE SOUZA
Universidade Federal Rural da Amazônia / Belém/PA

Jaboticabal, 03 de fevereiro de 2020

DADOS CURRICULARES DO AUTOR

Valter Barbosa dos Santos- Filho de José Carlos dos Santos e Severina Barbosa (In memoriam). Nasceu em São José de Ribamar, Maranhão, no dia 22 de fevereiro de 1990. Técnico em Agropecuária pela Escola Agrotécnica Federal do Maranhão (2009), cursou engenharia Agrônômica na Universidade Estadual do Maranhão-UEMA, câmpus Paulo VI, de São Luís-MA, de 2010 a 2017. Bolsista de iniciação científica pela PIBIC/UEMA em 2014-2015, atuando principalmente nos seguintes temas: manejo, potássio, adubação orgânica, agricultura familiar, Fertilidade do Solo. Trabalhou como Servidor público, concursado no cargo de Fiscal Ambiental (2013-2018). Em março de 2018, ingressou no Curso de Mestrado em Agronomia (Ciência do Solo), na Faculdade de Ciências Agrárias e Veterinárias – UNESP. É integrante do grupo de pesquisa: “Group of Agrometeorological Studies” (GAS), da Unesp – Câmpus de Jaboticabal. Em fevereiro de 2020, submeteu-se à banca para a defesa de Dissertação.

DEDICO

A Deus, pela dádiva da vida pois sem ele eu não teria traçado o meu caminho e feito a minha escolha pela Agronomia.

Aos meus pais, José Carlos e Severina Barbosa **(In memoriam)** pelo apoio e confiança, mesmo distantes. A minha esposa e filho pela companhia nessa jornada.

OFEREÇO

À minha família, pelo incentivo e força nessa caminhada.

À professora Ana Maria Araújo da Silva, por ter acreditado no meu potencial desde 2014, despertando o lado científico das ciências agrônômicas através das orientações em projetos de iniciação científica.

AGRADECIMENTOS

Agradeço primeiramente a **Deus**, pela dádiva da vida pois sem ele eu não teria traçado o meu caminho e feito a minha escolha pela Agronomia.

A minha **família** pelo amor incondicional, especialmente meus pais e meus irmãos, acreditando, apoiando e confiando em mim mesmo distantes, aos meus sobrinhos e a toda família Moreno e Barbosa dos Santos.

A minha esposa e filho, Aline e Samuel pelo amor, companheirismo e amizade, e principalmente por me acompanharem nessa jornada.

Ao meu orientador professor Glauco de Souza Rolim, por sua amizade, dedicação pelos seus ensinamentos que me possibilitou realizar mais essa etapa da minha vida.

A professora Dr^a. Ana Maria Silva de Araújo, pelos ensinamentos, amizade, disposição, e paciência, por ter acreditado no meu potencial desde 2014 quando começamos a trabalhar juntos, me despertando o lado científico das ciências agrônomicas através das orientações em projetos de iniciação científica entre os anos de 2014 a 2015.

Aos professores Rogério Teixeira de Faria e Paulo Jorge de Oliveira Ponte de Souza membros da banca de qualificação pelas importantes sugestões.

A meus amigos e parceiros científicos, José Reinaldo da Silva Cabral de Moraes, Tayanara Tuany Borges Valeriano, Mary Jane Nunes Carvalho, Kárita Almeida Silva, Paulo Alexandre da Silva e Kamila Cunha de Meneses.

Ao Grupo de Pesquisa em Agrometeorologia da Unesp – GAS, pelo recebimento no grupo e pelos conhecimentos compartilhados.

A meus amigos e colegas de departamento, Gustavo André, Thiago Águas, pela disponibilidade em ajudar sempre que foi necessário.

Aos funcionários do Departamento de Ciências Exatas, Maria José Servidone Trizólio, Shirley Aparecida Martineli de Sousa, Adriana Elisabete Takakura, por me receberem bem no departamento, e pelo carinho.

Ao programa de Pós-Graduação em Agronomia (Ciência do Solo), pela oportunidade em cursar o mestrado.

À Fundação de Amparo à Pesquisa e ao Desenvolvimento Científico e Tecnológico do Maranhão - FAPEMA, pela concessão da bolsa de mestrado no país e, pelo apoio para o desenvolvimento da pesquisa.

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

SUMÁRIO

| | |
|--|-------------|
| RESUMO | x |
| ABSTRACT | xi |
| LISTA DE ABREVIATURAS E SIGLAS | xii |
| LISTA DE FIGURAS | xiii |
| LISTA DE TABELAS | xvi |
| CAPÍTULO 1 – Considerações gerais | 1 |
| 1.1 Introdução | 1 |
| 1.2 Revisão de Literatura | 2 |
| 1.2.1 Importância econômica do MATOPIBA..... | 2 |
| 1.2.2 Cultivo da Soja..... | 4 |
| 1.2.3 Importância da modelagem..... | 6 |
| 1.2.3.1 Modelagem | 7 |
| 1.2.3.2 Redes neurais artificiais (RNAs) | 9 |
| REFERÊNCIAS | 11 |
| CAPITULO 2 – ESTIMAÇÃO E PREVISÃO DE PRODUTIVIDADE DE SOJA POR REDES NEURAIIS | 19 |
| Resumo | 19 |
| Abstract | 20 |
| 2.1 Introdução | 21 |
| 2.2 Material e Métodos | 24 |
| 2.2.1 Localização e caracterização da área de estudo | 24 |
| 2.2.2 Seleção dos locais de estudo | 25 |
| 2.2.3 Dados Meteorológicos | 27 |
| 2.2.4 Balanço Hídrico | 28 |
| 2.2.5 Análise de dados | 28 |
| 2.2.6 Avaliação do modelo | 31 |
| 2.3 Resultados e Discussão | 32 |
| 2.4 Conclusões | 49 |
| Referências | 50 |

ESTIMAÇÃO E PREVISÃO DE PRODUTIVIDADE DE SOJA POR REDES NEURAIS NO MATOPIBA

RESUMO: A demanda por alimento se torna cada vez maior e para atendê-la em tempo hábil e quantidade suficiente, utilizamos redes neurais artificiais para previsão de produtividade. Dentre essas técnicas, a mais comumente empregada para prever a produtividade da soja é a Rede Neural Artificial. O objetivo dessa pesquisa foi estimar e prever a produtividade de soja utilizando redes neurais artificiais e informações climáticas mensais da temperatura do ar, precipitação, radiação global, e componentes do balanço hídrico como evapotranspiração de cultivo, armazenamento, evapotranspiração real de cultivo, deficiência e excedentes hídricos durante o ciclo do cultivo para os estados do Maranhão, Tocantins, Piauí e Bahia, que juntos formam a fronteira agrícola do MATOPIBA no Brasil, produtividade média para a região durante o período avaliado é de 2.575 kg ha⁻¹. Os resultados mostram que a RNA ajustada para estimação apresentou erros de até 100 kg ha⁻¹ (sub e superestimando) em grande parte (27) das localidades, 10 municípios apresentam produtividade subestimada, e 15 apresentam resultados superestimados. Enquanto que a RNA ajustada para previsão mostrou que 30 localidades apresentam produtividades com erros até 100 kg ha⁻¹, 8 e 14 municípios com produtividade subestimada e superestimada, respectivamente.

Palavras – Chave: inteligência artificial, balanço hídrico, soja, produtividade agrícola

ESTIMATION AND FORECAST OF SOYBEAN YIELD BY NEURAL NETWORKS IN MATOPIBA

ABSTRACT – The demand for food becomes increasing and to meet it in a timely and sufficient amount, we use artificial neural networks to predict productivity. Among these techniques, the most commonly used to predict soybean yield is the Artificial Neural Network. The objective of this research was to estimate and predict soybean yield using artificial neural networks and monthly climatic information of air temperature, precipitation, global radiation, and water balance components such as crop evapotranspiration, storage, real crop evapotranspiration, deficiency and water surpluses during the cultivation cycle for the states of Maranhão, Tocantins, Piauí and Bahia, which together form the agricultural frontier of MATOPIBA in Brazil, average productivity for the region during the period evaluated is 2,575 kg ha⁻¹. The results show that the RNA adjusted for estimation presented errors of up to 100 kg ha⁻¹ (under and overestimation) in large part (27) of the localities, 10 municipalities presented underestimated productivity, and 15 presented overestimated results. While the RNA adjusted for forecasting showed that 30 localities present productivity with errors up to 100 kg ha⁻¹, 8 and 14 municipalities with underestimated and overestimated productivity, respectively.

Keywords: Artificial Intelligence, water balance, soybeans, agricultural productivity

LISTA DE ABREVIATURAS E SIGLAS

ML – Machine learning

RNAs – Redes Neurais Artificiais

DNNs – Redes Neurais Profundas

CNNs- Redes Neurais Convolucionais

LSTMs - Long Short Term Memory

MLP - Multilayer Perceptron

GEE - Google Earth Engine

RUE - Eficiência no uso da radiação

T – Temperatura °C – Graus

Celsius

P – Precipitação

AWC – Capacidade de água disponível

CET – Evapotranspiração da cultura

AET – Evapotranspiração real

STO – Armazenamento

DEF – Déficit

EXC – Excedente

BH – Balanço Hídrico

CV – Cros - validation

LISTA DE FIGURAS

| | Página |
|--|---------------|
| Figura 1 (Capítulo 1) - Rede neural perceptron de múltiplas camadas..... | 22 |
| Figura 1 (Capítulo 2) - Localização da área de estudo, A) Brasil, B) Estados que compõem a Fronteira agrícola do MATOPIBA e C) Delimitação territorial do MATOPIBA. | 38 |
| Figura 2 (Capítulo 2) - Representação dos dados na camada de entrada (input) relativo a um ano | 42 |
| Figura 3 (Capítulo 2) - Esquema das Features para as Redes Neurais Artificiais para estimação (A) e previsão (B) de produtividade de soja para o MATOPIBA. | 43 |
| Figura 4 (Capítulo 2) - Precipitação dos meses de A) novembro, B) dezembro, C) janeiro, D) fevereiro, E) março e F) precipitação acumulada do período. 1) Anapurus-MA, 2) Sucupira do Norte-MA, 3) Balsas-MA, 4) Estreito-MA, 5) Guaraí-TO, 6) Alvorada-TO, 7) Dianópolis-TO, 8) Porto Nacional-TO, 9) B Jesus-PI, 10) Sebastião Leal-PI, 11) Alto Parnaíba-MA, 12) Formosa do Rio Preto-BA, 13) São Desidério-BA..... | 46 |
| Figura 5 (Capítulo 2) - Temperatura dos meses de A) novembro, B) dezembro, C) janeiro, D) fevereiro, E) março e F) temperatura média do período. 1) Anapurus-MA, 2) Sucupira do Norte-MA, 3) Balsas-MA, 4) Estreito-MA, 5) Guaraí-TO, 6) Alvorada-TO, 7) Dianópolis-TO, 8) Porto Nacional-TO, 9) B Jesus-PI, 10) Sebastião Leal-PI, 11) Alto Parnaíba-MA, 12) Formosa do Rio Preto-BA, 13) São Desidério-BA..... | 47 |
| Figura 6 (Capítulo 2) - Evapotranspiração da cultura dos meses de A) novembro, B) dezembro, C) janeiro, D) fevereiro, E) março e F) evapotranspiração media da cultura no período. 1) Anapurus-MA, 2) Sucupira do Norte-MA, 3) Balsas-MA, 4) Estreito-MA, 5) Guaraí-TO, 6) Alvorada-TO, 7) Dianópolis-TO, 8) Porto Nacional-TO, 9) B Jesus-PI, 10) Sebastião Leal-PI, 11) Alto Parnaíba-MA, 12) Formosa do Rio Preto-BA, 13) São Desidério-BA | 48 |
| Figura 7 (Capítulo 2) - Evapotranspiração real dos meses de A) novembro, B) dezembro, C) janeiro, D) fevereiro, E) março e F) evapotranspiração média do período. 1) Anapurus-MA, 2) Sucupira do Norte-MA, 3) Balsas-MA, 4) Estreito-MA, 5) Guaraí-TO, 6) Alvorada-TO, 7) Dianópolis-TO, 8) Porto Nacional-TO, 9) B Jesus-PI, 10) Sebastião Leal-PI, 11) Alto Parnaíba-MA, 12) Formosa do Rio Preto-BA, 13) São Desidério-BA..... | 49 |

Figura 8 (Capítulo 2) - Deficiência hídrica nos meses de A) novembro, B) dezembro, C) janeiro, D) fevereiro, E) março e F) Deficiência hídrica acumulada do período. 1) Anapurus-MA, 2) Sucupira do Norte-MA, 3) Balsas-MA, 4) Estreito-MA, 5) Guaraí-TO, 6) Alvorada-TO, 7) Dianópolis-TO, 8) Porto Nacional-TO, 9) B Jesus-PI, 10) Sebastião Leal-PI, 11) Alto Parnaíba-MA, 12) Formosa do Rio Preto-BA, 13) São Desidério-BA..... 50

Figura 9 (Capítulo 2) - Armazenamento de água no solo nos meses de A) novembro, B) dezembro, C) janeiro, D) fevereiro, E) março e F) armazenamento médio do período. 1) Anapurus-MA, 2) Sucupira do Norte-MA, 3) Balsas-MA, 4) Estreito-MA, 5) Guaraí-TO, 6) Alvorada-TO, 7) Dianópolis-TO, 8) Porto Nacional-TO, 9) B Jesus-PI, 10) Sebastião Leal-PI, 11) Alto Parnaíba-MA, 12) Formosa do Rio Preto-BA, 13) São Desidério-BA..... 51

Figura 10 (Capítulo 2) - Excedente hídrico nos meses de A) novembro, B) dezembro, C) janeiro, D) fevereiro, E) março e F) excedente médio do período. 1) Anapurus-MA, 2) Sucupira do Norte-MA, 3) Balsas-MA, 4) Estreito-MA, 5) Guaraí-TO, 6) Alvorada-TO, 7) Dianópolis-TO, 8) Porto Nacional-TO, 9) B Jesus-PI, 10) Sebastião Leal-PI, 11) Alto Parnaíba-MA, 12) Formosa do Rio Preto-BA, 13) São Desidério-BA..... 52

Figura 11 (Capítulo 2) - Correlação entre as variáveis meteorológicas mensais e a produtividade de soja para o município de: A) Anapurus – MA com clima seco e B) Alvorada – TO com clima úmido 53

Figura 12 (Capítulo 2) - Desempenho do modelo de Redes Neurais Profundas (DNNs) para estimação de produtividade de soja para o MATOPIBA..... 56

Figura 13 (capítulo 2) – A) Produtividade observada (OBS), B) produtividade estimada (EST) e C) diferença entre a produtividade estimada e observada. 1) Anapurus-MA, 2) Sucupira do Norte-MA, 3) Balsas-MA, 4) Estreito-MA, 5) Guaraí-TO, 6) Alvorada-TO, 7) Dianópolis-TO, 8) Porto Nacional-TO, 9) B Jesus-PI, 10) Sebastião Leal-PI, 11) Alto Parnaíba-MA, 12) Formosa do Rio Preto-BA, 13) São Desidério-BA..... 57

Figura 14 (Capítulo 2) - Distribuição de Erros de estimação por redes neurais profundas para a região do MATOPIBA. Observação: as cores correspondem a escala da figura 13-C. 58

Figura 15 (capítulo 2) – Ajuste do modelo de Redes Neurais Profundas (DNNs) para previsão de produtividade de soja. 59

Figura 16 (Capítulo 2) - A) Produtividade observada, B) produtividade predita e C) diferença entre a produtividade predita e observada. 1) Anapurus-MA, 2) Sucupira do Norte-MA, 3) Balsas-MA, 4) Estreito-MA, 5) Guaraí-TO, 6) Alvorada-TO, 7) Dianópolis-TO, 8) Porto Nacional-TO, 9) B Jesus-PI, 10) Sebastião Leal-PI, 11) Alto Parnaíba-MA, 12) Formosa do Rio Preto-BA, 13) São Desidério-BA..... 60

Figura 17 (capítulo 2) – Distribuição de Erros de previsão por redes neurais profundas para a região do MATOPIBA. Observação: as cores correspondem a escala da figura 16-C. 61

LISTA DE TABELAS

Página

| | |
|--|-----|
| Tabela 1 (Capítulo 1) - Área (km ² e %) das características predominantes nos municípios do Matopiba com grande participação na produção nacional de soja, milho e algodão em 2012 | 16 |
| Tabela 2 (Capítulo 1) - - Estádios vegetativos e reprodutivos da soja. Adaptado de RITCHIE, 1998 | 19 |
| Tabela 1 (Capítulo 2) – Municípios selecionados conforme critério da série histórica nos estados do Maranhão - MA, Tocantins – TO, Piauí – PI e Bahia BA, seguidos de área plantada em ha e produtividade média da série histórica em kg ha ⁻¹ | 39 |
| Tabela 2 (Capítulo 2) - Coeficiente de cultivo da soja | 41 |
| Tabela 3 (Capítulo 2) – Hiperparâmetros ajustados na Rede Neural Artificial Multilayer Perceptron – RNA-MLP. O MLPRegressor é um método disponível no pacote Scikit-learn (autor, ano) da linguagem Python | 43 |
| Tabela 4 (Capítulo 2) – Hiperparâmetros ajustados e métricas de desempenho dos modelos de Redes Neurais Artificiais para estimação e previsão de produtividade de soja para o MATOPIBA | .55 |

CAPÍTULO 1 – Considerações gerais

1.1 Introdução

A inteligência artificial (IA) é definida como o campo científico que dá às máquinas a capacidade de aprender sem serem estritamente programadas. Este conhecimento poderá causar um efeito disruptivo em diferentes atividades da civilização humana. A IA aplicada em problemas agrícolas é recente, entretanto, já se vislumbra que o potencial preditivo, reconhecimento de padrões como os de imagem, robótica, entre outros, causará também uma transformação da agricultura mundial (Patricio, 2018).

A agricultura de precisão começou a ser desenvolvida na década de 1980 (Finger, 2019) e aborda o desafio de adaptar o gerenciamento das características do local, do cultivo e do ambiente (Lowenberg-DeBoer, 2015) lidando com heterogeneidades de áreas produtivas.

A união da IA com a agricultura de precisão define o que hoje é chamado de agricultura digital, que surge como um novo campo científico utilizando grande quantidade de dados, auxiliando o planejamento e decisões estratégicas para impulsionar a produtividade agrícola minimizando seu impacto ambiental (Liakos, 2018).

Uma das técnicas da inteligência artificial é o aprendizado de máquina (ML) que surgiu juntamente com as tecnologias de big data e a computação de alto desempenho para criar novas oportunidades para entender, quantificar e estimar/prever diversos processos ou fenômenos (Samuel, 2000).

Dentre as técnicas de ML, as Redes Neurais Artificiais (ANNs) são as mais utilizadas na literatura. As MLs são modelos supervisionados que normalmente são usados para problemas de regressão e classificação, geralmente utilizando a unidade de processamento chamada por Rosenblatt (1958) de 'perceptron'. As ANNs são divididas em duas categorias: ANNs tradicionais, com apenas uma camada escondida, e ANNs profundas, com mais camadas. As RNAs profundas são conhecidas como aprendizado profundo ou redes neurais profundas (DNNs) (Lecun, 2015).

O potencial preditivo possibilitado pelas DNNs poderá causar efeito significativo no melhoramento do monitoramento e das previsões agrícolas, tanto em áreas agrícolas tradicionais como em fronteiras agrícolas como o MATOPIBA

no Brasil. As previsões de produtividade agrícolas têm forte importância relacionadas à especulação de estoques, preços, sustentabilidade e segurança alimentar, regulando os mercados agrícolas, permitindo o planejamento da redução dos impactos ambientais e fazer frente à variabilidade e mudanças climáticas (Gusso et al, 2017).

