

**UNIVERSIDADE ESTADUAL PAULISTA
FACULDADE DE CIÊNCIAS AGRÁRIAS E VETERINÁRIAS
CÂMPUS DE JABOTICABAL**

**UTILIZAÇÃO DE DIFERENTES TÉCNICAS DE MACHINE
LEARNING NO PROCESSO DE DETERMINAÇÃO DA
EMIÇÃO DA CO₂ DO SOLO, EM FUNÇÃO DOS
ATRIBUTOS DO SISTEMA SOLO-PLANTA-ATMOSFERA**

PAULO ALEXANDRE DA SILVA
Mestre em Agronomia

2023

**UNIVERSIDADE ESTADUAL PAULISTA
FACULDADE DE CIÊNCIAS AGRÁRIAS E VETERINÁRIAS
CÂMPUS DE JABOTICABAL**

**UTILIZAÇÃO DE DIFERENTES TÉCNICAS DE MACHINE
LEARNING NO PROCESSO DE DETERMINAÇÃO DA
EMIÇÃO DA CO₂ DO SOLO, EM FUNÇÃO DOS
ATRIBUTOS DO SISTEMA SOLO-PLANTA-ATMOSFERA**

PAULO ALEXANDRE DA SILVA

Orientador: Prof. Dr. Glauco de Souza Rolim

Coorientador: Prof. Dr. Alan Rodrigo Panosso

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

2023

S586u

Silva, Paulo Alexandre da

Utilização de diferentes técnicas de machine learning no processo de determinação da emissão da CO₂ do solo, em função dos atributos do sistema solo-planta-atmosfera /Paulo Alexandre da Silva. -- Jaboticabal, 2023
209 p. : il., tabs., mapas, 2 v.

Tese (doutorado) - Universidade Estadual Paulista (Unesp), Faculdade de Ciências Agrárias e Veterinárias, Jaboticabal

Orientador: Glauco de Souza Rolim

Coorientador: Alan Rodrigo Panosso

1. Gases estufa. 2. Modelagem de dados. 3. Inteligência artificial. 4. Solos
Manejo. 5. Climatologia agrícola. I. Título.

CERTIFICADO DE APROVAÇÃO

TÍTULO DA TESE: UTILIZAÇÃO DE DIFERENTES TÉCNICAS DE MACHINE LEARNING NO PROCESSO DE DETERMINAÇÃO DA EMISSÃO DA CO₂ DO SOLO, EM FUNÇÃO DOS ATRIBUTOS DO SISTEMA SOLO-PLANTA-ATMOSFERA

AUTOR: PAULO ALEXANDRE DA SILVA

ORIENTADOR: GLAUCO DE SOUZA ROLIM

COORIENTADOR: ALAN RODRIGO PANOSSO

Aprovado como parte das exigências para obtenção do Título de Doutor em Agronomia (Ciência do Solo), pela Comissão Examinadora:

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

Prof. Dr. JOSÉ MARQUES JUNIOR (Participação Virtual)
Departamento de Ciências da Produção Agrícola / FCAV UNESP Jaboticabal

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

Angélica Santos Rabelo de Souza Bahia
Pós-doutoranda ANGÉLICA SANTOS RABELO DE SOUZA BAHIA (Participação Virtual)
Departamento de Solos e Adubos / FCAV / UNESP - Jaboticabal

Diego Silva Siqueira
Dr. DIEGO SILVA SIQUEIRA (Participação Virtual)
Bayer S/A / Ribeirão Preto/SP

Jaboticabal, 22 de setembro de 2022

Documento assinado digitalmente
 LUCAS EDUARDO DE OLIVEIRA APARECIDO
Data: 05/10/2022 13:55:08-0300
Verifique em <https://verificador.ig.br>

DADOS CURRICULARES DO AUTOR

PAULO ALEXANDRE DA SILVA – Filho do metalúrgico José Paulo da Silva, o qual teve a oportunidade de trabalhar na mesma empresa e o tem como o primeiro mestre profissional, e também da técnica de enfermagem Solange Aparecida Brandão Silva, que sempre foi sua grande fonte de inspiração pela sua força de vontade e determinação, sendo uma referência de resiliência e cuidado com as suas atividades, onde ambos não mediram esforços para que os seus três filhos, negros, estudantes da rede pública de ensino pudessem superar as barreiras sociais e realizar os sonhos de se formarem em universidades públicas renomadas (Universidade Estadual Paulista - UNESP, Universidade Estadual do Mato Grosso do Sul - UEMS e Universidade Federal Tecnológica do Paraná - UFTPR), onde sempre enfatizaram que a educação pode ser um agente transformador. Nasceu em São Carlos, São Paulo, no dia 18 de outubro de 1988. Em fevereiro de 2010, ingressou no curso de Engenharia Agrônômica pela Universidade Estadual Paulista – Faculdade de Ciências Agrárias e Veterinárias – Câmpus de Jaboticabal – SP. Foi bolsista do Programa de Educação Tutorial – PET do MEC, durante o período de agosto de 2011 até dezembro de 2013. Iniciou, em março de 2017, o Curso de Mestrado em Agronomia (Ciência do Solo), na Faculdade de Ciências Agrárias e Veterinárias – UNESP. No mês de julho de 2018, submeteu-se à banca para a defesa de Dissertação, sendo eu aprovado como Mestre em Agronomia. Em agosto de 2018 iniciou o Curso de Doutorado na mesma área, em Agronomia (Ciência do Solo), na Faculdade de Ciências Agrárias e Veterinárias – UNESP. Nesta mesma universidade, entre os anos de 2019 a 2022 foi o professor substituto das disciplinas de Estatística e Informática, onde ministrou aulas para as turmas de Biologia, Agronomia, Zootecnia, Administração e Medicina Veterinária. Trabalhou em um projeto de sustentabilidade agrícola na empresa Bayer no ano de 2022, que visa incentivar o manejo e práticas agronômicas que promovam o sequestro do carbono no solo e a maior produtividade das culturas. No mês de setembro de 2022, submeteu-se à banca para a defesa a Tese, como parte do processo para obter o título Doutor em Agronomia pelo programa de Ciência do Solo, na Faculdade de Ciências Agrárias e Veterinárias – UNESP Jaboticabal.

“Seja você quem for, seja qual for a posição social que você tenha na vida, a mais alta ou a mais baixa, tenha sempre como meta muita força, muita determinação e sempre faça tudo com muito amor e fé em Deus, que um dia você chega lá. De alguma maneira, você chega lá”.

(Ayrton Senna)

DEDICO

Primeiramente a Deus, aos meus pais José Paulo da Silva e Solange Aparecida Brandão Silva, os meus grandes exemplos e heróis. Aos meus irmãos Diego Bruno da Silva e Leonardo Augusto da Silva, que sempre me apoiaram e minha avó Maria Olga de Campos Brandão pelo carinho e palavras de motivação.

OFEREÇO

In memoriam aos meus amados avôs Lázaro Emílio da Silva, Alípio Paulino Brandão e minha avó Tereza Emílio da Silva pelos ensinamentos e pela simplicidade.

AGRADECIMENTOS

Agradeço em primeiro lugar a Deus, sem cuja vontade eu não teria realizado e concluído o curso, nem tido a chance de mudar minha vida e a de outras pessoas que estiveram e estarão envolvidas em meu trabalho.

Em segundo lugar, agradeço a duas pessoas que foram fundamentais para que eu concluísse esta fase de minha vida, os meus pais, Solange Aparecida Brandão Silva e José Paulo da Silva, que nestes anos dedicaram cada minuto a minha formação e a de meus irmãos e que, nas horas de dificuldades e turbulências, usaram de palavras, comportamentos e ensinamentos que foram essenciais para o meu crescimento e desenvolvimento.

Também agradeço aos meus irmãos, que sempre estiveram comigo, Diego Bruno e Leonardo Augusto, que nestes últimos anos também estudaram com o objetivo de se formarem em uma universidade de qualidade e agora estão colhendo seus frutos. E meus familiares, que estiveram presentes com gestos ou palavras de apoio, dando-me força para seguir em frente e estar aqui hoje.

A minha noiva Bruna, pela ajuda moral, companhia, carinho, colaboração, palavras de incentivo e paciência, que foram fundamentais para que eu pudesse concluir meu objetivo, e sendo assim muito importante para mim. Agradeço também a sua família, por me tratarem muito bem e ajudarem-me neste processo.

Ao meu orientador Prof. Dr. Glauco de Souza Rolim, que confiou em mim e que através do convívio, de seus conhecimentos e conselhos, ajudou-me neste processo de grande crescimento intelectual e pessoal, que com certeza serão de suma importância na minha carreira profissional. Além disso, também agradeço pelo apoio nos momentos de dificuldades, sendo bastante atencioso e estando sempre disposto a tirar minhas dúvidas e ajudar-me e a solucionar os problemas, com muita paciência e confiança, sendo um parceiro que colaborou muito com meu crescimento pessoal.

Ao meu coorientador Prof. Dr. Alan Rodrigo Panosso, antes meu orientador do mestrado e um grande exemplo de profissional, que durante todos esses anos de formação foi um grande incentivador e que através do seu conhecimento profissional e pessoal, pode me trazer diversos conhecimentos e conselhos, ajudando-me no processo da minha formação intelectual para o desenvolvimento deste trabalho e que, nos momentos de dificuldades, estive ali sempre à disposição para tirar minhas

dúvidas com muita paciência e confiança, disposto sempre a ajudar-me e a solucionar os problemas, sendo um grande parceiro que me auxiliou muito no minha jornada pessoal e acadêmica.

Ao professor Newton La Scala Júnior e Nelson José Peruzzi, pelos conselhos, confiança, e, principalmente, por todos os ensinamentos teóricos de física, matemática e emissão dos gases geradores do efeito estufa, que foram determinantes para a execução desta Dissertação.

Aos professores Euclides Braga Malheiros, Antônio Sergio Ferraudo, Danísio Prado Munari, que através de cujos ensinamentos teóricos em modelagem, programas estatísticos e seus conhecimentos sobre a manipulação de dados sempre estiveram dispostos quando os procurei e são grandes exemplos de profissionais e docentes que admiro muito.

Aos grupos Neuron - Data Science and Artificial Intelligence, Caracterização do Solo para fins de Manejo Específico (CSME) e Group of Agrometeorological Studies (GAS) e seus respectivos membros que me proporcionaram diversos ensinamentos que usarei em minha vida profissional e pessoal nas áreas da ciência de dados, gestão e relacionamento interpessoal.

Ao projeto Cursinho Ativo, que me proporcionou diversos ensinamentos, experiências que me ajudaram a crescer e a desenvolver habilidades que usarei em minha vida profissional e pessoal.

Não poderia esquecer daquelas pessoas especiais que, mesmo longe, estiveram sempre presentes e apoiando-me: Marjorie, Eduardo, minha avó Maria Olga, minhas madrinhas Abigail e Sônia e minhas tias Sílvia e Sandra.

Ao grupo de pesquisa, em especial a Maria Elisa, Ludhanna Veras, Katharine Batista, Kamila Meneses e Deise Nogueira, que através da amizade e dos estudos fizeram os dias de trabalho melhores, ajudando-me muito nesta fase com suas experiências.

E por fim, à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (Capes) – Código de Financiamento 001, que foi o órgão que financiou O as minhas pesquisas, tanto no mestrado, quanto no doutorado.

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

SUMÁRIO

RESUMO.....	iv
ABSTRACT.....	v
LISTA DAS ABREVIACÕES	vii
LISTA DAS TABELAS	ix
LISTA DE FIGURAS	xiii
CAPÍTULO 1 – CONSIDERAÇÕES GERAIS.....	1
1.1 INTRODUÇÃO	1
1.1.1. Gases do Efeito Estufa	1
1.1.2. Aspectos Gerais sobre o Fluxo de CO ₂	2
1.1.3. Emissão de gases causadores do efeito estufa nos sistemas Agrícolas.....	3
1.1.4. Impactos mundiais ligados à agricultura e aos Gases do Efeito Estufa	5
1.1.5. Uso da modelagem para a emissão de CO ₂ do solo	7
1.1.6. Algoritmos de aprendizado de máquina	8
1.2 OBJETIVO.....	13
1.3 JUSTIFICATIVA	13
CAPÍTULO 2 – DESENVOLVIMENTO DE UM MODELO MECANÍSTICO PARA PREDIZER A EMISSÃO DO CO ₂ DO SOLO, A PARTIR DA INFLUÊNCIA DAS VARIÁVEIS QUE COMPÕEM O SISTEMA SOLO-PLANTA-ATMOSFERA ..	14
RESUMO.....	14
CHAPTER 2 – DEVELOPMENT OF A MECHANISTIC MODEL TO PREDICT SOIL CO ₂ EMISSIONS, FROM THE INFLUENCE OF THE VARIABLES THAT COMPOSE THE SOIL-PLANT-ATMOSPHERE SYSTEM	15
ABSTRACT.....	15
2.1 INTRODUÇÃO	16
2.2 OBJETIVO.....	22
2.3 MATERIAL E MÉTODOS.....	22
2.3.1. Caracterização da Área de Estudo.....	22
2.3.2. Medições de FCO ₂ , Us e Ts	24
2.3.3. Análise de atributos físicos e químicos do solo	24
2.3.4. Estoque de carbono	25
2.3.5. Evapotranspiração potencial	26

2.3.6. Balanço hídrico climatológico	26
2.3.7. Obtenções das variáveis ambientais e variáveis do solo para a elaboração do modelo da FCO ₂	28
2.3.8. Metodologia da seleção dos pontos amostrais para a elaboração do modelo mecanístico da emissão de CO ₂ do solo	31
2.3.9. Elaboração do modelo mecanístico para a estimação do CO ₂ do solo	35
2.3.10. Medidas estatísticas de avaliação de desempenho	40
2.4 RESULTADOS E DISCUSSÃO	42
2.4.1. Distribuição dos dados e Estatística Descritiva	42
2.4.2. Correlações	46
2.4.3. Análise de Regressão Linear Múltipla	48
2.4.4. Regressão Linear Múltipla da MOS, em função das variáveis P, CTC, Tair, Ts, Dp e Us	49
2.4.5. MOS observado versus MOS estimado	57
2.4.6. Regressão Polinomial Quadrática e curva de caracterização da CAD, em função da Umidade do Solo	60
2.4.7. Umidade para a CAD de 40 mm observado versus umidade para a CAD de 40 mm estimada	69
2.4.8. Regressão Linear Múltipla da CAD, em função das variáveis Us, Ds, Dp e ARM	72
2.4.9. CAD observada versus CAD estimada	80
2.4.10. Regressão Linear Múltipla da Umidade do solo (Us), em função das variáveis ETP, Precipitação e ARM	82
2.4.11. Umidade do solo observada versus Umidade do solo estimada ...	90
2.4.12. Regressão Linear Múltipla da PLA, em função das variáveis PLA_BH e Dp	93
2.4.13. PLA observada versus PLA estimada	99
2.4.14. Predição do modelo final	101
2.4.15. Distribuição dos dados e Estatística Descritiva das variáveis preditas	101
2.4.16. Correlações das variáveis preditas	104
2.4.17. Regressão Linear Múltipla da FCO ₂ , em função das variáveis PLA_prevista e MOS_prevista	107
2.4.18. FCO ₂ observada versus FCO ₂ prevista	113
2.5 CONCLUSÕES	117

CAPÍTULO 3 – USO DAS TÉCNICAS DE MACHINE LEARNING PARA PREDIZER A EMISSÃO DE CO₂ DO SOLO DE CANAVIAIS COMERCIAIS, EM FUNÇÃO DOS ATRIBUTOS DO SISTEMA SOLO-PLANTA-ATMOSFERA.....	119
RESUMO.....	119
CHAPTER 3 – USE OF MACHINE LEARNING TECHNIQUES TO PREDICT CO₂ EMISSIONS FROM THE SOIL OF COMMERCIAL CANE FIELDS, AS A RESULT OF THE ATTRIBUTES OF THE SOIL-PLANT-ATMOSPHERE SYSTEM.....	120
ABSTRACT.....	120
3.1 INTRODUÇÃO	121
3.2 OBJETIVO.....	124
3.3 MATERIAL E MÉTODOS.....	124
3.3.1. Caracterização da Área de Estudo.....	124
3.3.2. Mensurações do CO₂, temperatura e umidade do solo	125
3.3.3. Análise dos Atributos Físicos e Químicos do Solo.....	126
3.3.4. Evapotranspiração potencial	127
3.3.5. Balanço hídrico climatológico	127
3.3.6. Obtenções das variáveis oriundas do sistema solo-planta-atmosfera que foram usadas nas predições dos modelos da FCO₂.....	129
3.3.7. Análise estatística	130
3.3.8. Software Computacional	132
3.3.9. Algoritmos de aprendizado de máquina	132
3.3.10. Medidas estatísticas de avaliação de desempenho	141
3.4 RESULTADOS E DISCUSSÃO	142
3.4.1. Distribuição dos dados e Estatística Descritiva.....	142
3.4.2. Correlações.....	147
3.4.3. Desempenho dos Algoritmos Preditivos	149
3.4.4. Treinamento dos modelos.....	150
3.4.5. Validação dos modelos.....	156
3.5 CONCLUSÕES	164
CAPÍTULO 4 – CONSIDERAÇÕES FINAIS.....	165
AGRADECIMENTOS.....	165
REFERÊNCIAS BIBLIOGRÁFICAS	166

RESUMO: Neste trabalho, foram gerados modelos para prever a emissão de CO_2 do solo (FCO_2), em função das variáveis que compõem o sistema solo-planta-atmosfera, utilizando técnicas de machine learning em áreas agrícolas cultivadas com cana-de-açúcar, nos estados de São Paulo e Mato Grosso do Sul. Os estudos foram realizados a partir de um conjunto de dados, oriundos de experimentos conduzidos em áreas de cana-de-açúcar nos municípios de Guariba (SP), Pradópolis (SP) e Aparecida do Taboado (MS). As determinações da FCO_2 foram registradas nas áreas de estudo pelo sistema LI-COR (LI-8100). No estudo 1 foi obtido o seguinte modelo mecanístico: $\text{FCO}_2 = 2,5938 + 0,0065 \cdot \text{matéria orgânica prevista} - 0,0401 \cdot \text{porosidade livre d'água prevista}$. O valor do coeficiente de determinação do modelo foi de $R^2 = 0,824$, e o seu respectivo valor ajustado foi $R^2_{\text{ajustado}} = 0,807$. No estudo 2 foram testados 11 algoritmos de machine learning para determinar a emissão do CO_2 do solo, em função dos atributos que compõem o sistema solo-planta-atmosfera e os algoritmos testados para prever a FCO_2 foram: RFT_GS = random forest com a seleção de variáveis pelo Grid Search, RFT = random forest regressor, TREE = decision tree regression, BOO = gradient boosting regressor, MLP = rede neural Multi-layer Perceptron regressor, KNN = k-nearest neighbors para regressão, SVM = regressor support vector machines, SVM_kpoli = support vector machines pelo método de kernel polinomial, RLM = regressão linear múltipla, SVM_VC = regressor support vector machines com validação cruzada, SGD = regressor Stochastic Gradient Descent. O algoritmo RFT_GS foi o que melhor estimou a FCO_2 , em função dos atributos do sistema solo-planta-atmosfera, seguido do RFT, da TREE, do MLP, do BOO, da KNN, do SVM, do SVM_kpoli, da RLM, da SVM_VC e da SGD. Conclui-se que estes estudos são importantes para os agricultores, poder público e privado, pois as informações geradas podem ser usadas no gerenciamento e nas tomadas de decisões do processo produtivo e financeiro, sendo uma opção para a caracterização não só da FCO_2 , mas também de outros atributos do sistema solo-planta-atmosfera, promovendo a redução de custos, o uso eficiente dos recursos infinitos e finitos, produtividade e maior sustentabilidade para o sistema produtivo.

Palavras-chave: atributos físicos e químicos do solo, machine learning, gases do efeito estufa, respiração do solo, emissão de CO_2 do solo.

ABSTRACT: In this work, models were generated to predict soil CO₂ emission (FCO₂), as a function of the variables that make up the soil-plant-atmosphere system, using machine learning techniques in agricultural areas cultivated with sugarcane, in the states of São Paulo and Mato Grosso do Sul. The studies were carried out from a set of data from experiments carried out in sugarcane areas in the municipalities of Guariba (SP), Pradópolis (SP) and Aparecida do Taboado (MS). FCO₂ determinations were recorded in the study areas by the LI-COR system (LI-8100). In study 1, the following mechanistic model was obtained: $FCO_2 = 2,5938 + 0,0065 \cdot \text{predicted organic matter} - 0,0401 \cdot \text{predicted water-free porosity}$. The value of the coefficient of determination of the model was $R^2 = 0,824$, and its respective adjusted value was $R^2 \text{ adjusted} = 0,807$. In study 2, 11 machine learning algorithms were tested to determine the emission of CO₂ from the soil, depending on the attributes that make up the soil-plant-atmosphere system and the algorithms tested to predict the FCO₂ were: RFT_GS = random forest with the selection of variables by Grid Search, RFT = random forest regressor, TREE = decision tree regression, BOO = gradient boosting regressor, MLP = Multi-layer neural network Perceptron regressor, KNN = k-nearest neighbors for regression, SVM = regressor support vector machines, SVM_kpoli = support vector machines by the kernel polynomial method, RLM = multiple linear regression, SVM_VC = regressor support vector machines with cross validation, SGD = regressor Stochastic Gradient Descent. The RFT_GS algorithm was the one that best estimated the FCO₂, as a function of the attributes of the soil-plant-atmosphere system, followed by RFT, TREE, MLP, BOO, KNN, SVM, SVM_kpoli, RLM, SVM_VC and the SGD. It is concluded that these studies are important for farmers, public and private authorities, as the information generated can be used in the management and decision-making of the production and financial process, being an option for the characterization not only of FCO₂, but also of other attributes of the soil-plant-atmosphere system, promoting cost reduction, efficient use of infinite and finite resources, productivity and greater sustainability for the production system.

Keywords: physical and chemical soil attributes, machine learning, greenhouse gases, soil respiration, soil CO₂ emission.

LISTA DAS ABREVIATÖES

A - Altos valores de FCO_2

ABC - Agricultura de Baixo Carbono

Al - Alumínio

APT - Aparecida do Taboado

Aw - Clima tropical com estação seca de Inverno

B - Baixos valores de F_m

C/N - Relação entre o carbono e o nitrogênio

Ca - Cálcio

Capes - Coordenação de Aperfeiçoamento de Pessoal de Nível Superior

CFCs - Clorofluorcarbonetos

CH₄ - Metano

CO - Teor de carbono orgânico oxidável

CO₂ - Dióxido de carbono

CONAB - Companhia Nacional de Abastecimento

CP1 - Componente principal 1

CP2 - Componente principal 2

CPs - Componentes Principais

CTC - Capacidade de troca de cátions

CV - Coeficiente de variação

D_p - Densidade das partículas

Dr. - Doutor

D_s - Densidade do solo

E - Espessura da camada estudada

EMBRAPA - *Empresa Brasileira de Pesquisa Agropecuária*

Estc - *Estoque do carbono no solo*

EUA - *Estados Unidos da América*

FBN - *Fixação Biológica de Nitrogênio*

FCO₂ - *Emissão do carbono via CO₂ do solo*

Fm - *Emissão do carbono via CO₂ média*

GEE - *Gases do efeito estufa*

GS+ - *Geostatistics Software*

GUA - *Guariba*

H - *Hidrogênio*

HCFCs - *Hidroclorofluorcarbonetos*

ILP - *Integração Lavoura-Pecuária*

ILPF - *Integração Lavoura-Pecuária-Floresta*

IPCC - *Intergovernmental Panel on Climate Change*

k - *Constante de decaimento do carbono*

K - *Potássio*

LI-8100 - *LI-COR 8100*

M - *Médios valores de Fm*

Macro - *Macroporosidade*

MCT - *Ministério da Ciência e Tecnologia*

MEC - *Ministério da Educação*

Mg - *Magnésio*

Micro – *Microporosidade*

MOS – *Matéria Orgânica do solo*

MOT - *Motuca*

MS - *Mato Grosso do Sul*

N₂O - *Óxido nitroso*

P - *Fósforo disponível no solo*

PAD - *Pradópolis*

PAG - *Potencial de aquecimento global*

PBMC - *Painel Brasileiro de Mudanças Climáticas*

PFCs - *Perfluorcarbonetos*

PLA - *Porosidade livre da água*

PLAm - *Porosidade livre da água média*

PmoA - *Metano mono-oxigenase A*

Prof. - *Professor*

PT - *Porosidade total*

PVC - *Policloreto de polivinila*

SB - *Soma de bases*

SF₆ - *Hexafluoreto de enxofre*

SP - *São Paulo*

TDR - *Time Domain Reflectometry*

T_m - *Temperatura média do solo*

T_s - *Temperatura do solo*

U_m - *Umidade do Solo*

UNESP - *Universidade Estadual Paulista*

U_s - *Umidade do solo*

V - *Índice de saturação por bases*

LISTA DAS TABELAS

Tabela 2.1: Teste de normalidade pelo método de Shapiro-wilk da emissão de CO ₂ , atributos atmosféricos e características físico-químicos do solo de todas as áreas estudadas.....	42
Tabela 2.2: Estatísticas descritivas da emissão de CO ₂ , atributos atmosféricos e características físico-químicos do solo de todas as áreas estudadas, N = 24	44
Tabela 2.3: Estatísticas descritivas da emissão de CO ₂ , atributos atmosféricos e características físico-químicos do solo, referentes às áreas de Motuca – SP (N = 3), Guariba – SP (N = 7), Pradópolis – SP (N = 8) e Aparecida do Taboado (N = 6)	45
Tabela 2.4: Coeficientes métricos do modelo da MOS, N = 24	50
Tabela 2.5: Demonstrações dos coeficientes lineares e angulares, resultados do p-valor em função da estatística t para o nível de significância $\alpha = 0,05$ com os seus respectivos valores críticos, para as variáveis P, CTC, Tair, Ts, Dp e Us que foram usadas na determinação do modelo da MOS, N = 24.....	52
Tabela 2.6: Coeficientes métricas para a determinação do modelo da MOS, N = 24	58
Tabela 2.7: Demonstrações do coeficiente linear e angular, resultados do p-valor em função da estatística T para o nível de significância $\alpha = 0,05$ com os seus respectivos valores críticos, para as variáveis MOS observados, versus MOS estimados.....	59
Tabela 2.8: Modelos utilizados para a estimação da CAD ideal na fase inicial do desenvolvimento da cultura da cana-de-açúcar	62
Tabela 2.9: Coeficientes métricos do modelo da umidade do solo para a CAD de 40 mm, N = 24.....	64
Tabela 2.10: Demonstrações dos coeficientes lineares e angulares, resultados do p-valor em função da estatística t para o nível de significância $\alpha = 0,05$ com os seus respectivos valores críticos, para a umidade do solo observada na análise temporal da emissão de CO ₂ (FCO ₂), que foi usada na determinação da umidade do solo obtida, a partir do balanço hídrico da CAD de 40 mm, N = 24	65
Tabela 2.11: E Estatística descritiva da umidade do solo observada na análise temporal da emissão de CO ₂ (FCO ₂) e da umidade do solo obtida a partir do balanço hídrico da CAD de 40 mm para todas as áreas estudadas, N = 24....	68
Tabela 2.12: Estatística descritiva da umidade do solo observada na análise temporal da emissão de CO ₂ (FCO ₂) e da umidade do solo obtida a partir do balanço hídrico da CAD de 40 mm, referentes às áreas de Motuca – SP (N = 3),	

Guariba – SP (N = 7), Pradópolis – SP (N = 8) e Aparecida do Taboado (N = 6)	69
Tabela 2.13: Coeficientes métricas para a determinação do modelo da umidade do solo para a CAD40 mm, N = 24	70
Tabela 2.14: Demonstrações do coeficiente linear e angular, resultados do p-valor em função da estatística t para o nível de significância $\alpha = 0,05$ com os seus respectivos valores críticos, para as variáveis umidade do solo observada a partir da CAD 40 mm, versus umidade do solo estimada a partir da CAD 40 m	71
Tabela 2.15: Coeficientes métricos do modelo da CAD, N = 24.....	73
Tabela 2.16: Demonstrações dos coeficientes lineares e angulares, resultados do p-valor em função da estatística t para o nível de significância $\alpha = 0,05$ com os seus respectivos valores críticos, para as variáveis Us, Ds, DP e ARM que foram usadas na determinação do modelo da CAD, N = 24	74
Tabela 2.17: Coeficientes métricas para a determinação do modelo da CAD, N = 24	80
Tabela 2.18: Demonstrações do coeficiente linear e angular, resultados do p-valor em função da estatística T para o nível de significância $\alpha = 0,05$ com os seus respectivos valores críticos, para as variáveis CAD observada, versus CAD estimada.....	81
Tabela 2.19: Coeficientes métricos do modelo da umidade do solo, N = 24 ...	83
Tabela 2.20: Demonstrações dos coeficientes lineares e angulares, resultados do p-valor em função da estatística t para o nível de significância $\alpha = 0,05$ com os seus respectivos valores críticos, para as variáveis ETP, Precipitação e ARM que foram usadas na determinação do modelo da Us, N = 24.....	84
Tabela 2.21: Coeficientes métricas para a determinação do modelo da umidade do solo, N = 24	91
Tabela 2.22: Demonstrações do coeficiente linear e angular, resultados do p-valor em função da estatística t para o nível de significância $\alpha = 0,05$ com os seus respectivos valores críticos, para as variáveis umidade do solo observadas, versus umidade do solo estimadas	92
Tabela 2.23: Coeficientes métricos do modelo da porosidade livre d'água do solo, N = 24, ...	94
Tabela 2.24: Demonstrações dos coeficientes lineares e angulares, resultados do p-valor em função da estatística t para o nível de significância $\alpha = 0,05$ com os seus respectivos valores críticos, para as variáveis PLA_BH e Dp que foram usadas na determinação do modelo da PLA, N = 24	95
Tabela 2.25: Coeficientes métricas para a determinação do modelo da PLA, N = 24	99

Tabela 2.26: Demonstrações do coeficiente linear e angular, resultados do p-valor em função da estatística t para o nível de significância $\alpha = 0,05$ com os seus respectivos valores críticos, para as variáveis PLA observada, versus PLA estimada.....	100
Tabela 2.27: Teste de normalidade pelo método de Shapiro-wilk das variáveis previstas MOS_prevista, Us_prevista, PLA_prevista, CAD40_prevista, FCO ₂ _prevista e FCO ₂ para todas as áreas estudadas	102
Tabela 2.28: Estatísticas descritivas da das variáveis previstas MOS_prevista, Us_prevista, PLA_prevista, CAD40_prevista, FCO ₂ _prevista e FCO ₂ para todas as áreas estudadas, N = 24	103
Tabela 2.29: Estatísticas descritivas das variáveis previstas MOS_prevista, Us_prevista, PLA_prevista, CAD40_prevista, FCO ₂ _prevista e FCO ₂ , referentes às áreas de Motuca – SP (N = 3), Guariba – SP (N = 7), Pradópolis – SP (N = 8) e Aparecida do Taboado (N = 6)	104
Tabela 2.30: Correlações de Pearson das variáveis previstas MOS_prevista, Us_prevista, PLA_prevista, CAD40_prevista, FCO ₂ _prevista e FCO ₂ para todas as áreas estudadas N = 24	105
Tabela 2.31: Coeficientes métricos do modelo da FCO ₂ , N = 24.....	108
Tabela 2.32: Demonstrações dos coeficientes lineares e angulares, resultados do p-valor em função da estatística t para o nível de significância $\alpha = 0,05$ com os seus respectivos valores críticos, para as variáveis MOS_prevista e PLA_prevista que foram usadas na determinação do modelo da FCO ₂ , N = 24	109
Tabela 2.33: Coeficientes métricas para a determinação do modelo da FCO ₂ , N = 24	113
Tabela 2.34: Demonstrações do coeficiente linear e angular, resultados do p-valor em função da estatística t para o nível de significância $\alpha = 0,05$ com os seus respectivos valores críticos, para as variáveis PLA observada, versus PLA estimada.....	114
Tabela 3.1 Hiperparâmetros ajustados para os algoritmos RFT_GS = random forest com a seleção de variáveis pelo método de grid Search, RFT = random forest regressor, TREE = decision tree regression, BOO = gradient boosting regressor, MLP = rede neural Multi-layer Perceptron regressor, KNN = regressão baseada no método k-nearest neighbors, SVM = método de regressão support vector machines, SVM_kpoli = support vector machines com o método de kernel polinomial, RLM = regressão linear múltipla, SVM_VC = método de regressão Support vector machines com validação cruzada, SGD = método de regressão da Stochastic Gradient Descent, disponíveis no pacote Scikit-learn (Pedregosa et al, 2011) da linguagem Python.....	136

Tabela 3.2 Teste de normalidade pelo método de Shapiro-wilk da emissão de CO ₂ , atributos atmosféricos e características físico-químicos do solo de todas as áreas estudadas.....	142
Tabela 3.3 Estatísticas descritivas da emissão de CO ₂ , atributos atmosféricos e características físico-químicos do solo de todas as áreas estudadas, N = 3019	145
Tabela 3.4 Índices estatísticos dos algoritmos utilizados para a previsão da FCO ₂ do solo, em função dos atributos atmosféricos e características físico-químicos do solo de todas as áreas estudadas, N = 3019	150

LISTA DE FIGURAS

Figura 1.1: Machine Learning, suas principais classes de algoritmos e atividades relacionadas. Fonte: Autor, adaptado de Mitchell (1997)	9
Figura 1.2: Modelo da multicamada baseado em perceptron (Ture, 2005)	11
Figura 2.1: Emissões antropogênicas totais de gases de efeito estufa (GEE) (gigatonelada de CO ₂ - equivalente por ano, GtCO ₂ -eq / ano) dos setores econômicos em 2010. Fonte: IPCC, 2014.....	16
Figura 2.2: Infográfico da cadeia produtiva da cana-de-açúcar do Brasil e dos estados de São Paulo e Mato Grosso do Sul, conforme os dados da safra 2019/2020 da Companhia Nacional de Abastecimento – CONAB. Fonte própria.	17
Figura 2.3: Mapas dos municípios das áreas experimentais do estado de São Paulo - SP (Pradópolis - PAD, Guariba - GUA e Motuca - MOT) e Mato Grosso do Sul - MS (Aparecida do Taboado - APT)	23
Figura 2.4: Fluxograma do cálculo do balanço Hídrico, conforme Thornthwaite e Mather modificado por Rolim et al. (2020).	27
Figura 2.5: Influência das variáveis do sistema solo-planta-atmosfera no processo da emissão do CO ₂ do solo (FCO ₂). Fonte: Própria	29
Figura 2.6: Modelo da análise empírica da influência dos atributos que compõem o sistema solo-planta-atmosfera, no processo da FCO ₂ . Fonte: Própria.....	30
Figura 2.7: GRID amostral das áreas estudados para A) Motuca – SP, B) Guariba – SP, C) Pradópolis – SP e D) Aparecida do Taboado – MS, com os seus respectivos pontos selecionados para a análise. Pontos amostrados dentro das áreas (azuis). Nos pontos selecionados utilizaram as variáveis para a construção do modelo	34
Figura 2.8: Sistematização das predições dos modelos da capacidade de água disponível do solo (CAD, 40 mm) e da emissão de CO ₂ do solo (FCO ₂).....	39
Figura 2.9: Histograma da emissão de CO ₂ , atributos atmosféricos e características físico-químicos do solo de todas as áreas estudadas	43

Figura 2.10: Emissão de CO ₂ , atributos atmosféricos e características físico-químicas do solo de todas as áreas estudadas, N = 24.....	47
Figura 2.11: Sistematização das etapas de modelagem da emissão de CO ₂ do solo, em função dos processos formados pela matéria orgânica e porosidade livre de água do solo	49
Figura 2.12: Boxplots da temperatura do ar (T _{air} , °C) e temperatura do solo (T _s , °C) para todas as áreas estudadas, N = 24	53
Figura 2.13: Boxplots da temperatura do ar (T _{air} , °C) e temperatura do solo (T _s , °C) para os municípios A) Motuca – SP, B) Guariba – SP, C) Pradópolis – SP e D) Aparecida do Taboado – MS.....	54
Figura 2.14: Boxplot da capacidade de troca de cátions (CTC, mmolc dm ⁻³) para todas as áreas estudadas, N = 24.....	55
Figura 2.15: Boxplot dos teores de fósforo disponível (P, mg dm ⁻³) para todas as áreas estudadas, N = 24.	56
Figura 2.16: Boxplot da densidade das partículas (D _p , kg dm ⁻³) para todas as áreas estudadas, N = 24	56
Figura 2.17: Boxplot da matéria orgânica do solo (MOS, g dm ⁻³) para todas as áreas estudadas, N = 24	57
Figura 2.18: Matéria Orgânica de solo (MOS) observado versus estimado pela equação de regressão.....	60
Figura 2.19: Determinações das curvas das regressões polinomiais quadráticas das capacidades de água disponível (CAD) para 100, 60, 45, 40, 35, 30 e 20 mm, em função da umidade do solo observada na análise temporal da emissão de CO ₂ (FCO ₂) e da umidade do solo obtida por meio do balanço hídrico pelo método de Thornthwaite e Mather (1955), na produção da cultura da cana-de-açúcar	62
Figura 2.20: Distribuição espacial (A) e o boxplot da umidade do solo observada (B) na análise temporal da emissão de CO ₂ (FCO ₂) e da umidade do solo obtida a partir do balanço hídrico da CAD de 40 mm, para todas as áreas estudadas, N = 24	67

Figura 2.21: Boxplots da umidade do solo observada na análise temporal da emissão de CO ₂ (FCO ₂) e da umidade do solo obtida a partir do balanço hídrico da CAD de 40 mm para os municípios A) Motuca – SP, B) Guariba – SP, C) Pradópolis – SP e D) Aparecida do Taboado – MS	68
Figura 2.22: Umidade do solo observada a partir da CAD 40 mm versus umidade do solo estimada a partir da CAD 40 mm pela equação de regressão	72
Figura 2.23: Boxplot da umidade do solo (Us, mm) para todas as áreas estudadas, N = 24	75
Figura 2.24: Boxplots das umidades do solo (Us, mm) para os municípios A) Motuca – SP, B) Guariba – SP, C) Pradópolis – SP e D) Aparecida do Taboado – MS.....	76
Figura 2.25: Boxplot da densidade do solo (Ds, kg dm ⁻³) para todas as áreas estudadas, N = 24	77
Figura 2.26: Boxplot da densidade das partículas (Dp, kg dm ⁻³) para todas as áreas estudadas, N = 24	77
Figura 2.27: Boxplot do armazenamento de água no solo (ARM, mm) para todas as áreas estudadas, N = 24	78
Figura 2.28: Boxplots dos armazenamentos de água no solo (ARM, mm) para os municípios de A) Motuca – SP, B) Guariba – SP, C) Pradópolis – SP e D) Aparecida do Taboado – MS.....	79
Figura 2.29: Capacidade de água disponível no solo (CAD) observada versus CAD estimada pela equação de regressão	81
Figura 2.30: Boxplot da evapotranspiração potencial (ETP, mm) para todas as áreas estudadas, N = 24	85
Figura 2.31: Boxplots da evapotranspiração potencial (ETP, mm) para os municípios A) Motuca – SP, B) Guariba – SP, C) Pradópolis – SP e D) Aparecida do Taboado – MS.....	85
Figura 2.32: Boxplot da precipitação (Precipitação, mm) para todas as áreas estudadas, N = 24	86

Figura 2.33: Boxplots das precipitações (Precipitação, mm) para os municípios de A) Motuca – SP, B) Guariba – SP, C) Pradópolis – SP e D) Aparecida do Taboado – MS.....	87
Figura 2.34: Boxplot do armazenamento de água no solo (ARM, mm) para todas as áreas estudadas, N = 24	88
Figura 2.35: Boxplots dos armazenamentos de água no solo (ARM, mm) para os municípios de A) Motuca – SP, B) Guariba – SP, C) Pradópolis – SP e D) Aparecida do Taboado – MS.....	88
Figura 2.36: Boxplot da umidade do solo (Us, mm) para todas as áreas estudadas, N = 24	89
Figura 2.37: Boxplots das umidades do solo (Us, mm) para os municípios A) Motuca – SP, B) Guariba – SP, C) Pradópolis – SP e D) Aparecida do Taboado – MS.....	90
Figura 2.38: Umidade do solo (Us) observado versus US estimada pela equação de regressão	92
Figura 2.39: Boxplot da porosidade livre de água decorrente do balanço hídrico (PLA_BH, %) para todas as áreas estudadas, N = 24	96
Figura 2.40: Boxplots da porosidade livre de água decorrente do balanço hídrico (PLA_BH, %) para os municípios A) Motuca – SP, B) Guariba – SP, C) Pradópolis – SP e D) Aparecida do Taboado – MS	96
Figura 2.41: Boxplot da densidade de partículas (Dp, kg dm ⁻³) para todas as áreas estudadas, N = 24	97
Figura 2.42: Boxplot da porosidade livre de água do solo (PLA, %) para todas as áreas estudadas, N = 24	98
Figura 2.43: Boxplots da porosidade livre de água do solo (PLA, %) para os municípios de A) Motuca – SP, B) Guariba – SP, C) Pradópolis – SP e D) Aparecida do Taboado – MS.....	98
Figura 2.44: Porosidade livre d'água do solo observado (PLA) versus PLA estimado pela equação de regressão	101

Figura 2.45: Histograma das variáveis previstas MOS_prevista, Us_prevista, PLA_prevista, CAD40_prevista, FCO ₂ _prevista e FCO ₂ para todas as áreas estudadas.....	102
Figura 2.46: Correlações de Pearson das variáveis previstas. FCO ₂ = Fator de emissão de CO ₂ (μmol CO ₂ m ⁻² s ⁻¹); MOS_prevista = matéria orgânica do solo prevista (g dm ⁻³); Us_prevista = umidade do solo prevista (% do volume); PLA_prevista = porosidade livre de água prevista (%); CAD = capacidade de água disponível prevista (mm); FCO ₂ _prevista = Fator de emissão de CO ₂ (μmol CO ₂ m ⁻² s ⁻¹), para todas as áreas estudadas, N = 24.....	106
Figura 2.47: Boxplot da porosidade livre de água prevista (PLA_prevista, %) para todas as áreas estudadas, N = 24.....	110
Figura 2.48: Boxplots da porosidade livre de água prevista (PLA_prevista, %) para os municípios A) Motuca – SP, B) Guariba – SP, C) Pradópolis – SP e D) Aparecida do Taboado – MS.....	110
Figura 2.49: Boxplot da matéria orgânica do solo prevista (MOS_prevista, g dm ⁻³) para todas as áreas estudadas, N = 24	111
Figura 2.50: Boxplot da emissão do CO ₂ do solo (FCO ₂ , μmol CO ₂ m ⁻² s ⁻¹) para todas as áreas estudadas, N = 24.....	112
Figura 2.51: Boxplots da emissão do CO ₂ do solo (FCO ₂ , μmol CO ₂ m ⁻² s ⁻¹) para os municípios de A) Motuca – SP, B) Guariba – SP, C) Pradópolis – SP e D) Aparecida do Taboado – MS.....	112
Figura 2.52: Emissão de CO ₂ do solo (FCO ₂) observado versus FCO ₂ estimado pela equação de regressão.....	115
Figura 2.53: Boxplot da emissão do CO ₂ do solo prevista (FCO ₂ _prevista, μmol CO ₂ m ⁻² s ⁻¹) para todas as áreas estudadas, N = 24	115
Figura 2.54: Boxplots da emissão do CO ₂ do solo (FCO ₂ , μmol CO ₂ m ⁻² s ⁻¹) para os municípios de A) Motuca – SP, B) Guariba – SP, C) Pradópolis – SP e D) Aparecida do Taboado – MS.....	116

Figura 3.1: Mapas dos municípios das áreas experimentais do Estado de São Paulo - SP (Pradópolis - PAD, Guariba - GUA e Motuca - MOT) e Mato Grosso do Sul - MS (Aparecida do Taboado - APT).....	124
Figura 3.2: Fluxograma do cálculo do balanço Hídrico, conforme Thornthwaite e Mather, modificado por Rolim et al. (2020)	128
Figura 3.3: Representação da análise empírica da influência dos atributos que compõem o sistema solo-planta-atmosfera, com relação ao processo de FCO ₂ . Fonte: Própria.....	130
Figura 3.4: Resumo do processo de desenvolvimento dos modelos para prever a emissão de CO ₂ do solo, por meio dos algoritmos de aprendizado de máquina	133
Figura 3.5: Representações dos oito algoritmos de aprendizado não supervisionado utilizados para a prever a emissão de CO ₂ do solo (FCO ₂).135	
Figura 3.6: Histograma da emissão de CO ₂ , atributos atmosféricos e características físico-químicos do solo de todas as áreas estudadas, N = 3019	144
Figura 3.7: Boxplot da emissão do CO ₂ do solo (FCO ₂ , $\mu\text{mol CO}_2 \text{ m}^{-2} \text{ s}^{-1}$) para todas as áreas estudadas, N = 3019.....	147
Figura 3.8: Análise das Correlações pelo método de Pearson das variáveis da Emissão de CO ₂ , atributos atmosféricos e características físico-químicos do solo de todas as áreas estudadas, N = 3019.....	148
Figura 3.9: Esquematização dos fluxos dos dados na execução dos algoritmos de aprendizado supervisionado.....	149
Figura 3.10: Visão geral da distribuição de métricas estatísticas em um gráfico de radar escalonado para emissão de CO ₂ do solo, em função dos atributos atmosféricos e características físico-químicos do solo de todas as áreas estudadas, na fase de calibração dos modelos, n = 2415.....	153
Figura 3.11: Gráficos de dispersão da emissão da CO ₂ do solo observado versus predita dos modelos estudados nos conjuntos de dados de calibração (n = 2415). A) RFT_GS = random forest com a seleção de variáveis pelo método de grid Search; B) RFT = random forest regressor; C) TREE = decision tree regression;	

D) BOO = gradient boosting regressor; E) MLP = rede neural Multi-layer Perceptron regressor; F) KNN = regressão baseada no método k-nearest neighbors; G) SVM = método de regressão support vector machines; H) SVM_kpoli = support vector machines com o método de kernel polinomial; I) RLM = regressão linear múltipla; J) SVM_VC = método de regressão Support vector machines com validação cruzada; K) SGD = método de regressão da Stochastic Gradient Descent. 155

Figura 3.12: Uma visão geral da distribuição de métricas estatísticas em um gráfico de radar escalonado para emissão de CO₂ do solo, em função dos atributos atmosféricos e características físico-químicos do solo de todas as áreas estudadas, na fase de validação dos modelos, n = 604..... 158

Figura 3.13: Gráficos de dispersão de CO₂ do solo observado versus predita dos modelos estudados nos conjuntos de dados de teste (n = 604) 160

Figura 3.14: Análise de resíduos dos dados observados FCO₂ versus os dados estimados de FCO₂, para todos os algoritmos preditivos estudados (n = 3019).. 161

Figura 3.15: Dados observados FCO₂ versus os dados estimados de FCO₂, para todos os algoritmos preditivos estudados (n = 604) 163

1. CAPÍTULO 1 – CONSIDERAÇÕES GERAIS

1.1. INTRODUÇÃO

1.1.1. Gases do Efeito Estufa

O efeito estufa é um processo natural que acontece devido à incidência da radiação solar na superfície do globo terrestre, que logo é refletida e absorvida pela atmosfera e redirecionada para a Terra, mantendo as condições adequadas de temperatura para a sobrevivência dos seres vivos.

Os principais gases causadores do efeito estufa (GEE) são: dióxido de carbono (CO_2), metano (CH_4), óxido nitroso (N_2O), clorofluorcarbonetos (CFCs), hidroclorofluorcarbonetos (HCFCs), perfluorcarbonetos (PFCs), hexafluoreto de enxofre (SF_6) e o vapor d'água (Forster et al., 2007). Os gases que causam maior impacto para o efeito estufa são CO_2 , CH_4 e N_2O , que representam cerca de 50, 15,5 e 5% do forçamento radioativo provocador do efeito estufa, respectivamente (IPCC, 2007).

O Primeiro Relatório de Avaliação Nacional do Painel Brasileiro de Mudanças Climáticas (PBMC, 2013) mostrou que a temperatura do ar próximo à superfície terrestre, até o final do século XXI, terá aumento sistemático com variação aproximada de 2 °C (no cenário de baixa emissão) a 4 °C (no cenário de alta emissão). Usando modelos biogeográficos ou modelos de biomas, os pesquisadores estimam grandes variações devido ao aquecimento do planeta, tendo como expectativas futuras nos biomas brasileiros o aumento de suas temperaturas, a alteração do regime de chuvas e, conseqüentemente, as mudanças no processo produtivo da agricultura (Pivetta, 2013).

Segundo previsões realizadas pelo IPCC (2007), no período de 1980 a 1999, as mudanças de temperatura provocadas pelo efeito estufa mostram que o aumento da concentração dos gases indutores desse processo na atmosfera poderá registrar, no final do século 21, o aumento de até 7,5°C na temperatura em alguns locais do planeta Terra.

A emissão de GEE é favorecida por diversas atividades antrópicas. Entre 1970 e 2004, o principal setor responsável por esse processo foi o de geração de energia, com 49%, seguido do industrial (19,4%), da degradação florestal e do desmatamento (17,4%), da construção civil (7,9%) e do tratamento de resíduos (2,8%). A agricultura

colaborou nesse período com 13,5% da emissão global de GEE, sendo a quarta atividade mais impactante (IPCC, 2007).

No Brasil, a emissão de GEE é diferente da realidade mundial, onde o setor energético tem baixa participação nesse processo (MCT, 2004; Cerri et al., 2009). Segundo o Sistema de Estimativas de Emissões e Remoções de Gases de Efeito Estufa (SEEG), no ano de 2016 as mudanças do uso do solo e floresta (54%) e a agropecuária (23%) totalizavam 77%, sendo os principais responsáveis pela emissão de GEE no país, enquanto o setor energético era responsável por 16% da emissão brasileira de GEE (SEEG, 2016).

No Brasil, os setores das mudanças do uso da terra e de floresta, e a agropecuária são os principais responsáveis pelo processo de emissão dos GEEs. O uso das práticas agronômicas de manejo que visam promover o processo de sequestro do carbono é muito importante, pois pode fazer dos solos agrícolas uma fonte natural de armazenamento dos GEEs. A cana-de-açúcar, por meio da cobertura formada pela palhada, proporciona uma condição que favorece o acúmulo de carbono no solo (Galdos et al., 2009), colaborando assim com a mitigação dos GEEs, envolvida na produção agrícola do etanol e do açúcar. Estudos recentes com diferentes tipos de preparo do solo (convencional, subsolagem localizada e subsolagem convencional) indicaram distintas perdas de carbono, via emissão de CO_2 , mostrando maiores emissões no preparo convencional ($2.864,3 \text{ kg de CO}_2 \text{ ha}^{-1}$), seguidos das subsolagens: convencional ($1970,9 \text{ kg CO}_2 \text{ ha}^{-1}$) e localizada ($1707,7 \text{ kg CO}_2 \text{ ha}^{-1}$) (Iamaguti et al., 2015).

Contudo, para que as medidas mitigadoras sejam efetivas, é de suma importância o conhecimento e a aplicação das práticas agrícolas conservacionistas, melhorando assim a produção e diminuindo os impactos gerados pelo setor sucroenergético em relação ao balanço das emissões dos GEEs gerados pelo setor agrícola.

1.1.2. Aspectos Gerais sobre o Fluxo de CO_2

Na Terra, pode-se encontrar o carbono em cinco compartimentos: atmosférico, oceânico, pedológico, biótico e geológico. O maior reservatório de carbono do Planeta é o geológico, com aproximadamente 90000000 de petagramas de carbono (Pg),

porém apenas 0,005% é usado na ciclagem do carbono. Esta fração efetiva está na forma de carvão, óleo e gás (Lal, 2004). O compartimento oceânico apresenta 38.000 Pg de carbono, seguido do pedológico com 2500 Pg. Os menores reservatórios estão presentes no compartimento biótico e atmosférico, com 560 e 760 Pg de carbono, respectivamente. No compartimento pedológico, o banco de carbono dos primeiros 30 cm de solo é de 800 Pg (Cerri et al., 2006), mesma quantidade estocada no compartimento atmosférico. Segundo Eswaran et al. (1993), em solos com até 1 metro de profundidade, estima-se o armazenamento de 1576 Pg de carbono. Manejos conservacionistas que buscam a mínima movimentação do solo favorecem a permanência deste carbono estocado, impedindo sua emissão para a atmosfera (Lal, 2004).

De acordo com Denman et al. (2007), as indústrias, o uso de combustíveis fósseis e as mudanças do uso da terra são as principais fontes de emissão de CO₂ da atmosfera. Além das atividades ligadas ao homem, o carbono do solo pode ser emitido para a atmosfera por meio da decomposição de resíduos orgânicos e também pelas respirações dos seres vivos e das raízes das plantas (Carvalho et al., 2010). A respiração radicular totaliza 20% do aumento da concentração de CO₂ no solo (quando comparado com a atmosfera) e 80% são provenientes da atividade biológica do solo. Por causa disso, vê-se que a concentração de CO₂ no solo é maior do que na atmosfera (Melillo et al., 2002).

O transporte da molécula de CO₂ no solo acontece primeiramente devido ao processo de difusão; assim o gás flui de uma região de maior concentração para uma de menor (Ball e Smith, 1991), onde a umidade e a temperatura do solo são os fatores que mais interferem no fluxo de CO₂ (Duiker e Lal, 2000).

1.1.3. Emissão de gases causadores do efeito estufa nos sistemas Agrícolas

Na agricultura, as práticas agrícolas ligadas às alterações do manejo, do uso do solo e da queima dos combustíveis fósseis estão intimamente relacionadas com as emissões dos GEEs. Muitas vezes, essas atividades geram impactos nocivos ao meio ambiente, como as emissões de metano na agropecuária, do óxido nitroso em solos fertilizados e o desmatamento. Estima-se que aproximadamente 1/4 dos GEEs

sejam oriundos da agricultura praticada em todo o planeta (Lamb et al., 2016). Por isso, a importância do uso adequado de técnicas de produção, visando manejar da melhor forma possível as áreas agrícolas e fornecendo assim, as menores quantidades de GEE para a atmosfera. Entre os anos de 1995 e 2010, a agricultura mundial contribuiu com cerca de 24% das emissões de CO₂, 85% de N₂O e 55% de CH₄, conforme o quarto relatório do IPCC (2007). As médias anuais da emissão CO₂ tiveram taxa de crescimento de 1,94 ppm quando comparado com o período de 1960 a 2005, com média em torno de 1,46 ppm (ESRL, 2016).

Além disso, as condições climáticas também impactam na intensidade da emissão de CO₂ do solo, interferindo diretamente no processo de respiração das raízes, na atividade microbiana e, conseqüentemente, na produção de CO₂ nos solos agrícolas (Tsai et al., 1992; Schlömer et al., 2013; Xie et al., 2015). Outras propriedades do solo também interferem no processo de produção de CO₂, sendo elas: a densidade do solo, os atributos do solo (Teixeira et al., 2012), a textura e a porosidade livre de água (Panosso et al., 2011), a temperatura do solo (Ussiri e Lal, 2009), o preenchimento de água nos poros (Smart e Peñuelas, 2005; Linn e Doran, 1984) e a umidade do solo (Moitinho et al., 2013).

O Brasil é um dos principais produtores de alimentos, apresentando um grande papel na segurança alimentar mundial. Além disso, também é o maior produtor mundial da cultura da cana-de-açúcar. Estima-se que a área plantada na safra 2021/2022 será em torno de 8,26 milhões hectares e a produção 568,4 milhões de toneladas. O estado de São Paulo será o maior produtor do país, com 4,20 milhões de hectares e uma produção de 287,44 milhões de toneladas, compreendendo 50,57 % da cana processada. Já o estado do Mato Grosso do Sul é o quarto maior produtor do país, com 653,7 mil de hectares e produção de 45,42 milhões de toneladas, onde será responsável por 7,99% da produção brasileira de acordo com estimativas (CONAB, 2022).

As pesquisas indicam que as práticas usadas na reforma do canavial, como a calagem para a correção do pH do solo (Marcelo et al., 2012; Silva et al., 2013; Figueiredo et al., 2014) e o preparo intensivo (Silva-Olaya et al., 2013; Iamaguti et al., 2015) promovem perdas expressivas do carbono presente no solo, na forma de CO₂, principalmente em períodos curtos; assim, como consequência, observam-se

alterações dos atributos químicos, físicos e biológicos do solo, além das variações de umidade e de temperatura do solo (Moitinho et al., 2013).

As interações da umidade e da temperatura do solo influenciam nas variações temporais da emissão de CO₂ do solo, principalmente em estações do ano em que os teores de água no solo são baixos (Panosso et al., 2011). Schwendenmann et al. (2003), estudaram a variação espaçotemporal do fluxo de CO₂ em áreas florestais e obtiveram que o fator principal que explicou as variações temporais do fluxo de CO₂ foi a quantidade de água no solo. Os autores concluem que o fluxo, a difusão e a taxa de produção de CO₂ diminuíram quando o solo apresentava elevada umidade. Dessa forma, observa-se que o processo de emissão do gás CO₂ do solo para a atmosfera é complexo e depende das condições edafoclimáticas locais e da grande variabilidade no tempo e no espaço.

1.1.4. Impactos mundiais ligados à agricultura e aos Gases do Efeito Estufa

O clima está ligado diretamente ao desenvolvimento da agricultura, e uma pequena alteração pode gerar grande mudança no processo produtivo. No mundo, estima-se que o aquecimento global e o aumento de GEE podem gerar impactos positivos ou negativos na produção agrícola, apesar dos efeitos negativos serem mais evidentes, conforme o estudo do grupo II do trabalho do IPCC. Os autores relatam que os efeitos positivos são: a extensão da produção agrícola devido à diminuição dos impactos do congelamento em regiões limitadas pela temperatura e também aos benefícios da fertilização de CO₂ da matéria orgânica sobre as culturas. Já os efeitos negativos são aqueles relacionados com o desequilíbrio do regime de precipitação, com a baixa umidade do solo e com os extremos de calor (Porter et al., 2014).

Segundo Tilman et al. (2011), a demanda de alimentos no mundo deverá aumentar 100% até 2050, aumentando também a demanda por recursos naturais e por novas áreas agrícolas. Vermeulen et al. (2012) concluem que a probabilidade de inundações devido à elevação da temperatura, vai dificultar a capacidade dos agricultores no processo de distribuição e de armazenagem segura dos alimentos, podendo trazer problemas e os riscos de doenças alimentares. Segundo os autores, a população de baixa renda e os pequenos produtores estão mais sujeitos às

alterações climáticas. Devido à sua limitada capacidade de investimento em equipamentos e nas tecnologias adaptadas para os riscos climáticos crescentes.

Thornton (2012) mostrou que a produção das mais comuns commodities, como o trigo, o milho e o arroz, devido ao aquecimento global e às alterações do clima, deverão passar por mudanças no processo produtivo, principalmente no controle fitossanitário, filotécnico e nutricional.

No final do ano de 2009, foi instituída a Política Nacional sobre Mudança do Clima – PNMC, que estabeleceu o “Plano para a consolidação de uma economia de Baixa Emissão de Carbono na Agricultura”. O Plano da Agricultura de Baixo Carbono (ABC), publicado no ano de 2015, é formado por sete programas ligados às tecnologias de mitigação e às ações de adaptação às mudanças climáticas. Esse plano abrange todo o território brasileiro e espera-se que até 2020 sejam cumpridos os compromissos firmados na COP-15, compostos pelas seguintes ações:

- Recuperar 15 milhões de hectares de pastagens degradadas, realizando o manejo adequado e a adubação;
- Ampliar o uso dos sistemas de Integração Lavoura-Pecuária-Floresta (ILPF) e de Sistemas Agroflorestais (SAFs) em 4 milhões de hectares;
- Aumentar a utilização do Sistema Plantio Direto (SPD) em 8 milhões de hectares;
- Expandir o uso da Fixação Biológica de Nitrogênio (FBN) em 5,5 milhões de hectares;
- Incentivar as ações de reflorestamento no Brasil, aumentando as áreas com Florestas Plantadas, as produções de fibras, de madeira e de celulose em 3,0 milhões de hectares;
- Usar tecnologias para tratar 4,4 milhões de m³ de dejetos de animais para a geração de energia e produção de composto orgânico (ABC, 2015).

O Brasil é um grande celeiro mundial, com grande relevância nas exportações de alimentos e na segurança alimentar do planeta. Porém, ainda são utilizadas práticas agrícolas no processo de produção inadequadas, que geram GEE e que colaboram com aproximadamente 70% da emissão de CO₂ equivalente produzido no país (Cerri et al., 2013), ranqueando o Brasil como o 4º maior emissor mundial (IPCC, 2014).

Visando diminuir esses impactos, diversas práticas agrícolas são sugeridas para diminuir a concentração de GEE na atmosfera (La Scala et al., 2006; Figueiredo et al., 2010; Panosso et al., 2011; Bicalho et al., 2014; Iamaguti et al., 2015), como o uso do sistema de plantio direto, a redução de queimadas, o controle e a redução de desmatamento e a substituição dos combustíveis fósseis por biocombustíveis, potencializando o acúmulo de carbono no solo, que também é conhecido como sequestro de carbono. Segundo Smith et al. (2008), o potencial de mitigação de GEE proveniente do setor agrícola pode atingir valores de 5500-6000 Mt CO₂ - eq. ano⁻¹ até o ano de 2030.

Globalmente, os solos com até 1 m de profundidade podem conter aproximadamente até 1500 GtC (gigatonelada de carbono) (Bruce et al., 1999; Machado, 2005). São estimados que o incremento de 10% no acúmulo de carbono no solo compensaria 30 anos de emissões de CO₂ de origem antropogênica (Kirschbaum, 2000). Dessa forma, o entendimento do ciclo do carbono nos ecossistemas agrícolas é imprescindível e de suma importância para contribuir com as reduções das emissões de GEE nas áreas agrícolas.

1.1.5. Uso da modelagem para a emissão de CO₂ do solo

Desde a década de 1960, métodos matemáticos e estatísticos estão sendo utilizados em pesquisas na ciência do solo, visando proporcionar a melhor compreensão das dinâmicas do solo e gerando ferramentas de avaliação que ajudam na resolução dos problemas agrícolas e ambientais (McBratney e De Gruijter, 1992).

Devido a não linearidade das relações encontradas entre as variáveis estudadas e à emissão dos GEE, observa-se uma alta complexidade, exigindo o uso de técnicas matemáticas e estatísticas que consigam extrair estas informações de maneira clara e robusta. O aumento do uso das técnicas de machine learning dentro da agricultura tem gerado grandes resultados, que por sua vez, colaboraram com o entendimento da variabilidade espaçotemporal do CO₂, por meio dos modelos lineares generalizados (MLG) e modelos de machine learning como Regressão Linear Múltipla (RLM), K Neighbors Regressor (KNN), Regressor Florestal Aleatório (RFT), Redes neurais artificiais – Multi-layer Perceptron (MLP), Regressor arvores de decisão

(TREE), Máquina de vetores de suporte (SVM), Gradient Boosting para regressão (BOO) e Descida do Gradiente Estocástico (SGD).

Analizando as pesquisas de outros autores, observam-se estudos que usam as técnicas de aprendizado de máquina para a previsão da emissão do CO₂ do solo, como por exemplo: Predição do fluxo de CO₂ no solo em sistemas de manejo de cana-de-açúcar utilizando a abordagem Random Forest (Tavares et al., 2018), Uso de técnicas de mineração de dados para classificar a emissão de CO₂ do solo induzida pelo manejo de culturas em canaviais (Farhate et al., 2018). No entanto, faltam estudos sobre a previsão da CO₂ utilizando variáveis oriundas do solo, planta e atmosfera que usam as técnicas de aprendizado de máquina.

1.1.6. Algoritmos de aprendizado de máquina

O aprendizado de máquina (ML) é um fragmento da inteligência artificial (IA) utilizado para prever processos de classificação, previsão, regressão e decisão, por meio de diversos modelos matemáticos (Alpaydın, 2020). O ML possui grande poder de precisão devido às variáveis usadas e alto nível de sucesso treinamento (Mitchell, 2006). Basicamente, temos uma primeira fase na qual o algoritmo passa por um aprendizado, por meio de dados da amostra denominados de dados de treinamento. Logo depois, devido a essas experiências são previstas as respostas desconhecidas.

Analizando as aplicações das diferentes técnicas de ML, observam-se os usos dos algoritmos em diversas áreas, como por exemplo na física, matemática, saúde, militar, engenharia e economia, imitando neurônios no cérebro, obtendo resultados de grande importância e impactando na vida dos seres humanos e do planeta como um todo (Ağbulut et al., 2021; Senturk 2020; Gürel et al., 2020; Başarslan e Argun, 2019; Salam e Verma, 2020; Karstoft et al., 2015).

Analizando as categorias dos modelos de predição, existem três principais classes de algoritmos que são comumente utilizados nos estudos relacionados de ML, conforme é descrito abaixo:

- **Aprendizado supervisionado:** o algoritmo utiliza uma quantidade de dados pré-determinada de entrada, identificados com suas respectivas saídas, para treinar e gerar os padrões dos dados, e com base nisto, determina as regras para classificar as novas entradas;

- **Aprendizado não supervisionado:** os algoritmos usam os dados de entrada não rotulados com o intuito de determinar a melhor identificação, em função da estrutura do algoritmo, evidenciando agrupamentos determinantes ou características especialmente úteis para a sua caracterização;
- **Aprendizado por reforço:** os algoritmos são treinados a partir da interação repetida com o ambiente, resultando em respostas positivas ou negativas para o algoritmo, direcionando-o nas decisões ou nos próximos passos.

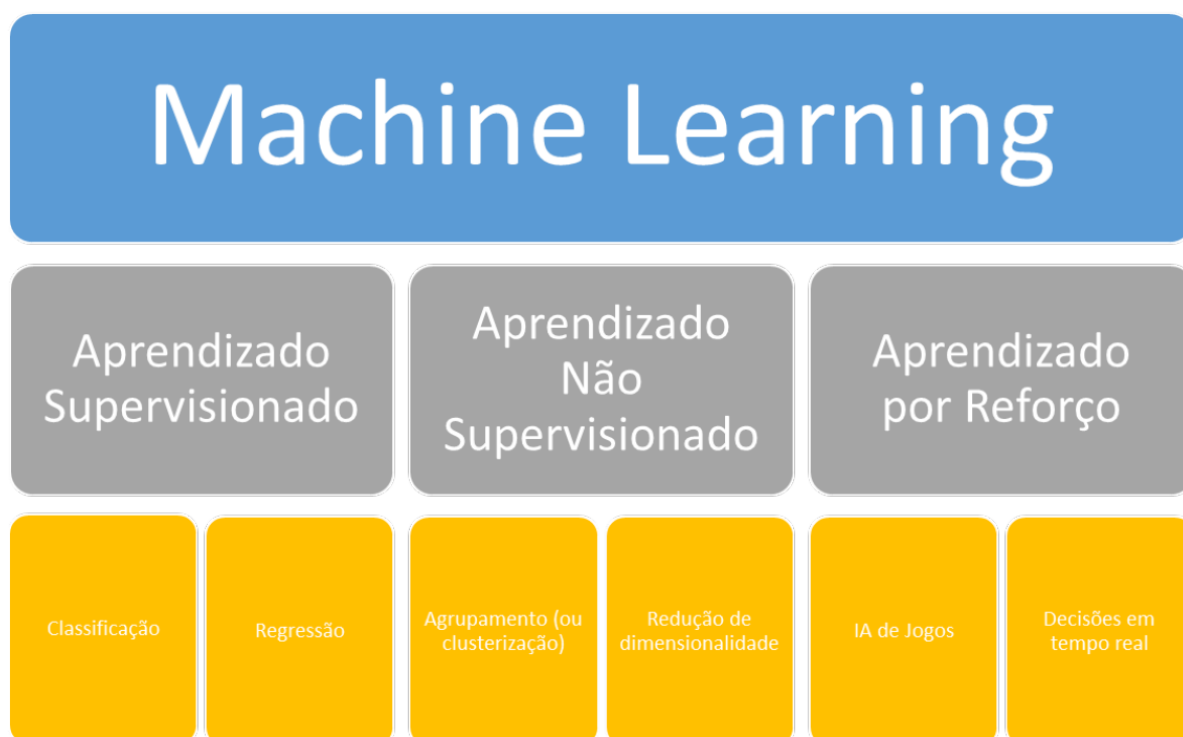


Figura 1.1: Machine Learning, suas principais classes de algoritmos e atividades relacionadas. Fonte: Autor, adaptado de Mitchell (1997).

Os algoritmos que resolvem problemas por meio da regressão fazem parte da categoria de aprendizado supervisionado. A seguir será realizada uma explanação dos principais modelos utilizados para estimar a FCO₂.

No modelo de regressão linear múltipla (RLM) tem-se uma quantidade das variações da variável dependente (Y), que é explicada pelo conjunto das variáveis independentes (X_i , $i = 1, 2, \dots, n$), o fragmento que não pode ser explicado por essas variações é caracterizado pelo erro aleatório (ϵ) (Santana, 2003). O RLM é um método que já foi usado em estudos com o CO₂ (Morita et al., 2019; Lemordant et al., 2018,

Yang et al., 2019). Foi utilizado o método Ridge Regression que suaviza atributos que sejam correlacionados e que aumentam os erros no modelo (multicolinearidade). A retirada de determinados atributos do modelo resulta em uma previsão mais estável, sem alterar acurácia do modelo. Esse processo gera um mecanismo que penaliza o seu viés e que vai diminuindo os valores dos betas até zero. As variáveis que contribuem pouco com o poder de predição do modelo são consideradas irrelevantes, por meio desse mecanismo de penalização do viés.

O algoritmo K-Vizinhos mais próximo (KNN) é um método não paramétrico utilizado para estimar a emissão de CO₂ (Gupta et al., 2020; Martinsen et al., 2020; Rubia e Nandhini, 2021). KNN é uma metodologia bastante flexível, simples e de fácil aplicação. Esta técnica seleciona um grupo de k amostras (dados de treinamento) mais próxima de amostras desconhecidas (dados de teste). As determinações da média da variável resposta resultam nas amostras desconhecidas para as k amostras. O parâmetro k é encontrado pelo do critério do cotovelo.

