



MINISTÉRIO DA CIÊNCIA E TECNOLOGIA
INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS

sid.inpe.br/mtc-m21d/2024/05.02.13.39-TDI

**ESTIMATIVA DA TROCA LÍQUIDA DE CO₂ NA
AMAZÔNIA LEGAL COM BASE EM APRENDIZADO
DE MÁQUINA**

Aline Andrade do Nascimento

Dissertação de Mestrado do Curso de Pós-Graduação em Computação Aplicada, orientada pelos Drs. Alan James Peixoto Calheiros, e Luciana Varanda Rizzo, aprovada em 29 de abril de 2024.

URL do documento original:

<<http://urlib.net/8JMKD3MGP3W34T/4B8EPS2>>

INPE
São José dos Campos
2024

PUBLICADO POR:

Instituto Nacional de Pesquisas Espaciais - INPE
Coordenação de Ensino, Pesquisa e Extensão (COEPE)
Divisão de Biblioteca (DIBIB)
CEP 12.227-010
São José dos Campos - SP - Brasil
Tel.:(012) 3208-6923/7348
E-mail: pubtc@inpe.br

CONSELHO DE EDITORAÇÃO E PRESERVAÇÃO DA PRODUÇÃO INTELLECTUAL DO INPE - CEPPII (PORTARIA Nº 176/2018/SEI-INPE):

Presidente:

Dra. Marley Cavalcante de Lima Moscati - Coordenação-Geral de Ciências da Terra (CGCT)

Membros:

Dra. Ieda Del Arco Sanches - Conselho de Pós-Graduação (CPG)
Dr. Evandro Marconi Rocco - Coordenação-Geral de Engenharia, Tecnologia e Ciência Espaciais (CGCE)
Dr. Rafael Duarte Coelho dos Santos - Coordenação-Geral de Infraestrutura e Pesquisas Aplicadas (CGIP)
Simone Angélica Del Ducca Barbedo - Divisão de Biblioteca (DIBIB)

BIBLIOTECA DIGITAL:

Dr. Gerald Jean Francis Banon
Clayton Martins Pereira - Divisão de Biblioteca (DIBIB)

REVISÃO E NORMALIZAÇÃO DOCUMENTÁRIA:

Simone Angélica Del Ducca Barbedo - Divisão de Biblioteca (DIBIB)
André Luis Dias Fernandes - Divisão de Biblioteca (DIBIB)

EDITORAÇÃO ELETRÔNICA:

Ivone Martins - Divisão de Biblioteca (DIBIB)
André Luis Dias Fernandes - Divisão de Biblioteca (DIBIB)



MINISTÉRIO DA CIÊNCIA E TECNOLOGIA
INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS

sid.inpe.br/mtc-m21d/2024/05.02.13.39-TDI

**ESTIMATIVA DA TROCA LÍQUIDA DE CO₂ NA
AMAZÔNIA LEGAL COM BASE EM APRENDIZADO
DE MÁQUINA**

Aline Andrade do Nascimento

Dissertação de Mestrado do Curso de Pós-Graduação em Computação Aplicada, orientada pelos Drs. Alan James Peixoto Calheiros, e Luciana Varanda Rizzo, aprovada em 29 de abril de 2024.

URL do documento original:

<<http://urlib.net/8JMKD3MGP3W34T/4B8EPS2>>

INPE
São José dos Campos
2024

Dados Internacionais de Catalogação na Publicação (CIP)

Nascimento, Aline Andrade do.

Na17e Estimativa da troca líquida de CO₂ na Amazônia Legal com base em aprendizado de máquina / Aline Andrade do Nascimento. – São José dos Campos : INPE, 2024. xxii + 140 p. ; (sid.inpe.br/mtc-m21d/2024/05.02.13.39-TDI)

Dissertação (Mestrado em Computação Aplicada) – Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2024.

Orientadores : Drs. Alan James Peixoto Calheiros, e Luciana Varanda Rizzo.