1.2 Revisão de Literatura

1.2.1 Importância econômica do MATOPIBA

O Brasil é um dos maiores produtores agropecuários do mundo, o setor agrícola foi responsável por 23,5% do produto interno bruto do país em 2017 (IBGE, 2018). A atividade gerou, aproximadamente, 18,2 milhões de empregos diretos em 2018 (CEPEA, 2018). Em termos de comércio exterior, a agropecuária contribuiu com mais de US\$ 97 bilhões em exportações em 2017, exercendo uma grande importância para balança comercial brasileira (MAPA, 2018).

Uma área de expansão agrícola que tem contribuído fortemente com o agronegócio brasileiro é a região do MATOPIBA, acrônimo referente à área de intersecção dos estados do Maranhão, Tocantins, Piauí e Bahia com área de 73.173.485 hectares, correspondendo aproximadamente a 1,3 vezes a área da França, envolvendo 337 municípios. Foi instituído pelo Decreto Presidencial nº 8.447, de 2015, o Plano de Desenvolvimento Agrícola do MATOPIBA (Pereira; Pauli 2016).

Esta fronteira agrícola no bioma cerrado responde por grande parte da produção brasileira de grãos, especialmente a soja (EMBRAPA, 2019). Uma fronteira agrícola é definida como uma região dominada por vegetação natural que começou a enfrentar intensa ocupação da terra relacionada à agricultura (Araújo et al., 2019).

É uma região que inclui características diferenciadas de clima, solo e relevo, que possibilitam com que municípios do MATOPIBA tenham grande participação na produção nacional de soja, milho e algodão (TABELA 1). Os estados pertencentes ao MATOPIBA foram responsáveis por 11% das 115 milhões de toneladas de soja da safra 2017/2018 (Buainain; Buainain; Vieira Filho, 2018). Por essa razão inúmeras pesquisas nos mais diversos campos vêm sendo realizadas no MATOPIBA, como seleção de cultivares adaptadas,

técnicas de plantio e manejo e previsão de produtividade agrícola (MINGOTI et al., 2014).

Estudos como o de Matricardi et al. (2018) fizeram o monitoramento do desmatamento por meio de modelagem utilizando dados de sensoriamento remoto e técnicas de SIG, na região do MATOPIBA. Silva et al. (2020) calibraram e validaram o modelo AquaCrop para simular a produtividade da soja adaptado às condições de solo e clima da região do MATOPIBA e estimaram a produtividade da soja para as condições do clima atual e futuras mudanças climáticas. Silva et al. (2019) pesquisaram a relação da intensidade-duração-frequência de chuvas para análise de vulnerabilidade à erosão com base na erosividade e erodibilidade em áreas agrícolas.

Tabela 1 - Área (km² e %) das características predominantes nos municípios do Matopiba com grande participação na produção nacional de soja, milho e algodão em 2012.

| Biomias | | |
|----------------------------|-------------------------|-----------------|
| Nome do bioma | Área (km ²) | Porcentagem (%) |
| Cerrado | 234.400,65 | 93,89 |
| Transição Cerrado-Caatinga | 7.765,51 | 3,11 |
| Transição Amazônia-Cerrado | 6.838,77 | 2,74 |
| Outros | 659,84 | 0,26 |
| Climas | | |
| Período sem chuva | Área (km ²) | Porcentagem (%) |
| 4 a 5 meses | 222.002,00 | 88,92 |
| 6 meses | 27.662,77 | 11,08 |
| Relevos | | |
| Unidade geomorfológica | Área (km ²) | Porcentagem (%) |
| Chapadas | 123.107,99 | 49,31 |
| Depressões | 84.017,18 | 33,65 |
| Patamares | 18.885,13 | 7,56 |
| Outros | 23.654,47 | 9,47 |
| Solos | | |
| Ordem | Área (km ²) | Porcentagem (%) |
| Latossolo | 126.966,84 | 50,85 |
| Neossolo | 85.729,52 | 34,34 |
| Plintossolo | 22.461,20 | 9,00 |
| Outros | 14.507,21 | 5,81 |

Fonte: Mingoti et al., 2014. Disponível em:

<https://www.alice.cnptia.embrapa.br/alice/bitstream/doc/991059/1/20140721NotaTecnica6.pdf>.

Apesar de toda a atividade econômica, existe grande preocupação na preservação ambiental. O MATOPIBA possui 50 unidades de conservação federais, estaduais e municipais (7,2 milhões de ha) e 23 terras indígenas (3,6 milhões de ha) (Araújo et al., 2019).

O latossolo é o tipo de solo dominante, cobrindo 27,8 milhões de ha (38% da região) (Magalhães e Miranda, 2014). São solos contendo menos de 20% de argila com profundidade de 0,00 a 0,50 m (Santos et al., 2013), com caráter distrófico, pH ácido, geralmente entre 4,8 e 5,2 (Cruz, 2016), altos níveis de intemperismo, baixos teores de fósforo disponível e alto teor de fósforo capacidade de adsorção, causando limitações de fertilidade para produção agrícola (Soares e Alleoni, 2008). No entanto, são solos que podem fornecer altas produtividades desde que feito a correção de pH e fertilização (Gmach et al., 2018).

A produção agrícola no MATOPIBA deverá aumentar cerca de 15 % na próxima década, atingindo 6,25 milhões de toneladas na Safra 2026/2027, correspondendo a aproximadamente 8,5 a 11 milhões de ha de área plantada (BRASIL, 2017). Promovendo vantagens econômicas e sociais para o desenvolvimento para a região (Cruz, 2016) e colaborando para superar a demanda global pela produção de alimentos.

1.2.2 Cultivo da Soja

A soja (*Glycine max* (L.) Merrill) é uma das oleaginosas cultivadas mais importantes no mundo devido a seu potencial produtivo, sua composição química e seu valor nutritivo, conferindo-lhe multiplicidade de aplicações na alimentação humana e animal (Hirakuri et al., 2017). Caracterizada por ter um alto valor nutricional, onde 40% da massa total de sementes são proteínas, 21% de óleo e 34% de carboidratos (Kerry et al., 2018).

Na maioria dos sistemas produtivos, o cultivo tem garantido posição de destaque, com produtividade média de grãos em torno de 3.000 kg ha⁻¹ (CONAB, 2015).

Em uma escala global a soja é a leguminosa mais cultivada no mundo e a quarta produção mais importante após o trigo, milho e arroz (FAO, 2018).

Atualmente três países são responsáveis pela produção global de soja, Estados Unidos, Brasil e Argentina (Yu et al., 2019). A sustentabilidade da produção de soja em todas as regiões do mundo sofre ameaças, o estresse hídrico por exemplo foi identificado como o principal fator ambiental que limita a produção (Zipper et al., 2016).

Nas últimas três décadas foi o cultivo agrícola que mais cresceu no Brasil, alcançando 49% da área plantada com grãos e uma produção da ordem de 114 milhões de toneladas, gerando de forma direta receita superior a 120 bilhões de reais (MAPA, 2014).

Devido à importância econômica desse cultivo, diversas pesquisas são realizadas na busca constante de cultivares adaptadas e mais produtivas, com maior porcentagem de proteína e melhor qualidade de óleo, maior resistência a doenças e pragas e outras características que refletem em maior produtividade. A semeadura por exemplo tem um papel fundamental na lavoura, quando feita corretamente favorece altas produtividades (Fiss et al., 2018).

As cultivares de soja são classificadas quanto ao seu hábito de crescimento (forma e estrutura morfológica) e pelos seus requerimentos em comprimento de dia e temperatura, necessários para iniciar o desenvolvimento floral ou reprodutivo (Ritchie, 1998; Urano et al., 2007).

O ciclo da soja, da germinação à maturação fisiológica, pode variar de 75 a 200 dias, e podem ser classificados em grupos de maturação precoce, semiprecoce, médio e tardio (Silva, 2019). O desenvolvimento da planta é dividido em duas fases, vegetativa (V) e reprodutiva (R) (Tabela 2).

Os elementos climáticos, temperatura, fotoperíodo, e disponibilidade hídrica são os que mais afetam o desenvolvimento e a produtividade da soja. A época de plantio é favorecida pela ocorrência de temperaturas compreendidas entre 20 °C e 30 °C, principalmente às mais próximas de 30 °C, além de uma precipitação de 400 a 800 mm ciclo⁻¹ (Farias et al., 2007).

Tabela 2 - Estádios vegetativos e reprodutivos da soja

| Estádios vegetativos | Estádios reprodutivos |
|------------------------------|--|
| VE - Emergência | R ₁ - Início do florescimento |
| VC - Cotilédone | R ₂ - Pleno florescimento |
| V ₁ - Primeiro nó | R ₃ - Início da formação das vagens |
| V ₂ - Segundo nó | R ₄ - Plena formação das vagens |
| V ₃ - Terceiro nó | R ₅ - Início do enchimento das sementes |
| * | R ₆ - Pleno enchimento das vagens |
| * | R ₇ - Início da maturação |
| V(n) - enésimo nó | R ₈ - Maturação plena |

Adaptado de Ritchie, 1998.

Sendo assim, a melhor compreensão das exigências climáticas do cultivo da soja e das relações da água no sistema solo-planta-atmosfera pode contribuir para redução dos riscos e incertezas da produção dos cultivos (FARIAS et al., 2007).

1.2.3 Importância da modelagem

Com o crescimento populacional há uma intensificação na demanda por alimentos, o que faz com que pesquisadores, empresas e o setor público busquem por novas respostas, conceitos e tecnologias para quantificar as relações da produtividade agrícola com o clima, o solo (Srivastava et al., 2019), as econômicas e de sustentabilidade para o agronegócio (Moreto, 2019).

A partir da agricultura digital ou “smart farming”, tornou-se possível combinar ciência de dados e agricultura na tentativa de relacionar o grande número de dados disponíveis de modo a favorecer na gestão antecipada das atividades agrícolas (Wolfert et al., 2017), políticas nacionais de alimentos e comércio internacional de produtos agropecuários (Kim et al., 2019).

Por esses motivos, o uso de modelos apresenta-se como ferramentas úteis em ambientes dinâmicos e complexos, como agricultura atual e pode ser usado para estimar o impacto de climas atuais e futuros (Van Keulen et al., 2019),

cultivar, data de plantio, tipo de solo sobre a produção agrícola (Dokoohaki et al., 2016) e a segurança alimentar (Ovando, 2018).

1.2.3.1 Modelagem

Um modelo é uma representação matemática de algum fenômeno ou conjunto de fenômenos (sistema), com objetivo de entendê-lo e buscar suas respostas para diferentes problemas. Ao se construir um modelo, faz-se a tentativa de ganho de conhecimento sobre determinado fenômeno que ainda não seja completamente entendido (Manziona et al., 2018).

É cada vez mais comum o uso de modelos que realizam a estimativa de crescimento e que avaliam o desenvolvimento de cultivos, dessa forma contribuindo para previsão da produtividade agrícola e tornando compreensível os fatores que estão envolvidos nas diferentes respostas dos cultivos ao ambiente (Anar et al., 2019).

Os modelos podem ser (i) matemáticos, quando o sistema é descrito por meio de equações, (ii) empíricos ou estatísticos, quando descrevem correlações entre variáveis sem necessariamente se referir ao processo modelado, e (iii) mecanísticos ou explicativos, que expressam as relações de causa e efeito entre as variáveis (Dourado Neto et al., 1998).

A modelagem apresenta-se como uma ferramenta poderosa, por ser considerada uma excelente forma de transferir o conhecimento da pesquisa para o produtor ou profissional da área (Zhao et al., 2019; Sihag e Prakash, 2019). O uso de modelos de previsão de safras por exemplo, podem ajudar a melhorar a qualidade da produção, elevar a produtividade, diminuir os riscos por incidências de pragas e doenças e melhorar o custo de produção (Avadhani et al., 2019).

Existem vários modelos que realizam a estimativa de produtividade agrícola, dentre os mais usados cita-se os modelos mecanísticos DSSAT - Decision Support System for Agrotechnology Transfer (Jones et al., 2003), o APSIM – Agricultural Production System siMulator (Mccown et al., 1996) e o modelo AQUACROP-FAO (Steduto et al., 2009).

Estudos como o de Ovando (2018) mostra a precisão do modelo DSSAT para estimar a produtividade de soja usando dados meteorológicos de satélite,

Balboa et al (2019) usaram o modelo APSIM para avaliar a produtividade em nível de sistema da rotação milho-soja e Adeboye et al (2019), simularam o armazenamento da água no solo, e a produtividade de soja de sequeiro, por meio do modelo AquaCrop.

Existem também os modelos de Inteligência Artificial sendo que os métodos de aprendizagem máquina (machine learning, ML) são os mais comuns. Esses métodos consistem em análise de dados automatizada a partir da construção de modelos gerais (Storm et al., 2019).

Nessa técnica, a máquina (computadores ou software) é programada para aprender padrões a partir dos dados. O aprendizado baseia-se em um conjunto de regras matemáticas e suposições estatísticas, cujo objetivo é desenvolver um modelo baseado em associações estatísticas entre recursos de um determinado conjunto de dados. O modelo ajustado pode ser usado para prever qualquer intervalo de saídas, como respostas binárias, rótulo categóricos ou valores contínuos (Camacho et al., 2018).

Uma das principais vantagens das técnicas de ML é que elas são capazes de resolver autonomamente grandes problemas não lineares usando conjuntos de dados de várias fontes (Chlingaryan et al., 2018)

Técnicas de ML são amplamente utilizadas: Mehra et al. (2016) usaram técnicas de ML como Redes Neurais Artificiais (RNAs), árvores categóricas e de regressão e Florestas Aleatórias (RFs) para abordar o problema de prever o risco de *Stagonospora nodorum blotch* (SNB) no pré-plantio em trigo de inverno. Técnicas de aprendizado de máquina aplicadas a dados de imagens hiperespectrais pode ser usado para revelar características fisiológicas e estruturais em plantas e permitir o rastreamento da dinâmica fisiológica devido a efeitos ambientais (Wahabzada et al., 2016). Goldstein et al. (2018) demonstraram que dados de campo, como umidade do solo, clima, irrigação características e a produtividade resultante poderiam ser fundidos através de técnicas de ML para fornecer recomendações automatizadas para irrigação.

1.2.3.2 Redes neurais artificiais (RNAs)

As redes neurais são unidades de processamento interconectadas que foram inspiradas pelas estruturas biológicas no cérebro humano (Haykin, 1999; Khoshroo et al., 2018).

Cada uma das unidades de processamento é chamada neurônio (biológico) ou perceptron (computacional). Os perceptrons são organizados de maneira a definir uma arquitetura de rede sendo que o perceptron de várias camadas (MLP) é o tipo mais comum de redes neurais atualmente (Figura 1) (Khoshroo et al., 2018).

Em uma MLP, os neurônios são frequentemente organizados como uma camada de entrada, uma ou mais camadas ocultas e uma camada de saída (Catalão et al., 2011). A saída do neurônio é produzida processando as entradas ponderadas através de funções de transferência linear ou não linear (Basheer e Hajmeer, 2000). O erro calculado durante a etapa de treinamento é distribuído pela rede e ajusta os pesos de conexão entre os neurônios (Haykin, 1999).

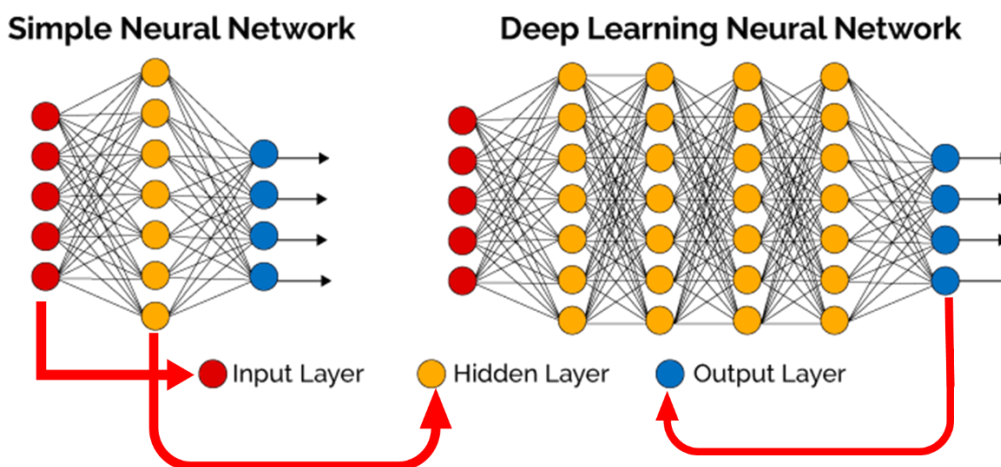


Figura 1 – Rede neural Perceptron de múltiplas camadas

Cada neurônio do MLP só pode transferir informações de sua camada para a próxima por uma função de transferência (TF). A TF determina o grau de excitação de cada camada. O processo pelo qual o conjunto de pesos sinápticos que compõe a RNA é conhecido como processo de treinamento (Perea et al., 2019)

Um problema que é muito frequente durante o treinamento da RNA é a memorização dos padrões de treinamento e, como consequência, a rede não é capaz de se adaptar ou reproduzir novas situações. Para evitar esse problema, chamado *overfitting*, são necessários procedimentos de generalização de rede. Existem vários métodos para obter uma generalização de rede apropriada. A parada antecipada é um desses métodos e consiste em dividir os dados disponíveis em três subconjuntos (Perea et al., 2019). Autores como González Perea (2015) mostram que a divisão do conjunto de dados pode ser feita em treinamento, validação e teste.

As RNAs fornecem uma ferramenta poderosa e flexível para modelar sistemas complexos (Catalão et al., 2011). RNAs são orientadas por dados e sem distribuição; portanto, eles podem aproximar funções não lineares e resolver os problemas em que a relação entrada-saída não é facilmente computável (Sözen, 2009).

A principal vantagem das RNAs é o fato de que a resolução de tarefas é feita, apresentando sinais de entrada estimulando a capacidade da rede de aprender e reconhecer padrões. Às vezes, a RNA é preferível a algoritmos complexos ou baseados em regras programação para resolver várias tarefas (Samborska et al., 2014).

As RNAs têm sido aplicada em vários campos de pesquisa, a agricultura vem sendo beneficiada por este método. Pesquisas mostram o uso de RNA para prever produtividade de cultivos, Michelon et al., (2018) usaram para estimar a produtividade de soja e milho por meio de leituras de clorofila, Alves et al., (2017) estimaram a produtividade de soja avaliando diferentes hábitos de crescimento e submetidas a diferentes densidades de semeadura, Abraham et al., (2019) também estimou a produtividade de soja, Taheri-rad et al., (2017) modelou os fluxos de energia para produção de diferentes cultivares de arroz em casca e Rostami et al., (2017) investigou a produção de desempenho de manjeriço de estufa e gases de efeito emitidos por este produto.

Devido ao grande potencial de estimativas e de previsões feitas por RNAs e a escassez de literatura nacional, esta dissertação tem como objetivo final a aplicação de RNAs para estimativas e previsões de produtividade de soja na região da fronteira agrícola do MATOPIBA.

REFERÊNCIAS

Abraham ER, dos Reis JGM, Toloi RC, de Souza AE, Colossetti AP (2019) Estimativa da produção da soja brasileira utilizando redes neurais artificiais. **Agrarian**, 2019, 12.44: 261-271.

Adeboye OB, Schultz B, Adekalu KO, Prasad KC (2019) Performance evaluation of AquaCrop in simulating soil water storage, yield, and water productivity of rainfed soybeans (*Glycine max L. merr*) in Ile-Ife, Nigeria. **Agricultural water management**, v. 213, p. 1130-1146.

Alves GR, Teixeira IR, Melo FR, Souza RTG, Silva AG (2018) Estimating soybean yields with artificial neural networks. **Acta Scientiarum. Agronomy**, 2018, 40.