A regressão de floresta aleatória (RF) é um algoritmo de árvore de decisão amplamente utilizado, em que diversas árvores de decisão são criadas, a partir de um conjunto de dados de entrada determinado. O conjunto de dados é dividido ao acaso pelo algoritmo em várias subpartes e para cada subparte são construídas diversas árvores de decisão. Logo após, o algoritmo mistura a saída prevista de cada árvore de decisão, para conseguir uma estimativa mais acurada. Na regressão RF, o valor de saída de um subconjunto ou qualquer outra entrada é uma média dos valores estimados por várias árvores de decisão. Outros autores também utilizaram o algoritmo RF para a previsão de CO₂ (Payen et al., 2021; Nguemezi et al., 2021; Terrer et al., 2021).

Outro método de rede neural artificial (RNA) utilizado foi o Multi-layer Perceptron (MLP). A estrutura de uma rede neural MLP contém três tipos de camadas: camada de entrada, camada oculta e a camada de saída. O MLP realiza um mapeamento não linear entre um vetor de entrada e um vetor de saída, que é formado por um conjunto de neurônios interconectados em sua estrutura (Pal e Mitra, 1992). Durante o processo de treinamento, todos os dados de entrada passam pela rede de neurônios de forma feed-forward, conforme pode ser observado na Figura 1.2. Outros

autores também utilizaram essa técnica de RNA pelo método MLP para prever a emissão de CO₂ (Cheshmberah et al., 2020; Siabi et al., 2019; Kolasa-Więcek, 2018).

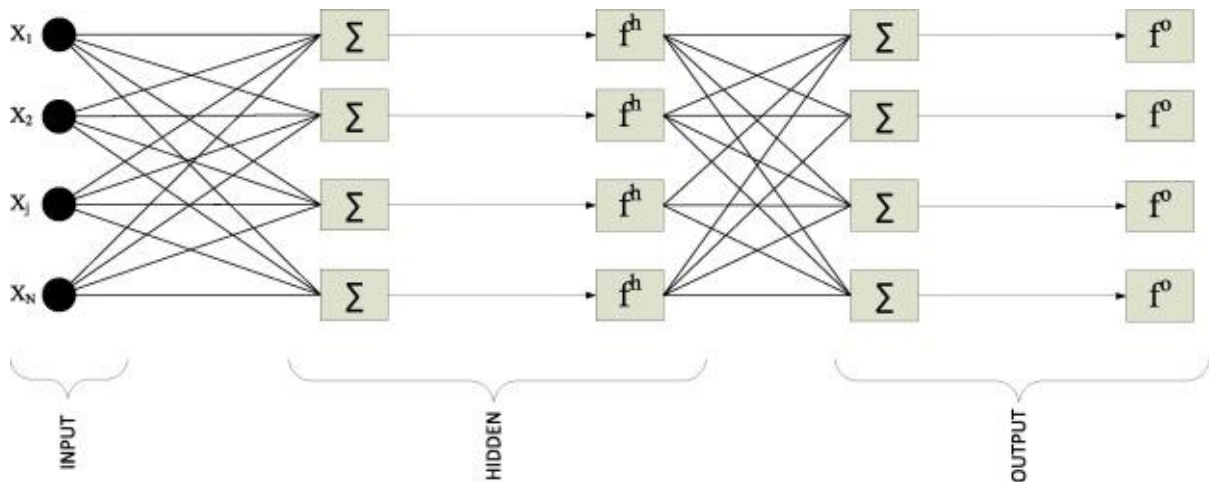


Figura 1.2: Modelo da multicamada baseado em perceptron (Ture, 2005).

O TREE é uma técnica de aprendizado de máquina supervisionado não paramétrico, bastante utilizado em processos de classificação e regressão. Na literatura são encontrados trabalhos que usaram o algoritmo TREE para estudar o comportamento do CO₂ no solo (Hamrani et al., 2020; Warner et al., 2019 a; Tavares et al., 2018; Ransom et al., 2019). O TREE realiza a classificação e previsão de variáveis em números finitos de classes, através da sua divisão em grupos e das regras de hierarquias, resultando em uma visão real da natureza do processo (Quinlan, 1986).

O algoritmo de máquina de vetores de suporte (SVM) é uma técnica que pode ser aplicado em problemas de regressão e classificação. É um algoritmo bastante novo, onde o seu primeiro registro foi no ano de 1992 (Boser et al., 1992). O SVM trabalha conforme a teoria da aprendizagem estatística e princípios de minimização do risco estrutural. Assim, ele mostra uma tentativa de recuperar o limite superior do erro na generalização, em vez de retirar o erro no estágio de treinamento local. Essa é a característica mais óbvia, que distingue o algoritmo SVM de outros algoritmos de máquina baseados em aprendizado (Chen et al., 2013). Alguns autores também utilizaram essa técnica de SVM em estudos relacionados com a emissão de CO₂ (Yan et al., 2020; Li et al., 2017; Deiss et al., 2017; Ehteram et al., 2021).

Os primeiros registros das máquinas impulsionadas por gradiente (BOO) foram feitos por Friedman (Friedman, 2001; Friedman, 2002), é um modelo de aprendizado

de máquina impulsador, que adere sequências de aprendizes 'fracos' para elaborar um aprendiz fortemente preciso ao acaso (Han et al., 2011). BOO são modelos de aprendizagem fundamentados em gradientes aditivos, onde um novo aprendiz fraco é adicionado e treinado para diminuir o erro geral de todo o algoritmo, sem alterar os aprendizes fracos presentes no modelo. GBMs usam as árvores de regressão como aprendizes fracos para fazer as interações, o algoritmo faz esse processo de maneira sequencial, onde a nova árvore de aprendizado gerada tem o intuito de diminuir o erro geral de treinamento. Este procedimento é repetido até que seja construído um modelo com a acurácia desejada e o erro global atinja o nível esperado (Breiman et al., 1984). Outros estudos foram encontrados nos quais utilizaram os BOO, em pesquisas ligadas à emissão de CO₂ e solo (Cai et al., 2020; Wen et al., 2021; Li et al., 2022).

O SGD utiliza os gradientes aproximados previstos por meio dos subconjuntos das observações de treinamento e atualiza os pesos desses recursos. Essa estrutura de aprendizado está chamando a atenção porque requer muito menos tempo de treinamento, quando comparado com outros algoritmos, principalmente quando se tem uma grande quantidade de dados (Pritzker e Feuer, 1991). Apesar dessas vantagens, a aplicação direta da regularização L1 neste framework de aprendizagem acaba resultando em treinamento ineficiente. Na L1 a penalidade o SGD nem sempre consegue ser percebida pelo modelo, podendo as suas flutuações serem muito baixas ou consideradas como zero e, conseqüentemente, diminuindo a acurácia do algoritmo. Analisando na literatura, poucos trabalhos utilizaram o algoritmo para SGD para estudar o CO₂ no solo (Kheradmand et al., 2022; Lenka e Mohapatra, 2015; Pan et al., 2021).

A escolha de hiperparâmetros específicos para os algoritmos ML são fundamentais para as obtenções dos melhores desempenhos dos modelos. Os hiperparâmetros de um algoritmo são os parâmetros que são configurados para aumentar a eficiência do modelo, o que resulta no aumento da precisão dos modelos estudados. A escolha otimizada dos hiperparâmetros pode ser realizada pela função "GridSearchCV" do Scikit-learn. A função GridSearchCV combina todas as configurações de hiperparâmetros dados, visando obter o melhor conjunto de

hiperparâmetros que possam proporcionar o melhor desempenho, dentro de uma configuração determinada pelo pesquisador (Yang e Shami, 2020).

1.2. OBJETIVO

O objetivo desta tese foi determinar a emissão de CO₂ do solo (FCO₂), em função das variáveis que compõem o sistema solo-planta-atmosfera, utilizando técnicas de machine learning, em áreas agrícolas cultivadas com cana-de-açúcar, nos estados de São Paulo e Mato Grosso do Sul.

1.3. JUSTIFICATIVA

O entendimento da emissão de CO₂ do solo (FCO₂) e suas previsões por meio do uso das técnicas de machine learning, podem auxiliar nas tomadas de decisões, pois devido ao aumento da preocupação com a sustentabilidade agrícola e as relações existentes entre os atributos do sistema solo-planta-atmosfera pode-se escolher técnicas de produção e manejos específicos do solo cultivados com cana, proporcionando a redução dos efeitos das mudanças climáticas globais, por meio da mitigação dos GEEs para a atmosfera.

O entendimento do processo da FCO₂, em função dos atributos do sistema solo-planta-atmosfera é importante pois pode servir de base para o estudo em outras culturas agrícolas, proporcionando a redução dos efeitos das mudanças climáticas globais, por meio da mitigação dos GEEs para a atmosfera.

CAPÍTULO 2 – DESENVOLVIMENTO DE UM MODELO MECANÍSTICO PARA PREDIZER A EMISSÃO DO CO₂ DO SOLO, A PARTIR DA INFLUÊNCIA DAS VARIÁVEIS QUE COMPÕEM O SISTEMA SOLO-PLANTA-ATMOSFERA

RESUMO: A emissão de CO₂ do solo é um processo muito complexo, dependente das interações dos atributos que compõem o sistema solo-planta-atmosfera. Além disso, o aumento das concentrações de CO₂ na atmosfera tem ocasionado grande preocupação mundial, incentivando a geração de pesquisas relacionadas a emissão do CO₂ do solo e também de outras fontes, onde o uso dos modelos matemáticos estão se mostrando uma ferramenta muito eficiente, devido a sua alta capacidade de classificação e predição, produzindo modelos confiáveis e acurados. O objetivo deste trabalho foi desenvolver um modelo mecanístico-empírico, utilizando a regressão linear múltipla para prever, quantificar e determinar a influência de cada elemento do sistema solo-planta-atmosfera no processo da FCO₂, em áreas comerciais cultivadas com a cultura cana-de-açúcar no Brasil. O estudo foi realizado a partir de um conjunto de dados, oriundos de experimentos conduzidos em áreas de cana-de-açúcar nos municípios de Motuca (SP), Guariba (SP), Pradópolis (SP) e Aparecida do Taboado (MS), e o algoritmo usado foi o de regressão linear múltipla. As determinações da FCO₂ foram registradas nas áreas de estudo pelo sistema LI-COR (LI-8100). O modelo obtido foi: $FCO_2 = 2,5938 + 0,0065 \cdot \text{matéria orgânica prevista} - 0,0401 \cdot \text{porosidade livre d'água prevista}$. O valor do coeficiente de determinação do modelo foi de $R^2 = 0,824$, e o seu respectivo valor ajustado foi $R^2_{\text{ajustado}} = 0,807$. Estudos como este, relacionados à determinação de parâmetros do solo são importantes para os agricultores, pois as informações geradas podem ser utilizadas no gerenciamento e nas tomadas de decisões do processo produtivo e financeiro, sendo uma alternativa para a caracterização não só da FCO₂, mas também de outros atributos do solo, promovendo a redução de custos e uma produção mais sustentável.

Palavras-chave: atributos físicos e químicos do solo, regressão linear múltipla, gases do efeito estufa, respiração do solo, emissão de CO₂ do solo.

CHAPTER 2 – DEVELOPMENT OF A MECHANISTIC MODEL TO PREDICT SOIL CO₂ EMISSIONS, FROM THE INFLUENCE OF THE VARIABLES THAT COMPOSE THE SOIL-PLANT-ATMOSPHERE SYSTEM

ABSTRACT: The emission of CO₂ from the soil is a very complex process, dependent on the interactions of the attributes that make up the soil-plant system and atmosphere. In addition, the increase in CO₂ concentrations in the atmosphere has caused great concern worldwide, encouraging the generation of research related to the emission of CO₂ from the soil and also from other sources, where the use of mathematical models is proving to be a very efficient tool, due to its high capacity for classification and prediction, producing reliable and accurate models. The objective of this work was to develop a mechanistic-empirical model, using multiple linear regression to predict, quantify and determine the influence of each element of the soil-plant-atmosphere system on the FCO₂ process in commercial areas cultivated with sugarcane. sugar in Brazil. The study was carried out from a set of data from experiments carried out in sugarcane areas in the municipalities of Motuca (SP), Guariba (SP), Pradópolis (SP) and Aparecida do Taboado (MS) and the algorithm used was the multiple linear regression. FCO₂ determinations were recorded in the study areas by the LI-COR system (LI-8100). The model obtained was: $FCO_2 = 2,5938 + 0,0065 \cdot \text{predicted organic matter} - 0,0401 \cdot \text{predicted water-free porosity}$. The value of the coefficient of determination of the model was $R^2 = 0,824$, and its respective adjusted value was $R^2_{\text{adjusted}} = 0,807$. Studies like this one, related to the determination of soil parameters are important for farmers, because the information generated can be used in the management and decision-making of the production and financial process, being an alternative for the characterization not only of FCO₂, but also of other soil attributes, promoting cost reduction and more sustainable production.

Keywords: soil physical and chemical attributes, multiple linear regression, greenhouse gases, soil respiration, soil CO₂ emission.

2.1. INTRODUÇÃO

Nos últimos anos houve um aumento significativo nos níveis de concentração de CO₂ na atmosfera. A intensificação desse processo vem aumentando desde a era pré-industrial, devido ao aumento do uso de combustíveis fósseis e da queima de biomassa para a sua produção. De maneira geral, esses aumentos trouxeram grandes desenvolvimentos tecnológicos e melhorias na qualidade de vida dos seres humanos, porém quando se analisa o ambiente, observa-se que está havendo um desequilíbrio nas diferentes esferas do ecossistema, principalmente devido ao aumento das emissões de gases do efeito estufa (GEE). Segundo o IPCC (2014), o setor de eletricidade é o que mais colabora com as emissões globais de GEE, estima-se que a quantidade emitida por esse setor representa 25%, já a agricultura, floresta e outros usos da terra (AFOLU) representam cerca de 24% das emissões totais (Figura 2.1).

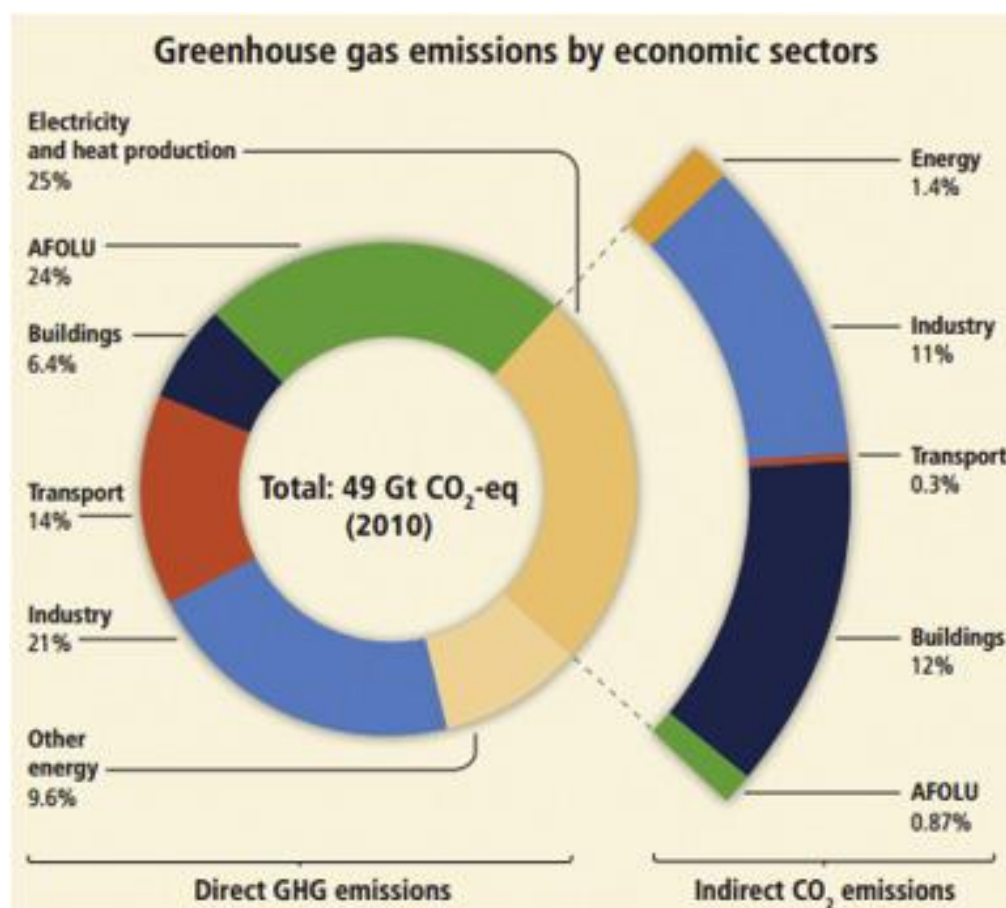


Figura. 2.1: Emissões antropogênicas totais de gases de efeito estufa (GEE) (gigatonelada de CO₂ - equivalente por ano, GtCO₂-eq / ano) dos setores econômicos em 2010. Fonte: IPCC, 2014.

Segundo Xie et al. (2015), o aumento de CO₂ na atmosfera colabora com o armazenamento de energia térmica oriunda da radiação solar, aumentando a temperatura do planeta, e indiretamente, resultando no aumento da atividade microbiana e na intensidade da respiração das raízes, causando maior produção de CO₂ no solo e na agricultura.

Visto isto, a agricultura apresenta grande importância neste cenário e a adoção de práticas de manejo sustentáveis podem colaborar com o sequestro de carbono, fazendo com que os solos agrícolas sejam um dreno de GEE.

O Brasil é um grande produtor de alimentos, sendo conhecido informalmente como o grande celeiro do mundo. Além disso, também é o maior produtor mundial de cana-de-açúcar (*Saccharum spp.*) e o segundo maior produtor de etanol (FAO, 2017). Na Figura 2.2, segundo os dados da Companhia Nacional de Abastecimento - CONAB, na safra 2019-2020 a área total foi em torno de 10,04 milhões de hectares e a produção foi de 642,72 milhões de toneladas (CONAB, 2020).



Figura. 2.2: Infográfico da cadeia produtiva da cana-de-açúcar do Brasil e nos estados de São Paulo e Mato Grosso do Sul, conforme os dados da safra 2019/2020 da Companhia Nacional de Abastecimento – CONAB. Fonte própria.

A produção total de etanol da cana-de-açúcar no Brasil foi de 34 bilhões de litros. O estado de São Paulo foi o maior produtor do país, com 5,08 milhões de hectares, 14,13 bilhões de litros de etanol e produção de 342,61 milhões de toneladas, representando 53,31% da cana processada. Já o estado de Mato Grosso do Sul foi o quarto maior produtor, com 809,2 mil hectares, 2,94 bilhões de litros de etanol e produção de 47,52 milhões de toneladas, sendo responsável por 7,39% da produção brasileira (CONAB, 2020).

A produção sustentável de alimento e energia extraídos da cana-de-açúcar são os grandes objetivos do setor sucroenergético brasileiro. A grande importância ambiental ocorre devido aos produtos originados da sua produção como por exemplo o etanol e a bioenergia. O etanol é um biocombustível originado desta cultura e é uma das grandes alternativas que podem ser usadas no processo de redução da emissão de gases causadores do efeito estufa (Mello et al., 2014). Além da produção de etanol, também são obtidos os produtos bagaço e a palhada, que juntas podem ser usadas como biomassa para a geração de calor e bioeletricidade. Diversos investimentos em pesquisas e tecnologias estão sendo realizados, visando buscar melhorias nos processos de produção do etanol de primeira e segunda geração (Walter et al., 2014).

Analizando economicamente o setor sucroenergético, estima-se que o valor bruto anual ultrapasse US\$ 100 bilhões, colaborando com PIB em torno de US\$ 40 bilhões, representando cerca de 2% do PIB brasileiro (Gerardi et al., 2017). Esses valores estão fortemente ligados à grande sustentabilidade agrícola do processo de produção da cana-de-açúcar e a sua grande importância na economia brasileira, pois o etanol produzido pode ser usado como um combustível alternativo, com relação ao consumo de combustíveis fósseis, principalmente devido ao seu grande potencial de mitigação de GEE (IPCC, 2014). No ano de 1966, iniciou-se uma regra governamental sobre a composição da gasolina no Brasil, que motivava o uso de até 25% de etanol na gasolina (Hira e Oliveira, 2009). Com o decorrer dos anos, foi observado uma grande valorização devido aos impactos ambientais positivos deste biocombustível brasileiro, pela Agência de Proteção Ambiental - EPA, sendo reconhecido como uma alternativa de combustível renovável avançada, podendo causar uma redução da emissão de CO₂ de até 61%, quando comparado com a gasolina (Milanez et al., 2012).

A variabilidade espacial das condições edafoclimáticas das regiões brasileiras interferem diretamente na produtividade da cana-de-açúcar (Everingham et al., 2015). Além disso, o uso da terra e as mudanças de cobertura do solo causam impactos significativos nos atributos químicos, físicos e biológicos do solo, afetando na dinâmica do carbono orgânico do solo e, conseqüentemente, nas emissões dos GEE (Tavares et al., 2018). Dentre estes atributos, a temperatura do solo (T_s) e a umidade do solo (U_s) são variáveis que geralmente governam os padrões da variabilidade temporal da FCO_2 . Por outro lado, a densidade do solo, textura, teor de MOS, estoque de carbono (EstC), biomassa microbiana do solo (BMS), potencial hidrogeniônico (pH), capacidade de troca de cátions (CTC) e mineralogia do solo tem maior destaque em termos de variabilidade espacial da emissão de CO_2 do solo (FCO_2) (Bicalho et al., 2014).

As práticas de eliminação da soqueira em áreas de cana-de-açúcar pelo preparo do solo podem causar grandes perdas de carbono via FCO_2 . Quando os campos de cana eram queimados, praticamente toda a palhada era perdida, ocorrendo a liberação de uma grande quantidade de gases aerossóis na atmosfera (De Figueiredo et al., 2015). No entanto, no manejo ecológico observa-se que na colheita mecanizada ocorre o retorno dos resíduos da cultura na camada mais superficial do solo, proporcionando o acúmulo de matéria orgânica e a redução da emissão dos gases, em comparação ao sistema de cana queimado. Estima-se que o sistema de colheita mecanizada da cana crua pode retornar cerca de $15 \text{ t ha}^{-1} \text{ ano}^{-1}$ de massa seca, representando um grande potencial de acúmulo de carbono nas camadas mais superficiais do solo (Carvalho et al., 2016). Awe et al. (2015) estudaram a adição da cobertura morta e seu efeito residual no preparo do solo, por meio da variabilidade temporal e da covariância da temperatura do solo de uma lavoura de cana-de-açúcar, na região Sul do Brasil, durante o período 2011/2012 e concluíram que a cobertura da palha influenciou no regime térmico de solo, durante os períodos críticos do verão e do inverno. Além disso, também observaram que o teor da umidade do solo foi menor nos tratamentos sem a camada superficial de palhada, quando comparado com aqueles que apresentavam palhada.

A modelagem é uma ferramenta que está sendo amplamente aplicada em várias áreas de estudo como na agricultura, medicina, física, química, biologia,

economia e engenharias. A modelagem é um processo de simulação de fenômenos, que tem como objetivo diminuir o tempo de execução dos projetos e acelerar o processo de inovação, causando a diminuição dos custos e aumentando o rendimento da empresa, promovendo a solução dos problemas e a otimização dos procedimentos. A sua representação geralmente é feita através de um conjunto de equações que descrevem com precisão os processos, cujo o modelo possui a sua peculiaridade, em função das características analisadas (Castiglioni et al., 2013).

Dependendo do fenômeno simulado, os modelos são classificados em empíricos ou mecanísticos. Os modelos empíricos usam a relação empírica existente entre atributos, como por exemplo, as características do solo, variáveis meteorológicas, variedades de culturas, índices de vegetação e outras variáveis para descrever os dados observados. Geralmente, as relações das variáveis para estes modelos são realizadas por técnicas estatísticas, como por exemplo, correlação e regressão. Já os modelos mecanísticos apresentam maior complexidade, comparado aos empíricos e têm como principais objetivos representar os processos fisiológicos da cultura, como por exemplo, a fotossíntese, respiração, transpiração (Thornley e Johnson, 1990), além de explicar como cada variável influencia no crescimento, produtividade e demais processos que envolvem o desenvolvimento das culturas. Existem diversos modelos mecanísticos que simulam o crescimento, desenvolvimento e processos das culturas (He et al., 2012; Keating et al., 1999; Singels e Bezuidenhout, 2002). Entre os modelos existentes para a cana-de-açúcar, pode-se citar o modelo biofísico AgrolBIS que simula o crescimento da cana-de-açúcar por meio de um sistema de produção agroecológico (Cuadra et al., 2012) e o CANEGRO, que foi criado com o objetivo de auxiliar no planejamento da produção da cultura na África do Sul, por pesquisadores da região (Inman-Bamber, 2000).

O estudo da modelagem matemática, aplicado na emissão de CO₂ desperta grande interesse e preocupação da população mundial, visto que este é um dos parâmetros que são usados para estimar o aquecimento global. Além disso, é necessário entender a dinâmica do CO₂ no sistema solo-planta-atmosfera e como isso interfere no dia-a-dia das pessoas, nos ecossistemas, nos fenômenos naturais e no globo terrestre. Sendo assim, a modelagem da emissão do CO₂ pode ser uma ferramenta estratégica na estruturação das tomadas de decisões sobre o manejo e

uso do solo, a partir do cenário vivido atualmente e da perspectiva que se almeja a curto e a longo prazo, visando diminuir os impactos das mudanças climáticas no planeta Terra.

As emissões de CO_2 do solo podem ser influenciadas por diversos atributos do solo que influenciam na transferência e produção de CO_2 no solo, como por exemplo a densidade do solo, a textura do solo, os espaços de poros cheios de ar, a temperatura do solo e o teor de água no solo (Da Silva et al., 2020).

O processo de FCO_2 pode ser medido por um modelo matemático a partir da produção de CO_2 no solo, em intervalos de tempo, logo após a preparação do solo (La Scala et al., 2008). O modelo assume que a decomposição de C na matéria orgânica prontamente decomposta é uma equação de primeira ordem, que é influenciada pelo cultivo e pelos tipos de preparação do solo, pois a preparação do solo acelera o processo de oxidação orgânica do C, resultando na liberação de grandes quantidades de CO_2 na atmosfera, em um curto período de semanas (La Scala et al., 2006). Esse fator contém informações sobre o potencial de emissão de CO_2 em um ponto específico da área de estudo, determinado pela estabilidade do carbono, e pode ser utilizado para a tomada de decisão, em relação ao uso e manejo do solo, podendo ser observados diferentes potenciais de emissão ou sequestro de CO_2 .

A determinação precisa da perda de carbono em áreas agrícolas é de grande relevância para o Brasil, pois pode auxiliar o país no cumprimento dos seus objetivos em relação à redução de emissões de GEE. De Figueiredo et al. (2015) estudaram as interferências do manejo da palha, em diferentes sistemas de colheita e áreas de cana-de-açúcar e observaram que a FCO_2 adicional foi induzida pelo plantio direto. A mudança dos sistemas de colheita de cana queimada para cana crua pode melhorar o estoque do carbono no solo, especialmente quando a palha foi mantida na superfície do solo.

De Souza et al. (2019) caracterizaram as FCO_2 , em função da distribuição dos poros do solo em Latossolo e Argissolo, em área total e sob escarificação na linha de plantio para o cultivo da cana-de-açúcar e obtiveram modelos, usando a técnica de análise multivariada de exploração de dados da regressão linear múltipla, em que as variáveis independentes foram a umidade do solo e a temperatura do solo, carbono

orgânico associado aos minerais e partículas orgânicas de carbono, já a variável dependente foi a variação da emissão do solo média. Além disso, observaram que a intensidade do preparo do solo em textura mais argilosa colaborou com a FCO_2 , provavelmente por causa do alto teor de carbono disponível para atividade microbiana, em comparação aos solos de textura mais arenosa.

Nos últimos anos, estes tipos de estudos têm sido amplamente utilizados dentro das ciências agrárias e ambientais, sendo uma grande ferramenta para a compreensão da FCO_2 (Panosso et al., 2011; Ma e Feng, 2013; Li et al., 2014; Da Silva et al., 2020; Moitinho et al., 2021). Embora alguns pesquisadores realizaram estudos sobre as emissões de CO_2 na agricultura, mais resultados e pesquisas são necessários para o maior entendimento desse processo. Além disso, poucos trabalhos usam a modelagem para prever a emissão do CO_2 , utilizando os atributos do sistema solo-planta-atmosfera, principalmente devido à alta complexidade e a grande gama de variáveis que influenciam no sistema.

2.2. OBJETIVO

Sabendo que a FCO_2 é um processo dependente de muitas variáveis que fazem parte do sistema solo-planta-atmosfera e que modelagem, por meio da abordagem multivariada exploratória de dados é uma estratégia que pode auxiliar na extração das relações existentes entre as variáveis que estão fortemente associadas à FCO_2 , neste contexto, o objetivo deste trabalho foi desenvolver um modelo mecanístico-empírico, utilizando a análise multivariada exploratória de dados da regressão linear múltipla para prever, quantificar e determinar a influência de cada elemento do sistema solo-planta-atmosfera no processo da FCO_2 , usando as variáveis dos atributos físicos e químicos do solo, do balanço hídrico da cultura e os dados climáticos, em áreas comerciais cultivadas com a cultura cana-de-açúcar no Brasil.

2.3. MATERIAL E MÉTODOS

2.3.1. Caracterização da Área de Estudo

O estudo foi realizado em áreas canaviais comerciais, sob o sistema de cultivo da cana crua, nas cidades de Aparecida do Taboado (APT), em Mato Grosso do Sul (MS), e Motuca (MOT), Guariba (GUA) e Pradópolis (PAD), no estado de São Paulo

(Figura 2.3). Os solos do estado de São Paulo foram classificados como Latossolo Vermelho eutroférrico. A textura do solo de Motuca foi interpretada como muito argilosa (argila > 600 g kg⁻¹). Em Guariba e Pradópolis, o solo foi definido como argiloso (350 g kg⁻¹ < argila < 600 g kg⁻¹). Já a classificação do solo de Aparecida do Taboado (MS) foi a Latossolo Vermelho distroférrico de textura argilosa, conforme o Sistema Brasileiro de Classificação de Solos - SIBCS (Santos et al., 2013 b).

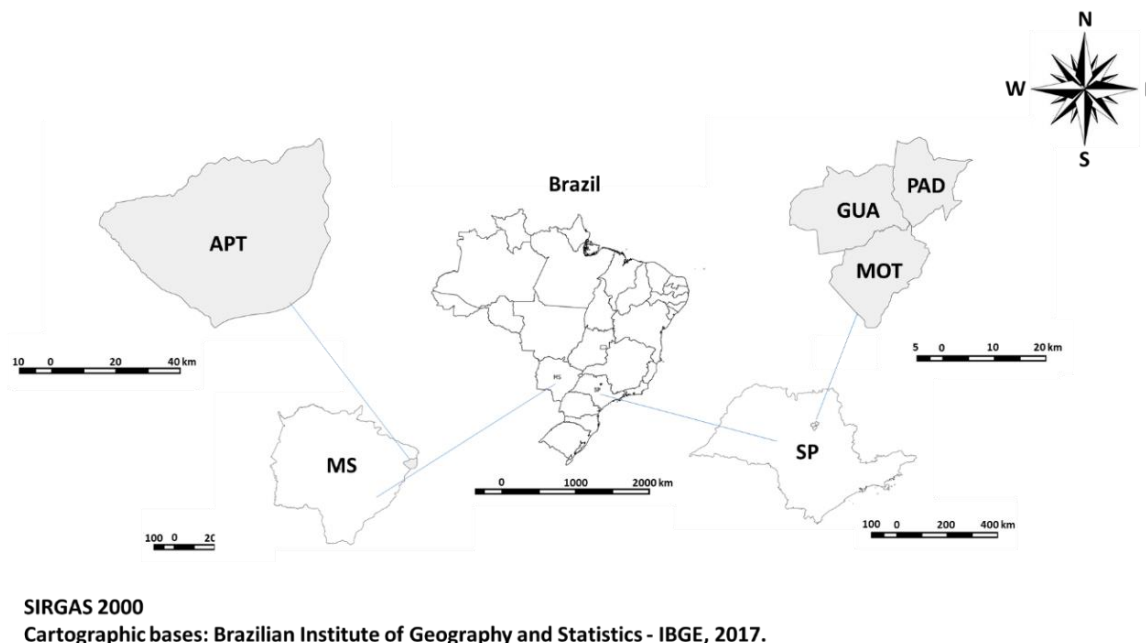


Figura 2.3: Mapas dos municípios das áreas experimentais do estado de São Paulo - SP (Pradópolis - PAD, Guariba - GUA e Motuca - MOT) e Mato Grosso do Sul - MS (Aparecida do Taboado - APT).

(PAD – Pradópolis, GUA – Guariba, MOT – Motuca, APT – Aparecida do Taboado, SP – São Paulo, MS – Mato Grosso do Sul, km – kilometers, N – north, S – south, E – east, W – west, SIRGAS - Geocentric Reference System for the Americas and IBGE – Brazilian Institute of Geography and Statistics). Fonte: Da Silva et al. (2020).

Conforme a metodologia de Köppen, as classificações climáticas das áreas estudadas foram Aw. As precipitações anuais das cidades do estado de São Paulo foram 1400 mm, com chuvas concentradas de novembro até fevereiro, com temperaturas médias anuais de 22,2 °C e as elevações dos municípios em relação ao nível do mar foram 550 (MOT), 620 (GUA) e 515 (PAD) metros. Para o ATP, a elevação do nível do mar foi de 370 metros, as temperaturas médias anuais foram de

23,7 °C e a precipitação anual de 1300 mm, com pluviosidade concentrada em novembro a fevereiro.

2.3.2. Medições de FCO₂, Us e Ts

Todas as mensurações de FCO₂, Ts e Us foram realizadas concomitantemente nos pontos da grade, no período da manhã (7h00-9h00). As FCO₂ foram medidas no início do desenvolvimento da cana-de-açúcar, através do sistema LI-COR (LI-8100). As câmaras usadas para as mensurações da FCO₂ apresentaram áreas de contatos circulares de 83,7 cm² e volumes internos de 854,2 cm³, essas câmaras eram encaixas em colares de PVC que foram previamente instaladas nos pontos amostrais no solo, a uma profundidade de 3 cm. A mensuração foi obtida por meio das mudanças nas concentrações do CO₂ dentro das câmaras, que foram registradas pelo sistema de espectroscopia, utilizando um refratômetro infravermelho de gás (IRGA). Os tempos utilizados para as obtenções das concentrações do CO₂ nos interiores das câmaras foram de 1,5 minutos para determinar as interpolações das concentrações do CO₂ no solo (contada a cada 2,5 segundos).

Além disso, a LI-8100 possui um sensor térmico integrado que foi utilizado para medir a temperatura do solo (Ts). O sensor é formado por uma haste de 0,20 m que foi introduzido no interior do solo, ao redor dos colares de PVC. Para medir a umidade do solo (Us) foi utilizado o equipamento de TDR (Time Domain Reflectometry - Hydrosense TM, Campbell Scientific, Austrália), que é formado por uma sonda composta por duas hastes de 0,12 m, que foram inseridas no interior do solo, em torno dos colares de PVC.

2.3.3. Análise de atributos físicos e químicos do solo

Após a finalização das avaliações de FCO₂, Ts e Us, foram coletadas as amostras de solo na profundidade de 0 a 0,10 m e secas ao ar. Posteriormente, as amostras foram destorroadas e passaram por uma peneira com malha de 2 mm para a determinação da saturação por bases (BS), soma de bases (V%), potencial de hidrogênio (pH), cálcio (Ca), magnésio (Mg), potássio (K), fósforo (P), capacidade de troca catiônica (CTC) e acidez potencial (H+Al), seguindo a metodologia sugerida pela EMBRAPA (1997). O fósforo disponível foi obtido conforme a metodologia da resina

de troca iônica (Raij, 2001). Já o teor de carbono orgânico total (COT) nas amostras de solo foi obtida pelo analisador de carbono orgânico total (COT-V, Shimadzu Labs), utilizando o módulo de amostra sólida (modelo SSM-5000, Shimadzu Labs). A obtenção da matéria orgânica do solo (MOS) foi realizada por meio da multiplicação do COT por 1,72 (EMBRAPA, 1997).

Para a determinação da densidade do solo (D_s) e a densidade de partículas do solo (D_p) foram utilizadas as amostras indeformadas extraídas por meio do amostrador adaptado a cilindros, que tinha dimensões médias de 4,0 cm de altura e 5,0 cm de diâmetro interno (EMBRAPA, 1997).

Para classificar os poros em microporosidade (Micro) macroporosidade (Macro) foi utilizado um funil de placa porosa, sob uma tensão de saturação amostral inicial de 60 cm de altura de coluna d'água. O volume de água que permaneceu na amostra, após este processo, corresponde aos microporos e os macroporos foram calculados pela diferença (EMBRAPA, 1997). Já a porosidade total do solo (PT) foi obtida pela quantificação do valor da densidade. A fração do espaço poroso preenchido com ar (PLA) foi determinada pela diferença entre a fração da porosidade preenchida pela água (PPA) e o volume total de poros (VTP), resultando no teor de água do solo (U_s).

2.3.4. Estoque de carbono

O estoque de carbono foi medido para uma profundidade de 0,10 metros, com relação à massa equivalente do solo para determinar a variabilidade da densidade do solo, nas diferentes áreas de estudo, usando a equação 1 (Bayer et al., 2000):

$$C_{stock} = \frac{(CO \cdot D_s \cdot h)}{10} \quad (1)$$

Em que:

C_{stock} = estoque de carbono ($Mg \text{ ha}^{-1}$);

CO = é a concentração encontrada pela análise de amostras de solo ($g \text{ kg}^{-1} = \text{SOM} / 1,724$);

D_s = densidade do solo ($kg \text{ dm}^{-3}$);

h = espessura do solo no intervalo de profundidade amostrado (cm).

2.3.5. Evapotranspiração potencial

Para o cálculo da Evapotranspiração potencial (ETP, mm) foram usados os seguintes dados climáticos: radiação solar no topo da atmosfera (Q_0 , MJ m⁻² dia⁻¹), temperatura média do ar (T_m , °C), precipitação pluvial (Prec, mm), umidade relativa (UR, %) e a velocidade do vento (U_s , m s⁻¹). Esses dados foram obtidos por meio da plataforma National Aeronautics and Space Administration/Prediction of World Wide Energy Resources - NASA/POWER (Stackhouse et al., 2016). A plataforma disponibiliza dados meteorológicos diários, com resolução espacial de 1°, o que representa um grid de 110,57 km aproximadamente.

O método utilizado para a determinação da Evapotranspiração Potencial (ETP) foi o Penman e Monteith (Allen et al., 1998) na escala diária, conforme as Equações 2 a 5.

$$ETP = \frac{0,408 \times s \times (Rn - G) + \frac{\gamma \times 900 \times U_2 \times (es - ea)}{T + 273}}{s + \gamma \times (1 + 0,34 \times U_2)} \quad (2)$$

$$s = \frac{4098 \times es}{(T + 273)^2} \quad (3)$$

$$ea = \frac{UR \times es}{100} \quad (4)$$

$$es = 0,6108 \times e^{\frac{17,27 \times T}{237,3 + T}} \quad (5)$$

Em que:

ETP é a evapotranspiração potencial (mm);

Rn é a radiação líquida (MJ m⁻² d⁻¹);

UR é umidade relativa (%);

G é o fluxo de calor no solo igual a 0 MJ m⁻² d⁻¹;

T é a temperatura do ar (°C);

γ é a constante psicrométrica igual a 0,063 kPa °C⁻¹;

s é a declividade da curva da pressão de vapor *versus* temperatura (kPa °C⁻¹);

U_2 é a velocidade do vento (m s⁻¹) na altura de 2 m;

es é a pressão da saturação de vapor (kPa);

ea é a pressão de vapor atual (kPa).

2.3.6. Balanço hídrico climatológico

A deficiência hídrica (DEF, mm), armazenamento de água no solo (STO, mm), excedente hídrico (EXC, mm), porosidade livre d'água (PLA, mm) e umidade do solo

diário (Um, mm) foram estimados na escala diária para todas as áreas de estudo, conforme a metodologia de Thornthwaite e Mather (1955) (Figura 2.4).

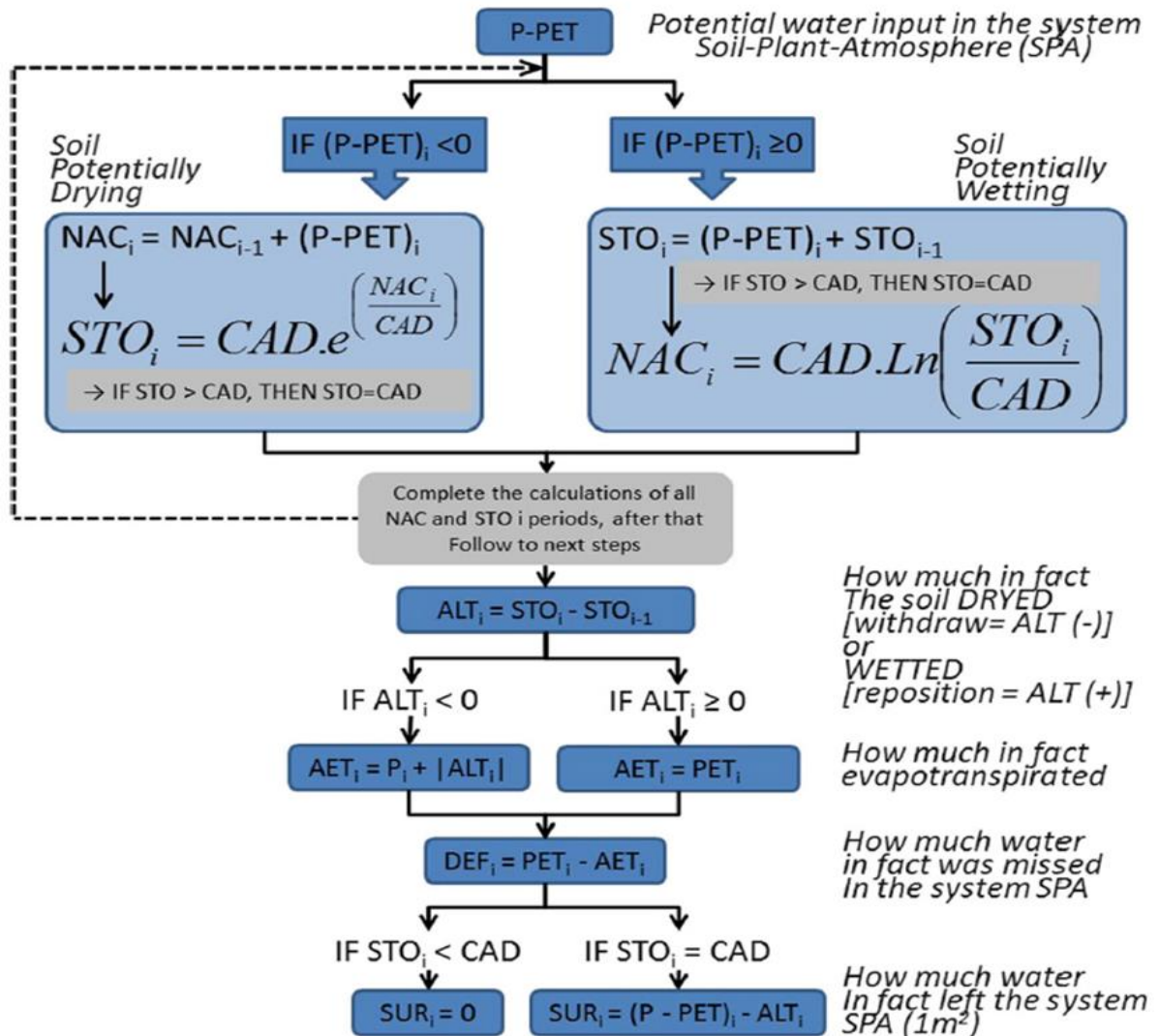


Figura 2.4: Fluxograma do cálculo do balanço Hídrico, conforme Thornthwaite e Mather modificado por Rolim et al. (2020).

Legenda:

AET é a evapotranspiração real (mm);

ALT é o armazenamento de água no solo do mês atual - armazenamento de água no solo do mês anterior (mm);

i é o período do mês;

NAC é negativo acumulado (mm);

Prec é a precipitação (mm);

PET é a evapotranspiração potencial (mm);

WC é a capacidade de água disponível (mm);

STO (ARM) é o armazenamento de água no solo (mm);

EXC é o excedente de água no sistema solo-planta-atmosfera (mm);

CAD é a capacidade de água disponível do solo (mm);

DEF é a deficiência de água no sistema solo-planta-atmosfera (mm).

2.3.7. Obtenções das variáveis ambientais e variáveis do solo para a elaboração do modelo da FCO₂

Inicialmente, foi criado um modelo empírico que teve como objetivo relacionar as influências que os atributos do sistema solo-planta-atmosfera possuem no processo de FCO₂. O sistema foi dividido em três partes: física do solo, química do solo, planta e atmosfera. Caracterizando os três sistemas, observa-se que a parte física solo e as variáveis da planta e atmosfera representam o processo de transporte de CO₂, tanto no solo, como também por meio da respiração das plantas e da fotossíntese. Já a parte química do solo representou o processo de formação de CO₂, por meio das reações que são observadas entre a microbiota do solo e de decomposição da matéria orgânica do solo (Figura 2.5).

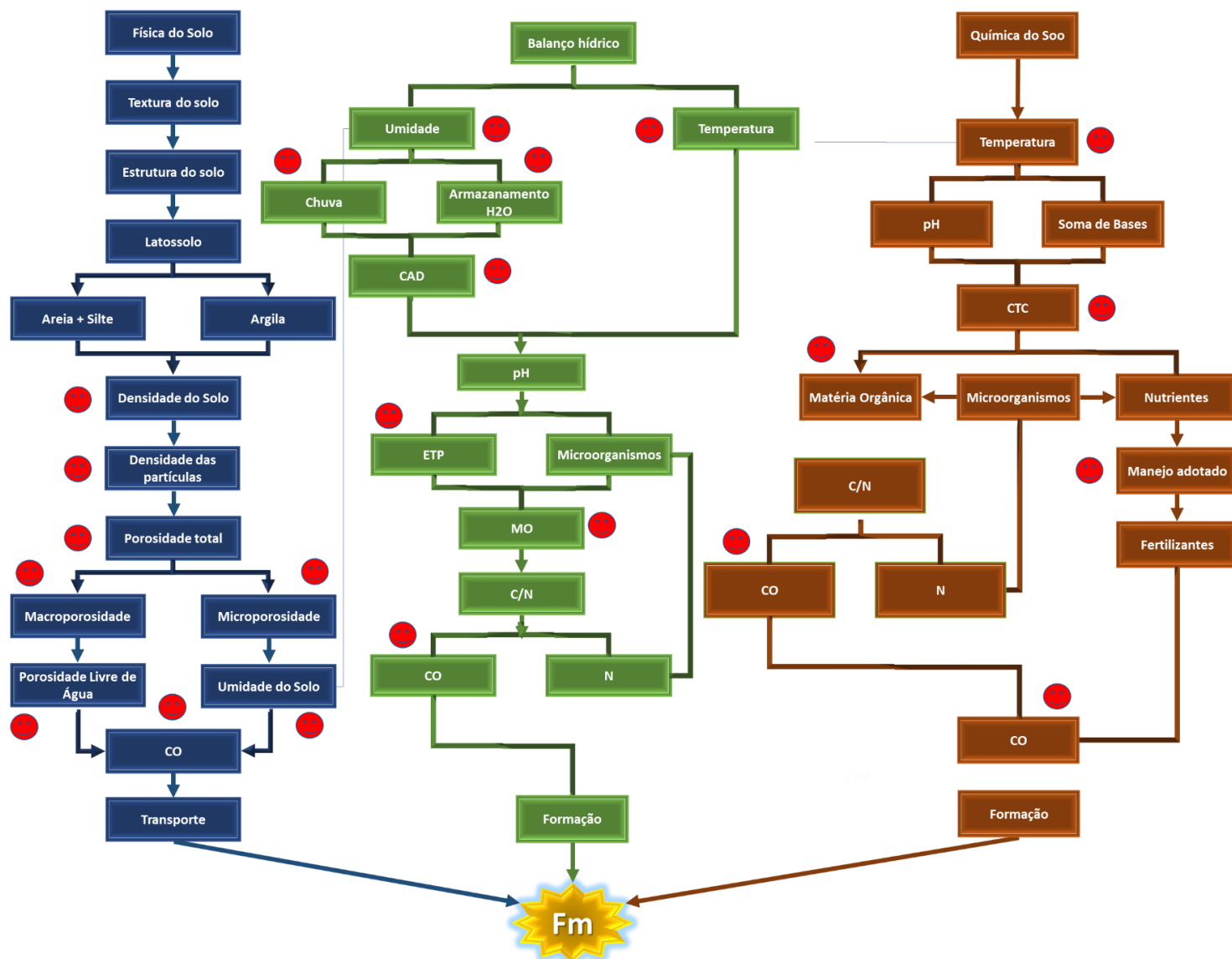


Figura 2.5: Influência das variáveis do sistema solo-planta-atmosfera no processo da FCO₂. Fonte: Própria.

CO = carbono orgânico no solo; H₂O = água; CAD = capacidade de água disponível; pH = potencial hidrogeniônico; ETP = evapotranspiração potencial; MO = matéria orgânica do solo; C/N = relação carbono/nitrogênio no solo; N = nitrogênio no solo; CTC = capacidade de troca de cátions.

As propriedades físicas do solo estão ligadas às estimativas dos processos físicos que ocorrem no solo, principalmente aqueles relacionados à movimentação de matéria e fluídos, interferindo na locomoção e na transformação de energia no solo. Para estimar o processo de transporte de CO₂ no solo, foram selecionados os seguintes atributos com base em pesquisas anteriores (Da Silva et al., 2020; Bicalho et al., 2014): porosidade livre d'água (PLA, mm), umidade solo (Us, mm), densidade das partículas (kg dm⁻³) e densidade do solo (kg dm⁻³).

As propriedades químicas dos solos, juntamente com a microbiota, são os principais causadores de transformações de energia no solo, resultantes praticamente dos processos de adsorção, fixação química, oxidação, troca e a neutralização. De maneira geral, observa-se que as características químicas do solo resultam em processos de transformação e formação de CO_2 , e os atributos que mais se relacionaram com esse processo foram a capacidade de troca de cátions, a matéria orgânica e o teor de fósforo disponível (Moitinho et al., 2015; Vicentini et al., 2019).

O balanço hídrico também pode auxiliar nos estudos dessas relações, essa metodologia permite analisar os efeitos da radiação solar, temperatura, água no sistema vegetal, balanço de energia, evapotranspiração vegetal e o balanço hídrico e hidrológico no sistema. Para este estudo foi realizado o balanço hídrico diário sequencial, por meio da metodologia de Thornthwaite e Mather (1955) para extrair os atributos que mais se relacionam ao processo de transporte, e de forma secundária a formação de CO_2 no solo. As variáveis escolhidas foram a capacidade de armazenamento de água no solo (ARM, mm), evapotranspiração potencial e real (ETP e ETR, mm), a porosidade livre d'água (PLA, mm), temperatura do ar (T_a , °C), temperatura do solo (T_s , °C) e umidade do solo (U_s , mm). Logo abaixo observa-se o modelo do impacto do sistema solo-planta-atmosfera (Figura 2.6) no desenvolvimento das plantas.

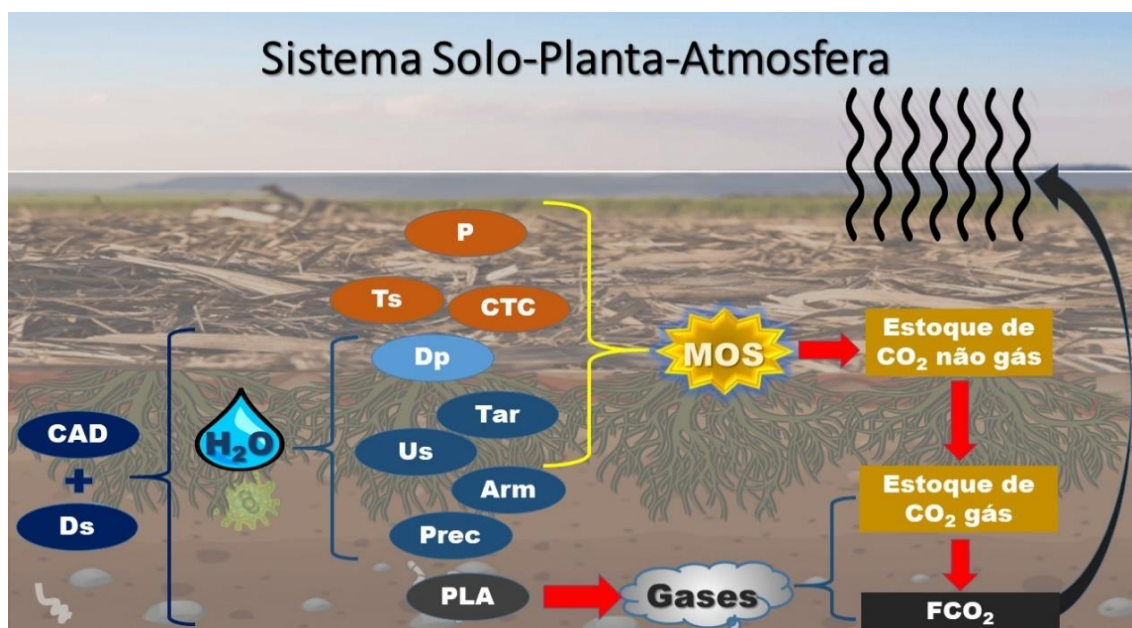


Figura 2.6: Modelo da análise empírica da influência dos atributos que compõem o sistema solo-planta-atmosfera, no processo da FCO_2 . Fonte: Própria.

FCO₂ = Fator de emissão de CO₂ (μmol CO₂ m⁻² s⁻¹); Dp = densidade das partículas (kg dm⁻³); Ds = densidade do solo (kg dm⁻³); PLA = porosidade livre de água (%); MOS = matéria orgânica do solo (g dm⁻³); P = fósforo disponível (mg dm⁻³); CTC = capacidade de troca de cátions (mmol_c dm⁻³); CAD = capacidade de água disponível (mm); ARM = armazenamento de água no solo (mm); Ts = temperatura do solo (°C); Us = umidade do solo (% do volume); Tar = temperatura do ar até 2 metros do solo (°C); Prec = precipitação (mm), H₂O = água (mm).

Logo após a criação do modelo empírico, as variáveis foram submetidas as técnicas de correlação pelo método de Pearson, que posteriormente foram utilizadas para estimar as subvariáveis MOS e PLA estimadas para serem usadas na predição da FCO₂.

2.3.8. Metodologia da seleção dos pontos amostrais para a elaboração do modelo mecanístico da emissão de CO₂ do solo

Para obter a melhor acurácia do modelo foram selecionados para cada localidade (MOT, PAD, GUA e ATP), os pontos amostrais espaciais que apresentaram maior similaridade, dentro de cada área de estudo, sendo agrupados pelas relações existentes entre os atributos do solo. Para essa determinação foi utilizada a metodologia de agrupamento dos grafos aleatórios, conforme método proposto por Anis et al. (2014). Um grafo G é formado pelo par (V, E), onde V representa o conjunto de objetos classificados como vértices e, já E são os subconjuntos que formam o conjunto V de ordem 2, denominadas arestas. Analisando o modelo grafo aleatório G (n, p), em que 0 < p < 1, um grafo aleatório com o grupo de vértices V = {1, 2, 3, 4, . . ., n} é formado por meio da inclusão de cada aresta independente e com probabilidade p, ou seja, sendo H um grafo com o conjunto de vértices V e com m arestas, dessa forma observa-se na Equação 6:

$$\mathbb{P}(G = H) = p^m (1 - p)^{N - m} \quad (6)$$

Em que: $N = \binom{k}{2}$, essa notação representa as quantidades máximas de arestas possíveis de um grafo com k vértices (Bollobás, 2001).

Para a predição dos modelos, foram selecionados os pontos espaciais que apresentaram maior similaridade dentro das áreas estudadas nas cidades de MOT, GUA, PAD e ATP, em função dos seus respectivos números de interações, conforme pode ser observado nos Quadros 2.1 e Figura 2.7.

Logo após a seleção dos pontos amostrais, foram obtidos os valores médios diários das variáveis estudadas, corresponde ao período de extração dos dados e foi

realizada a estatística descritiva e análises exploratória da distribuição das variáveis, utilizando métodos gráficos e analíticos para auxiliar na seleção da técnica que seria usada para a elaboração do modelo. Para encontrar possíveis lacunas e outliers usaram os métodos gráficos histograma, boxplot e gráfico de probabilidade normal. Já por meio dos métodos analíticos foram analisadas as correlações entre as variáveis pelo método e Pearson.

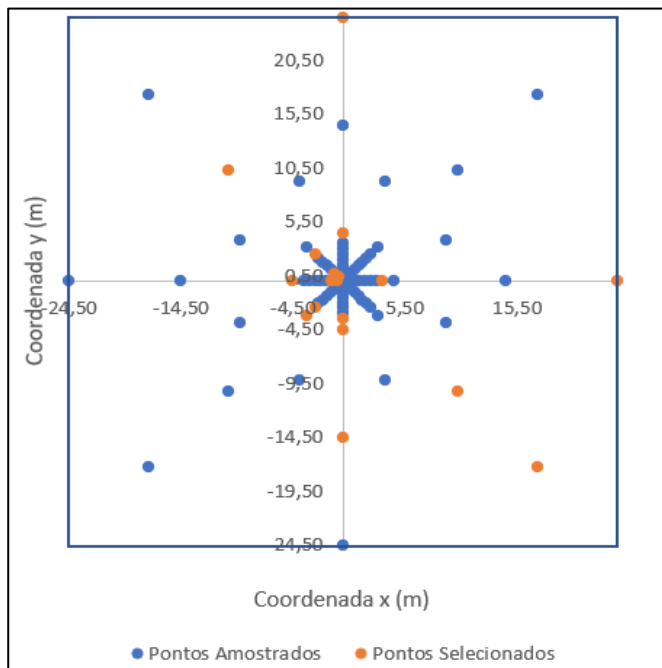
Para analisar a variabilidade geral dos dados, foram realizados os testes de normalidade e homocedasticidade por meio dos testes Shapiro-Wilk e Levene, respectivamente. A avaliação foi feita pelo parâmetro do valor de prova, também conhecido como valor-p, p-valor ou significância. Também foi gerado um histograma de barras, com o objetivo de delimitar as suas respectivas classes e frequências. Os testes de normalidade e homocedasticidade foram realizados ao nível de significância (α) de 5% (grau de confiança de 95%).

Para o teste Shapiro-Wilk a sua interpretação foi feita através do nível de concordância entre a hipótese nula (H_0) e as observações, sendo H_0 referente à distribuição Normal. Quanto maior for o p-valor, maior é a solidez entre os dados e a hipótese nula. Dessa forma, realizou-se a seguinte regra de decisão: H_0 : (i) se p-valor $\leq \alpha$, rejeita-se H_0 , isto é, não se aceita que o conjunto de dados tenha distribuição Normal; (ii) se valor-p $> \alpha$, não se rejeita H_0 , isto é, os dados tendem a ter uma distribuição normal. A interpretação do teste de Levene é bastante semelhante ao que foi explicado anteriormente: H_0 : (i) se p-valor $\leq \alpha$, rejeita-se H_0 , isto é, os dados tendem a apresentar heterocedasticidade; (ii) se valor-p $> \alpha$, não se rejeita H_0 , isto é, os dados tendem a apresentar homoscedasticidade.

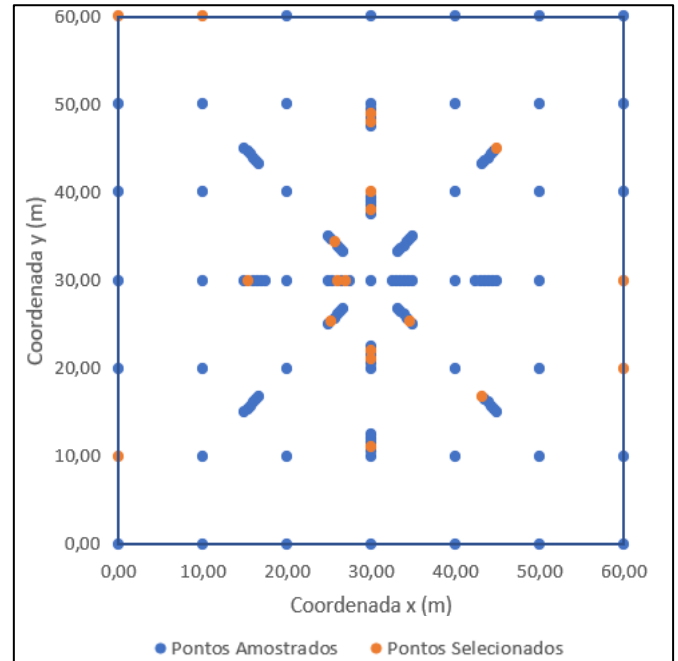
Quadro 1: Seleção dos pontos utilizados para a realização das análises estatísticas, em função das características espaço-temporais das áreas comerciais da cultura de cana-de-açúcar (*Saccharum spp.*), por meio da análise de grafos aleatórios.

Área de estudo	Dias julianos escolhidos	Número de pontos coletados (n)	Pontos com maior número de Vértices	Número de Interações
Motuca (MOT)	299, 301, 308	89	3, 10	8
			9	7
			18, 44, 59, 61, 70, 71, 80	6
			21, 38, 39, 43, 45, 63, 65, 81	5
Guariba (GUA)	195, 196, 197, 200, 201, 204, 207	141	41	9
			70, 104, 126	8
			54	7
			86, 134, 137	6
			2, 7, 8, 42, 48, 57, 97, 103, 108, 112, 114, 116	5
Pradópolis (PAD)	240, 242, 244, 247, 249, 251, 255, 258	133	74	4
			28, 40, 42, 43, 125	2
			23, 75, 101, 110, 112, 118, 126, 132	1
Aparecida do Taboado (APT)	184, 185, 188, 190, 191, 193	102	94	46
			50	45
			76, 91	44
			14, 18, 79, 96	43
			6, 41, 80	42
			68, 63	41
			97	40
			7, 43	39
			2	38

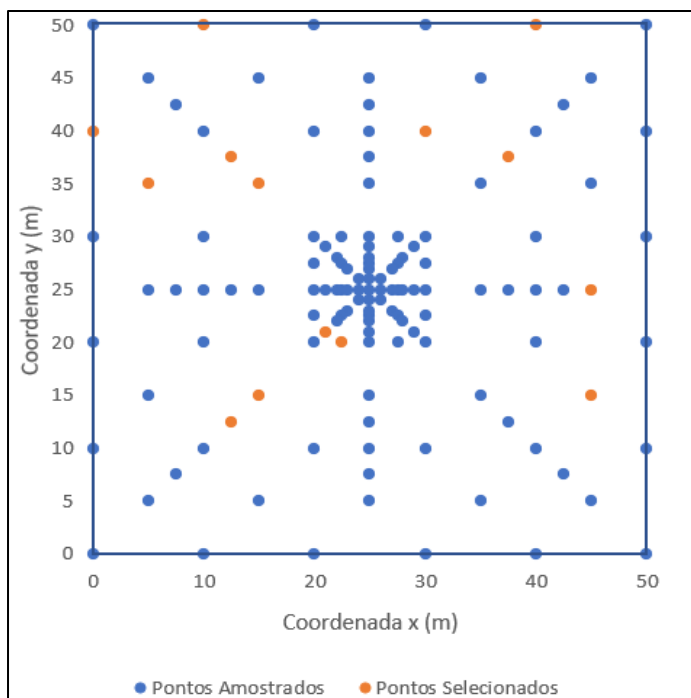
A) Motuca – SP



B) Guariba – SP



C) Pradópolis – SP



D) Aparecida do Taboado – MS

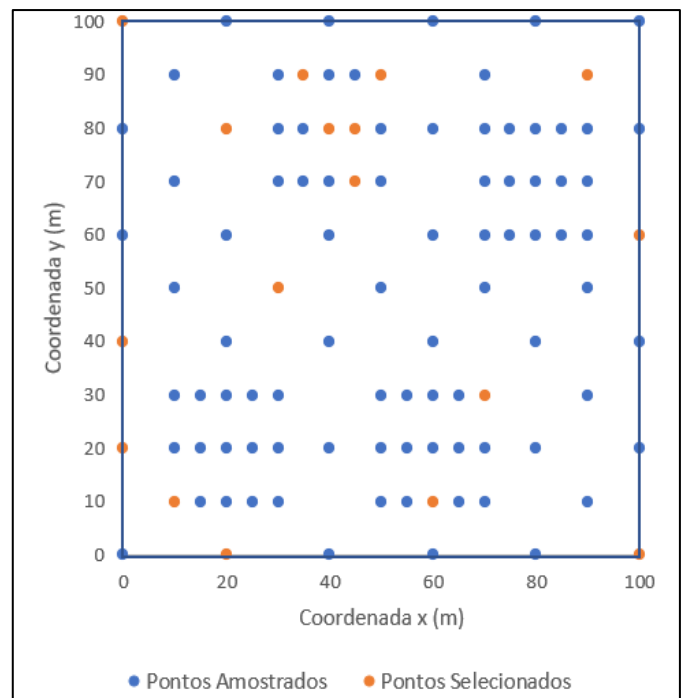


Figura 2.7: GRID amostral das áreas estudados para A) Motuca – SP, B) Guariba – SP, C) Pradópolis – SP e D) Aparecida do Taboado – MS, com os seus respectivos pontos selecionados para a análise. Pontos amostrados dentro das áreas (azuis). Nos pontos selecionados utilizaram as variáveis para a construção do modelo.

2.3.9. Elaboração do modelo mecanístico para a estimação do CO₂ do solo

A técnica estatística utilizada para a elaboração do modelo mecanístico-empírico foi a análise exploratória de dados da regressão linear múltipla, que tem como principal objetivo estabelecer uma relação matemática entre a variável dependente ou resposta (y_i) e as variáveis independentes ou explicativas (x_i), visando diminuir a quantidade de variáveis, causando o mínimo de perda de informação possível, encontrando as correlações e os principais padrões de similaridade. Neste estudo a variável dependente foi a FCO₂, já as variáveis independentes foram as variáveis MOS estimada e a PLA Estimada.

A principal utilidade da regressão linear múltipla é produzir valores para a variável dependente quando se têm as variáveis independentes, obtendo resultados que podem ser usados para aplicações específicas. Ou seja, ela pode ser usada na predição de resultados, por meio da regra estatística dos mínimos quadrados ordinários.

A regressão polinomial fundamenta-se no ajustar de uma reta a um conjunto de pares de dados observados, onde y é a variável dependente, x_1 é a variável regressora ou explanatória, e o termo de erro e_i , o indicador da i -ésima observação. Geralmente a representação de um modelo de regressão polinomial de grau m é apresentada na Equação 7:

$$(E y / x_i) = \hat{y}_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \dots + \beta_j x_i^j + \dots + \beta_m x_i^m \quad (7)$$

Na qual:

$E(y/x_i)$ expressa o valor da variável dependente y na observação $i = 1, \dots, n$;

x_i é o valor da variável independente x na observação $i = 1, \dots, n$;

β_0 é o coeficiente linear que indica o local que a função corta o eixo dos y ;

β_j representa o coeficiente angular ligado à j -ésima potência da variável independente x .

Também foi utilizado o MMQ, a minimização da soma dos quadrados dos erros para descobrir os parâmetros do modelo (Equação 8).

$$q = \sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (8)$$

Para analisar a precisão do modelo com relação a distribuição dos dados, foram usados os métodos analíticos do coeficiente de determinação (Equação 9) e seu valor

ajustado (Equação 10). Para avaliar a qualidade do ajuste foi realizado o teste da distribuição de Fischer-Snedecor (Quadro 2.2), que teve como objetivo verificar se a hipótese nula, ou seja, se todos os coeficientes do modelo eram nulos: $H_0: \beta_i = 0$ para cada valor de i .

$$R^2 = \left[\frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \right] \quad (9)$$

$$R^2_{aj} = \left[1 - \frac{(1-R^2) \times (n-1)}{N-k-1} \right] \quad (10)$$

Quadro 2.2: Análise de variância da regressão utilizando o teste de Fischer-Snedecor.

Fontes de Variação	Soma dos Quadrados (SQ)	Graus de Liberdade (G.L.)	Quadrados Médios (QM)	F
Regressão	$\sum_{i=1}^n (\hat{y}_i - \bar{y})^2$	K	$\frac{SQ_{regressão}}{GL_{regressão}}$	$\frac{QM_{regressão}}{QM_{erro}}$
Erro (resíduo)	$\sum_{i=1}^n (y_i - \hat{y}_i)^2$	n - k - 1	$\frac{SQ_{erro}}{GL_{regressão}}$	
Total	$\sum_{i=1}^n (y_i - \bar{y}_i)^2$	n - 1		

Fonte: adaptado de Zar, 1999.

Nas equações e no Quadro 2, tem-se que:

n é a número de observações;

y_i é a i -ésima observação;

$\bar{y}$ é a média das observações;

$\hat{y}_i$ é o valor estimado para a i -ésima observação;

k é o grau do polinômio predito.

Para as análises dos resíduos e dos valores estimados foram usados os métodos gráficos histograma, boxplot, gráfico de probabilidade normal e diagrama de dispersão. Além disso, para todas as estatísticas descritivas e análises exploratórias multivariadas, todos os dados foram padronizados com média zero e variância σ^2 .