1. NEE. 2. Carbono. 3. Ampliação de escala. 4. Amazônia. 5. Sensoriamento Remoto. I.Título.

CDU 004.85:504.7(81)



Esta obra foi licenciada sob uma Licença Creative Commons Atribuição-NãoComercial 3.0 Não Adaptada.

This work is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported License.



MINISTÉRIO DA
CIÊNCIA, TECNOLOGIA
E INOVAÇÃO



INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS

DEFESA FINAL DE DISSERTAÇÃO DE ALINE ANDRADE DO NASCIMENTO BANCA Nº 070/2024, REG. 408637/2022

No dia 29 de abril de 2024, às 08:30h, por teleconferência, o(a) aluno(a) mencionado(a) acima defendeu seu trabalho final (apresentação oral seguida de arguição) perante uma Banca Examinadora, cujos membros estão listados abaixo. O(A) aluno(a) foi APROVADO(A) pela Banca Examinadora, por unanimidade, em cumprimento ao requisito exigido para obtenção do Título de Mestre em Computação Aplicada, com a exigência de que o trabalho final a ser publicado deverá incorporar as correções sugeridas pela Banca Examinadora, com revisão pelo(s) orientador(es).

Novo título: "Estimativa da troca líquida de CO₂ na Amazônia Legal com base em aprendizado de máquina"

Membros da Banca:

Dr. Haroldo Fraga de Campos Velho – Presidente – INPE
Dr. Alan James Peixoto Calheiros – Orientador(a) – INPE
Dra. Luciana Varanda Rizzo – Orientador(a) – USP
Dr. Luiz Eduardo Oliveira e Cruz de Aragão – Membro Interno – INPE
Dr. Marcio José Teixeira – Membro Externo – Unicamp



Documento assinado eletronicamente por **Alan James Peixoto Calheiros, Tecnologista**, em 06/05/2024, às 13:24 (horário oficial de Brasília), com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Haroldo Fraga de Campos Velho, Pesquisador Titular**, em 06/05/2024, às 15:08 (horário oficial de Brasília), com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Márcio José Teixeira (E), Usuário Externo**, em 10/05/2024, às 15:50 (horário oficial de Brasília), com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **LUCIANA VARANDA RIZZO (E), Usuário Externo**, em 11/05/2024, às 15:07 (horário oficial de Brasília), com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Luiz Eduardo Oliveira E Cruz de Aragão, Pesquisador**, em 02/07/2024, às 14:58 (horário oficial de Brasília), com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no site <https://sei.mcti.gov.br/verifica.html>, informando o código verificador **11913338** e o código CRC **9408D4E0**.

Referência: Processo nº 01340.003473/2024-56

SEI nº 11913338

A meus pais Raimundo Pereira do Nascimento e Jania Fátima Andrade do Nascimento, à minha irmã Michelle Andrade do Nascimento e ao meu parceiro João Lucas de Faria Gonçalves, sem vocês nada disso seria possível.

AGRADECIMENTOS

Gostaria de expressar meus sinceros agradecimentos a todos que contribuíram para a realização desta dissertação .

Primeiramente, desejo expressar minha profunda gratidão à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), ao Instituto Nacional de Pesquisas Espaciais (INPE) e à Pós-graduação em Computação Aplicada pelo apoio financeiro e institucional fornecido durante este período de estudo.

Um agradecimento especial aos meus orientadores, Alan Calheiros e Luciana Rizzo, por sua orientação, incentivo e valiosos insights ao longo deste processo de pesquisa.

Quero expressar minha profunda gratidão aos meus pais, Raimundo Pereira e Jania Fatima, minha irmã Michelle Andrade, e meu parceiro João Lucas, por seu amor, apoio, suporte e compreensão e por serem meus suspiros de todos os dias.

Gostaria de agradecer também à minha melhor amiga, Brenda Axel, e à pequena Iolanda Axel, por trazerem leveza e alegria à minha vida, mesmo nos momentos mais desafiadores. Um agradecimento especial também às minhas cachorras, Lola, Geléia e Dandara, por seus momentos de brincadeiras que ajudaram a diminuir minha ansiedade e trazer um pouco de calma ao caos da vida.