Anar MJ, Lin Z, Hoogenboom G, Shelia V, Batchelor WD, Teboh JM, Khan M (2019) Modeling growth, development and yield of Sugarbeet using DSSAT. **Agricultural systems**, 2019, 169: 58-70.

Araújo MLS, Sano EE, Bolfe ÉL, Santos JRN, dos Santos JS, Silva FB (2019) Spatiotemporal dynamics of soybean crop in the Matopiba region, Brazil (1990–2015). **Land Use Policy**, v. 80, p. 57-67.

Avadhani SS, Arun AB, Govinda V, Inamdar JSI (2019) Crop Prediction Models—A Review. **In: Emerging Technologies in Data Mining and Information Security**. Springer, Singapore, 2019. p. 13-17.

Balboa G R, Archontoulis SV, Salvagiotti F, Garcia FO, Stewart WM, Francisco E, Ciampitti IA (2019) A systems-level yield gap assessment of maize-soybean rotation under high-and low-management inputs in the Western US Corn Belt using APSIM. **Agricultural Systems**, v. 174, p. 145-154.

Basheer IA, Hajmeer M (2000) Artificial neural networks: fundamentals, computing, design, and application. **Journal of microbiological methods**, 43.1: 3-31.

Brasil. Ministério da Agricultura, Pecuária e Abastecimento (2017) **Projeções do agronegócio**: Brasil 2016/17 a 2026/27 (103p). Brasília, DF: MAPA/SPA.

Buainain AM, Garcia JR, Vieira Filho JER (2018) A economia agropecuária do Matopiba. **Estudos Sociedade e Agricultura**, 2018, 26.2: 376-401.

Camacho DM, Collins KM, Powers RK, Costello JC, Collins JJ (2018) Nextgeneration machine learning for biological networks. **Cell**, v. 173, n. 7, p. 15811592.

Catalão JPDS, Pousinho HMI, Mendes VMF (2011) Short-term wind power forecasting in Portugal by neural networks and wavelet transform. **Renewable energy**, 2011, 36.4: 1245-1251.

CEPEA – Centro de Estudos Avançados em Economia Aplicada. **Mercado de trabalho do Agronegócio Brasileiro**. São Paulo - SP: CEPEA/ESALQ/USP, p.1-6, 2018. Disponível em: <
https://www.cepea.esalq.usp.br/upload/kceditor/files/2018_Relatorio%20MERCADODETR.> Acesso em: 03 mar. 2019.

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**, 2018, 151: 61-69.

Companhia Nacional De Abastecimento – CONAB (2015) **Acompanhamento da safra brasileira de grãos, Quarto Levantamento – Safra 2014/15**, v. 2, n. 4, p.1-90. Disponível em: <http://file:///C:/Users/amfal/Downloads/Perspectivas_para_a_Agropecuaria_-_V.5__Safra_2017-2018.pdf>. Acesso em: 25 outubro de 2019.

Companhia Nacional De Abastecimento – CONAB (2015) **Perspectivas para a agropecuária – Produtos de verão – Safra 2017-2018**, v. 5, p. 93-95. Disponível em: <[http://file:///C:/Users/amfal/Downloads/Perspectivas_para_a_Agropecuaria__V.5_-_Safra_2017-2018%20\(1\).pdf](http://file:///C:/Users/amfal/Downloads/Perspectivas_para_a_Agropecuaria__V.5_-_Safra_2017-2018%20(1).pdf)>. Acesso em: 25 outubro de 2019.

Cruz NNL (2016) Dependência espacial de atributos químicos, físicos e suscetibilidade magnética na região piauiense do MATOPIBA. Dissertação de Mestrado (Agronomia) - UFP, Bom Jesus.

Dokoohaki H, Gheysari M, Mousavi S, Zand-Parsa S, Miguez F, Archontoulis S,

Hoogenboom G (2016) Coupling and testing a new soil water module in DSSAT CERES-Maize model for maize production under semi-arid condition. **Agric. Water Manag.** 163, 90–99.

Dourado-Neto D, Teruel DA, Reichardt K, Nielsen DR, Frizzone JA, Bacchi OOS (1998) Principles of crop modeling and simulation: I. Uses of mathematical models in agricultural science. **Scientia Agricola**, 1998, 55.SPE: 46-50.

Empresa Brasileira de Pesquisa Agropecuária - EMBRAPA (2019) **MATOPIBA**. Disponível em:<<https://www.embrapa.br/tema-matopiba>> . Acesso em: 14 julho. 2019.

Fao Faostat (2018) [citado em 18 de outubro de 2019]. [Internet] Disponível em <<http://www.fao.org/faostat/en/#data/QC>>. Acesso em: 25 outubro de 2019.

Farias JRB, Nepomuceno AL, Neumaier N (2007) **Ecofisiologia da soja**. Londrina: Embrapa Soja, 1º ed., 9 p. (Circular Técnica, nº 48).

FINGER, Robert, et al. Precision farming at the nexus of agricultural production and the environment. 2019.

Fiss G, Schuch LOB, Peske ST, Castellanos CIS, Meneghello GE, Aumonde TZ (2018) Produtividade e características agrônomicas da soja em função de falhas na semeadura. Revista de Ciências Agrárias **Amazonian Journal of Agricultural and Environmental Sciences**, 61.

Gmach MR, Dias BO, Silva CA, Nóbrega JC, Lustosa-Filho JF, Siqueira-Neto M (2018) Soil organic matter dynamics and land-use change on Oxisols in the Cerrado, Brazil. **Geoderma regional**, 2018, 14: e00178.

Gmach MR, Dias BO, Silva CA, Nóbrega JC, Lustosa-Filho JF, Siqueira-Neto M (2018) Soil organic matter dynamics and land-use change on Oxisols in the Cerrado, Brazil. **Geoderma regional**, 2018, 14: e00178.

Goldstein A, Fink L, Meitin A, Bohadana S, Lutenberg O, Ravid G (2018) Applying machine learning on sensor data for irrigation recommendations: revealing the agronomist’s tacit knowledge. **Precision agriculture**, 19.3: 421-444.

González PP, Poyato EC, Montesinos P, Díaz JR (2015) Irrigation demand forecasting using artificial neuro-genetic networks. **Water Resources Management**, 29.15: 5551-5567.

Gusso, A., Arvor, D., & Ducati, J. R. (2017). Model for soybean production forecast based on prevailing physical conditions. **Pesquisa Agropecuária Brasileira**, 52(2), 95-103.

Haykin S (1999). Multilayer perceptrons. **Neural networks: a comprehensive foundation**, 1999, 2: 156-255.

Hirakuri M, Conte O, Balbinot Junior AA (2017) Análise econômica de diferentes arranjos espaciais de plantas de soja. **Embrapa Soja-Circular Técnica (INFOTECA-E)**.

IBGE – Instituto Brasileiro de Geografia e Estatística. **Estatísticas econômicas**. Disponível em: <<https://www.ibge.gov.br/estatisticas/economicas/agricultura-epecuaria.html>> Acesso em: 03 mar. 2019.

Jones MR, Singels A (2018) Refining the Canegro model for improved simulation of climate change impacts on sugarcane. **European Journal of Agronomy**, 100(1):76-86.

Kerry R G, Mahapatra GP, Patra S, Sahoo SL, Pradhan C, Padhi BK, Rout JR (2018) Proteomic and genomic responses of plants to nutritional stress. **Biometals**, 2018, 31.2: 161-187.

Khoshroo A, Emrouznejad A, Ghaffarizadeh A, Kasraei M, Omid M (2018) Sensitivity analysis of energy inputs in crop production using artificial neural networks. **Journal of cleaner production**, 2018, 197: 992-998.

Kim N, Ha KJ, Park NW, Cho J, Hong S, Lee YW (2019) A Comparison Between Major Artificial Intelligence Models for Crop Yield Prediction: Case Study of the Midwestern United States, 2006–2015. **ISPRS International Journal of GeoInformation**, 2019, 8.5: 240.

LECUN, Yann; BENGIO, Yoshua; HINTON, Geoffrey. Deep learning. **nature**, v. 521, n. 7553, p. 436-444, 2015.

Lowenberg-DeBoer J. The Precision Agriculture Revolution. *Foreign Aff.*. 2015;94:105.

Magalhães LA, Miranda ED (2014) MATOPIBA: quadro natural. **Nota Técnica**, v. 5. Disponível em: Grupo de Inteligência Territorial Estratégica (GITE) (2014) <https://www.embrapa.br/gite/publicacoes/NT5_Matopiba_Quadro_Natural.pdf>. (Acessado em 14 de julho de 2019).

Manzione RL (2018) Physical-based time series model applied on water table depths dynamics characteristics simulation. **RBRH**, 2018, 23.

Matricardi EAT, Aguiar AS, Miguel EP, Angelo H, Gaspar RDO (2018) Deforestation modeling in the MATOPIBA region. **Nativa: Pesquisas Agrárias e Ambientais**, 2018, 6.2: 198-206.

MATRICARDI EAT, MENDES TJ, PEREIRA EM, de Andrade VASCONCELOS PG, ÂNGELO H, da COSTA OB (2019). Dinâmica no uso e cobertura da terra na região do MATOPIBA entre 2000 e 2016. **Nativa**, Sinop, v. 7, n. 5, p. 547-555.

McCown RL, Hammer GL, Hargreaves JNG, Holzworth DP, Freebairn DM (1996) APSIM: a novel software system for model development, model testing and simulation in agricultural systems research. **Agricultural systems**, 1996, 50.3: 255-271.

Mehra LK, Cowger C, Gross K, Ojiambo PS, 2016 Predicting pre-planting risk of stagonospora nodorum blotch in winter wheat using machine learning models. *Front. Plant Sci.* 7, 390.

Michelon GK, de Menezes PL, Bazzi CL, Jasse EP, Magalhães PS, Borges LF (2018) Artificial neural networks to estimate the productivity of soybeans and corn by chlorophyll readings. **Journal of plant nutrition**, 2018, 41.10: 1285-1292.

Mingoti R, Brasco MA, Holler WA, Lovisi Filho E, Spadotto CA (2014) Matopiba: caracterização das áreas com grande produção de culturas anuais. **Embrapa Gestão Territorial-Nota Técnica/Nota Científica (ALICE)**.

Ministério Da Agricultura, Pecuária e Abastecimento – MAPA (2014) **Culturas: soja**. Ministério da Agricultura, Pecuária e Abastecimento. Disponível em: <<https://bit.ly/2P148xt>>. Acesso em: 22 setembro de 2019.

Ministério da Agricultura, Pecuária e Abastecimento – MAPA (2019) **Importação e exportação**. Disponível em:

<<http://www.agricultura.gov.br/assuntos/importacao-e-exportacao> > Acesso em: 03 mar. 2019.

Moreto VB (2019) Modelagem para auxiliar na otimização do sistema "ClimateSmart-Agriculture" para cultivo de cana-de-açúcar. Tese (Agronomia – Produção Vegetal) – UNESP, Jaboticabal.

Ovando G, Sayago S, Bocco M (2018) Evaluating accuracy of DSSAT model for soybean yield estimation using satellite weather data. **ISPRS journal of photogrammetry and remote sensing**, v. 138, p. 208-217.

Patrício, D. I., & 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.

Perea RG, Poyato EC, Montesinos P, Díaz JAR (2019) Optimisation of water demand forecasting by artificial intelligence with short data sets. **Biosystems engineering**, 2019, 177: 59-66.

Pereira LI, Pauli L (2016) O processo de estrangeirização da terra e expansão do agronegócio na região do Matopiba. **Campo-Território: Revista de Geografia Agrária**, v. 11, n. 23 jul.

Ritchie SW (1998) **Como a planta de soja se desenvolve**. POTAFOS, 1998.

ROSENBLATT, Frank. The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological review*, v. 65, n. 6, p. 386, 1958.

Rostami S, Choobin S, Samani BH, Esmaeili Z, Zareiforoush H (2017) Analysis and modeling of yield, CO₂ emissions, and energy for basil production in Iran using artificial neural networks. **International Journal of Agricultural Management and Development (IJAMAD)**, 2017, 7.1047-2017-1714: 47-58.

Samborska IA, Alexandrov V, Sieczko L, Kornatowska B, Goltsev V, Cetner MD, Kalaji HM (2014) Artificial neural networks and their application in biological and agricultural research. *J. NanoPhotoBioSciences*, 2014, 2: 14-30.

SAMUEL, Arthur L. Some studies in machine learning using the game of checkers. **IBM Journal of research and development**, v. 44, n. 1.2, p. 206-226, 2000.

Santos HG, Jacomine PKT, Anjos LHC, Oliveira VA, Lumbrieras JF, Coelho MR, Almeida JA, Cunha TJF, Oliveira JB (2013) **Sistema brasileiro de classificação de solos** (3 ed. rev. e ampl., 353p). Brasília: Embrapa Solos.

Sihag J, Prakash D (2019) uma revisão: Importância de várias técnicas de modelagem na agricultura / produção agrícola. Em: **Soft Computing: Teorias e Aplicações**. Springer, Singapura, 2019. p. 699-707.

Silva LVD, Casaroli D, Evangelista AWP, Alves Júnior J, Battisti R (2019) Relação Intensidade-Duração-Frequência de Chuva para Análise de Risco na Região do Matopiba, Brasil. **Revista Brasileira de Meteorologia**, 2019, 34.2: 247-254.

Silva MSL (2019) PRINCIPAIS DOENÇAS DA CULTURA DA SOJA (*Glycine max* (L.) Merrill). Trabalho de conclusão de curso (Agronomia) - IF Goiano, Rio Verde.

Silva VDP, Maciel GF, Souza EPD, Braga CC, Holanda RMD (2020). Soybean yield in the Matopiba region under climate changes. **Revista Brasileira de Engenharia Agrícola e Ambiental**, 24.1: 8-14.

Soares MR, Alleoni LRF (2008) Contribution of soil organic carbon to the ion exchange capacity of tropical soils. **Journal of Sustainable Agriculture**, 32 (3), 439-462.

Sözen A (2009). Future projection of the energy dependency of Turkey using artificial neural network. **Energy policy**, 2009, 37.11: 4827-4833.

Srivastava P, Singh R, Bhadouria R, Tripathi S, Singh H, Raghubanshi AS (2019) Understanding Soil Aggregate Dynamics and Its Relation With Land Use and Climate Change. In: **Climate Change and Agricultural Ecosystems**. Woodhead Publishing, 2019. p. 331-354.

Steduto P, Hsiao TC, Fereres E (2009). AquaCrop-the FAO crop model to simulate yield response to water: I. Concepts and underlying principles. **Agronomy Journal**, 101(3), 426-437.

Storm H, Baylis K, Heckelei T (2019). Machine learning in agricultural and applied economics. **European Review of Agricultural Economics**.

Taheri-Rad A, Khojastehpour M, Rohani A, Khoramdel S, Nikkhah A (2017) Energy flow modeling and predicting the yield of Iranian paddy cultivars using artificial neural networks. **Energy**, 2017, 135: 405-412.

Urano EOM, Kurihara CH, Maeda S, Vitorino ACT, Gonçalves MC, Marchetti ME (2007) Determinação de teores ótimos de nutrientes em soja pelos métodos chance matemática, sistema integrado de diagnose e recomendação e diagnose da composição nutricional. **Revista Brasileira de Ciência do Solo**, v. 31, n. 1, p. 63-72.

van Keulen H, Asseng S (2019). Simulation models as tools for crop management. **Crop Science**, 433-452.

Wahabzada M, Mahlein AK, Bauckhage C, Steiner U, Oerke EC, Kersting K (2016) Plant phenotyping using probabilistic topic models: uncovering the hyperspectral language of plants. **Scientific reports**, 6: 22482.

Wolfert S, Ge L, Verdouw C, Bogaardt MJ (2017) Big data in smart farming—a review. **Agricultural Systems**, 2017, 153: 69-80.

Yu Q, Xiang M, Wu W, Tang H (2019) Changes in global cropland area and cereal production: An inter-country comparison. **Agriculture, Ecosystems & Environment**, 2019, 269: 140-147.

Zhao X, van der Mensbrugge DY, Keeney RM, Tyner WE (2019) Improving the way land use change is handled in economic models. **Economic Modelling**.

Zipper SC, Qiu J, Kucharik CJ (2016) Drought effects on US maize and soybean production: spatiotemporal patterns and historical changes. **Environmental Research Letters**, v. 11, n. 9, p. 094021.

CAPÍTULO 2 – ESTIMAÇÃO E PREVISÃO DE PRODUTIVIDADE DE SOJA POR REDES NEURAS PROFUNDAS NO MATOPIBA

RESUMO – A utilização de Inteligência Artificial por meio de redes neurais artificiais profundas (DNNs) tem grande potencial para melhorias na acurácia em estimativas e previsões de produtividade agrícola de diversos cultivos no mundo em comparação a modelos de cultivo tradicionais baseados em processos. As DNNs tornam-se uma importante aliada não só aos produtores, mas também para instituições governamentais para decisões estratégicas em todos os níveis do sistema agrícola. Sabendo-se que as condições climáticas afetam sobremaneira a disponibilidade de água no solo e a produtividade dos cultivos de sequeiro, este trabalho propõe a utilização de DNNs a partir de dados meteorológicos mensais como temperatura do ar, precipitação e radiação global, e componentes do balanço hídrico como evapotranspiração de cultivo, armazenamento, evapotranspiração real, deficiência e excedentes hídricos durante o ciclo do cultivo para a estimativa e previsão de produtividade da soja para os estados do Maranhão, Tocantins, Piauí e Bahia que juntos formam a fronteira agrícola do MATOPIBA no Brasil. Esta área tem aproximadamente 73.173.485 hectares, correspondendo aproximadamente a 1,3 vezes a área da França. A avaliação da DNN para estimativa teve $R^2=0.88$ e $RMSE=167.85 \text{ kg ha}^{-1}$, enquanto que a DNN para previsão obteve $R^2=0.86$ e $RMSE=185.85 \text{ kg ha}^{-1}$. Conseguiu-se uma previsão de até 2 meses antes da colheita a partir das RNAs para o MATOPIBA.

Palavras-Chave: redes neurais artificiais, modelo de cultivos, agrometeorologia, MATOPIBA, Python, scikit-learn.

ESTIMATION AND FORECAST OF SOYBEAN YIELD BY NEURAL NETWORKS IN MATOPIBA

ABSTRACT – The use of Artificial Intelligence through deep artificial neural networks (NDS) has great potential for improvements in accuracy in estimates and forecasts of agricultural productivity of various crops in the world compared to cultivation models process-based traditional ones. NDS become an important ally not only to producers, but also for government institutions for strategic decisions at all levels of the agricultural system. Knowing that weather conditions greatly affect the availability of water in the soil and the productivity of dry land crops, this work proposes the use of NDs from monthly weather data such as air temperature, precipitation and global radiation, and components of the water balance such as evapotranspiration of cultivation, storage, real evapotranspiration, deficiency and water surpluses during the cultivation cycle for estimation and forecast soybean productivity for the states of the Maranhão, Tocantins, Piauí and Bahia that together form the agricultural frontier of MATOPIBA in Brazil. This area has approximately 73,173,485 hectares, corresponding approximately 1.3 times the area of France. The evaluation of NND for estimation had $R^2= 0.88$ and $RMSE=167.85$ kg ha⁻¹, while the DNN for forecast obtained $R^2= 0.86$ and $RMSE=185.85$ kg ha⁻¹. A forecast of up to 2 months before harvest from the RNAs to MATOPIBA was obtained.

Keywords: artificial neural networks, crop model, agrometeorology, MATOPIBA, Python scikit-learn.