Para determinar a melhor capacidade de água disponível (CAD) foram testados 7 níveis: 100, 60, 45, 40, 35, 30 e 20 mm, que foram mensuradas a partir da umidade do solo. Para a calibração utilizaram referências literárias (CIIAGRO, IAC 2020, Santos et al., 2013 a, De Oliveira et al., 2020) que corresponderem a CAD no estágio de desenvolvimento pós colheita e início do crescimento da cana-soca e encontraram o valor de 40 mm, que posteriormente foi assumido para todos locais estudados. Logo abaixo, observa-se o modelo da regressão polinomial quadrática, utilizado para determinar a melhor relação da CAD (mm, variável dependente), com a variável independente umidade do solo (mm), conforme a Equação 11.

$$(CAD_estimado) = \beta_0 x^2 + \beta_1 x + \beta_2 \quad (11)$$

Em que:

CAD_estimado é o valor da variável dependente y na observação $i = 1, \dots, n$;

x = coeficiente angular de ordem de 1;

x^2 = coeficiente angular de ordem de 2.

β_2 é o coeficiente linear que indica o local que a função corta o eixo dos y para CAD_estimado;

β_1 e β_2 , representam os coeficientes angulares das variáveis da umidade do solo.

Para estimar as variáveis independentes do modelo mecanístico – empírico foi usada a técnica da regressão linear múltipla. Para estimar a variável independente capacidade de água disponível do solo (CAD, mm) foram utilizadas os valores médios dos pontos amostrais da umidade do solo oriundo do balanço hídrico (U_s , mm), densidade do solo (D_s , kg dm^{-3}), densidade das partículas D_p (kg dm^{-3}) e armazenamento de água no solo (ARM, mm) (Equação 12). Para a estimação da U_s (mm) foram utilizadas evapotranspiração potencial (ETP, mm), precipitação (Prec, mm) e armazenamento de água do solo (ARM, mm) (Equação 13).

$$(CAD_estimado) = \beta_0 + \beta_1 U_s + \beta_2 D_s + \beta_3 D_p + \beta_4 ARM \quad (12)$$

Em que:

CAD_estimado é o valor da variável dependente y na observação $i = 1, \dots, n$;

β_0 é o coeficiente linear que indica o local que a função corta o eixo dos y para CAD_estimado;

$\beta_1, \beta_2, \beta_3$ e β_4 representam os coeficientes angulares das variáveis Us, Ds, Dp e Arm, respectivamente.

$$(Us_estimado) = \beta_0 + \beta_1ETP + \beta_2PREC + \beta_3ARM \quad (13)$$

Em que:

Us_estimado é o valor da variável dependente y na observação $i = 1, \dots, n$;

β_0 é o coeficiente linear que indica o local que a função corta o eixo dos y para Us_estimado;

β_1, β_2 e β_3 representam os coeficientes angulares das variáveis ETP, PREC e ARM, respectivamente.

Para estimar a variável independente porosidade livre de água decorrente do balanço hídrico (PLA, %) foram utilizadas os valores médios dos pontos amostrais da Dp (kg dm^{-3}) e a porosidade livre de água decorrente do balanço hídrico (PLA_BH, %), conforme a Equação 14.

$$(PLA_estimado) = \beta_0 + \beta_1Dp + \beta_2PLA_BH \quad (14)$$

Em que:

PLA_estimado é o valor da variável dependente y na observação $i = 1, \dots, n$;

β_0 é o coeficiente linear que indica o local que a função corta o eixo dos y para PLA_estimado;

β_1 e β_2 representam os coeficientes angulares das variáveis Dp e PLA_BH, respectivamente.

Já para a predição da matéria orgânica do solo (MOS, g dm^{-3}) foi utilizada a capacidade de troca de cátions (CTC, mmolc dm^{-3}), fósforo disponível (P, mg dm^{-3}), temperatura do ar até 2 metros do solo (T_{air} , °C), temperatura do solo (T_s , °C); umidade do solo (Us, % do volume), densidade das partículas (Dp, kg dm^{-3}), conforme a Equação 15.

$$(MOS_estimado) = \beta_0 + \beta_1CTC + \beta_2P + \beta_3T_{air} + \beta_4T_s + \beta_5Us + \beta_6Dp \quad (15)$$

Sendo:

MOS_estimado é o valor da variável dependente y na observação $i = 1, \dots, n$;

β_0 é o coeficiente linear que indica o local que a função corta o eixo dos y para MOS_estimado;

$\beta_1, \beta_2, \beta_3, \beta_4, \beta_5$ e β_6 representam os coeficientes angulares das variáveis CTC, P, Tair, Ts, Us e Dp, respectivamente.

Após as predições anteriores, foi criado o modelo mecanístico-empírico de predição do FCO_2 , por meio da técnica de regressão linear múltipla, utilizando os valores preditos de PLA e MOS, sendo que as variáveis independentes foram a PLA e a MOS estimados e a variável dependente foi a FCO_2 medida em campo, conforme a Equação 16.

$$(\text{FCO}_2\text{-estimado}) = \beta_0 + \beta_1 \text{PLA_estimado} + \beta_2 \text{MOS_estimado} \quad (16)$$

Em que:

$\text{FCO}_2\text{-estimado}$ é o valor da variável dependente y na observação $i = 1, n$;

β_0 é o coeficiente linear que indica o local que a função corta o eixo dos y para $\text{FCO}_2\text{-estimado}$;

β_1 e β_2 representam os coeficientes angulares das variáveis PLA_estimado e MOS_estimado , respectivamente.

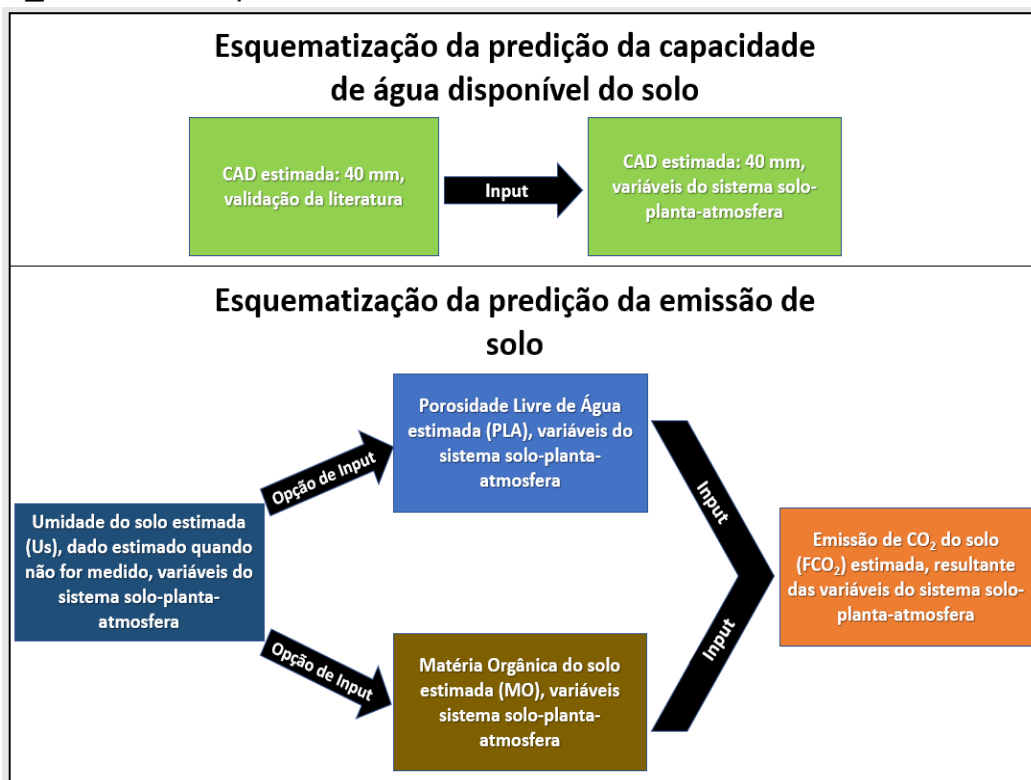


Figura 2.8: Sistematização das predições dos modelos da capacidade de água disponível do solo (CAD, 40 mm) e da emissão de CO_2 do solo FCO_2).

2.3.10. Medidas estatísticas de avaliação de desempenho

Para a avaliar o desempenho dos algoritmos calibrados foram utilizados os seguintes índices estatísticos:

- Correlação de Pearson (r), Equação 17;
- Coeficiente de determinação (R^2), Equação 18;
- Coeficiente de determinação ajustado (R^2 adj), Equação 19;
- Erro quadrático médio (MSE), Equação 20;
- Raiz quadrada do erro-médio (RMSE), Equação 21;
- Média Percentual Absoluta do Erro (MAPE), Equação 22;
- Critério de Informação de Akaike (AIC), Equação 23 (Akaike, 1974);
- Critério Bayesiano de Schwarz (BIC), Equação 24 (Schwarz, 1978).

Para a obtenção da maior confiabilidade e acurácia das regressões também foram realizados o teste F a 5% de probabilidade.

$$r = \frac{\sum_{i=1}^n (Y_{obs_i} - \bar{Y}_{obs}) \times (Y_{est_i} - \bar{Y}_{est})}{\sqrt{\sum_{i=1}^n (Y_{obs_i} - \bar{Y}_{obs})^2 \times \sum_{i=1}^n (Y_{est_i} - \bar{Y}_{est})^2}} \quad (17)$$

$$R^2_{aj} = \left[1 - \frac{(1-R^2) \times (n-1)}{N-k-1} \right] \quad (18)$$

$$R^2_{aj} = \left[1 - \frac{(1-R^2) \times (n-1)}{N-k-1} \right] \quad (19)$$

$$MSE = \frac{\sum_{i=1}^N (Y_{obs_i} - Y_{est_i})^2}{N} \quad (20)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (Y_{obs_i} - Y_{est_i})^2}{N}} \quad (21)$$

$$MAPE(\%) = \frac{\sum_{i=1}^n \left(\left| \frac{Y_{est_i} - Y_{obs_i}}{Y_{obs_i}} \right| \times 100 \right)}{N} \quad (22)$$

$$AIC = -2 \ln L \left(\frac{x}{\hat{\theta}} \right) + 2(p) \quad (23)$$

$$BIC = -2 \log p_i(x_n) \quad (24)$$

Em que:

Yest_i: variável estimada na posição i;

Yobs_i: variável observada na posição i;

N: número de dados;

k: número de variáveis independentes na regressão;

$L \left(\frac{x}{\hat{\theta}} \right)$: função de máxima verossimilhança;

p : probabilidade de sucesso;

$\hat{\theta}_i$: estimador de máxima verossimilhança para o k_i -ésimo vetor paramétrico θ_i
do modelo $f_i\left(\frac{x_n}{\theta_i}\right)$

x_n : n -ésimo de x .

Quanto menor for os valores da MSE, RMSE, MAPE, AIC e BIC, melhores são os ajustes dos modelos. No entanto, para os parâmetros r , R^2 , R^2 adj, quanto maior for os valores em módulo, maior será a correlação entre as variáveis.

2.4. RESULTADOS E DISCUSSÃO

2.4.1. Distribuição dos dados e Estatística Descritiva

Quando se analisou os valores do p-valor das variáveis (Tabela 2.1) ficou evidente que a grande maioria apresentou índices inferiores a 0,05, apontando que as variáveis possuem uma tendência de não apresentarem distribuição normal, com exceção a Us (0,3456), a CAD (0,4117) e a Ts (0,0945) que apresentaram p-valor superior ao α de 5%, conforme pode ser observado na Figura 2.9 e confirmada pelo teste de homocedasticidade de Levene (p-valor = $3,36 \times 10^{-36}$).

Tabela 2.1: Teste de normalidade pelo método de Shapiro-wilk da emissão de CO₂, atributos atmosféricos e características físico-químicos do solo de todas as áreas estudadas.

Atributos	Estatística do teste	P-valor
FCO₂	0,85	0,0020*
Dp	0,84	0,0017*
Ds	0,57	2,88e-07*
PLA	0,76	6,83e-05*
MOS	0,77	8,59e-05*
P	0,81	0,0004*
CTC	0,78	0,0002*
EstC	0,74	3,68e-05*
CAD	0,96	0,4117
ARM	0,87	0,0047*
PLA_BH	0,87	0,0047*
Ts	0,93	0,0945
Us	0,96	0,3456
Tair	0,86	0,0030*
Precipitação	0,39	6,04e-09*
ETP	0,84	0,0017*

N = 24; Métricas seguidas pelo * foi significativa ao nível de 5% de probabilidade, pelo teste *Shapiro-Wilk*; FCO₂ = Fator de emissão de CO₂ ($\mu\text{mol CO}_2 \text{ m}^{-2} \text{ s}^{-1}$); Dp = densidade das partículas (kg dm^{-3}); Ds = densidade do solo (kg dm^{-3}); PLA = porosidade livre de água (%); MOS = matéria orgânica do solo (g dm^{-3}); P = fósforo disponível (mg dm^{-3}); CTC = capacidade de troca de cátions ($\text{mmol}_c \text{ dm}^{-3}$); Estc = estoque de carbono no solo (Mg ha^{-1}); CAD = capacidade de água disponível (mm); ARM = armazenamento de água no solo (mm); PLA_BH = porosidade livre de água decorrente do balanço hídrico (%); Ts = temperatura do solo ($^{\circ}\text{C}$); Us = umidade do solo (% do volume); Tair = temperatura do ar até 2 metros do solo ($^{\circ}\text{C}$); Precipitação = precipitação (mm); ETP = evapotranspiração potencial (mm).

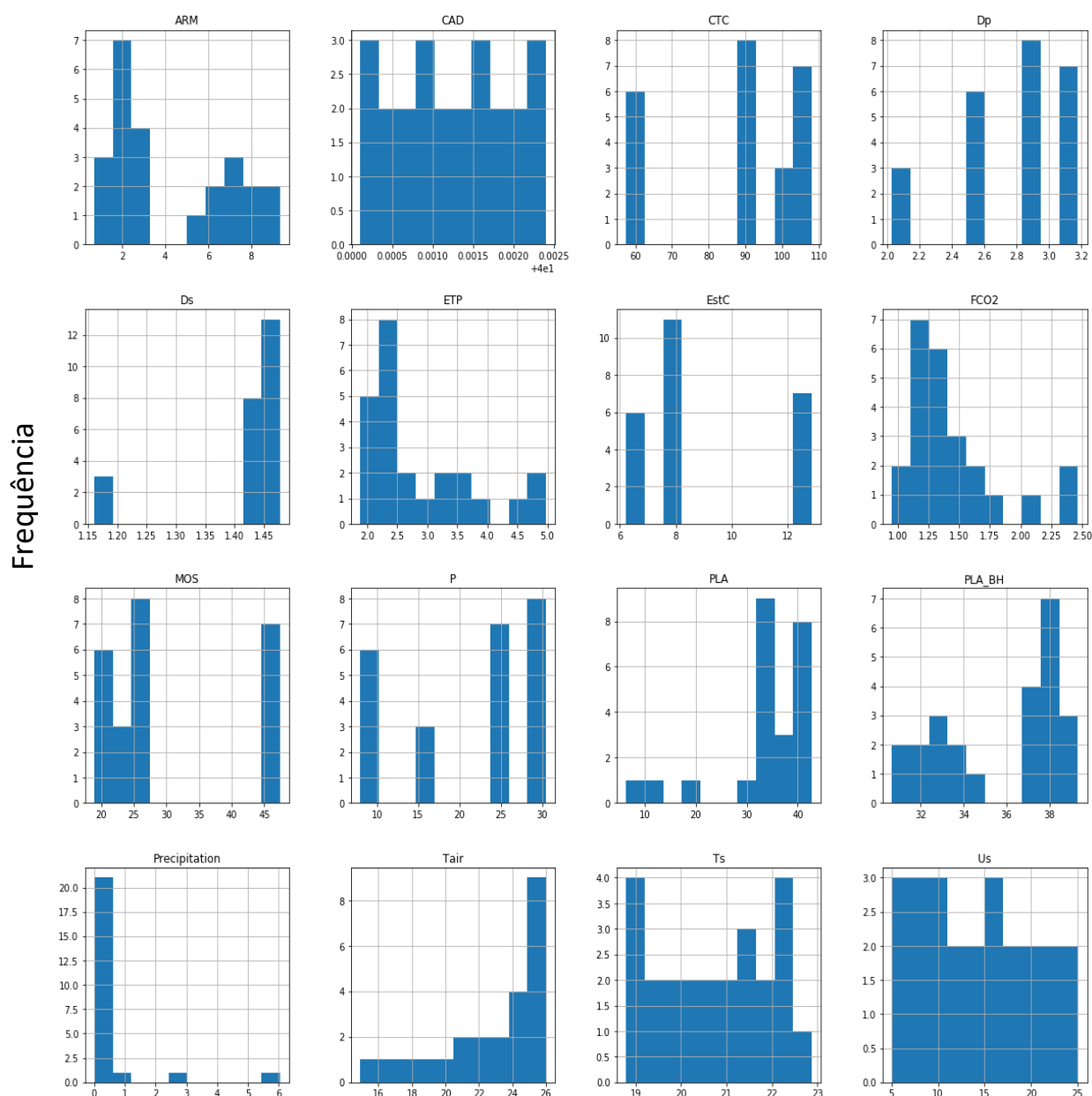


Figura 2.9: Histograma da emissão de CO₂, atributos atmosféricos e características físico-químicos do solo de todas as áreas estudadas.

N = 24; FCO₂ = Fator de emissão de CO₂ ($\mu\text{mol CO}_2 \text{ m}^{-2} \text{ s}^{-1}$); Dp = densidade das partículas (kg dm^{-3}); Ds = densidade do solo (kg dm^{-3}); PLA = porosidade livre de água (%); MOS = matéria orgânica do solo (g dm^{-3}); P = fósforo disponível (mg dm^{-3}); CTC = capacidade de troca de cátions ($\text{mmol}_c \text{ dm}^{-3}$); EstC = estoque de carbono no solo (Mg ha^{-1}); CAD = capacidade de água disponível (mm); ARM = armazenamento de água no solo (mm); PLA_BH = porosidade livre de água decorrente do balanço hídrico (%); Ts = temperatura do solo ($^{\circ}\text{C}$); Us = umidade do solo (% do volume); Tair = temperatura do ar até 2 metros do solo ($^{\circ}\text{C}$); Precipitation = precipitação (mm); ETP = evapotranspiração potencial (mm).

Além disso, também foi realizada a estatística descritiva para determinar as variações que cada atributo possuía dentro do conjunto de dados. De maneira geral, observou-se que a maioria das variáveis apresentaram valores de coeficiente de variação inferiores a 30% ($\text{FCO}_2 = 27,01\%$, $\text{Dp} = 13,11\%$, $\text{Ds} = 7,22\%$, $\text{PLA} = 27,92\%$, $\text{CTC} = 22,42\%$, $\text{EstC} = 29,88\%$, $\text{CAD} = 0\%$, $\text{PLA_BH} = 8,03\%$, $\text{Ts} = 6,37\%$ e $\text{Tair} = 13,88\%$), indicando que os dados possuíram uma baixa ou média dispersão, o que resultou em dizer que os dados tendem a ser homogêneos. Porém as variáveis MOS ($37,35\%$), P ($42,56\%$), ARM ($68,07\%$), Us ($43,16\%$), Precipitação ($297,37\%$) e ETP ($31,22\%$) apresentaram valores de CV superiores a 30%, indicando que os dados tendem a ter uma alta dispersão (heterogêneos).

Tabela 2.2: Estatísticas descritivas da emissão de CO_2 , atributos atmosféricos e características físico-químicos do solo de todas as áreas estudadas, $N = 24$.

Atributos	Média	Desvpd	Máx / Min	CV
<i>FCO₂</i>	1,44	0,39	2,46 / 0,95	27,01
<i>Dp</i>	2,78	0,36	3,18 / 2,03	13,11
<i>Ds</i>	1,42	0,10	1,48 / 1,16	7,22
<i>PLA</i>	33,74	9,42	42,77 / 6,34	27,92
<i>MOS</i>	30,67	11,46	47,42 / 18,93	37,35
<i>P</i>	21,24	9,04	30,46 / 7,97	42,56
<i>CTC</i>	88,62	19,87	108,10 / 57,43	22,42
<i>EstC</i>	8,93	2,67	12,86 / 6,21	29,88
<i>CAD</i>	40,00	0,00	40,00 / 40,00	0,00
<i>ARM</i>	4,22	2,87	9,33 / 0,69	68,07
<i>PLA_BH</i>	35,78	2,87	39,30 / 30,67	8,03
<i>Ts</i>	20,69	1,32	22,88 / 18,79	6,37
<i>Us</i>	14,27	6,16	25,03 / 5,03	43,16
<i>Tair</i>	22,94	3,18	26,00 / 14,90	13,88
<i>Precipitação</i>	0,45	1,35	6,05 / 0,00	297,37
<i>ETP</i>	2,85	0,89	4,98 / 1,88	31,22

Desvpd = desvio padrão; Máx = Máximo; Mín = Mínimo; CV = coeficiente de variação (%); FCO_2 = Fator de emissão de CO_2 ($\mu\text{mol CO}_2 \text{ m}^{-2} \text{ s}^{-1}$); Dp = densidade das partículas (kg dm^{-3}); Ds = densidade do solo (kg dm^{-3}); PLA = porosidade livre de água (%); MOS = matéria orgânica do solo (g dm^{-3}); P = fósforo disponível (mg dm^{-3}); CTC = capacidade de troca de cátions ($\text{mmol}_c \text{ dm}^{-3}$); EstC = estoque de carbono no solo (Mg ha^{-1}); CAD = capacidade de água disponível (mm); ARM = armazenamento de água no solo (mm); PLA_BH = porosidade livre de água decorrente do balanço hídrico (%); Ts = temperatura do solo ($^{\circ}\text{C}$); Us = umidade do solo (% do volume); Tair = temperatura do ar até 2 metros do solo ($^{\circ}\text{C}$); Precipitação = precipitação (mm); ETP = evapotranspiração potencial (mm).

Analisando de maneira mais específica as principais variáveis usadas para a determinação do modelo, observou-se que a variação dos valores da FCO_2 foram

entre 0,95 e 2,46 $\mu\text{mol CO}_2 \text{ m}^{-2} \text{ s}^{-1}$, sendo que os maiores valores observados foram na área de Motuca e os menores em Pradópolis (Tabela 2.3).

Tabela 2.3: Estatísticas descritivas da emissão de CO_2 , atributos atmosféricos e características físico-químicos do solo, referentes às áreas de Motuca – SP (N = 3), Guariba – SP (N = 7), Pradópolis – SP (N = 8) e Aparecida do Taboado (N = 6).

Motuca					Guariba			
Atributos	Média	Desvpd	Máx / Min	CV	Média	Desvpd	Máx / Min	CV
FCO₂	2,30	0,18	2,46 / 2,11	7,72	1,48	0,19	1,84 / 1,29	12,81
Dp	2,03	0,00	2,03 / 2,03	0,00	3,18	0,00	3,18 / 3,18	0,00
Ds	1,16	0,00	1,16 / 1,16	0,00	1,48	0,00	1,48 / 1,48	0,00
PLA	11,82	6,32	18,73 / 6,34	53,43	34,49	1,39	36,73 / 32,13	4,03
MOS	23,72	0,00	23,72 / 23,72	0,00	47,42	0,00	47,42 / 47,42	0,00
P	16,11	0,00	16,11 / 16,11	0,00	24,28	0,00	24,28 / 24,28	0,00
CTC	102,48	0,00	102,48 / 102,48	0,00	108,10	0,00	108,10 / 108,10	0,00
EstC	7,85	0,00	7,85 / 7,85	0,00	12,86	0,00	12,86 / 12,86	0,00
CAD	40,00	0,00	40,00 / 40,00	0,00	40,00	0,00	40,00 / 40,00	0,00
ARM	7,05	0,78	7,87 / 6,31	11,12	7,48	1,55	9,33 / 5,07	20,70
PLA_BH	32,95	0,78	33,69 / 32,13	2,38	32,53	1,55	34,93 / 30,67	4,76
Ts	21,66	1,18	22,88 / 20,53	5,44	19,11	0,34	19,76 / 18,79	1,79
Us	23,97	1,06	25,03 / 22,91	4,42	18,93	2,09	21,88 / 16,07	11,05
Tair	16,39	1,47	17,84 / 14,90	8,97	22,14	1,97	24,58 / 19,14	8,89
Precipitação	0,51	0,48	0,96 / 0,00	94,67	1,30	2,36	6,05 / 0,00	181,08
ETP	4,68	0,29	4,98 / 4,40	6,13	2,18	0,19	2,48 / 1,88	8,83
Pradópolis					Aparecida do Taboado			
FCO₂	1,20	0,14	1,39 / 0,95	11,58	1,29	0,22	1,59 / 1,06	17,09
Dp	2,86	0,00	2,86 / 2,86	0,00	2,59	0,00	2,59 / 2,59	0,00
Ds	1,42	0,00	1,42 / 1,42	0,00	1,47	0,00	1,47 / 1,47	0,00
PLA	41,23	0,73	42,77 / 40,50	1,77	33,84	3,24	37,66 / 28,36	9,57
MOS	27,43	0,00	27,43 / 27,43	0,00	18,93	0,00	18,93 / 18,93	0,00
P	30,46	0,00	30,46 / 30,46	0,00	7,97	0,00	7,97 / 7,97	0,00
CTC	89,78	0,00	89,78 / 89,78	0,00	57,43	0,00	57,43 / 57,43	0,00
EstC	7,93	0,00	7,93 / 7,93	0,00	6,21	0,00	6,21 / 6,21	0,00
CAD	40,00	0,00	40,00 / 40,00	0,00	40,00	0,00	40,00 / 40,00	0,00
ARM	1,78	0,80	2,96 / 0,69	44,69	2,26	0,42	2,90 / 1,76	18,80
PLA_BH	38,22	0,80	39,30 / 37,04	2,08	37,74	0,42	38,25 / 37,10	1,13
Ts	21,06	0,96	22,27 / 19,61	4,58	21,57	0,90	22,36 / 20,06	4,16
Us	12,18	2,04	15,16 / 9,33	16,76	6,77	1,32	8,57 / 5,03	19,55
Tair	25,69	0,31	26,00 / 25,09	1,22	23,48	1,37	25,11 / 21,46	5,85
Precipitação	0,00	0,00	0,01 / 0,00	282,59	0,03	0,05	0,13 / 0,00	158,36
ETP	3,20	0,46	3,78 / 2,61	14,41	2,26	0,11	2,40 / 2,12	4,78

Desvpd = desvio padrão; Máx = Máximo; Mín = Mínimo; CV = coeficiente de variação (%); FCO₂ = Fator de emissão de CO₂ ($\mu\text{mol CO}_2 \text{ m}^{-2} \text{ s}^{-1}$); Dp = densidade das partículas (kg dm^{-3}); Ds = densidade do solo (kg dm^{-3}); PLA = porosidade livre de água (%); MOS = matéria orgânica do solo (g dm^{-3}); P = fósforo disponível (mg dm^{-3}); CTC = capacidade de troca de

cátions ($\text{mmol}_c \text{ dm}^{-3}$); Estc = estoque de carbono no solo (Mg ha^{-1}); CAD = capacidade de água disponível (mm); ARM = armazenamento de água no solo (mm); PLA_BH = porosidade livre de água decorrente do balanço hídrico (%); Ts = temperatura do solo ($^{\circ}\text{C}$); Us = umidade do solo (% do volume); Tair = temperatura do ar até 2 metros do solo ($^{\circ}\text{C}$); Precipitação = precipitação (mm); ETP = evapotranspiração potencial (mm).

Já para a MOS, observou-se que houve uma variação de 18,93 a 47,42 g dm^{-3} , no município de Aparecida do Taboado foi observado os menores valores (18,93 g dm^{-3}) e os maiores valores foram em Guariba (47,42 g dm^{-3}). Com relação à PLA, foram mensurados valores entre 6,34% e 42,77%, sendo que os maiores índices foram registrados em Pradópolis e os menores em Motuca. Dessa forma, foi possível observar que ou houve uma forte influência das variáveis físicas e químicas do solo na FCO_2 , impulsionadas principalmente pela PLA e MOS. Da Silva et al. (2020) também observaram que as mesmas variáveis tiveram uma grande influência no processo de emissão e estoque de carbono no solo.

2.4.2. Correlações

As correlações da FCO_2 foram fortemente influenciadas pelos atributos atmosféricos e também pelas características físico-químicas do solo. Devido à complexidade do processo, foi realizado um estudo mais aprofundado para as variáveis FCO_2 , PLA, PLA_BH e MOS e após isso foram realizadas as correlações com as demais variáveis, cujo os valores de correlações usados para a discussão foram aqueles que apresentaram correlações menores que -0,50 e maiores que 0,50.

De maneira geral, a FCO_2 apresentou fortes correlações negativas com a Dp (-0,56), Ds (-0,77), PLA (-0,92), CAD (-0,68), PLA_BH (-0,62) e a Tair (-0,90), e positivas com a Precipitação (0,62) e Us (0,71), mostrando que as variáveis físicas do solo e também a temperatura do ar apresentaram uma forte interferência nos processos de deslocamento do CO_2 no solo (Figura 2.10). As práticas de eliminação da soqueira, pelo preparo do solo, em áreas de cana-de-açúcar podem causar grandes perdas de carbono do solo pela FCO_2 . Moitinho et al. (2015) também estudaram a influência dos atributos do solo no processo de produção e transferência de CO_2 no solo e observaram que a densidade do solo, textura do solo, porosidade livre de água, temperatura do solo e umidade do solo foram as que mais se correlacionaram com este processo.

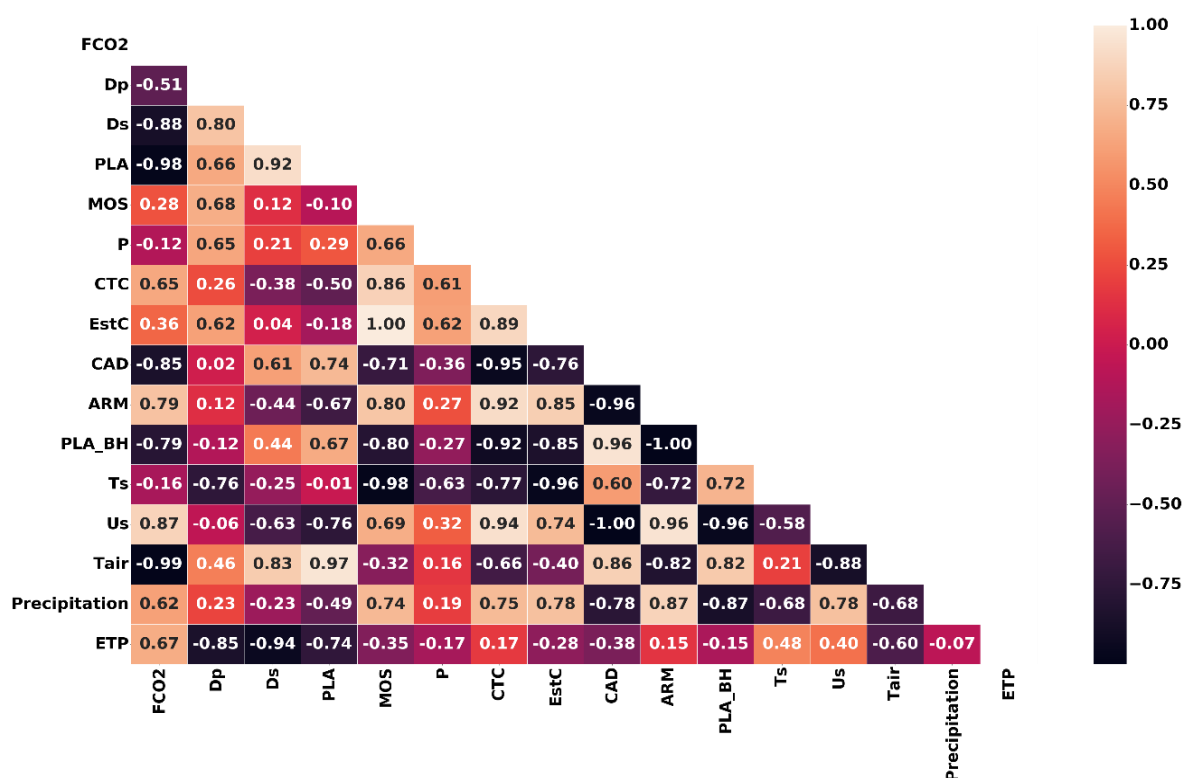


Figura 2.10: Emissão de CO₂, atributos atmosféricos e características físico-químicas do solo de todas as áreas estudadas, N = 24.

FCO₂ = Fator de emissão de CO₂ ($\mu\text{mol CO}_2 \text{ m}^{-2} \text{ s}^{-1}$); Dp = densidade das partículas (kg dm^{-3}); Ds = densidade do solo (kg dm^{-3}); PLA = porosidade livre de água (%); MOS = matéria orgânica do solo (g dm^{-3}); P = fósforo disponível (mg dm^{-3}); CTC = capacidade de troca de cátions ($\text{mmol}_c \text{ dm}^{-3}$); EstC = estoque de carbono no solo (Mg ha^{-1}); CAD = capacidade de água disponível (mm); = armazenamento de água no solo (mm); PLA_BH = porosidade livre de água decorrente do balanço hídrico (%); Ts = temperatura do solo ($^{\circ}\text{C}$); Us = umidade do solo (% do volume); Tair = temperatura do ar até 2 metros do solo ($^{\circ}\text{C}$); Precipitation = precipitação (mm); ETP = evapotranspiração potencial (mm).

Houve uma forte correlação com as variáveis físicas do solo e com as variáveis derivadas do processo do balanço hídrico. De maneira geral, a PLA correlacionou-se positivamente com a Dp (0,71), Ds (0,79), CAD (0,51), PLA_BH (0,53) e Tair (0,89), e negativamente com a FCO₂ (-0,92), ARM (-0,53), Us (-0,55) e ETP (-0,52). Já para a PLA_BH, obtida através do balanço hídrico, por meio do método de Thornthwaite e Mather (1955), observou-se as correlações negativas com a FCO₂ (-0,62), MOS (-0,69), CTC (-0,69), EstC (-0,75), ARM (-1,00), Us (-0,86) e Precipitação (-0,50). Já se as correlações positivas foram com as variáveis PLA (0,53), CAD (0,85), Ts (0,54) e

Tair (0,71). De maneira geral, analisando as interferências dessas variáveis, considerou-se que a porosidade do solo e as variáveis ligadas ao fluxo de água no solo e atmosfera, impactaram diretamente no processo de formação dos PLA. Silva et al. (2020) investigaram as variações e interações espaço-temporais da FCO₂, temperatura, umidade e porosidade livre de água durante a reforma do canavial e também observaram fortes relações da PLA com a FCO₂ e com outros atributos físicos e químicos do solo.

Para a MOS foram observadas correlações positivas com a Dp (0,77), CTC (0,79), ARM (0,69) e US (0,57). As correlações negativas foram relacionadas com as variáveis CAD (-0,60), PLA_BH (-0,69) e Ts (-0,80). Esses atributos mostraram que a fertilidade do solo, a porosidade do solo e a água no solo impactaram diretamente nos processos de formação e decomposição da matéria orgânica do solo. Vicentini et al. (2019) estudaram a caracterização da variabilidade temporal das emissões de CO₂ do solo, fluxo de O₂ do solo, teor de água e temperatura do solo, bem como suas relações em áreas reflorestadas, em Latossolo no bioma Cerrado do brasileiro e também observaram que as mesmas variáveis influenciaram no processo de produção de MOS, em função das áreas reflorestadas com pinus (PI), eucaliptos (EU) e espécies nativas (NS).

2.4.3. Análise de Regressão Linear Múltipla

Logo abaixo serão descritas as respectivas análises e equações das regressões lineares múltiplas que foram preditas para estimar os modelos das variáveis MOS (estrutura química do modelo), PLA (estrutura física do modelo) e FCO₂, em função dos atributos atmosféricos e características físico-químicos do solo, extraídas das áreas estudadas. Porém antes foram estimados a CAD 40 mm, umidade do solo e a porosidade livre de água, visando diminuir a limitação de dados medidos em campo para que pudessem serem usados nos modelos finais. Na Figura 2.11, segue uma esquematização simplificada dos modelos criados para predizer a FCO₂ do solo.

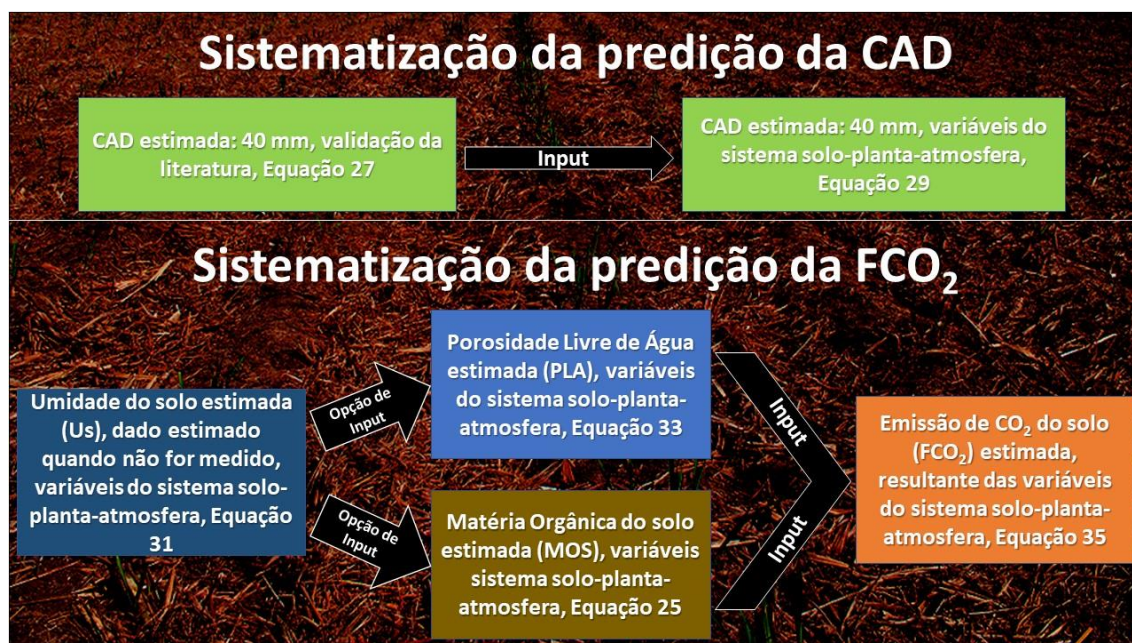


Figura 2.11: Sistematização das etapas de modelagem da Emissão de CO₂ do solo, em função dos processos formados pela matéria orgânica e porosidade livre de água do solo.

2.4.4. Regressão Linear Múltipla da MOS, em função das variáveis P, CTC, Tair, Ts, Dp e Us

A MOS se origina de todas as substâncias orgânicas que são formadas pelos resíduos animais e vegetais, em diferentes fases de decomposição (Silva e Mendonça 1997). A grande parte da MOS que é gerada no sistema de produção de cana-de-açúcar é oriunda do processo de decomposição da palhada deixada sobre o solo após a colheita, fortemente impactadas pelas interações que ocorrem entre a palhada e os atributos atmosféricos, físicos e químicos do solo.

A equação de regressão linear múltipla que foi estimada para explicar a MOS, em função das variáveis P, CTC, Tair, Ts, Dp e Us foi representada pela Equação 25:

$$MOS = -62,8932 - 0,5957P + 0,4946CTC - 3,2200e^{-15}Tair - 5,3290e^{-15}Ts + 22,4228Dp - 2,2200e^{-15}Us \quad (25)$$

Na qual:

MOS: matéria orgânica do solo (g dm⁻³);

P: fósforo disponível (mg dm⁻³);

CTC: capacidade de troca de cátions (mmol_c dm⁻³);

Tair: temperatura do ar até 2 metros do solo (°C);

Ts: temperatura do solo (°C);

Dp: densidade das partículas (kg dm⁻³);

Us: Umidade do solo (% do volume).

O valor do coeficiente de determinação deste modelo foi $R^2 = 0,999$, e o seu respectivo valor ajustado foi $R_a^2 = 0,985$, ou seja, 98,5% das variações na estimativa da MOS podem ser explicadas pelas variações dos valores das variáveis independentes. O restante, 1,5% são fontes de variações que são explicados por outros fatores (Tabela 2.4).

Tabela 2.4: Coeficientes métricos do modelo da MOS, N = 24.

Estatísticas	Coeficientes
R-squared	0,999
Adj. R-squared	0,985
F-statistic	8,263e ⁺²⁹
Prob (F-statistic)	1,77e ⁻²⁴⁹
Log-Likelihood	722,08
AIC	-1430
BIC	-1422

R-squared = Coeficiente de Determinação; Adj. R-squared = Coeficiente de Determinação ajustado; F-statistic = valor do teste de Fischer – Snedecor; Prob (F-statistic) = p-valor do teste F; Log-Likelihood = probabilidade natural logarítmica; AIC = Critério de Informação de Akaike; BIC = Critério Bayesiano de Schwarz.

Para a verificação dos resultados obtidos para a determinação da equação foi realizado o teste F, visando verificar se $H_0: \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = \beta_7 = 0$ ou H_1 : que pelo menos um β é diferente dos demais os valores, por meio do teste F e do p-valor onde foram obtidos os valores de 8,263e⁺²⁹ e 1,77e⁻²⁴⁹, respectivamente (Tabela 2.4). Esses valores mostraram que foi rejeitada a hipótese nula de que todos os coeficientes independentes seriam iguais, ao nível de significância de 0,05, indicando que pelo menos uma das variáveis independentes apresentaram relação com a MOS.

Para descrever a relação entre a variável dependente e as variáveis independentes, visando escolher o modelo que proporcione a minimização das suas

divergências, foram realizados dois testes para determinar a maior verossimilhança para o ajustamento. O primeiro teste realizado foi o teste de Akaike (AIC), cujo o valor obtido foi -1430. Além disso, também foi utilizado o Critério Bayesiano de Schwarz (BIC) para descrever as interações entre as variáveis que maximizam a probabilidade da escolha do modelo verdadeiro, cujo o valor mensurado foi -1422 (Tabela 2.4). Nos dois casos observou que as variáveis independentes apresentaram fortes relações com a variável dependente, mostrando que existem uma grande relação entre as variáveis estudadas.

Em função do modelo previsto para a MOS, o valor do coeficiente linear foi -62.8932 e apresentou significância estatística dentro do modelo (p -valor = 0,000), mostrando que esse coeficiente impacta fortemente na determinação do modelo (Tabela 2.5). O valor negativo do coeficiente linear indicou que inicialmente o solo tende a não possuir matéria orgânica, porém quando há o incremento dos resíduos animais e vegetais sob o solo, no nosso estudo, a palhada oriunda da colheita da cana-de-açúcar, observa-se a mineralização desses resíduos, que são fortemente influenciadas pelos processos ligados ao intemperismo e as características físico-químicas do solo. Os solos tropicais são conhecidos por apresentarem baixos teores de matéria orgânica e baixa fertilidade natural, principalmente devido às fortes influências causadas pelo intemperismo, o que pode resultar no maior custo de produção. A manutenção da palhada na lavoura permite o aumento da matéria orgânica do solo, podendo gerar uma quantidade de resíduos de 10 - 20 t ha⁻¹ (Trivellin et al., 2013).

Quando foram analisados os coeficientes angulares obtiveram os seguintes resultados para as variáveis dependentes: $P = -0,5957$, $CTC = 0,4946$, $T_{air} = -3,2200e^{-15}$, $T_s = -5,3290e^{-15}$, $D_p = 22,4228$ e $U_s = -2,2200e^{-15}$. Com relação ao nível de significância foram observados os p -valores de 0,000, 0,000, 0,477, 0,433, 0,000 e 0,516, respectivamente (Tabela 2.5). As variáveis que apresentaram forte poder explicativo no modelo, por meio do nível de significância foram o P , CTC e D_p . As variáveis U_s , T_{air} e T_s não apresentaram nível significativo que conseguissem impactar no modelo, porém caso seja replicado esse modelo para outro conjunto de dados, pode-se observar uma maior relação.

Tabela 2.5: Demonstrações dos coeficientes lineares e angulares, resultados do p-valor em função da estatística t para o nível de significância $\alpha = 0,05$ com os seus respectivos valores críticos, para as variáveis P, CTC, Tair, Ts, Dp e Us que foram usadas na determinação do modelo da MOS, N = 24.

	<i>Coeficientes</i>	<i>Erro Padrão</i>	<i>Resultado do teste t</i>	<i>p-valor</i>	<i>IC_{5%} Limite Inferior</i>	<i>IC_{95%} Limite Superior</i>
Constante	-62,8932	2,21e ⁻¹³	-2,84e ⁺¹⁴	0,000*	-62,893	-62,893
P	-0,5957	1,5e ⁻¹⁵	-3,98e ⁺¹⁴	0,000*	-0,596	-0,596
CTC	0,4946	1,19e ⁻¹⁵	4,14e ⁺¹⁴	0,000*	0,495	0,495
Tair	-3,22e ⁻¹⁵	4,42e ⁻¹⁵	-0,728	0,477 ^{NS}	-1,26e ⁻¹⁴	6,11e ⁻¹⁵
Ts	-5,329e ⁻¹⁵	6,64e ⁻¹⁵	-0,803	0,433 ^{NS}	-1,93e ⁻¹⁴	8,67e ⁻¹⁵
Dp	22,4228	3e ⁻¹⁴	7,47e ⁺¹⁴	0,000*	22,423	22,423
Us	-2,2200 e ⁻¹⁵	3,35e ⁻¹⁵	-0,663	0,516 ^{NS}	-9,28e ⁻¹⁵	4,84e ⁻¹⁵

Erro Padrão = Standard error ou erro padrão; Resultado teste t = resultado do teste de hipótese da média T Student; * = significativo no teste T a 5% de probabilidade; NS = não significativo no teste T a 5% de probabilidade; IC_{5%} Limite Inferior = Distribuição, razão de prevalência e intervalo de confiança de 5%; IC_{95%} Limite superior = Distribuição, razão de prevalência e intervalo de confiança de 95%; P = fósforo disponível (mg dm⁻³); CTC = capacidade de troca de cátions (mmol_c dm⁻³); Tair = temperatura do ar até 2 metros do solo (°C); Ts = temperatura do solo (°C); Dp = densidade das partículas (kg dm⁻³); Us = umidade do solo (mm).

Quando são deixados os resíduos da cultura da cana-de-açúcar na superfície do solo, observa-se o aumento significativo do teor da matéria orgânica, o equilíbrio dos agregados do solo, a diminuição da temperatura, a maior retenção de água e taxas de infiltração, que por sua vez, impactam na densidade das partículas do solo (Moitinho et al., 2013). Além disso, Moitinho et al. (2021) estudaram a caracterização dos padrões da variabilidade espacial do FCO₂ e dos atributos físicos, químicos e microbiológicos do solo, após a reforma do canavial e observaram fortes influências

das características químicas do solo, como o P e a CTC no processo da manutenção da matéria orgânica do solo.

Para analisar a variabilidade das variáveis utilizadas no modelo, foram usadas a estatística descritiva e a técnica de boxplot. O boxplot ou diagrama de caixa é uma ferramenta gráfica comparativa que permite visualizar a distribuição e valores discrepantes (outliers) dos dados, fornecendo assim um meio complementar para desenvolver uma perspectiva sobre o caráter dos dados. Essas técnicas foram utilizadas nos estudos das variáveis MOS, P, CTC, Tair, Ts, Dp e Us.

A Tair apresentou uma maior amplitude, quando comparado com a Ts (Figura 2.12), a variação dos valores encontrados da Tair foi de 14,9 °C a 26,0 °C, sendo que os maiores valores aconteceram em Pradópolis e os menores em Motuca (Tabela 2.3 e Figura 2.12). A temperatura média geral da Tair foi 22,9 °C e apresentou o valor do desvio padrão de 3,2 °C, Motuca foi a que apresentou a menor média (16,4 °C), já Pradópolis apresentou a maior média de temperatura (25,7 °C), dentre todas as áreas estudadas (Tabela 2.2 e Tabela 2.3). Com relação aos valores de coeficiente de variação, quando foram analisados todos os dados das áreas estudadas, foi encontrado o valor de 13,88%, indicando que as Tair apresentaram baixas amplitudes. Quando foi analisado cada local, também foram observados valores inferiores a 15%, sendo que Motuca apresentou o maior valor (8,97%) e Pradópolis o menor (1,22%) (Tabela 2.2, Tabela 2.3, Figura 2.12 e Figura 2.13).

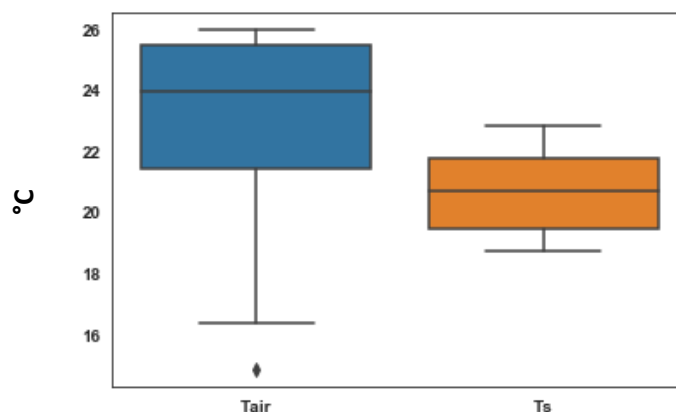


Figura 2.12: Boxplots da temperatura do ar (Tair, °C) e temperatura do solo (Ts, °C) para todas as áreas estudadas, N = 24.

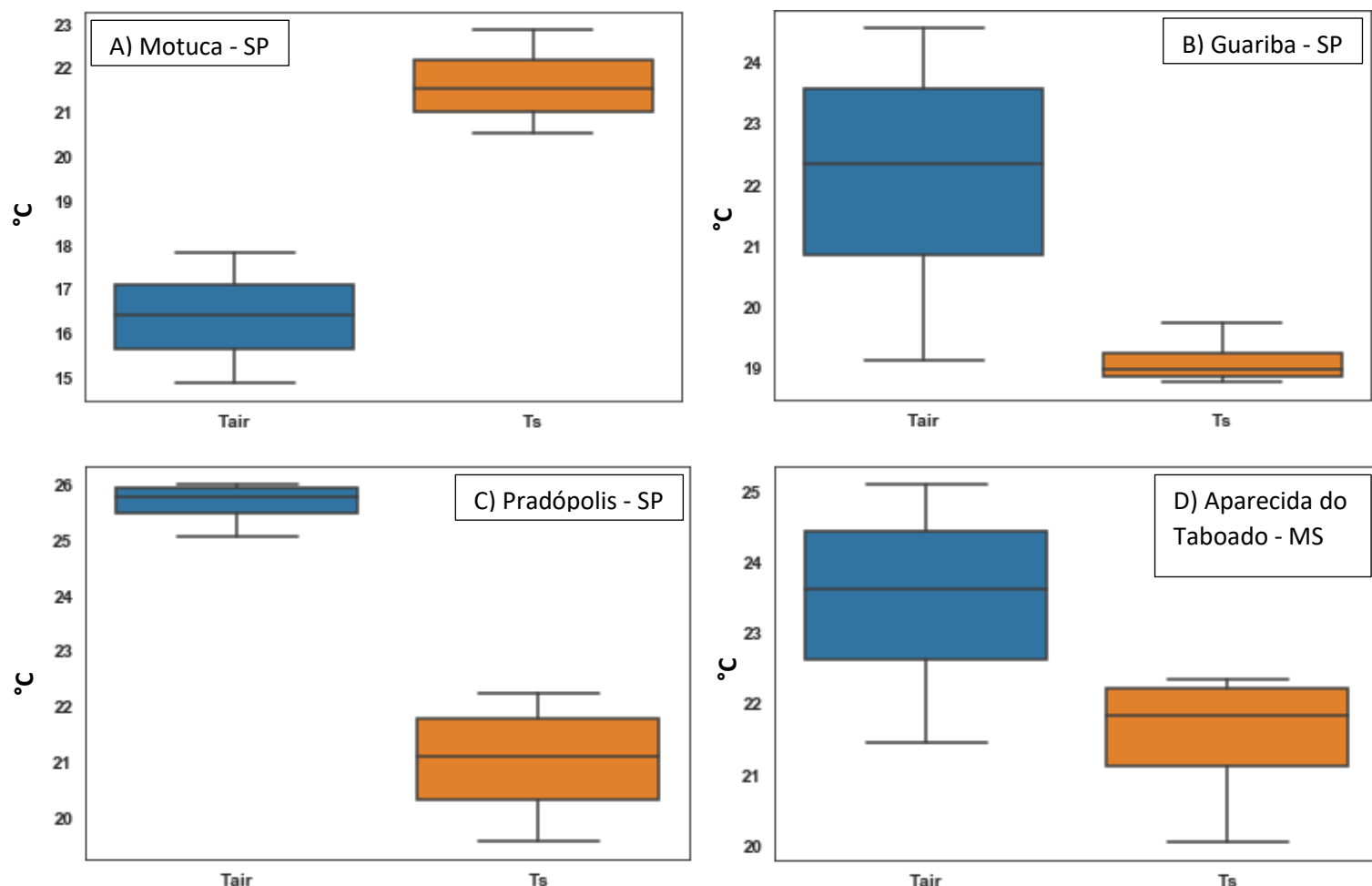


Figura 2.13: Boxplots da temperatura do ar (Tair, °C) e temperatura do solo (Ts, °C) para os municípios A) Motuca – SP, B) Guariba – SP, C) Pradópolis – SP e D) Aparecida do Taboado – MS.

Os gráficos das Figuras 2.12 e 2.13 e as Tabelas 2.2 e 2.3 mostraram as distribuições dos dados da Ts. Para a temperatura média geral da Ts foi observado o valor 20,7 °C e apresentou desvio padrão de 1,3 °C. Quando foram avaliados cada local de estudo, Guariba indicou a menor média (19,1 °C) e também o menor desvio padrão (0,3 °C). Já no município de Motuca, registrou-se a maior média de temperatura (21,7 °C) e maior desvio padrão (1,2 °C) dentre todas as áreas estudadas. Com relação a amplitude dos valores da Ts nas áreas estudadas foram observadas temperaturas de 18,8 °C e 22,9 °C. Quando analisados o coeficiente de variação para todos os dados das áreas estudadas foi encontrado o valor de 6,37%, indicando a baixa amplitude da Ts. Para cada local de estudo também observaram

valores inferiores a 15%, sendo que Guariba apresentou o menor valor (1,79%) e Motuca foi o maior (5,44%).

A fertilidade do solo também impactou na MOS, que neste estudo foram representadas pela CTC e pelo teor de P. Inicialmente foi realizada a análise da CTC do solo. Na Figura 2.14 e nas Tabelas 2.2 e 2.3 foi possível observar que a média geral da CTC foi de $88,62 \text{ mmol}_c \text{ dm}^{-3}$ e desvio padrão de $19,87 \text{ mmol}_c \text{ dm}^{-3}$. Analisando individualmente as áreas estudadas, Guariba registrou a maior média ($108,10 \text{ mmol}_c \text{ dm}^{-3}$), tendendo ser a área com maior fertilidade do solo. Já Aparecida do Taboado foi o local que apresentou a menor fertilidade do solo ($\text{CTC} = 57,43 \text{ mmol}_c \text{ dm}^{-3}$). Com relação aos desvios padrões, foram observados que em dentro das áreas não houve variação da CTC, porém quando foi analisado as variações entre todas as áreas estudadas, foi observado um coeficiente de variação de 22,42%, registrando moderada variabilidade.

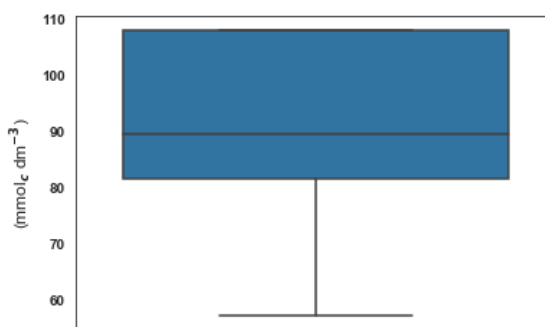


Figura 2.14: Boxplot da capacidade de troca de cátions (CTC, $\text{mmol}_c \text{ dm}^{-3}$) para todas as áreas estudadas, N = 24.

A outra variável analisada dentro da fertilidade do solo foi o teor de P. Na Figura 2.15 e nas Tabelas 2.2 e 2.3 observou-se que a média geral da P foi de $21,24 \text{ mg dm}^{-3}$ e desvio padrão de $9,04 \text{ mg dm}^{-3}$. Quando foi analisando cada área de estudo, foi observado os maiores valores em Pradópolis ($30,46 \text{ mg dm}^{-3}$), já em Aparecida do Taboado observaram os menores valores ($7,97 \text{ mg dm}^{-3}$), mostrando que se juntarmos as informações da CTC e do teor de P, podem concluir que Aparecida do Taboado foi o local que apresentou a menor fertilidade do solo, dentre todas as áreas estudadas. Devido ao tamanho das áreas usadas para a medição da FCO_2 , não foram observadas variações do teor de P, por isso que os desvios padrões foram 0. Porém quando foi analisado as variações entre todas as áreas estudadas, foi mensurado um coeficiente

de variação de 42,56%, registrando alta variabilidade entre as áreas estudadas, principalmente devido à grande amplitude dos valores observados.

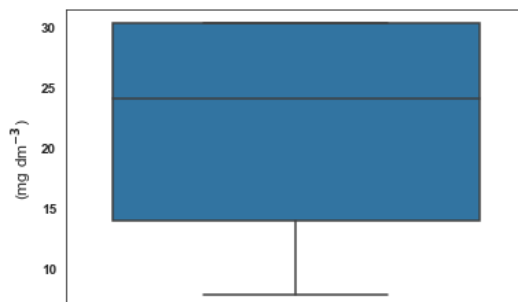


Figura 2.15: Boxplot dos teores de fósforo disponível (P, mg dm⁻³) para todas as áreas estudadas, N = 24.

Além dos efeitos da fertilidade do solo e da temperatura no processo da formação da matéria orgânica do solo, também foi observado o efeito da porosidade do solo, que na predição do modelo foi expressada pela Dp. Na Figura 2.16 e nas Tabelas 2.2 e 2.3 foram registrados a média geral da Dp (2,78 kg dm⁻³) e desvio padrão (0,36 kg dm⁻³). Para os locais estudados observaram que os menores valores foram registrados em Motuca (2,03 kg dm⁻³), já em Guariba observaram os maiores valores (3,18 kg dm⁻³), isso mostrou que Guariba tende a ter as maiores estruturas solidas, dentre as áreas estudadas, sendo uma característica relacionada aos minerais constituintes do solo. Observou-se que para o tamanho das áreas usadas para a medição da FCO₂, não foram registradas variações dos indices de Dp, fazendo com que os desvios padrões fossem 0. Entretanto, quando foi comparado as variações entre todas as áreas estudadas, foi obtido um coeficiente de variação de 13,11%, mostrando que ocorreu uma baixa variabilidade entre as áreas estudadas.

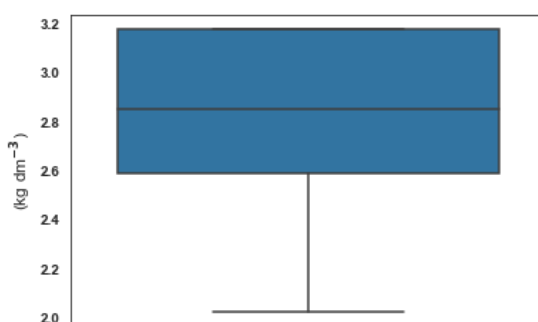


Figura 2.16: Boxplot da densidade das partículas (Dp, kg dm⁻³) para todas as áreas estudadas, N = 24.

Quanto a principal variável, a MOS, foram observadas na Figura 2.17 e nas Tabelas 2.2 e 2.3 que a média geral foi $30,67 \text{ g dm}^{-3}$ e o desvio padrão foi $11,46 \text{ g dm}^{-3}$. Quando foi analisado os locais estudados, foram registrados os menores valores em Aparecida do Taboado ($18,93 \text{ g dm}^{-3}$), já em Guariba observaram os maiores valores ($47,42 \text{ g dm}^{-3}$), mostrando que Aparecida do Taboado tende a ter um solo menos estruturado, dentre todas as áreas estudadas. Assim como as outras variáveis estudadas anteriormente, observou-se que para o tamanho das áreas usadas para a medição da FCO_2 , não foram registradas variações dos índices de MOS, fazendo com que os desvios padrões fossem 0. Porém, quando compararam as variações entre as áreas estudadas, foi mensurado um coeficiente de variação de 37,35%, mostrando que ocorreu uma alta variabilidade entre as áreas estudadas.

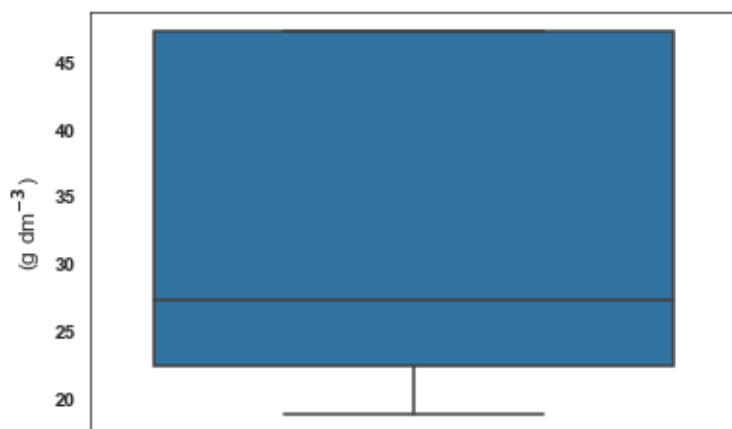


Figura 2.17: Boxplot da matéria orgânica do solo (MOS, g dm^{-3}) para todas as áreas estudadas, $N = 24$.

Geralmente as leituras da FCO_2 do solo são realizadas *in loco*, em períodos curtos e sequenciais de dias, geralmente no início do dia. Essas características, na sua maioria, resultam nas baixas amplitudes dos valores das variáveis físicas e químicas do solo, principalmente quando analisaram a T_s , P , CTC , Dp e a MOS. Outros pesquisadores estudaram o processo de FCO_2 e também observaram baixos valores de coeficientes de variação (Bicalho et al., 2017; Tavares et al., 2018).

2.4.5. MOS observado versus MOS estimado

Para realizar a avaliação do modelo do MOS, foi feita a relação entre os dados estimados a partir dos coeficientes obtidos anteriormente e dos valores da MOS

observados em campo. Para essa verificação foi utilizada a técnica da regressão linear. Conforme a Associação Brasileira de Normas Técnicas (2004), o poder de predição do modelo pode ser verificado por meio do gráfico da MOS observado, observado no eixo da abscissa, versus a MOS estimado pelo modelo no eixo da ordenada, que deve apresentar pontos próximos à bissetriz do primeiro quadrante, conforme a Equação 26:

$$MOS_{\text{estimado}} = 1,243e^{-14} + 1,000 MOS_{\text{observado}} \quad (26)$$

Em que:

MOS_{estimado} : matéria orgânica do solo (g dm^{-3}) estimado;

$MOS_{\text{observado}}$: matéria orgânica do solo (g dm^{-3}) observado.

Semelhante ao estudo da predição da MOS, o valor do coeficiente de determinação deste modelo foi $R^2 = 0,999$, e o valor do coeficiente de determinação ajustado foi $R_a^2 = 0,985$, ou seja, 98,5% das variações na estimativa da MOS podem ser explicadas pelas variações dos valores das variáveis independentes. O restante, 1,5% são fontes de variações que são explicados por fatores (Tabela 2.6).

Tabela 2.6: Coeficientes métricas para a determinação do modelo da MOS, N = 24.

Estatísticas	Coeficientes
R-squared	0,99
Adj. R-squared	0,985
F-statistic	2,088e ⁺³¹
Prob (F-statistic)	0
Log-Likelihood	736,23
AIC	-1468
BIC	-1466

R-squared = Coeficiente de Determinação; Adj. R-squared = Coeficiente de Determinação ajustado; F-statistic = valor do teste de Fischer – Snedecor; Prob (F-statistic) = p-valor do teste F; Log-Likelihood = probabilidade natural logarítmica; AIC = Critério de Informação de Akaike; BIC = Critério Bayesiano de Schwarz.

Para a verificação dos resultados obtidos na determinação da equação, foi realizado o teste F e confrontado o seu resultado com o p-valor, onde foram

encontrados os valores $2,088e^{+31}$ e 0, respectivamente (Tabela 2.6), mostrando que a variável independente apresentou relação com a MOS estimada.

Além disso, também foram realizados o teste de Akaike (AIC), cujo o valor obtido foi -1468. Já o valor encontrado pelo Critério Bayesiano de Schwarz (BIC) para descrever as interações entre as variáveis que maximizam a probabilidade da escolha do modelo verdadeiro, cujo o valor mensurado foi -1466 (Tabela 2.6). Para os dois parâmetros avaliativos foram observadas fortes interações da variável independente (MOS observada) com a variável dependente (MOS estimada).

Com relação aos coeficientes encontrados e parâmetros de qualidade da regressão linear, foi observado que o valor coeficiente linear foi igual a $-1,243e^{-14}$ e apresentou o p-valor = 0,096, indicando que este coeficiente não impactou significante na determinação do modelo ao nível de 5% (Tabela 2.7).

Tabela 2.7: Demonstrações do coeficiente linear e angular, resultados do p-valor em função da estatística T para o nível de significância $\alpha = 0,05$ com os seus respectivos valores críticos, para as variáveis MOS observados, versus MOS estimados.

	<i>Coeficientes</i>	<i>Erro Padrão</i>	<i>Resultado do teste t</i>	<i>p-valor</i>	<i>IC_{5%} Limite Inferior</i>	<i>IC_{95%} Limite Superior</i>
Constante	$1,243e^{-14}$	$7,15e^{-15}$	1,740	0,096 ^{NS}	$-2,39e^{-15}$	$2,73e^{-14}$
MOS_observado	1,0000	$2,19e^{-16}$	$4,57e^{+15}$	0,000*	1,000	1,000

Erro Padrão = Standard error ou erro padrão; Resultado teste t = resultado do teste de hipótese da média T Student; * = significativo no teste T a 5% de probabilidade; NS = não significativo no teste T a 5% de probabilidade; IC_{5%} Limite Inferior = Distribuição, razão de prevalência e intervalo de confiança de 5%; IC_{95%} Limite superior = Distribuição, razão de prevalência e intervalo de confiança de 95%; MOS_observado = matéria orgânica do solo observada ($g\ dm^{-3}$).

Quando foi estudado o coeficiente angular foi obtido o valor de 1,000 e p-valor de 0,000, onde pode-se concluir que o coeficiente angular apresentou uma relação significativa com o MOS previsto ao nível de 5% (Tabela 2.7), mostrando que o coeficiente angular interferiu na relação de como os valores estimados são fortemente influenciados pelos valores observados para esta série de dados.

Na Figura 2.18 foi representado graficamente a plotagem das MOS observadas versus MOS estimadas. Observa-se que o ângulo da reta, com relação ao ajuste linear foi 45 graus, caracterizando a forte relação entre a MOS observado e a MOS estimando. Além disso, também foram obtidos os valores da RMSE (*root mean squared error*) que significa a “raiz quadrática média dos erros ” e também a MSE (Mean squared error) que significa a “média quadrática dos erros”, foram encontrados os valores de 0 para os dois parâmetros, confirmando a existência das altas correlações entre os valores observados e estimados apresentaram.

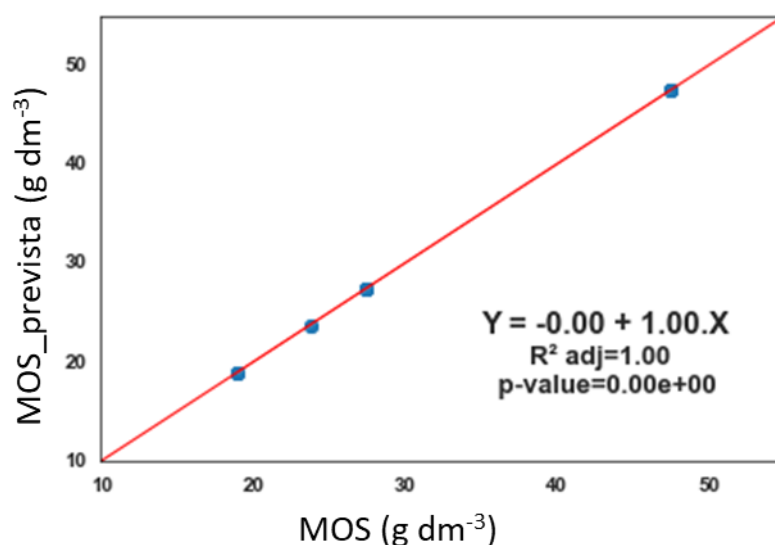


Figura 2.18: Matéria Orgânica de solo (MOS) observado versus estimado pela equação de regressão.

2.4.6. Regressão Polinomial Quadrática e curva de caracterização da CAD, em função da Umidade do Solo

Os atributos do solo são dependentes das propriedades químicas, físicas e microbiológicas do solo, que por sua vez, impactam diretamente na produtividade das culturas agrícolas e também na dinâmica dos gases no solo.

Estudos apontam que em solos tropicais as emissões de CO₂ são dependentes da umidade do solo (Panosso et al., 2008), que por sua vez, dependem da capacidade de armazenamento da água no solo (CAD). De maneira geral, a umidade do solo correlaciona-se positivamente com a variabilidade temporal e negativamente com a variabilidade espacial no estudo FCO₂ (La Scala et al., 2010).

Pereira et al. (1997) observaram que o estudo dos processos de acúmulo e das entradas e saídas de água no solo, bem como as determinações dos períodos de deficiência hídrica, podem ser estimados a partir da metodologia sugerida por Thornthwaite e Mather (1955), que usa como base a dinâmica do armazenamento de água no solo, utilizando somente os dados climáticos de temperatura do ar e precipitação. O balanço hídrico climatológico de Thornthwaite e Mather (1955) foi criado com o intuito de definir o regime hídrico local, não havendo a necessidade de utilizar as medidas diretas das características do solo.