A todos vocês, meu mais profundo obrigado por fazerem parte deste importante capítulo da minha vida.

RESUMO

As medições de troca líquida de carbono (NEE) na Amazônia enfrentam desafios devido à escassez de torres de fluxo, à heterogeneidade na cobertura e uso da terra, e às limitações temporais e espaciais das medições disponíveis. Neste estudo, busca-se superar esses obstáculos utilizando dados meteorológicos (METEO) dos produtos ERA5 e Merge, juntamente com dados de sensoriamento remoto (RS) dos produtos MODIS e CERES como preditores em modelos de aprendizado de máquina. A série temporal diária de NEE foi estimada no período de 2002 a 2011 na Célula K67, que engloba as torres de fluxo BrSa1, K67, K77 e K83 na Floresta Nacional dos Tapajós. Os valores de NEE medidos na torre foram utilizados como alvo nos modelos, que foram validados e testados nas torres K77 e K34, esta última localizada em Manaus, para avaliar sua capacidade de generalização. O modelo de floresta aleatória, com preditores do conjunto RS e variável alvo da torre K77, demonstrou o melhor desempenho. Em seguida, o modelo foi aplicado para realizar a ampliação de escala na Amazônia Legal, visando subsidiar estudos sobre a compreensão dos fluxos de carbono nesses ecossistemas. Este estudo visa contribuir para o avanço na compreensão dos padrões de troca de carbono na Amazônia e destacar o potencial dos modelos de aprendizado de máquina na estimativa de NEE na região.

Palavras-chave: NEE. Carbono. Ampliação de escala. Amazônia. Sensoriamento Remoto. Modelos de Aprendizado de Máquina. Floresta Aleatória.

ESTIMATE OF NET CO₂ EXCHANGE IN THE LEGAL AMAZON BASED ON MACHINE LEARNING

ABSTRACT

Net ecosystem exchange (NEE) measurements in the Amazon face challenges due to the scarcity of flux towers, heterogeneity in land cover and use, and the temporal and spatial limitations of available measurements. This study aims to overcome these obstacles by using meteorological data (METEO) from ERA5 and Merge products, along with remote sensing (RS) data from MODIS and CERES products as predictors in machine learning models. The daily time series of NEE was estimated from 2002 to 2011 in Cell K67, encompassing the BrSa1, K67, K77, and K83 flux towers in the Tapajós National Forest. NEE values measured at the towers were used as targets in the models, which were validated and tested at the K77 and K34 towers, the latter located in Manaus, to assess their generalization capability. The random forest model, with predictors from the RS set and the target variable from the K77 tower, demonstrated the best performance. Subsequently, the model was applied to upscale NEE in the Legal Amazon, aiming to support studies on the understanding of carbon fluxes in these ecosystems. This study aims to contribute to advancing the knowledge of carbon exchange patterns in the Amazon and to highlight the potential of machine learning models in estimating NEE in the region.

Keywords: NEE. Carbon. Upscaling. Amazon. Remote Sensing. Machine Learning Models. Random Forest.