2.1 Introdução

O uso da tecnologia nas atividades agrícolas tornou-se uma ferramenta fundamental de auxílio à tomada de decisões e que conseqüentemente proporciona aumento e sustentabilidade da produção agrícola em relação às condições ambientais, ecológicas e de mercado. A capacidade de prever a produtividade final antes do período da colheita é importante para tomadas de decisões em relação a épocas de plantio e colheita e determinação de preços de vendas (Filippi, 2019)

Várias estratégias para previsão de produtividade de soja foram realizadas, Monteiro e Sentelhas (2014) verificaram que o modelo agrometeorológico da Food and Agriculture Organization (FAO), tem alta precisão na estimativa da produtividade da soja para diferentes regiões do Brasil, Martin et al. (2012) estimaram a produtividade da soja através do modelo de previsão de cultivos modificados e parâmetros gerados no Brasil. Já Araújo et al. (2014) utilizaram modelo de regressão espacial autorregressivo misto e modelo do erro espacial para a região oeste do estado do Paraná. Enquanto, Alves (2018) utilizou a Rede Neural Artificial (RNA) do tipo Multilayer Perceptron (MLP) para estimar a produtividade da soja em Anápolis, Estado de Goiás, Brasil.

Khaki et al. (2019) usaram redes neurais profundas para prever e diferenciar a produtividade dos cultivos usando dados de genótipo e ambiente no Estados Unidos e no Canadá. You et al (2017) propôs uma abordagem baseada em ideias modernas de aprendizagem de máquina que levaram a melhorias maciças em várias tarefas de visão computacional, contornando a escassez de dados de treinamento empregando técnica de redução de dimensionalidade e tratando imagens brutas como histogramas de contagens de pixels usando aproximação de campo médio para obter rastreabilidade dos pixels utilizando arquiteturas de aprendizado profundo, incluindo redes neurais convolucionais – CNNs e Long Short Term Memory- LSTMs, um tipo de rede neural recorrente, para treinamento e previsão de produtividade dos cultivos nos Estados Unidos.

Shook et al. (2018) utilizou rede neural baseada em LSTM, dados de séries temporais multivariadas para prever o valor anual da produtividade da colheita da soja para os EUA e Canadá. Sun et al (2019) propôs um modelo de aprendizagem profundo de CNN-LSTM para previsão da produtividade final da

soja nos Estados Unidos a nível do condado. Com base no Google Earth Engine - GEE, várias variáveis monitoradas de longo prazo, incluindo dados climáticos, MODIS LST e MODIS SR que foram transformados em tensores para o treinamento do modelo, além disso, foram utilizados dados históricos de produtividade da soja para identificação e validação.

A previsão se difere da estimacão na ciência. A estimacão é o cálculo de uma condicão atual a partir de variáveis atuais, enquanto que a previsão é o cálculo de uma condicão futura (Aparecido et al., 2017). As previsões agrícolas constituem o meio mais importante de desenvolver a agricultura. Grande parte das técnicas de previsão numérica baseia-se na estatística, fundamentando-se na relacão entre as variáveis dependentes como produtividade dos cultivos que deverão ser estimadas, e as variáveis independentes agrometeorológicas (Hoogenboom et al., 2010).

É comum o uso de modelos que realizam a estimativa de crescimento e que avaliam o desenvolvimento de cultivos, dessa forma contribuindo para estimativa da produtividade agrícola e tornando compreensível os fatores que estão envolvidos nas diferentes respostas dos cultivos ao ambiente em cada estágio fenológico (Jones et al., 2017).

Nath et al, (2018) utilizaram o modelo de simulacão CROPGRO-soybean para estimar acuradamente a produtividade da soja sob diferentes faixas de temperatura e concentrações de CO₂ e avaliar os impactos das mudanças climáticas na Índia. Paredes et al., (2015) avaliou o modelo AquaCrop quanto a estimativa da produtividade da soja ao usar parâmetros calibrados e padrão na China, enquanto Yuan et al., (2016) examinou o desempenho de um modelo de eficiência no uso da radiação por satélite (RUE) para a previsão da biomassa de cultivos e produtividade, em várias espécies de cultivos em todo o mundo.

No entanto, a principal desvantagem da aplicacão de modelos de cultivos em é que esses modelos geralmente são calibrados para as condicões locais de campo, seu uso em outras regiões exigiria uma recalibração (Ceglar et al., 2016).

A implementacão em larga escala de modelos de cultivos requer dados detalhados das variáveis climáticas, produtividades anteriores, solos indicadores genéticos entre outras variáveis o que se configura como big data agrícola e requer técnicas e ferramentas apropriadas para o gerenciamento de um grande conjunto de dados (Kamilaris et al., 2017).

Essa desvantagem pode ser superada utilizando-se de Redes Neurais Artificiais (RNAs) que possuem a capacidade de executar modelagem não-linear sem conhecimento prévio sobre as relações entre variáveis de entrada e saída, portanto, RNAs é uma ferramenta de modelagem mais geral e flexível para previsão (Lamba et al., 2014).

Segundo Oikawa e Ishiki (2013), as Redes Neurais Artificiais podem ser definidas como uma ferramenta estatística cujo princípio de funcionamento é regido por um modelo matemático inspirado no funcionamento dos elementos básicos que formam a estrutura neural de organismos inteligentes, que adquirem conhecimento através de experiência e, por meio do processamento de informações, gera uma saída (dados preditos ou target) a partir de uma ou mais entradas apresentadas (preditores ou features). As RNA são compostas por neurônios ou unidades de processamento, que computam determinadas funções matemáticas (ANOCHI et al., 2016).

A solução de problemas através de redes neurais artificiais é bastante atrativa, já que a forma como estes são apresentados internamente pela rede cria a possibilidade de um desempenho superior ao dos modelos convencionais como citado por Alves et al. (2018) que afirmam que o emprego de sistemas computacionais inteligentes na estimação da produtividade da soja é viável.

O relatório de Perspectivas para a Agropecuária para a Safra 2019/2020 produzido pela Companhia Nacional de Abastecimento - CONAB, com base nos dados do Departamento de Agricultura dos Estados Unidos (USDA) mostra que o Brasil se tornará o maior produtor mundial de soja em 2020. As previsões do USDA mostram uma redução na área de plantio do Estados Unidos e o aumento das áreas de plantios no Brasil tendo com previsão de produtividade 123 milhões de toneladas para a safra 2019/2020 e mantendo o posto de maior exportador mundial (CONAB, 2019).

No Brasil a soja é cultivada em 14 estados: Rondônia, Pará, Tocantins, Maranhão, Piauí, Bahia, Minas Gerais, São Paulo, Paraná, Santa Catarina, Rio Grande do Sul, Mato Grosso do Sul, Mato Grosso, Goiás e no Distrito Federal com produção regular e com expectativas de crescimento (IBGE, 2018).

A safra 2017/2018 foi de aproximadamente 114,0 milhões de toneladas. Com produção liderada pelos estados de Mato Grosso, com mais de 27,3% da produção nacional, e Paraná com 16,3%, além do Maranhão, Tocantins, Pará,

Rondônia, Piauí e Bahia, que em 2017/18 responderam por 14,0% da produção brasileira, que corresponde a uma produção de 16,4 milhões de toneladas de soja (CONAB, 2018).

Os estados do Maranhão, Tocantins, Piauí e Bahia, juntos formam o MATOPIBA que totalizam quase 143 milhões de hectares e mais de 25 milhões de habitantes, a delimitação geográfica e operacional do MATOPIBA é de aproximadamente 73 milhões de hectares, 5,9 milhões de habitantes em 337 municípios (Censos agropecuários de 2006 e demográfico de 2010, Miranda, 2015).

Como um dos principais cultivos do Brasil e do MATOPIBA, a soja tem papel de destaque entre as pesquisas, sendo necessário prever sua produtividade para o planejamento da lavoura e tomadas de decisões.

O objetivo deste estudo é desenvolver um modelo de redes neurais artificiais de aprendizado profundo para a estimativa e previsão da produtividade da soja na região do MATOPIBA a partir de dados meteorológicos, do cultivo e do balanço hídrico.

2.2 Material e Métodos

2.2.1 Localização e caracterização da área de estudo

A área de estudo consistiu nas localidades pertencente aos estados que fazem parte da região do MATOPIBA (Figura 1). Os municípios foram identificados através do relatório de Delimitação e Caracterização Territorial do MATOPIBA de 2015 do Grupo de Inteligência Territorial Estratégica da Embrapa e a portaria N^o- 244, de 12 de novembro de 2015 do Ministério da Agricultura, Pecuária e Abastecimento - MAPA que aborda o Plano de Desenvolvimento Agropecuário para a região (Figura 1).

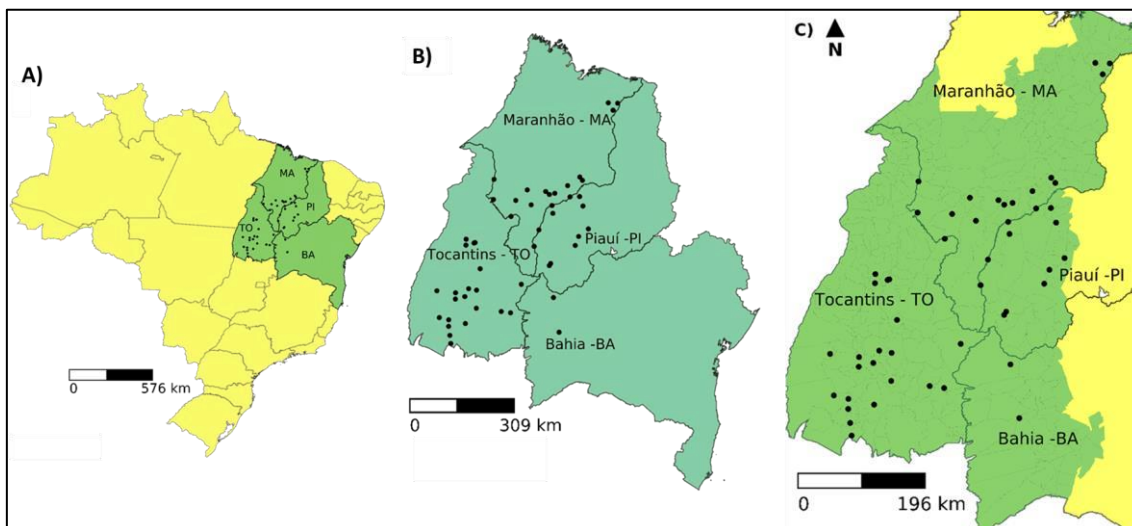


Figura 1 - Localização da área de estudo, A) Brasil, B) Estados que compõem a Fronteira agrícola do MATOPIBA e C) Delimitação territorial do MATOPIBA.

2.2.2 Seleção dos locais de estudo

Os dados de produtividade foram obtidos no Sistema de Recuperação Automática do Instituto Brasileiro de Geografia e Estatística – SIDRA/IBGE, na base de dados da Produção Agrícola Municipal – PAM, que investiga um conjunto de produtos das lavouras temporárias e permanentes do Brasil anualmente (IBGE,2019).

Foram selecionadas 52 localidades do MATOPIBA pois apresentaram 18 anos de dados consistentes e ininterruptos de produtividade de soja. A produtividade média foi de 2.575 kg ha^{-1} , a região apresenta variabilidade na produtividade média na série histórica desde 2.274 kg ha^{-1} em Santa Rita do Tocantins – TO a 2.837 kg ha^{-1} em São Desidério – BA, a estimativa da produtividade média nacional para a safra 2019/2020 será de 3.330 kg ha^{-1} (CONAB, 2019). As falhas de dados foram corrigidas por meio de média móvel (Tabela 1).

Tabela 1 – Municípios selecionados conforme critério da série histórica nos estados do Maranhão - MA, Tocantins – TO, Piauí – PI e Bahia -BA, seguidos de área plantada em ha e produtividade média da série histórica em kg ha⁻¹.

| Local | UF | Área Plantada (ha) | Prod. Média (kg ha ⁻¹) | Local | UF | Área Plantada (ha) | Prod. Média (kg ha ⁻¹) |
|------------------------------|----|--------------------------|--|-------------------------|----|--------------------------|--|
| Alto Parnaíba | MA | 24.806 | 2.669 | Formoso do Araguaia | TO | 16.353 | 2.498 |
| Anapurus | MA | 4.944 | 2.449 | Fortaleza do Tabocão | TO | 2.106 | 2.544 |
| Balsas | MA | 121.428 | 2.632 | Guaraí | TO | 7.747 | 2.618 |
| Brejo | MA | 12.638 | 2.623 | Lagoa da Confusão | TO | 18.391 | 2.516 |
| Buriti | MA | 8.540 | 2.728 | Mateiros | TO | 28.595 | 2.602 |
| Carolina | MA | 15.850 | 2.617 | Monte do Carmo | TO | 12.299 | 2.723 |
| Estreito | MA | 550 | 2.729 | Pedro Afonso | TO | 26.410 | 2.506 |
| Fortaleza dos Nogueiras | MA | 15.886 | 2.640 | Peixe | TO | 12.785 | 2.314 |
| Loreto | MA | 16.897 | 2.604 | Porto Nacional | TO | 17.580 | 2.495 |
| Pastos Bons | MA | 2.470 | 2.711 | Santa Rita do Tocantins | TO | 2.334 | 2.274 |
| Riachão | MA | 31.245 | 2.614 | Santa Rosa do Tocantins | TO | 13.504 | 2.531 |
| Sambaíba | MA | 37.944 | 2.536 | Talismã | TO | 4.086 | 2.381 |
| São Domingos do Azeitão | MA | 15.131 | 2.648 | Tupirama | TO | 8.293 | 2.464 |
| São Raimundo das Mangabeiras | MA | 19.903 | 2.583 | Alvorada do Gurguéia | PI | 4.465 | 2.303 |
| Sucupira do Norte | MA | 1.273 | 2.800 | Antônio Almeida | PI | 3.464 | 2.281 |
| Tasso Fragoso | MA | 99.910 | 2.520 | Baixa Grande do Ribeiro | PI | 73.696 | 2.737 |
| Almas | TO | 2.296 | 2.659 | Bom Jesus | PI | 34.091 | 2.353 |
| Alvorada | TO | 9.765 | 2.684 | Gilbués | PI | 15.873 | 2.381 |
| Aparecida do Rio Negro | TO | 7.901 | 2.862 | Monte Alegre do Piauí | PI | 9.741 | 2.316 |

Continua na próxima

| Local | UF | Área Plantada (ha) | Prod. Média (kg ha ⁻¹) | Local | UF | Área Plantada (ha) | Prod. Média (kg ha ⁻¹) |
|------------------------|----|--------------------------|--|----------------------|----|--------------------------|--|
| Bom Jesus do Tocantins | TO | 5.369 | 2.779 | Palmeira do Piauí | PI | 8.259 | 2.537 |
| Brejinho de Nazaré | TO | 8.914 | 2.545 | Ribeiro Gonçalves | PI | 38.937 | 2.648 |
| Campos Lindos | TO | 44.584 | 2.606 | Santa Filomena | PI | 24.607 | 2.630 |
| Cariri do Tocantins | TO | 6.284 | 2.663 | Sebastião Leal | PI | 10.364 | 2.499 |
| Crixás do Tocantins | TO | 1.985 | 2.506 | Uruçuí | PI | 76.019 | 2.721 |
| Dianópolis | TO | 15.132 | 2.591 | Formosa do Rio Preto | BA | 218.785 | 2.582 |
| Figueirópolis | TO | 8.285 | 2.591 | São Desidério | BA | 265.779 | 2.837 |

Fonte: IBGE, 2019

2.2.3 Dados Meteorológicos

Foi usada a linguagem de programação em Python 3.6 para coletar os dados agrometeorológicos de todos os locais estudados por meio de um API disponibilizado na plataforma NASA-POWER (Stackhouse et al., 2017). Todos os produtos de dados são classificados na resolução espacial 1 x 1°. No entanto, a resolução temporal varia dependendo do uso do produto de dados (Stackhouse et al., 2016).

Foram coletados dados diários de precipitação (P , mm dia⁻¹), temperatura máxima e mínima do ar (T , °C), temperatura no ponto de orvalho (T_o , °C), umidade relativa do ar (UR %), radiação solar no topo da atmosfera (Q_o , Mj m⁻² dia⁻¹), insolação (Q_g , Mj m⁻² dia⁻¹) e velocidade média do vento (V_{Vm} , m s⁻¹) de uma série histórica de 18 anos para cada localidade. A evapotranspiração potencial de referência foi calculada pelo método de Penman-Monteith (FAO 56, Allen et al, 1998) e em seguida obteve-se a evapotranspiração de cultivo, multiplicando a evapotranspiração potencial com o coeficiente de cultivo k_c segundo Evangelista (2017).

Tabela 2: Coeficiente de cultivo (Kc) da soja. Legenda: S-semeadura; V-vegetativo; R-reprodutivo.

| Estádios | KC(Evangelista, 2017) |
|-------------------------|------------------------------|
| Estabelecimento –S-V2 | 0.58 |
| Desenvolvimento – V2-R1 | 1.06 |
| Floração – R1 – R3 | 1.39 |
| Frutificação – R3 - R7 | 1.09 |
| Maturação – R7- R8 | 0.55 |

2.2.4 Balanço Hídrico

O balanço hídrico mensal foi realizado pelo método de Thornthwaite e Mather (1955), com capacidade de água disponível (AWC) igual a 80 mm de acordo com o proposto por Evangelista (2017). Em seguida foram selecionadas, Evapotranspiração de cultivo (mm) (CET), Evapotranspiração real (mm) (AET), Armazenamento (mm) (STO), Déficit (mm) (DEF) e Excedente (mm) (EXC) hídricos, para compor o conjunto de variáveis independentes nos modelos de redes neurais artificiais profundas (DNNs)

2.2.5 Análise de dados

A análise foi dividida em 2 etapas: a primeira buscou-se compreender as relações existentes entre as variáveis climáticas e os componentes do balanço hídrico com a produtividade da soja na região do MATOPIBA.

A segunda etapa foi a aplicação de algoritmos de machine learning, Redes Neurais Artificial Profundas do tipo Multilayer Perceptron (DNN) com o algoritmo MLPRegressor do pacote scikit-learn (Pedregosa et al, 2011) com diferentes topologias para A) estimar e B) prever a produtividade da soja a partir das variáveis climáticas e componentes do BH. Os dados de entrada (features) foram todos padronizados (média nula e variância constante).

A MLP é um algoritmo de aprendizado supervisionado. Ela aprende uma função de treinamento em um conjunto de dados para proporcionar uma ou mais saídas (dependente ou targets). Para se obter índices ótimos de previsão pela DNNs, foram ajustadas diferentes topologias de redes (Figura 2).

A camada de entrada (input) foi formada pelo banco de dados composto por uma matriz de 36 colunas por 936 linhas. Cada uma das variáveis independentes T, P, CET, AET, STO, DEF e EXC foi estratificada em cinco meses referentes ao ciclo da soja (Figura 2).

| Estádios | Variáveis independentes (Features) | | | | | Variável dependente |
|------------|------------------------------------|------------------|------------------|------------------|------------------|---------------------------|
| | NOV | DEZ | JAN | FEV | MAR | |
| | S-V6 | R1-R2 | R3-R5 | R6-R8 | R9 | |
| T | T ₁ | T ₂ | T ₃ | T ₄ | T ₅ | Produtividade média anual |
| P | P ₁ | P ₂ | P ₃ | P ₄ | P ₅ | |
| CET | CET ₁ | CET ₂ | CET ₃ | CET ₄ | CET ₅ | |
| AET | AET ₁ | AET ₂ | AET ₃ | AET ₄ | AET ₅ | |
| STO | STO ₁ | STO ₂ | STO ₃ | STO ₄ | STO ₅ | |
| DEF | DEF ₁ | DEF ₂ | DEF ₃ | DEF ₄ | DEF ₅ | |
| EXC | EXC ₁ | EXC ₂ | EXC ₃ | EXC ₄ | EXC ₅ | |

Figura 2: Representação dos dados na camada de entrada (input) relativo a um ano. Legenda: T – Temperatura do ar (°C); P – Precipitação (mm); CET – Evapotranspiração de cultivo (mm); AET – Evapotranspiração real de cultivo (mm); STO – Armazenamento (mm); DEF – Déficit (mm); EXC – Excedente (mm).