Uma vez escolhido o local de estudo, são obtidos os valores da capacidade de água disponível (CAD) no solo, da precipitação (P) e da evapotranspiração de referência (ET_o) para um período estabelecido. A partir dessas informações, para o mesmo período, foram estimados o armazenamento da água no solo (ARM), a evapotranspiração real (ETR), deficiência hídrica (DEF) e o excedente hídrico (EXC). Nesta etapa do estudo foi estabelecida uma relação entre a umidade do solo (Us) para determinar qual a melhor a capacidade de água disponível (CAD) do solo. Diversos autores, em diferentes áreas de estudo observaram a relação da água no solo com a capacidade de água disponível (Souza et al., 2018; Barbosa et al., 2017).

Foram estimadas sete equações, utilizando a técnica da regressão polinomial quadrática. A variável dependente foi a umidade obtida pelo balanço hídrico de cada CAD e a variável independente foi a umidade do solo. Este estudo foi realizado com o intuito de estimar o melhor valor da CAD para a cultura da cana de açúcar, no estágio de brotação da cultura, em função da umidade do solo, visto que, as leituras observadas da FCO₂ foram analisadas nesse período. As equações foram geradas para as CAD de 100, 60, 45, 40, 35, 30, 20 mm. Na Tabela 2.8 e na Figura 2.19 são observadas as equações das regressões polinomiais quadráticas e as curvas das capacidades de água disponível (CAD) para 100, 60, 45, 40, 35, 30 e 20 mm, que foram mensuradas a partir da umidade do solo para as CADs, por meio das umidades do balanço hídrico obtidas pelo método de Thornthwaite e Mather (1955), em função da umidade do solo observada na análise temporal da FCO₂, para o sistema de produção da cultura da cana-de-açúcar.

Tabela 2.8. Modelos utilizados para a estimação da CAD ideal na fase inicial do desenvolvimento da cultura da cana-de-açúcar.

CADs (mm)	Equações (y)	R ²	R ² ajustado	Teste F	P-valor	AIC	BIC
CAD 100	$-0,0258x^2 + 1,1411x + 15,1876$	0,1939	0,117	2,525 ^{NS}	0,104	146,8	150,3
CAD 60	$-0,0252x^2 + 1,5054x - 0,0848$	0,69	0,66	23,37 ^{**}	4,56E-06	128,6	132,1
CAD 45	$-0,0311x^2 + 1,8717x - 7,3029$	0,7707	0,749	35,28 ^{**}	1,93E-07	129,6	133,1
CAD 40	$-0,0346x^2 + 2,0422x - 9,9035$	0,7718	0,75	35,51 ^{**}	1,83E-07	132,3	135,8
CAD 35	$-0,0453x^2 + 2,4259x - 13,7267$	0,7459	0,722	30,83 ^{**}	5,65E-07	137	140,5
CAD 30	$-0,0451x^2 + 2,4854x - 15,4068$	0,7341	0,709	28,99 ^{**}	9,10E-07	141,8	145,3
CAD 20	$-0,0644x^2 + 3,2113x - 21,9205$	0,6089	0,572	16,35 ^{**}	5,23E-05	160,1	163,6

CAD = capacidade de água disponível no solo, mm; R² = coeficiente de determinação; AIC = teste de Akaike; BIC = Critério Bayesiano de Schwarz; Teste F = teste de Fisher-Snedecor; NS = não foi significativo ao nível de 1% de probabilidade, pelo teste Fisher-Snedecor; ** significativo ao nível de 5% de probabilidade, pelo teste Fisher-Snedecor, x = coeficiente angular de ordem de 1, x² = coeficiente angular de ordem de 2.

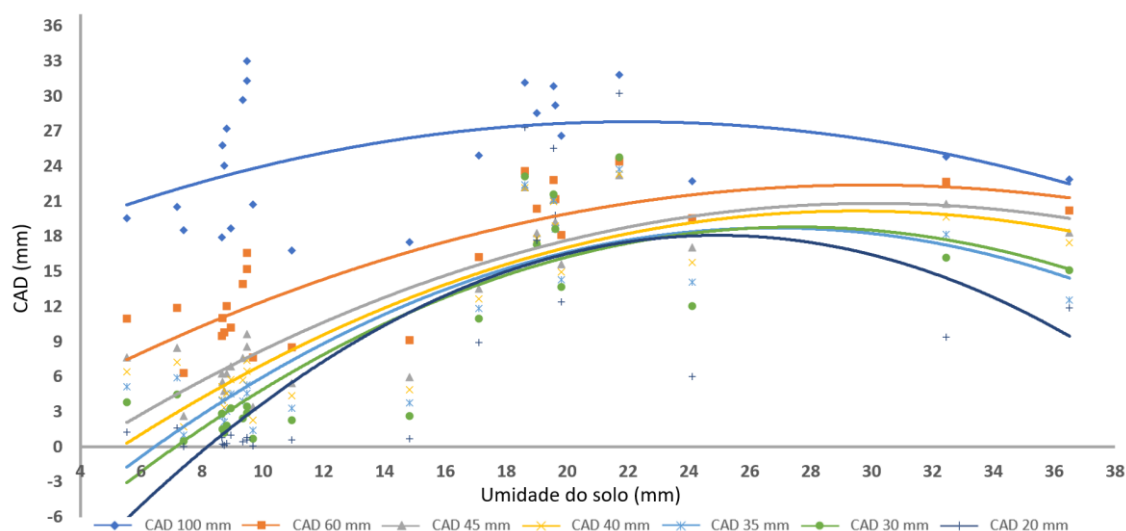


Figura 2.19: Determinações das curvas das regressões polinomiais quadráticas das capacidades de água disponível (CAD) para 100, 60, 45, 40, 35, 30 e 20 mm, em função da umidade do solo observada na análise temporal da emissão de CO₂ (FCO₂) e da umidade do solo obtida por meio do balanço hídrico pelo método de Thornthwaite e Mather (1955), na produção da cultura da cana-de-açúcar.

Quando foram analisadas as equações geradas, observou-se que apenas a regressão da CAD 100 mm não foi significativa, ao nível de 5% de probabilidade pelo teste de Fisher-Snedecor. O valor do coeficiente de determinação deste modelo foi $R^2 = 0,1939$, e o seu respectivo valor ajustado foi de $R_a^2 = 0,117$, ou seja, 11,7% das

variações da umidade do solo obtida por meio do balanço hídrico foram explicadas pelas variações dos valores da umidade do solo observada na análise temporal da FCO_2 . A grande maioria, 88,3% foram fontes de variações explicadas por outros fatores, mostrando que esse modelo possui um baixo poder de estimação. Com relação ao teste F e o p-valor foram obtidos os valores de 2,525 e 0,104, respectivamente (Tabela 2.8). Esse valor mostrou que as relações entre as umidades não foram significativas ao nível de 0,05. Além disso, o valor obtido pelo teste de Akaike (AIC) foi de 146,8 e o valor encontrado pelo Critério Bayesiano de Schwarz (BIC) foi 150,3. Quando foi analisado esses dois parâmetros, observou-se que a variável independente apresentou baixa relação com a variável dependente, mostrando que existiu uma fraca relação entre as variáveis estudadas.

Quando analisaram as equações estimadas para as CADs de 60, 45, 40, 35, 30 e 20 mm, todas as regressões foram significativas ao nível de 5% de probabilidade pelo teste de Fisher-Snedecor, sendo que os valores do teste F e p-valor foram: 23,37 e $4,56\text{E}^{-06}$; 35,28 e $1,93\text{E}^{-07}$; 35,51 e $1,83\text{E}^{-07}$; 30,83 e $5,65\text{E}^{-07}$; 28,99 e $9,10\text{E}^{-07}$; 16,35 e $5,23\text{E}^{-05}$, respectivamente. Com relação aos coeficientes de determinação (R^2) e o coeficientes de determinação ajustado (R_a^2) observaram os valores de 0,69 e 0,66; 0,7707 e 0,749; 0,7718 e 0,75; 0,7459 e 0,722; 0,7341 e 0,709; 0,6089 e 0,572, respectivamente. Os valores obtidos pelo teste de AIC foram: 128,6; 129,6; 132,3; 137; 141,8; 160,1. Para o Critério Bayesiano de Schwarz (BIC), os valores foram: 132,1; 133,1; 135,8; 140,5; 145,3 e 163,6, respectivamente. Para todas as CADs foi observado que a variável independente apresentou alta relação com a variável dependente, mostrando que existiu uma forte relação entre as variáveis estudadas (Tabela 2.8).

Como o modelo criado com a umidade da CAD 40 mm apresentou o melhor desempenho, em função do R_a^2 e do p-valor, a CAD 40 mm foi selecionada e será feito o estudo da equação do modelo. Outros autores também utilizaram a CAD 40 mm quando fizeram seus estudos no estágio inicial da brotação da cana-de-açúcar (Vieira et al., 2019; Sanches et al., 2018; Perin et al., 2019), visto que foi as leituras da FCO_2 e da umidade foram realizadas neste estágio de desenvolvimento. Na equação 27, segue o modelo polinomial quadrático que foi estimado para explicar as variações da umidade do solo, obtida por meio do balanço hídrico para a CAD 40 mm, em função

das variações dos valores da umidade do solo observada na análise temporal da FCO_2 .

$$Us \text{ para CAD } 40 \text{ mm} = -0,0346x^2 + 2,0422x - 9,9035 \quad (27)$$

Em que:

Umidade do solo CAD 40 mm = Umidade do solo (mm) observada a partir da CAD 40 mm;

x = coeficiente angular de ordem de 1;

x^2 = coeficiente angular de ordem de 2.

O coeficiente de determinação do modelo foi de $R^2 = 0,772$, e o seu respectivo valor ajustado foi $R_a^2 = 0,750$, ou seja, a regressão criada com a umidade da CAD 40 mm conseguiu explicar as relações entre as variáveis dependentes e independentes, explicando 75% das interações, o restante, 25%, são decorrentes de outros fatores que não foram explicadas pelo modelo (Tabela 2.9).

Tabela 2.9: Coeficientes métricos do modelo da umidade do solo para a CAD de 40 mm, N = 24.

Estatísticas	Coeficientes
R-squared	0,772
Adj. R-squared	0,750
F-statistic	35,51
Prob (F-statistic)	1,83e ⁻⁰⁷

R-squared = Coeficiente de Determinação; Adj. R-squared = Coeficiente de Determinação ajustado; F-statistic = valor do teste de Fischer – Snedecor; Prob (F-statistic) = p-valor do teste F.

Para analisar a independência dos coeficientes do modelo foi realizada o teste F, onde foi verificado se $H_0: \beta_1 = \beta_2 = \beta_3 = 0$ ou se H_1 : se algum $\beta_1, \beta_2, \beta_3$ é diferente de 0. Os valores do teste F e do p-valor foram 35,51 e 1,87e⁻⁰⁷, respectivamente (Tabela 2.9). A partir dos valores obtidos tem-se a rejeição da hipótese nula ao nível de significância de 0,05, implicando em dizer que existiu uma relação entre a umidade do solo obtida a partir do balanço hídrico da CAD de 40 mm e a umidade do solo observada na análise temporal da FCO_2 .

O valor do coeficiente linear do modelo foi -9,9035 e apresentou significância estatística ao nível de significância 5% (p - valor = 0,006), mostrando que esse

coeficiente impactou diretamente no modelo (Tabela 2.10). O valor negativo do coeficiente linear mostrou que o solo tende a ter baixos valores ou não ter umidade, sendo bastante coerente com relação ao processo de coleta dos dados, visto que as leituras foram realizadas à uma profundidade de 0 a 10 cm e com solos que apresentaram baixas umidades.

Tabela 2.10: Demonstrações dos coeficientes lineares e angulares, resultados do p-valor em função da estatística t para o nível de significância $\alpha = 0,05$ com os seus respectivos valores críticos, para a umidade do solo observada na análise temporal da emissão de CO₂ (FCO₂), que foi usada na determinação da umidade do solo obtida, a partir do balanço hídrico da CAD de 40 mm, N = 24.

	<i>Coeficientes</i>	<i>Erro Padrão</i>	<i>Resultado do teste t</i>	<i>p-valor</i>	<i>IC_{5%} Limite Inferior</i>	<i>IC_{95%} Limite Superior</i>
Constante	-9,9035	3,259	-3,038	0,006*	-16,682	-3,125
X	2,0422	0,394	5,181	0,000*	1,222	2,862
x²	-0,0346	0,010	-3,493	0,002*	-0,055	-0,014

Erro Padrão = Standard error ou erro padrão; Resultado teste t = resultado do teste de hipótese da média T Student; * = significativo no teste T a 5% de probabilidade; NS = não significativo no teste T a 5% de probabilidade; IC_{5%} Limite Inferior = Distribuição, razão de prevalência e intervalo de confiança de 5%; IC_{95%} Limite superior = Distribuição, razão de prevalência e intervalo de confiança de 95%; x = coeficiente angular de ordem de 1, x² = coeficiente angular de ordem de 2.

Quando foi analisado o coeficiente angular de ordem de 1 (x) foi obtido o valor de 2,0422 e o seu p-valor foi de 0,000. Já o valor do coeficiente angular de ordem 2 (x²) foi -0,0346 e seu p-valor foi de 0,002 (Tabela 2.10). Em função dessas informações pode ser dito que os dois coeficientes apresentaram forte poder explicativo no modelo, mostrando a forte relação existente entre a umidade do solo observada na análise temporal da FCO₂ e a umidade do solo obtida a partir do balanço hídrico da CAD de 40 mm. Outros autores utilizaram a umidade do solo para obter valores de umidade por meio do balanço hídrico pelo método em diversas áreas de estudo (Mammoliti et al., 2021; Al-Sudani et al., 2020; Galvão et al., 2018).

Para analisar a variabilidade das variáveis utilizadas no modelo, foram usadas a estatística descritiva e a técnica de boxplot. Na figura 2.20 observa-se o boxplot e a variação temporal da umidade do solo observada na análise temporal da FCO₂ e da umidade do solo obtida a partir do balanço hídrico da CAD de 40 mm. De maneira geral, a umidade do solo observada na análise temporal da FCO₂ apresentou uma maior amplitude, quando comparado com a umidade do solo obtida a partir do balanço hídrico da CAD de 40 mm.

A variação dos valores encontrados da umidade do solo observada na análise temporal da FCO₂ foi entre 5,53 e 36,5 mm, sendo que o maior valor observado ocorreu em Motuca e o menor em Aparecida do Taboado. O valor médio geral da umidade do solo observada na análise temporal da FCO₂ foi de 14,83 mm, Pradópolis foi a que apresentou a menor média (8,94 mm) e Motuca foi a maior (31,02 mm), dentre as áreas estudadas. Já o valor encontrado do desvio padrão geral foi de 8,17 mm. Motuca também apresentou o maior desvio padrão (6,32 mm) e Pradópolis o menor (0,73 mm). Com relação aos valores de coeficiente de variação, quando foram analisados todos os dados das áreas estudadas, foi encontrado o valor de 55,05%, indicando uma grande amplitude dos dados para essa variável. Porém quando foi realizado o estudo de cada local, foi observado que Aparecida do Taboado apresentou o maior valor (34,65%) e alta dispersão de dados, seguido de Motuca (20,36%) que apresentou moderada dispersão e por fim Pradópolis (8,18%) e Guariba (7,18%) que foram considerados com baixa dispersão dos dados (Tabela 2.11, Tabela 2.12, Figura 2.20 e Figura 2.21).

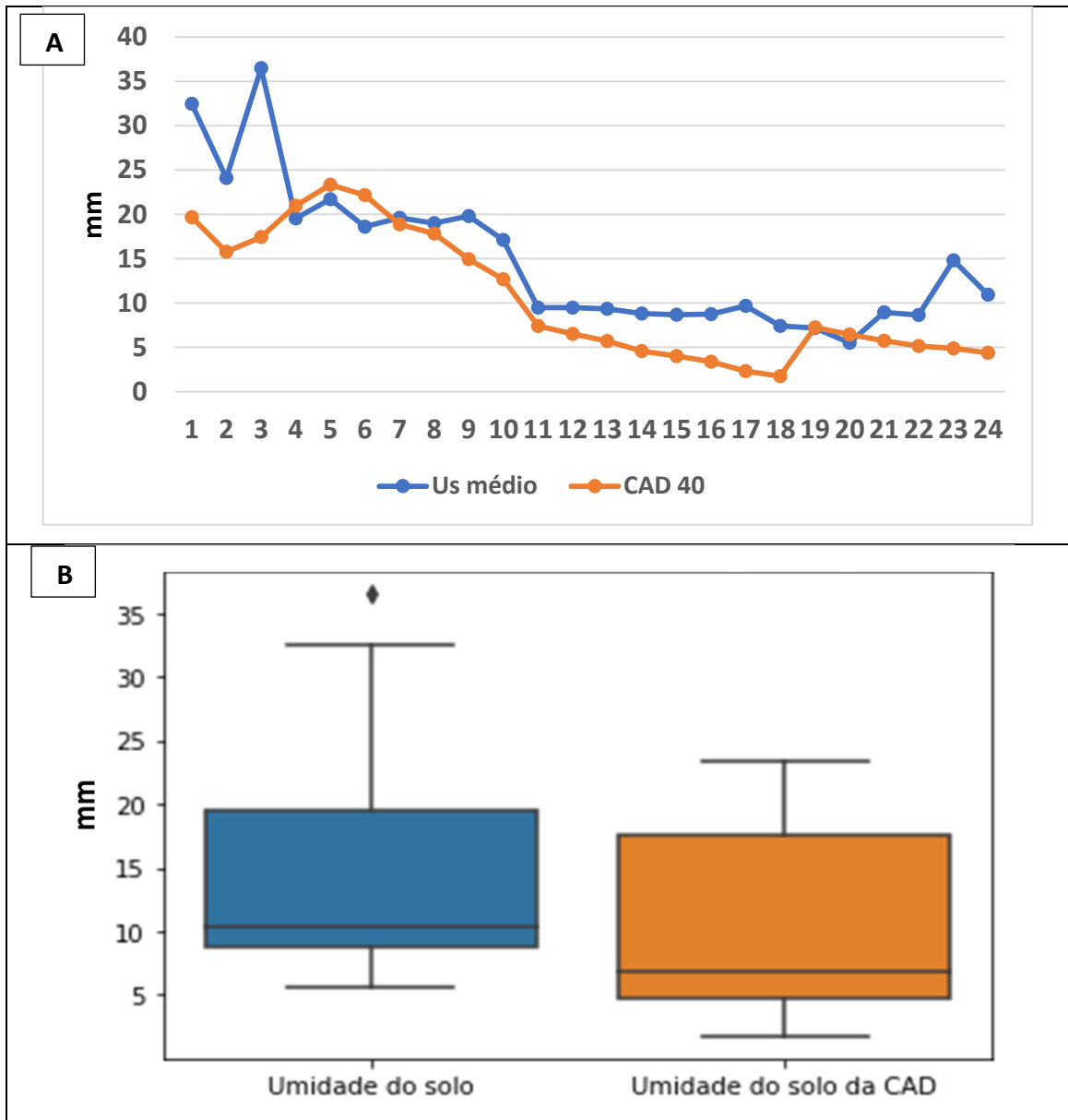


Figura 2.20: Distribuição espacial (A) e o boxplot da umidade do solo observada (B) na análise temporal da emissão de CO_2 (FCO_2) e da umidade do solo obtida a partir do balanço hídrico da CAD de 40 mm, para todas as áreas estudadas, $N = 24$.

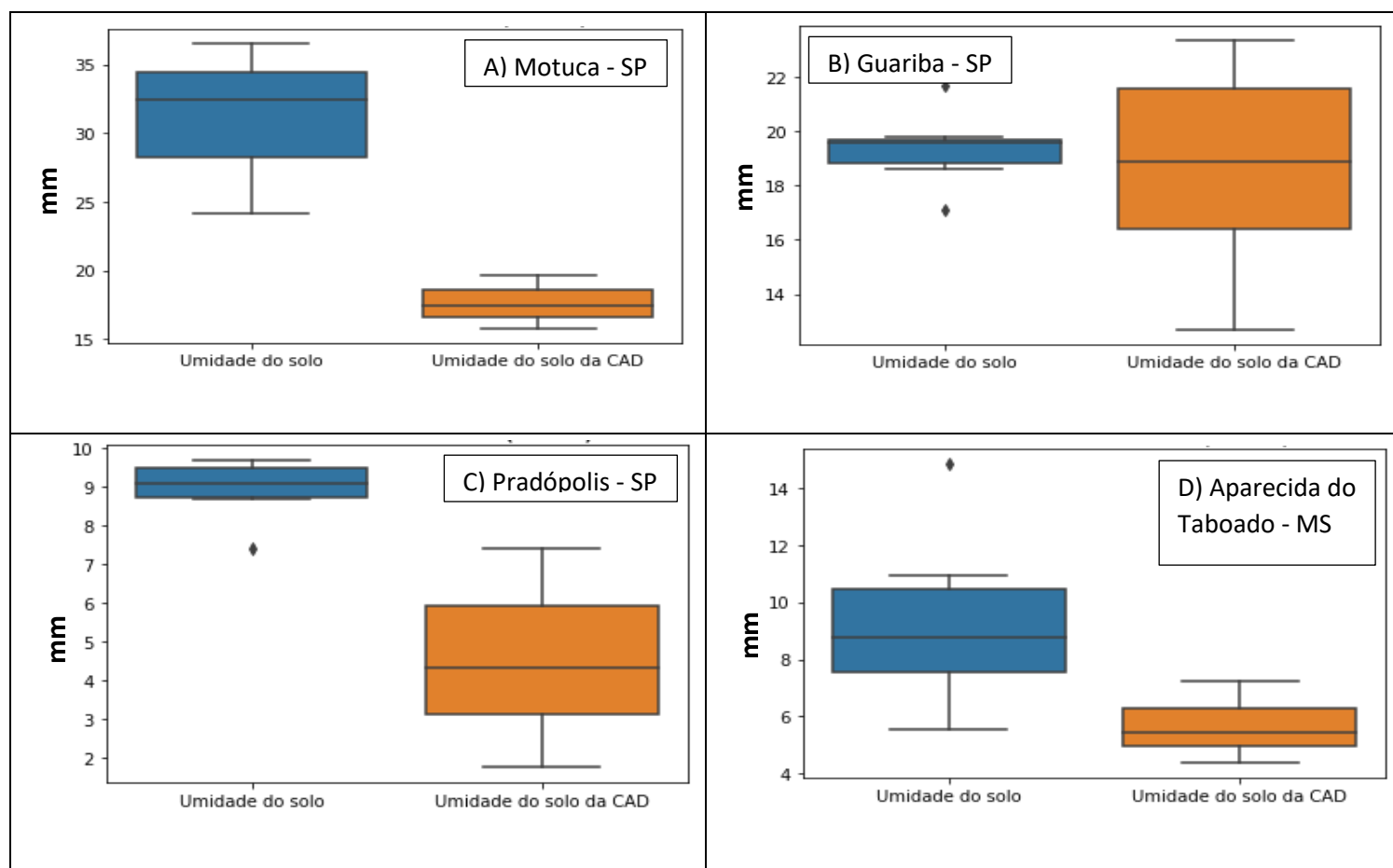


Figura 2.21: Boxplots da umidade do solo observada na análise temporal da emissão de CO_2 (FCO_2) e da umidade do solo obtida a partir do balanço hídrico da CAD de 40 mm para os municípios A) Motuca – SP, B) Guariba – SP, C) Pradópolis – SP e D) Aparecida do Taboado – MS.

Tabela 2.11: Estatística descritiva da umidade do solo observada na análise temporal da emissão de CO_2 (FCO_2) e da umidade do solo obtida a partir do balanço hídrico da CAD de 40 mm para todas as áreas estudadas, $N = 24$.

<i>Atributos</i>	<i>Média</i>	<i>Dp</i>	<i>Máx / Mín</i>	<i>CV</i>
<i>Us</i>	14,83	8,17	36,5 / 5,53	55,05
<i>Us CAD 40 mm</i>	18,18	7,18	23,33 / 1,75	39,52

Dp = desvio padrão; Máx = Máximo; Mín = Mínimo; CV = coeficiente de variação (%); Us = umidade do solo observada na análise temporal da emissão de CO_2 (FCO_2); Us CAD 40 mm = umidade do solo obtida a partir do balanço hídrico da CAD de 40 mm.

Tabela 2.12: Estatística descritiva da umidade do solo observada na análise temporal da emissão de CO₂ (FCO₂) e da umidade do solo obtida a partir do balanço hídrico da CAD de 40 mm, referentes às áreas de Motuca – SP (N = 3), Guariba – SP (N = 7), Pradópolis – SP (N = 8) e Aparecida do Taboado (N = 6).

Motuca					Guariba			
Atributos	Média	Dp	Máx / Min	CV	Média	Dp	Máx / Min	CV
Us	31,02	6,32	36,50 / 24,11	20,36	19,34	1,39	21,70 / 17,10	7,18
Us CAD 40	17,63	1,96	19,67 / 15,77	11,12	18,69	3,87	23,33 / 12,69	20,70
Pradópolis					Aparecida do Taboado			
Us	8,94	0,73	9,67 / 7,40	8,18	9,35	3,24	10,94 / 5,53	34,65
Us CAD 40	4,45	1,99	7,41 / 1,75	44,69	5,65	1,06	7,25 / 4,39	18,80

Dp = desvio padrão; Máx = Máximo; Mín = Mínimo; CV = coeficiente de variação (%); Us = umidade do solo observada na análise temporal da emissão de CO₂ (FCO₂); Us CAD 40 mm = umidade do solo obtida a partir do balanço hídrico da CAD de 40 mm.

Os gráficos das Figuras 2.20 e 2.21 e as Tabelas 2.11 e 2.12 indicaram as distribuições dos dados da umidade do solo obtida a partir do balanço hídrico da CAD de 40 mm. A média geral da umidade do solo obtida a partir do balanço hídrico da CAD de 40 mm foi 18,18 mm e apresentou desvio padrão de 7,18 mm. Quando foram avaliados os locais estudados, Pradópolis indicou a menor média (4,45 mm) e Aparecida do Taboado registrou o menor desvio padrão (1,06 mm). O município de Guariba apresentou a maior média de umidade (18,69 mm) e maior desvio padrão (3,87 mm) dentre todas as áreas estudadas. Com relação a amplitude geral dos dados, a variação foi de 1,75 a 23,33 mm. Com relação aos valores de coeficiente de variação, foram observados que para todos os dados das áreas estudadas o valor de 39,52%, mostrando uma grande amplitude para a variável. Quando foi realizado o estudo de cada local, Motuca registrou baixa dispersão dos dados (11,12%), Aparecida do Taboado (18,80%) e Guariba (20,70%) apresentou moderada dispersão e Pradópolis foi considerado como alta dispersão (44,69%).

2.4.7. Umidade para a CAD de 40 mm observado versus umidade para a CAD de 40 mm estimada

Para a avaliação da predição da umidade para a CAD de 40 mm foi feito a relação entre os dados estimados, a partir dos coeficientes obtidos anteriormente e os valores da umidade de CAD 40 mm observados em campo. Para essa verificação foi

utilizada a técnica da regressão linear. O poder de predição do modelo foi verificado por meio do gráfico dos dados observados no eixo da abscissa, versus os dados estimados pelo modelo na ordenada, que deve apresentar pontos próximos à bissetriz do primeiro quadrante, conforme a Equação 28:

$$Umidade\ do\ solo\ CAD\ 40\ mm_{estimado} = 2,408 + 0,772\ Umidade\ do\ solo\ CAD\ 40\ mm_{observado} \quad (28)$$

Em que:

$Umidade\ do\ solo\ CAD\ 40\ mm_{estimado}$: Umidade do solo (mm) estimada a partir da CAD 40 mm;

$Umidade\ do\ solo\ CAD\ 40\ mm_{observado}$: Umidade do solo (mm) observado a partir da CAD 40 mm.

O valor do coeficiente de determinação deste modelo foi $R^2 = 0,77$, e o valor do coeficiente de determinação ajustado foi $R_a^2 = 0,761$, ou seja, 76,1% das variações da unidade do solo estimada para a CAD 40 mm podem ser explicadas pelas variações dos valores das variáveis independentes. O restante, 23,9% são fontes de variações que são explicados por outros fatores (Tabela 2.13).

Tabela 2.13: Coeficientes métricas para a determinação do modelo da umidade do solo para a CAD 40 mm, N = 24.

Estatísticas	Coeficientes
R-squared	0,772
Adj. R-squared	0,761
F-statistic	74,40
Prob (F-statistic)	1,65e ⁻⁰⁸
Log-Likelihood	-60,024
AIC	124,0
BIC	126,4

R-squared = Coeficiente de Determinação; Adj. R-squared = Coeficiente de Determinação ajustado; F-statistic = valor do teste de Fischer – Snedecor; Prob (F-statistic) = p-valor do teste F; Log-Likelihood = probabilidade natural logarítmica; AIC = Critério de Informação de Akaike; BIC = Critério Bayesiano de Schwarz.

Para a verificação dos resultados obtidos na determinação da equação 28 foi realizado o teste F e obtido o p-valor, cujo os valores encontrados foram 74,40 e 1,65e⁻

⁰⁸, respectivamente, mostrando que a variável independente apresentou relação com a umidade do solo estimada a partir da CAD 40 mm. Além disso, também foi realizado o teste de Akaike (AIC), cujo o valor obtido foi 124. Já o valor encontrado pelo Critério Bayesiano de Schwarz (BIC) para descrever as interações entre as variáveis que maximizam a probabilidade da escolha do modelo verdadeiro, foi -126,4 (Tabela 2.13). Para os dois parâmetros avaliativos foram observadas fortes interações da variável independente com a variável dependente, evidenciando a forte interação da umidade do solo observada a partir da CAD 40 mm com a umidade do solo estimada a partir da CAD 40 mm.

Analizando os parâmetros de qualidade da regressão linear, foi observado que o coeficiente linear foi igual a 2,0480 e apresentou o p-valor = 0,045. Para o coeficiente angular foi encontrado o valor de 0,7718 e p-valor de 0,000, indicando que os dois coeficientes afetaram significante ao nível de 5% na determinação do modelo, respectivamente (Tabela 2.14).

Tabela 2.14: Demonstrações do coeficiente linear e angular, resultados do p-valor em função da estatística t para o nível de significância $\alpha = 0,05$ com os seus respectivos valores críticos, para as variáveis umidade do solo observada a partir da CAD 40 mm, versus umidade do solo estimada a partir da CAD 40 mm.

	<i>Coeficientes</i>	<i>Erro Padrão</i>	<i>Resultado do teste t</i>	<i>p-valor</i>	<i>IC_{5%} Limite Inferior</i>	<i>IC_{95%} Limite Superior</i>
Constante	2,0480	1,134	2,123*	0,045*	0,055	4,761
Umidade do solo (mm) observada a partir da CAD 40 (mm)	0,7718	0,089	8,625*	0,000*	0,586	0,957

Erro Padrão = Standard error ou erro padrão; Resultado teste t = resultado do teste de hipótese da média T Student; * = significativo no teste T a 5% de probabilidade; NS = não significativo no teste T a 5% de probabilidade; IC_{5%} Limite Inferior = Distribuição, razão de prevalência e intervalo de confiança de 5%; IC_{95%} Limite superior = Distribuição, razão de prevalência e intervalo de confiança de 95%; Umidade do solo (mm) observada a partir da CAD 40 (mm).

O valor positivo do coeficiente linear mostrou que a constante, inicialmente, inicia-se com uma umidade de 2,0480 mm e por meio do coeficiente angular, tende

que a cada unidade acrescida, tem a adição de 0,7718 mm, mostrando que a determinação do valor estimado é fortemente influenciada pelos valores observados (Tabela 2.14).

Na Figura 2.22 foi representado graficamente a plotagem da umidade do solo observada a partir da CAD 40 mm, versus umidade do solo estimada a partir da CAD 40 mm. Observa-se que o ângulo da reta com relação ao ajuste linear foi próximo de 45 graus, caracterizando uma forte interação entre as variáveis. Além disso, também foram obtidos os valores da RMSE e a MSE, onde encontraram os valores de 11,28 e 3,36, respectivamente, indicando que os valores observados e estimados apresentaram altas correlações.

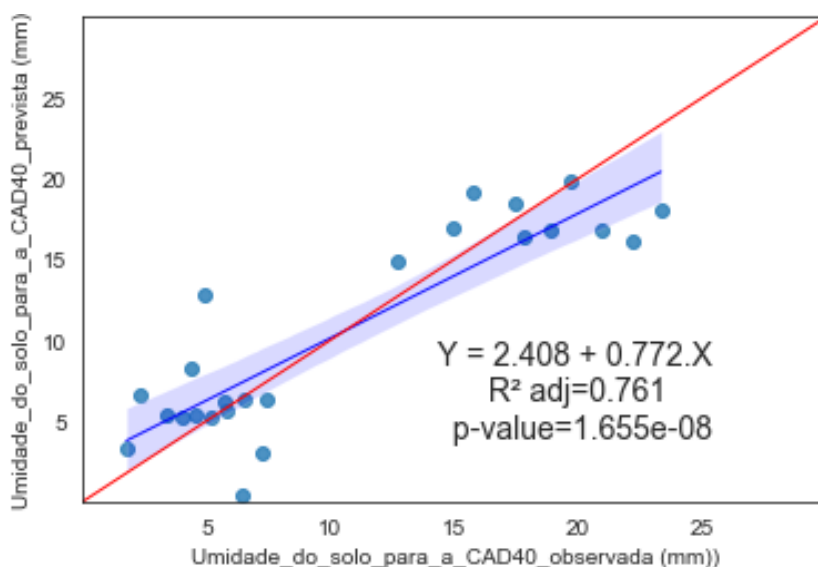


Figura 2.22: Umidade do solo observada a partir da CAD 40 mm versus umidade do solo estimada a partir da CAD 40 mm pela equação de regressão.

2.4.8. Regressão Linear Múltipla da CAD, em função das variáveis Us, Ds, Dp e ARM

A equação de regressão linear múltipla que foi estimada para explicar a CAD que foi calculada para 40 mm, em função das variáveis Us, Ds, Dp e ARM foi representada pela Equação 29:

$$CAD\ 40\ mm = 40,0035 - 0,0001Us - 0,0003Ds - 5,72e^{-05}Dp + 2,37e^{-05}ARM \quad (29)$$

Em que:

CAD 40 mm: capacidade de água disponível no solo de 40 mm;

Us: umidade do solo (mm);

Ds: densidade do solo (kg dm^{-3});

Dp: densidade das partículas (kg dm^{-3});

ARM: armazenamento de água no solo (mm).

O valor do coeficiente de determinação deste modelo foi $R^2 = 0,999$ e o seu respectivo valor ajustado foi $R_a^2 = 0,985$, ou seja, 98,5% das variações na estimativa da CAD para 40 mm podem ser explicadas pelas variações das variáveis independentes. O restante, 1,5% são fontes de variações que são explicados por outros fatores (Tabela 2.15).

Tabela 2.15: Coeficientes métricos do modelo da CAD, N = 24.

Estatísticas	Coeficientes
R-squared	0,999
Adj. R-squared	0,985
F-statistic	5677
Prob (F-statistic)	$6,05e^{-29}$
Log-Likelihood	225,60
AIC	-441,2
BIC	-435,3

R-squared = Coeficiente de Determinação; Adj. R-squared = Coeficiente de Determinação ajustado; F-statistic = valor do teste de Fischer – Snedecor; Prob (F-statistic) = p-valor do teste F; Log-Likelihood = probabilidade natural logarítmica; AIC = Critério de Informação de Akaike; BIC = Critério Bayesiano de Schwarz.

Para a verificar os resultados obtidos na determinação da equação 29 foi realizado o teste F e obtido o p-valor, visando verificar se $H_0: \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = 0$ ou H_1 : que pelo menos um β é diferente dos demais os valores, cujo os valores encontrados foram 5677 e $6,05e^{-29}$, respectivamente, mostrando que a variável independente apresentou relação com a CAD estimada, a partir da Us, Ds, Dp e ARM. Além disso, também foi realizado o teste de Akaike (AIC), cujo o valor obtido foi -441,2. Já o valor encontrado pelo Critério Bayesiano de Schwarz (BIC) para descrever as interações entre as variáveis que maximizam a probabilidade da escolha do modelo verdadeiro, cujo o valor mensurado foi -435,3 (Tabela 2.15). Para os dois parâmetros avaliativos foram observadas fortes interações da variável independente com a

variável dependente. Os dois parâmetros mostraram fortes interações entre as variáveis independentes e a dependente, indicando uma alta relação da Us, Ds, DP e ARM com CAD.

Quando se analisou o valor do coeficiente linear do modelo da CAD prevista, obtiveram o valor de 40,0035 que foi significativo ao nível de 5% pelo teste t, apresentando relevância no modelo (p-valor = 0,000, Tabela 2.16).

Tabela 2.16: Demonstrações dos coeficientes lineares e angulares, resultados do p-valor em função da estatística t para o nível de significância $\alpha = 0,05$ com os seus respectivos valores críticos, para as variáveis Us, Ds, DP e ARM que foram usadas na determinação do modelo da CAD, N = 24.

	<i>Coeficientes</i>	<i>Erro Padrão</i>	<i>Resultado do teste t</i>	<i>p-valor</i>	<i>IC_{5%} Limite Inferior</i>	<i>IC_{95%} Limite Superior</i>
Constante	40,0035	0,000	1,25e ⁺⁰⁵	0,000*	40,003	40,004
Us	-0,0001	5,16e ⁻⁰⁶	-24,515	0,000*	-0,000	-0,000
Ds	-0,0003	0,000	-0,833	0,415 ^{NS}	-0,001	0,000
Dp	-5,72e ⁻⁰⁵	6,27e ⁻⁰⁵	-0,913	0,373 ^{NS}	-0,000	7,4e ⁻⁰⁵
ARM	2,37e ⁻⁰⁵	6,38e ⁻⁰⁶	3,715	0,001*	1,03e ⁻⁰⁵	3,71e ⁻⁰⁵

Erro Padrão = Standard error ou erro padrão; Resultado teste t = resultado do teste de hipótese da média T Student; * = significativo no teste T a 5% de probabilidade; NS = não significativo no teste T a 5% de probabilidade; IC_{5%} Limite Inferior = Distribuição, razão de prevalência e intervalo de confiança de 5%; IC_{95%} Limite superior = Distribuição, razão de prevalência e intervalo de confiança de 95%; Us = umidade do solo (mm); Ds = densidade do solo (kg dm⁻³); Dp = densidade das partículas (kg dm⁻³); ARM = armazenamento de água no solo (mm).

O valor positivo e próximo de 40 indicou que esse modelo conseguiu prever com precisão a CAD inicial da cultura da cana-de-açúcar no início da brotação, visto que as leituras da FCO₂ e a coleta dos demais dados foram realizadas neste estágio de desenvolvimento (Tabela 2.16).

Com relação aos coeficientes angulares, foram adquiridos os seguintes resultados para as variáveis dependentes: Us = -0,0001, Ds = -0,0003, Dp = -5,72e⁻

⁰⁵ e $ARM = 2,37e^{-05}$. Os p-valores foram 0,000; 0,415; 0,373 e 0,001, respectivamente (Tabela 2.16). A Us e o ARM apresentaram forte poder explicativo no modelo, no entanto, a Ds e a Dp não apresentaram níveis significativos que conseguissem impactar no modelo ao nível de significância de 5%. Porém para outro conjunto de dados pode ser observada uma maior relação dessas variáveis.

Para analisar a variabilidade das variáveis utilizadas no modelo, foram usadas a estatística descritiva e a técnica de boxplot. Essas técnicas foram utilizadas nos estudos das variáveis Us, Ds, Dp, ARM.

Quando se avaliou a Us, foram observadas nas Figuras 2.23 e 2.24 e nas Tabelas 2.2 e 2.3 que a média geral foi de 14,27 mm e desvio padrão de 6.16 mm. Quando analisaram os locais estudados foram registrados os menores valores em Aparecida do Taboado (5,03 mm), já em Motuca os maiores valores (25,03 mm), mostrando que Motuca tende a ter um solo mais úmido, dentre todas as áreas estudadas. Quando foram analisados os valores dos coeficientes de variação, observaram uma alta variabilidade para todos os dados das áreas estudadas, sendo obtido o valor de 43,16%. Quando estudaram cada local foi observado que Motuca (4,42%) e Guariba (11,05%) apresentaram baixas dispersões dos dados, já Pradópolis (16,76%) e Aparecida do Taboado (19,55%) apresentaram moderadas dispersões.

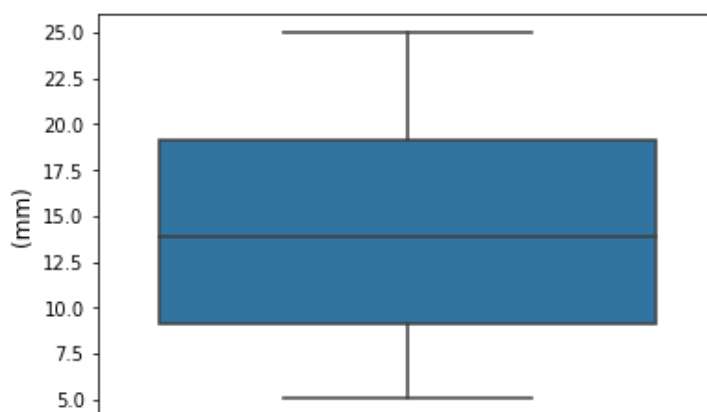


Figura 2.23: Boxplot da umidade do solo (Us, mm) para todas as áreas estudadas, N = 24.

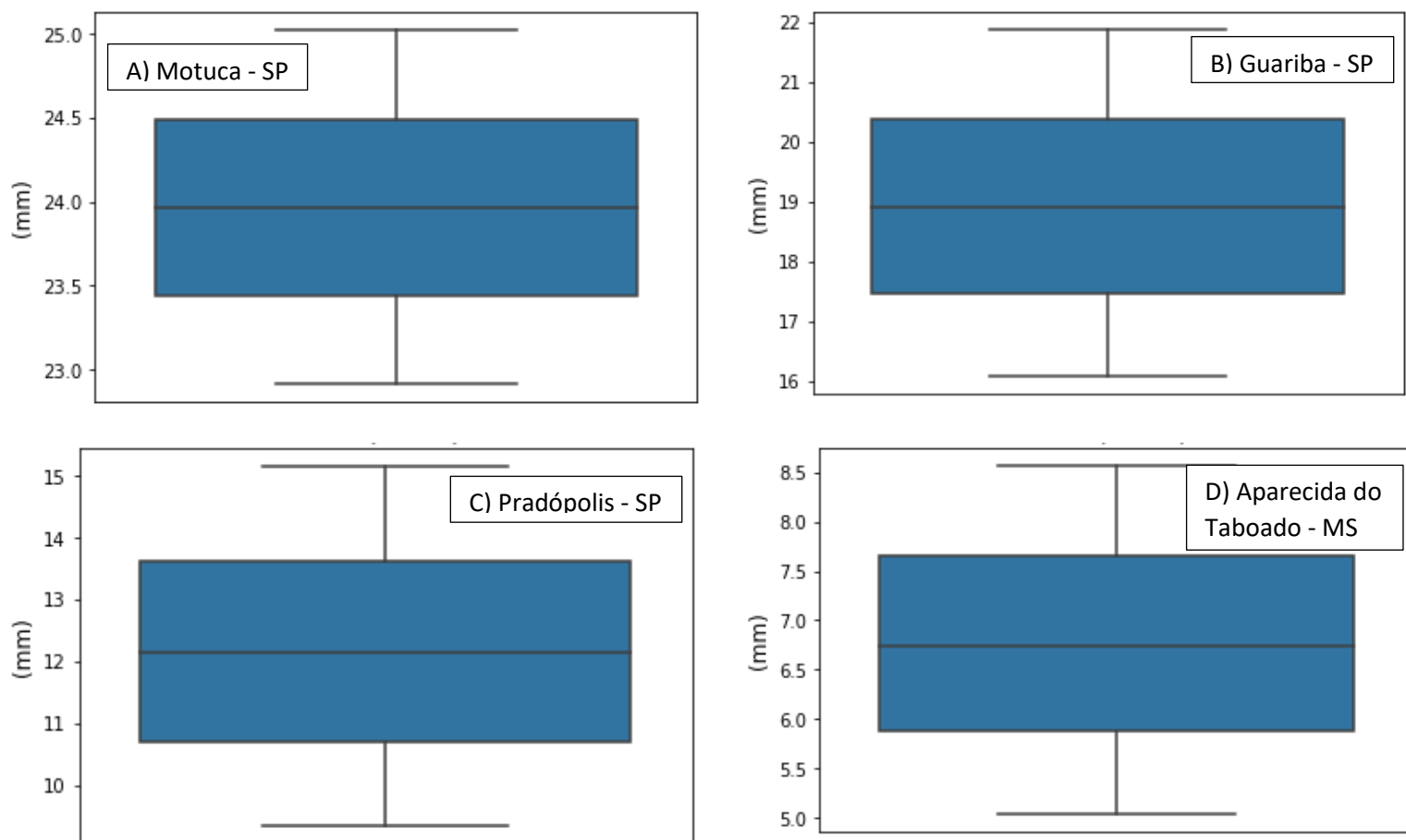


Figura 2.24: Boxplots das umidades do solo (U_s , mm) para os municípios A) Motuca – SP, B) Guariba – SP, C) Pradópolis – SP e D) Aparecida do Taboado – MS.

Um dos principais elementos que interferem na umidade, porosidade do solo e na estruturação do solo também foi considerada na predição do modelo, a D_s . Na Figura 2.25 e nas Tabelas 2.2 e 2.3 foram exibidas a média geral da D_s que foi $1,42 \text{ kg dm}^{-3}$ e desvio padrão de $0,10 \text{ kg dm}^{-3}$. Para cada local de estudo, observaram que os menores valores foram registrados em Motuca ($1,16 \text{ kg dm}^{-3}$) e em Guariba observaram os maiores valores ($1,48 \text{ kg dm}^{-3}$), isso mostrou que Guariba tende a ter um solo mais estruturado dentre as áreas estudadas. Assim como algumas outras variáveis estudadas anteriormente, observou-se que para o tamanho das áreas usadas para a medição da FCO_2 , não foram registradas variações nos índices das D_s dentro de cada área estudada, fazendo com que os desvios padrões fossem 0. Entretanto, quando foi comparado as variações entre todas as áreas estudadas, foi obtido um

coeficiente de variação de 7,22%, mostrando que ocorreu uma baixa variabilidade entre os locais de estudo.

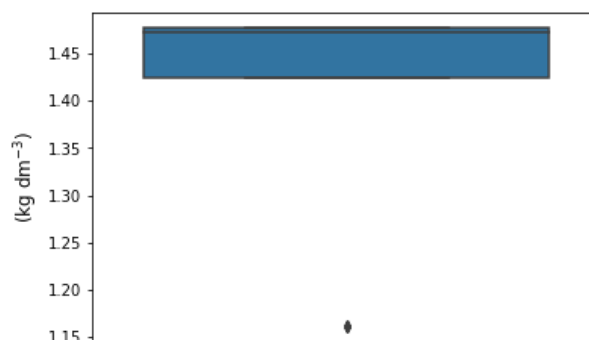


Figura 2.25: Boxplot da densidade do solo (D_s , kg dm^{-3}) para todas as áreas estudadas, $N = 24$.

A D_p também foi estudada, visto que é impactada pela porosidade e umidade do solo. Na Figura 2.26 e nas Tabelas 2.2 e 2.3 foram registrados a média geral da D_p , que foi 2,78 kg dm^{-3} e desvio padrão 0,36 kg dm^{-3} . Para cada local de estudo, observaram que os menores valores foram registrados em Motuca (2,03 kg dm^{-3}), já em Guariba observaram os maiores valores (3,18 kg dm^{-3}), isso mostrou que Guariba tende a ter as maiores estruturas sólidas do solo, dentre todas as áreas estudadas. Assim como as outras variáveis estudadas anteriormente, observou-se que para o tamanho das áreas usadas para a medição da FCO_2 , não foram registradas variações nos índices de D_p dentro das áreas estudadas, fazendo com que os desvios padrões fossem 0. Entretanto, quando foi comparado as variações entre os locais, foi obtido um coeficiente de variação de 13,11%, mostrando que ocorreu uma baixa variabilidade entre as áreas estudadas.

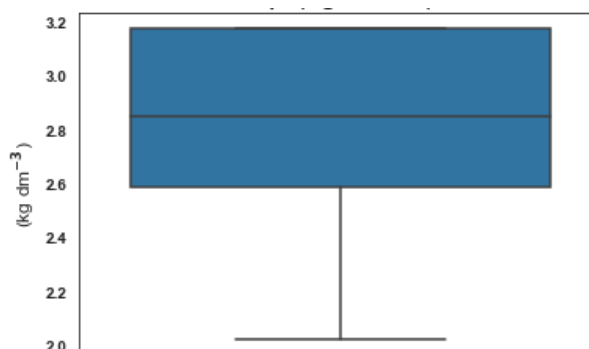


Figura 2.26: Boxplot da densidade das partículas (D_p , kg dm^{-3}) para todas as áreas estudadas, $N = 24$.

Outra variável utilizada para a predição foi a ARM. Nas Figuras 2.27 e 2.28 e nas Tabelas 2.2 e 2.3 foram registrados a média geral da ARM, que foi 4,22 mm e desvio padrão 2,87 mm. Estudando cada local de estudo, observaram que os menores valores foram obtidos em Pradópolis (0,69 mm), já em Guariba observaram os maiores valores (9,33 mm), isso mostrou que Guariba tende a armazenar mais água no solo, dentre todas as áreas estudadas. Quando analisaram os valores dos coeficientes de variação, foram observados uma alta variabilidade para todos os dados nas áreas estudadas, apresentando o valor de 68,07%. Quando foi estudado cada local, observaram que Pradópolis apresentou o maior valor (44,69%) e alta dispersão dos dados, seguido de Guariba (20,70%) e Aparecida do Taboado (18,80%) que apresentaram moderadas dispersões e por fim Motuca (11,12%) que foi considerada com baixa dispersão dos dados.

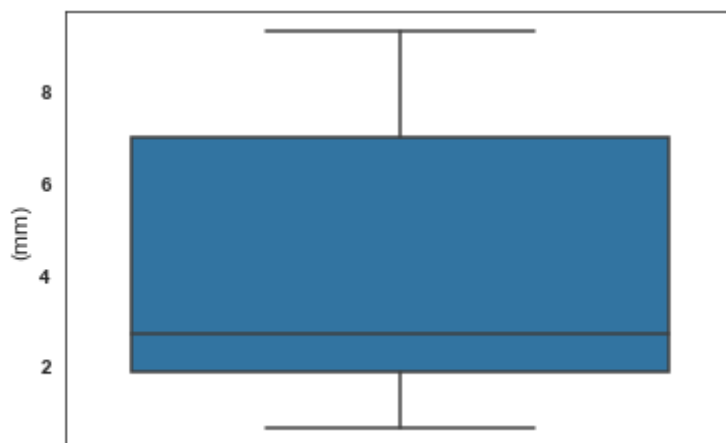


Figura 2.27: Boxplot do armazenamento de água no solo (ARM, mm) para todas as áreas estudadas, N = 24.

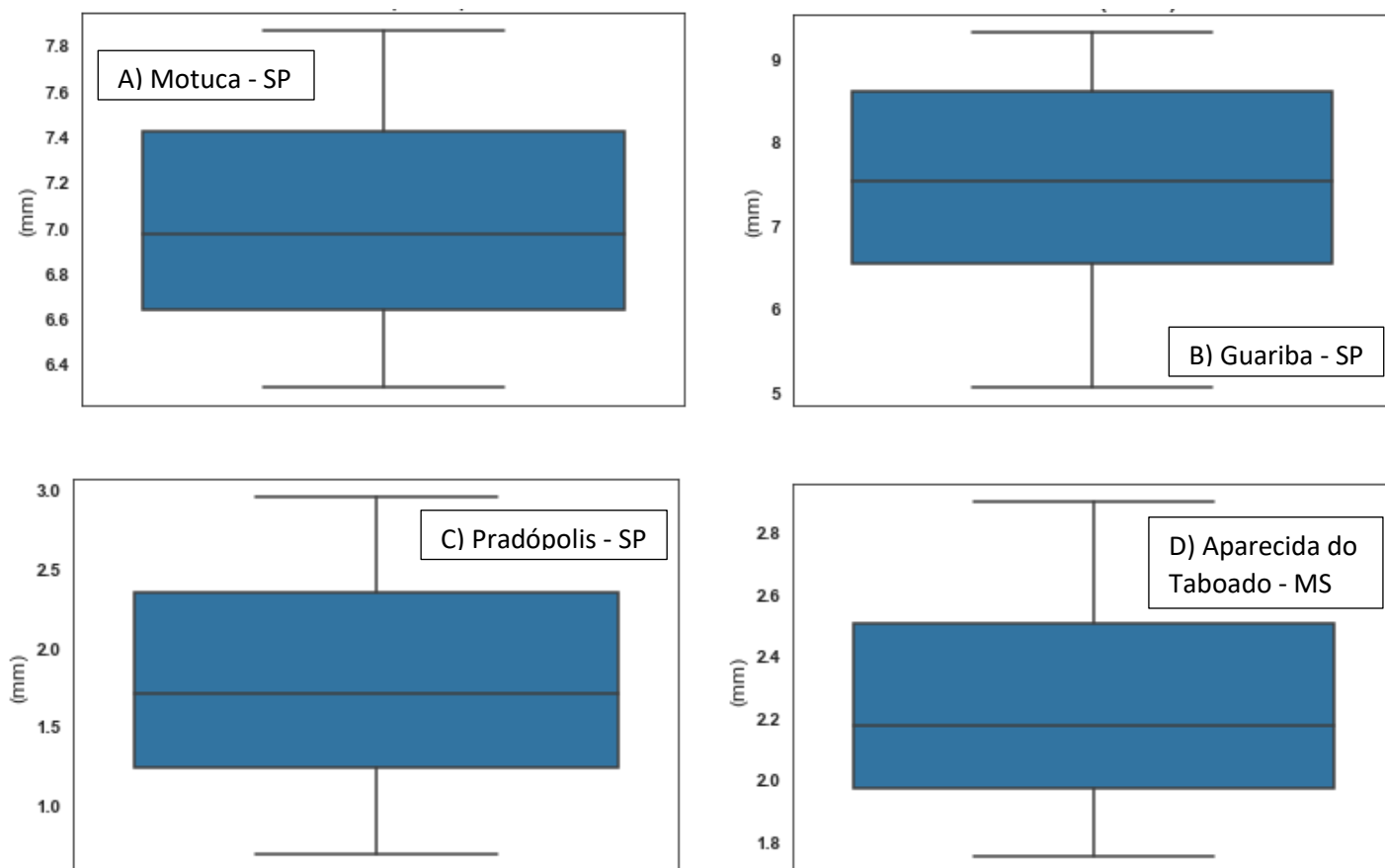


Figura 2.28: Boxplots dos armazenamentos de água no solo (ARM, mm) para os municípios de A) Motuca – SP, B) Guariba – SP, C) Pradópolis – SP e D) Aparecida do Taboado – MS.

2.4.9. CAD observada versus CAD estimada

A avaliação do modelo do CAD foi realizada em relação aos dados estimados dos coeficientes obtidos anteriormente e dos valores da CAD determinada para o estágio de desenvolvimento da brotação da cultura da cana-de-açúcar. Para essa verificação, foi utilizada a técnica da regressão linear. Na equação 30 tem-se a validação os valores da CAD estimada, em função da CAD observada.

$$CAD_{estimada} = 0,0334 + 0,9992 CAD_{observada} \quad (30)$$

Em que:

CAD_{estimado}: capacidade de água disponível no solo estimada (mm);

CAD_{observado}: capacidade de água disponível no solo observada (mm).

O valor do coeficiente de determinação deste modelo foi $R^2 = 0,999$, e o valor do coeficiente de determinação ajustado foi $R_a^2 = 0,999$, ou seja, 99,9% das variações

na estimativa da CAD podem ser explicadas pelas variações dos valores das variáveis independentes. O restante, 0,001% são fontes de variações que são explicados por outros fatores (Tabela 2.17).

A verificação do modelo foi feita por meio do teste F e sua avaliação pelo resultado do p-valor, cujos os valores foram $2,629e^{+04}$ e $2,35e^{-35}$, respectivamente. Além disso, também realizaram o teste de Akaike (AIC), cujo o valor encontrado foi -447,2. Já o valor obtido pelo Critério Bayesiano de Schwarz (BIC) foi -444,9 (Tabela 2.17). Para os dois parâmetros avaliados foram observadas fortes interações da variável independente com a variável dependente, evidenciando que a variável independente apresentou relação com a CAD estimada.

Tabela 2.17: Coeficientes métricas para a determinação do modelo da CAD, N = 24.

Estatísticas	Coeficientes
R-squared	0,999
Adj. R-squared	0,999
F-statistic	$2,629e^{+04}$
Prob (F-statistic)	$2,35e^{-35}$
Log-Likelihood	225,61
AIC	-447,2
BIC	-444,9

R-squared = Coeficiente de Determinação; Adj. R-squared = Coeficiente de Determinação ajustado; F-statistic = valor do teste de Fischer – Snedecor; Prob (F-statistic) = p-valor do teste F; Log-Likelihood = probabilidade natural logarítmica; AIC = Critério de Informação de Akaike; BIC = Critério Bayesiano de Schwarz.

Com relação aos coeficientes encontrados e parâmetros de qualidade da regressão linear, o valor coeficiente linear foi igual a 0,0334 e apresentou o p-valor = 0,893, indicando que este coeficiente não impactou significante na determinação do modelo ao nível de 5%. Já o valor do coeficiente angular foi 0,9992 e o p-valor foi 0,000, onde pode-se concluir que o coeficiente angular apresentou uma relação significativa com a CAD prevista ao nível de 5% (Tabela 2.18), mostrando que o

coeficiente angular interferiu na relação de como os valores estimados são fortemente influenciados pelos valores observados.

Tabela 2.18: Demonstrações do coeficiente linear e angular, resultados do p-valor em função da estatística T para o nível de significância $\alpha = 0,05$ com os seus respectivos valores críticos, para as variáveis CAD observada, versus CAD estimada.

	<i>Coeficientes</i>	<i>Erro Padrão</i>	<i>Resultado do teste t</i>	<i>p-valor</i>	<i>IC_{5%} Limite Inferior</i>	<i>IC_{95%} Limite Superior</i>
Constante	0,0334	0,246	0,136	0,893 ^{NS}	-0,478	0,545
CAD_observada	0,9992	0,006	162,154	0,000*	0,986	1,012

Erro Padrão = Standard error ou erro padrão; Resultado teste t = resultado do teste de hipótese da média T Student; * = significativo no teste T a 5% de probabilidade; NS = não significativo no teste T a 5% de probabilidade; IC_{5%} Limite Inferior = Distribuição, razão de prevalência e intervalo de confiança de 5%; IC_{95%} Limite superior = Distribuição, razão de prevalência e intervalo de confiança de 95%; CAD_observada = capacidade de Armazenamento de Água no Solo observada (mm).

Na Figura 2.29 é observada a validação da escolha da CAD, por meio da representação gráfica da plotagem da CAD observada versus CAD estimada. Observa-se que o ângulo da reta, com relação ao ajuste linear foi muito próximo de 45 graus, indicando que existiu uma forte relação entre a CAD observada e a CAD estimada. Além disso, também mensuraram os valores da RMSE e da MSE, onde foram encontrados os valores de 0 para os dois parâmetros, mostrando que os valores observados e estimados apresentaram altas correlações.

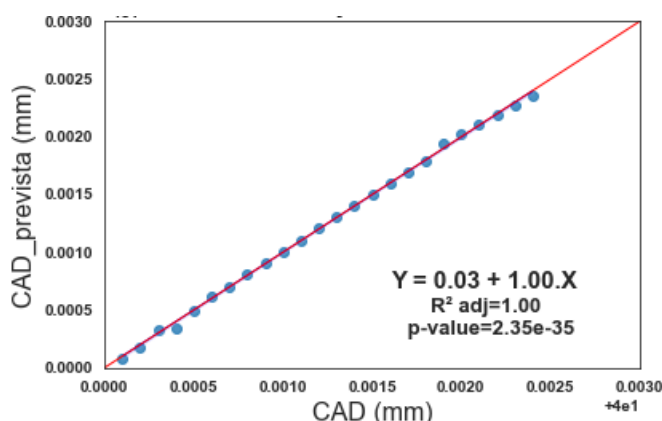


Figura 2.29: Capacidade de água disponível no solo (CAD) observada versus CAD estimada pela equação de regressão.

2.4.10. Regressão Linear Múltipla da Umidade do solo (Us), em função das variáveis ETP, Precipitação e ARM

Os processos de transporte e de produção de CO₂ dos solos agrícolas são extremamente complexos e são resultantes dos atributos físicos e químicos do solo que apresentam grandes variabilidades espaciais e temporais, tornando-se complexa sua explicação. De maneira geral, observa-se que os processos de difusão e emissão de CO₂ do solo para a atmosfera é impactado principalmente pela interação entre os espaços porosos e o teor de umidade do solo (Dossou-Yovo et al., 2016), que por sua vez estão diretamente relacionados com a capacidade de armazenagem de água do solo.

Outros fatores são importantes para a determinação da umidade do solo, tais como a ETP, ARM e a precipitação. Nos últimos 50 anos o clima do planeta sofreu com o seu aquecimento, essas mudanças térmicas resultaram em importantes variações na agricultura, nos sistemas ambientais, na saúde dos seres humanos e na segurança alimentar. Além disso, também impactaram diretamente no ciclo hidrológico, pois quanto maior for a incidências de temperaturas mínimas, observa-se a diminuição da amplitude térmica diária, colaborando diretamente com o aumento da evapotranspiração potencial (ETP) desta região (Salviano et al., 2016) e também nos processos de transporte e de produção de CO₂ dos solos agrícolas.

A equação de regressão linear múltipla que foi estimada para explicar a umidade do solo, em função das variáveis ETP, precipitação e ARM foi representada pela Equação 31:

$$Us = -1,6591 + 2,8635ETP + 0,3824PREC + 1,7984ARM \quad (31)$$

Em que:

Us: umidade do solo (mm);

ETP: evapotranspiração potencial (mm);

PREC: precipitação (mm);

ARM: armazenamento de água no solo (mm).

O valor do coeficiente de determinação do modelo foi de $R^2 = 0,904$, e o seu respectivo valor ajustado foi $R_a^2 = 0,889$, ou seja, 88,9% das variações na estimativa da umidade do solo podem ser explicadas pelas variações dos valores das variáveis

independentes. O restante, 11,1% são fontes de variações que são explicados por outros fatores (Tabela 2.19).

Tabela 2.19: Coeficientes métricos do modelo da umidade do solo, N = 24.

Estatísticas	Coeficientes
R-squared	0,904
Adj. R-squared	0,889
F-statistic	62,50
Prob (F-statistic)	2,45e ⁻¹⁰
Log-Likelihood	-49,106
AIC	106,2
BIC	110,9

R-squared = Coeficiente de Determinação; Adj. R-squared = Coeficiente de Determinação ajustado; F-statistic = valor do teste de Fischer – Snedecor; Prob (F-statistic) = p-valor do teste F; Log-Likelihood = probabilidade natural logarítmica; AIC = Critério de Informação de Akaike; BIC = Critério Bayesiano de Schwarz.

Para verificar os resultados obtidos para a determinação da equação, foi realizado o teste de hipóteses para $H_0: \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$ ou H_1 : pelo menos um β é diferente dos demais por meio dos parâmetros do teste F e do p-valor, onde foram encontrados os valores de 62,50 e 2,45e⁻¹⁰, respectivamente. Esses valores indicaram que foi rejeitada a hipótese nula de que todos os coeficientes independentes seriam iguais, ao nível de significância de 0,05, indicando que pelo menos uma das variáveis independentes apresentaram relação com a umidade do solo. Também foi realizado o teste de Akaike (AIC), cujo o valor obtido foi 106,2 e o valor encontrado pelo Critério Bayesiano de Schwarz (BIC) foi 110,9 (Tabela 2.19). Os dois parâmetros avaliativos mostraram fortes interações entre as variáveis independentes e a dependente, indicando uma alta relação da ETP, PREC e da ARM com Us.

O valor do coeficiente linear do modelo da umidade do solo prevista foi -1,6591 e não apresentou significância estatística (p-valor = 0,311), Tabela 2.20. O valor negativo do coeficiente linear confirmou que o solo no momento da coleta dos dados tende a ter baixos valores ou não ter umidade, esse valor pode ser confirmado pois as leituras foram realizadas à uma profundidade de 0 a 10 cm e com solos que aparentemente estavam secos superficialmente ou com baixas umidades.

Tabela 2.20: Demonstrações dos coeficientes lineares e angulares, resultados do p-valor em função da estatística t para o nível de significância $\alpha = 0,05$ com os seus respectivos valores críticos, para as variáveis ETP, Precipitação e ARM que foram usadas na determinação do modelo da Us, N = 24.

	<i>Coeficientes</i>	<i>Erro Padrão</i>	<i>Resultado do teste t</i>	<i>p-valor</i>	<i>IC_{5%} Limite Inferior</i>	<i>IC_{95%} Limite Superior</i>
Constante	-1,6591	1,596	-1,039	0,311 ^{NS}	-4,989	1,671
ETP	2,8635	0,493	5,803	0,000*	1,834	3,893
Precipitação	0,3824	0,376	1,018	0,321 ^{NS}	-0,401	1,166
ARM	1,7984	0,172	10,448	0,000*	1,439	2,157

Erro Padrão = Standard error ou erro padrão; Resultado teste t = resultado do teste de hipótese da média T Student; * = significativo no teste T a 5% de probabilidade; NS = não significativo no teste T a 5% de probabilidade; IC_{5%} Limite Inferior = Distribuição, razão de prevalência e intervalo de confiança de 5%; IC_{95%} Limite superior = Distribuição, razão de prevalência e intervalo de confiança de 95%; ETP = evapotranspiração potencial (mm); Precipitação = precipitação (mm); ARM = armazenamento de água no solo (mm).

Para os coeficientes angulares foram obtidos os seguintes resultados: ETP = 2,8635, Precipitação = 0,3824 e ARM = 1,7984. Os valores de p-valor para as variáveis foram 0,000, 0,321 e 0,000, respectivamente (Tabela 2.20). A Precipitação apresentou baixo poder explicativo no modelo, entretanto, a ETP e a ARM foram significativas e conseguiram impactar no modelo ao nível de significância de 5%. Porém para outro conjunto de dados, pode ser que a precipitação apresente um maior desempenho.

A técnica de boxplot e a estatística descritiva foram usadas para analisar as variabilidades das variáveis utilizadas no modelo. Nos gráficos das Figura 2.30 e 2.31 e nas Tabelas 2.2 e 2.3 observam-se os boxplots e a variação temporal da ETP. A média geral da ETP observada foi 2,85 mm e apresentou desvio padrão de 0,89 mm. Avaliando os locais de estudo individualmente, Guariba mostrou a menor média (1,88 mm), já o município de Motuca foi o de maior média de temperatura (4,98 mm) dentre todas as áreas estudadas. Com relação aos valores de coeficiente de variação, foram observados que para todos os dados das áreas estudadas o valor de 31,22%,

mostrando altas amplitudes da ETP. Quando se analisou cada local de estudo foram observados valores inferiores a 15%, sendo que Aparecida do Taboado apresentou o menor valor (4,78%), seguido de Motuca (6,13%), Guariba (8,83%) e Pradópolis (14,41%).

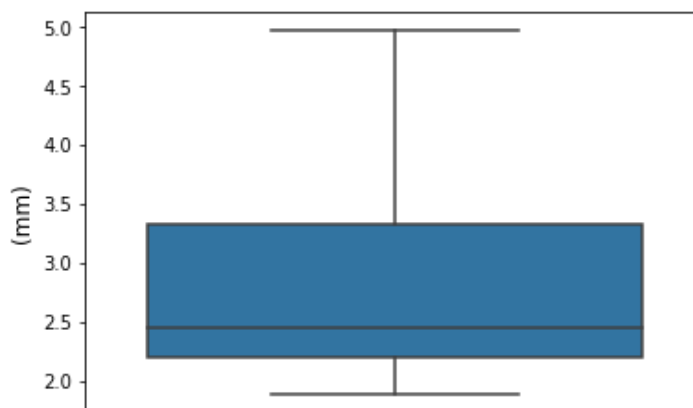


Figura 2.30: Boxplot da evapotranspiração potencial (ETP, mm) para todas as áreas estudadas, N = 24.

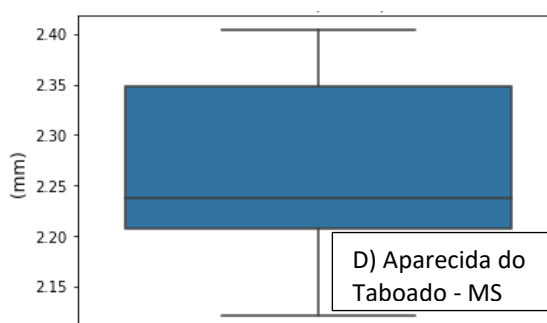
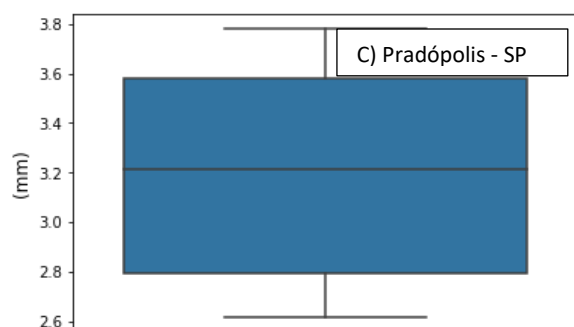
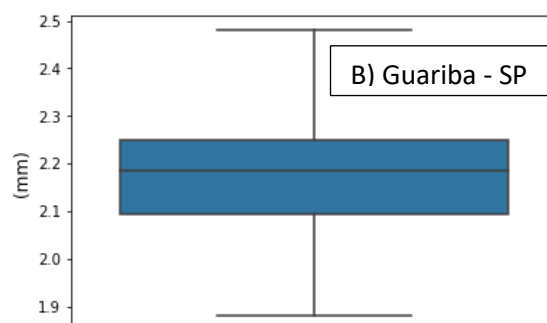
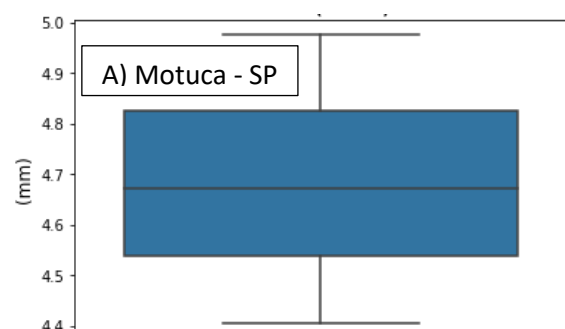


Figura 2.31: Boxplots da evapotranspiração potencial (ETP, mm) para os municípios A) Motuca – SP, B) Guariba – SP, C) Pradópolis – SP e D) Aparecida do Taboado – MS.