LISTA DE FIGURAS

	<u>Pág.</u>
2.1 Taxa de desmatamento anual na Amazônia Legal em milhares de quilômetros.	7
2.2 Representação da arquitetura geral de uma MLP.	18
2.3 Demonstração da técnica híbrida: Lógica Fuzzy + Redes Neurais.	19
2.4 Arquitetura da Rede Neural ANFIS.	20
2.5 Representação do grafo e sua matriz de adjacência.	22
2.6 Representação do grafo de entrada de uma rede neural à esquerda e representação da propagação e transformação de informação no grafo à direita.	23
3.1 Delimitação do Bioma Amazônia.	28
3.2 Mapeamento Mapbiomas para as células K67 (à esquerda) e K34 (à direita) no ano de 2011.	29
3.3 Fluxograma Fluxnet.	33
3.4 Sítios de torres de fluxo da iniciativa LBA.	34
3.5 Fluxograma de extração dos dados LBA.	35
3.6 Fluxograma de extração dos produtos MODIS.	39
3.7 Fluxograma de extração dos produtos CERES.	41
3.8 Fluxograma de extração dos produtos ERA5.	43
3.9 Cálculo do Deficit de Pressão de Vapor.	44
3.10 Fluxograma de extração dos dados MERGE.	45
4.1 Fluxograma geral da metodologia utilizada neste trabalho.	47
4.2 Arquivo netcdf.	48
4.3 Estrutura de cada atributo no netcdf.	48
4.4 Grade Bioma Amazônia , Célula K34 e Célula K67, em vermelho e laranja, respectivamente.	50
4.5 Grafo base METEO.	58
4.6 Grafo base RS.	58
4.7 Método de ampliação de escala.	63
5.1 Comparação da temperatura do ar a 2m (t2m) do ERA5 e temperatura do ar (ta) da Torre K67.	66
5.2 Comparação da temperatura do ar a 2m (t2m) do modelo ERA5 com a temperatura do ar (ta) da Torre K34.	66

5.3	Gráficos de distribuições das médias diárias dos atributos do conjunto METEO para a Célula K67.	68
5.4	Gráficos de distribuições das médias diárias dos atributos do conjunto METEO para a Célula K34.	69
5.5	Gráficos de distribuições das médias diárias dos atributos do conjunto RS para a Célula K67.	71
5.6	Gráficos de Distribuições das médias diárias dos atributos do conjunto RS para a Célula K34.	72
5.7	Séries Temporais de NEE para as Torres K77, K67, K83 e BrSa1. E Gráfico de Comparação entre as séries K67 e BrSa1 ao fim da Figura. . .	74
5.8	Boxplots mostrando a distribuição dos atributos de NEE das bases Fluxnet BrSa1e LBA K67.	77
5.9	BoxPlot das variáveis NEE de cada uma das torres.	79
5.10	Gráfico dos comportamento mensal anual de NEE nas Torres K77, K67, K34, K83 e BrSa1.	80
5.11	Gráfico da matriz diagonal da correlação de Spearman dos atributos METEO da Célula K67.	83
5.12	Gráfico da matriz diagonal da correlação de Spearman dos atributos METEO da Célula K34.	84
5.13	Comparação da Distribuição de Dados Meteorológicos entre a Célula K67 e a Célula K34: BoxPlot.	87
5.14	Gráfico da matriz diagonal da Correlação de Spearman dos atributos RS da Célula K67.	89
5.15	Gráfico da matriz diagonal da Correlação de Spearman dos atributos RS da Célula K34.	90
5.16	BoxPlot RS K34 x BrSa1.	91
5.17	Gráfico da matriz diagonal da Correlação de Spearman dos atributos METEORS da Célula K67.	93
5.18	Gráfico da matriz diagonal da Correlação de Spearman dos atributos METEORS da Célula K34.	94
5.19	Gráficos de Pares Comparativos das Métricas de Validação dos Preditores separados por cores para todos os modelos utilizados neste trabalho. . . .	100
5.20	Gráficos de Dispersão Comparativos dos Dados Observados e Simulados para as Etapas de Validação e Testes na estimativa de NEE para a Torre BrSa1 do conjunto FLUXNET. Referente aos modelos da Tabela 5.10. . .	102
5.21	Gráficos de Pares Comparativos das Métricas de Validação dos Modelos RF e Gboost separados por cores por Variável Alvo do Modelo.	103