Cada conjunto de 36 dados são referentes a um ano para um local específico, considerando a série histórica estabelecida de 18 anos o total de valores por local é de 648. É importante ressaltar que o ciclo de cultivo da soja considerado neste trabalho é de 5 meses, conforme o Calendário de Plantio e Colheita de Grãos no Brasil (Conab, 2019), que fornece os meses nos quais se realizam a semeadura e a colheita de diversos cultivos agrícolas ao longo do ano de acordo com a região do país. A soja de ciclo médio varia entre 126 a 137 dias (EMBRAPA, 2014) o que é comumente adotado no MATOPIBA.

O banco de dados para estimação foi composto por variáveis referente aos 5 meses do ciclo (Novembro, Dezembro, Janeiro, Fevereiro e Março), já o banco de dados para previsão utilizou apenas dados referente aos 3 primeiros meses (Novembro, Dezembro e Janeiro) (Figura 3).

A) Estimation



B) Forecasting



Figura 3: Esquema das Features para as Redes Neurais Artificiais Profundas para estimação (A) e previsão (B) de produtividade de soja para o MATOPIBA.

A escolha do melhor ajuste de hiperparâmetros para o modelo se deu por meio do Algoritmo de busca exaustiva `sklearn.model_selection.GridSearchCV`, que testa todas as diferentes de topologias e parâmetros solicitados de forma ordenada em busca da melhor otimização, ou seja, determina o mínimo geral da função da soma de quadrados entre os dados observados e estimados (Pedregosa et al, 2011).

Tabela 3: Hiperparâmetros ajustados na Rede Neural Artificial Profunda Multilayer Perceptron – DNN-MLP. O MLPRegressor é um método disponível no pacote Scikit-learn (Pedregosa et al, 2011) da linguagem Python.

| Hiperparâmetros ajustados | Função |
|---------------------------------|--|
| <code>hidden_layer_sizes</code> | i-ésimo (quantidade camadas) elemento representa o número de neurônios na i-ésima camada oculta. |
| <code>activation</code> | Função de ativação para a camada oculta. |
| <code>solver</code> | Solucionador de otimização de peso |
| <code>alpha</code> | Parâmetro de penalidade L2 (termo de regularização). |
| <code>learning_rate</code> | Programação da taxa de aprendizado para atualizações de peso. |
| <code>learning_rate_init</code> | A taxa de aprendizado inicial usada. Controla o tamanho da etapa na atualização dos pesos. |
| <code>max_iter</code> | Número máximo de iterações. |
| <code>random_state</code> | Gerador de números aleatórios. |

Para realizar a pesquisa exaustiva inicialmente determinou-se o número de camadas ocultas da DNN, em seguida criou-se um gerador de números aleatórios, onde por meio desse, gerou-se três listas e em seguida agregou-se os números em tuplas (listas imutáveis) de forma que cada tupla tivesse três valores correspondendo ao número de camadas ocultas. Os valores dessas tuplas representam a quantidade de neurônios em cada camada oculta da DNN.

Utilizou-se as funções de ativações 'relu': função de unidade linear retificada, retorna $f(x) = \max(0, x)$ e 'logistic': função sigmóide logística, retorna $f(x) = 1 / (1 + \exp(-x))$. Os solucionadores Stochastic Gradient Descent - 'sgd': refere-se à descida do gradiente estocástico (Robbins e Monro, 1951), 'adam': refere-se a um otimizador estocástico baseado em gradiente proposto por Kingma, Diederik e Jimmy Ba (Kingma et al., 2014) e o algoritmo de otimização de memória limitada Broyden-Fletcher-Goldfarb-Shanno - 'lbfgs' que é um otimizador na família de métodos quasi-Newton.

Determinou-se a taxa de aprendizado 'constant': que é uma taxa de aprendizado constante fornecida por 'learning_rate_init' e 'adaptive': que mantém a taxa de aprendizado constante como 'learning_rate_init', desde que a perda de treinamento continue diminuindo. E a taxa de aprendizado inicial para 0.1 e 0.01, além do parâmetro de penalidade alpha para 0.0001 e 0.05 para número de interações para 4000 e o Random_state igual a 0 (zero).

Todos esses parâmetros foram combinados pelo algoritmo de pesquisa exaustiva. O resultado foi filtrado por meio de um loop afim de se separar as melhores configurações para obter os melhores ajustes.

2.2.6 Avaliação do modelo

Nesta pesquisa utilizamos a validação cruzada como método de avaliação do desempenho do modelo. Usamos o pacote Cross-validation (CV) da biblioteca Scikit-Learn que usa uma abordagem denominada k-fold CV, onde o conjunto de dados é dividido em k conjuntos menores. A medida de desempenho relatada pela validação cruzada é então a média dos valores calculados no loop. (Pedregosa et al, 2011).

A CV utilizando o k-fold dividiu o conjunto de dados disponíveis aleatoriamente em 10 subconjuntos (k-fold=10). Para cada execução, um subconjunto foi utilizado para avaliação dos resultados classificados derivados dos outros 9, que são repetidos 10 vezes até que todos os 10 subconjuntos sejam usados uma vez como subconjunto de teste, (Xu, 2019). Após os testes obteve-se o valor médio da validação cruzada para precisão (R^2) e o erro quadrático médio (MSE) e a raiz do erro quadrático médio (RMSE).

$$MSE = \frac{1}{n} \sum_{i=1}^n (Y_{obs_i} - \bar{Y}_{est_i})^2 \quad (1)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (Y_{obs_i} - \bar{Y}_{est_i})^2}{n}} \quad (2)$$

$$R^2 = \frac{\sum_{i=1}^N (Y_{est} - \bar{Y}_{obs})^2}{\sum_{i=1}^N (Y_{est} - \bar{Y}_{obs})^2} \quad (3)$$

Onde, Y_{obs_i} são os dados observados e Y_{est_i} os dados estimados e ou previstos.

2.3 Resultados e Discussão

O MATOPIBA apresenta locais com características altamente distintas, considerando o conjunto de variáveis climáticas e componentes do balanço hídrico. Essas diferenças ocorrem dentro de um mesmo estado e de um estado para outro, evidenciando áreas com maior e menor potencial produtivo.

As Figuras de 4 a 10 mostram os mapas de distribuição espacial das variáveis climáticas e componentes do balanço hídrico entre novembro e março, respectivamente, seguido do acumulado do período observado (Figuras 4, 9 e 10 – F) e da distribuição média (Figuras 5, 6 e 7 – F) para o MATOPIBA durante série histórica de 18 anos.

Volumes mais altos de precipitação (Figura 4) ocorrem na região Oeste e Central do MATOPIBA durante todo período analisado. De novembro a janeiro a região leste apresenta baixos volumes precipitação com aumento considerável em fevereiro e março. Na região sul do MATOPIBA, correspondendo ao oeste Baiano, os maiores volumes de chuva ocorrem entre novembro e janeiro, distinguindo-se da região leste.

Porém, considerando o acumulado de precipitação para o período analisado (Figura 4-F), o MATOPIBA pode ser dividido em região com altos volumes de precipitação (região oeste) com valores variando de 1020 a 1280 mm, região de transição de precipitação (faixa central –norte – sul) com valores variando de 890 a 1020 mm e região de baixo volume de precipitação (região leste) com valores variando de 630 a 890 mm.

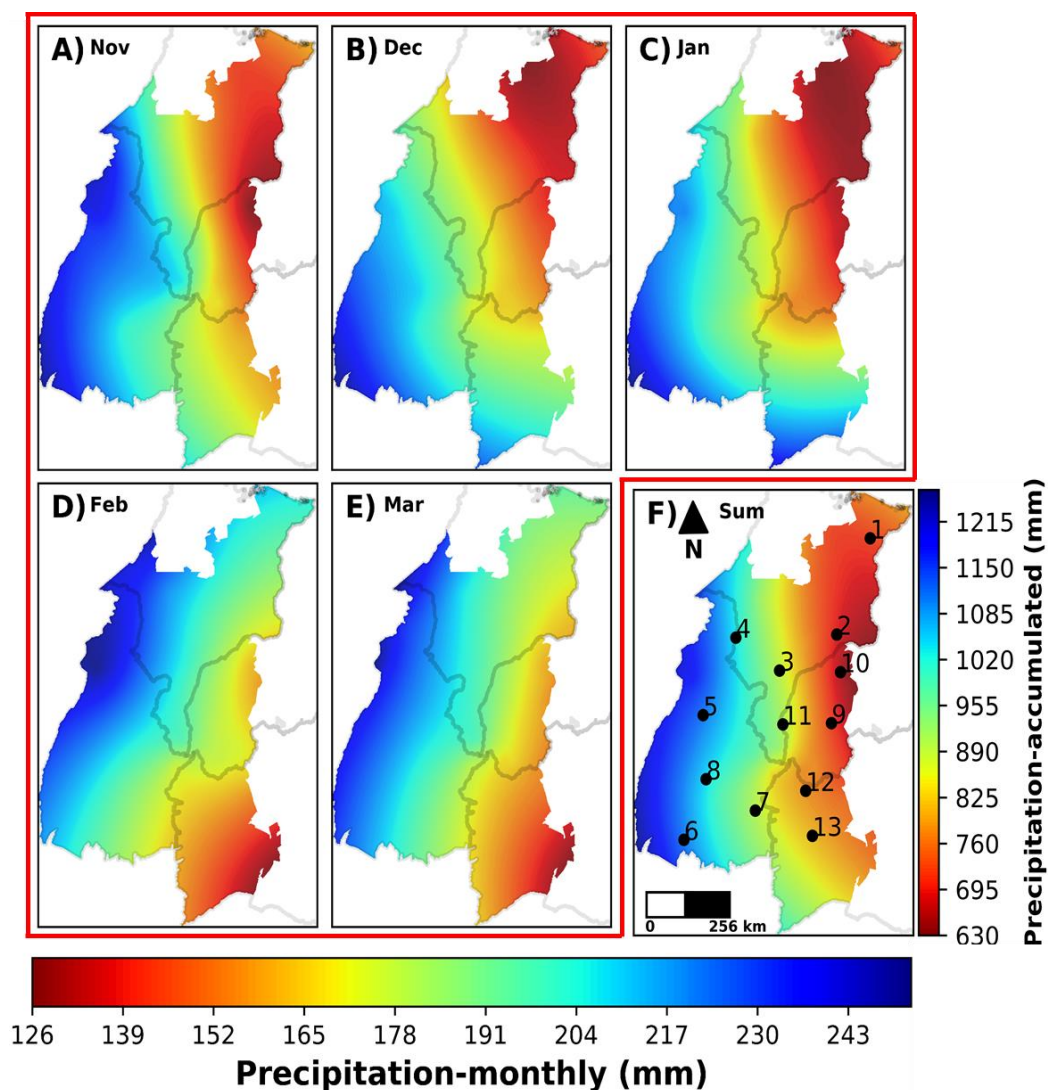


Figura 4: Precipitação dos meses de A) novembro, B) dezembro, C) janeiro, D) fevereiro, E) março e F) precipitação acumulada do período. 1) Anapurus-MA, 2) Sucupira do Norte-MA, 3) Balsas-MA, 4) Estreito-MA, 5) Guarái-TO, 6) Alvorada-TO, 7) Dianópolis-TO, 8) Porto Nacional-TO, 9) B Jesus-PI, 10) Sebastião Leal-PI, 11) Alto Parnaíba-MA, 12) Formosa do Rio Preto-BA, 13) São Desidério-BA.

A temperatura no MATOPIBA (Figura 5) apresenta média de 28.8 °C, variando de 24.6 a 33 °C entre novembro e março. A temperatura não apresenta grandes variações entre os estados que compõem o MATOPIBA (Figura 5-F).

Nota-se a alta variabilidade espacial principalmente das condições hídricas (AET, DEF, STO e EXC) (Figuras 7,8,9 e 10).

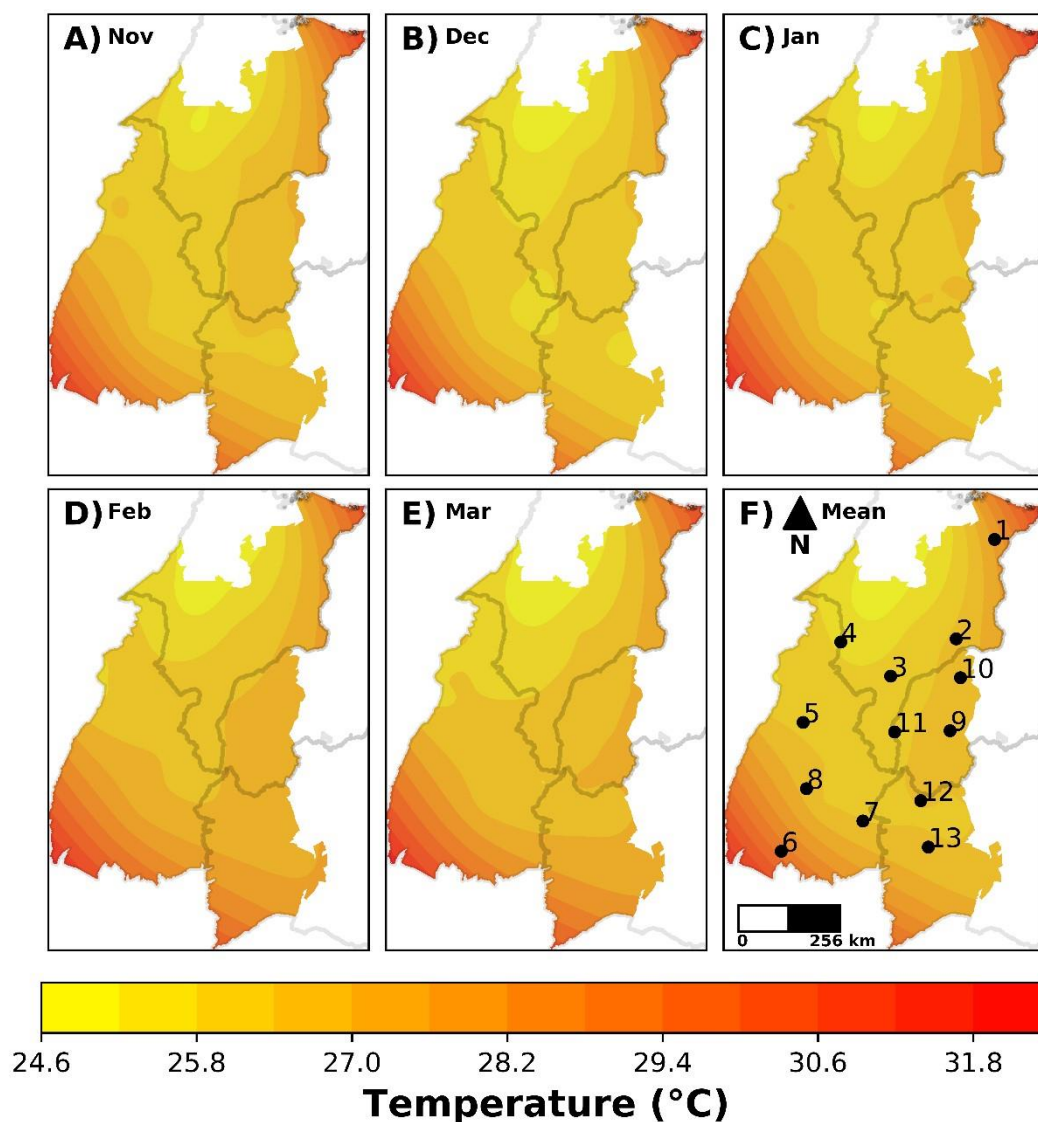


Figura 5: Temperatura dos meses de A) novembro, B) dezembro, C) janeiro, D) fevereiro, E) março e F) temperatura média do período. 1) Anapurus-MA, 2) Sucupira do Norte-MA, 3) Balsas-MA, 4) Estreito-MA, 5) Guaraí-TO, 6) Alvorada-TO, 7) Dianópolis-TO, 8) Porto Nacional-TO, 9) B Jesus-PI, 10) Sebastião Leal-PI, 11) Alto Parnaíba-MA, 12) Formosa do Rio Preto-BA, 13) São Desidério-BA.

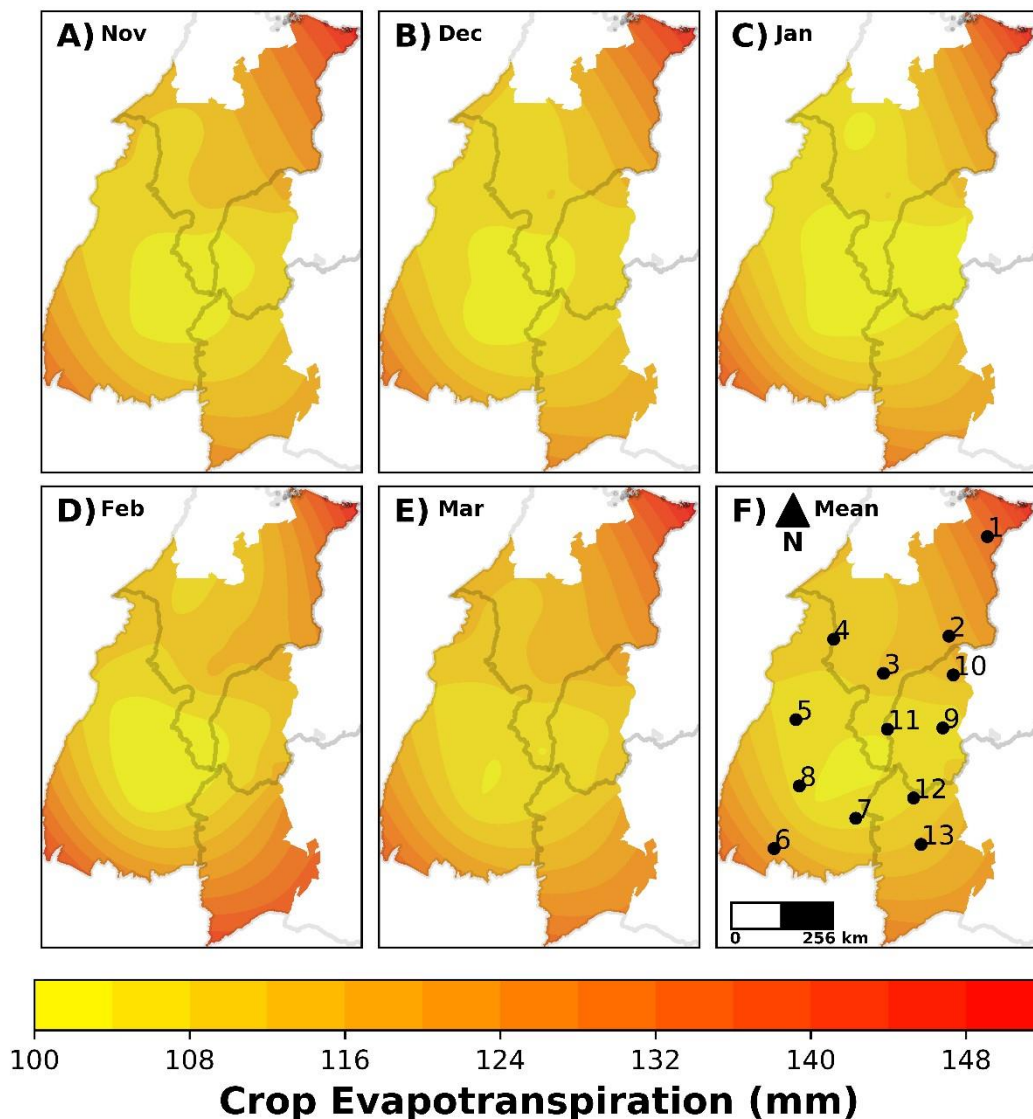


Figura 6: Evapotranspiração da cultura dos meses de A) novembro, B) dezembro, C) janeiro, D) fevereiro, E) março e F) evapotranspiração média da cultura no período. 1) Anapurus-MA, 2) Sucupira do Norte-MA, 3) Balsas-MA, 4) Estreito-MA, 5) Guaraí-TO, 6) Alvorada-TO, 7) Dianópolis-TO, 8) Porto Nacional-TO, 9) B Jesus-PI, 10) Sebastião Leal-PI, 11) Alto Parnaíba-MA, 12) Formosa do Rio Preto-BA, 13) São Desidério-BA.