A outra variável analisada que foi utilizada para determinar a umidade do solo foi a precipitação (Precipitação, mm). Nas Figuras 2.32 e 2.33 e as Tabelas 2.2 e 2.3, observou-se que a média geral da Precipitação foi de 0,45 mm e desvio padrão 1,35 mm. Analisando cada área de estudo, foi observado os maiores valores em Guariba (6,05 mm) e em todos os municípios registraram ausência de chuva (0,00 mm). Devido ao tamanho das áreas usadas para a medição da FCO₂, não foram observadas variações dos teores de P, por isso que os desvios padrões foram 0. Quando analisaram os valores dos coeficientes de variação, foram observados uma alta variabilidade para todos os dados das áreas estudadas, registrando o valor de 297,37%. O mesmo aconteceu nas análises dos locais estudados, onde foram registrados valores superiores a 30%, sendo que Pradópolis apresentou o maior valor (282,59%), seguido de Guariba (181,08%), Aparecida do Taboado (158,36%) e Motuca (94,67%).

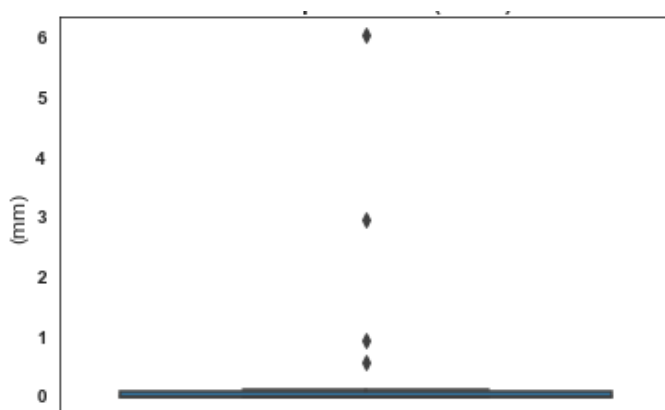


Figura 2.32: Boxplot da precipitação (Precipitação, mm) para todas as áreas estudadas, N = 24.

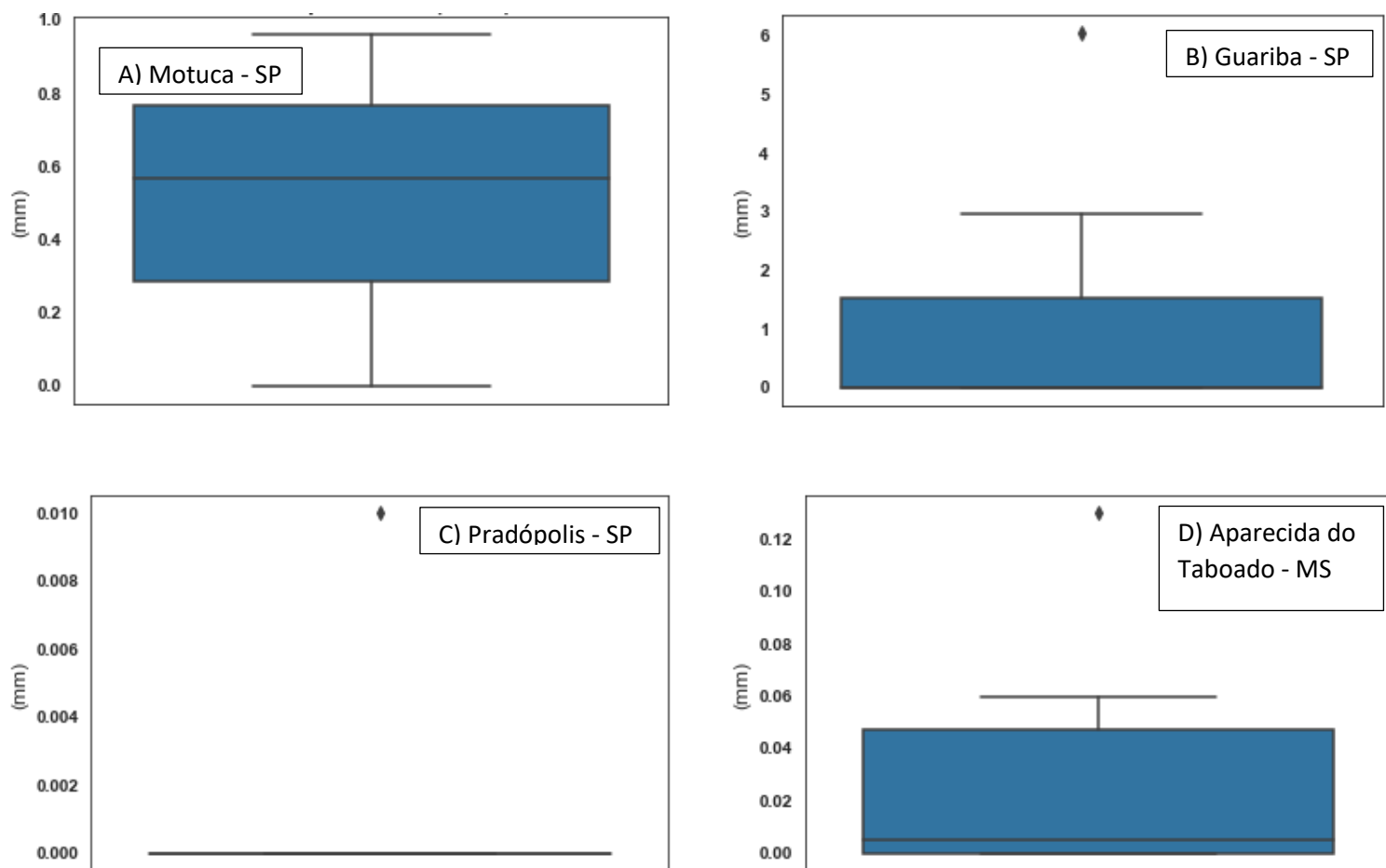


Figura 2.33: Boxplots das precipitações (Precipitação, mm) para os municípios de A) Motuca – SP, B) Guariba – SP, C) Pradópolis – SP e D) Aparecida do Taboado – MS.

Outra variável utilizada para a predição da umidade do solo foi a ARM. Nas Figuras 2.34 e 2.35 e as Tabelas 2.2 e 2.3 foi registrada a média geral da ARM que foi 4,22 mm e o desvio padrão de 2,87 mm. Estudando cada local de estudo, observaram que os menores valores foram obtidos em Pradópolis (0,69 mm), já em Guariba observaram os maiores valores (9,33 mm), isso mostrou que Guariba tende a armazenar mais água no solo, dentre todas as áreas estudadas. Quando analisaram os valores dos coeficientes de variação, foram observados uma alta variabilidade para todos os dados das áreas estudadas, apresentando o valor de 68,07%. Quando foi estudado cada local, observaram que Pradópolis apresentou o maior valor (44,69%) e alta dispersão dos dados, seguido de Guariba (20,70%) e Aparecida do Taboado

(18,80%) que apresentaram moderadas dispersões, e por fim Motuca (11,12%) que foi considerada com baixa dispersão dos dados.

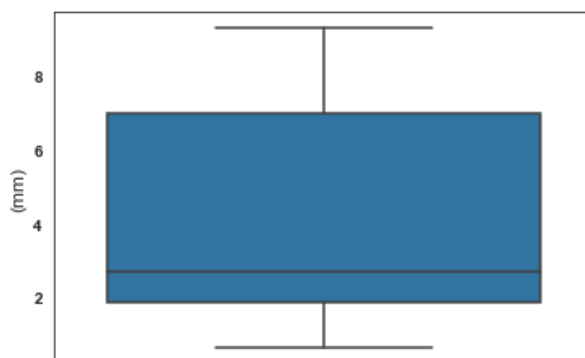


Figura 2.34: Boxplot do armazenamento de água no solo (ARM, mm) para todas as áreas estudadas, N = 24.

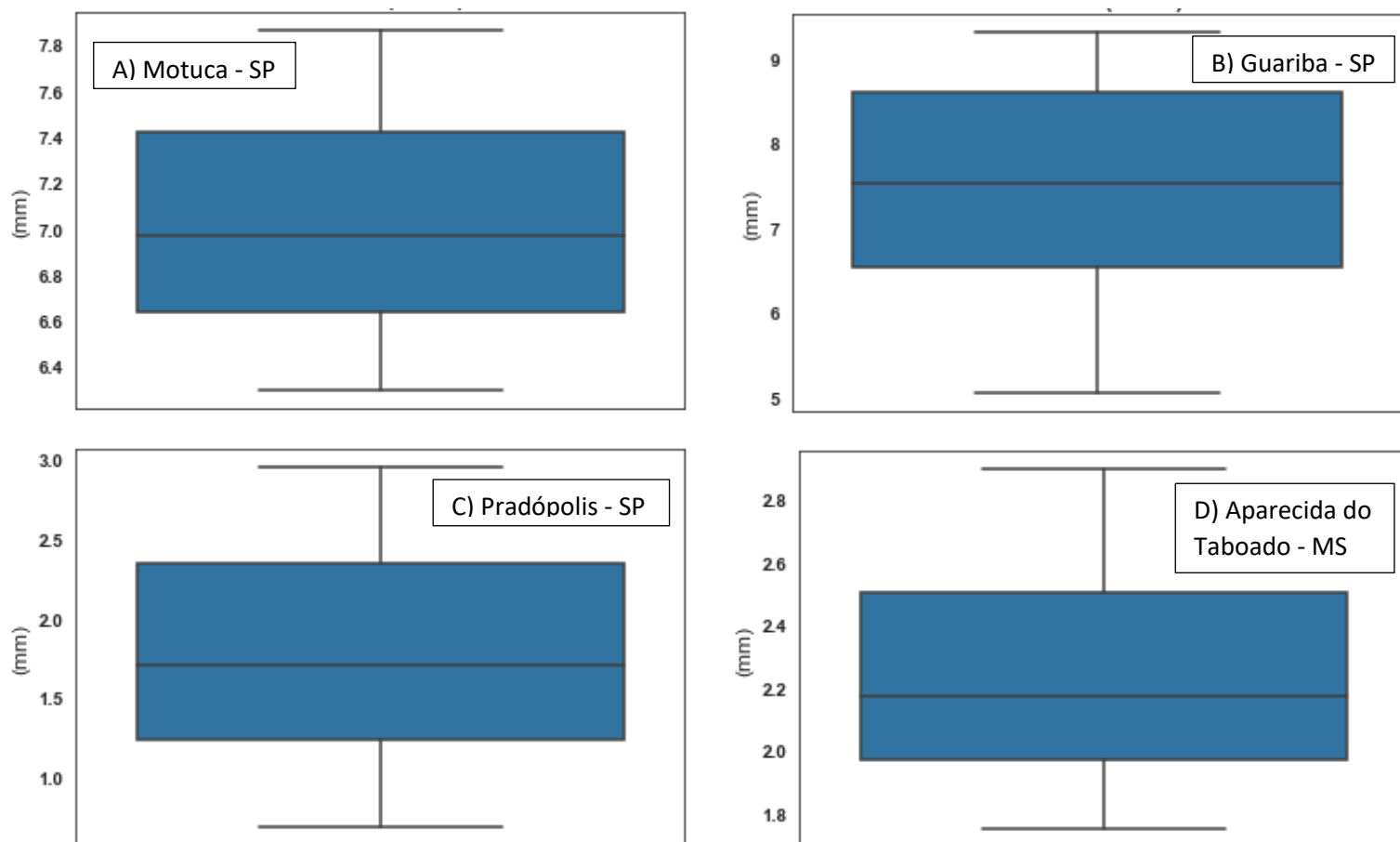


Figura 2.35: Boxplots dos armazenamentos de água no solo (ARM, mm) para os municípios de A) Motuca – SP, B) Guariba – SP, C) Pradópolis – SP e D) Aparecida do Taboado – MS.

Quando se avaliou a principal variável, a U_s , foram observadas nas Figuras 2.36 e 2.37 e nas Tabelas 2.2 e 2.3 que a média geral foi de 14,27 mm e desvio padrão 6,16 mm. Quando analisaram os locais estudados, foram registrados os menores valores em Aparecida do Taboado (5,03 mm), já em Motuca os maiores valores (25,03 mm), mostrando que Motuca tende a ter um solo mais úmido, dentre todas as áreas estudadas. Quando foram analisados os valores dos coeficientes de variação, observaram uma alta variabilidade para todos os dados das áreas estudadas, sendo obtido o valor de 43,16%. Quando estudaram cada local, foi observado que Motuca (4,42%) e Guariba (11,05%) apresentaram baixas dispersões dos dados, já Pradópolis (16,76%) e Aparecida do Taboado (19,55%) apresentaram moderadas dispersões.

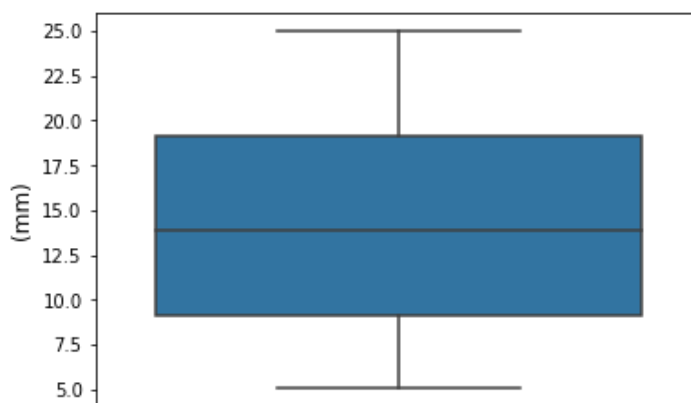


Figura 2.36: Boxplot da umidade do solo (U_s , mm) para todas as áreas estudadas, N = 24.

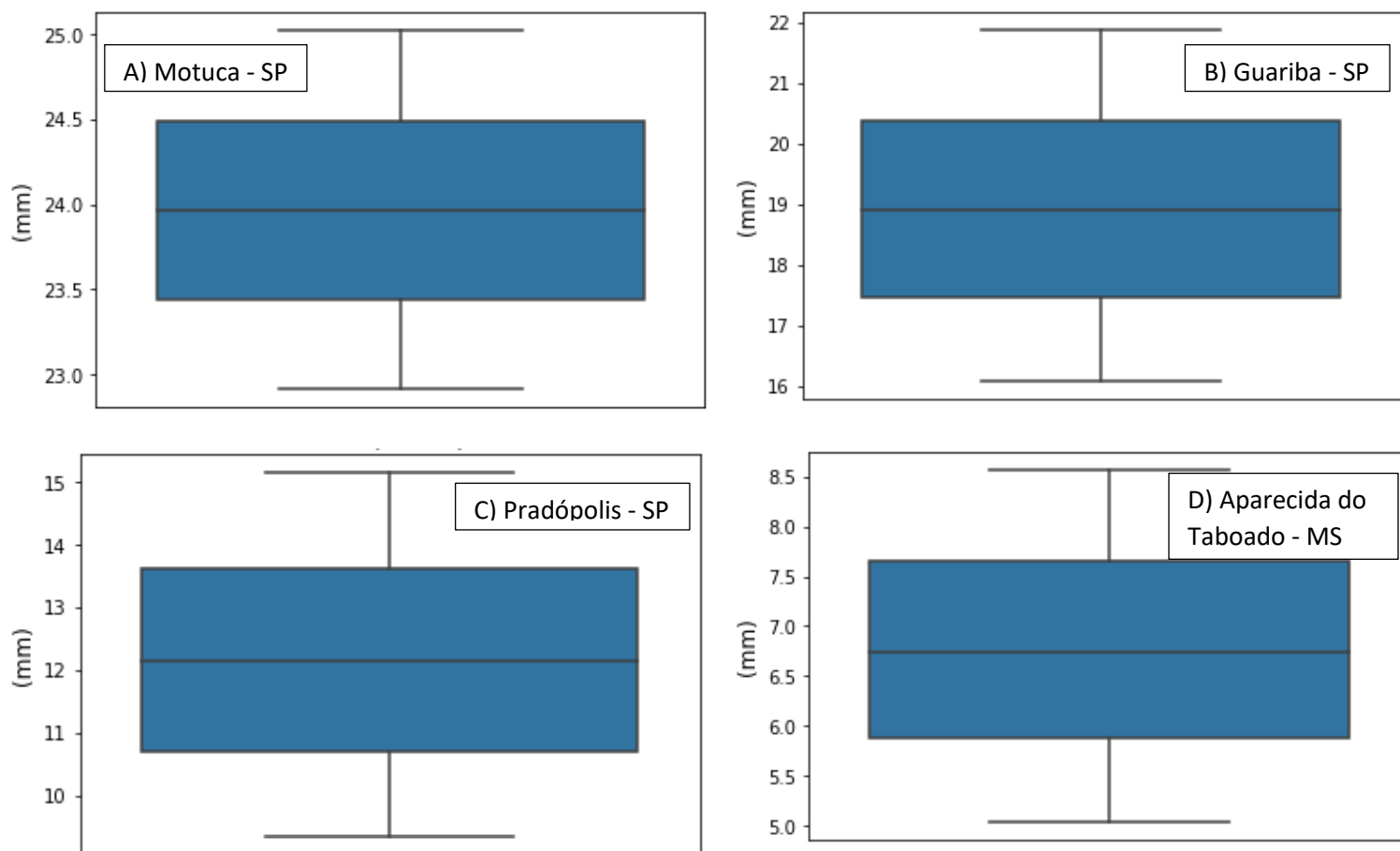


Figura 2.37: Boxplots das umidades do solo (U_s , mm) para os municípios A) Motuca – SP, B) Guariba – SP, C) Pradópolis – SP e D) Aparecida do Taboado – MS.

2.4.11. Umidade do solo observada versus Umidade do solo estimada

A avaliação do modelo da umidade do solo foi feita por meio da relação entre os dados estimados e os dados observados em campo. A técnica utilizada para essa verificação foi a regressão linear. A equação da regressão linear que foi estimada para validar os valores da umidade do solo estimada, em função dos valores observados foi representada pela Equação 32:

$$U_{s_{estimada}} = -2,132e^{-14} + 1,000 U_{s_{observada}} \quad (32)$$

Em que:

$U_{s_{estimada}}$: umidade do solo estimada (mm);

$U_{s_{observada}}$: umidade do solo observada (mm).

O valor do coeficiente de determinação deste modelo foi $R^2 = 0,904$, e o valor do coeficiente de determinação ajustado foi $R_a^2 = 0,899$, ou seja, 8,9% das variações

na estimativa da umidade do solo foram explicadas pela variação dos valores da variável independente. O restante, 10,1% são fontes de variações que são explicadas por outros fatores (Tabela 2.21).

Tabela 2.21: Coeficientes métricas para a determinação do modelo da umidade do solo, N = 24.

Estatísticas	Coeficientes
R-squared	0,904
Adj. R-squared	0,899
F-statistic	206,2
Prob (F-statistic)	$1,18e^{-12}$
Log-Likelihood	-47,890
AIC	99,78
BIC	102,1

R-squared = Coeficiente de Determinação; Adj. R-squared = Coeficiente de Determinação ajustado; F-statistic = valor do teste de Fischer – Snedecor; Prob (F-statistic) = p-valor do teste F; Log-Likelihood = probabilidade natural logarítmica; AIC = Critério de Informação de Akaike; BIC = Critério Bayesiano de Schwarz.

Para realizar a verificação dos resultados obtidos para a determinação da equação, foi feito o teste F e a partir do seu resultado obteve-se o p-valor. Foram encontrados os valores 206,2 e $1,18e^{-12}$, respectivamente (Tabela 2.21), evidenciando que a variável independente teve uma alta relação com a umidade do solo estimada. Também realizaram o teste de Akaike (AIC), cujo o valor encontrado foi de 99,78. Já o valor obtido pelo Critério Bayesiano de Schwarz (BIC) foi de 102,1 (Tabela 2.21). Para os dois parâmetros avaliativos foram observadas fortes interações da variável independente com a variável dependente, evidenciando a forte interação da umidade do solo observada com a umidade do solo estimada.

Analisando os coeficientes e parâmetros da qualidade da regressão linear, observaram que o valor coeficiente linear foi igual a 1,3756 e apresentou o p-valor = 0,172, indicando que este coeficiente não impactou significante na determinação do modelo. Porém quando foi estudado o coeficiente angular obtiveram o valor de 0,9036 e 0,000, respectivamente, evidenciando que o coeficiente angular apresentou uma relação significativa ao nível de 5% com a umidade do solo estimada (Tabela 2.22).

Tabela 2.22: Demonstrações do coeficiente linear e angular, resultados do p-valor em função da estatística t para o nível de significância $\alpha = 0,05$ com os seus respectivos valores críticos, para as variáveis umidade do solo observadas, versus umidade do solo estimadas.

	<i>Coeficientes</i>	<i>Erro Padrão</i>	<i>Resultado do teste t</i>	<i>p-valor</i>	<i>IC_{5%} Limite Inferior</i>	<i>IC_{95%} Limite Superior</i>
Constante	1,3756	0,975	1,411	0,172 ^{NS}	-0,646	3,397
Umidade do solo observada (mm)	0,9036	0,063	14,361	0,000*	0,773	1,034

Erro Padrão = Standard error ou erro padrão; Resultado teste t = resultado do teste de hipótese da média T Student; * = significativo no teste T a 5% de probabilidade; NS = não significativo no teste T a 5% de probabilidade; IC_{5%} Limite Inferior = Distribuição, razão de prevalência e intervalo de confiança de 5%; IC_{95%} Limite superior = Distribuição, razão de prevalência e intervalo de confiança de 95%.

Segue a representação gráfica dos dados da umidade do solo observados versus umidade do solo estimados (Figura 2.38). Observaram que o ângulo da reta, com relação ao ajuste linear foi praticamente de 45 graus, caracterizando a forte relação entre a umidade do solo observada e a umidade do solo estimada. Além disso para a RMSE e a MSE foram encontrados os valores de 1,87 e 3,51, respectivamente, indicando a existência de uma alta correlação entre os valores observados e estimados.

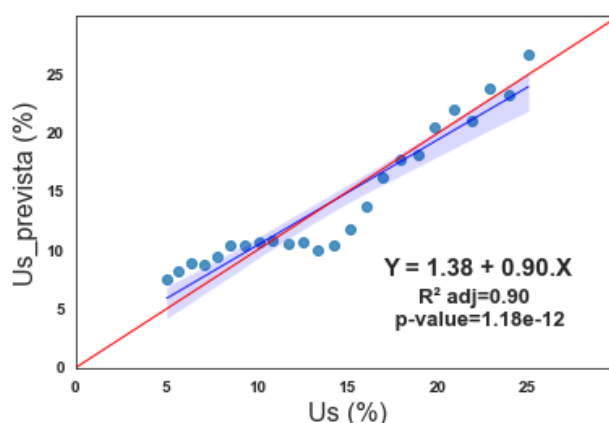


Figura 2.38: Umidade do solo (Us) observado versus US estimada pela equação de regressão.

2.4.12. Regressão Linear Múltipla da PLA, em função das variáveis PLA_BH e Dp

O solo é um conglomerado de diferentes partículas que se encontram ligadas por meio de uma estrutura padronizada, formando os agregados. As frações granulométricas são componentes importantes de diversos modelos ambientais que estudam fenômenos de pedotransferência. A textura do solo é amplamente usada na previsão de variáveis do solo que são dificilmente mensurados, principalmente quando são estudados parâmetros relacionados com os processos de transportes hídricos (Patil e Sing, 2016) e de gases. Sabendo da complexidade do processo da FCO_2 e das fortes relações existentes entre textura e a porosidade do solo, será estabelecida uma relação entre a porosidade livre de água e a densidade das partículas. Outros autores também observaram as relações entre essas variáveis (Almeida et al., 2018; Cerri et al., 2013, Silva-Olaya et al., 2013).

A equação de regressão linear múltipla que foi estimada para explicar a PLA, em função das variáveis PLA_BH e Dp foi representada pela equação 39:

$$PLA = -98,8014 + 2,1009PLA_{BH} + 20,6205Dp \quad (33)$$

Em que:

PLA: porosidade livre de água (%);

PLA_BH: porosidade livre de água decorrente do balanço hídrico (%);

Dp: densidade de partículas ($kg\ dm^{-3}$).

O valor do coeficiente de determinação do modelo foi de $R^2 = 0,902$, e o seu respectivo valor ajustado foi $R_a^2 = 0,893$, ou seja, 89,3% das variações da umidade do solo podem ser explicadas pelas variações dos valores das variáveis independentes. O restante, 10,7% são fontes de variações que são explicados por outros fatores. A verificação da qualidade do modelo foi realizada por meio do teste de hipóteses, sendo $H_0: \beta_1 = \beta_2 = \beta_3 = 0$ ou H_1 : pelo menos um β é diferente dos demais. Os valores do teste F e do p-valor foram 97,04 e $2,46e^{-11}$, respectivamente. Esses valores indicaram que foi rejeitada a hipótese nula de que todos os coeficientes independentes seriam iguais, ao nível de significância de 0,05, mostrando que pelo menos uma das variáveis independentes apresentaram relação com a PLA (Tabela 2.23).

Tabela 2.23: Coeficientes métricos do modelo da umidade do solo, N = 24.

Estatísticas	Coeficientes
R-squared	0,902
Adj. R-squared	0,893
F-statistic	97,04
Prob (F-statistic)	2,46e ⁻¹¹
Log-Likelihood	-59,457
AIC	124,9
BIC	128,4

R-squared = Coeficiente de Determinação; Adj. R-squared = Coeficiente de Determinação ajustado; F-statistic = valor do teste de Fischer – Snedecor; Prob (F-statistic) = p-valor do teste F; Log-Likelihood = probabilidade natural logarítmica; AIC = Critério de Informação de Akaike; BIC = Critério Bayesiano de Schwarz.

Além disso, também foi realizado o teste de Akaike (AIC), cujo o valor obtido foi 124,9. Já o valor encontrado pelo Critério Bayesiano de Schwarz (BIC) foi 128,4 (Tabela 2.23). Os dois parâmetros avaliativos indicaram que existiram fortes interações entre as variáveis independentes e a dependente, mostrando uma alta relação da PLA_BH e Dp com a PLA.

Quando foi analisado o valor do coeficiente linear do modelo da PLA prevista e seu p-valor foram obtidos os valores de -98,8014 e 0,000, respectivamente, evidenciando que este parâmetro apresentou significância estatística no modelo. Para os coeficientes angulares foram encontrados os seguintes resultados: PLA_BH = 2,1009 e Dp = 20,6205. Os valores de p-valor para as variáveis foram 0,000 e 0,000, respectivamente (Tabela 2.24). As duas variáveis apresentaram forte poder explicativo dentro do modelo, impactando ao nível de significância de 5%.

Tabela 2.24: Demonstrações dos coeficientes lineares e angulares, resultados do p-valor em função da estatística t para o nível de significância $\alpha = 0,05$ com os seus respectivos valores críticos, para as variáveis PLA_BH e Dp que foram usadas na determinação do modelo da PLA, N = 24.

	<i>Coeficientes</i>	<i>Erro Padrão</i>	<i>Resultado do teste t</i>	<i>p-valor</i>	<i>IC_{5%} Limite Inferior</i>	<i>IC_{95%} Limite Superior</i>
Constante	-98,8014	10,080	-9,802	0,000*	-119,764	-77,839
PLA_BH	2,1009	0,226	9,301	0,000*	1,631	2,571
Dp	20,6205	1,779	11,589	0,000*	16,920	24,321

Erro Padrão = Standard error ou erro padrão; Resultado teste t = resultado do teste de hipótese da média T Student; * = significativo no teste T a 5% de probabilidade; NS = não significativo no teste T a 5% de probabilidade; IC_{5%} Limite Inferior = Distribuição, razão de prevalência e intervalo de confiança de 5%; IC_{95%} Limite superior = Distribuição, razão de prevalência e intervalo de confiança de 95%; PLA_BH = porosidade livre de água decorrente do balanço hídrico (%); Dp = densidade das partículas (kg dm⁻³).

A estatística descritiva e os gráficos boxplots foram utilizados para avaliar as variabilidades das variáveis utilizadas no modelo. Nos gráficos das Figuras 2.39 e 2.40 e nas Tabelas 2.2 e 2.3 foram registrados os valores da PLA_BH. A média geral da PLA_BH observada foi 35,78% e apresentou desvio padrão de 2,87%. Quando avaliado os locais de estudo individualmente, Guariba registrou o menor valor (30,67%) e Pradópolis o maior (39,30%), dentre todas as áreas estudadas. Com relação aos valores de coeficiente de variação, foram observados que para todos os dados das áreas estudadas o valor de 8,03%, indicando uma baixa amplitude da PLA_BH. O mesmo aconteceu quando se analisou cada local de estudo, onde foram observados valores inferiores a 15%, sendo que Aparecida do Taboado apresentou o menor valor (1,13%), seguido de Pradópolis (2,08%), Motuca (2,38%) e Guariba (4,76%).

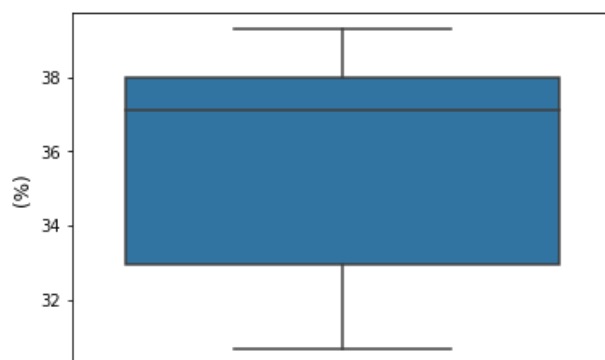


Figura 2.39: Boxplot da porosidade livre de água decorrente do balanço hídrico (PLA_BH, %) para todas as áreas estudadas, N = 24.

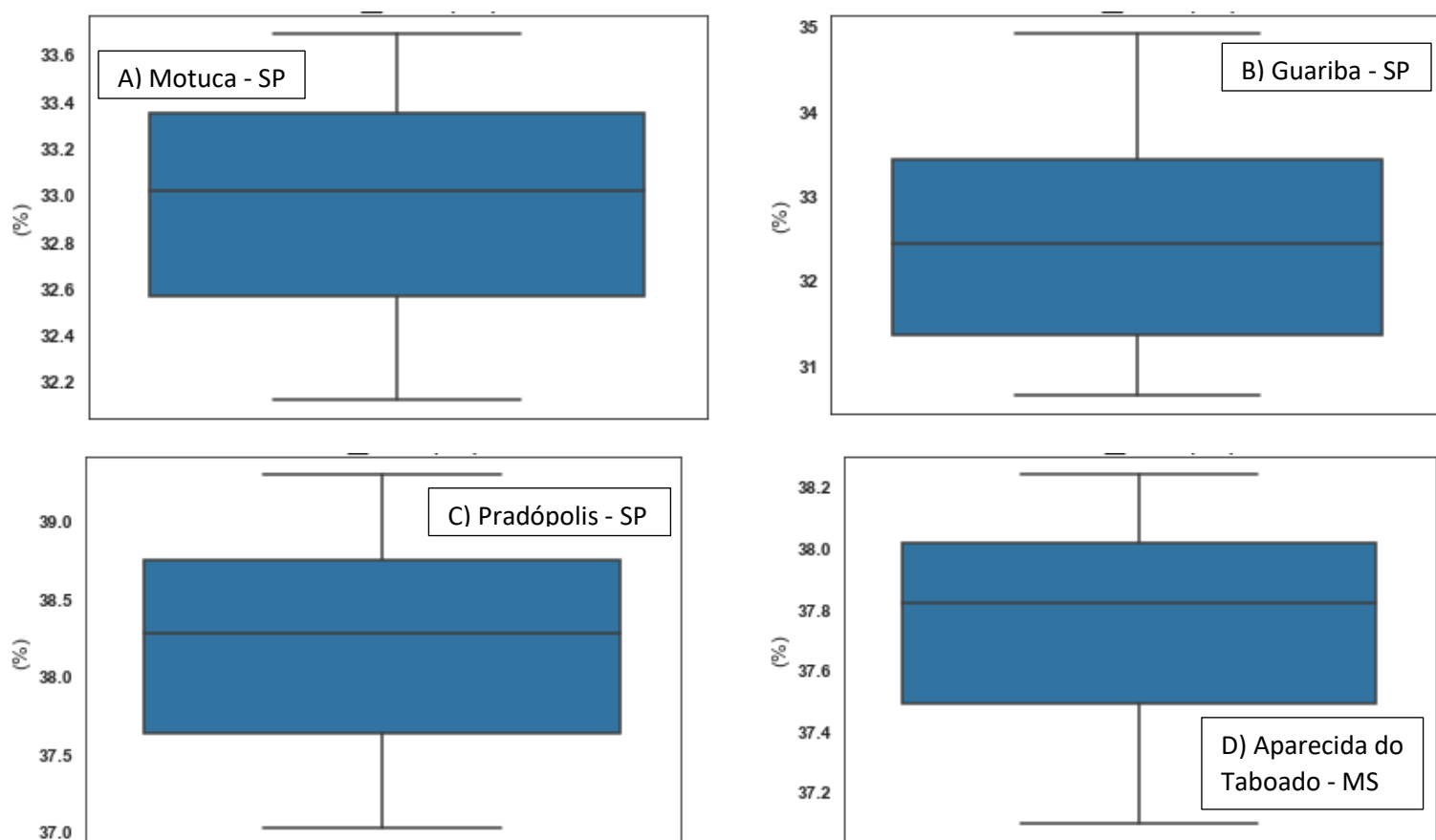


Figura 2.40: Boxplots da porosidade livre de água decorrente do balanço hídrico (PLA_BH, %) para os municípios A) Motuca – SP, B) Guariba – SP, C) Pradópolis – SP e D) Aparecida do Taboado – MS.

A outra variável que foi analisada e utilizada na construção do modelo foi a Dp. Na Figura 2.41 e nas Tabelas 2.2 e 2.3 observa-se que a média geral da Dp foi 2,78

kg dm^{-3} e desvio padrão $0,36 \text{ kg dm}^{-3}$. Analisando cada área de estudo, foi observado os maiores valores em Guariba ($3,18 \text{ kg dm}^{-3}$) e os menores em Motuca ($2,03 \text{ kg dm}^{-3}$), dentre todas as áreas estudadas. Quando analisaram os valores dos coeficientes de variação para todas as áreas estudadas, foram observados uma baixa variabilidade registrando o valor de 13,11%. Quando foi analisado cada local de estudo a variação foi de 0%, mostrando que não ocorreu variação dessa variável *in loco*.

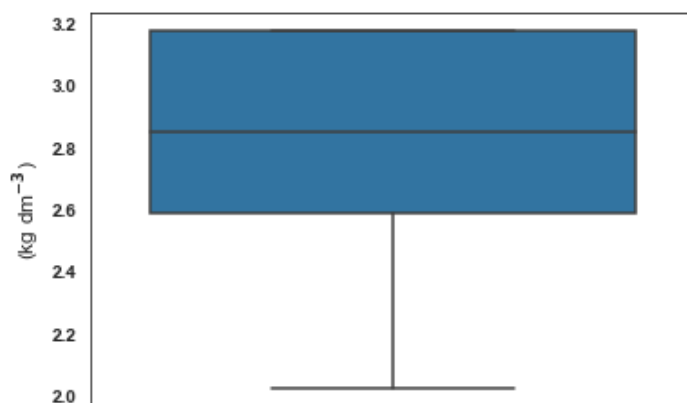


Figura 2.41: Boxplot da densidade de partículas (D_p , kg dm^{-3}) para todas as áreas estudadas, $N = 24$.

A variabilidade da variável PLA está representada nas Figuras 2.42 e 2.43 e nas Tabelas 2.2 e 2.3. A média geral da PLA foi 33,74% e o desvio padrão foi 9,42%. Estudando cada local de estudo observaram que os menores valores foram obtidos em Motuca (6,34%), já em Pradópolis observaram os maiores valores (42,77%), dentre todas as áreas estudadas. Quando analisaram os valores dos coeficientes de variação, foram observados uma moderada variabilidade para todos os dados das áreas estudadas, apresentando o valor de 27,92%. Quando foi estudado cada local observaram que Motuca apresentou o maior valor (53,43%) e alta dispersão dos dados, seguido de Aparecida do Taboado (9,57%), Guariba (4,03%) e Pradópolis (1,77%) que apresentaram baixas dispersões dos dados.

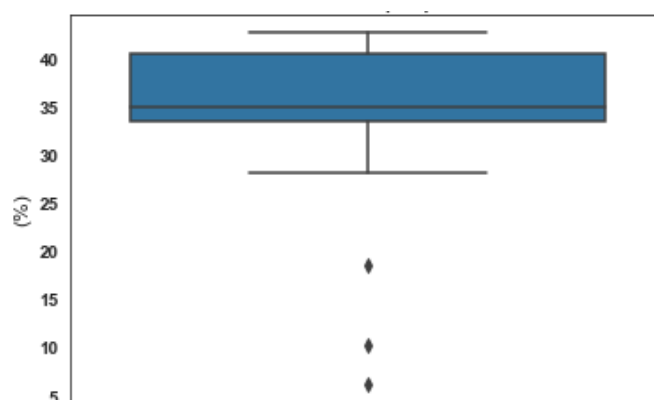


Figura 2.42: Boxplot da porosidade livre de água do solo (PLA, %) para todas as áreas estudadas, N = 24.

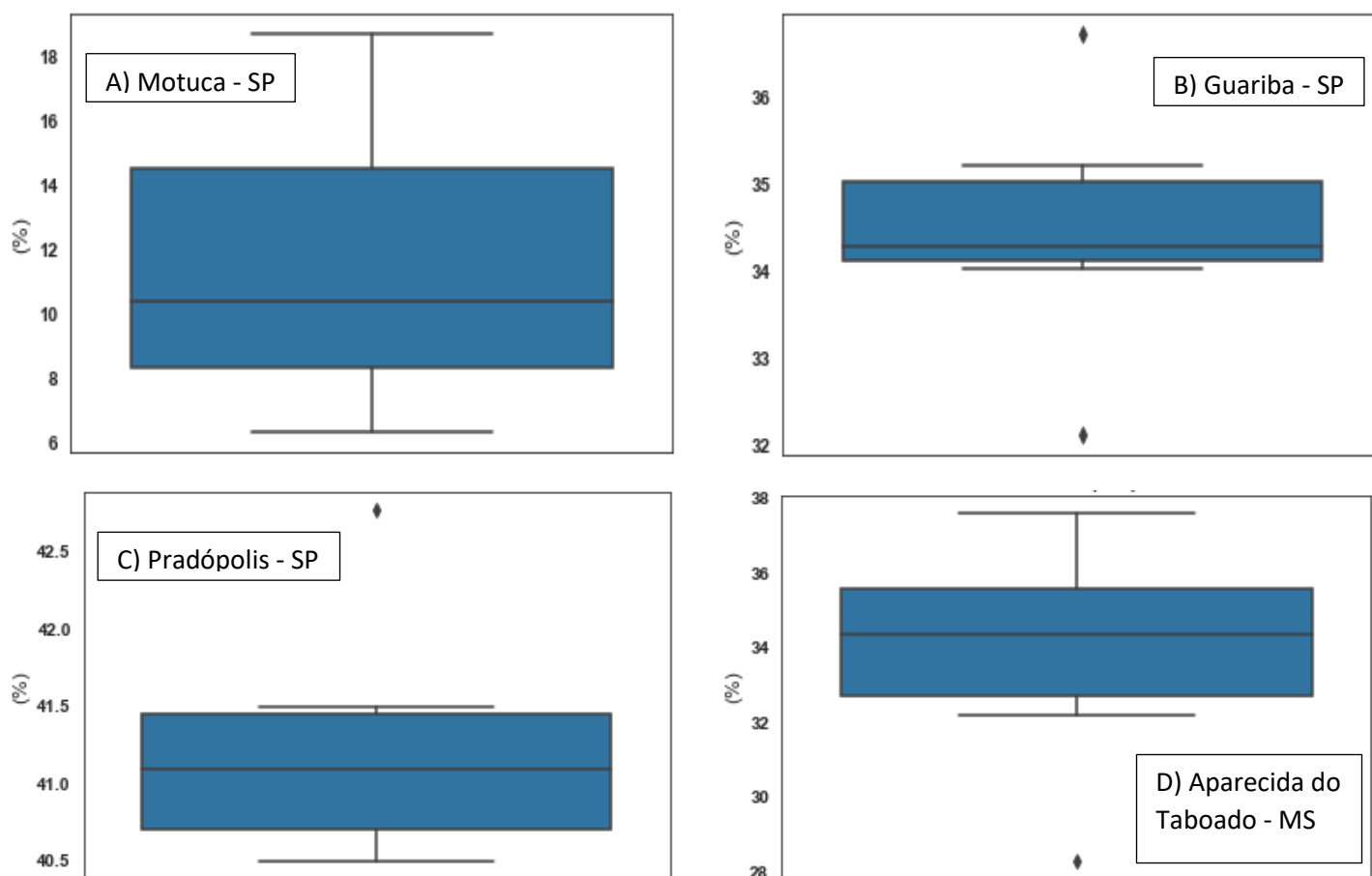


Figura 2.43: Boxplots da porosidade livre de água do solo (PLA, %) para os municípios de A) Motuca – SP, B) Guariba – SP, C) Pradópolis – SP e D) Aparecida do Taboado – MS.

2.4.13. PLA observada versus PLA estimada

A avaliação do modelo da PLA foi realizada por meio da relação entre os dados estimados e os dados observados em campo. A técnica utilizada para essa verificação foi a regressão linear. A fórmula da regressão linear que foi estimada para validar os valores da umidade do solo estimada, em função dos valores observados foi representada pela Equação 34:

$$PLA_{estimada} = 3,2945 + 0,9024 PLA_{observada} \quad (34)$$

Em que:

$PLA_{estimada}$: porosidade livre de água estimada (%);

$PLA_{observada}$: porosidade livre de água observada (%).

O valor do coeficiente de determinação deste modelo foi $R^2 = 0,902$, e o valor do coeficiente de determinação ajustado foi $R_a^2 = 0,898$, ou seja, 89,8% das variações na estimativa da PLA foram explicadas pela variação dos valores da variável independente. O restante, 10,2% são fontes de variações que são explicadas por outros fatores (Tabela 2.25).

Tabela 2.25: Coeficientes métricas para a determinação do modelo da PLA, N = 24.

Estatísticas	Coeficientes
R-squared	0,902
Adj. R-squared	0,898
F-statistic	203,3
Prob (F-statistic)	$1,36e^{-12}$
Log-Likelihood	-58,224
AIC	120,4
BIC	122,8

R-squared = Coeficiente de Determinação; Adj. R-squared = Coeficiente de Determinação ajustado; F-statistic = valor do teste de Fischer – Snedecor; Prob (F-statistic) = p-valor do teste F; Log-Likelihood = probabilidade natural logarítmica; AIC = Critério de Informação de Akaike; BIC = Critério Bayesiano de Schwarz.

A realização da verificação dos resultados obtidos para a determinação da equação foi feita por meio do teste F e a partir do seu resultado foi obtido o p-valor,

onde foram encontrados os valores 203,3 e $1,36e^{-12}$, respectivamente (Tabela 2.25), indicando que a variável independente teve uma alta relação com a PLA estimada. Também fizeram o teste de Akaike (AIC), cujo o valor observado foi de 120,4. Já o valor obtido pelo Critério Bayesiano de Schwarz (BIC) foi de 122,8 (Tabela 2.25). Para os dois parâmetros avaliativos foram constatadas fortes interações da variável independente com a variável dependente, evidenciando a forte interação da PLA observada com a PLA estimada.

Analisando os coeficientes e parâmetros de qualidade da regressão linear, verificaram que o valor do coeficiente linear foi igual a 3,2945 e apresentou o p-valor = 0,151, mostrando que este coeficiente não interferiu significante na determinação do modelo. Porém quando foi estudado o coeficiente angular, obtiveram o valor de 0,9024 e 0,000, respectivamente, inferindo que o coeficiente angular apresentou uma relação significativa ao nível de 5% com a PLA do solo estimada (Tabela 2.26).

Tabela 2.26: Demonstrações do coeficiente linear e angular, resultados do p-valor em função da estatística t para o nível de significância $\alpha = 0,05$ com os seus respectivos valores críticos, para as variáveis PLA observada, versus PLA estimada.

	<i>Coeficientes</i>	<i>Erro Padrão</i>	<i>Resultado do teste t</i>	<i>p-valor</i>	<i>IC_{5%} Limite Inferior</i>	<i>IC_{95%} Limite Superior</i>
Constante	3,2945	2,214	1,488	0,151 ^{NS}	-1,296	7,885
PLA (%)	0,9024	0,063	14,259	0,000*	0,771	1,034

Erro Padrão = Standard error ou erro padrão; Resultado teste t = resultado do teste de hipótese da média T Student; * = significativo no teste T a 5% de probabilidade; NS = não significativo no teste T a 5% de probabilidade; IC_{5%} Limite Inferior = Distribuição, razão de prevalência e intervalo de confiança de 5%; IC_{95%} Limite superior = Distribuição, razão de prevalência e intervalo de confiança de 95%;

Na Figura 2.44 segue a representação gráfica dos dados da PLA observada versus PLA estimada. Constataram que o ângulo da reta, com relação ao ajuste linear foi praticamente de 45 graus, evidenciando a forte relação entre a PLA observada e a PLA estimada. Além disso, para a RMSE e a MSE foram obtidos os valores de 2,88 e 8,31, respectivamente, indicando a existência de uma alta correlação entre os valores observados e estimados.

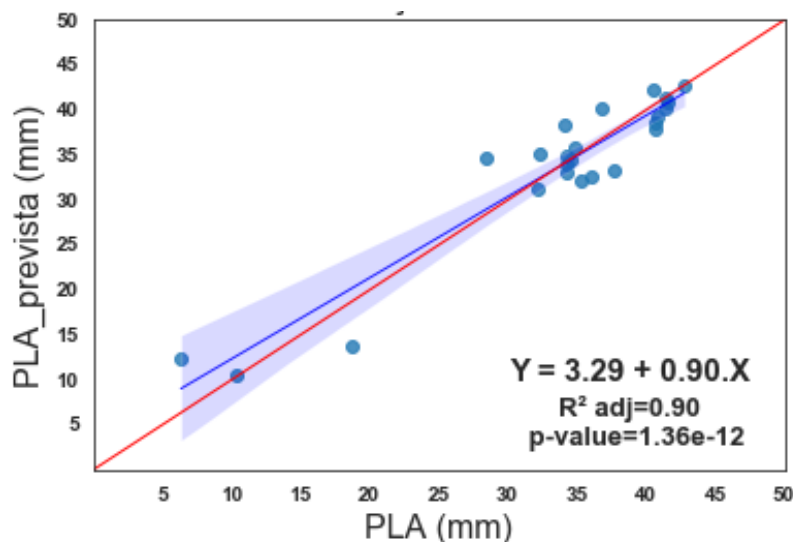


Figura 2.44: Porosidade livre d'água do solo observado (PLA) versus PLA estimado pela equação de regressão.

2.4.14. Predição do modelo final

2.4.15. Distribuição dos dados e Estatística Descritiva das variáveis preditas

Inicialmente, ao nível de significância (α) de 5% foram realizadas as análises das variabilidades gerais dos dados previstos por meio dos testes de normalidade e homoscedasticidade, pelos testes Shapiro-Wilk e Levene, respectivamente. Também foi gerado um histograma de barras para delimitar as suas variáveis e suas respectivas classes e frequências. Na Tabela 2.27 foram observados os resultados do teste de Shapiro-Wilk, para os dados previstos. Quando se analisou os valores do p-valor das variáveis ficou evidente que a grande maioria apresentou índices inferiores a 0,05, apontando que as variáveis possuem uma tendência de não apresentarem distribuição normal, com exceção a CAD40 mm (0,3521) que apresentou p-valor superior ao α de 5%, conforme pode ser observado na Figura 2.45 e confirmada pelo teste de homocedasticidade de Levene (p-valor = $3,69 \times 10^{-10}$).

Tabela 2.27: Teste de normalidade pelo método de Shapiro-wilk das variáveis previstas MOS_prevista, Us_prevista, PLA_prevista, CAD40_prevista, FCO₂_prevista e FCO₂ para todas as áreas estudadas.

Atributos	Estatística do teste	P-valor
MOS_prevista	0,77	8,59e ⁻⁰⁵ *
Us_prevista	0,86	0,0031*
PLA_prevista	0,75	5,29e ⁻⁰⁵ *
CAD_prevista	0,96	0,3521 ^{NS}
FCO₂_prevista	0,81	0,0004*
FCO₂	0,85	0,0019*

N = 24; Métricas seguidas pelo * foi significativa ao nível de 5% de probabilidade, pelo teste *Shapiro-Wilk*; MOS_prevista = matéria orgânica do solo prevista (g dm⁻³); Us_prevista = umidade do solo prevista (% do volume); PLA_prevista = porosidade livre de água prevista (%); CAD = capacidade de água disponível prevista (mm); FCO₂_prevista = Fator de emissão de CO₂ (μmol CO₂ m⁻² s⁻¹); FCO₂ = Fator de emissão de CO₂ (μmol CO₂ m⁻² s⁻¹).

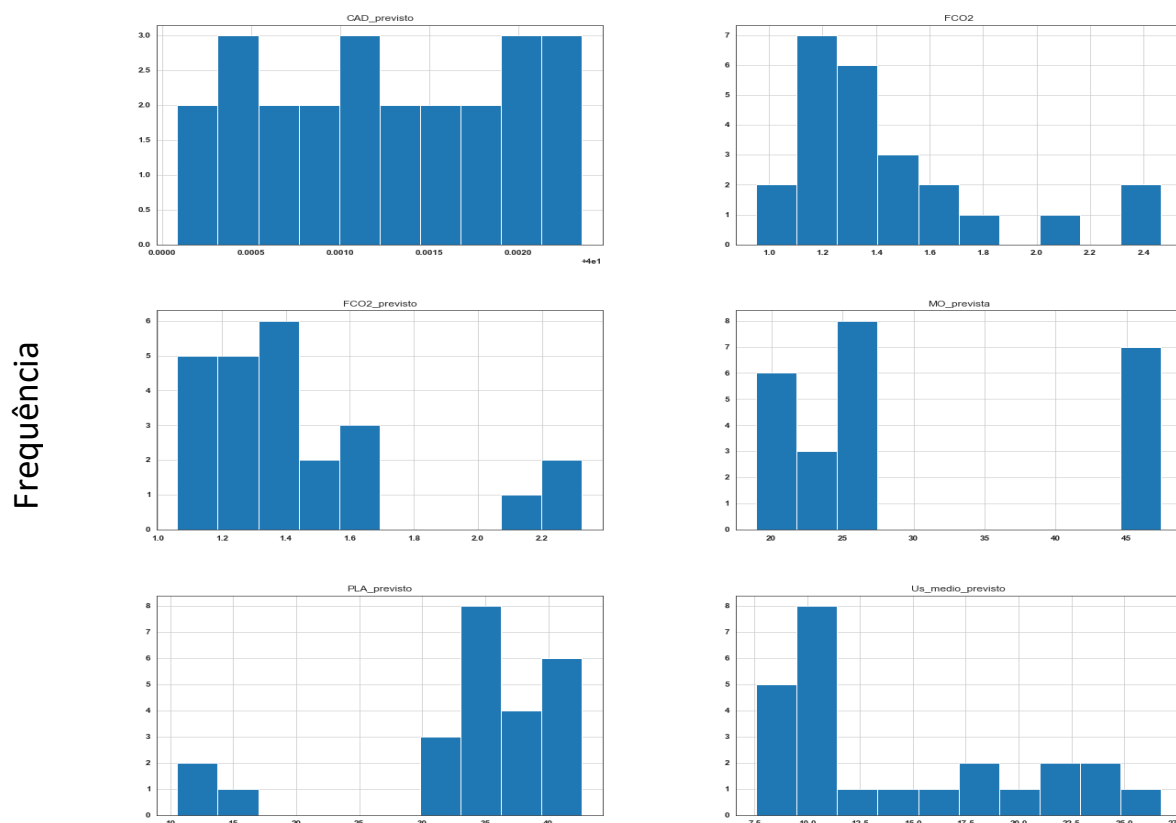


Figura 2.45: Histograma das variáveis previstas MOS_prevista, Us_prevista, PLA_prevista, CAD40_prevista, FCO₂_prevista e FCO₂ para todas as áreas estudadas.

N = 24; MOS_prevista = matéria orgânica do solo prevista (g dm⁻³); Us_prevista = umidade do solo prevista (% do volume); PLA_prevista = porosidade livre de água prevista (%); CAD = capacidade de água disponível prevista (mm); FCO₂_prevista = Fator de emissão de CO₂ (μmol CO₂ m⁻² s⁻¹); FCO₂ = Fator de emissão de CO₂ (μmol CO₂ m⁻² s⁻¹).

Os resultados da estatística descritiva, expressos na Tabela 2.28, mostraram que a maioria das variáveis apresentaram valores de coeficiente de variação inferiores a 30% (PLA_prevista = 26,52%, CAD_prevista = 0,00%, FCO₂_prevista = 24,52% e FCO₂ = 27,01%), evidenciando que os dados possuíram uma baixa ou média dispersão, implicando na conclusão que as variáveis tendem a ser homogêneas. No entanto, para as variáveis MOS_prevista (37,35%) e Us_prevista (41,03%) foram observados valores de CV superiores a 30%, indicando que os dados tendem a ter uma alta dispersão (heterogêneos).

Tabela 2.28: Estatísticas descritivas das variáveis previstas MOS_prevista, Us_prevista, PLA_prevista, CAD40_prevista, FCO₂_prevista e FCO₂ para todas as áreas estudadas, N = 24.

Atributos	Média	Desvpd	Máx / Mín	CV
MOS_prevista	30,67	11,46	47,42 / 18,93	37,35
Us_prevista	14,27	5,86	26,75 / 7,58	41,03
PLA_prevista	33,74	8,95	10,57 / 42,68	26,52
CAD_prevista	40,00	7,07 e ⁻⁰⁴	40,00 / 40,00	1,77 e ⁻⁰³
FCO₂_prevista	1,44	0,35	2,32 / 1,06	24,52
FCO₂	1,44	0,39	2,46 / 0,95	27,01

Desvpd = desvio padrão; Máx = Máximo; Mín = Mínimo; CV = coeficiente de variação (%); MOS_prevista = matéria orgânica do solo prevista (g dm⁻³); Us_prevista = umidade do solo prevista (% do volume); PLA_prevista = porosidade livre de água prevista (%); CAD = capacidade de água disponível prevista (mm); FCO₂_prevista = Fator de emissão de CO₂ (μmol CO₂ m⁻² s⁻¹); FCO₂ = Fator de emissão de CO₂ (μmol CO₂ m⁻² s⁻¹).

Estudando de maneira mais específica as variáveis preditas no modelo, observou-se que para a FCO₂_prevista a variação dos valores foram entre 1,06 e 2,32 μmol CO₂ m⁻² s⁻¹, sendo que os maiores valores observados foram na área de Motuca e os menores em Pradópolis (Tabelas 2.28 e 2.29). Já para a MOS_prevista, observou-se que dentro das áreas estudadas houve uma variação de 18,93 a 47,42 g dm⁻³. No município de Aparecida do Taboado, foi observado os menores valores (18,93 g dm⁻³) e os maiores valores foram em Guariba (47,42 g dm⁻³). Com relação à PLA_prevista, foram mensurados valores entre 10,57% e 42,68%, sendo que os maiores índices foram registrados em Pradópolis e os menores em Motuca. Dessa forma, foi possível observar que ou houve uma forte influência das variáveis físicas e químicas do solo na FCO₂, impulsionadas principalmente pela PLA e MOS. Da Silva

e t al. (2020) também observaram que as mesmas variáveis tiveram uma grande influência no processo de emissão e estoque de carbono no solo.

Tabela 2.29: Estatísticas descritivas das variáveis previstas MOS_prevista, Us_prevista, PLA_prevista, CAD40_prevista, FCO₂_prevista e FCO₂, referentes às áreas de Motuca – SP (N = 3), Guariba – SP (N = 7), Pradópolis – SP (N = 8) e Aparecida do Taboado (N = 6).

<i>Atributos</i>	<i>Motuca</i>				<i>Guariba</i>			
	<i>Média</i>	<i>Desvpd</i>	<i>Máx / Mín</i>	<i>CV</i>	<i>Média</i>	<i>Desvpd</i>	<i>Máx / Mín</i>	<i>CV</i>
<i>MO_prevista</i>	23,72	0,00	23,72 / 23,72	2,59e ⁻¹⁴	47,42	5,02e ⁻¹⁵	47,42 / 47,42	1,06e ⁻¹⁴
<i>Us_prevista</i>	24,63	1,85	26,75 / 23,28	7,53	18,52	2,94	22,04 / 13,78	15,88
<i>PLA_prevista</i>	12,29	1,65	13,85 / 10,57	13,41	35,10	3,25	40,15 / 31,19	9,26
<i>CAD_prevista</i>	40,00	0,00	40,00 / 40,00	3,11e ⁻⁰⁴	40,00	2,32e ⁻⁰⁴	40,00 / 40,00	5,80e ⁻⁰⁴
<i>FCO₂_prevista</i>	2,26	0,07	2,32 / 2,19	2,93	1,50	0,13	1,65 / 1,30	8,71
<i>FCO₂</i>	2,30	0,18	2,46 / 2,11	7,72	1,48	0,19	1,84 / 1,30	12,81
<i>Atributos</i>	<i>Pradópolis</i>				<i>Aparecida do Taboado</i>			
	<i>Média</i>	<i>Desvpd</i>	<i>Máx / Mín</i>	<i>CV</i>	<i>Média</i>	<i>Desvpd</i>	<i>Máx / Mín</i>	<i>CV</i>
<i>MO_prevista</i>	27,43	5,00e ⁻¹⁵	27,43 / 27,43	1,96e ⁻¹⁴	18,93	1,18e ⁻¹⁴	18,93 / 18,93	6,22e ⁻¹⁴
<i>Us_prevista</i>	10,70	0,50	11,79 / 10,09	4,64	8,90	1,00	10,44 / 7,58	11,19
<i>PLA_prevista</i>	40,41	1,67	42,68 / 37,92	4,14	33,99	0,89	35,04 / 32,64	2,63
<i>CAD_prevista</i>	40,00	2,00e ⁻¹⁴	40,00 / 40,00	5,99e ⁻⁰⁴	40,00	1,58e ⁻⁰⁴	40,00 / 40,00	3,94e ⁻⁰⁴
<i>FCO₂_prevista</i>	1,15	0,07	1,25 / 1,06	5,82	1,35	0,04	1,41 / 1,31	2,64
<i>FCO₂</i>	1,20	0,14	1,39 / 0,95	11,58	1,29	0,22	1,59 / 1,06	17,09

Desvpd = desvio padrão; Máx = Máximo; Mín = Mínimo; CV = coeficiente de variação (%); MOS_prevista = matéria orgânica do solo prevista (g dm⁻³); Us_prevista = umidade do solo prevista (% do volume); PLA_prevista = porosidade livre de água prevista (%); CAD = capacidade de água disponível prevista (mm); FCO₂_prevista = Fator de emissão de CO₂ (μmol CO₂ m⁻² s⁻¹); FCO₂ = Fator de emissão de CO₂ (μmol CO₂ m⁻² s⁻¹).

Analisando as demais variáveis preditas, para a Us_prevista os obtidos ficaram entre 7,58 e 26,75%, sendo que os maiores índices ocorreram na área de Motuca e os menores em Aparecida do Taboado (Tabela 2.28 e 2.29). Já para a CAD_prevista houve uma variação de 40,000085 a 40,002358 mm, em que o município de Motuca foi observado os menores valores (40,000085 mm) e os maiores valores foram em Aparecida do Taboado (40,002358 mm).

2.4.16. Correlações das variáveis preditas

Devido à complexidade do processo, foram consideradas na discussão as correlações menores que -0,50 e maiores que 0,50. A emissão de CO₂ é oriunda de um processo complexo, resultante dos atributos físicos, químicos, biológicos e

climáticos (Tavares et al., 2016), onde pode-se destacar a umidade (Costa et al., 2017), o teor de matéria orgânica, densidade e porosidade do solo (Brito et al., 2009), PLA e na CAD do solo.

Para a FCO_2 foram observadas fortes correlações negativas com a PLA_prevista (-0,89) e com a CAD_prevista (-0,68), e positivas com a Us (0,79) e FCO_2 _prevista (0,91), conforme pode ser observado na Tabela 2.30 e Figura 2.46, mostrando que as variáveis físicas do solo impactaram fortemente no deslocamento do CO_2 no solo. Lal (2009) comentou que a umidade do solo, a atmosfera e a temperatura estão intrinsicamente ligadas aos processos de produção e transporte do CO_2 no solo. Semelhantes índices foram encontrados para a FCO_2 _prevista. As correlações fortes e positivas ocorreram com as variáveis FCO_2 (0,91) e com a Us_prevista (0,86). Já as correlações negativas foram observadas com a PLA_prevista (-0,98) e com a CAD_prevista (-0,70). A MO_prevista apresentou baixa correlação tanto com a FCO_2 (0,03), quanto com a FCO_2 _prevista (0,03).

Tabela 2.30: Correlações de Pearson das variáveis previstas MOS_prevista, Us_prevista, PLA_prevista, CAD40_prevista, FCO_2 _prevista e FCO_2 para todas as áreas estudadas, N = 24.

	<i>FCO₂</i>	<i>MO_prevista</i>	<i>Us_prevista</i>	<i>PLA_prevista</i>	<i>CAD_prevista</i>	<i>FCO₂_prevista</i>
<i>FCO₂</i>	1,00	0,03	0,79	-0,89	-0,68	0,91
<i>MO_prevista</i>	0,03	1,00	0,49	0,18	-0,60	0,03
<i>Us_prevista</i>	0,79	0,49	1,00	-0,74	-0,93	0,86
<i>PLA_prevista</i>	-0,89	0,18	-0,74	1,00	0,56	-0,98
<i>CAD_prevista</i>	-0,68	-0,60	-0,93	0,56	1,00	-0,70
<i>FCO₂_prevista</i>	0,91	0,03	0,86	-0,98	-0,70	1,00

FCO_2 = Fator de emissão de CO_2 ($\mu\text{mol } CO_2 \text{ m}^{-2} \text{ s}^{-1}$); MOS_prevista = matéria orgânica do solo prevista (g dm^{-3}); Us_prevista = umidade do solo prevista (% do volume); PLA_prevista = porosidade livre de água prevista (%); CAD = capacidade de água disponível prevista (mm); FCO_2 _prevista = Fator de emissão de CO_2 ($\mu\text{mol } CO_2 \text{ m}^{-2} \text{ s}^{-1}$).

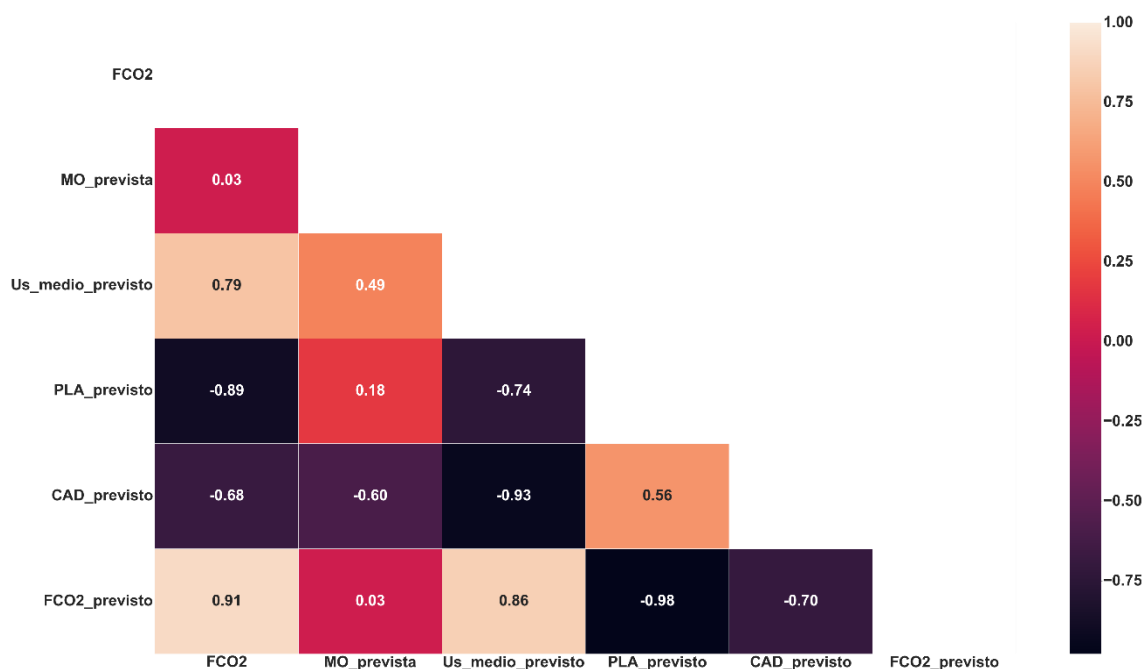


Figura 2.46: Correlações de Pearson das variáveis previstas. FCO₂ = Fator de emissão de CO₂ ($\mu\text{mol CO}_2 \text{ m}^{-2} \text{ s}^{-1}$); MOS_prevista = matéria orgânica do solo prevista (g dm^{-3}); Us_prevista = umidade do solo prevista (% do volume); PLA_prevista = porosidade livre de água prevista (%); CAD = capacidade de água disponível prevista (mm); FCO₂_prevista = Fator de emissão de CO₂ ($\mu\text{mol CO}_2 \text{ m}^{-2} \text{ s}^{-1}$), para todas as áreas estudadas, N = 24.

De maneira geral, a MO_prevista apresentou fracas correlações com praticamente todas as variáveis preditas (FCO₂ = 0,03; Us_prevista = 0,49; PLA_prevista = 0,18; FCO₂_prevista = 0,03), exceto com a CAD_prevista, onde foi registrado uma correlação negativa de -0,6 (Tabela 2.30 e Figura 2.46).

A Us_prevista também apresentou fortes correlações com praticamente todas as variáveis, exceto com a MO_prevista (0,49). As correlações positivas foram com a FCO₂ (0,79) e FCO₂_prevista (0,86), indicando que a umidade colabora positivamente com o transporte e emissão do CO₂ no solo. Já as correlações negativas com a PLA_prevista (-0,74) e com a CAD_prevista (-0,93). Panosso et al. (2009) também obtiveram correlações positivas entre a emissão e umidade do solo, nos solos descobertos, logo após irrigação artificial (Tabela 2.30 e Figura 2.43).

Semelhante a variável anterior, a PLA_prevista apresentou fortes correlações com praticamente todas as variáveis preditas, com exceção a MO_prevista (0,18). A

CAD_prevista foi a única variável que apresentou uma correlação positiva e forte (0,56), já as variáveis FCO₂, Us_prevista e FCO₂_prevista apresentaram as correlações negativas e fortes, cujo os valores foram -0,89; -0,74 e -0,98; respectivamente (Tabela 2.30 e Figura 2.46).

Para a CAD_prevista, todas as variáveis apresentaram fortes correlações. As variáveis que apresentaram correlações negativas foram a FCO₂ (-0,68), MO_prevista (-0,60), Us_prevista (-0,93), e FCO₂_prevista (-0,70). A única correlação que foi positiva foi a da PLA_prevista cujo o valor foi 0,56 (Tabela 2.30 e Figura 2.46).

2.4.17. Regressão Linear Múltipla da FCO₂, em função das variáveis PLA_prevista e MOS_prevista

Nos últimos anos, diversas pesquisas científicas foram desenvolvidas com o intuito de descrever o potencial da agricultura no processo de mitigação das emissões de gases do efeito estufa. Entretanto, foram desenvolvidos poucos experimentos e modelos matemáticos robustos que conseguiram estimar o efeito de fatores bióticos e abióticos no balanço do carbono no solo. Para este estudo foi usada a técnica da regressão linear múltipla. Logo abaixo segue a equação da regressão linear múltipla que foi estimada para explicar a FCO₂, em função das variáveis PLA_prevista e MOS_prevista que foi representada pela Equação 35.

$$FCO_2 = 2,5938 + 0,0065MOS_{prevista} - 0,0401PLA_{prevista} \quad (35)$$

Em que:

FCO₂: Fator de emissão de CO₂ (μmol CO₂ m⁻² s⁻¹).

MOS_prevista: matéria orgânica do solo prevista (g dm⁻³);

PLA_prevista: porosidade livre de água prevista (%).

O valor do coeficiente de determinação do modelo foi de $R^2 = 0,824$, e o seu respectivo valor ajustado foi $R_a^2 = 0,807$, ou seja, 80,7% das variações na estimativa da FCO₂ podem ser explicadas pelas variações dos valores das variáveis independentes. O restante, 19,3% são fontes de variações que são explicados por outros fatores (Tabela 2.31).

Tabela 2.31: Coeficientes métricos do modelo da FCO₂, N = 24.

Estatísticas	Coeficientes
R-squared	0,824
Adj. R-squared	0,807
F-statistic	49,14
Prob (F-statistic)	1,20e ⁻⁰⁸
Log-Likelihood	9,9401
AIC	-13,88
BIC	-10,35

R-squared = Coeficiente de Determinação; Adj. R-squared = Coeficiente de Determinação ajustado; F-statistic = valor do teste de Fischer – Snedecor; Prob (F-statistic) = p-valor do teste F; Log-Likelihood = probabilidade natural logarítmica; AIC = Critério de Informação de Akaike; BIC = Critério Bayesiano de Schwarz.