5.22	Gráficos de dispersão comparativos dos dados observados e estimados para as etapas de validação e testes na estimativa de NEE para a Torre K77.	105
5.23	Gráficos de Dispersão Comparativos dos Dados Observados e Simulados para a Etapa de Testes na estimativa de NEE para a Torre K34.	107
5.24	Valores de baixa qualidade por célula para cada produto MODIS.	108
5.25	Valores máximos, mínimos e médios de NEE (ampliação de escala) no bioma Amazônia. Estimativa realizada com preditores RS e modelo RF.	109
5.26	Tendência de Mann Kendall da estimativa de NEE para toda a Amazônia Legal com valor $p < 0,05$ para o modelo RF com preditores RS.	110
5.27	Valores máximos, mínimos e médios de NEE (Ampliação de escala) no Bioma Amazônia. Estimativa realizada com Preditores METEORS e modelo RF.	111
5.28	Tendência de Mann Kendall da estimativa de NEE para todo o Bioma Amazônia com valor $p < 0,05$ para o modelo RF com preditores METEORS.	111
5.29	Comportamento mensal médio de NEE criado a partir de dados preditores RS no modelo RF.	113
5.30	Comportamento mensal médio de NEE criado a partir de dados preditores RS - FLUXCOM.	114
5.31	Comportamento mensal médio de NEE criado a partir de dados preditores METEORS no modelo RF.	115
5.32	Comportamento mensal médio de NEE do criado a partir de dados preditores METEORS - FLUXCOM.	116
5.33	Gráficos de Dispersão Comparativos dos Dados Observados da Torre K77 e Simulados pelos modelos Fluxcom e RF, com preditores METEORS, para a Etapa de Testes na estimativa de NEE.	118
5.34	Gráficos de Dispersão Comparativos dos Dados Observados da Torre K77 e Simulados pelos modelos Fluxcom e RF, com preditores RS, para a Etapa de Testes na estimativa de NEE.	118
A.1	Comportamento Anual médio de NEE com o modelo RF e preditores RS.	136
A.2	Comportamento Anual da ampliação de escala realizada pelo Projeto Fluxcom com preditores RS.	137
A.3	Resultados da Tendência de Mann Kendall para atributos referentes à disponibilidade de água.	138
A.4	Resultados dos Testes de Mann Kendall para atributos referentes à atividade fotossintética.	139
A.5	Resultados dos Testes de Mann-Kendall para atributos referentes à Radiação Difusa.	140

LISTA DE TABELAS

	<u>Pág.</u>
3.1 Dados de Fluxos Turbulentos de CO ₂	31
3.2 Produtos do MODIS e suas variáveis.	36
3.3 Produtos do CERES e suas variáveis.	37
3.4 Dados Meteorológicos do Modelo Era5 do ECMWF Climate Reanalysis e Precipitação do Produto MERGE.	42
4.1 Conjuntos de Dados Preditores.	48
4.2 Conjuntos de Atributos utilizados no treinamento.	55
4.3 Conjuntos de treinamento.	56
4.4 Características dos grafos de entradas na GNN.	58
4.5 Hiperparâmetros Ajustados dos modelos de aprendizado de máquina.	60
5.1 Métricas de Intercomparação entre as temperaturas ' t2m ' do modelo ERA5 e ' ta ' das Torres K67 e K34.	65
5.2 Métricas estatísticas dos atributos do conjunto METEO para as células K67 e K34. A descrição dos atributos pode ser encontrada na Tabela 3.4.	67
5.3 Métricas estatísticas dos atributos do conjunto RS para a célula K67. A descrição dos atributos pode ser encontrada nas Tabelas 3.2 e 3.3.	70
5.4 Tendência de Mann Kendall de cada série temporal das variáveis NEE de suas respectivas Torres.	75
5.5 Estatísticas dos Atributos de NEE da Fluxnet (BrSa1) e LBA (K67) e de comparação entre as séries temporais do Fluxnet BrSa1 e LBA K67.	76
5.6 Estatísticas dos Atributos de NEE das Torres do LBA e Fluxnet.	78
5.7 Tendências avaliadas com o teste de Mann Kendall dos preditores do conjunto METEO nas Células K67 e K34.	86
5.8 Tendências avaliadas com o teste de Mann Kendall dos preditores do conjunto RS nas Células K67 e K34.	92
5.9 Hiperparâmetros dos Modelos com melhor desempenho.	98
5.10 Métricas de validação e teste do modelo com melhor desempenho utili- zando dados do conjunto Fluxnet - Torre BrSa1 com filtro de qualidade de >0,7.	102
5.11 Métricas de validação e teste dos modelos com melhor desempenho em relação a Torre K77.	105
5.12 Métricas de teste dos modelos com melhor desempenho - Teste com a Torre K34.	107