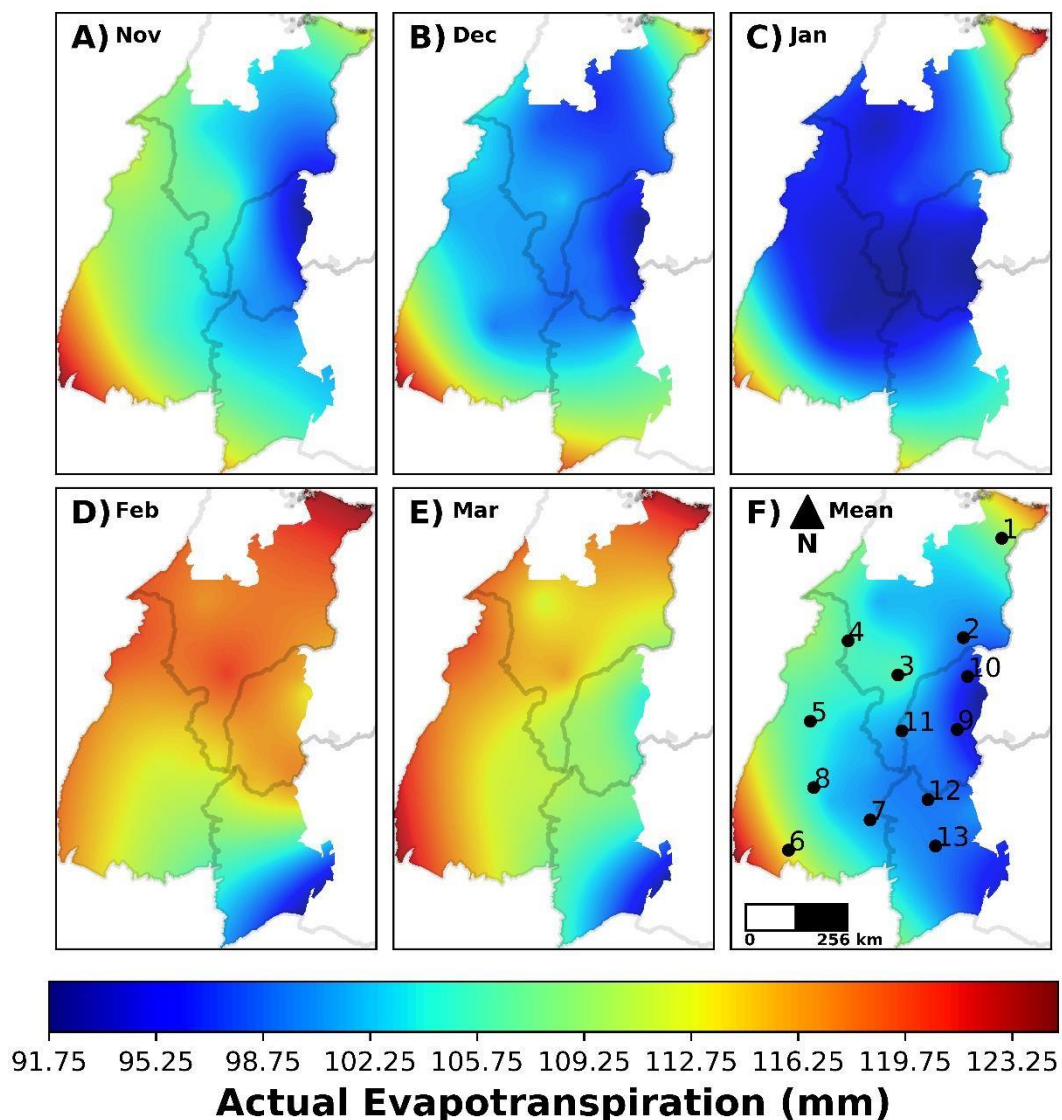


Figura 7: Evapotranspiração real dos meses de A) novembro, B) dezembro, C) janeiro, D) fevereiro, E) março e F) evapotranspiração média do período. 1) Anapurus-MA, 2) Sucupira do Norte-MA, 3) Balsas-MA, 4) Estreito-MA, 5) Guaraí-TO, 6) Alvorada-TO, 7) Dianópolis-TO, 8) Porto Nacional-TO, 9) B JesusPI, 10) Sebastião Leal-PI, 11) Alto Parnaíba-MA, 12) Formosa do Rio Preto-BA, 13) São Desidério-BA.

O déficit hídrico (Figura 8) é mais acentuado na região leste do Maranhão entre novembro e janeiro. Em fevereiro e março o déficit diminui de acordo com os volumes de chuva. O acumulado de déficit hídrico para o período (Figura 8F) evidencia as regiões leste do Maranhão e oeste da Bahia como regiões de elevado déficit hídrico.

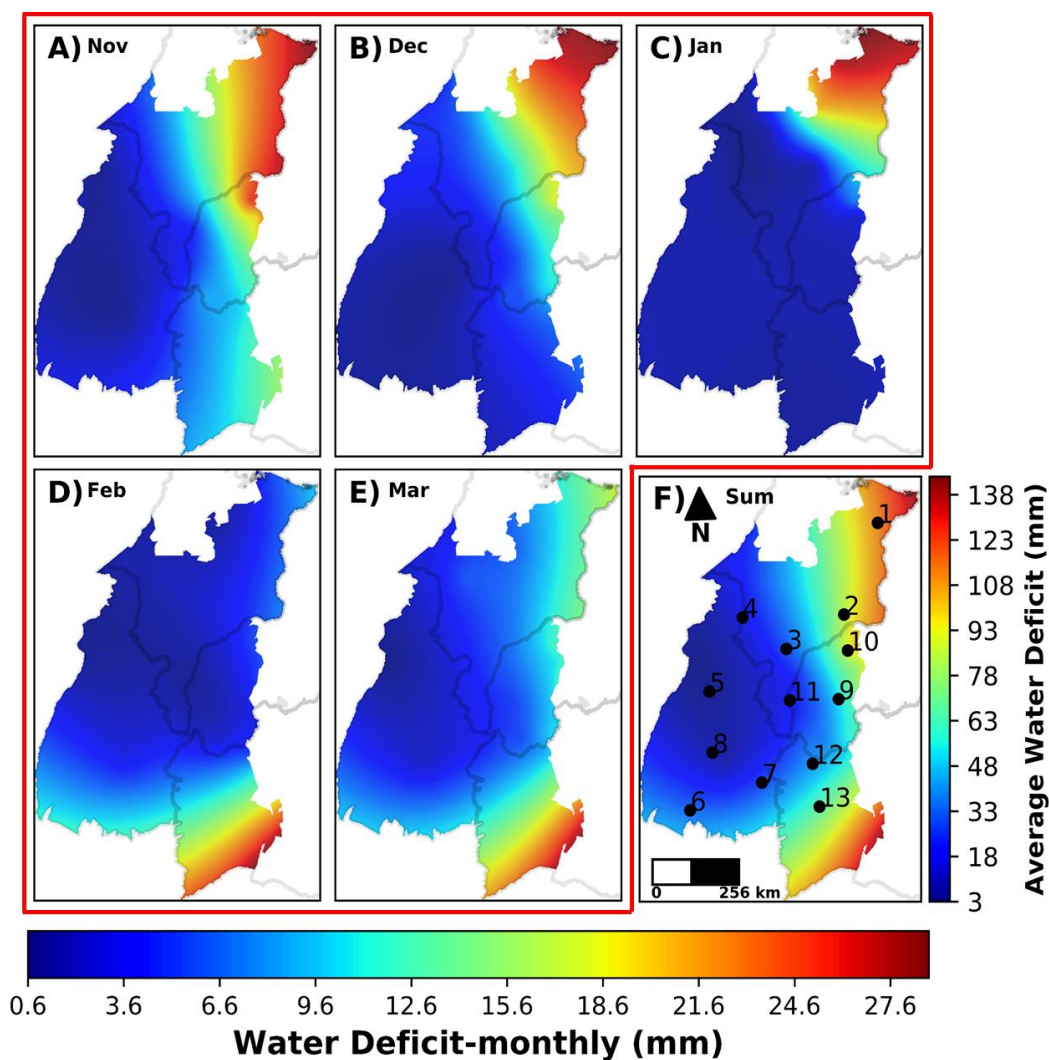


Figura 8: Deficiência hídrica nos meses de A) novembro, B) dezembro, C) janeiro, D) fevereiro, E) março e F) Deficiência hídrica acumulada do período. 1) Anapurus-MA, 2) Sucupira do Norte-MA, 3) Balsas-MA, 4) Estreito-MA, 5) Guarai-TO, 6) Alvorada-TO, 7) Dianópolis-TO, 8) Porto Nacional-TO, 9) B Jesus-PI, 10) Sebastião Leal-PI, 11) Alto Parnaíba-MA, 12) Formosa do Rio Preto-BA, 13) São Desidério-BA.

A média do armazenamento de água no solo e a média de excedente hídrico (Figura 9 – F e Figura 10 – F) mostram que o MATOPIBA está dividido em duas regiões, uma úmida (região oeste) e uma seca (região leste) acompanhando a distribuição da precipitação.

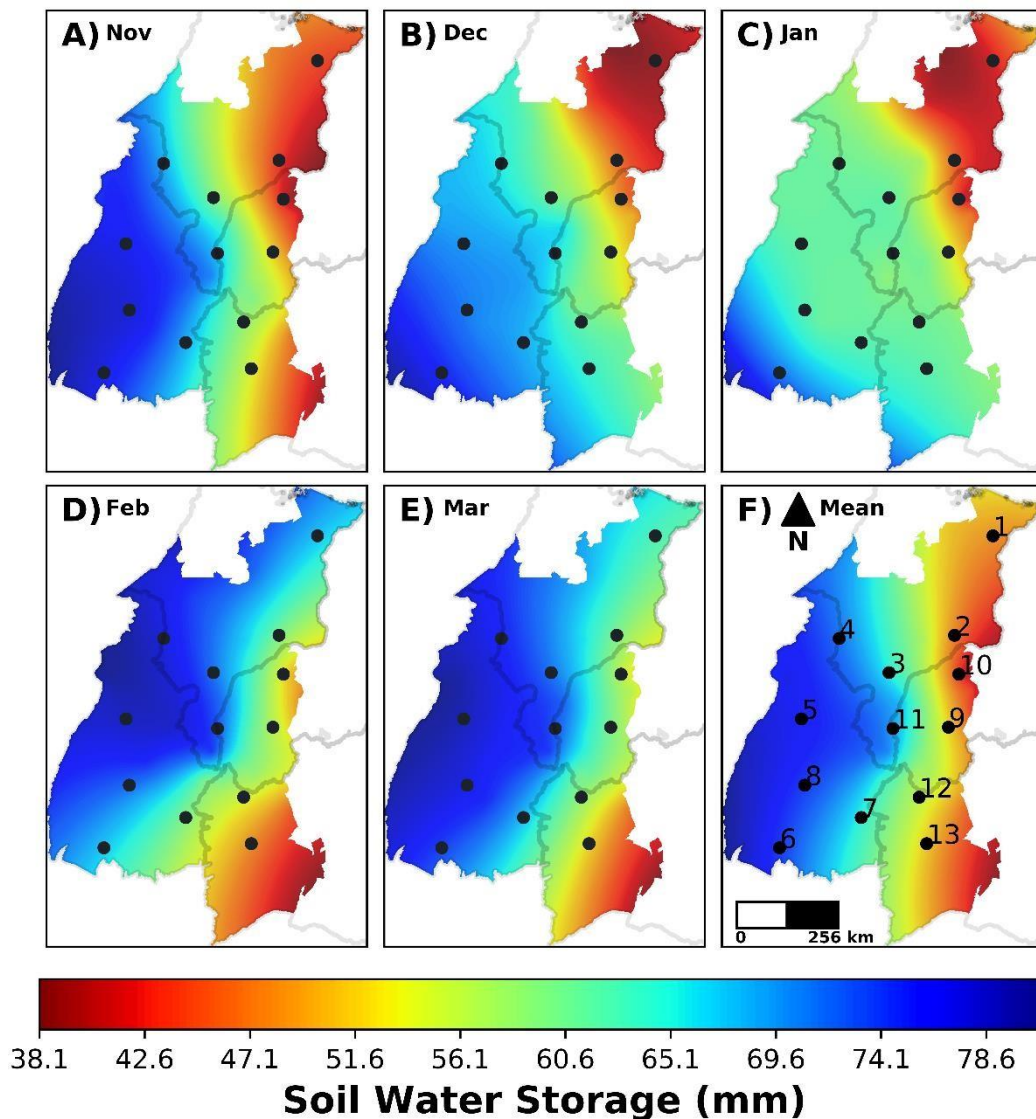


Figura 9: Armazenamento de água no solo nos meses de A) novembro, B) dezembro, C) janeiro, D) fevereiro, E) março e F) armazenamento médio do período. 1) Anapurus-MA, 2) Sucupira do Norte-MA, 3) Balsas-MA, 4) Estreito-MA, 5) Guaraí-TO, 6) Alvorada-TO, 7) Dianópolis-TO, 8) Porto Nacional-TO, 9) B Jesus-PI, 10) Sebastião Leal-PI, 11) Alto Parnaíba-MA, 12) Formosa do Rio Preto-BA, 13) São Desidério-BA.

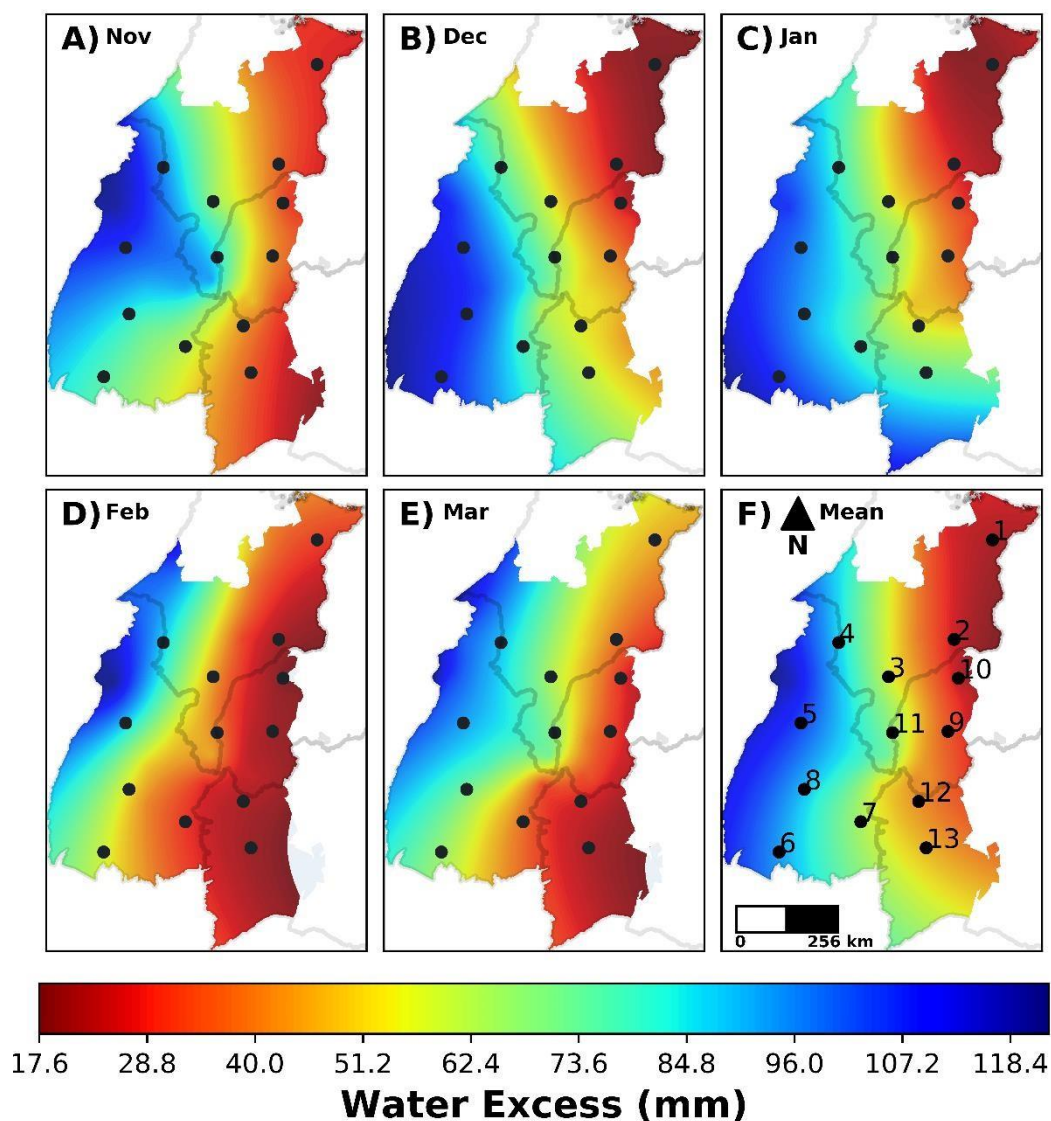


Figura 10: Excedente hídrico nos meses de A) novembro, B) dezembro, C) janeiro, D) fevereiro, E) março e F) excedente médio do período. 1) AnapurusMA, 2) Sucupira do Norte-MA, 3) Balsas-MA, 4) Estreito-MA, 5) Guaraí-TO, 6) Alvorada-TO, 7) Dianópolis-TO, 8) Porto Nacional-TO, 9) B Jesus-PI, 10) Sebastião Leal-PI, 11) Alto Parnaíba-MA, 12) Formosa do Rio Preto-BA, 13) São Desidério-BA.

Os estádios fenológicos são extremamente importantes para a avaliação dos efeitos e interferências das variáveis climáticas e componentes do balanço hídrico no ciclo do cultivo. Compreender a relação entre o clima e a produtividade dos cultivos é um primeiro passo crítico antes que modelos mais elaborados possam ser usados. (Schlenker, 2009).

A correlação entre as variáveis climáticas para cada estágio fenológico e a produtividade de soja para o local mais seco (Anapurus, MA) e para o mais úmido (Alvorada, TO) do MATOPIBA apresentaram características distintas. Para Anapurus (MA) (Figura 11 – A) representando um local seco, a correlação entre as condições meteorológicas mensais e a produtividade indicou correlação negativa de P e STO e a produtividade durante a semeadura e emergência. Durante o estágio reprodutivo R1 a R5, P e STO passa a ter correlação positiva, indicando que para um possível aumento no armazenamento haverá efeitos benéficos em relação a produtividade final da soja. Como esta localidade tem um clima seco, a correlação entre DEF e produtividade foi negativa deste o estágio S-E até o R7 como já esperado, indicando que sempre que houver uma diminuição do DEF haverá um aumento da produtividade.