A qualidade do modelo foi validada por meio do teste de hipóteses, em que H_0 : $\beta_1 = \beta_2 = \beta_3 = 0$ ou H_1 : pelo menos um β é diferente dos demais. Por meio da análise de variância foram realizados os testes F e mensurado p-valor, onde encontraram os valores de 49,14 e 1,20e⁻⁰⁸, respectivamente. Esses valores evidenciaram que foi rejeitada a hipótese nula que todos os coeficientes independentes seriam iguais, ao nível de significância de 0,05, mostrando que pelo menos uma das variáveis independentes apresentaram relação com a FCO₂. Além disso, também foi realizado o teste de Akaike (AIC), cujo o valor obtido foi -13,88. Já o valor encontrado pelo Critério Bayesiano de Schwarz (BIC) foi -10,35 (Tabela 2.31). Os dois parâmetros avaliativos mostraram fortes interações entre as variáveis independentes e a dependente, mostrando uma alta relação da MOS_prevista e PLA_prevista com a FCO₂.

Quando analisaram o valor do coeficiente linear do modelo da FCO₂ e seu p-valor obtiveram os valores de 2,5938 e 0,000, respectivamente, evidenciando que este parâmetro apresentou significância estatística no modelo (Tabela 2.32). O valor positivo do coeficiente linear confirmou a influência dos coeficientes angulares e mostrou que inicialmente o solo tem um potencial de FCO₂ natural de 2,5938 $\mu\text{mol CO}_2 \text{ m}^{-2} \text{ s}^{-1}$.

Tabela 2.32: Demonstrações dos coeficientes lineares e angulares, resultados do p-valor em função da estatística t para o nível de significância $\alpha = 0,05$ com os seus respectivos valores críticos para as variáveis MOS_prevista e PLA_prevista que foram usadas na determinação do modelo da FCO₂, N = 24.

	<i>Coeficientes</i>	<i>Erro Padrão</i>	<i>Resultado do teste t</i>	<i>p-valor</i>	<i>IC_{5%} Limite Inferior</i>	<i>IC_{95%} Limite Superior</i>
Constante	2,5938	0,157	16,542	0,000*	2,268	2,920
MOS_prevista	0,0065	0,003	2,065	0,051 ^{NS}	-4,54e ⁻⁰⁵	0,013
PLA_prevista	-0,0401	0,004	-9,908	0,000*	-0,049	-0,032

Erro Padrão = Standard error ou erro padrão; Resultado teste t = resultado do teste de hipótese da média T Student; * = significativo no teste T a 5% de probabilidade; NS = não significativo no teste T a 5% de probabilidade; IC_{5%} Limite Inferior = Distribuição, razão de prevalência e intervalo de confiança de 5%; IC_{95%} Limite superior = Distribuição, razão de prevalência e intervalo de confiança de 95%; FCO₂ = Fator de emissão de CO₂ ($\mu\text{mol CO}_2 \text{ m}^{-2} \text{ s}^{-1}$); MOS_prevista = matéria orgânica do solo prevista (g dm^{-3}); PLA_prevista = porosidade livre de água prevista (%).

Para os coeficientes angulares obtiveram os seguintes resultados: MOS_prevista = 0,0065 e PLA_prevista = -0,0401. Os valores de p-valor para as variáveis foram 0,051 e 0,000, respectivamente (Tabela 2.32). A PLA_prevista apresentou forte poder explicativo dentro do modelo, impactando ao nível de significância de 5%, já a MOS_prevista não apresentou valor significativo ao nível de 5%, porém para outro conjunto de dados pode haver maior poder explicativo.

A estatística descritiva e os gráficos boxplots foram usados para explicar a variabilidade das variáveis utilizadas no modelo. Nos gráficos das Figuras 2.47 e 2.48 e nas Tabelas 2.28 e 2.29 foram registrados os valores da PLA_prevista. A média geral da PLA_prevista observada foi 33,74% e apresentou desvio padrão de 8,95%. Avaliando os locais de estudo individualmente, Motuca registrou o menor valor (10,57%) e Pradópolis o maior (42,68%), dentre todas as áreas estudadas. Com relação aos valores de coeficiente de variação, observaram que para todos os dados das áreas estudadas o valor de 26,52%, indicando uma moderada amplitude dos dados da PLA_prevista. No entanto, quando se analisou cada local de estudo foram

observados valores inferiores a 15%, sendo que Aparecida do Taboado apresentou o menor valor (2,63%), seguido de Pradópolis (4,14%), Guariba (9,26%) e Motuca (13,41%).

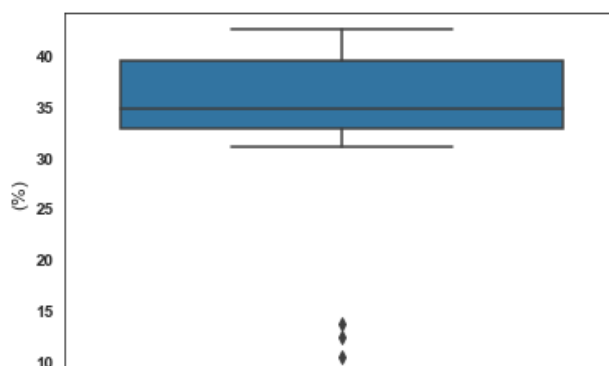


Figura 2.47: Boxplot da porosidade livre de água prevista (PLA_prevista, %) para todas as áreas estudadas, N = 24.

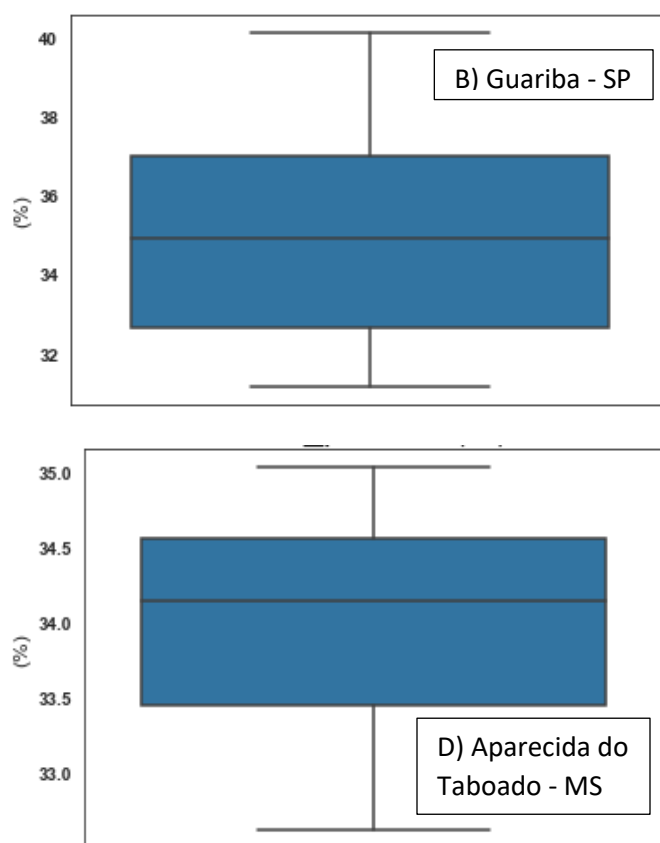
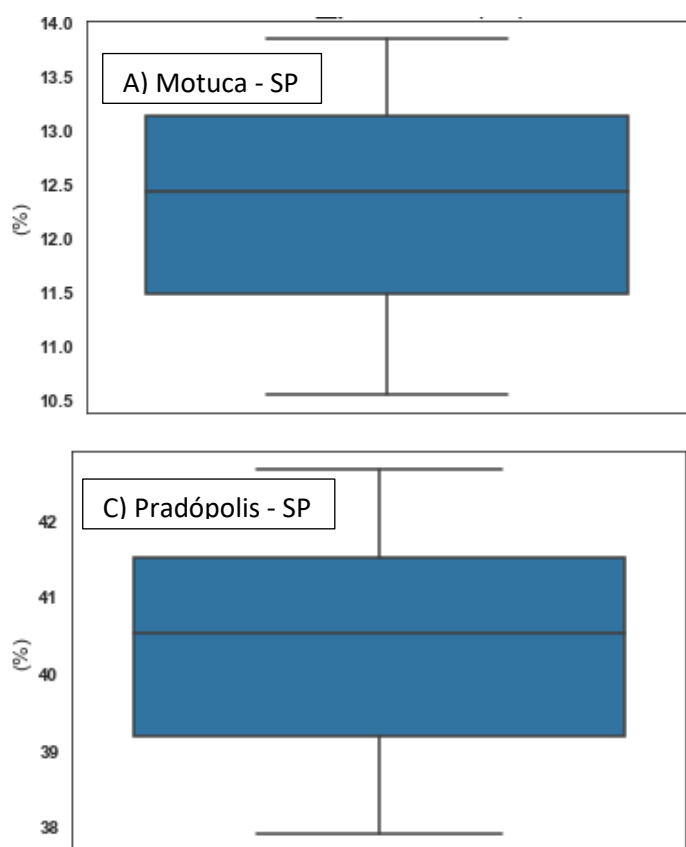


Figura 2.48: Boxplots da porosidade livre de água prevista (PLA_prevista, %) para os municípios A) Motuca – SP, B) Guariba – SP, C) Pradópolis – SP e D) Aparecida do Taboado – MS.

Outra variável que foi analisada e utilizada na construção do modelo foi a MOS_prevista. Na Figura 2.49 e as Tabelas 2.28 e 2.29 observa-se que a média geral da MOS_prevista foi $30,67 \text{ g dm}^{-3}$ e desvio padrão $11,46 \text{ g dm}^{-3}$. Quando se analisou cada área de estudo, foi observado os maiores valores em Guariba ($47,42 \text{ g dm}^{-3}$) e os menores em Aparecida do Taboado ($18,93 \text{ g dm}^{-3}$), dentre todas as áreas estudadas. Quando analisaram os valores dos coeficientes de variação para todas as áreas estudadas, foram observados uma alta variabilidade, registrando o valor de 37,35%. Quando foi analisado cada local de estudo a variação foi de 0%, mostrando que não ocorreu variação dessa variável *in loco*.

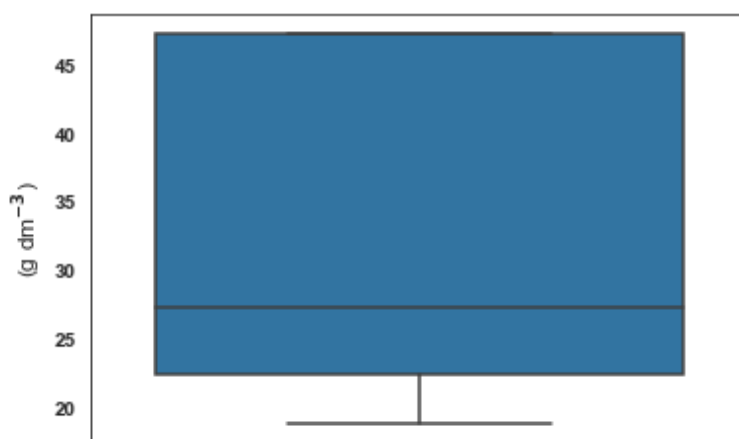


Figura 2.49: Boxplot da matéria orgânica do solo prevista (MOS_prevista, g dm^{-3}) para todas as áreas estudadas, N = 24.

A variabilidade da principal variável, a FCO_2 , está representada nas Figuras 2.50 e 2.51 e as Tabelas 2.28 e 2.29. A média geral da FCO_2 foi $1,44 \mu\text{mol CO}_2 \text{ m}^{-2} \text{ s}^{-1}$ e o desvio padrão foi $0,39 \mu\text{mol CO}_2 \text{ m}^{-2} \text{ s}^{-1}$. Estudando cada local de estudo, observaram que os menores valores foram obtidos em Pradópolis ($0,95 \mu\text{mol CO}_2 \text{ m}^{-2} \text{ s}^{-1}$), já em Motuca observaram os maiores valores ($2,46 \mu\text{mol CO}_2 \text{ m}^{-2} \text{ s}^{-1}$). Analisando os valores dos coeficientes de variação, foram observados uma moderada variabilidade para todos os dados das áreas estudadas, apresentando o valor de 27,01%. Quando foi estudado cada local, observaram baixas dispersões para Motuca (7,72%), Pradópolis (11,58%) e Guariba (12,81%). Já Aparecida do Taboado apresentou moderada dispersão dos dados (17,09%).

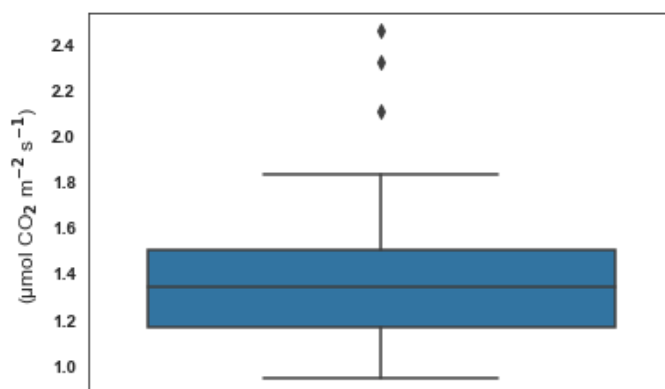


Figura 2.50: Boxplot da emissão do CO₂ do solo (FCO₂, $\mu\text{mol CO}_2 \text{ m}^{-2} \text{ s}^{-1}$) para todas as áreas estudadas, N = 24.

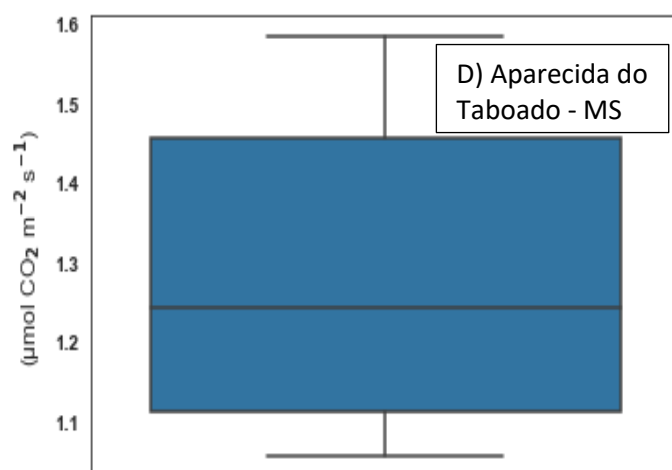
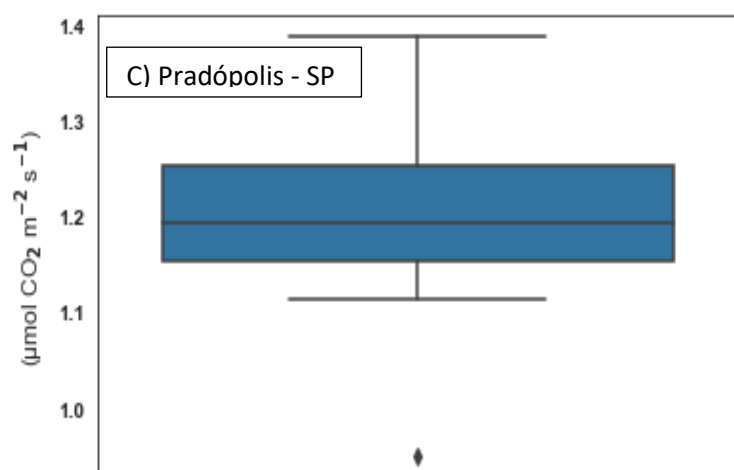
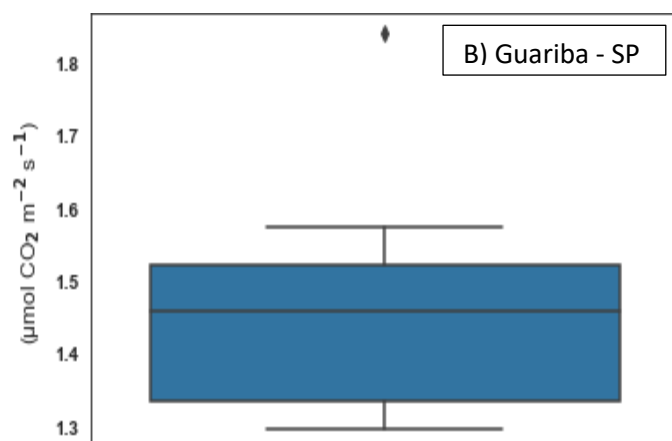
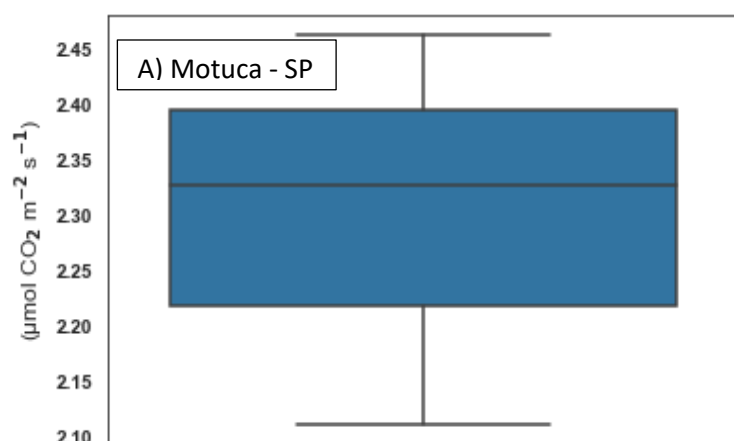


Figura 2.51: Boxplots da emissão do CO₂ do solo (FCO₂, $\mu\text{mol CO}_2 \text{ m}^{-2} \text{ s}^{-1}$) para os municípios de A) Motuca – SP, B) Guariba – SP, C) Pradópolis – SP e D) Aparecida do Taboado – MS.

2.4.18. FCO₂ observada versus FCO₂ prevista

A avaliação do modelo da FCO₂ foi realizada por meio da relação entre os dados estimados e os dados observados em campo. A técnica utilizada para essa verificação foi a regressão linear. A fórmula da regressão linear que foi estimada para validar os valores da FCO₂ estimada, em função dos valores observados foi representada pela Equação 36.

$$FCO_{2_estimada} = 0,2537 + 0,8239 FCO_{2_observada} \quad (36)$$

Em que:

FCO_{2_estimada}: emissão de CO₂ do solo estimada (μmol CO₂ m⁻² s⁻¹);

FCO_{2_observada}: emissão de CO₂ do solo observada (μmol CO₂ m⁻² s⁻¹).

O valor do coeficiente de determinação deste modelo foi $R^2 = 0,824$, e o valor do coeficiente de determinação ajustado foi $R_a^2 = 0,816$, implicando em dizer que 81,6% das variações na estimativa da FCO₂ foram explicadas pelas variações dos valores da variável independente. O restante, 18,4% são fontes de variações que são explicadas por outros fatores (Tabela 2.33).

Tabela 2.33: Coeficientes métricas para a determinação do modelo da FCO₂, N = 24.

Estatísticas	Coeficientes
R-squared	0,824
Adj. R-squared	0,816
F-statistic	103,0
Prob (F-statistic)	9,26e ⁻¹⁰
Log-Likelihood	12,264
AIC	-20,53
BIC	-18,17

R-squared = Coeficiente de Determinação; Adj. R-squared = Coeficiente de Determinação ajustado; F-statistic = valor do teste de Fischer – Snedecor; Prob (F-statistic) = p-valor do teste F; Log-Likelihood = probabilidade natural logarítmica; AIC = Critério de Informação de Akaike; BIC = Critério Bayesiano de Schwarz.

A realização da verificação dos resultados obtidos para a determinação da equação foi feita por meio do teste F e a partir do seu resultado foi obtido o p-valor,

onde foram encontrados os valores 103,0 e $9,26e^{-10}$, respectivamente, mostrando que a variável independente teve uma alta relação com a FCO_2 estimada. No teste de Akaike (AIC) o valor encontrado foi de -20,53. Já o valor registrado pelo Critério Bayesiano de Schwarz (BIC) foi -18,17 (Tabela 2.33). Para os dois parâmetros avaliativos foram constatadas fortes interações da variável independente com a variável dependente, evidenciando a forte interação da FCO_2 observada com a FCO_2 estimada.

Analisando os coeficientes e parâmetros de qualidade da regressão linear, verificaram que o valor coeficiente linear foi igual a 0,25437 e apresentou o p-valor = 0,048. Já para o coeficiente angular obtiveram os valores de 0,9024 e 0,000, respectivamente. Os dois coeficientes interferiram significante na determinação do modelo ao nível de significância de 5% (Tabela 2.34).

Tabela 2.34: Demonstrações do coeficiente linear e angular, resultados do p-valor em função da estatística t para o nível de significância $\alpha = 0,05$ com os seus respectivos valores críticos, para as variáveis PLA observada, versus PLA estimada.

	<i>Coeficientes</i>	<i>Erro Padrão</i>	<i>Resultado do teste t</i>	<i>p-valor</i>	<i>IC5% Limite Inferior</i>	<i>IC95% Limite Superior</i>
Constante	0,2537	0,121	2,096	0,048*	0,003	0,505
FCO_2	0,8239	0,081	10,146	0,000*	0,656	0,992

Erro Padrão = Standard error ou erro padrão; Resultado teste t = resultado do teste de hipótese da média T Student; * = significativo no teste T a 5% de probabilidade; NS = não significativo no teste T a 5% de probabilidade; IC5% Limite Inferior = Distribuição, razão de prevalência e intervalo de confiança de 5%; IC95% Limite superior = Distribuição, razão de prevalência e intervalo de confiança de 95%; FCO_2 = emissão de CO_2 do solo ($\mu\text{mol } CO_2 \text{ m}^{-2} \text{ s}^{-1}$).

Na Figura 2.52 segue a representação gráfica dos dados da FCO_2 observada versus FCO_2 estimada. Constataram que o ângulo da reta, com relação ao ajuste linear foi praticamente de 45 graus, evidenciando a forte relação entre a FCO_2 observada e a FCO_2 estimada. Além disso, para a RMSE e a MSE foram obtidos os valores de 0,03 e 0,16, respectivamente, indicando a existência de uma alta correlação entre os valores observados e estimados.

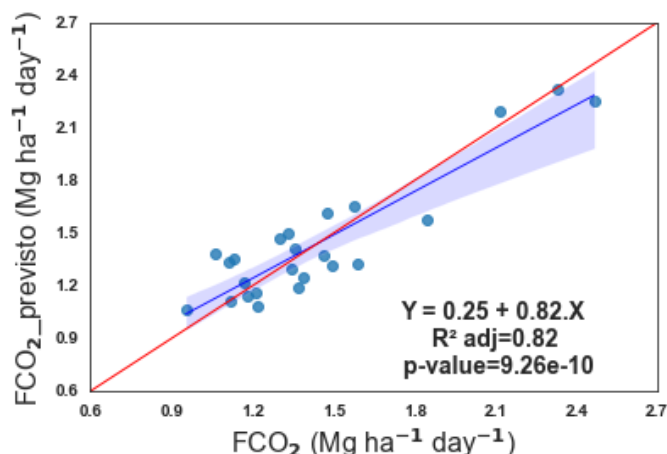


Figura 2.52: Emissão de CO₂ do solo (FCO₂) observado versus FCO₂ estimado pela equação de regressão.

Analisando a variabilidade dos valores estimados da FCO₂ prevista, observa-se que a média geral da FCO₂ foi 1,44 $\mu\text{mol CO}_2 \text{ m}^{-2} \text{ s}^{-1}$ e o desvio padrão foi 0,35 $\mu\text{mol CO}_2 \text{ m}^{-2} \text{ s}^{-1}$. Quando foi estudado cada local, observaram que os menores valores foram obtidos em Pradópolis (1,06 $\mu\text{mol CO}_2 \text{ m}^{-2} \text{ s}^{-1}$), já em Motuca observaram os maiores valores (2,32 $\mu\text{mol CO}_2 \text{ m}^{-2} \text{ s}^{-1}$), dentre todas as áreas estudadas. Analisando os valores dos coeficientes de variação, foram observados uma moderada variabilidade para todos os dados das áreas estudadas, apresentando o valor de 24,52%. Quando foi estudado cada local, foram observadas baixas dispersões para todos os municípios, sendo que Aparecida do Taboado (2,64%) apresentou o menor valor, seguido de Motuca (2,93%), Pradópolis (5,82%) e Guariba (8,71%), conforme as Figuras 2.53 e 2.54 e Tabelas 2.29 e 2.30.

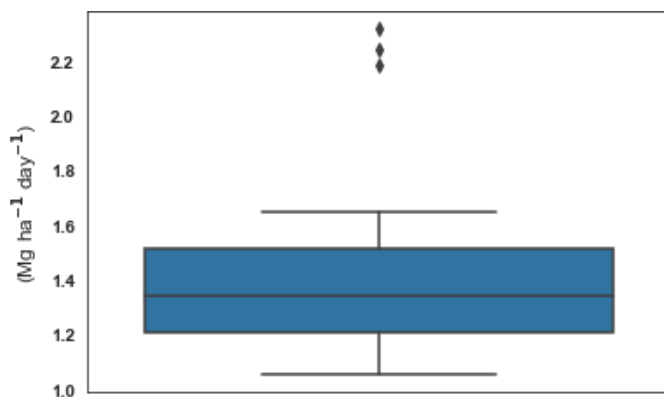


Figura 2.53: Boxplot da emissão do CO₂ do solo prevista (FCO₂_prevista, $\mu\text{mol CO}_2 \text{ m}^{-2} \text{ s}^{-1}$) para todas as áreas estudadas, N = 24.

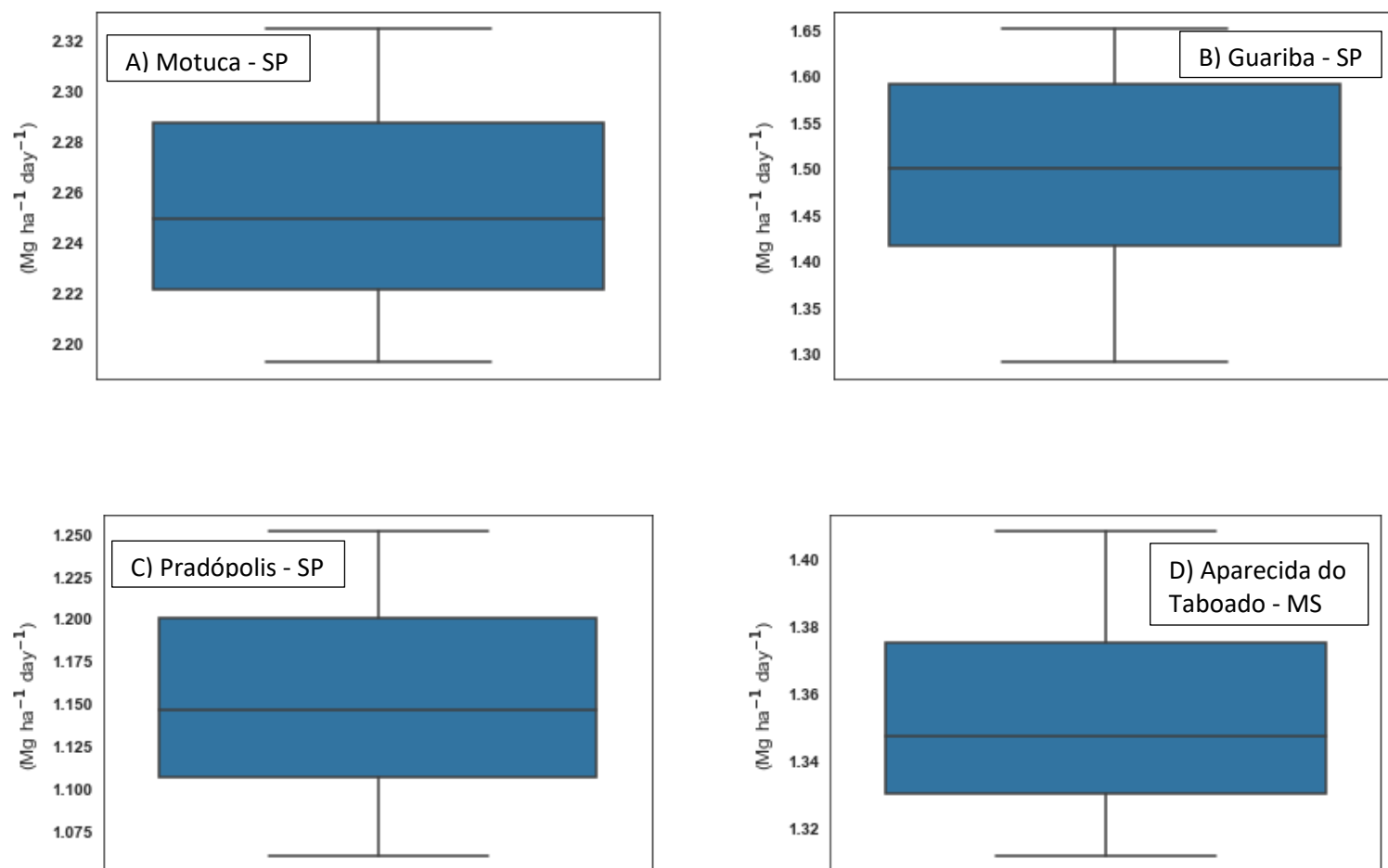


Figura 2.54: Boxplots da emissão do CO_2 do solo (FCO_2 , $\mu\text{mol CO}_2 \text{ m}^{-2} \text{ s}^{-1}$) para os municípios de A) Motuca – SP, B) Guariba – SP, C) Pradópolis – SP e D) Aparecida do Taboado – MS.

2.5. CONCLUSÕES

De maneira geral, foi observado que o processo de emissão da CO₂ do solo é um processo bastante complexo, dependente do sistema solo-planta-atmosfera que por sua vez, impactam nos atributos físicos, químicos e biológicos do solo.

Com relação as correlações, a FCO₂ apresentou fortes correlações negativas com a Dp (-0,56), Ds (-0,77), PLA (-0,92), CAD (-0,68), PLA_BH (-0,62) e a Tair (-0,90), e positivas com a ARM (0,62) e Us (0,71).

A PLA medida em campo correlacionou-se positivamente com a Dp (0,71), Ds (0,79), CAD (0,51), PLA_BH (0,53) e Tair (0,89), e negativamente com a FCO₂ (-0,92), ARM (-0,53), Us (-0,55) e ETP (-0,52).

Já para a PLA_BH, obtida através do balanço hídrico, por meio do método de Thornthwaite e Mather (1955), observou-se as correlações negativas com a FCO₂ (-0,62), MOS (-0,69), CTC (-0,69), EstC (-0,75), ARM (-1,00), Us (-0,86) e Precipitação (-0,50). Já se as correlações positivas foram com as variáveis PLA (0,53), CAD (0,85), Ts (0,54) e Tair (0,71).

Para a MOS foram observadas correlações positivas com a Dp (0,77), CTC (0,79), ARM (0,69) e US (0,57). As correlações negativas foram relacionadas com as variáveis CAD (-0,60), PLA_BH (-0,69) e Ts (-0,80).

De maneira geral, as variáveis físicas do solo interferem nos processos de deslocamento do CO₂ no solo, já as variáveis químicas no processo de formação de CO₂.

A técnica de regressão linear múltipla se mostrou eficiente na predição das variáveis estudadas, promovendo a diminuição da dependência de dados medidos em campo, possibilitando a estimativa de parâmetros que podem ser difíceis de medir ou que o custo de obtenção seja elevado. Nos estudos dos modelos preliminares foram obtidas as seguintes variáveis:

- CAD 40 mm em função da Us, Ds, Dp e ARM, apresentando um $R_a^2 = 0,999$;
- Us em função da ETP, Precipitação e ARM, apresentando um $R_a^2 = 0,889$.

Além disso, observou-se que a técnica de regressão linear múltipla foi capaz de estimar as variáveis principais utilizadas na predição da FCO₂. Para a MOS foi obtida uma equação cuja as variáveis independentes foram a P, CTC, Tair, Ts, Dp e

Us ($R_a^2 = 0,985$). Para a PLA foi obtida uma equação cuja as variáveis independentes foram a Dp e PLA_BH ($R_a^2 = 0,898$).

Depois de obter os valores preditos, chegou-se a seguinte equação da regressão linear múltipla que foi estimada para explicar a FCO₂, em função das variáveis PLA_prevista e MOS_prevista:

$$FCO_2 = 2,5938 + 0,0065MOS_{prevista} - 0,0401PLA_{prevista}$$

O valor do coeficiente de determinação do modelo foi de $R^2 = 0,824$, e o seu respectivo valor ajustado foi $R_a^2 = 0,807$, ou seja, 80,7% das variações na estimativa da FCO₂ podem ser explicadas pelas variações dos valores das variáveis independentes.

O teste F e o p-valor (49,14 e $1,20e^{-08}$, respectivamente) da análise de variância do modelo da FCO₂ mostrou que pelo menos uma das variáveis independentes apresentaram relação com a FCO₂. O valor do coeficiente linear do modelo da FCO₂ e seu p-valor foram 2,5938 e 0,000, respectivamente. Para os coeficientes angulares obtiveram os seguintes resultados: MOS_prevista = 0,0065 e PLA_prevista = -0,0401. Os valores de p-valor para as variáveis foram 0,051 e 0,000. A PLA_prevista apresentou forte poder explicativo dentro do modelo, impactando ao nível de significância de 5%, já a MOS_prevista não apresentou valor significativo ao nível de 5%, porém para outro conjunto de dados pode haver maior poder explicativo.

Estudos como este, relacionados à determinação de parâmetros do solo são importantes para os agricultores, pois as informações geradas podem ser utilizadas no gerenciamento e nas tomadas de decisões do processo produtivo e financeiro, sendo uma alternativa para a caracterização não só de FCO₂, mas também de outros atributos do solo, promovendo a redução de custos e uma produção mais sustentável.

CAPÍTULO 3 – USO DAS TÉCNICAS DE MACHINE LEARNING PARA PREDIZER A EMISSÃO DE CO₂ DO SOLO DE CANAVIAIS COMERCIAIS, EM FUNÇÃO DOS ATRIBUTOS DO SISTEMA SOLO-PLANTA-ATMOSFERA

RESUMO: A emissão de CO₂ do solo (FCO₂) é um processo muito complexo, dependente das interações dos atributos que compõem o sistema solo-planta-atmosfera. Os algoritmos de aprendizado de máquina estão se mostrando uma ferramenta muito eficiente, devido a sua alta capacidade de classificação e predição da FCO₂, produzindo modelos confiáveis e acurados. Objetivo: estimar a FCO₂, em função dos atributos do sistema solo-planta-atmosfera, em áreas de cana-de-açúcar, nos estados de São Paulo e Mato Grosso do Sul. A base de dados foi oriunda de experimentos conduzidos em áreas de cana-de-açúcar, nos municípios de Motuca, (SP), Guariba (SP), Pradópolis (SP) e Aparecida do Taboado (MS). Os algoritmos testados para prever a FCO₂ foram: RFT_GS = random forest com a seleção de variáveis pelo Grid Search, RFT = random forest regressor, TREE = decision tree regression, BOO = gradient boosting regressor, MLP = rede neural Multi-layer Perceptron regressor, KNN = k-nearest neighbors para regressão, SVM = regressor support vector machines, SVM_kpoli = support vector machines pelo método de kernel polinomial, RLM = regressão linear múltipla, SVM_VC = regressor support vector machines com validação cruzada, SGD = regressor Stochastic Gradient Descent. O algoritmo RFT_GS foi o que melhor estimou a FCO₂, em função dos atributos do sistema solo-planta-atmosfera, seguido do RFT, da TREE, do MLP, do BOO, da KNN, do SVM, do SVM_kpoli, da RLM, da SVM_VC e da SGD. Conclui-se que estes estudos são importantes para os agricultores, poder público e privado, pois as informações geradas podem ser usadas no gerenciamento e nas tomadas de decisões do processo produtivo e financeiro, sendo uma opção para a caracterização não só da FCO₂, mas também de outros atributos do sistema solo-planta-atmosfera, promovendo a redução de custos, o uso eficiente dos recursos infinitos e finitos, produtividade e maior sustentabilidade para o sistema produtivo.

Palavras-chave: random forest; machine learning, respiração do solo; sistema solo-planta-atmosfera; efeito estufa, agricultura, mitigação de CO₂, cana.

CHAPTER 3 – USE OF MACHINE LEARNING TECHNIQUES TO PREDICT CO₂ EMISSIONS FROM THE SOIL OF COMMERCIAL CANE FIELDS, AS A RESULT OF THE ATTRIBUTES OF THE SOIL-PLANT-ATMOSPHERE SYSTEM

ABSTRACT: The emission of CO₂ from soil (FCO₂) is a very complex process, dependent on the interactions of attributes that make up the soil-plant-atmosphere system. Machine learning algorithms are proving to be a very efficient tool, due to their high ability to classify and predict FCO₂, producing reliable and accurate models. Objective: to estimate the FCO₂, as a function of the attributes of the soil-plant-atmosphere system, in sugarcane areas, in the states of São Paulo and Mato Grosso do Sul. The database came from experiments conducted in sugarcane areas, in the municipalities of Motuca (SP), Guariba (SP), Pradópolis (SP) and Aparecida do Taboado (MS). The algorithms tested to predict the FCO₂ were: RFT_GS = random forest with the selection of variables by the Grid Search, RFT = random forest regressor, TREE = decision tree regression, BOO = gradient boosting regressor, MLP = Multi-layer Perceptron regressor neural network, KNN = k-nearest neighbors for regression, SVM = support vector machines regressor, SVM_kpoli = support vector machines by the polynomial kernel method, RLM = multiple linear regression, SVM_VC = support vector machines regressor with cross validation, SGD = Stochastic Gradient Descent regressor. The RFT_GS algorithm was the one that best estimated the FCO₂, as a function of the attributes of the soil-plant-atmosphere system, followed by RFT, TREE, MLP, BOO, KNN, SVM, SVM_kpoli, RLM, SVM_VC and the SGD. It is concluded that these studies are important for farmers, public and private authorities, as the information generated can be used in the management and decision-making of the production and financial process, being an option for the characterization not only of FCO₂, but also of other attributes of the soil-plant-atmosphere system, promoting cost reduction, efficient use of infinite and finite resources, productivity and greater sustainability for the production system.

Keywords: random forest; machine learning, ground breathing; soil-plant-atmosphere system; greenhouse effect, agriculture, CO₂ mitigation, sugarcane.

3.1. INTRODUÇÃO

A mitigação da emissão dos gases do efeito estufa (GEEs) diminui os impactos das mudanças climáticas no planeta. Nos últimos anos foram observados preocupantes patamares históricos das concentrações atmosféricas dos gases CO₂ (dióxido de carbono), CH₄ (metano) e N₂O (óxido nitroso) em todo o globo terrestre. Conforme a Organização Meteorológica Mundial, a concentração atmosférica da CO₂ foi $400,0 \pm 0,1$ ppm, a CH₄ foi em torno de 1845 ± 2 ppb e N₂O foi cerca de $328,0 \pm 0,1$ ppb, representando, respectivamente, 144, 256 e 121% em relação aos níveis que foram observados nas eras pré-industriais (OMM, 2018).

As condições climáticas interferem na intensidade da emissão de CO₂ do solo, impactando diretamente no processo de respiração das raízes, na atividade microbiana e, conseqüentemente, na produção de CO₂ nos solos agrícolas (Schlömer et al., 2013; Xie et al., 2015). Além disso, as pesquisas apontam que as propriedades do solo também interferem no processo de produção de CO₂, dentre as quais mais citadas têm-se: a densidade do solo, os atributos do solo (Teixeira et al., 2012), a textura e a porosidade livre de água (Panosso et al., 2011), a temperatura do solo (Ussiri e Lal, 2009), o preenchimento de água nos poros (Smart e Peñuelas, 2005; Linn e Doran, 1984) e a umidade do solo (La Scala, et al., 2006; Moitinho et al., 2013).

A agricultura pode ser um importante agente colaborativo na mitigação das emissões do GEE. Para que esse objetivo seja atendido, é necessário que durante os processos produtivos sejam utilizadas práticas de manejo que favoreçam a drenagem dos GEE no solo. Existem diversas propostas que podem ser sugeridas para a mitigação da emissão de CO₂ e acúmulo de carbono no solo, como por exemplo o uso do sistema de plantio direto, a redução de queimadas, o controle e redução de desmatamento, a substituição dos combustíveis fósseis por biocombustíveis, o uso de práticas conservacionistas e o cultivo de plantas sustentáveis.

O Brasil é um dos principais produtores de alimentos, apresentando um grande papel na segurança alimentar mundial. Também é o maior produtor mundial da cultura da cana-de-açúcar. Estima-se que a área plantada na safra 2021/2022 será em torno de 8,26 milhões hectares e a produção 568,4 milhões de toneladas. O estado de São Paulo será o maior produtor do país, com 4,20 milhões de hectares e terá uma produção de 287,44 milhões de toneladas, compreendendo 50,57 % da cana

processada. Já o Estado do Mato Grosso do Sul é o quarto maior produtor do país, com 653,7 mil de hectares e produção de 45,42 milhões de toneladas, onde será responsável por 7,99 % da produção brasileira (CONAB, 2022).

Os estudos das práticas de manejo do solo, relacionadas a cultura da cana de açúcar são de grande importância para o processo de mitigação das emissões de GEE. As pesquisas apontam que as práticas usadas na reforma do canavial, como a calagem para a correção do pH do solo (Figueiredo et al., 2014) e o preparo intensivo (Iamaguti et al., 2015) promovem perdas expressivas do carbono presente no solo, na forma de CO_2 , principalmente em períodos curtos; assim, como consequência, observam-se alterações dos atributos químicos, físicos e biológicos do solo, além das variações de umidade e de temperatura do solo (Moitinho et al., 2013).

Uma das práticas que são utilizadas para promover a mitigação dos GEE é a colheita mecanizada que, dentre os seus benefícios, deposita sob a superfície do solo uma cobertura de palha da cana-de-açúcar, promovendo uma proteção física no solo contra o processo de erosão e a perda de carbono via respiração do solo e, consequentemente, aumentando o acúmulo de carbono por meio da matéria orgânica (Panosso et al., 2012). Iamaguti et al. (2015) estudaram a influência dos diferentes tipos de preparo do solo (convencional, subsolagem localizada e subsolagem convencional) nas perdas de carbono, via emissão de CO_2 e observaram que o preparo convencional ($2864,3 \text{ kg de CO}_2 \text{ ha}^{-1}$) é o mais nocivo, seguido pela subsolagem convencional ($1970,9 \text{ kg CO}_2 \text{ ha}^{-1}$) e pela subsolagem localizada ($1707,7 \text{ kg CO}_2 \text{ ha}^{-1}$),

Devido à grande relevância e a preocupação mundial com relação ao aumento das emissões do GEE, diversos pesquisadores pelo mundo estão estudando esses processos e uma linha de pesquisa que está crescente é a estimativa e a quantificação das emissões do CO_2 do solo, por meio da modelagem matemática. Desde a década de 1960, métodos matemáticos e estatísticos estão sendo utilizados em pesquisas da ciência do solo, com o intuito de promover o melhor entendimento das dinâmicas do solo e gerando ferramentas de avaliação, que auxiliam na resolução de problemas agrícolas e ambientais (McBratney e De Gruijter, 1992).

A não linearidade das relações existentes entre as variáveis utilizadas nos estudos da emissão dos GEE resultam em processos de grande complexidade, sendo

necessário o uso de técnicas matemáticas e estatísticas que consigam extrair as informações de maneira sólida e estruturada. Em função desta demanda, tem-se observado o aumento do uso das técnicas de machine learning dentro da agricultura, devido a sua grande robustez, são gerados resultados que podem auxiliar na compreensão da variabilidade espaçotemporal do CO₂, por meio dos modelos lineares generalizados (MLG) e modelos de machine learning como Regressão Linear Múltipla (RLM), K Neighbors Regressor (KNN), Regressor Florestal Aleatório (RFT), Redes neurais artificiais – Multi-layer Perceptron (MLP), Regressor de árvores de decisão (TREE), Máquina de vetores de suporte (SVM), Gradient Boosting para regressão (BOO) e Descida do Gradiente Estocástico (SGD).

A emissão de CO₂ do solo é dependente das variáveis oriundas do sistema solo-planta-atmosfera, que por sua vez, podem ser utilizadas na construção de um modelo de previsão. Quando analisam os impactos biológicos, observa-se as influências das mudanças que podem ocorrer na decomposição da matéria orgânica, na população microbiana, no volume e na densidade das raízes. Além disso, analisando os efeitos químicos e físicos do solo, são observadas as interferências com relação ao manejo adotado na área, podendo causar diferentes níveis de revolvimento do solo, diferentes incrementos de matéria orgânica, seja por meio de uma planta de cobertura, ou especificamente para a cana, pelo uso da colheita mecanizada e depósito da palhada sobre a superfície do solo. Também pode-se observar o efeito das condições climáticas no processo de emissão do CO₂, pois variáveis como temperatura, umidade, luminosidade e precipitação podem impactar diretamente no processo natural da emissão de CO₂, fazendo com que as suas variações promovam impactos nas análises temporais e espaciais.

Com relação as pesquisas, existem estudos usando aprendizado de máquina para a previsão da emissão do CO₂ do solo, como por exemplo: Predição do fluxo de CO₂ no solo em sistemas de manejo de cana-de-açúcar usando a abordagem Random Forest (Tavares et al., 2018), Uso de técnicas de mineração de dados para classificar a emissão de CO₂ do solo, induzida pelo manejo de culturas em canaviais (Farhate et al., 2018). No entanto, faltam estudos sobre a previsão da CO₂ utilizando variáveis oriundas do solo, planta e atmosfera, sendo analisadas por técnicas de aprendizado de máquina.

3.2. OBJETIVO

Visto este contexto, o objetivo do trabalho foi estimar a emissão de CO₂ do solo, em função dos atributos climáticos, físicos, químicos e biológicos do sistema solo-planta-atmosfera, em áreas de cana-de-açúcar, nos estados de São Paulo e Mato Grosso do Sul. O estudo foi realizado a partir de um conjunto de dados, oriundo de experimentos conduzidos em áreas de cana-de-açúcar, localizadas nos municípios de Guariba (SP), Pradópolis (SP) e Aparecida do Taboado (MS).

3.3. MATERIAIS E MÉTODOS

3.3.1. Caracterização da Área de Estudo

O estudo foi realizado em áreas canaviais comerciais, sob o sistema de cultivo da cana crua, nas cidades de Aparecida do Taboado (APT), em Mato Grosso do Sul (MS), e Motuca (MOT), Guariba (GUA) e Pradópolis (PAD), no estado de São Paulo (Figura 3.1).

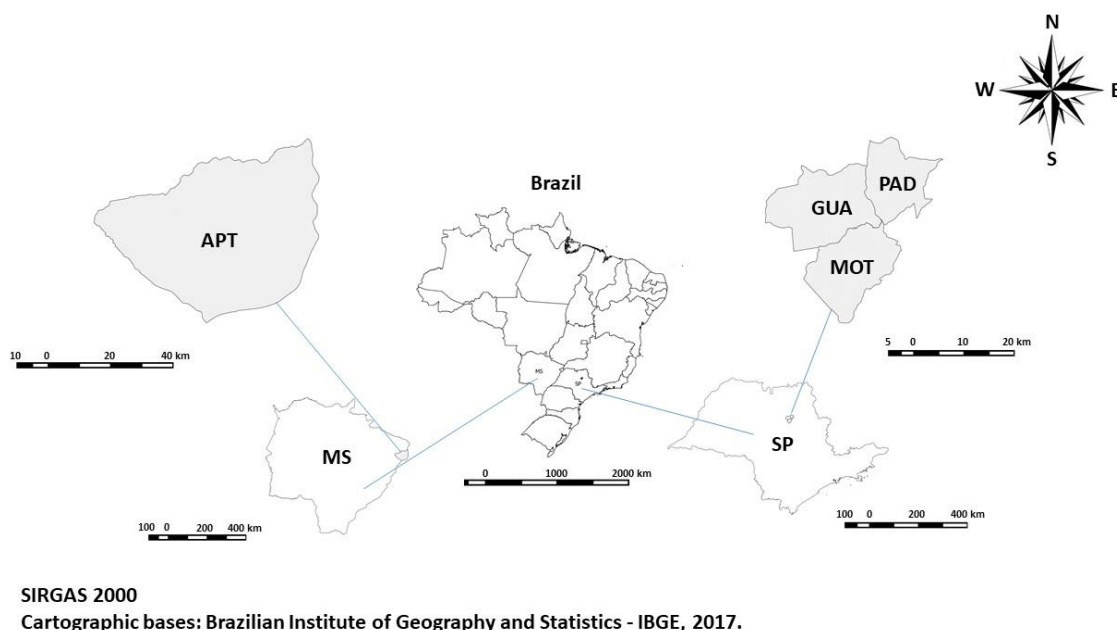


Figura 3.1: Mapas dos municípios das áreas experimentais do Estado de São Paulo - SP (Pradópolis - PAD, Guariba - GUA e Motuca - MOT) e Mato Grosso do Sul - MS (Aparecida do Taboado - APT).

(PAD – Pradópolis, GUA – Guariba, MOT – Motuca, APT – Aparecida do Taboado, SP – São Paulo, MS – Mato Grosso do Sul, km – kilometers, N – north, S – south, E – east, W – west, SIRGAS - Geocentric Reference System for the Americas and IBGE – Brazilian Institute of Geography and Statistics). Fonte: Da Silva et al. (2020).

Os solos do estado de São Paulo foram classificados como Latossolo Vermelho eutroférico. A textura do solo de Motuca foi interpretada como muito argilosa (argila > 600 g kg⁻¹). Em Guariba e Pradópolis, o solo foi definido como argiloso (350 g kg⁻¹ < argila < 600 g kg⁻¹). Já a classificação do solo de Aparecida do Taboado (MS) foi a Latossolo Vermelho distroférico de textura argilosa, conforme o Sistema Brasileiro de Classificação de Solos - SIBCS (Santos et al., 2013 b).

Conforme a metodologia de Köppen, as classificações climáticas das áreas estudadas foram Aw. As precipitações anuais das cidades do estado de São Paulo foram 1400 mm, com chuvas concentradas de novembro até fevereiro, com temperaturas médias anuais de 22,2 °C e as elevações dos municípios em relação ao nível do mar foram 550 (MOT), 620 (GUA) e 515 (PAD) metros. Para o ATP, a elevação do nível do mar foi de 370 metros, as temperaturas médias anuais foram de 23,7 °C e a precipitação anual de 1.300 mm, com pluviosidade concentrada em novembro a fevereiro.

3.3.2. Mensurações do CO₂, temperatura e umidade do solo

A emissão de CO₂ do solo (FCO₂) foi registrada, no início do desenvolvimento da cultura, por meio do sistema LI-COR (LI-8100), cuja medição foi realizada através das alterações da concentração do CO₂ da câmara de solos, monitoradas pelo sistema, por meio da espectroscopia na região do infravermelho. As câmaras utilizadas para as medições no solo possuíam um volume interno de 854,2 cm³ e área de contato circular de 83,7 cm². Para a coleta das informações, previamente no solo foram inseridos colares de PVC em cada ponto da malha, na profundidade de 3 cm. O tempo de duração para a determinação da concentração de CO₂ dentro da câmara foi de 1,5 minutos e a interpolação entre a concentração de CO₂ da câmara e a concentração do solo é realizada a cada 2,5 segundos.

Durante as avaliações da respiração do solo, também foi realizado o monitoramento da temperatura do solo (Ts), por meio de um sensor térmico integrado ao sistema da LI-8100. O sensor possui uma haste de 0,20 m que foi inserida no interior do solo, em torno dos colares de PVC.

A umidade do solo (Us) foi mensurada pelo equipamento de TDR (Time Domain Reflectometry - Hydrosense TM, Campbell Scientific, Austrália). O aparelho de TDR é

formado por uma sonda que apresenta duas hastes de 0,12 m, que foram introduzidas no interior do solo, próximos aos colares de PVC. As medidas de FCO_2 , Ts e Us foram realizadas simultaneamente em ambos os pontos da gride, no período da manhã (7:00–9:00 AM).

3.3.3. Análise dos Atributos Físicos e Químicos do Solo

Ao final do período das medições de FCO_2 , amostras de solo foram coletadas em cada ponto amostral, na profundidade de 0 a 0,10 m e secas ao ar, posteriormente, foram destorroadas e passadas pela peneira de malha 2 mm. Conforme o método da resina trocadora de íons de Raij (2001), foram determinados as extrações de fósforo disponível, cálcio, magnésio e potássio trocáveis. Foram calculadas a soma de bases (SB) e a capacidade de troca de cátions (CTC). O índice de saturação por bases (V%) foi mensurado através da relação entre SB e CTC. A densidade do solo (D_s) e a densidade de partículas do solo (D_p) foram obtidas em amostras indeformadas, coletadas com um amostrador adaptado a cilindros com dimensões médias de 4,0 cm de altura e 5,0 cm de diâmetro interno (EMBRAPA, 1997). A porosidade total do solo (PT) foi mensurada por meio do valor de densidade. As classificações dos poros pelos tamanhos, tais como a macroporosidade (Macro) e a microporosidade (Micro) foram realizadas usando um funil de placa porosa, sob a tensão de 60 cm de altura de coluna d'água em saturação amostral prévia. O volume de água que permaneceu na amostra após este processo, corresponde aos microporos e aos macroporos que foram calculados pela diferença (EMBRAPA, 1997).

O estoque de carbono foi calculado para a profundidade de 0,10 m e, com base na massa de solo equivalente para contabilizar as variações da densidade do solo, nas diferentes áreas de estudo pela equação 1 (Bayer et al., 2000):

$$\text{EstC} = (\text{CO} \cdot D_s \cdot E) / 10 \quad (1)$$

Em que: EstC = estoque de carbono (Mg ha^{-1});

CO = teor de carbono orgânico oxidável (g kg^{-1});

D_s = densidade do solo (kg dm^{-3});

E = espessura da camada estudada (cm).

3.3.4. Evapotranspiração potencial

Para o cálculo da Evapotranspiração potencial (ETP, mm) foram usados os seguintes dados climáticos: radiação solar no topo da atmosfera (Q_0 , MJ m⁻² dia⁻¹), temperatura média do ar (T_m , °C), precipitação pluvial (Prec, mm), umidade relativa (UR, %) e a velocidade do vento (U_s , m s⁻¹). Esses dados foram obtidos por meio da plataforma National Aeronautics and Space Administration/Prediction of World Wide Energy Resources - NASA/POWER (Stackhouse et al., 2016). A plataforma disponibiliza dados meteorológicos diários, com resolução espacial de 1°, o que representa um grid de 110,57 km aproximadamente.

O método utilizado para a determinação da Evapotranspiração Potencial (ETP) foi o Penman e Monteith (Allen et al., 1998) na escala diária, conforme as Equações 2 a 5.

$$ETP = \frac{0,408 \times s \times (Rn - G) + \frac{\gamma \times 900 \times U_2 \times (es - ea)}{T + 273}}{s + \gamma \times (1 + 0,34 \times U_2)} \quad (2)$$

$$s = \frac{4098 \times es}{(T + 273)^2} \quad (3)$$

$$ea = \frac{UR \times es}{100} \quad (4)$$

$$es = 0,6108 \times e^{\frac{17,27 \times T}{237,3 + T}} \quad (5)$$

Em que:

ETP é a evapotranspiração potencial (mm);

Rn é a radiação líquida (MJ m⁻² d⁻¹);

UR é umidade relativa (%);

G é o fluxo de calor no solo igual a 0 MJ m⁻² d⁻¹;

T é a temperatura do ar (°C);

γ é a constante psicrométrica igual a 0,063 kPa °C⁻¹;

s é a declividade da curva da pressão de vapor versus temperatura (kPa °C⁻¹);

U_2 é a velocidade do vento (m s⁻¹) na altura de 2 m;

es é a pressão da saturação de vapor (kPa);

ea é a pressão de vapor atual (kPa).

3.3.5. Balanço hídrico climatológico

A deficiência hídrica (DEF, mm), armazenamento de água no solo (STO, mm), excedente hídrico (EXC, mm), porosidade livre d'água (PLA, mm) e umidade do solo

diário (Um, mm) foram estimados na escala diária para todas as áreas de estudo, conforme a metodologia de Thornthwaite e Mather (1955) (Figura 3.2).

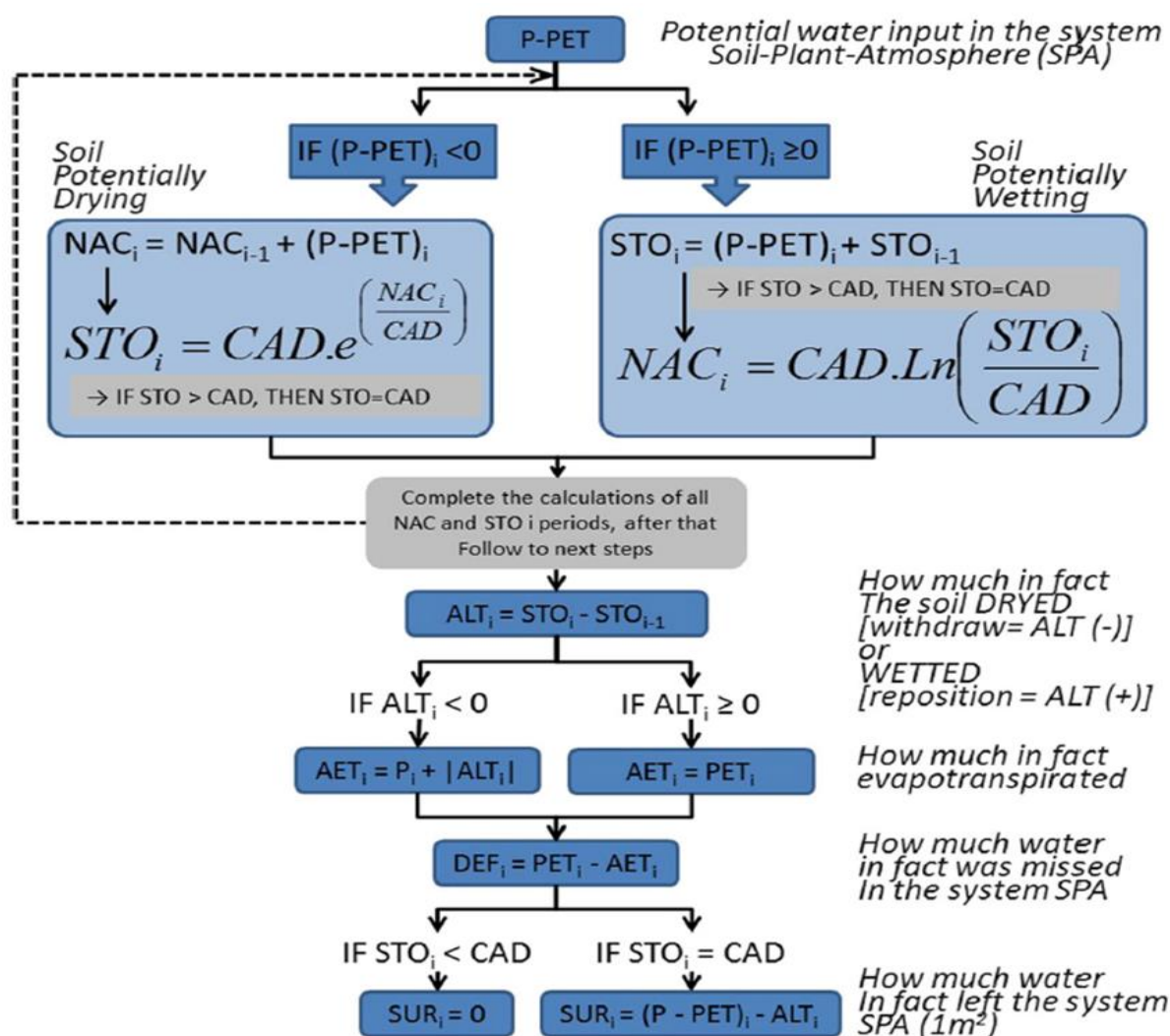


Figura 3.2: Fluxograma do cálculo do balanço Hídrico, conforme Thornthwaite e Mather, modificado por Rolim et al. (2020).

Em que:

AET é a evapotranspiração real (mm);

ALT é o armazenamento de água no solo do mês atual - armazenamento de água no solo do mês anterior (mm);

i é o período do mês;

NAC é negativo acumulado (mm);

Prec é a precipitação (mm);

PET é a evapotranspiração potencial (mm);

WC é a capacidade de água disponível (mm);

STO (ARM) é o armazenamento de água no solo (mm);

EXC é o excedente de água no sistema solo-planta-atmosfera (mm);

CAD é a capacidade de água disponível do solo (mm);

DEF é a deficiência de água no sistema solo-planta-atmosfera (mm).

3.3.6. Obtenções das variáveis oriundas do sistema solo-planta-atmosfera que foram usadas nas predições dos modelos da FCO₂

Inicialmente, foi criado um modelo empírico, que teve como objetivo relacionar as influências dos atributos do sistema solo-planta-atmosfera no processo de FCO₂. O sistema foi dividido em três partes: física do solo, química do solo, planta-atmosfera.

As propriedades físicas do solo estão ligadas as estimações de processos físicos que ocorrem no solo, principalmente aqueles relacionados à movimentação de matéria e fluídos, interferindo no processo de locomoção e transformação de energia no solo. Para estimar o processo de transporte de CO₂ no solo, foram selecionados os seguintes atributos com base em pesquisas anteriores (Da Silva et al., 2020; Bicalho et al., 2014): porosidade livre d'água (PLA, mm), umidade solo (Us, mm), densidade das partículas (kg dm⁻³) e densidade do solo (kg dm⁻³).

As propriedades químicas dos solos, juntamente com a microbiota, são os principais causadores de transformações de energia no solo, resultantes praticamente dos processos de adsorção, fixação química, oxidação, troca e a neutralização. De maneira geral, observa-se que as características químicas do solo resultam em processos de transformação e formação de CO₂, e os atributos que mais se relacionaram com esse processo foram a capacidade de troca de cátions, a matéria orgânica e o teor de fósforo disponível (Moitinho et al., 2015; Vicentini et al., 2019).

O balanço hídrico também pode auxiliar nos estudos dessas relações, essa metodologia permite analisar os efeitos da radiação solar, temperatura, água no sistema vegetal, balanço de energia, evapotranspiração vegetal e o balanço hídrico e hidrológico no sistema. Para este estudo foi realizado o balanço hídrico diário sequencial, por meio da metodologia de Thornthwaite e Mather (1955) para extrair os atributos que mais se relacionam ao processo de transporte, e de forma secundária a formação de CO_2 no solo. As variáveis escolhidas foram a capacidade de armazenamento de água no solo (ARM, mm), evapotranspiração potencial e real (ETP e ETR, mm), a porosidade livre d'água (PLA, mm), temperatura do ar (T_{ar} , °C), temperatura do solo (T_s , °C) e umidade do solo (U_s , mm). Logo abaixo observa-se o modelo do impacto do sistema solo-planta-atmosfera (Figura 3.3) no desenvolvimento das plantas.

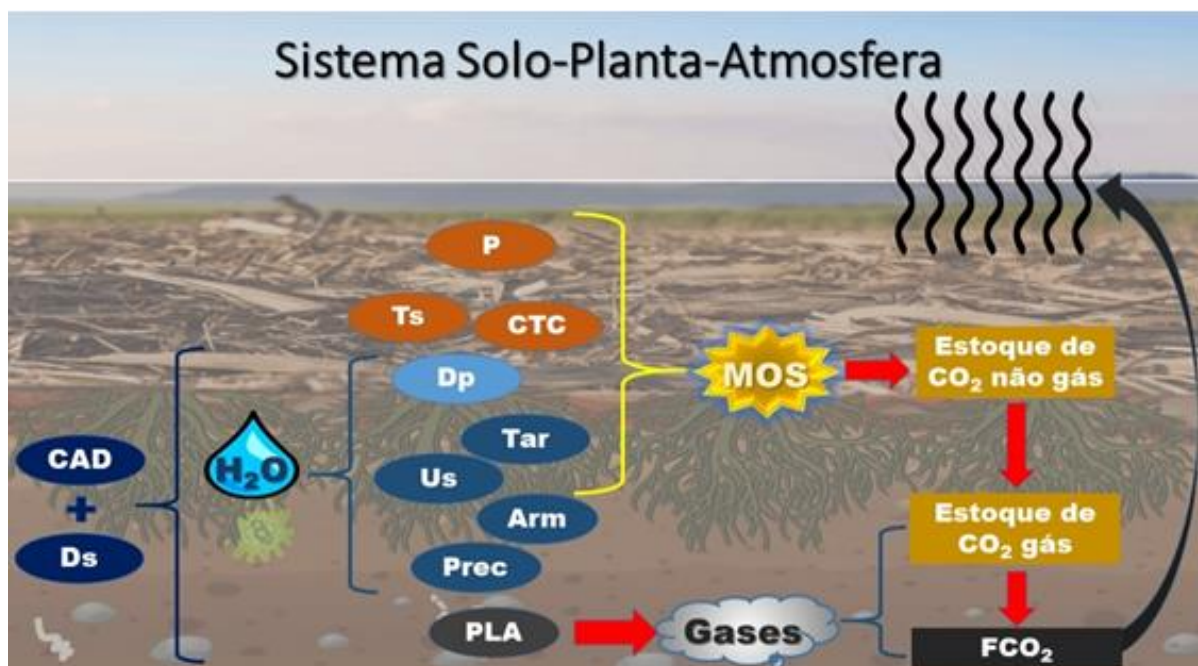


Figura 3.3: Representação da análise empírica da influência dos atributos que compõem o sistema solo-planta-atmosfera, com relação ao processo de FCO_2 . Fonte: Própria.

3.3.7. Análise estatística

Os dados foram analisados, inicialmente por meio da estatística descritiva (média, desvio-padrão, máximo, mínimo e coeficiente de variação). Os testes de normalidade e homocedasticidade foram realizados por meio dos testes Shapiro-Wilk

e Levene, respectivamente. Utilizaram as análises das variabilidades gerais dos dados, onde sua qualidade foi medida pelo p-valor. Também foi gerado um histograma de barras, visando determinar as suas respectivas classes e frequências. Os testes de normalidade e homocedasticidade foram realizados ao nível de significância (α) de 5% (grau de confiança de 95%).

A interpretação do teste Shapiro-Wilk foi feita por meio do nível de concordância entre a hipótese nula (H_0) e as observações, sendo H_0 referente à distribuição Normal dos dados. Quanto maior for o p-valor, maior é a solidez entre os dados e a hipótese nula. Sendo assim, realizou-se a seguinte regra de decisão: H_0 : (i) se $p\text{-valor} \leq \alpha$, rejeita-se H_0 , isto é, não se aceita que o conjunto de dados tenha distribuição Normal; (ii) se $\text{valor-p} > \alpha$, não se rejeita H_0 , isto é, os dados tendem a ter uma distribuição Normal. A interpretação do teste de Levene é bastante semelhante ao que foi explicado anteriormente: H_0 : (i) se $p\text{-valor} \leq \alpha$, rejeita-se H_0 , isto é, os dados tendem a apresentar heterocedasticidade; (ii) se $\text{valor-p} > \alpha$, não se rejeita H_0 , isto é, os dados tendem a apresentar homoscedasticidade.

Para essas análises, foram avaliados os seguintes conjuntos dos atributos do solo: emissão de CO_2 do solo (FCO_2 , $\mu\text{mol CO}_2 \text{ m}^{-2} \text{ s}^{-1}$), densidade das partículas (D_p , kg dm^{-3}), densidade do solo (D_s , kg dm^{-3}), macroporosidade (Macro, %), microporosidade (Micro, %), porosidade total (PT, %), porosidade livre de água (PLA, %), teor de areia total (Areia, g kg^{-1}), teor de silte (Silte, g kg^{-1}), teor de argila (Argila, g kg^{-1}), potencial hidrogeniônico (pH, CaCl_2), matéria orgânica do solo (MOS, g dm^{-3}), fósforo disponível (P, mg dm^{-3}), potássio (K, $\text{mmol}_c \text{ dm}^{-3}$), cálcio (Ca, $\text{mmol}_c \text{ dm}^{-3}$), magnésio (Mg, $\text{mmol}_c \text{ dm}^{-3}$), acidez potencial ($\text{H}+\text{Al}$, $\text{mmol}_c \text{ dm}^{-3}$), saturação de bases (V, %), capacidade de troca de cátions (CTC, $\text{mmol}_c \text{ dm}^{-3}$), soma de bases (SB, $\text{mmol}_c \text{ dm}^{-3}$), carbono orgânico oxidável (Corg, g kg^{-1}), estoque de carbono no solo (Estc, Mg ha^{-1}), temperatura do solo (T_s , $^{\circ}\text{C}$) e umidade do solo (U_s , % do volume). Também foram usadas as seguintes variáveis relacionadas ao clima e à planta: capacidade de água disponível do solo (CAD, mm), armazenamento de água no solo (ARM, mm), porosidade livre de água decorrente do balanço hídrico (PLA_BH, %), temperatura do ar até 2 metros do solo (T_{air} , $^{\circ}\text{C}$), precipitação (Prect, mm), evapotranspiração potencial (ETP, mm), evapotranspiração real (ETR, mm), umidade do solo oriunda do balanço hídrico (U_{mi} , mm), deficiência de água no sistema solo-planta-atmosfera

(DEF, mm), armazenamento de água no solo (ARM, mm), armazenamento de água no solo do mês atual - armazenamento de água no solo do mês anterior (ALT, mm), equilíbrio hidrológico potencial do sistema solo-planta-atmosfera (P-ETP, mm).

3.3.8. Software Computacional

O software computacional utilizado neste trabalho foi o Python (linguagem de programação, versão 3.8.5). A principal vantagem em usar o Python é que esse programa é gratuito e de código aberto, isso facilita muito, pois é atualizado constantemente e os bugs são corrigidos quando comunicado para os programadores (McKinney, 2012).