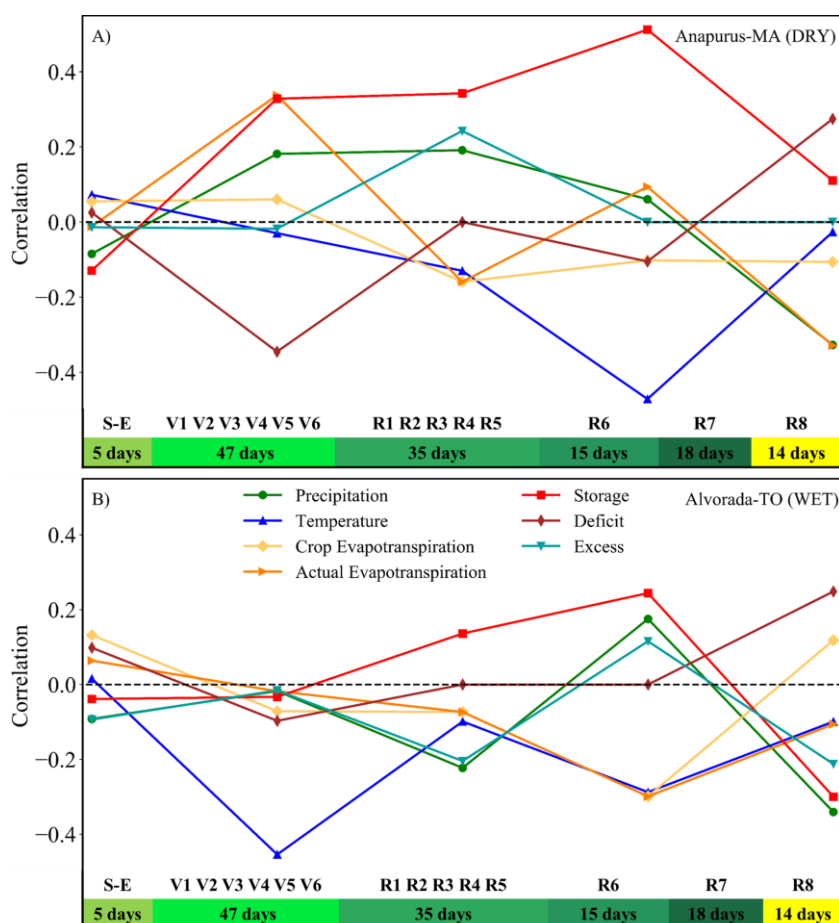


Figura 11: Correlação entre as variáveis meteorológicas mensais e a produtividade de soja para o município de: A) Anapurus – MA com clima seco e B) Alvorada – TO com clima úmido.

Para Alvorada (TO) (Figura 11 – B), uma localidade representativamente úmida, tanto P quanto STO durante a semeadura e germinação, a correlação é negativa indicando que se houver aumento no volume de chuva promoverá efeitos negativos a produtividade do cultivo isso ocorre porque o armazenamento de água no solo já pode ser considerado alto. Nota-se que P perde importância no final do ciclo. A temperatura para esta localidade apresenta correlação negativa durante todo o ciclo com exceção do estágio S – E.

Em Alvorada (TO) o armazenamento de água no solo já é mais que suficiente para suprir as necessidades do cultivo, e que seu aumento durante os estádios S-E e vegetativo, provocaria um possível encharcamento levando a redução de oxigênio para as raízes. O encharcamento modifica as relações bióticas do solo, pois promove deficiência de oxigênio, reduz a respiração aeróbica e promove indisponibilidade de determinados nutrientes para a planta (Dutra et al., 1995).

Por outro lado, durante o estágio reprodutivo a correlação entre STO e produtividade apresenta-se positiva demonstrando provável aumento na produtividade da soja.

STO é importante, principalmente, em dois períodos de desenvolvimento da soja: germinação-emergência e floração-enchimento de grãos. Durante o primeiro período, tanto o EXC quanto o DEF são prejudiciais ao estabelecimento do cultivo e à obtenção de uma boa uniformidade na população de plantas, sendo o EXC mais limitante do que o DEF. (Monteiro et al 2009)

Para os dois locais Anapurus – MA e Alvorada – TO a correlação entre a temperatura e a produtividade demonstrou-se negativa durante os estádios vegetativo e reprodutivo. Choi (2016), observa que os componentes da produção de soja são negativamente correlacionados e dependentes da temperatura e que a variação da temperatura entre os estádios fenológicos da soja principalmente nos estádios reprodutivos R1 a R7 não provoca efeitos significativos na produtividade final desde que não ocorra deficiência hídrica.

A compreensão da dinâmica das variáveis independentes (features) é essencial para escolha dos métodos mais adequados para o ajuste dos hiperparâmetros de uma rede neural artificial. O ajuste final dos hiperparâmetros configura-se como a minimização do RMSE e R^2 próximo a 1 (Tabela 4).

Tabela 4: Hiperparâmetros ajustados e métricas de desempenho dos modelos de Redes Neurais Artificiais Profundas para estimação e previsão de produtividade de soja para o MATOPIBA.

| Hyperparameters | Estimation | Forecasting |
|--|-------------------|--------------------|
| hidden_layer_sizes (number of neurons) | (18,17,6) | (31, 16, 20) |
| learning_rate_init | 0.01 | 0.001 |
| learning_rate | adaptive | constant |
| activation | relu | relu |
| solver | lbfgs | lbfgs |
| alpha | 0.05 | 0.05 |
| random_state | 11 | 0 |
| max_iter | 4000 | 4000 |
| R ² | 0.88 | 0.86 |
| RMSE | 167.85 | 185.85 |

O ajuste do modelo de DNN para estimação (Figura 12) de produtividade apresentou $R^2 = 0.88$ e $RMSE = 167.85 \text{ Kg ha}^{-1}$, demonstrando robustez das DNNs na estimação de produtividade da soja e identificação da relação entre as variáveis meteorológicas e de balanço hídrico e produtividade final.

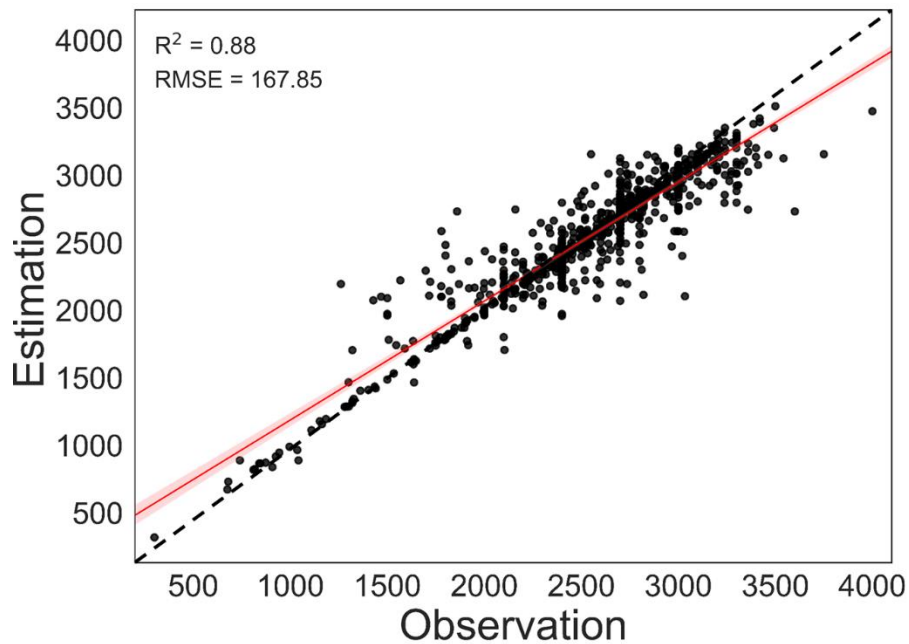


Figura 12: Desempenho do modelo de Redes Neurais Profundas (DNNs) para estimação de produtividade de soja para o MATOPIBA.

A espacialização dos resultados estimados (Figura 13) mostra que a DNN estima com erros (OBS-EST) próximos a zero em grande parte do território do MATOPIBA (Figura 13 – B). Em algumas localidades a DNN subestima a produtividade mais acentuadamente na região norte do Tocantins em até 188 kg ha^{-1} . Na região leste do MATOPIBA caracterizada por sua baixa produtividade entre 2280 a 2448 kg ha^{-1} , a DNN prevê resultados próximos aos observados.

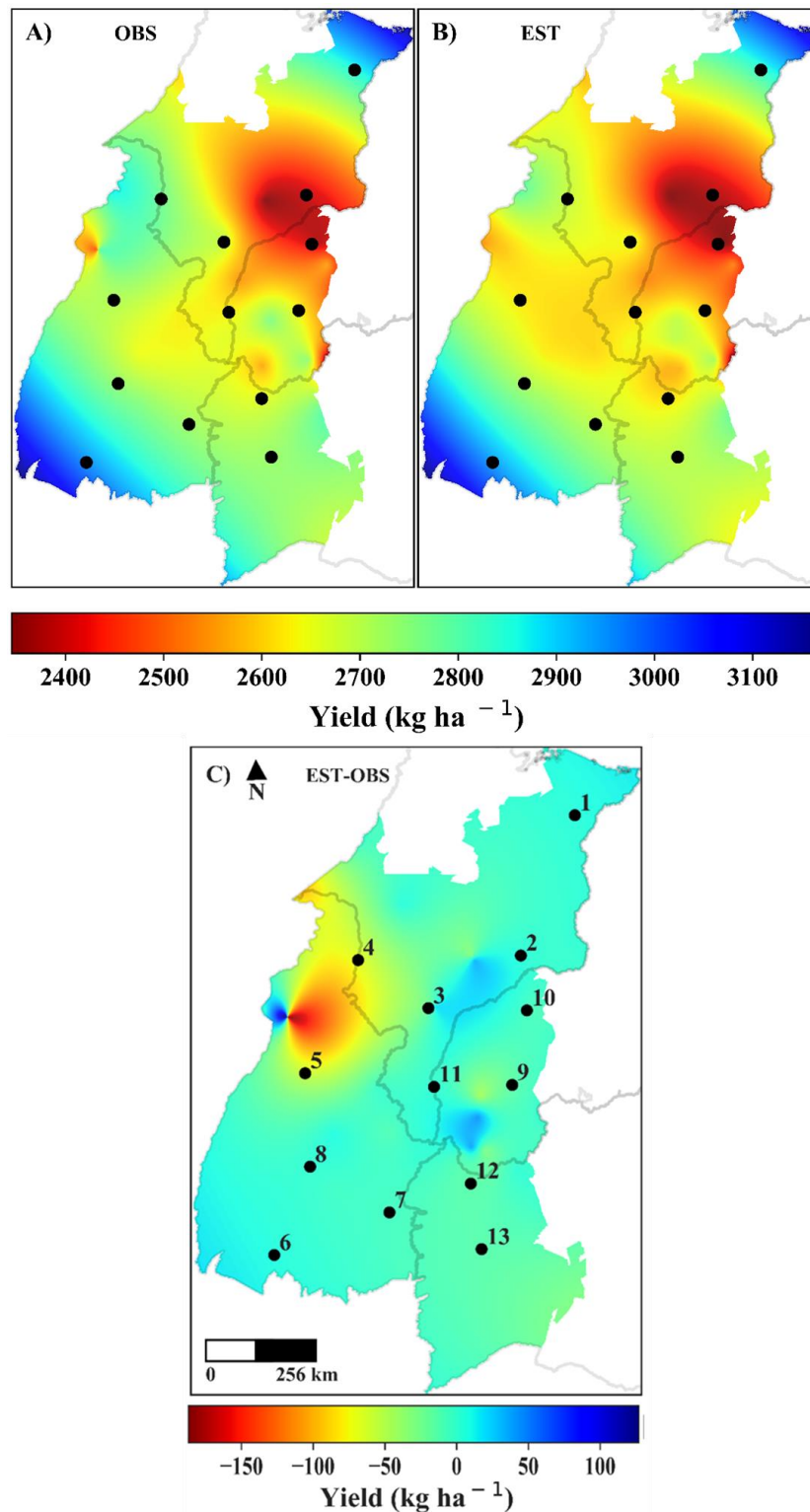


Figura 13: A) Produtividade observada (OBS), B) produtividade estimada (EST) e C) diferença entre a produtividade estimada e observada. 1) Anapurus-MA, 2) Sucupira do Norte-MA, 3) Balsas-MA, 4) Estreito-MA, 5) Guaraí-TO, 6) Alvorada-TO, 7) Dianópolis-TO, 8) Porto Nacional-TO, 9) B Jesus-PI, 10) Sebastião Leal-PI, 11) Alto Parnaíba-MA, 12) Formosa do Rio Preto-BA, 13) São Desidério-BA.

O histograma de erros da estimação da DNN (Figura 14) apresentou uma distribuição normal com grande parte das. Cerca de 27 localidades apresentaram erro superestimados em até 100 kg ha⁻¹. 24 municípios apresentam produtividade subestimada em até 50 kg ha⁻¹, e apenas 2 municípios a DNN ajustada para estimação, subestimou os resultados entre 100 e 200 kg ha⁻¹.

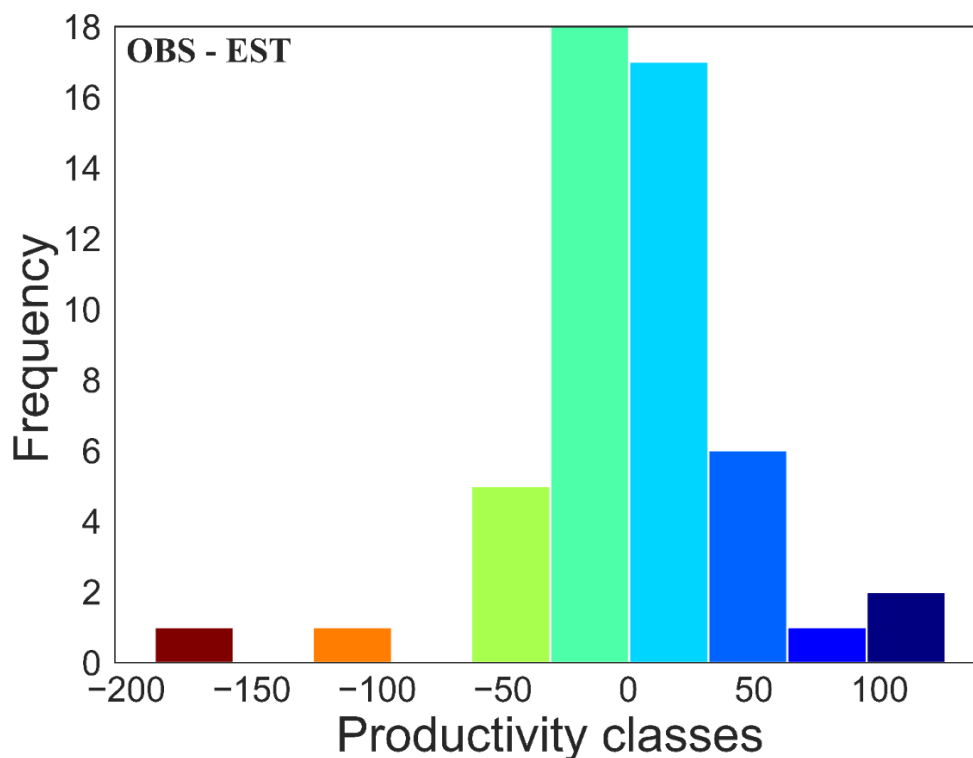


Figura 14: Distribuição de Erros de estimação por redes neurais profundas para a região do MATOPIBA. Observação: as cores correspondem a escala da figura 13-C.

Já o ajuste do modelo de DNN (Figura 15) para previsão de produtividade apresentou $R^2 = 0.86$ e $RMSE = 185.86 \text{ Kg ha}^{-1}$, subestimando a produtividade de soja para as localidades com produtividades inferiores a 1500 kg ha^{-1} .

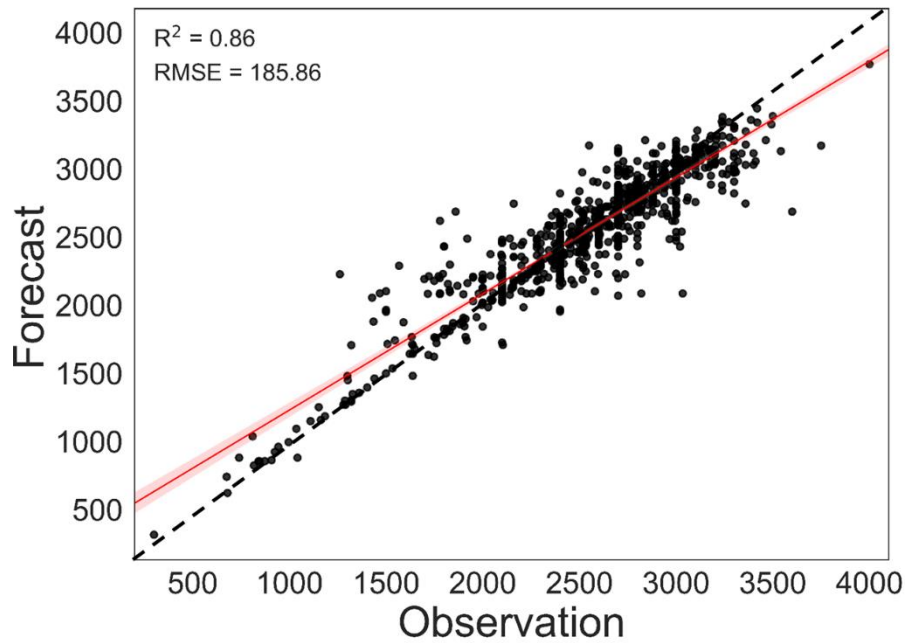


Figura15: Ajuste do modelo de Redes Neurais Profundas (DNNs) para previsão de produtividade de soja.

A espacialização dos resultados previstos (Figura 16) mostra que a DNN, mantém o bom resultado de RMSE, com diferença de aproximadamente 18 kg ha⁻¹ entre RMSE de estimação e previsão. Com isso, existem poucas diferenças entre as Figuras 13 – C e 16 – C.

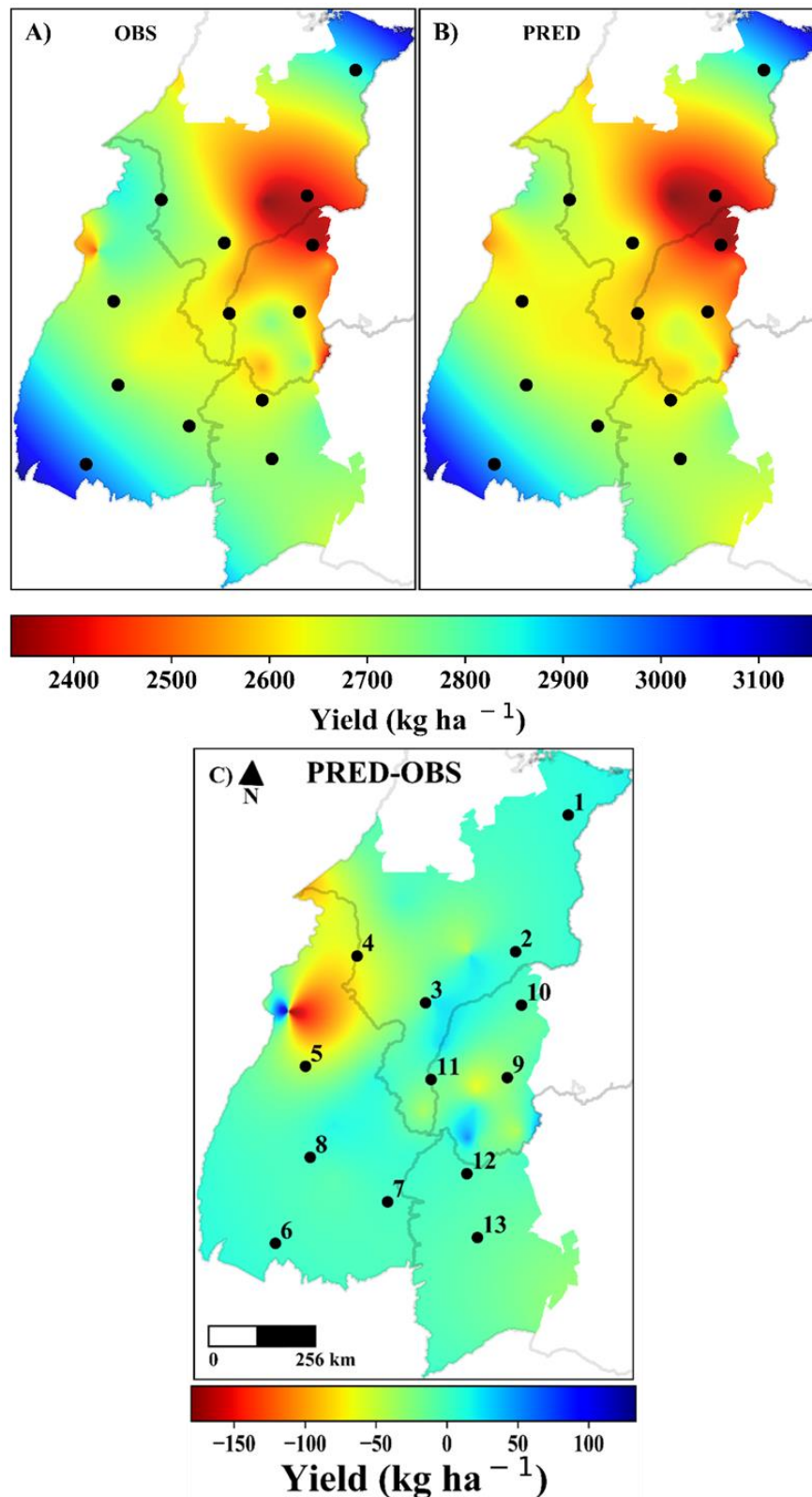


Figura 16: A) Produtividade observada, B) produtividade predita e C) diferença entre a produtividade predita e observada. 1) Anapurus-MA, 2) Sucupira do Norte-MA, 3) Balsas-MA, 4) Estreito-MA, 5) Guaraí-TO, 6) Alvorada-TO, 7) Dianópolis-TO, 8) Porto Nacional-TO, 9) B Jesus-PI, 10) Sebastião Leal-PI, 11) Alto Parnaíba-MA, 12) Formosa do Rio Preto-BA, 13) São Desidério-BA.

O histograma de erros da previsão da DNN (Figura 17) mostrou que 22 localidades obtiveram produtividades com erros até 100 kg ha^{-1} , 25 municípios apresentaram produtividade subestimada em até 50 kg ha^{-1} e cerca de 5 municípios com produtividade subestimada em mais de 150 kg ha^{-1} .

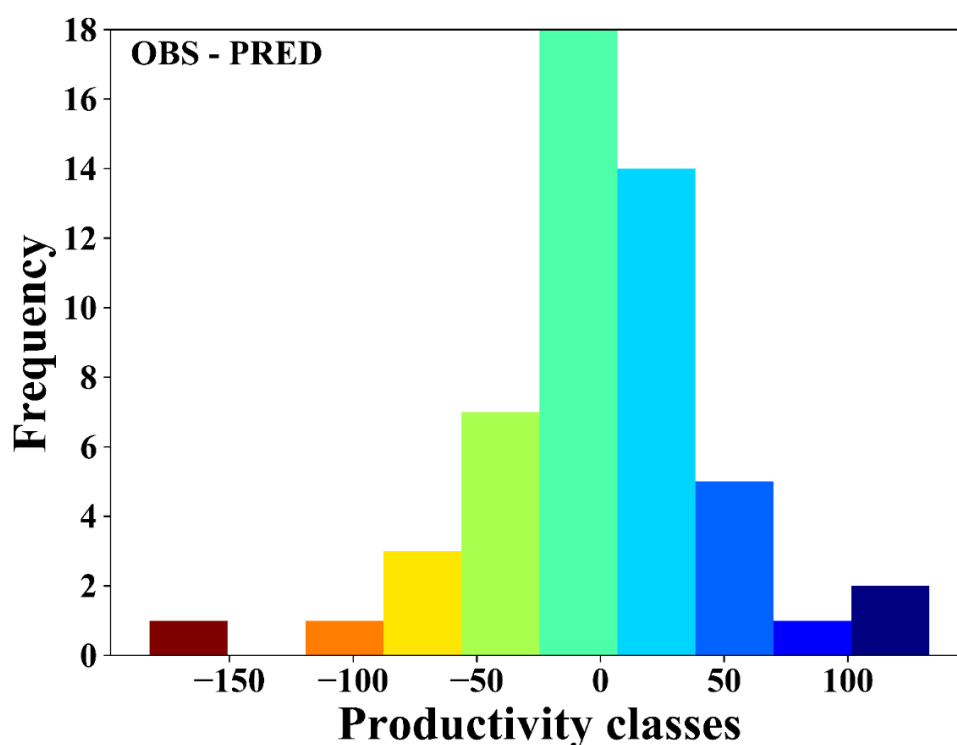


Figura 17: Distribuição de Erros de previsão por redes neurais profundas para a região do MATOPIBA. Observação: as cores correspondem a escala da figura 16-C.

Os valores de RMSE tanto para estimação quanto para previsão estão abaixo dos encontrados por Teixeira (2019), utilizando o modelo CROPGRO – soja (DSSAT) para avaliar entre outros parâmetros a produtividade da soja para localidades do Brasil obtendo RMSE de 347 kg ha^{-1} e também por Silva et al., (2019) que obtiveram RMSE de aproximadamente 218 kg ha^{-1} utilizando o modelo AquaCrop (FAO) para o MATOPIBA.

Seria interessante a aplicação de DNNs de forma regionalizada dentro da área do MATOPIBA para a minimização dos erros e melhoria das respostas, entretanto o número de dados disponíveis não suporta essa análise. De forma

geral é interessante ter pelo menos 10 dados de repetição para cada feature no modelo como sugerido por Haykin (1999).

As diferenças entre os resultados da estimação e da previsão são visíveis nas figuras 12 e 14 nos itens B, confirmando mais uma vez que as DNN são capazes de compreender as correlações entre as variáveis em uma série histórica, sendo possível resolver problemas complexos envolvendo um grande número de variáveis e obter bom desempenho do modelo. Essa avaliação é corroborada por Rhee et al., (2017) e por Chunjing et al., (2017) que afirmam que os modelos de aprendizado de máquina compreendem as limitações do conjunto de dados para a previsão além de possuírem uma superioridade única no reconhecimento de padrões em sistemas complexos.

2.4 Conclusões

As Redes Neurais Artificiais Profundas (DNNs) ajustadas foram capazes de estimar com $R^2=0.88$ e $RMSE= 167.85 \text{ kg ha}^{-1}$ e prever com $R^2=0.86$ e $RMSE= 185.86 \text{ kg ha}^{-1}$ a produtividade da soja para o MATOPIBA.

Foi possível a previsão de produtividade em até 2 meses antes da colheita utilizando DNNs.

A DNN mostrou-se robusta, uma vez que atingiu resultados acurados usando apenas variáveis de clima e componentes do balanço hídrico. Isso mostra a importância dos fatores climáticos na produtividade de soja na região do MATOPIBA.

REFERÊNCIAS

Allen RG, Pereira LS, Raes D, Smith M. Crop evapotranspiration: Guidelines for computing crop water requirements. Rome: FAO, 1998. 300 p. (FAO – Irrigation and Drainage Paper, 56).

Alves GR, Teixeira IR, Melo FR, Souza, RTG, Silva, AG (2018). Estimating soybean yields with artificial neural networks. **Acta Scientiarum Agronomy**, 40. <http://dx.doi.org/10.4025/actasciagron.v40i1.35250>

Anochi JÁ, Velho HC (2016). Mesoscale Precipitation Climate Prediction for Brazilian South Region by Artificial Neural Networks. **American Journal of Environmental Engineering**, 6(4A), 94-102. doi: 10.5923/s.ajee.201601.14

Aparecido LEO, Rolim GS, Lamparelli RAC, Souza PS, Santos, ER (2017). Agrometeorological models for forecasting coffee yield. **Agronomy Journal**, 109(1), 249-258. doi:10.2134/agronj2016.03.0166

Araújo, ECD, Uribe-Opazo MA, Johann JA (2014). Modelo de regressão espacial para estimativa da produtividade da soja associada a variáveis agrometeorológicas na região oeste do estado do Paraná. **Engenharia Agrícola**, 34(2), 286-299. <http://dx.doi.org/10.1590/S0100-69162014000200010>

CAMARGO AP. Balanço hídrico no Estado de São Paulo. Campinas: IAC, 1971. 24 p. (Boletim Técnico, 116)

Ceglar A, Toreti A, Lecerf R, Van der Velde M, Dentener F (2016). Impact of meteorological drivers on regional inter-annual crop yield variability in France. **Agricultural and forest meteorology**, 216, 58-67. <https://doi.org/10.1016/j.agrformet.2015.10.004>

Choi DH, Ban HY, Seo BS, Lee KJ, Lee BW (2016). Phenology and seed yield performance of determinate soybean cultivars grown at elevated temperatures in a temperate region. **PloS one**, 11(11), e0165977. <https://doi.org/10.1371/journal.pone.0165977>

Chunjing Y, Yueyao Z, Yaxuan Z, Liu H (2017). Application of convolutional neural network in classification of high resolution agricultural remote sensing images. The International Archives of the Photogrammetry, **Remote Sensing and Spatial Information Sciences**, 42. doi: 10.5194/isprs-archives-XLII-2-W7989-2017

Companhia Nacional De Abastecimento (CONAB). 2018. Perspectivas para a agropecuária, v.6, Safra 2018/2019, Brasília. Disponível em:

https://www.conab.gov.br/perspectivas-para-aagropecuaria/item/download/22780_ee707c6e6d44f06fe7b6a86ce6141652. Acesso em: 13 de novembro de 2019.

Companhia Nacional De Abastecimento (CONAB). 2019. Calendário de Plantio e Colheita de Grãos no Brasil 2019, Brasília. 73 p. Disponível em: <https://www.conab.gov.br/institucional/publicacoes/outras-publicacoes/item/7694-calendario-agricola-plantio-e-colheita.pdf>. Acesso em: 13 de novembro de 2019.

De Boeck H J, Kimball BA, Miglietta F, Nijs I (2012). Quantification of excess water loss in plant canopies warmed with infrared heating. **Global change biology**, 18(9), 2860-2868. <https://doi.org/10.1111/j.1365-2486.2012.02734.x>

Diogo AA, Ferraz SET, Streck NA, Cardoso ADO, Alencar JZ (2019). Indicadores Climáticos e a Produtividade de Soja no Rio Grande do Sul. **Revista Brasileira de Meteorologia**, 34(2), 191-200. <http://dx.doi.org/10.1590/0102-77863340024>

Dutra LF, Tavares S, Sartoretto L, Vahl L (1995). Resposta do feijoeiro ao fósforo em dois níveis de umidade no solo. **Current Agricultural Science and Technology**, 1(2). <http://dx.doi.org/10.18539/cast.v1i2.125>

Evangelista BA, da SILVA FAM, Simon J, Campos LJM (2017). Zoneamento de risco climático para determinação de épocas de semeadura da cultura da soja na região MATOPIBA. Embrapa Pesca e Aquicultura-Boletim de Pesquisa e Desenvolvimento (INFOTECA-E).

Farias JRB, Nepomuceno AL, Neumaier N (2007). Ecofisiologia da soja. Embrapa Soja-Circular Técnica (INFOTECA-E).

Filippi P, Jones EJ, Wimalathunge NS, Somarathna PD, Pozza LE, Ugbaje SU, Bishop TF (2019). An approach to forecast grain crop yield using multi-layered, multi-farm data sets and machine learning. **Precision Agriculture**, 1-15.

Haykin S (1999). Multilayer perceptrons. **Neural networks: a comprehensive foundation**, 1999, 2: 156-255.

Hoogenboom G, Jones JW, Wilkens PW, Porter CH, Hunt LA, Singh U, Koo J (2015). Decision Support System for Agrotechnology Transfer (DSSAT) version 4.5. DSSAT Foundation. Prosser, Washington. <http://dssat.net>.

Hussain MZ, Hamilton SK, Bhardwaj AK, Basso B, Thelen KD, Robertson GP (2019). Evapotranspiration and water use efficiency of continuous maize and maize and soybean in rotation in the upper Midwest US. **Agricultural Water Management**, 221, 92-98. <https://doi.org/10.1016/j.agwat.2019.02.049>

IBGE, (2018). Produção Agrícola Municipal – PAM In: IBGE. Sidra: sistema IBGE de recuperação automática. Rio de Janeiro, 2019. Disponível em: <https://sidra.ibge.gov.br/tabela/1612>

Jones JW, Antle JM, Basso B, Boote KJ, Conant RT, Foster I, Keating BA (2017). Toward a new generation of agricultural system data, models, and knowledge products: State of agricultural systems science. **Agricultural systems**, 155, 269288. <https://doi.org/10.1016/j.agsy.2016.09.021>

Jumrani K, Bhatia VS (2018). Impact of combined stress of high temperature and water deficit on growth and seed yield of soybean. **Physiology and Molecular biology of Plants**, 24(1), 37-50. <https://doi.org/10.1007/s12298-017-0480-5>

Kamilaris A, Kartakoullis A, Prenafeta-Boldú FX (2017). A review on the practice of big data analysis in agriculture. **Computers and Electronics in Agriculture**, 143, 23-37. <https://doi.org/10.1016/j.compag.2017.09.037>

Khaki S, Wang L (2019). Crop yield prediction using deep neural networks. **Frontiers in plant science**, 10. doi: 10.3389/fpls.2019.00621

Kingma DP, Ba J (2014). Adam: A method for stochastic optimization. **arXiv preprint arXiv:1412.6980**.

Lamba V, Dhaka VS (2014). Wheat yield prediction using artificial neural network and crop prediction techniques (A Survey). **IJRASET**, 2(IX).

Liakos KG, Busato P, Moshou D, Pearson S, Bochtis D (2018). Machine learning in agriculture: A review. **Sensors**, 18(8), 2674. <https://doi.org/10.3390/s18082674>

Martin TN, Dourado Neto D, Vieira Junior PA, Pereira AR, Manfron PA, Christoffoleti PJ (2012). Modified crop model estimation of depleted and potential soybean yield. **Acta Scientiarum. Agronomy**, 34(4), 369-378. <http://dx.doi.org/10.4025/actasciagron.v34i4.15274>

Miranda EE. Matopiba: desenvolver a agricultura ou os agricultores? *Correio Brasiliense*, 30 abr. 2015.

Monteiro JE (2009). Agrometeorologia dos cultivos: o fator meteorológico na produção agrícola. Brasília, DF: INMET.

Monteiro LA, Sentelhas PC (2014). Calibration and testing of an agrometeorological model for the estimation of soybean yields in different brazilian regions. **Acta Scientiarum. Agronomy** . 36(3), 265-272.

Nath A, Karunakar AP, Wanjari S, Kumar A (2018). Application of the

CROPGRO-Soybean Model (DSSAT V 4.5) for the Assessment of the Environmental Modification Effect on Soybean Productivity in the Akola region of Vidarbha, India. **International Journal of Bio-Resource & Stress Management**, 9(4). <https://doi.org/10.23910/IJBSM/2018.9.4.1865b>

Oikawa RT, Ishiki HM (2013). Modelos estatísticos e de redes neurais artificiais utilizados na previsão de precipitação. *Periódico Eletrônico Fórum Ambiental da Alta Paulista*, ISSN 1980-0827. <http://dx.doi.org/10.17271/19800827982013463>

Paredes P, Wei Z, Liu Y, Xu D, Xin Y, Zhang B, Pereira LS (2015). Performance assessment of the FAO AquaCrop model for soil water, soil evaporation, biomass and yield of soybeans in North China Plain. **Agricultural Water Management**, 152, 57-71. <https://doi.org/10.1016/j.agwat.2014.12.007>

Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Vanderplas J (2011). Scikit-learn: Machine learning in Python. **Journal of machine learning research**, 12(Oct), 2825-2830.

Reboita MS, Gan MA, da Rocha RP, Ambrizzi T (2010). Regimes de precipitação na América do Sul: uma revisão bibliográfica. **Revista Brasileira de Meteorologia**, 25(2).

Rhee J, Im J (2017). Meteorological drought forecasting for ungauged areas based on machine learning: Using long-range climate forecast and remote sensing data. **Agricultural and Forest Meteorology**, 237, 105-122. <https://doi.org/10.1016/j.agrformet.2017.02.011>

Robbins, H, Monro, S. A Stochastic Approximation Method. *The Annals of Mathematical Statistics*, Vol. 22, No. 3. (Sep., 1951), pp. 400-407.

Schlenker W, Roberts M J (2009). Nonlinear temperature effects indicate severe damages to US crop yields under climate change. **Proceedings of the National Academy of sciences**, 106(37), 15594-15598. <https://doi.org/10.1073/pnas.0906865106>

Sentelhas PC, Battisti R, Câmara GMS, Farias JRB., Hampf AC, Nendel C (2015). The soybean yield gap in Brazil—magnitude, causes and possible solutions for sustainable production. **The journal of agricultural science**, 153(8), 1394-1411. <https://doi.org/10.1017/S0021859615000313>

Silva, VPR, Silva, RA, Maciel, GF, Souza, EP, Braga, CC, Holanda, RM. (2020). Soybean yield in the Matopiba region under climate changes. **Revista Brasileira de Engenharia Agrícola e Ambiental**, 24(1), 8-14. Epub December 09, 2019. <https://dx.doi.org/10.1590/1807-1929/agriambi.v24n1p8-14>.

Shook J, Wu L, Gangopadhyay T, Ganapathysubramanian B, Sarkar S, Singh, AK (2018). Integrating genotype and weather variables for soybean yield prediction using deep learning. **bioRxiv**. doi: 10.1101/331561

Stackhouse PW, Westberg D, Hoell JM, Chandler WS, Zhang T (2015). Prediction Of Worldwide Energy Resource (POWER)-Agroclimatology Methodology-(1.0 Latitude by 1.0 Longitude Spatial Resolution). Technical Report of NASA Langley Research Center and SSAI/NASA Langley Research Center, 1-46.

Stackhouse PW, Jr, Minnis P, Perez R, et al., 2016: An Assessment of New Satellite Data Products for the Development of a Long-term Global Solar Resource at 10-100 km. ASES National Solar Conference, San Francisco, CA, 10-13 July, 2016.

Sun J, Di L, Sun Z, Shen Y, Lai Z (2019). County-Level Soybean Yield Prediction Using Deep CNN-LSTM Model. **Sensors**, 19(20), 4363. <https://doi.org/10.3390/s19204363>

Teixeira, WWR, Battisti, R, Sentelhas, PC, de Moraes, MF, de Oliveira Junior, A. (2019). Uncertainty assessment of soya bean yield gaps using DSSAT-CSM-CROPGRO-Soybean calibrated by cultivar maturity groups. **Journal of Agronomy and Crop Science**, 205(5), 533-544.

Thornthwaite CW, Mather JR (1955). The water balance. Centerton, NJ: Drexel Institute of Technology - Laboratory of Climatology. 104p. (Publications in Climatology, vol. VIII, n.1)

Wiebbecke, CE, Graham MA, Cianzio SR, Palmer RG (2012). Day temperature influences the male-sterile locus ms9 in soybean. **Crop science**, 52(4), 15031510. doi:10.2135/cropsci2011.08.0410

Xu Y, Yu L, Cai Z, Zhao J, Peng D, Li C, Gong P (2019). Exploring intra-annual variation in cropland classification accuracy using monthly, seasonal, and yearly sample set. **International Journal of Remote Sensing**, 1-16. <https://doi.org/10.1080/01431161.2019.1620377>

You J, Li X, Low M, Lobell D, Ermon S (2017, February). Deep gaussian process for crop yield prediction based on remote sensing data. In Thirty-First AAAI Conference on Artificial Intelligence.

Yuan W, Chen Y, Xia J, Dong W, Magliulo V, Moors E, Zhang H (2016). Estimating crop yield using a satellite-based light use efficiency model. **Ecological indicators**, 60, 702-709. <https://doi.org/10.1016/j.ecolind.2015.08.013>