Para facilitar o entendimento e a construção dos modelos, o código para este estudo foi escrito em Jupyter Notebook (Kluyver, 2016). O Jupyter Notebook é um aplicativo da Web de código aberto, que propicia a criação e compartilhamento de código, modelagem de aprendizado de máquina, limpeza e visualização de dados. Foram baixados no ambiente Jupyter Notebook o Scikit-learn, que é um módulo que integra uma série de algoritmos de ML (Pedregosa et al., 2011). A visualização dos dados foi realizada por meio das bibliotecas matplotlib e seaborn.

3.3.9. Algoritmos de aprendizado de máquina

O aprendizado de máquina (ML) é um fragmento da inteligência artificial (IA) que pode prever processos de classificação, previsão, regressão e decisão, por meio de diversos modelos matemáticos (Alpaydm, 2020). O ML possui grande poder de precisão devido às variáveis usadas e o alto nível de sucesso do treinamento (Mitchell, 2006). Basicamente temos uma primeira fase em que o algoritmo passa por um aprendizado, via os registros amostrados denominados de dados de treinamento. Logo depois, devido a essas experiências são previstas as respostas desconhecidas.

Analisando as aplicações das técnicas de ML, observa-se que os algoritmos estão sendo usados em diversas áreas, como por exemplo na física, matemática, saúde, militar, engenharia e economia, imitando neurônios no cérebro, obtendo resultados de grande importância e impactando na vida dos seres humanos e do planeta como um todo (Ağbulut et al., 2021; Senturk 2020; Gürel et al., 2020; Başarslan e Argun, 2019; Salam e Verma, 2020; Karstoft et al., 2015).

Algumas técnicas de ML foram utilizadas para a predição da FCO₂. Em todos os casos, a FCO₂ foi a variável dependente e as variáveis oriundas do sistema solo-planta-atmosfera foram as independentes. Para todas as técnicas, foi separado aleatoriamente 80% dos dados para treinamento (2415) e 20% para calibração e validação (604) da precisão dos modelos, usando a biblioteca scikit-learn do Python (Pedregosa et al. 2011).

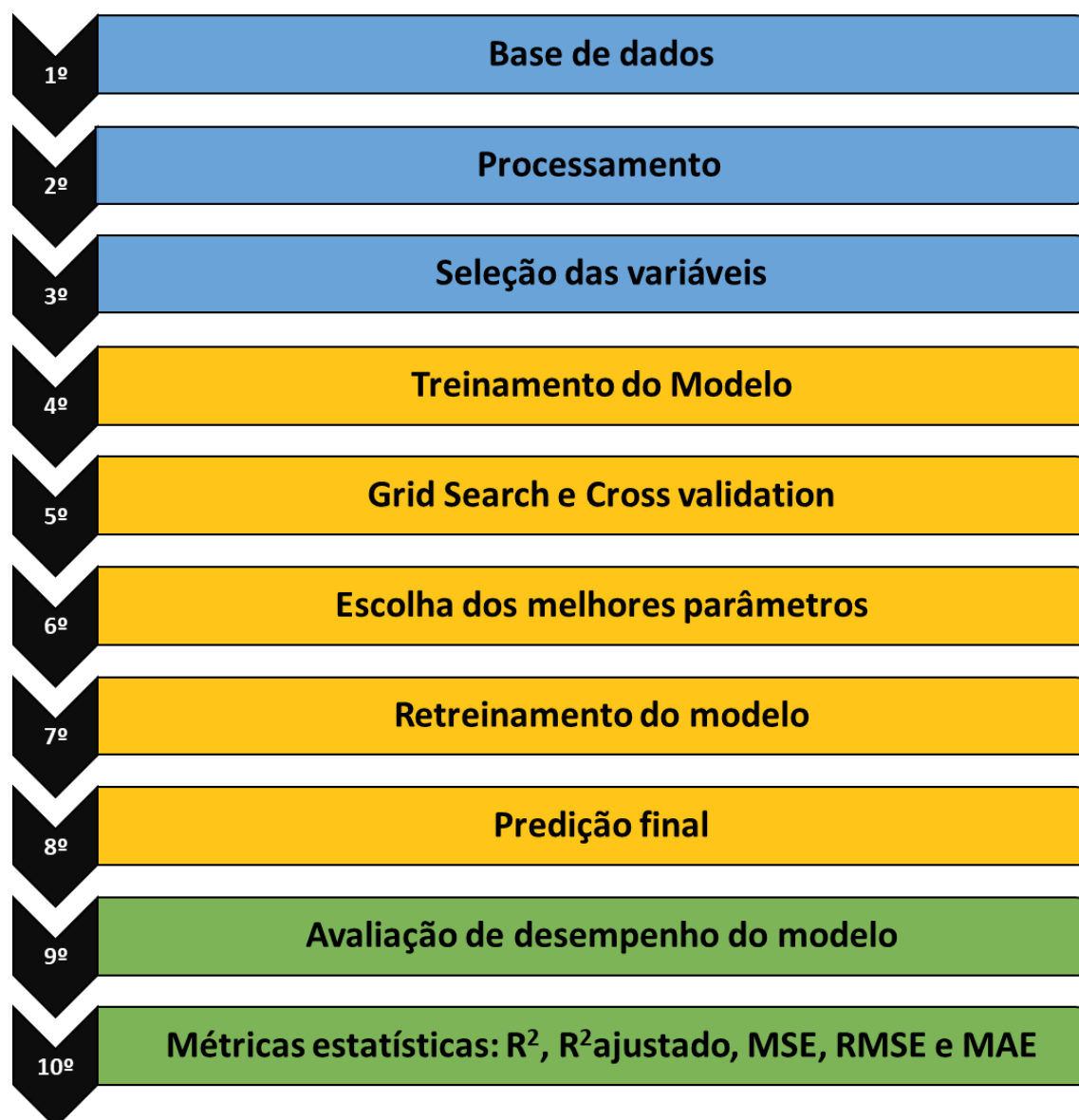


Figura 3.4: Resumo do processo de desenvolvimento dos modelos para prever a emissão de CO₂ do solo, por meio dos algoritmos de aprendizado de máquina.

R^2 = coeficiente de determinação; R^2_{ajust} = coeficiente de determinação ajustado; MSE = erro quadrático médio; RMSE = raiz quadrada do erro-médio; MAE = média dos erros absolutos.

Na Figura 3.4, observa-se um resumo das etapas realizadas para a determinação da construção dos algoritmos de ML. Após os processos de pré-processamento e seleção de recursos, os atributos de entrada foram dimensionados. Os dados de entrada foram normalizados e dimensionados para se enquadrarem em um intervalo cujo os valores variam de 0 a 1, subtraindo o valor mínimo e dividindo pelo intervalo entre o valor máximo e mínimo de uma variável de entrada. A padronização é alcançada subtraindo a média de uma variável de entrada de cada valor e dividindo o seu resultado pelo desvio padrão (Witten e Frank, 2002). Para diminuir o impacto das unidades das variáveis, a técnica de normalização das variáveis de entrada foi feita para proporcionar a igualdade da representação no framework ML.

Neste trabalho foram usados algoritmos que resolvem problemas por meio da regressão, que são uma forma de aprendizado supervisionado. Esses algoritmos geram os modelos por meio dos dados de treinamento, estabelecendo relações que serão aplicadas na predição de novos dados. Para estimar a FCO₂, em função das variáveis que compõem o sistema solo-planta-atmosfera, foram usados os seguintes algoritmos:

- 1) Regressão Linear Múltipla (RLM);
- 2) K Neighbors Regressor (KNN);
- 3) Regressor Florestal Aleatório (RFT);
- 4) Redes neurais artificiais – Multi-layer Perceptron (MLP);
- 5) Árvore de Decisão pelo método da regressão (TREE);
- 6) Máquina de vetores de suporte (SVM);
- 7) Gradient Boosting para regressão (BOO);
- 8) Descida do Gradiente Estocástico (SGD).

Logo abaixo segue a representação dos modelos utilizados para a predição da emissão de CO₂ do solo (Figura 3.5).

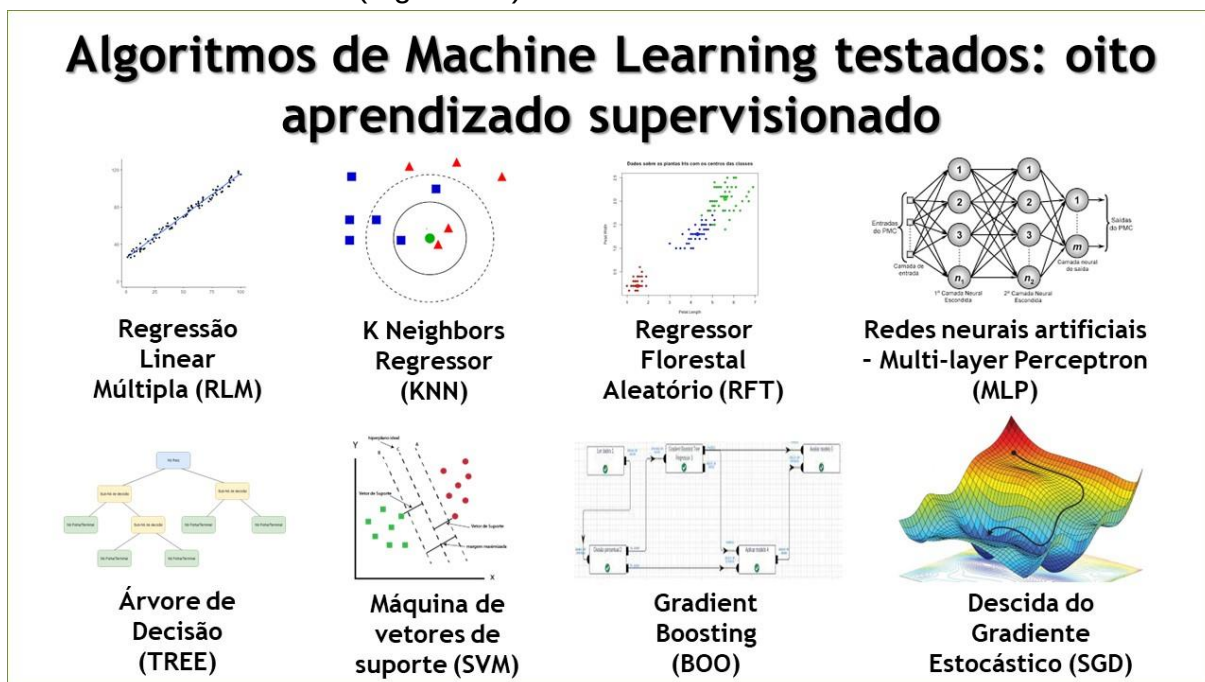


Figura 3.5: Representações dos oito algoritmos de aprendizado não supervisionado utilizados para a predizer a emissão de CO₂ do solo (FCO₂).

Após a fase de treinamento dos modelos, foram encontrados os hiperparâmetros otimizados por meio da função “GridSearchCV” no Scikit-learn. A função GridSearchCV combina todas as configurações de hiperparâmetros dados, visando obter o melhor conjunto de hiperparâmetros que possam proporcionar o melhor desempenho, dentro de uma configuração determinada pelo pesquisador (Yang e Shami, 2020). Os hiperparâmetros otimizados foram usados para treinar novamente os algoritmos e as estimativas finais da precisão dos modelos foram realizadas, usando o conjunto de dados destinados para a previsão e validação dos modelos. A técnica de seleção dos hiperparâmetros foi utilizada nos algoritmos RFT, BOO, TREE e SVM.

Logo abaixo segue a Tabela 3.1 com os hiperparâmetros utilizados para cada algoritmo desenvolvido neste estudo. Para os parâmetros que não foram listados, foram usados os parâmetros padrões dos algoritmos disponibilizados pelo pacote Scikit-learn.

Tabela 3.1: Hiperparâmetros ajustados para os algoritmos RFT_GS = random forest com a seleção de variáveis pelo método de grid Search, RFT = random forest regressor, TREE = decision tree regression, BOO = gradient boosting regressor, MLP = rede neural Multi-layer Perceptron regressor, KNN = regressão baseada no método k-nearest neighbors, SVM = método de regressão support vector machines, SVM_kpoli = support vector machines com o método de kernel polinomial, RLM = regressão linear múltipla, SVM_VC = método de regressão Support vector machines com validação cruzada, SGD = método de regressão da Stochastic Gradient Descent, disponíveis no pacote Scikit-learn (Pedregosa et al, 2011) da linguagem Python.

Regressão Linear Múltipla com Ridge (RLM)		
Hiperparâmetros ajustados	Função	Hiperparâmetros selecionados
Alfa	Força de regularização. A regularização melhora o condicionamento do problema e reduz a variância das estimativas.	1
Solver	Solucionador de otimização de peso	Cholesky
K Neighbors Regressor (KNN)		
n_neighbors	Número de vizinhos a serem usados por padrão das consultas.	8
Regressor Florestal Aleatório (RFT)		
random_state	Controla a aleatoriedade do das amostras usadas ao construir árvores e a amostragem dos recursos a serem considerados ao procurar a melhor divisão em cada nó.	0
Regressor Florestal Aleatório com a seleção de variáveis pelo método de grid Search (RFT_GS)		
n_estimators	O número de árvores na floresta.	(3,10,30,40,50) – Combinação 1 (0,1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20) – Combinação 2
max_features	O número de recursos a serem considerados ao procurar a melhor divisão	(10,14,18) – Combinação 1 (1,5,10,15,20,30,50) – Combinação 2
random_state	Controla a aleatoriedade do das amostras usadas ao construir árvores e a amostragem dos recursos a serem considerados ao procurar a melhor divisão em cada nó.	0
n_jobs	O número de jobs a serem executados em paralelo.	-1
Cv	Determina a estratégia de divisão de validação cruzada.	10
Stochastic Gradient Descent (SGD)		

max_iter	O número máximo de passagens sobre os dados de treinamento (épocas).	500
Penalty	A penalidade ou termo de regularização a ser usada.	12
eta0	A taxa de aprendizado inicial para os cronogramas 'constante', 'invscaling' ou 'adaptativo'	0.1
Gradient Boosting Regressor (BOO)		
max_depth	Profundidade máxima dos estimadores de regressão individuais.	(1,2,3)
n_estimators	O número de estágios de reforço a serem executados.	1200
max_features	O número de recursos a serem considerados ao procurar a melhor divisão	(2,4,6,8)
learning_rate	A taxa de aprendizado reduz a contribuição de cada árvore na taxa de aprendizado.	(0.1,0.2,0.5)
random_state	Controla a semente aleatória dada a cada estimador de Árvore em cada iteração de reforço.	0
Cv	Determina a estratégia de divisão de validação cruzada.	5
Regressor arvores de decisão (TREE)		
n_estimators	O número de árvores na floresta.	(3,10,30,40,50) – Combinação 1 (3,10) – Combinação 2
max_features	O número de recursos a serem considerados ao procurar a melhor divisão.	(2,4,6,8) – Combinação 1 (2,3,4) – Combinação 2
Cv	Determina a estratégia de divisão de validação cruzada.	10

Máquina de vetores de suporte (SVM)		
n_estimators	O número de árvores na floresta.	(3,10,30,40,50) – Combinação 1 (3,10) – Combinação 2
max_features	O número de recursos a serem considerados ao procurar a melhor divisão.	(2,4,6,8) – Combinação 1 (2,3,4) – Combinação 2
Cv	Determina a estratégia de divisão de validação cruzada.	10
Support vector machines com validação cruzada (SVM_VC)		
Kernel	Especifica o tipo de kernel a ser usado no algoritmo.	rbf – Validação cruzada
C	Parâmetro de regularização.	500 – Validação cruzada
Gamma	Coeficiente de kernel para 'rbf', 'poli' e 'sigmoid'.	0.1 – Validação cruzada
Épsilon	Epsilon no modelo epsilon-SVR. Ele especifica o tubo epsilon dentro do qual nenhuma penalidade é associada na função de perda de treinamento com pontos previstos dentro de uma distância epsilon do valor real.	0.1 – Validação cruzada
Cv	Determina a estratégia de divisão de validação cruzada.	10
Kernel	Especifica o tipo de kernel a ser usado no algoritmo.	Linear
C	Parâmetro de regularização.	100
Gamma	Coeficiente de kernel para 'rbf', 'poli' e 'sigmoid'.	Auto
Support vector machines com o método de kernel polinomial (SVM_kpoli)		
Kernel	Especifica o tipo de kernel a ser usado no algoritmo.	linear – Validação cruzada
C	Parâmetro de regularização.	10 – Validação cruzada

Gamma	Coeficiente de kernel para 'rbf', 'poli' e 'sigmoid'.	auto – Validação cruzada
Kernel	Especifica o tipo de kernel a ser usado no algoritmo.	Poly
C	Parâmetro de regularização.	10
Gamma	Coeficiente de kernel para 'rbf', 'poli' e 'sigmoid'.	Auto
Degree	Grau da função kernel polinomial ('poli'). Ignorado por todos os outros kernels.	3
Épsilon	Epsilon no modelo epsilon-SVR. Ele especifica o tubo epsilon dentro do qual nenhuma penalidade é associada na função de perda de treinamento com pontos previstos dentro de uma distância epsilon do valor real.	0.1
coef0	Termo independente na função kernel. Só é significativo em 'poli' e 'sigmóide'.	1
Redes neurais artificiais – Multi-layer Perceptron (MLP)		
Activation	Função de ativação para a camada oculta.	Relu
Alpha	Parâmetro de penalidade L2 (prazo de regularização).	0.0001
hidden_layer_sizes	O i-ésimo elemento representa o número de neurônios na i-ésima camada oculta.	(100, 50, 30, 40, 60)
learning_rate	Cronograma de taxa de aprendizado para atualizações de peso.	Constant
max_iter	Número máximo de iterações.	50
Solver	O solucionador para otimização de peso.	Adam

3.3.10. Medidas estatísticas de avaliação de desempenho

A avaliação da acurácia dos modelos da FCO₂ foi realizada por meio das medidas estatísticas de desempenho. Para avaliar o desempenho dos algoritmos calibrados e validados foram utilizados os seguintes índices estatísticos:

- Correlação de Pearson (r), equação 6;
- Coeficiente de determinação (R²), equação 7;
- Coeficiente de determinação ajustado (R²ajust), equação 8;
- Erro quadrático médio (MSE), equação 9;
- Raiz quadrada do erro-médio (RMSE), equação 10;
- Média dos erros absolutos (MAE), equação 11;

Para a obtenção da maior confiabilidade e acurácia das regressões também foram realizados o teste F a 5% de probabilidade.

$$r = \frac{\sum_{i=1}^n (Y_{obs_i} - \bar{Y}_{obs}) \times (Y_{est_i} - \bar{Y}_{est})}{\sqrt{\sum_{i=1}^n (Y_{obs_i} - \bar{Y}_{obs})^2} \times \sqrt{\sum_{i=1}^n (Y_{est_i} - \bar{Y}_{est})^2}} \quad (6)$$

$$R^2 = r^2 \quad (7)$$

$$R^2_{ajust} = \left[1 - \frac{(1 - R^2) \times (n - 1)}{N - k - 1} \right] \quad (8)$$

$$MSE = \frac{\sum_{i=1}^N (Y_{obs_i} - Y_{est_i})^2}{N} \quad (9)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (Y_{obs_i} - Y_{est_i})^2}{N}} \quad (10)$$

$$MAE = \frac{1}{N} \sum_{i=1}^N |Y_{obs_i} - Y_{est_i}| \quad (11)$$

Em que:

Yesti: variável estimada na posição i;

Yobsi: variável observada na posição i;

N: número de dados;

k: número de variáveis independentes na regressão;

Quanto menor for os valores da MSE, RMSE e MAE, melhores são os ajustes dos modelos. No entanto, para os parâmetros r, R², R²ajust, quanto maior for os valores em módulo, maior será a correlação entre as variáveis.

3.4. RESULTADOS E DISCUSSÃO

3.4.1. Distribuição dos dados e Estatística Descritiva

Na Tabela 3.2 estão descritos os resultados do teste de Shapiro-Wilk, para todos os dados coletados. Analisando os valores do p-valor das variáveis, observa-se que todas as variáveis apresentaram índices inferiores a 0,05, mostrando que as variáveis tendem a não apresentarem distribuição normal. Para o teste de Levene foi encontrado o valor de 2339,70 e o p-valor foi 0,00e⁺⁰⁰, indicando que as variáveis tendem a apresentar heterocedasticidade. Na Figura 3.6 são mostradas as distribuições das variáveis estudadas.

Tabela 3.2: Teste de normalidade pelo método de Shapiro-wilk da emissão de CO₂, atributos atmosféricos e características físico-químicos do solo de todas as áreas estudadas.

<i>Variável</i>	<i>Estatística do teste</i>	<i>P-valor</i>
<i>FCO₂</i>	0,87*	4,20e ⁻⁴⁵
<i>Ts</i>	0,73*	0,00e ⁺⁰⁰
<i>Us</i>	0,91*	8,24e ⁻³⁹
<i>Dp</i>	0,97*	2,39e ⁻²³
<i>Ds</i>	0,99*	2,23e ⁻¹⁴
<i>Macro</i>	0,98*	2,32e ⁻²¹
<i>Micro</i>	0,99*	1,17e ⁻¹⁶
<i>PT</i>	1,00*	1,26e ⁻⁰⁷
<i>PLA</i>	0,93*	1,93e ⁻³⁴
<i>Areia</i>	0,87*	4,20e ⁻⁴⁵
<i>Silte</i>	0,77*	0,00e ⁺⁰⁰
<i>Argila</i>	0,85*	0,00e ⁺⁰⁰
<i>pH</i>	0,99*	1,31e ⁻¹⁰
<i>MO</i>	0,92*	1,18e ⁻³⁶
<i>P</i>	0,97*	2,59e ⁻²⁵
<i>K</i>	0,89*	6,2e ⁻⁴²
<i>Ca</i>	0,98*	1,18e ⁻¹⁸
<i>Mg</i>	0,99*	3,02e ⁻¹⁴
<i>H-Al</i>	0,97*	2,26e ⁻²⁵

SB	0,99*	1,21e ⁻¹⁴
CTC	0,95*	6,52e ⁻³²
V	0,98*	3,80e ⁻²⁰
Corg	0,92*	1,18e ⁻³⁶
Estc	0,95*	3,94e ⁻³²
Prect	0,38*	0,00e ⁺⁰⁰
ETP	0,85*	0,00e ⁺⁰⁰
P-ETP	0,85*	0,00e ⁺⁰⁰
NAC	0,94*	7,96e ⁻³⁴
ARM	0,95*	8,91e ⁻³¹
ALT	0,71*	0,00e ⁺⁰⁰
ETR	0,79*	0,00e ⁺⁰⁰
DEF	0,90*	1,90e ⁻⁴⁰
Umi	0,95*	8,84e ⁻³¹
PLA_BH	0,95*	8,85e ⁻³¹
CAD	0,95*	8,82e ⁻³¹

N* = 3019; Medidas seguidas pelo * foi significativa ao nível de 5% de probabilidade, pelo teste *Shapiro-Wilk*; FCO₂ = Fator de emissão de CO₂ (μmol CO₂ m⁻² s⁻¹); Dp = densidade das partículas (kg dm⁻³); Ds = densidade do solo (kg dm⁻³); Macro = macroporosidade (%), Micro = microporosidade (%), PT = porosidade total (%), PLA = porosidade livre de água (%); Areia = teor de areia total (g kg⁻¹); Silte = teor de silte (g kg⁻¹); Argila = teor de argila (g kg⁻¹); pH = potencial hidrogeniônico (CaCl₂); MOS = matéria orgânica do solo (g dm⁻³); P = fósforo disponível (mg dm⁻³); K = potássio (mmol_c dm⁻³); Ca = cálcio (mmol_c dm⁻³); Mg = magnésio (mmol_c dm⁻³); H-Al = acidez potencial (mmol_c dm⁻³); V= saturação de bases (%); CTC = capacidade de troca de cátions (mmol_c dm⁻³); SB = soma de bases (mmol_c dm⁻³); Corg = carbono orgânico oxidável (g kg⁻¹); Estc = estoque de carbono no solo (Mg ha⁻¹); CAD = capacidade de água disponível (mm); ARM = armazenamento de água no solo (mm); PLA_BH = porosidade livre de água decorrente do balanço hídrico (%); Ts = temperatura do solo (°C); Us = umidade do solo (% do volume); Tair = temperatura do ar até 2 metros do solo (°C); Prect = precipitação (mm); ETP = evapotranspiração potencial (mm), ETR = evapotranspiração real (mm); Umi = umidade do solo oriunda do balanço hídrico (mm); DEF = deficiência de água no sistema solo-planta-atmosfera (mm); ARM = armazenamento de água no solo (mm); ALT = armazenamento de água no solo do mês atual - armazenamento de água no solo do mês anterior (mm); P-ETP = equilíbrio hidrológico potencial do sistema solo-planta-atmosfera (mm).

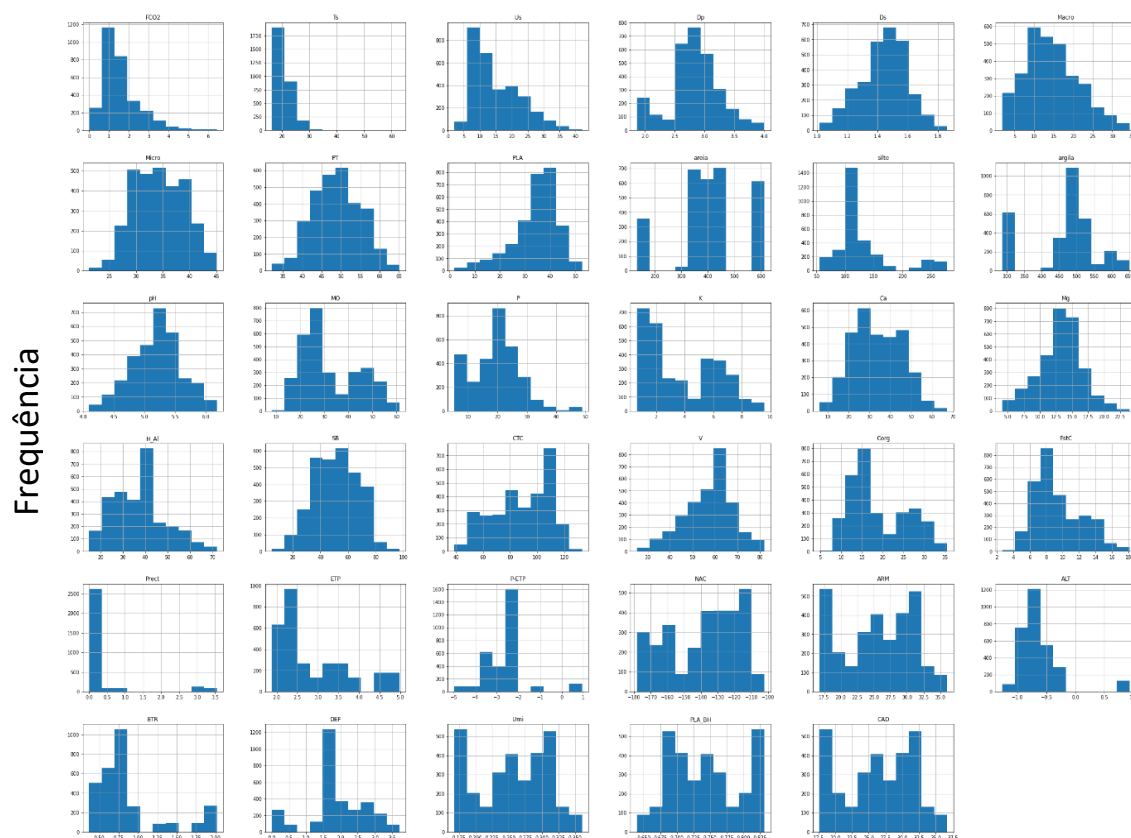


Figura 3.6: Histograma da emissão de CO₂, atributos atmosféricos e características físico-químicos do solo de todas as áreas estudadas, N = 3019.

FCO₂ = Fator de emissão de CO₂ ($\mu\text{mol CO}_2 \text{ m}^{-2} \text{ s}^{-1}$); Dp = densidade das partículas (kg dm^{-3}); Ds = densidade do solo (kg dm^{-3}); Macro = macroporosidade (%), Micro = microporosidade (%), PT = porosidade total (%), PLA = porosidade livre de água (%); Areia = teor de areia total (g kg^{-1}); Silte = teor de silte (g kg^{-1}); Argila = teor de argila (g kg^{-1}); pH = potencial hidrogeniônico (CaCl₂); MOS = matéria orgânica do solo (g dm^{-3}); P = fósforo disponível (mg dm^{-3}); K = potássio ($\text{mmol}_c \text{ dm}^{-3}$); Ca = cálcio ($\text{mmol}_c \text{ dm}^{-3}$); Mg = magnésio ($\text{mmol}_c \text{ dm}^{-3}$); H-Al = acidez potencial ($\text{mmol}_c \text{ dm}^{-3}$); V = saturação de bases (%); CTC = capacidade de troca de cátions ($\text{mmol}_c \text{ dm}^{-3}$); SB = soma de bases ($\text{mmol}_c \text{ dm}^{-3}$); Corg = carbono orgânico oxidável (g kg^{-1}); Estc = estoque de carbono no solo (Mg ha^{-1}); CAD = capacidade de água disponível (mm); ARM = armazenamento de água no solo (mm); PLA_BH = porosidade livre de água decorrente do balanço hídrico (%); Ts = temperatura do solo ($^{\circ}\text{C}$); Us = umidade do solo (% do volume); Tair = temperatura do ar até 2 metros do solo ($^{\circ}\text{C}$); Prect = precipitação (mm); ETP = evapotranspiração potencial (mm), ETR = evapotranspiração real (mm); Umi = umidade do solo oriunda do balanço hídrico (mm); DEF = deficiência de água no sistema solo-planta-atmosfera (mm); ARM = armazenamento de água no solo (mm); ALT = armazenamento de água no solo do mês atual - armazenamento de água no solo do mês anterior (mm); P-ETP = equilíbrio hidrológico potencial do sistema solo-planta-atmosfera (mm).

Também foi realizada a estatística descritiva, visando demonstrar as variações dos atributos estudados. De maneira geral, observou-se que a maioria das variáveis apresentaram valores de coeficiente de variação superiores a 30% (Prect = 289,66 %, K = 63,97 %, ALT = 62,9 %, FCO₂ = 59,75 %, ETR = 50,57 %, Us = 50,41 %, Macro

= 46,66 %, DEF = 44,68 %, P-ETP = 40,93 %, P = 39,42 %, silte = 38,19 %, Corg = 36,57 %, MO = 36,54 %, Ca = 34,38 %, areia = 32,53 %, H-Al = 31,5 %, SB = 30,99 %, EstC = 30,85 %), indicando que os dados tendem a ter uma alta dispersão (heterogêneos). Porém as variáveis ETP (29,47 %), PLA (26,12 %), Mg (25,65 %), CTC (23,3 %), argila (21,36 %), ARM (20,47 %), CAD (19,7 %), Umi (19,23 %), V (18,48 %), NAC (15,37 %), Dp (15,14 %), Ts (13,85 %), Micro (13,63 %), PT (12,51 %), Ds (10,42 %), pH (7,68 %) e PLA_BH (6,76 %) apresentaram valores de CV inferiores a 30 %, indicando que os dados possuíram uma baixa ou média dispersão, o que resultou em dizer que os dados tendem a ser homogêneos (Tabela 3.3).

Tabela 3.3: Estatísticas descritivas da emissão de CO₂, atributos atmosféricos e características físico-químicos do solo de todas as áreas estudadas, N = 3019.

Atributos	Média	Desvio Padrão	Max / Min	CV
FCO₂	1,59	0,95	6,44 / 0	59,75
Ts	21,08	2,92	62,78 / 16,32	13,85
Us	14,66	7,39	42 / 2	50,41
Dp	2,84	0,43	4,01 / 1,86	15,14
Ds	1,44	0,15	1,86 / 1,01	10,42
Macro	14,53	6,78	34,43 / 1,79	46,66
Micro	34,33	4,68	45,1 / 21,22	13,63
PT	48,91	6,12	65,15 / 32,06	12,51
PLA	34,34	8,97	52,63 / 1,83	26,12
Areia	409,65	133,25	613 / 132,5	32,53
Silte	126,8	48,43	278 / 55,22	38,19
Argila	462,36	98,77	651,5 / 286	21,36
pH	5,21	0,4	6,18 / 4,1	7,68
MO	32,54	11,89	61,26 / 8,21	36,54
P	19,71	7,77	48,8 / 5,33	39,42
K	3,83	2,45	9,6 / 0,65	63,97
Ca	33,65	11,57	67,33 / 5,98	34,38
Mg	13,41	3,44	24 / 4,11	25,65
H-Al	37,24	11,73	72 / 14,81	31,50
SB	50,96	15,79	97,11 / 4,7	30,99

CTC	88,92	20,72	133,1 / 38,72	23,30
V	57,24	10,58	82,07 / 25,17	18,48
Corg	18,87	6,9	35,53 / 4,76	36,57
EstC	9,4	2,9	18,14 / 2,62	30,85
Prect	0,29	0,84	3,55 / 0	289,66
ETP	2,85	0,84	4,98 / 1,88	29,47
P-ETP	-2,59	1,06	0,94 / -4,98	40,93
NAC	-138,47	21,28	-102,06 / -178,35	15,37
ARM	25,6	5,24	36,04 / 16,81	20,47
ALT	-0,62	0,39	0,94 / -1,27	62,90
ETR	0,87	0,44	2,02 / 0,37	50,57
DEF	1,88	0,84	3,71 / 0	44,68
Umi	0,26	0,05	0,36 / 0,17	19,23
PLA_BH	0,74	0,05	0,83 / 0,64	6,76
CAD	26,34	5,19	36,68 / 17,64	19,70

Desvpd = desvio padrão; Máx = Máximo; Mín = Mínimo; CV = coeficiente de variação (%); FCO₂ = Fator de emissão de CO₂ (μmol CO₂ m⁻² s⁻¹); Dp = densidade das partículas (kg dm⁻³); Ds = densidade do solo (kg dm⁻³); Macro = macroporosidade (%), Micro = microporosidade (%), PT = porosidade total (%), PLA = porosidade livre de água (%); Areia = teor de areia total (g kg⁻¹); Silte = teor de silte (g kg⁻¹); Argila = teor de argila (g kg⁻¹); pH = potencial hidrogeniônico (CaCl₂); MOS = matéria orgânica do solo (g dm⁻³); P = fósforo disponível (mg dm⁻³); K = potássio (mmol_c dm⁻³); Ca = cálcio (mmol_c dm⁻³); Mg = magnésio (mmol_c dm⁻³); H-Al = acidez potencial (mmol_c dm⁻³); V = saturação de bases (%); CTC = capacidade de troca de cátions (mmol_c dm⁻³); SB = soma de bases (mmol_c dm⁻³); Corg = carbono orgânico oxidável (g kg⁻¹); Estc = estoque de carbono no solo (Mg ha⁻¹); CAD = capacidade de água disponível (mm); ARM = armazenamento de água no solo (mm); PLA_BH = porosidade livre de água decorrente do balanço hídrico (%); Ts = temperatura do solo (°C); Us = umidade do solo (% do volume); Tair = temperatura do ar até 2 metros do solo (°C); Prect = precipitação (mm); ETP = evapotranspiração potencial (mm), ETR = evapotranspiração real (mm); Umi = umidade do solo oriunda do balanço hídrico (mm); DEF = deficiência de água no sistema solo-planta-atmosfera (mm); ARM = armazenamento de água no solo (mm); ALT = armazenamento de água no solo do mês atual - armazenamento de água no solo do mês anterior (mm); P-ETP = equilíbrio hidrológico potencial do sistema solo-planta-atmosfera (mm).

Analisando a variabilidade dos valores da FCO₂, observa-se que a média geral da FCO₂ foi 1,59 μmol CO₂ m⁻² s⁻¹ e o desvio padrão foi 0,95 μmol CO₂ m⁻² s⁻¹. Interpretando a amplitudes dos dados encontrados para essa variável, o valor mínimo encontrado foi 0 e o máximo foi 6,44 μmol CO₂ m⁻² s⁻¹. O coeficiente de variação evidenciou que uma alta variabilidade para a FCO₂, apresentando o valor de 59,75 %,

na Figura 3.7 observa-se o gráfico boxplot, mostrando a distribuição dos dados da FCO_2 .

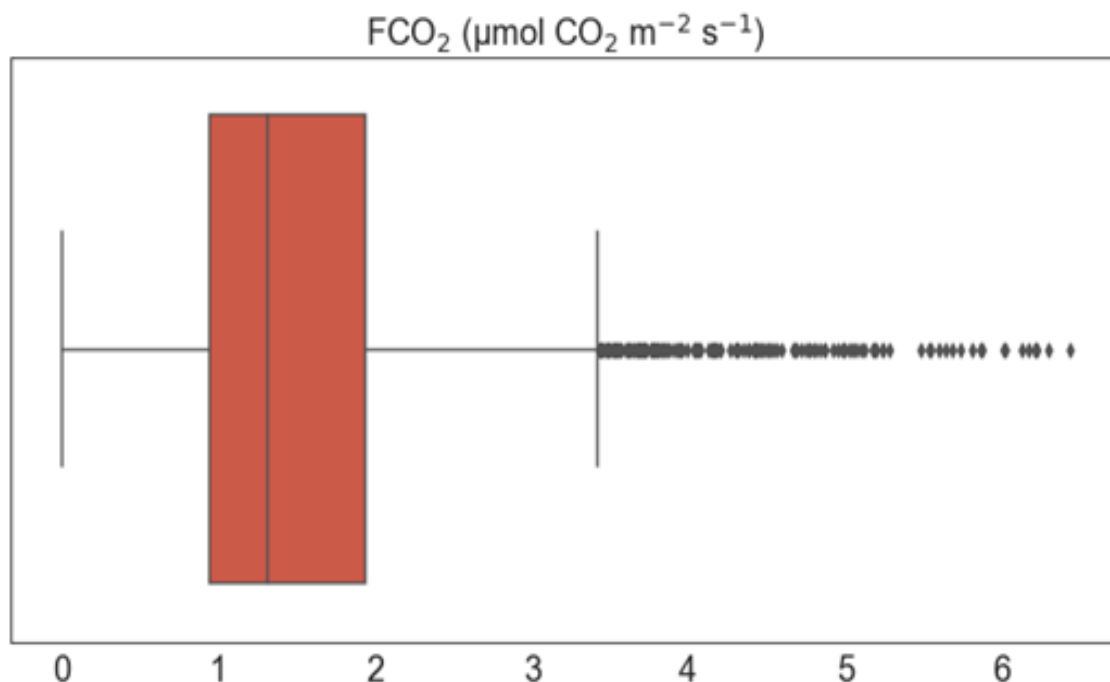


Figura 3.7: Boxplot da emissão do CO₂ do solo (FCO_2 , $\mu\text{mol CO}_2 \text{ m}^{-2} \text{ s}^{-1}$) para todas as áreas estudadas, N = 3019.

3.4.2. Correlações

As correlações variaram entre -1 a 1. Quando as variáveis apresentaram correlações negativas, observou-se que as relações foram inversas, ou seja, se uma variável aumentou o seu índice, a outra diminuiu na mesma proporção. Já quando as relações são positivas, tem-se o aumento diretamente proporcional para as duas variáveis. Pensando nisso, observou que as correlações da FCO_2 foram influenciadas pelos atributos atmosféricos e também pelas características físico-químicas do solo. Devido à complexidade do processo, foi realizado um estudo mais aprofundado para as variáveis que mais correlacionaram com a FCO_2 . Em função da complexidade e das relações encontradas, para a discussão foram estudadas as correlações que apresentaram valores menores que -0,15 e maiores que 0,15.

A FCO_2 apresentou correlações negativas com a D_p (-0,18), D_s (-0,37), PLA (-0,16) e pH (-0,22). As correlações positivas foram observadas com as variáveis U_s (0,26), silte (0,22), K (0,17) e $H\text{-Al}$ (0,17), conforme pode ser observado na Figura 3.8,

mostrando que as variáveis físicas do solo apresentaram uma forte interferência nos processos de deslocamento do CO₂ no solo. O preparo do solo na cultura da cana-de-açúcar é um processo bastante agressivo e causa grandes perdas de carbono do solo, via FCO₂. Moitinho et al. (2015) também investigaram os impactos que os atributos do solo causam no processo de produção e transferência de CO₂ no solo e observaram que a densidade do solo, textura do solo, porosidade livre de água, temperatura do solo e umidade do solo foram as que mais se correlacionaram com este processo.

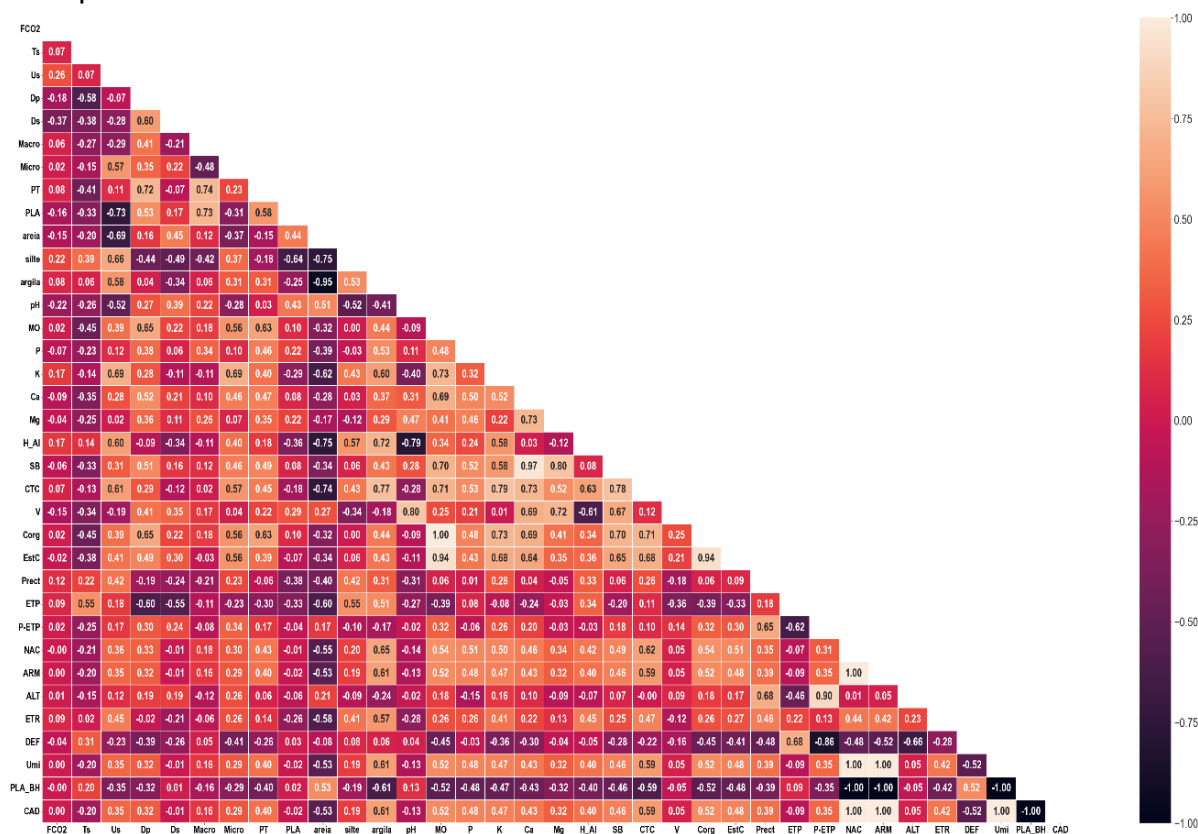


Figura 3.8: Análise das Correlações pelo método de Pearson das variáveis da Emissão de CO₂, atributos atmosféricos e características físico-químicos do solo de todas as áreas estudadas, N = 3019.

FCO₂ = Fator de emissão de CO₂ ($\mu\text{mol CO}_2 \text{ m}^{-2} \text{ s}^{-1}$); Dp = densidade das partículas (kg dm^{-3}); Ds = densidade do solo (kg dm^{-3}); Macro = macroporosidade (%), Micro = microporosidade (%), PT = porosidade total (%), PLA = porosidade livre de água (%); Areia = teor de areia total (g kg^{-1}); Silte = teor de silte (g kg^{-1}); Argila = teor de argila (g kg^{-1}); pH = potencial hidrogeniônico (CaCl_2); MOS = matéria orgânica do solo (g dm^{-3}); P = fósforo disponível (mg dm^{-3}); K = potássio ($\text{mmol}_c \text{ dm}^{-3}$); Ca = cálcio ($\text{mmol}_c \text{ dm}^{-3}$); Mg = magnésio ($\text{mmol}_c \text{ dm}^{-3}$); H-Al = acidez potencial ($\text{mmol}_c \text{ dm}^{-3}$); V = saturação de bases (%); CTC = capacidade de troca de cátions ($\text{mmol}_c \text{ dm}^{-3}$); SB = soma de bases ($\text{mmol}_c \text{ dm}^{-3}$); Corg = carbono orgânico oxidável (g kg^{-1}); Estc = estoque de carbono no solo (Mg ha^{-1}); CAD = capacidade de água disponível (mm);

ARM = armazenamento de água no solo (mm); PLA_BH = porosidade livre de água decorrente do balanço hídrico (%); Ts = temperatura do solo (°C); Us = umidade do solo (% do volume); Tair = temperatura do ar até 2 metros do solo (°C); Prect = precipitação (mm); ETP = evapotranspiração potencial (mm), ETR = evapotranspiração real (mm); Umi = umidade do solo oriunda do balanço hídrico (mm); DEF = deficiência de água no sistema solo-planta-atmosfera (mm); ARM = armazenamento de água no solo (mm); ALT = armazenamento de água no solo do mês atual - armazenamento de água no solo do mês anterior (mm); P-ETP = equilíbrio hidrológico potencial do sistema solo-planta-atmosfera (mm).

3.4.3. Desempenho dos Algoritmos Preditivos

Visando obter a maior acurácia nos modelos, foram usadas todas as variáveis estudadas. Além disso, foram analisados os seguintes algoritmos de machine learning (ML): random forest com a seleção de variáveis pelo método de grid Search (RFT_GS), random forest regressor (RFT), decision tree regression (TREE), gradient boosting regressor (BOO), rede neural Multi-layer Perceptron regressor (MLP), regressão baseada no método k-nearest neighbors (KNN), método de regressão Support vector machines (SVM), Support vector machines com o método de kernel polinomial SVM_kpoli, regressão linear múltipla (RLM), método de regressão Support vector machines com validação cruzada (SVM_VC) e o método de regressão da Stochastic Gradient Descent (SGD). Logo abaixo, na Figura 3.9 observa-se os fluxos dos dados para os algoritmos de aprendizado supervisionado.

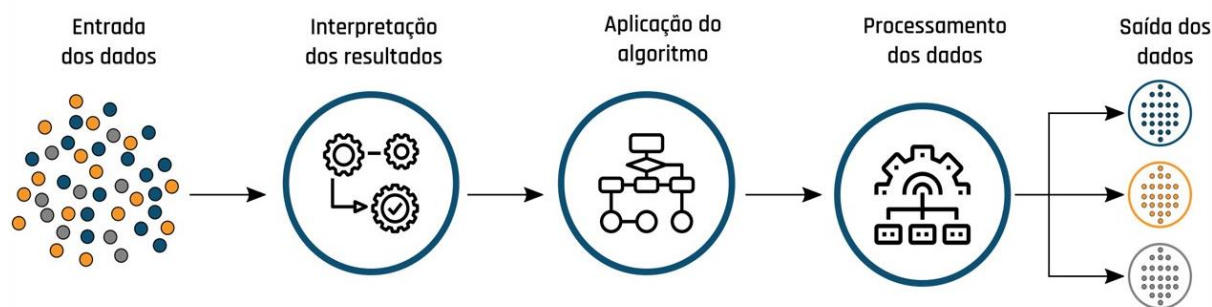


Figura 3.9: Esquematização dos fluxos dos dados na execução dos algoritmos de aprendizado supervisionado.

Em todos os casos, foram realizadas as etapas de treino para calibrar o modelo e depois foram realizados os testes com o intuito de validar os resultados. Para medir as acurácias dos modelos, foram utilizadas as seguintes métricas qualitativas: coeficiente de determinação (R^2), coeficiente de determinação ajustado (R^2_{ajust}), erro quadrático médio (MSE), raiz quadrada do erro-médio (RMSE), média dos erros

absolutos (MAE). Na Tabela 3.4 são apresentados os resultados numéricos da análise qualitativa, por meio das métricas estatísticas obtidas para cada algoritmo de aprendizado de máquina.

Tabela 3.4: Índices estatísticos dos algoritmos utilizados para a previsão da FCO₂ do solo, em função dos atributos atmosféricos e características físico-químicos do solo de todas as áreas estudadas, N = 3019.

INDEXES	RFT_GS	RFT	TREE	MLP	BOO	KNN	SVM	SVM_kpoli	RLM	SVM_VC	SGD
TREINO (n = 2415)											
R²	0,9390	0,9402	1,0000	0,6972	0,8405	0,4760	0,9137	0,5832	0,2580	0,1845	0,1831
R²ajust	0,9383	0,9395	1,0000	0,6938	0,8387	0,4701	0,9127	0,5785	0,2496	0,1752	0,1738
MSE	0,0563	0,0552	5,5787E ⁻³¹	0,2751	0,1472	0,4835	0,0797	0,3846	0,6280	0,6902	0,6914
RMSE	0,2372	0,2349	7,4691E ⁻¹⁶	0,5145	0,3837	0,6953	0,2823	0,6202	0,7924	0,8308	0,8315
MAE	0,1436	0,1422	5,3644E ⁻¹⁶	0,3428	0,2509	0,4581	0,1356	0,3368	0,5737	0,5619	0,5956
TESTE (n = 604)											
R²	0,5996	0,5879	0,5589	0,5051	0,5013	0,3633	0,3268	0,3049	0,2370	0,1760	0,1057
R²ajust	0,5950	0,5832	0,5539	0,4995	0,4956	0,3560	0,3192	0,2970	0,2283	0,1666	0,0955
MSE	0,3389	0,3488	0,3733	0,4474	0,4221	0,5389	0,5698	0,5883	0,7041	0,7604	0,8253
RMSE	0,5822	0,5906	0,6110	0,6689	0,6497	0,7341	0,7548	0,7670	0,8391	0,8720	0,9084
MAE	0,3762	0,3827	0,3945	0,4552	0,4179	0,5111	0,5084	0,4775	0,5920	0,5597	0,6415

R² = coeficiente de determinação; R²ajust = coeficiente de determinação ajustado; MSE = erro quadrático médio; RMSE = raiz quadrada do erro-médio; MAE = média dos erros absolutos; RFT_GS = random forest com a seleção de variáveis pelo método de grid Search; RFT = random forest regressor; TREE = decision tree regression; BOO = gradient boosting regressor; MLP = rede neural Multi-layer Perceptron regressor; KNN = regressão baseada no método k-nearest neighbors; SVM = método de regressão support vector machines; SVM_kpoli = support vector machines com o método de kernel polinomial; RLM = regressão linear múltipla; SVM_VC = método de regressão Support vector machines com validação cruzada; SGD = método de regressão da Stochastic Gradient Descent. Números em negrito referem-se ao melhor resultado das métricas estatísticas relevantes entre os algoritmos de aprendizado de máquina estudados.

3.4.4. Treinamento dos modelos

Todos os resultados aqui discutidos podem ser observados na Tabela 3.4. Inicialmente foi analisado o R² que é a métrica utilizada para determinar a precisão dos dados estimados, quando comparados com os dados observados, geralmente expressados nas curvas de previsões. Considerando o valor de R² dos diferentes algoritmos preditivos, relacionados com as variáveis atmosféricas, da planta e dos atributos físico-químicos do solo, verificou-se que os valores de R² variaram de 1,0000 a 0,1831 para a FCO₂. De maneira, foi observado que a TREE foi o algoritmo que

apresentou o melhor desempenho, apresentando a maior acurácia para a estimativa da FCO₂ (1,0000), seguido da RFT (0,9402), da RFT_GS (0,9390), da SVM (0,9137), do BOO (0,8405), da KNN (0,4760), da MLP (0,6972), da SVM_kpoli (0,5832), da RLM (0,2580), da SVM_VC (0,1845) e da SGD (0,1831). Esses resultados obtidos se assemelharam a outros resultados obtidos em pesquisas relacionadas a emissão de CO₂ no solo (Hamrani et al., 2020; Ağbulut, 2022; Adjuik e Davis, 2022).

O R^2_{ajust} é derivado do R^2 , calculado e ajustado em função do número de preditores no modelo. O aumento do R^2_{ajust} só será observado se um novo fator melhorar o modelo mais do que seria previsto pela casualidade. Em contrapartida, observa-se a sua diminuição quando uma variável não colabora com a melhoria esperada do modelo devido à casualidade. Vale a pena confirmar, que é sempre menor que o R^2 (Piepho, 2019). Analisando o R^2_{ajust} , os resultados variaram de 1,0000 a 0,1738, observando-se um padrão semelhante ao R^2 . O TREE também apresentou a maior performance (1,0000), seguido da RFT (0,9395), da RFT_GS (0,9385), da SVM (0,9127), do BOO (0,8387), da MLP (0,6938), da SVM_kpoli (0,5785), da KNN (0,4701), da RLM (0,2496), da SVM_VC (0,1752) e da SGD (0,1738).

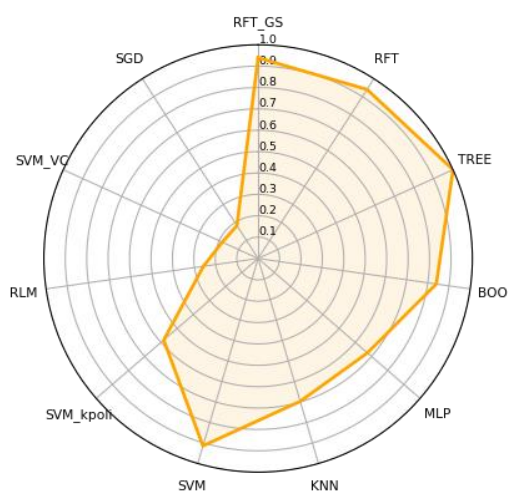
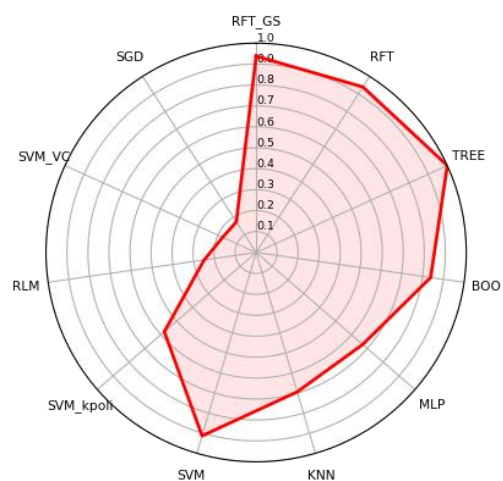
Outra métrica estatística utilizada neste estudo foi a métrica MSE. Essa medida penaliza os resultados que sejam muito distintos entre o estimado e o real. Dessa forma, quanto maior é o valor de MSE, menos o modelo performou em função das previsões (Bruce e Bruce, 2019). Os resultados numéricos da métrica MSE encontrados apresentaram valores baixos, muito próximos de zero, o que é esperado para a MSE de cada algoritmo, variando de 0,6914 a $5,5787E^{-31}$. O algoritmo que apresentou o melhor resultado foi novamente foi o TREE ($5,5787E^{-31}$), seguido da RFT (0,0552), da RFT_GS (0,0563), da SVM (0,0797), do BOO (0,1472), da MLP (0,2751), da SVM_kpoli (0,3846), da KNN (0,4835), da RLM (0,6280), da SVM_VC (0,6902) e da SGD (0,6914). Foi observado que os algoritmos MLP, SVM_poli, KNN, SVM alteraram as suas posições, quando comparado com o R^2 e R^2 ajustado, porém os restantes mantiveram as mesmas ordens.

Além dessas métricas, o RMSE também foi avaliado para determinar o sucesso dos algoritmos. Para a sua obtenção, realiza-se a raiz quadrada do MSE, visando obter a penalização entre diferenças grandes do valor previsto e o real, fazendo com

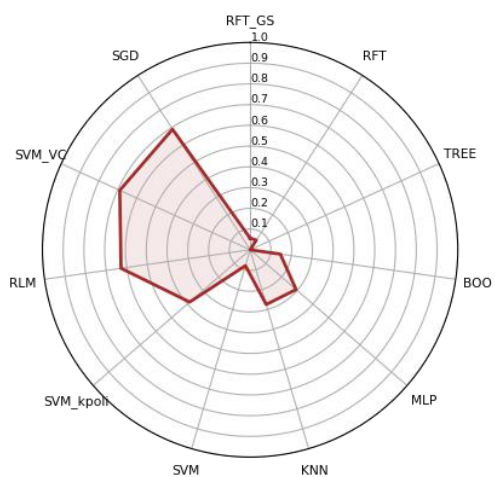
que a unidade permaneça igual à do dado original, auxiliando no processo de interpretação do resultado. O RMSE leva os valores de zero a ∞ , e os pequenos valores de RMSE são desejáveis (Bakay e Ağbulut, 2021; Zang et al., 2020). Conforme a Tabela 3.4, pode ser observado o mesmo padrão dos resultados da MSE, a variação observada foi de 0,8315 a $7,4691E^{-16}$. O algoritmo que apresentou o melhor desempenho foi a TREE ($7,4691E^{-16}$), seguido da RFT (0,2372), da RFT_GS (0,2349), da SVM (0,2823), do BOO (0,3837), da MLP (0,5145), da SVM_kpoli (0,6202), da KNN (0,6953), da RLM (0,7924), da SVM_VC (0,8308) e da SGD (0,8315).

Por fim, também foi analisado o MAE que calcula a média da diferença entre o valor real e o valor estimado. Essa medida é dada por meio do módulo entre os valores obtidos, pois são observados valores positivos e negativos. Além disso, esta métrica não é influenciada pelos outliers, que comumente identificados como valores discrepantes (Ma et al., 2019). A variação deste coeficiente métrico foi de 0,5956 a $5,3644E^{-16}$, novamente a TREE apresentou o melhor desempenho ($5,3644E^{-16}$), porém os modelos que estavam na região mais central sofreram alterações de posições, quando comparado com as outras métricas, sendo observado o SVM (0,1356) como o segundo algoritmo, seguido da RFT (0,1422), RFT_GS (0,1436), do BOO (0,2509), da SVM_kpoli (0,3368), da MLP (0,3428), da KNN (0,4581), da SVM_VC (0,5619), da RLM (0,5737) e da SGD (0,5956).

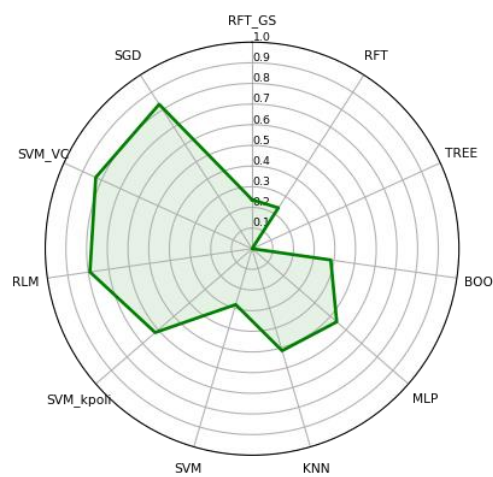
Logo abaixo, na Figura 3.10, pode ser observado no gráfico de radar uma visão resumida dos resultados de cada métrica estatística para a FCO_2 , em função dos atributos atmosféricos e características físico-químicos do solo de todas as áreas estudadas, obtida em cada algoritmo estudado, na fase de calibração dos modelos.

A) R^2 B) R^2_{ajust} 

C) MSE



D) RMSE



E) MAE

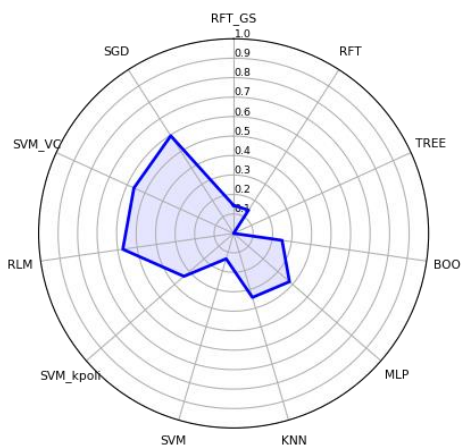


Figura 3.10: Visão geral da distribuição de métricas estatísticas em um gráfico de radar escalonado para emissão de CO₂ do solo, em função dos atributos atmosféricos e características físico-químicos do solo de todas as áreas estudadas, na fase de calibração dos modelos, n = 2415.

A) R^2 = coeficiente de determinação; B) R^2_{ajust} = coeficiente de determinação ajustado; C) MSE = erro quadrático médio; D) RMSE = raiz quadrada do erro-médio; E) MAE = média dos erros absolutos; RFT_GS = random forest com a seleção de variáveis pelo método de grid Search; RFT = random forest regressor; TREE = decision tree regression; BOO = gradient boosting regressor; MLP = rede neural Multi-layer Perceptron regressor; KNN = regressão baseada no método k-nearest neighbors; SVM = método de regressão support vector machines; SVM_kpoli = support vector machines com o método de kernel polinomial; RLM = regressão linear múltipla; SVM_VC = método de regressão Support vector machines com validação cruzada; SGD = método de regressão da Stochastic Gradient Descent.

Após essa análise detalhada de todas métricas estatísticas, observou-se que os algoritmos apresentaram diferentes níveis de desempenho. Fazendo uma análise exploratória, e utilizando o critério do R^2 ajustado foram definidas 3 classes:

- Classe 1: $R^2_{ajustado} > 0,7$, bom desempenho, tem-se o TREE, RFT, RFT_GS, SVM e o BOO;
- Classe 2: $0,3 < R^2_{ajustado} < 0,7$, desempenho médio, tem-se o MLP, SVM_kpoli e o KNN;
- Classe 3: $R^2_{ajustado} < 0,3$, desempenho ruim, tem-se RLM, SVM_VC e o SGD.

Dessa forma, pode-se dizer que foram 5 algoritmos de aprendizado de máquina que apresentaram os melhores desempenhos, sendo que nessa fase de treinamento o TREE foi o modelo mais eficiente. Outros autores encontraram resultados semelhantes nas suas predições, em estudos relacionados com a emissão de CO_2 (Lu et al., 2021; Adjuik e Davis, 2022; Hamzacebi e Karakurt, 2015). Na Figura 3.11 pode ser observado o gráfico da comparação da dispersão dos dados observados versus os dados estimados, para os onze modelos que foram testados nesta fase de treinamento.

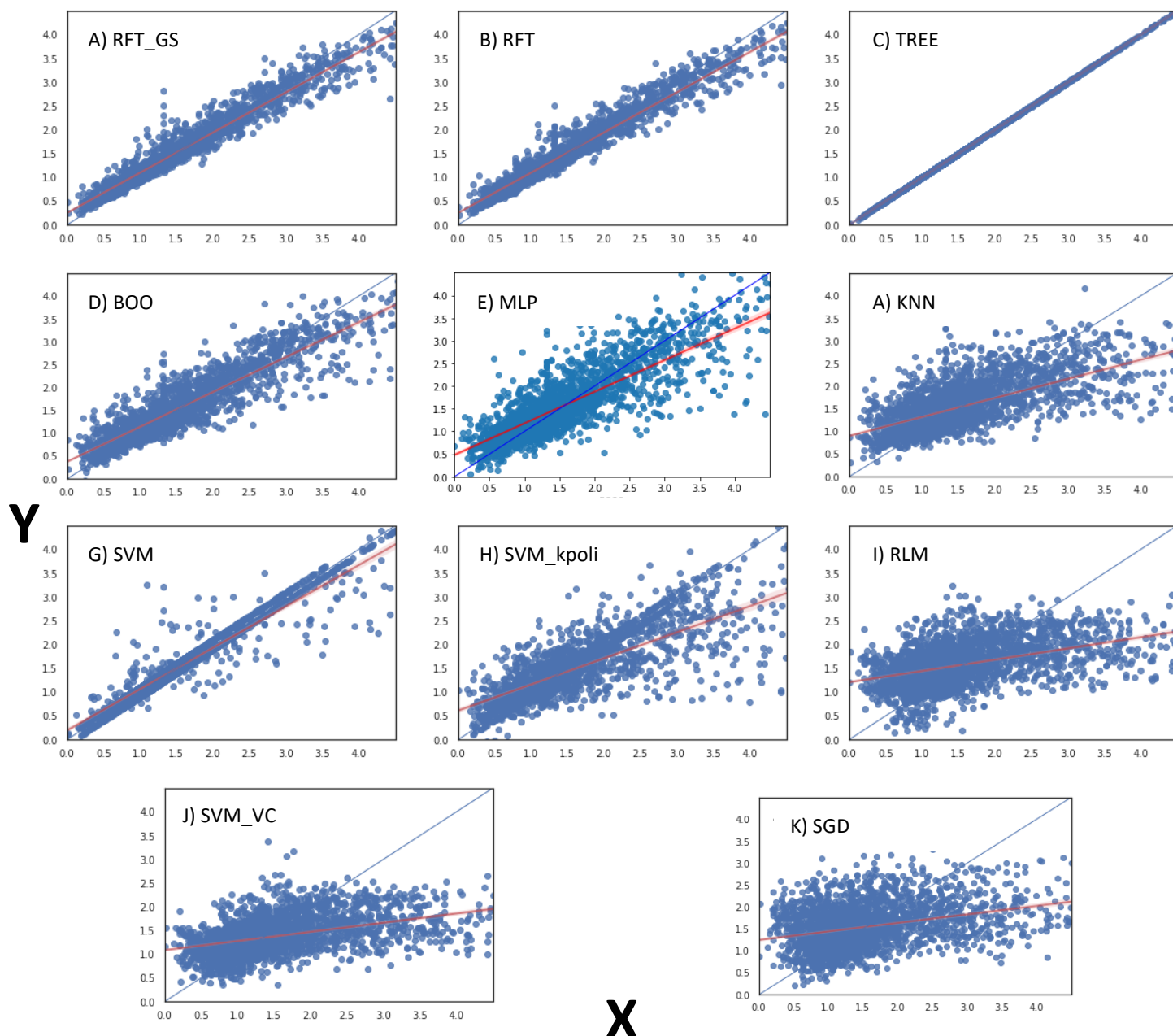


Figura 3.11: Gráficos de dispersão da emissão da CO₂ do solo observado versus predita dos modelos estudados nos conjuntos de dados de calibração ($n = 2415$). A) RFT_GS = random forest com a seleção de variáveis pelo método de grid Search; B) RFT = random forest regressor; C) TREE = decision tree regression; D) BOO = gradient boosting regressor; E) MLP = rede neural Multi-layer Perceptron regressor; F) KNN = regressão baseada no método k-nearest neighbors; G) SVM = método de regressão support vector machines; H) SVM_kpoli = support vector machines com o método de kernel polinomial; I) RLM = regressão linear múltipla; J)

SVM_VC = método de regressão Support vector machines com validação cruzada; K)
SGD = método de regressão da Stochastic Gradient Descent.

3.4.5. Validação dos modelos

De maneira geral, para todos os algoritmos foram observadas as reduções dos desempenhos das métricas estatísticas, quando comparado com a fase de treinamento (Tabela 3.4). Esse resultado é originado devido à penalização sobre o ajustamento dos dados. Em outras pesquisas que usaram os algoritmos de ML, em estudos relacionados ao CO₂ no solo, também observaram as diminuições das métricas estatísticas na fase validação dos dados (Saha et al., 2021; Chen et al., 2018; Warner et al., 2019 b).

Assim como foi visto na fase treinamento, todos os resultados aqui discutidos podem ser observados na Tabela 3.4. A primeira análise de desempenho analisada foi R^2 . Os valores encontrados de R^2 nos algoritmos para prever a FCO₂, a partir das variáveis atmosféricas, da planta e dos atributos físico-químicos do solo variaram de 0,5996 a 0,1057. A RFT_GS foi o algoritmo que apresentou o melhor desempenho (0,5996), seguido do RFT (0,5879), da TREE (0,5589), do MLP (0,5051), do BOO (0,5013), da KNN (0,3633), do SVM (0,3268), do SVM_kpoli (0,3049), da RLM (0,2370), da SVM_VC (0,1760) e da SGD (0,1057).

Quando foi analisado o R^2_{ajust} , observou-se o mesmo padrão do R^2 . A variação da R^2_{ajust} foi de 0,5950 a 0,0955. O RFT_GS também apresentou o melhor desempenho dentre os estudados (0,5950), seguido do RFT (0,5832), da TREE (0,5539), do MLP (0,4995), do BOO (0,4956), da KNN (0,3560), do SVM (0,3192), do SVM_kpoli (0,2970), da RLM (0,2283), da SVM_VC (0,1666) e da SGD (0,0955).

Para a MSE foram obtidos valores baixos, assim como foi observado na fase de treinamento do modelo. A variação encontrada foi de 0,8253 a 0,3389. O algoritmo que apresentou a melhor acurácia novamente foi o RFT_GS (0,3389), seguido do RFT (0,3488), da TREE (0,3733), do BOO (0,4221), do MLP (0,4474), da KNN (0,5389), do SVM (0,5698), do SVM_kpoli (0,5883), da RLM (0,7041), da SVM_VC (0,7604) e da SGD (0,8253). Foi observado que os algoritmos MLP e BOO trocaram de posições entre si, quando comparado com o R^2 e R^2_{ajust} , porém os restantes mantiveram as mesmas ordens.

Para o RMSE também foi observado que o mesmo padrão dos resultados da MSE, onde a variação foi de 0,5822 a 0,9084, sendo que o algoritmo que apresentou o melhor desempenho foi o RFT_GS (0,5822), seguido do RFT (0,5906), da TREE (0,6110), do BOO (0,6497), do MLP (0,6689), da KNN (0,7341), do SVM (0,7548), do SVM_kpoli (0,7670), da RLM (0,8391), da SVM_VC (0,8720) e da SGD (0,9084).

Finalizando, também foi analisado o MAE que apresentou uma variação de 0,3762 a 0,6415. Novamente a RFT_GS apresentou o melhor desempenho (0,3762), porém os modelos que estavam na região mais central sofreram alterações de posições, quando comparado com as outras métricas, sendo observado o RFT (0,3827) como o segundo algoritmo, seguido por TREE (0,3945), BOO (0,4179), do MLP (0,4552), da SVM_kpoli (0,4775), da SVM (0,5084), da KNN (0,5111), da SVM_VC (0,5597), da RLM (0,5920) e da SGD (0,6415).

A partir desta análise metódica de todas as métricas estatísticas, ficou evidente as diferenças nos níveis de desempenho dos algoritmos. Realizando uma análise exploratória, e utilizando o critério do R^2 ajustado para a definição de 3 classes, assim como foi feito para a fase de treinamento, temos as seguintes classes:

- Classe 1: $R^2_{ajustado} > 0,5$, bom desempenho, tem-se o RFT_GS, RFT e TREE;
- Classe 2: $0,5 > R^2_{ajustado} > 0,3$, desempenho médio, tem-se o MLP, BOO, KNN e SVM;
- Classe 3: $R^2_{ajustado} < 0,3$, desempenho ruim, tem-se o SVM_kpoli, RLM SVM_VC e o SGD.

Na Figura 3.12 observa-se o gráfico de radar que traz uma visão resumida dos coeficientes métricos estatísticos obtidos em cada algoritmo para a FCO₂, em função dos atributos atmosféricos e características físico-químicos do solo de todas as áreas estudadas, obtida em cada algoritmo estudado, na fase de validação dos modelos.

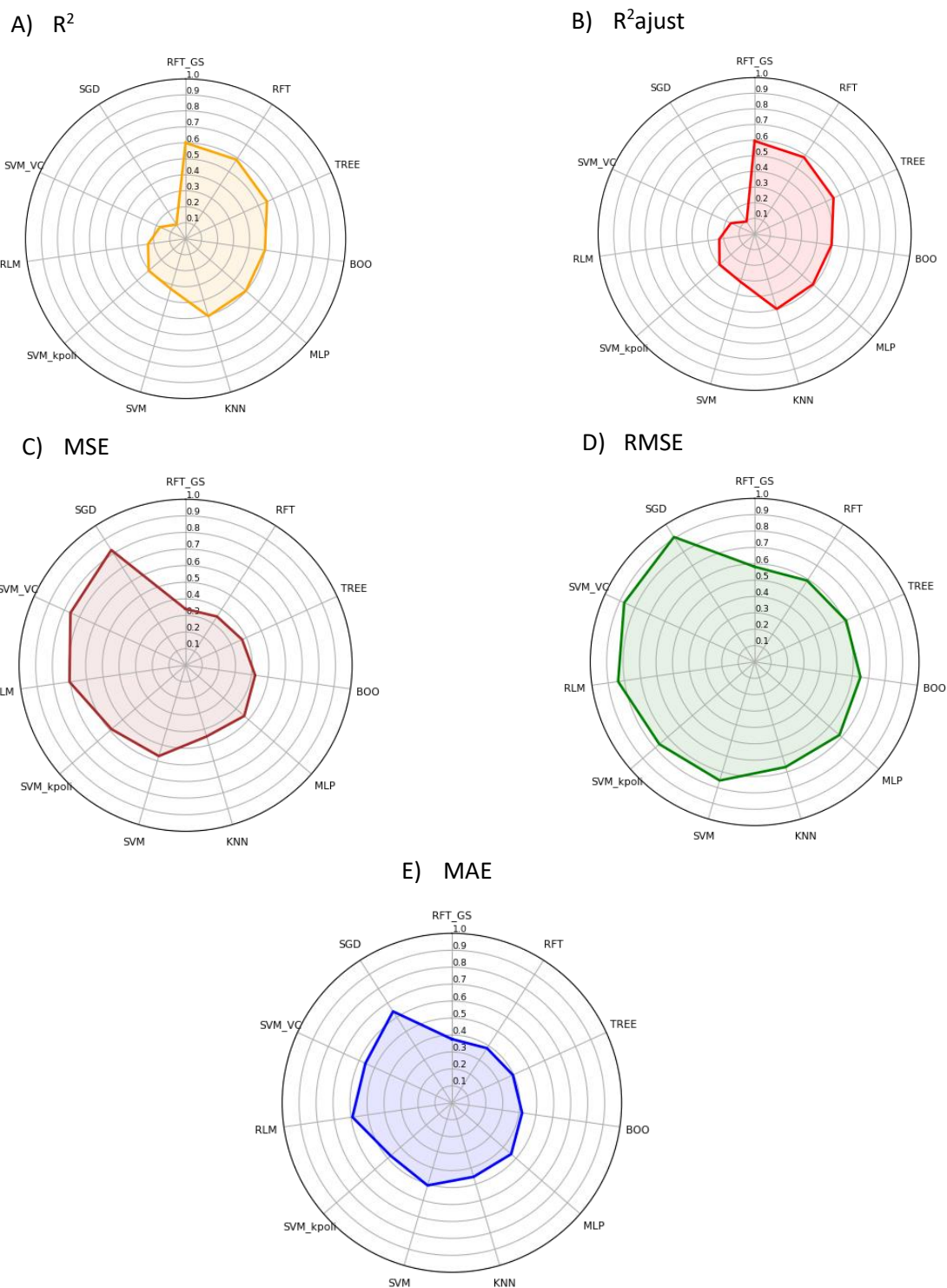


Figura 3.12: Uma visão geral da distribuição de métricas estatísticas em um gráfico de radar escalonado para emissão de CO_2 do solo, em função dos atributos atmosféricos e características físico-químicos do solo de todas as áreas estudadas, na fase de validação dos modelos, $n = 604$.

A) R^2 = coeficiente de determinação; B) R^2_{ajust} = coeficiente de determinação ajustado; C) MSE = erro quadrático médio; D) RMSE = raiz quadrada do erro-médio; E) MAE = média dos

erros absolutos; RFT_GS = random forest com a seleção de variáveis pelo método de grid Search; RFT = random forest regressor; TREE = decision tree regression; BOO = gradient boosting regressor; MLP = rede neural Multi-layer Perceptron regressor; KNN = regressão baseada no método k-nearest neighbors; SVM = método de regressão support vector machines; SVM_kpoli = support vector machines com o método de kernel polinomial; RLM = regressão linear múltipla; SVM_VC = método de regressão Support vector machines com validação cruzada; SGD = método de regressão da Stochastic Gradient Descent.

Conforme pode ser observado nas informações acima, houve a diminuição da amplitude dos coeficientes métricos obtidos, resultando na adaptação das partições das classes. Os melhores desempenhos foram observados em 5 algoritmos de aprendizado de máquina (RFT_GS, RFT, TREE), sendo que após a validação, o melhor modelo obtido neste estudo o RFT_GS. Abbasi et al. (2021) compararam 6 modelos de ML por meio do seu desempenho preditivo (operador de seleção e encolhimento mínimo absoluto (LASSO), rede neural de alimentação direta (FNN), rede neural de função de base radial (RBFNN), máquina de aprendizado extremo (ELM), SVM e RFT), para as medições no campo das emissões de CO₂, sobre os tratamentos de dois tipos de fertilizantes: fertilizante inorgânico (IF) e esterco sólido aditivado com fertilizante inorgânico (SCM), aplicado a uma rotação de milho-soja e observaram que a melhor previsão das emissões do CO₂ foi realizada pelo algoritmo RFT. Outros autores, em estudos relacionados as emissões de CO₂ do solo também observaram que o algoritmo RFT foi o que melhor conseguir predizer esse processo, quando comparado com outros modelos de ML (Houshmandfar et al., 2021; Chen et al., 2021; Pham et al., 2021).

Na Figura 3.13 pode ser observado o gráfico com as comparações das dispersões dos dados observados versus os dados estimados para os onze modelos que foram calibrados nesta fase de validação.

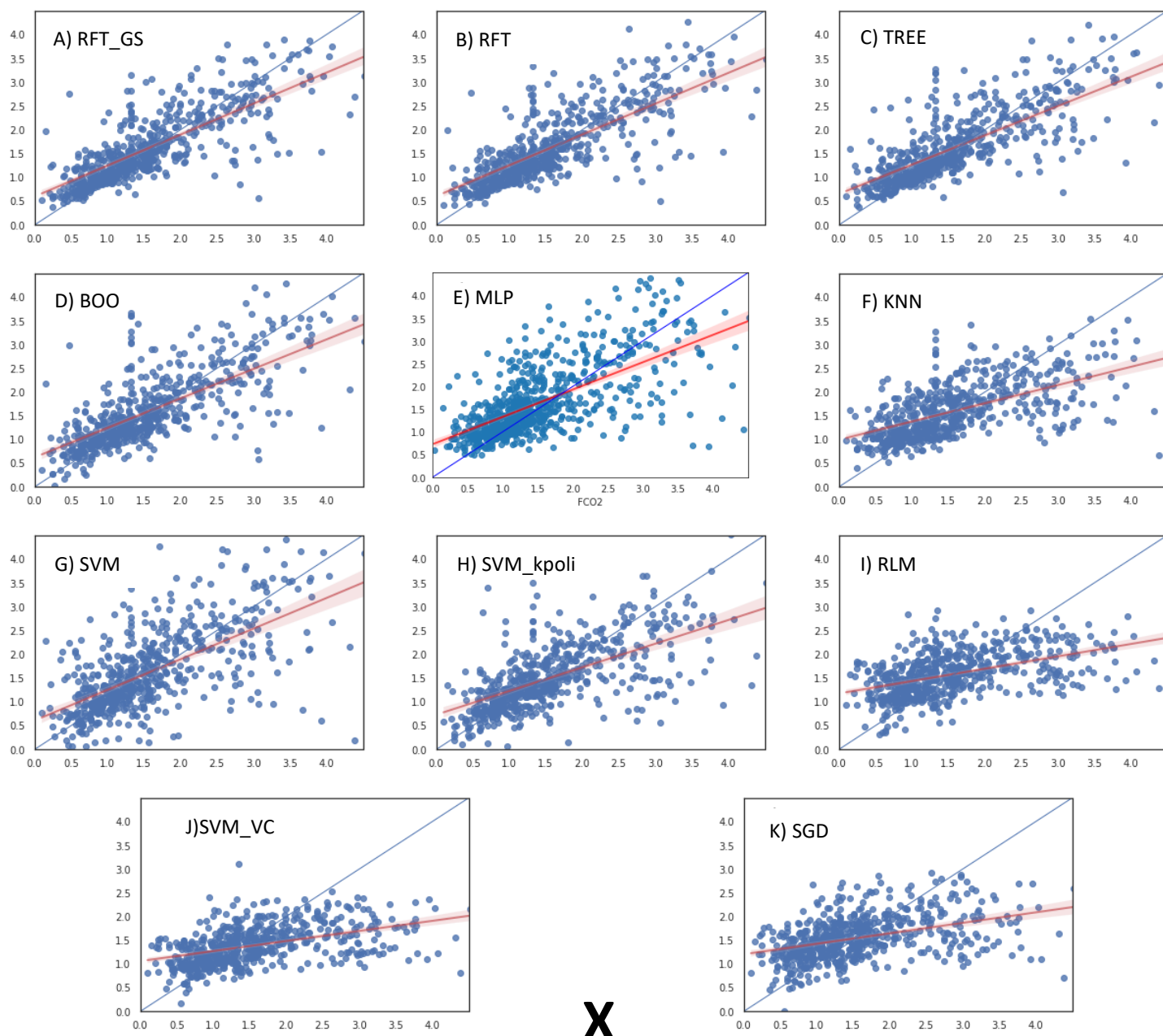


Figura 3.13: Gráficos de dispersão de CO₂ do solo observado versus predita dos modelos estudados nos conjuntos de dados de teste (n = 604).

A) RFT_GS = random forest com a seleção de variáveis pelo método de grid Search; B) RFT = random forest regressor; C) TREE = decision tree regression; D) BOO = gradient boosting regressor; E) MLP = rede neural Multi-layer Perceptron regressor; F) KNN = regressão baseada no método k-nearest neighbors; G) SVM = método de regressão support vector machines; H) SVM_kpoli = support vector machines com o método de kernel polinomial; I) RLM = regressão linear múltipla; J) SVM_VC = método de regressão Support vector machines com validação cruzada; K) SGD = método de regressão da Stochastic Gradient Descent.

Além disso, foi realizada a análise de resíduos, que nada mais é do que a relação entre o valor observado da variável dependente (y) e o valor previsto ($\hat{y}$). De

maneira geral, observa-se que os modelos apresentaram moderada performance, sendo que o modelo que melhor conseguiu prever foi o RFT_GS. Para o melhor modelo estimado (RFT_GS), foi observado que o valor médio estimado de FCO_2 na fase de treinamento foi de $1,5979 \mu\text{mol CO}_2 \text{ m}^{-2} \text{ s}^{-1}$, já o valor encontrado na fase de teste foi $1,6375 \mu\text{mol CO}_2 \text{ m}^{-2} \text{ s}^{-1}$, evidenciando que a validação do modelo sobrestimou os resultados. Na Figura 3.14, são observados os gráficos com o valor observado da variável dependente (y) versus valor previsto ($\hat{y}$), para todos os algoritmos preditivos estudados.

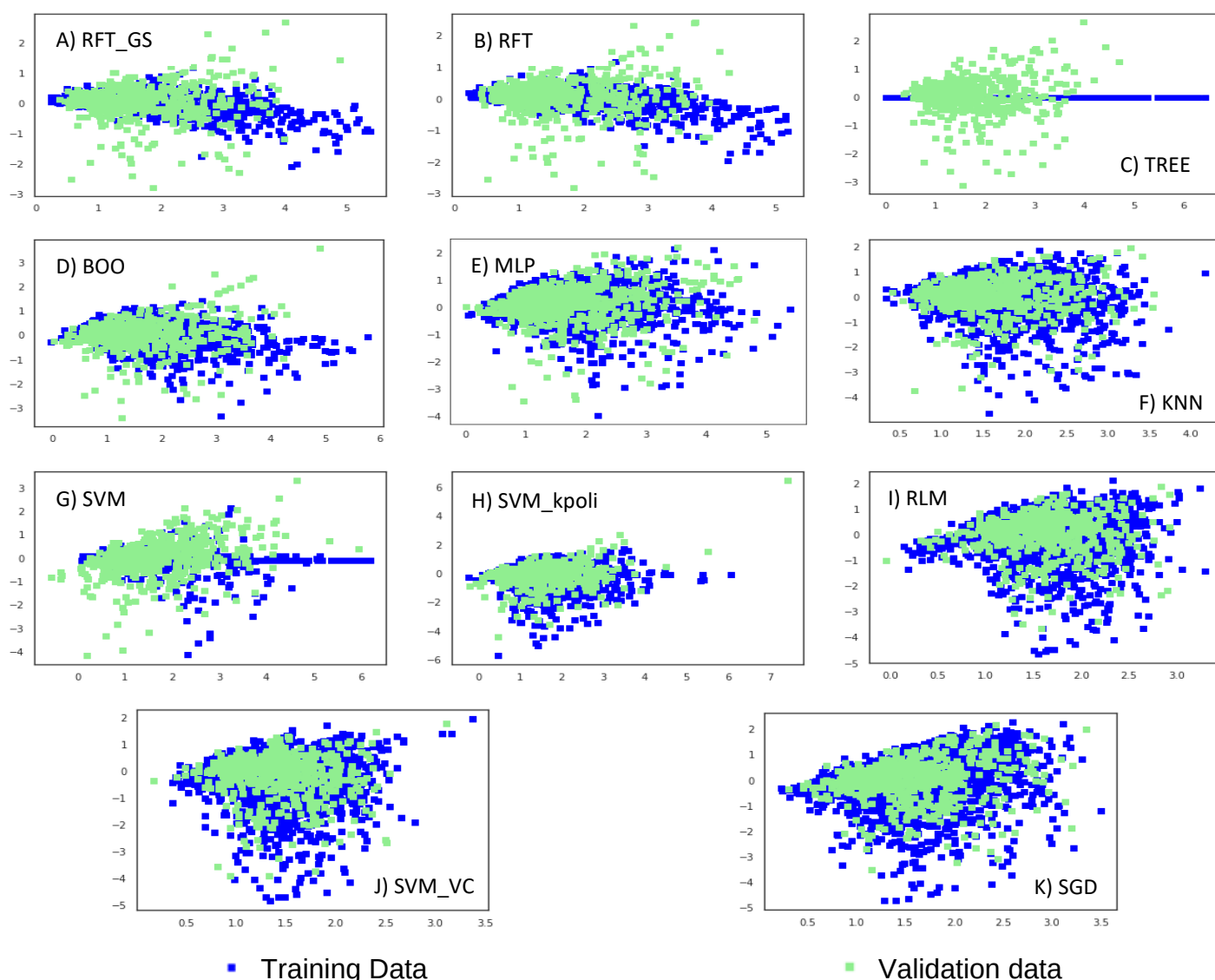


Figura 3.14: Análise de resíduos dos dados observados FCO_2 versus os dados estimados de FCO_2 , para todos os algoritmos preditivos estudados ($n = 3019$).

A) RFT_GS = random forest com a seleção de variáveis pelo método de grid Search; B) RFT = random forest regressor; C) TREE = decision tree regression; D) BOO = gradient boosting

regressor; E) MLP = rede neural Multi-layer Perceptron regressor; F) KNN = regressão baseada no método k-nearest neighbors; G) SVM = método de regressão support vector machines; H) SVM_kpoli = support vector machines com o método de kernel polinomial; I) RLM = regressão linear múltipla; J) SVM_VC = método de regressão Support vector machines com validação cruzada; K) SGD = método de regressão da Stochastic Gradient Descent.

Além disso, foi realizada a análise de resíduos, que nada mais é do que os estudos dos erros do valor observado da variável dependente (y) e o valor previsto ($\hat{y}$) pelo algoritmo. Também foram observadas diferentes performances quando comparados os algoritmos estudados, evidenciando que o algoritmo que apresentou os maiores erros médios foi o SGD, implicando dizer que esse é o modelo menos eficiente dentre os estudados. Já para o RFT_GS foi observado que o erro médio obtido na fase de treinamento foi $0,0069 \mu\text{mol CO}_2 \text{ m}^{-2} \text{ s}^{-1}$, já o erro médio encontrado na fase de validação do modelo foi de $0,0424 \mu\text{mol CO}_2 \text{ m}^{-2} \text{ s}^{-1}$, o que resultou na diminuição da acurácia do modelo na fase de validação, quando comparado com a etapa de calibração do modelo. O mesmo processo aconteceu com os demais modelos estudados. Na Figura 3.15, são observados os gráficos com o valor observado da variável dependente (y) versus valor previsto ($\hat{y}$), para todos os algoritmos preditivos estudados.

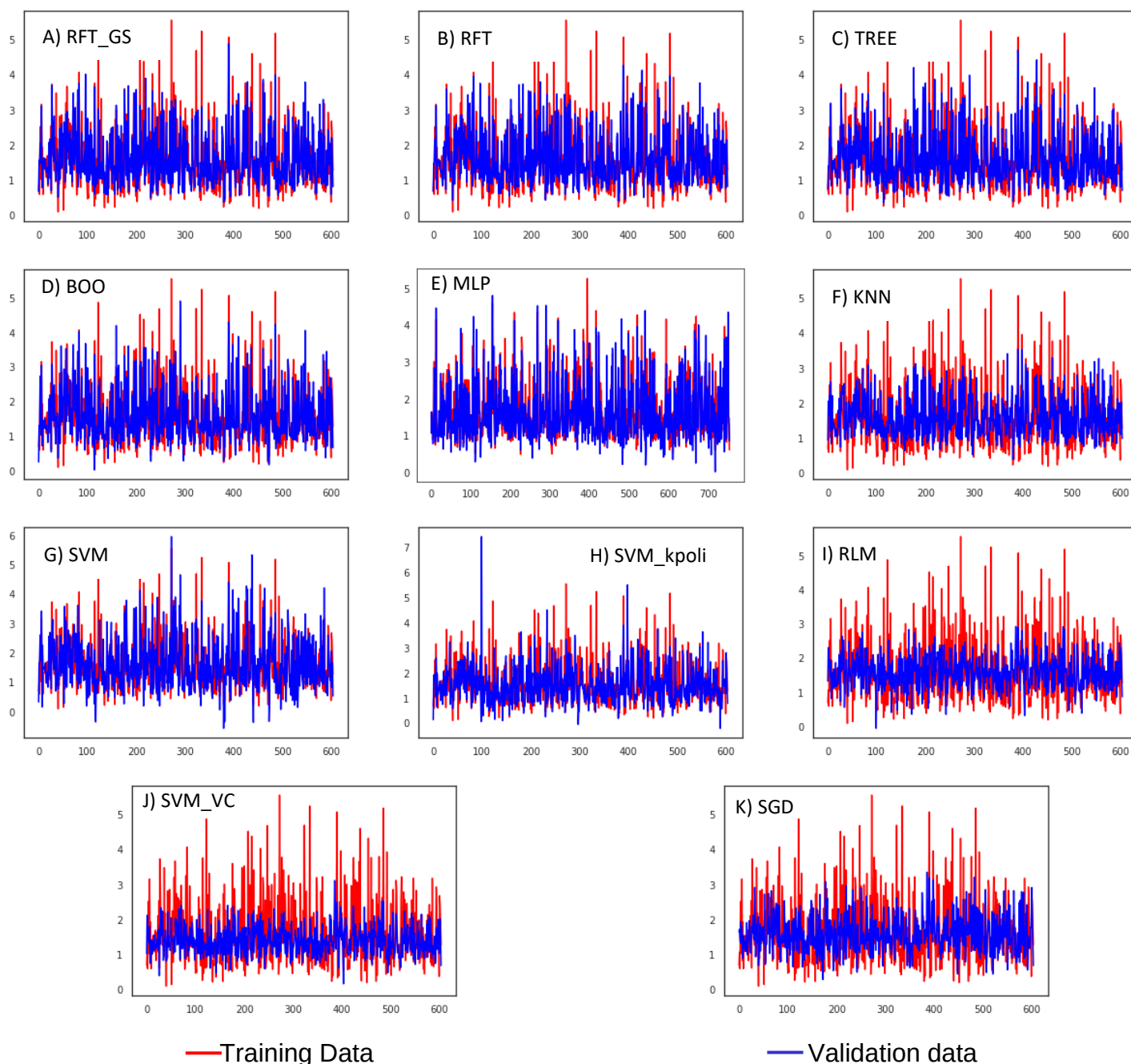


Figura 3.15: Dados observados FCO_2 versus os dados estimados de FCO_2 , para todos os algoritmos preditivos estudados ($n = 604$).

A) RFT_GS = random forest com a seleção de variáveis pelo método de grid Search; B) RFT = random forest regressor; C) TREE = decision tree regression; D) BOO = gradient boosting regressor; E) MLP = rede neural Multi-layer Perceptron regressor; F) KNN = regressão baseada no método k-nearest neighbors; G) SVM = método de regressão support vector machines; H) SVM_kpoli = support vector machines com o método de kernel polinomial; I) RLM = regressão linear múltipla; J) SVM_VC = método de regressão Support vector machines com validação cruzada; K) SGD = método de regressão da Stochastic Gradient Descent.

3.5. CONCLUSÕES

Foi observado que a emissão da CO₂ do solo é um processo bastante complexo e dependente do sistema solo-planta-atmosfera. Esses tipos de estudos colaboram com o entendimento e mitigação da emissão não só do CO₂ do solo, mas também de todos os gases causadores do efeito estufa.

Os processos de mineração, seleção, organização dos dados e as escolhas dos parâmetros nos algoritmos de ML são determinantes para as obtenções de modelos mais acurados, e o GridSearchCV se mostrou uma ferramenta bastante eficiente na seleção automatizada dos parâmetros dos algoritmos de ML, contribuindo com o aumento da acuraria dos modelos.

A FCO₂ apresentou correlações negativas com a Dp (-0,18), Ds (-0,37), PLA (-0,16) e pH (-0,22). As correlações positivas foram: Us (0,26), silte (0,22), K (0,17) e H-Al (0,17), mostrando que as variáveis físicas do solo foram as que mais interferiram nos processos de deslocamento do CO₂ no solo.

Quando foi analisado o desempenho dos modelos, por meio R²_{ajust}, na fase de calibração foi observado que TREE apresentou a maior performance (1,0000), seguido da RFT (0,9395), da RFT_GS (0,9385), da SVM (0,9127), do BOO (0,8387), da MLP (0,6938), da SVM_kpoli (0,5785), da KNN (0,4701), da RLM (0,2496), da SVM_VC (0,1752) e da SGD (0,1738). Já na fase validação o algoritmo RFT_GS apresentou o melhor desempenho dentre os estudados (0,5950), seguido do RFT (0,5832), da TREE (0,5539), do MLP (0,4995), do BOO (0,4956), da KNN (0,3560), do SVM (0,3192), do SVM_kpoli (0,2970), da RLM (0,2283), da SVM_VC (0,1666) e da SGD (0,0955), resultando na conclusão que o modelo que mais conseguiu prever a FCO₂, em função dos atributos do sistema solo-planta-atmosfera foi o RFT_GS.

Dessa forma, conclui-se que estudos como este são importantes para os agricultores, poder público e privado, pois as informações geradas podem ser utilizadas no gerenciamento e nas tomadas de decisões do processo produtivo e financeiro, sendo uma alternativa para a caracterização não só da FCO₂, mas também de outros atributos do sistema solo-planta-atmosfera, promovendo a redução de custos, o uso eficiente dos recursos infinitos e finitos, rentabilidade, produtividade e maior sustentabilidade para o sistema produtivo como um todo.

CAPÍTULO 4 – CONSIDERAÇÕES FINAIS

A emissão da CO₂ do solo é um processo bastante complexo e dependente do sistema solo-planta-atmosfera. Esses tipos de estudos colaboram com o entendimento e mitigação da emissão não só do CO₂ do solo, mas também de todos os gases causadores do efeito estufa.

Analizando as pesquisas, existem relativamente poucos trabalhos que predizem a emissão do CO₂ do solo, relacionando os atributos constituintes do sistema solo-planta-atmosfera. Além disso, outras variáveis importantes outras variáveis podem ser preditas, indicando que muitas implicações precisam ser estudadas e que existem grandes oportunidades e problemas para serem resolvidos.

A mineração de dados e as escolhas dos parâmetros para os algoritmos proporcionam modelos mais acurados, e o GridSearchCV foi uma ferramenta bastante eficiente na seleção automatizada dos parâmetros dos algoritmos. Além disso, devido ao aumento da capacidade de processamento dos computadores, outros algoritmos precisam ser testados com o objetivo de se obter modelos mais acurados, disponibilizando informações que possam auxiliar na maior produtividade, rentabilidade e sustentabilidade do processo produtivo.

Dessa forma, fica evidenciado que estes estudos são muito importantes e que essas informações podem ser consumidas pelos agricultores, setor público e privado, pois as informações geradas podem ser utilizadas no gerenciamento e nas tomadas de decisões do processo produtivo e financeiro, sendo uma alternativa para a caracterização não só da FCO₂, mas também de outros atributos do sistema solo-planta-atmosfera, impactando diretamente na qualidade de vida, na produtividade e na sustentabilidade do planeta.

AGRADECIMENTOS

Meus agradecimentos à Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) e a Universidade Estadual Paulista (UNESP) que colaboram com a minha formação e suporte nas atividades realizadas.

REFERÊNCIAS BIBLIOGRÁFICAS

Abbasi NA, Hamrani, A, Madramootoo, CA, Zhang, T, Tan, CS, Goyal, MK (2021) Modelling carbon dioxide emissions under a maize-soy rotation using machine learning. **Biosystems Engineering**, 212:1-18.

ABC (2015) **Agricultura de Baixo Carbono**. Disponível em <<http://www.agricultura.sp.gov.br/media/13376-plano-abc-publicacao.pdf>>. Acesso em: 26 set. 2018.

Adjuik TA e Davis SC (2022) Machine Learning Approach to Simulate Soil CO₂ Fluxes under Cropping Systems. **Agronomy**, 12: p. 197.

Ağbulut Ü (2022) Forecasting of transportation-related energy demand and CO₂ emissions in Turkey with different machine learning algorithms. **Sustainable Production and Consumption**, 29:141-157.

Ağbulut Ü, Gürel AE, Biçen Y (2021) Prediction of daily global solar radiation using different machine learning algorithms: Evaluation and comparison. **Renewable and Sustainable Energy Reviews**, 135: p. 110114.

Akaike H (1974) A new look at the statistical model identification. **IEEE Transactions on automatic control**, 19:716-723.

Al-Sudani HIZ (2020) Calculation of meteorological water balance in Iraq. **Resources Environment and Information Engineering**, 2:84-89.

Allen RG, Pereira LS, Raes D, Smith M (1998) Crop evapotranspiration-Guidelines for computing crop water requirements-FAO Irrigation and drainage paper 56. **Fao, Rome**, 300:D05109.

Almeida R, Teixeira DB, Montanari R, Bolonhezi AC, Teixeira EB, Moitinho MR, Panosso AR, Spokas KA, La Scala Júnior N (2018) Ratio of CO₂ and O₂ as index for categorising soil biological activity in sugarcane areas under contrasting straw management regimes. **Soil Research**, 56:373-381.

Alpaydin E (2020) Introduction to machine learning. **MIT press**.

Anis A, Gadde A, Ortega A (2014) Towards a sampling theorem for signals on arbitrary graphs. **IEEE International Conference on Signal Processing**, 3864-3868.

Awe GO, Reicherta JM, Wendrothb OO (2015) Temporal variability and covariance structures of soil temperature in a sugarcane field under different management practices in southern Brazil. Amsterdam, **Soil and Tillage Research**, 150:93–106.

Bakay MS e Ağbulut Ü (2021) Electricity production based forecasting of greenhouse gas emissions in Turkey with deep learning, support vector machine and artificial neural network algorithms. **Journal of Cleaner Production**, 285: p. 125324.

Ball BC e Smith KA (1991) **Gas movement**. In: Smith KA e MULLINS CE (Ed.). **Solis analysis: physical methods**, New York: Marcel Dekker. p. 511-549.

Barbosa JDS, Silva K DCR, Carducci CE, Santos KLD, Kohn LS, Fucks JS (2017) Atributos físico-hídricos de um Cambissolo Húmico sob sistema agroflorestal no planalto catarinense. **Floresta e Ambiente**, 24.

Başarslan MS e Argun İD (2019) **Prediction of potential bank customers: application on data mining**. In **The International Conference on Artificial Intelligence and Applied Mathematics in Engineering (pp. 96-106)**. Springer, Cham.

Bayer C, Mielniczuk J, Amado TJC, Martin-Neto L, Fernandes SV (2000) Organic matter storage in a Sandy clay loam Acrisol affected by tillage and cropping systems in southern Brazil. **Soil and Tillage Research**, 54:101-109.

Bicalho ES, Panosso AR, Teixeira DDB, Miranda JGV, Pereira GT, La Scala N (2014) Spatial variability structure of soil CO₂ emission and soil attributes in a sugarcane area. **Agriculture, Ecosystems & Environment**, 189:206-215.

Bicalho SDE, Moitinho MR, Bortoli DT, Panosso AR, Spokas KA, La Scala NJ (2017) Soil Greenhouse Gases: Relations to Soil Attributes in a Sugarcane Production Area. **Soil Science Society of America Journal**, 81(5), 1168-1178.

Bollobás, B. **Random Graphs**, 2nd Edition. Cambridge University Press, 2001.

Boser BE, Guyon IM, Vapnik VN (1992) **A training algorithm for optimal margin classifiers**. In **Proceedings of the fifth annual workshop on Computational learning theory (pp. 144-152)**.

Breiman L, Friedman JH, Olshen RA, Stone CJ (1984) **Classification and regression trees**. **Wadsworth & Brooks**. Cole Statistics/Probability Series.

Brito LF, Marques Júnior J, Pereira GT, Souza ZM (2009) Soil CO₂ emission of sugarcane fields as affected by topography. **Scientia Agricola**, 66:77-83.

Bruce JP, Frome, Haites E, Janzen H, Lal R (1999) Carbon sequestration in soils. Ankeny, **Journal of Soil and Water Conservation**, 54:382-389.

Bruce P, Bruce A (2019) **Estatística prática para cientistas de dados: 50 conceitos essenciais. 1ª edição**. Rio de Janeiro — RJ: Alta Books, 2019.

Cai J, Xu K, Zhu Y, Hu F, Li L (2020) Prediction and analysis of net ecosystem carbon exchange based on gradient boosting regression and random forest. **Applied energy**, 15; 262:114566.

Carvalho JLN (2010) Greenhouse gas mitigations in Brazil for land-use change, livestock and agriculture. **Scientia Agricola**, 67.

Carvalho JLN, Nogueirol RC, Menandro LMS, Bordonal RO, Borges CD, Cantarella H, Franco HCJ (2016) Agronomic and environmental implications of sugarcane straw removal: a major review. **Global Change Biology Bioenergy**, 9:1-15.

Castiglioni GL, Silva FA, Caliar M, Junior MSS (2013) Modelagem matemática do processo de secagem da massa fibrosa de mandioca. **Revista Brasileira de Engenharia Agrícola e Ambiental**, 17:987-994.

Cerri CC, Bernoux M, Cerri CEP, Lal R. Challenges and opportunities of soil carbon sequestration in Latin America. In: Lal R, Cerri CC, Bernoux M, Etchevers JD, Cerri CEP (Ed.) (2006) Carbon sequestration in soils of Latin America. New York: **The Haworth Press**, 1:41-47.

Cerri CC, Maia SMF, Galdos MV, Cerri CEP, Feigl BJ, Bernoux M (2009) Brazilian greenhouse gas emissions: the importance of agriculture and livestock. Piracicaba, **Scientia Agrícola**, 66:831-843.

Cerri CEP, Galdos MV, Carvalho JLN, Feigl BJ, Cerri CC (2013) Quantifying soil carbon stocks and greenhouse gas fluxes in the sugarcane agrosystem: point of view. **Scientia Agrícola**, 70:361-368.

Chen B, Harp DR, Lin Y, Keating EH, Pawar RJ (2018) Geologic CO₂ sequestration monitoring design: A machine learning and uncertainty quantification based approach. **Applied energy**, 225:332-345.

Chen JL, Li GS, Wu SJ (2013) Assessing the potential of support vector machine for estimating daily solar radiation using sunshine duration. **Energy conversion and management**, 75:311-318.

Chen T, Xu Z, Tang G, Chen X, Fang, H., Guo H, Yuan Y, Zheng G, Jiang L, Niu X (2021) Spatiotemporal Monitoring of Soil CO₂ Efflux in a Subtropical Forest during the Dry Season Based on Field Observations and Remote Sensing Imagery. **Remote Sensing**, 13:p.3481.

Cheshmberah F, Fathizad H, Parad GA, Shojaeifar S (2020) Comparison of RBF and MLP neural network performance and regression analysis to estimate carbon sequestration. **International Journal of Environmental Science and Technology**, 17:3891-3900.

CIIAGRO, IAC. Centro Integrado de Informações Agrometeorologias. **AGROCLIMATOLOGIA DO ESTADO DE SÃO PAULO OUTUBRO A NOVEMBRO, 2020.**

CONAB, 2020. **Acompanhamento da safra brasileira: cana-de-açúcar, segundo levantamento, maio/2020.** In: Brasília. Conab. Companhia Nacional de Abastecimento. Disponível em <https://www.conab.gov.br/info-agro/safras/cana>, [Safra 2019/2020](#). Acesso em 13 out. 2020.

CONAB, 2022. **BOLETIM DA SAFRA DE CANA-DE-AÇÚCAR: 3º Levantamento - Safra 2021/22, Boletim de cana-de-açúcar - Novembro de 2021, segundo a Companhia Nacional de Abastecimento, Brasília.** Disponível em < <https://www.conab.gov.br/info-agro/safras/cana/boletim-da-safra-de-cana-de-acucar?limitstart=0>>. Acesso em: 28 fev. 2022.

Costa DCC, Souza Bispo J, Oliveira GM, Angelotti F, Dantas BF (2017) Emergência e crescimento inicial de *Amburana cearensis* sob aumento da temperatura e dióxido de carbono. **Journal of Environmental Analysis and Progress**, 2:212-219.

Cuadra SV, Costa MH, Kucharik CJ, DA Rocha HR, Tatsch JD, Inman-Bamber G, Da Rocha RP, Leite CC, Cabral OMR (2012) A biophysical model of Sugarcane growth. **Global Change Biology Bioenergy**, 4:36-48.

Da Silva PA, De Lima BH, La Scala Jr N, Peruzzi NJ, Chavarette FR, Panosso AR (2020) Spatial variation of soil carbon stability in sugarcane crops, central-south of Brazil. **Soil and Tillage Research**, 202, p. 104667.

De Figueiredo EB, Panosso AR, Reicosky DC, La Scala, N Jr, (2015) Short-term CO₂-C emissions from soil prior to sugarcane (*Saccharum spp.*) replanting in southern Brazil. **Global Change Biology Bioenergy**, 7:316-327.

De Oliveira FS, Ferreira M, Sentelhas PC (2020) Favorabilidade para a aplicação de herbicidas na cultura da cana-de-açúcar com base em critérios agrometeorológicos. **Agrometeoros**, v. 28.

De Souza, LC, Fernandes C, Moitinho M R, da Silva Bicalho E, La Scala N (2019) Soil carbon dioxide emission associated with soil porosity after sugarcane field reform. **Mitigation and adaptation strategies for global change**, 24:113-127.

Deiss L, Franzluebbers AJ, De Moraes A (2017) Soil texture and organic carbon fractions predicted from near-infrared spectroscopy and geostatistics. **Soil Science Society of America Journal**, 81:1222-1234.

Denman KL, Brasseur G, Chidthaisong A, Ciais P, Cox PM, Dickinson RE, Hauglustaine D, Heinze C, Holland E, Jacob D, Lohmann U, Ramachandran S, Silva Dias PL da, Wofsy SC, Zhang X (2007) Couplings between changes in the climate system and biogeochemistry. In: Solomon S, Qin D, Manning M, Chen Z, Marquis M, Averyt KB, Tignor M, Miller HL (Ed.). *Climate change 2007: the physical science basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. **Cambridge**, New York: Cambridge University Press, 499-588.

Dossou-Yovo ER, Brüggemann N, Jesse N, Huat J, Ago EE, Agbossou EK (2016) Reducing soil CO₂ emission and improving upland rice yield with no-tillage, straw mulch and nitrogen fertilization in northern Benin. **Soil and Tillage Research**, 156:44-53.

Duiker SW e Lal R (2000) Carbon budge study using CO₂ flux measurements from a no till system in Central Ohio. Amsterdam, **Soil and Tillage Research**, 54:21-30.

Ehteram M, Sammen SS, Panahi F, Sidek LM (2021) A hybrid novel SVM model for predicting CO₂ emissions using Multiobjective Seagull Optimization. **Environmental Science and Pollution Research**, 28:66171-66192.

EMBRAPA (1997) **Manual de métodos de análise de solo**. 2 ed. Brasília: Ministério da Agricultura e do Abastecimento / EMBRAPA-CNPQ, p. 212.

Esrl (2016) Recent monthly average Mauna Loa CO₂. **Earth system research**. Disponível em <<https://www.esrl.noaa.gov/gmd/ccgg/trends/>>. Acesso em: 26 set. 2018.

Eswaran H, Berg EVD, Reich P (1993) Organic carbon in soils of the world. **Soil Science Society of America Journal**, 57:192-194.

Everingham Y, Sexton J, Robson A (2015) A statistical approach for identifying important climatic influences on sugarcane yields. In Proceedings of the 37th Conference of the Australian Society of Sugar Cane Technologists, 28-30 April 2015, Bundaberg, Queensland, Australia. **Australian Society of Sugar Cane Technologists**.

FAO (2017) **The Future of Food and Agriculture: Trends and Challenges**. Annual Report, v. 296.

Farhate CVV, Souza ZMD, Oliveira SRDM, Tavares RLM, Carvalho JLN (2018) Use of data mining techniques to classify soil CO₂ emission induced by crop management in sugarcane field. **Plos one**, 13:p.e0193537.

Figueiredo CC, Resck DVS, Carneiro MAC (2010) Labile and stable fractions of soil organic matter under management systems and native Cerrado. Viçosa-MG, **Revista Brasileira de Ciência do Solo**, 34:907-916.

Figueiredo EB, Panosso AR, Reicosky DC, La Scala N (2014) Short-term CO₂-C emissions from soil prior to sugarcane (*Saccharum spp.*) replanting in southern Brazil. Chichester, **Global Change Biology**, 7:316-327.

Forster P, Ramaswamy V, Artaxo P, Berntsen T, Betts R, Fahey DW, Haywood J, Lean J, Lowe DC, Myhre G, Nganga J, Prinn R, Raga G, Schulz M, Van Dorland R (2007) Changes in atmospheric constituents and in radiative forcing. In: Solomon S, Qin D, Manning M, Chen Z, Marquis M, Averyt KB, Tignor M, Miller HL (Ed.). Climate change 2007: the physical science basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. New York: Cambridge University Press **Cambridge**, p. 129-234.

Friedman JH (2001) Greedy function approximation: a gradient boosting machine. **Annals of statistics**, 1:1189-232.

Friedman JH (2002) Stochastic gradient boosting. **Computational statistics & data analysis**, 38:367-78.

Galdos MV, Cerri CC, Cerri CEP (2009) Soil carbon stocks under burned and unburned sugarcane in Brazil. Amsterdam, **Geoderma**, 153:347-352.

Galvão P, Hirata R, Conicelli B (2018) Estimating groundwater recharge using GIS-based distributed water balance model in an environmental protection area in the city of Sete Lagoas (MG), Brazil. **Environmental Earth Sciences**, 77:1-19.

Gerardi F, Neves M, Kalaki RB, Gali R (2017) **O Setor Sucroenergético em 2030-Dimensões, Investimentos e uma Agenda Estratégica**. Brasília: CNI-Confederação Nacional das Indústrias, v. 1.

Gupta G, Setia R, Meena A, Jaint B (2020) **Environment monitoring system for agricultural application using IoT and predicting crop yield using various data mining techniques**. In **2020 5th International Conference on Communication and Electronics Systems (ICCES)**, pp. 1019-1025.

Gürel AE, Ağbulut Ü, Biçen Y (2020) Assessment of machine learning, time series, response surface methodology and empirical models in prediction of global solar radiation. **Journal of Cleaner Production**, 277:p.122353.

Hamrani A, Akbarzadeh A, Madramootoo CA (2020) Machine learning for predicting greenhouse gas emissions from agricultural soils. **Science of The Total Environment**, 741:p. 140338.

Hamzacebi C e Karakurt I (2015) **Forecasting the energy-related CO₂ emissions of Turkey using a grey prediction model**. **Energy Sources, Part A: Recovery, Utilization, and Environmental Effects**, 37:1023-1031.

Han J, Pei J, Kamber M (2011) Data mining concepts and techniques. **Elsevier**.

He J, Dukes MD, Hochmuth GJ, Jones JW, Graham WD (2012) Identifying irrigation and nitrogen best management practices for sweet corn production on sandy soils using CERES-Maize model. **Agricultural Water Management**, 109:61-70.

Hira A, & De Oliveira LG (2009) No substitute for oil? How Brazil developed its ethanol industry. **Energy policy**, 37:2450-2456.

Houshmandfar A, O'Leary G, Fitzgerald GJ, Chen Y, Tausz-Posch S, Benke K, Uddin S, Tausz M (2021) Machine learning produces higher prediction accuracy than the Jarvis-type model of climatic control on stomatal conductance in a dryland wheat agro-ecosystem. **Agricultural and Forest Meteorology**, 304:p.108423.

Iamaguti JL, Moitinho MR, Teixeira DD, Bicalho EDS, Panosso AR, La Scala N (2015) Preparo do solo e emissão de CO₂, temperatura e umidade do solo em área

canavieira. Campina Grande, **Revista Brasileira de Engenharia Agrícola e Ambiental**, 19:497-504.

Inman–Bamber NG (2000) **History of the Canegro Model**. Proceedings of International Workshop, n. 4-7 August, p. 5-8.

Intergovernmental Panel on Climate Change – IPCC. Climate Change (2007) Mitigation. **Contribution of Working Group III. Fourth Assessment Report of the Intergovernmental Panel on Climate Change**. Cambridge University Press, United Kingdom and New York, NY, USA, Cambridge.

Intergovernmental Panel on Climate Change – IPCC. Climate Change (2014) Impacts, Adaptation, and Vulnerability. **Working Group II contribution to AR5**. Disponível em: < <http://www.ipcc.ch/report/ar5/wg2/>. Acesso em 13 abr. 2021.

Intergovernmental Panel on Climate Change – IPCC. Climate Change (2014) **Mitigation. Contribution of Working Group III. Fourth Assessment Report of the Intergovernmental Panel on Climate Change**.

Karstoft KI, Statnikov A, Andersen SB, Madsen T, Galatzer-Levy IR (2015) Early identification of posttraumatic stress following military deployment: Application of machine learning methods to a prospective study of Danish soldiers. **Journal of affective disorders**, 184:170-175.

Keating BA, Robertson MJ, Muchow RC, Huth NI (1999) Modelling sugarcane production systems I. Development and performance of the sugarcane module. **Field Crops Research**, 61:253-271.

Kheradmand S, Heidarzadeh N, Kia SH (2022) Prediction of the CO₂ emission across grassland and cropland using tower-based eddy covariance flux measurements: a machine learning approach. **Environment, Development and Sustainability**, 1-15.

Kirschbaum MUF (2000) Will changes in soil organic carbon act as a positive or negative feedback on global warming? Dordrecht, **Biogeochemistry**, 48:21-51.

Kluyver T, Ragan-Kelley B, Pérez F, Granger BE; Bussonnier M, Frederico J, Kelley K, Hamrick JB, Grout J, Corlay S (2016) **Jupyter Notebooks-um Formato de Publicação para Fluxos de Trabalho Computacionais Reprodutíveis**; IOS Press: Amsterdã, Holanda.

Kolasa-Więcek A (2018) Neural modeling of greenhouse gas emission from agricultural sector in European Union member countries. **Water, Air, & Soil Pollution**, 229:1-12.

La Scala Jr N, Mendonça ES, Souza JJ, Panosso AR, Simas FNB, Schaefer CEGR (2010) Spatial and temporal variability in soil CO₂-C emissions and relation to soil temperature at King George Island, Maritime Antarctica. **National Institute of Polar Research**, 4:479-487.

La Scala N, Bolonhezi D, Pereira GT (2006) Short-term soil CO₂ emission after conventional and reduced tillage of a no-till sugar cane area in southern Brazil. Amsterdam, **Soil and Tillage Research**, 91:244-248.

La Scala N, Lopes A, Spokas K, Bolonhezi D, Archer DW, Reicosky DC (2008) Short-term temporal changes of soil carbon losses after tillage described by a first-order decay model. Amsterdam, **Soil and Tillage Research**, 99:108-118.

Lal R (2004) Soil carbon sequestration impacts on global climate change and food security. Washington, **Science**, 304:1523-1627.

Lal R (2009) Challenges and opportunities in soil organic matter research. **European Journal of Soil Science**, 60:158-169.

Lamb A, Green R, Bateman I, Broadmeadow M, Bruce T, Burney J, Goulding K (2016) The potential for land sparing to offset greenhouse gas emissions from agriculture. London, **Nature Climate Change**.

Lemordant L, Gentine P, Swann AS, Cook BI, Scheff J (2018) Critical impact of vegetation physiology on the continental hydrologic cycle in response to increasing CO₂. **Proceedings of the National Academy of Sciences**, 115:4093-4098.

Lenka SK, Mohapatra AG (2015) **Gradient descent with momentum based neural network pattern classification for the prediction of soil moisture content in precision agriculture**. In 2015 IEEE International Symposium on Nanoelectronic and Information Systems, 63-66.

Li T, Ji Y, Zhang M, Li M (2017) Determining optimal CO₂ concentration of greenhouse tomato based on PSO-SVM. **Applied Engineering in Agriculture**, 33:157-166.

Li W, Ou Q, Chen Y (2014) Decomposition of China's CO₂ emissions from agriculture utilizing an improved Kaya identity. **Environmental Science and Pollution Research**, 21:13000–13006.

Li X, Ren A, Li Q (2022) Exploring Patterns of Transportation-Related CO₂ Emissions Using Machine Learning Methods. **Sustainability**, 14:p.4588.

Linn DM e Doran JW (1984) Effect of water-filled pore space on carbon dioxide and nitrous oxide production in tilled and non-tilled soils. Madison, **Soil Science Society of American Journal**, 48:1267-1272.

Lu H, Li S, Ma M, Bastrikov V, Chen X, Ciais P, Dai Y, Ito A, Ju W, Lienert S, Lombardozzi D, Lu X, Maignan F, Nakhavali M, Quine T, Schindlbacher A, Wang J, Wang Y, Wårlind D, Zhang S, Yuan W (2021) Comparing machine learning-derived global estimates of soil respiration and its components with those from terrestrial ecosystem models. **Environmental Research Letters**, 16:p.054048.

Ma S and Feng H (2013) Will the decline of efficiency in China's agriculture come to an end? An analysis based on opening and convergence. **China Economic Review**, 27:179-190.

Ma X, Mei X, Wu W, Wu X, Zeng B (2019) A novel fractional time delayed grey model with Grey Wolf Optimizer and its applications in forecasting the natural gas and coal consumption in Chongqing China. **Energy**, 178:487-507.

Machado Ploa (2005) Carbono do solo e a mitigação da mudança climática global. São Paulo, **Química Nova**, 28:329-334.

Mammoliti E, Fronzi D, Mancini A, Valigi D, Tazioli A (2021) WaterbalANce, a WebApp for Thornthwaite–Mather Water Balance Computation: Comparison of Applications in Two European Watersheds. **Hydrology**, v. 8, p. 34.

Marcelo AV, Cora JE, La Scala N (2012) Influence of liming on residual soil respiration and chemical properties in a tropical no-tillage system. Viçosa, **Revista Brasileira de Ciência do Solo**, 36:45-50.

Martinsen KT, Kragh T, Sand-Jensen K (2020) Carbon dioxide partial pressure and emission throughout the Scandinavian stream network. **Global Biogeochemical Cycles**, 34:p.e2020GB006703.

McBratney AB, De Grujter JJ (1992) A continuum approach to soil classification by modified fuzzy k-means with extragrades. **Journal of Soil Science**, 43:159-175.

McKinney W (2012) **Python for data analysis: Data wrangling with Pandas, NumPy, and IPython**. " O'Reilly Media, Inc."

MCT (2004) **Comunicação Nacional Inicial do Brasil à Convenção-Quadro das Nações Unidas sobre Mudança do Clima. Coordenação Geral de Mudanças Globais de Clima**, Ministério da Ciência e Tecnologia (MCT), Brasília, DF, Brasil, p. 271.

Melillo JM, Steudler PA, Aber JD, Newkirk K, Lux H, Bowles FP, Catricala C, Magill A, Ahrens T, Morrisseau S (2002) Soil warming and carbon-cycle feedbacks to the climate system. Washington, **Science**, 298:2173-2176.

Mello FFC, Cerri CEP, Davies CA, Holbrook NM, Paustian K, Maia SMF, Galdos MV, Bernoux M, Cerri CC (2014) Payback time for soil carbon and sugar-cane ethanol. **Nature Climate Change**, 4:605-609.

Milanez AY, Nyko D, Garcia JLF, Reis BLSFSD (2012) O déficit de produção de etanol no Brasil entre 2012 e 2015: determinantes, consequências e sugestões de política. **BNDES Setorial**, n. 35, mar. 2012, 277-302.

Mitchell T (1997) **Machine Learning. McGraw-Hill.** (McGraw-Hill International Editions). ISBN 9780071154673. Available from Internet: <<https://books.google.com.br/books?id=EoYBngEACAAJ>>.

Mitchell TM (2006) **The discipline of machine learning (Vol. 9). Pittsburgh: Carnegie Mellon University, School of Computer Science, Machine Learning Department.**

Moitinho MR, Ferraudo AS, Panosso AR, da Silva Bicalho E, Teixeira DDB, de Andrade Barbosa M, Tsai SM, Borges BMF, Cannavan FS, De Souza JAM, La Scala Jr N (2021) Effects of burned and unburned sugarcane harvesting systems on soil CO₂ emission and soil physical, chemical, and microbiological attributes. **CATENA**, 196, p. 104903.

Moitinho MR, Padovan MP, Panosso AR, La Scala N (2013) Efeito do preparo do solo e resíduo da colheita de cana-de-açúcar sobre a emissão de CO₂. Viçosa-MG, **Revista Brasileira de Ciência do Solo**, 37:1720-1728.

Moitinho MR, Padovan MP, Panosso AR, Teixeira DB, Ferraudo AS, La Scala N (2015) On the spatial and temporal dependence of CO₂ emission on soil properties in sugarcane (*Saccharum spp.*) production. Amsterdam. **Soil and Tillage Research**, 148:127-132.

Moitinho MR, Teixeira DDB, da Silva Bicalho E, Panosso AR, Ferraudo AS, Pereira GT, Tsai SM, Borges BMF, La Scala Jr N (2021) Soil CO₂ emission and soil attributes associated with the microbiota of a sugarcane area in southern Brazil. **Scientific Reports**, 11:1-15.

Morita M, Mori T, Yokoo A, Ohkura T, Morita Y (2019) Continuous monitoring of soil CO₂ flux at Aso volcano, Japan: the influence of environmental parameters on diffuse degassing. **Earth, Planets and Space**, 71:1-16.

Nguemezi C, Tematio P, Silatsa FB, Yemefack M (2021) Spatial variation and temporal decline (1985–2017) of soil organic carbon stocks (SOCS) in relation to land use types in Tombel area, South-West Cameroon. **Soil and Tillage Research**, 213:p.105114.

OMM. **Organização Mundial da Meteorologia: segundo o El aumento de la concentración de gases de efecto invernadero alcanza un nuevo récord.** Disponível: <<https://public.wmo.int/es/media/comunicados-de-prensa/el-aumento-de-la-concentraci%C3%B3n-de-gases-de-efecto-invernadero-alcanza-un>>. Acesso em: 12 jan. 2018.

Pal SK, Mitra S (1992). **Multilayer perceptron, fuzzy sets, classification.**

Pan B, Lam SK, Wang E, Mosier A, Chen D (2021) New approach for predicting nitrification and its fraction of N₂O emissions in global terrestrial ecosystems. **Environmental Research Letters**, 16:p.034053.

Panosso AR, Marques Júnior J, Milori DMBP, Ferraudo AS, Barbieri DM, Pereira GT, La Scala N (2011) Soil CO₂ emission and its relation to soil properties in sugarcane areas under Slash-and-burn and Green harvest. Amsterdam, **Soil Tillage Research**, 111:190-196.

Panosso AR, Pereira GT, Marques Júnior J, La Scala Júnior N (2008) Variabilidade espacial da emissão de CO₂ em Latossolos sob cultivo de cana-de-açúcar em diferentes sistemas de manejo. **Engenharia Agrícola**, 28:227-236.

Panosso AR, Perillo LI, Ferraudo AS, Pereira GT, Miranda J GV, La Scala JR N (2012) Fractal dimension and anisotropy of soil CO₂ emission in a mechanically harvested sugarcane production area. **Soil & Tillage Research**, 124: p. 8-16.

Panosso AR, Ribeiro CER, Zanini JR, Pavani LC, Pereira GT, La Scala Júnior N (2009) Variabilidade espacial da emissão de CO₂, da temperatura e umidade de um Latossolo desprovido de vegetação sob diferentes lâminas de molhamento. **Semina: Ciências Agrárias**, 30:1017-1034.

Patil NG and Singh SK (2016) Pedotransfer functions for estimating soil hydraulic properties: a review. Amsterdam, **Pedosphere**, 26:417-430.

Payen FT, Sykes A Aitkenhead M, Alexander, P., Moran, D., & MacLeod, M. (2021). Predicting the abatement rates of soil organic carbon sequestration management in Western European vineyards using random forest regression. **Cleaner Environmental Systems**, 2:p. 100024.

PBMC (2013) Contribuição do Grupo de Trabalho 1 ao **Primeiro Relatório de Avaliação Nacional do Painel Brasileiro de Mudanças Climáticas** – Sumário Executivo GT1. PBMC.

Pereira AR, Villa Nova NA, Sedyama GC (1997) **Evapo(transpi)ração**. Piracicaba, FEALQ, 183 p.

Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V. 2011. Scikit-learn: machine learning in Python. J Mach Learn Res. 12: 2825-2830. others.

Perin V, Sentelhas PC, Dias HB, Santos EA (2019) Sugarcane irrigation potential in Northwestern São Paulo, Brazil, by integrating Agrometeorological and GIS tools. **Agricultural Water Management**, 220:50-58.

Pham TD, Yokoya, N, Nguyen TTT, Le NN, Ha, N.T, Xia J, Takeuchi W.; Pham, T.D. (2021) Improvement of mangrove soil carbon stocks estimation in North Vietnam using Sentinel-2 data and machine learning approach. **GIScience & Remote Sensing**, 58:68-87.

Piepho HP (2019) A coefficient of determination (R^2) for generalized linear mixed models. **Biometrical Journal**, 61: 860-872.

Pivetta M (2013) Extremos do clima. São Paulo, **Revista Pesquisa FAPESP (online)**, 16p.

Porter Jr, Xie L, Challinor AJ, Cochrane K, Howden SM, Iqbal MM, Lobell DB, Travasso MI (2014) Food security and food production systems. In: Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part A: Global and Sectoral Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Field CB, Barros VR, Dokken DJ, Mach KJ, Mastrandrea MD, Bilir TE, Chatterjee M, Ebi KL, Estrada YO, Genova RC, Girma B, Kissel ES, Levy AN, MacCracken S, Mastrandrea PR, White LL (Eds.)]. United Kingdom and New York, Cambridge University Press, **Cambridge**, 485-533.

Pritzker Z, Feuer A (1991) Variable length stochastic gradient algorithm. **IEEE Transactions on signal processing**, 39: 997-1001.

Quinlan JR (1986). **Induction of decision trees. Mach Learn 1, 81–106.** Available at <https://doi.org/10.1007/BF00116251>.

R DEVELOPMENT CORE TEAM. **A language and environment for statistical computing. R Foundation for Statistical Computing**, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org/>.

Raij BV, Andrade JC, Cantarella H, Quaggio JA (2001) **Chemical Analysis to Evaluate the Fertility of Tropical Soils = Análise Química para Avaliação da Fertilidade de Solos Tropicais**. Campinas, SP, Instituto Agrônômico, Brazil (in Portuguese). 285 p.

Ransom CJ, Kitchen NR, Camberato JJ, Carter PR, Ferguson RB, Fernández FG, Franzen DW, Laboski CAM, Myers DB, Nafziger ED, Sawyer JE, Shanahan JF (2019) Statistical and machine learning methods evaluated for incorporating soil and weather into corn nitrogen recommendations. **Computers and Electronics in Agriculture**, 164, p.104872.

Rolim GS, de Oliveira Aparecido LE, de Souza PS, Augusto R, Lamparelli C, dos Santos ÉR (2020) Climate and natural quality of *Coffea arabica* L. drink. **Theoretical and Applied Climatology**, 141:87–98.

Rubia G e Nandhini M (2021). Predictive Soil Analytics Using Data Mining Techniques. INFOCOMP. **Journal of Computer Science**, 20(1).

Saha D, Basso B, & Robertson GP (2021) Machine learning improves predictions of agricultural nitrous oxide (N₂O) emissions from intensively managed cropping systems. **Environmental Research Letters**, 16:p.024004.

Salam S e Verma TN (2020) Analysis of significance of variables in IC engine operation: an empirical methodology. **Energy Conversion and Management**, 207:p.112520.

Salviano MF, Groppo JD, Pelegrino GQ (2016) Análise de tendências em dados de precipitação e temperatura no Brasil. **Revista Brasileira de Meteorologia**, 31:64-73.

Sanches MC, Zullo Junior J, Romani LAS (2018) Comparação do risco climático da soja, cana-de-açúcar e café arábica, para o estado de São Paulo, calculado com dados terrestres e orbitais de precipitação pluvial. **Embrapa Informática Agropecuária-Artigo em periódico indexado (ALICE)**.

Santana AC **Métodos quantitativos em economia: elementos e aplicações**. Belém: UFRA, 2003.

Santos GO, Hernandez FBT, Rossetti JC (2013) Balanço hídrico como ferramenta ao planejamento agropecuário para a região de Marinópolis, noroeste do estado de São Paulo. **Revista Brasileira de Agricultura Irrigada-RBAI**, v. 4.

Santos HG, Jacomine PKT, Anjos LHC, Oliveira VA, Lumberras JF, Coelho MR, Almeida JA, Cunha TJF, Oliveira JB (2013) **Brazilian System of Soil Classification. = Sistema Brasileiro de Classificação de Solos (3 ed.)**. Embrapa, Brasília, DF, Brazil (in Portuguese).

Schlömer S, Furche M, Dumke I, Poggenburg J, Bahr C, Seeger S, Vidal A, Faber E (2013) A review of continuous soil gas monitoring related to CCS – technical advances and lessons learned. London, **Applied Geochemistry**, 30:148-160.

Schwarz G (1978) Estimating the dimension of a model. **The annals of statistics**, 461-464.

Schwendenmann L, Veldkamp E, Brenes T, O'brien JJ, Mackensen J (2003) Spatial and temporal variation in soil CO₂ efflux in an old-growth neotropical rain forest, La Selva, Costa Rica. Dordrecht, **Biogeochemistry**, 64:111-28.

SEEG (2016) **Sistema de Estimativas de Emissões e Remoções de Gases de Efeito Estufa**. Disponível em < <http://seeg.eco.br/o-que-e-o-seeg/>>. Acesso em: 26 set. 2018.

Şentürk ZK (2020) Artificial Neural Networks based decision support system for the detection of diabetic retinopathy. **Sakarya University Journal of Science**, 24:424-431.

Siabi Z, Falahatkar S, Alavi SJ (2019). Spatial distribution of XCO₂ using OCO⁻² data in growing seasons. **Journal of environmental management**, 244:110-118.

Silva EF, Moitinho MR, Teixeira DDB, Pereira GT, La Scala N (2013) Soil CO₂ emission associated with liming in area of conversion of orange for sugarcane in Jaboticabal, **Engenharia Agrícola**, 34: 885-898.

Silva EFD, Moitinho MR, Teixeira DDB, Bicalho EDS, Castioni GAF, Pereira GT, La Scala N (2020) Short-term spatiotemporal variation of soil CO₂ emission, temperature,

moisture and aeration in sugarcane field reform areas under the influence of precipitation events. **Soil Use and Management**, 36:658-670.

Silva IR e Mendonça ES (1997) **Matéria orgânica do solo. Biologia dos solos dos cerrados**. Planaltina: Embrapa-CPAC.

Silva-Olaya AM, Cerri CEP, La Scala N, Dias CTS, Cerri CC (2013) Carbon dioxide emissions under different soil tillage systems in mechanically harvested sugarcane, **Environmental Research Letters**, v. 8, p. 015014.

Singels A and Bezuidenhout CN (2002) A new method of simulating dry matter partitioning in the Canegro sugarcane model, **Field Crops Research**, 78:151-164.

Smart DR e Peñuelas J (2005) Short-term CO₂ emissions from planted soil subject to elevated CO₂ and simulated precipitation. Amsterdam, **Applied Soil Ecology**, 28:247-57.

Smith P, Martino D, Cai Z, Gwary D, Janzen H, Kumar P, McCarl B, Ogle S, O'mara F, Rice C, Scholes B, Sirotenko O, Howden M, Mcallister T, Pan G, Romanenkov V, Schneider U, Towprayoon S, Wattenbach M, Smith J (2008) Greenhouse gas mitigation in agriculture. Philosophical Transactions of the Royal Society B: Rockville, **Biological Sciences**, 363:789-813.

Souza JLMD, Fezer KF, Gurski BC, Jerszurk D, Pachechenik PE, Evangelista AWP (2018) Atributos físicos e balanço hídrico do solo com floresta ombrófila mista, em latossolo vermelho-amarelo, em telêmaco Borba-PR, **Ciência Florestal**, 28:90-101.

Stackhouse PW, Chandler JrWS, Zhang T, Westberg D, Barnett AJ, Hoell JM (2016) Surface Meteorology and Solar Energy (SSE) Release 6.0 Methodology Version 3.2.0. **NASA, Langley Research Center**, pp. 76.

Tavares RLM, Oliveira SRDM, Barros FMMD, Farhate CVV, Souza ZMD, Junior S, La Scala JRN (2018) Prediction of soil CO₂ flux in sugarcane management systems using the Random Forest approach, **Scientia Agricola**, 75:281-287.

Tavares RLM, Siqueira DS, Panosso AR, Castioni GAF, Souza ZMD, Scala Jr, NL (2016) Soil management of sugarcane fields affecting CO₂ fluxes, **Scientia Agricola**, 73:543-551.

Teixeira DB, Bicalho ES, Panosso AR, Perillo LI, Iamaguti JL, Pereira GT, La Scala N (2012) Uncertainties in the prediction of spatial variability of soil CO₂ emissions and related properties. Viçosa-MG, **Revista Brasileira Ciência do Solo**, 36:1466-1475.

Terrer C, Phillips RP, Hungate BA, Rosende J, Pett-Ridge J, Craig ME, Van Groenigen KJ, Keenan TF, Sulman BN, Stocker BD, Reich PB, Pellegrini AFA, Pendall E, Zhang H, Evans RD, Carrillo Y, Fisher JB, Van Sundert K, Vicca S, Jackson RB (2021) A trade-off between plant and soil carbon storage under elevated CO₂. **Nature**, 591:599-603.

Thornley JH and Johnson IR (1990) **Plant and crop modelling**. Oxford: Clarendon.

Thornthwaite CW, Mather JR (1955) Publications in climatology, **The water balance**, 8:1-104.

Thornton P (2012) **Recalibrating food production in the developing world: global warming will change more than just the climate**. CCAFS Policy Brief n. 6. CGIAR Research Program on Climate Change, Agriculture and Food Security (CCAFS).

Tilman D, Balzer C, Hill J, Befort BL (2011) Global food demand and the sustainable intensification of agriculture. Washington-DC, **Proceedings of the National Academy of Sciences of the United States of America**, 108:20260-20264.

Trivellin PCO, Franco HCJ, Otto R, Ferreira DA, Vitti AC, Fortes C, Faron CE, Oliveira ECA, Cantarella H (2013) Impact of sugarcane trash on fertilizer requirements for São Paulo, Brazil, Piracicaba, **Scientia Agricola**, 70:345-352.

Tsai SM, Baraibar AVL, Romani VLM (1992) Efeitos de fatores físicos e químicos sobre os microrganismos do solo umidade e aeração. In: CARDOSO EJBN, TSAI SM, NEVES MCP (Coord.). **Microbiologia do solo**. Campinas: SBCS, p. 60-89.

Ture M, Kurt I, Kurum AT, Ozdamar K (2005) Comparing classification techniques for predicting essential hypertension. **Expert Systems with Applications**, 29:583-588.

Ussiri AN e Lal R (2009) Long-term tillage effects on soil carbon storage and carbon dioxide emissions in continuous corn cropping system from an alfisol in Ohio. Amsterdam, **Soil & Tillage Research**, 104:39-47.

Vermeulen SJ, Campbell BM, Ingram JSI (2012) Climate change and food systems. Palo Alto, **Annual Review of Environment and Resources**, 37:195-222.

Vicentini ME, Pinotti CR, Hirai WY, Moraes MLT, Montanari R, Filho MCMT, Milori DMBP, La Scala Jr N, Panosso AR (2019) CO₂ emission and its relation to soil temperature, moisture, and O₂ absorption in the reforested areas of Cerrado biome, Central Brazil, **Plant and Soil**, 444:193-211.

Vieira LH, Sentelhas PC, Pereira AB (2019) Sequence of workable days for mechanized harvest of sugarcane in Southern Brazil, Toronto, **Journal of Agricultural Science** 11:199-208.

Walter A, Galdos MV, Scarpore FV, Leal MRLV, Seabra JEA, Cunha MP, Picoli MCA, Oliveira COF (2014) Brazilian sugarcane ethanol: developments so far and challenges for the future, **Wiley Interdisciplinary Reviews: Energy and Environment**, 3:70-92.

Warner DL, Bond-Lamberty B, Jian J, Stell E, Vargas R (2019) Spatial predictions and associated uncertainty of annual soil respiration at the global scale. **Global Biogeochemical Cycles**, 33:1733-1745.

Warner DL, Guevara M, Inamdar S, Vargas R (2019) Upscaling soil-atmosphere CO₂ and CH₄ fluxes across a topographically complex forested landscape. **Agricultural and forest meteorology**, 264:80-91.

Wen HT, Lu JH, Jhang DS (2021) Features Importance Analysis of Diesel Vehicles' NO_x and CO₂ Emission Predictions in Real Road Driving Based on Gradient Boosting Regression Model. **International Journal of Environmental Research and Public Health**, 24:p.13044.

Witten IH e Frank E (2002) Data mining: practical machine learning tools and techniques with Java implementations. **Acm Sigmod Record**, 31:76-77.

Xie SP, Deser C, Vecchi GA, Collins M, Delworth TL, Hall A, Watanabe M (2015) Towards predictive understanding of regional climate change. London, **Nature Climate Change**, 5:921-930.

Yan H, Zhang J, Rahman SS, Zhou N, Suo Y (2020) Predicting permeability changes with injecting CO₂ in coal seams during CO₂ geological sequestration: A comparative study among six SVM-based hybrid models. **Science of the Total Environment**, 705:p.135941.

Yang L e Shami A (2020) Sobre otimização de hiperparâmetros de algoritmos de aprendizado de máquina: teoria e prática. **Neurocomputing**, 415:295-316.

Yang Y, Tilman D, Furey G, Lehman C (2019) Soil carbon sequestration accelerated by restoration of grassland biodiversity. **Nature communications**, 10:1-7.

Zang H, Cheng L, Ding T, Cheung KW, Wang M, Wei Z, Sun G (2020) Application of functional deep belief network for estimating daily global solar radiation: A case study in China. **Energy**, 191:p.116502.

Zar J. **Biostatistical Analysis**. New Jersey, USA: Prentice Hall, 1999. 663 p.