5.13 Métricas de comparação dos resultados dos modelos RF e Fluxcom com a Torre K77.	117
---	-----

LISTA DE ABREVIATURAS E SIGLAS

alnid	– Albedo no infravermelho próximo para radiação difusa
aluvd	– Albedo no ultravioleta do visível para radiação difusa
b02	– Refletância da superfície na banda do azul
b03	– Refletância da superfície na banda do infravermelho próximo
b07	– Refletância da superfície na banda do infravermelho médio
BIAS	– Viés
Br-Sa1	– Torre de fluxo localizada na região de Santarém (PA) do conjunto Fluxnet
CV	– Coeficiente de variação
d2m	– Temperatura de ponto de orvalho a 2 metros
e	– Evaporação
e_a	– Pressão de vapor
e_s	– Pressão de vapor de saturação
EDA	– Análise exploratória de dados
EVI	– Índice de vegetação melhorado
Fpar	– Fração de radiação fotossinteticamente ativa absorvida
FTP	– File transfer protocol
GPP	– Produtividade primária bruta (do inglês, gross primary production)
grib	– Formato de dados para previsão meteorológica (gridded binary)
IBGE	– Instituto brasileiro de geografia e estatística
INPE	– Instituto nacional de pesquisas espaciais
K34	– Torre de fluxo localizada na região de Manaus (AM) do conjunto LBA
K67	– Torre de fluxo localizada na região de Santarém (PA) do conjunto LBA
K77	– Torre de fluxo localizada na região de Santarém (PA) do conjunto LBA
K83	– Torre de fluxo localizada na região de Santarém (PA) do conjunto LBA
Lai	– Índice de área foliar
LBA	– Large biosphere
LST_Day	– Temperatura da superfície terrestre durante o dia
LST_Night	– Temperatura da superfície terrestre durante a noite
MAE	– Erro absoluto médio
max	– Máximo
mean	– Média
med	– Mediana
METEO	– Conjunto com dados preditores meteorológicos
METEO+RS	– Conjunto com dados preditores meteorológicos e de sensoriamento remoto
min	– Mínimo
MSE	– Erro quadrático médio
NDVI	– Índice de vegetação por diferença normalizada
NEE	– Troca líquida de carbono (em inglês, net ecosystem exchange)
Netcdf	– Formato de dados para troca (network common data form)
NPP	– Produtividade primária líquida (do inglês, net primary production)
Par	– Radiação fotossinteticamente ativa
par_all	– Fluxo difuso de PAR - todo o céu

par_clr	– Fluxo difuso de PAR - céu claro
par_pri	– Fluxo difuso de PAR - prístino
prec	– Precipitação
R	– Coeficiente de correlação
ROV	– Variabilidade observada relativa
R ²	– Coeficiente de determinação
RMSE	– Raiz do erro quadrático médio
ro	– Escoamento superficial
RS	– Conjunto com dados preditores sensoriamento remoto
Sk	– Coeficiente de assimetria
std	– Desvio padrão
stl1	– Temperatura do solo nível 1
sw_all	– Fluxo difuso de onda curta - todo o céu
sw_clr	– Fluxo difuso de onda curta - céu claro
sw_naer	– Fluxo difuso de onda curta - sem aerossol
sw_pri	– Fluxo difuso de onda curta - prístino
t2m	– Temperatura do ar a 2 metros
ta	– Temperatura do ar medida pela torre de fluxo
tcwv	– Vapor de água total na coluna
var	– Variância
VPD	– Déficit de pressão de vapor

SUMÁRIO

	<u>Pág.</u>
1 INTRODUÇÃO	1
1.1 Objetivos e hipóteses	5
1.1.1 Objetivo principal	5
1.1.2 Objetivos específicos	5
1.1.3 Hipótese	5
2 REFERENCIAL TEÓRICO	7
2.1 Trocas de carbono entre a biosfera e a atmosfera	8
2.2 Troca líquida de carbono (NEE)	11
2.3 Modelos baseados em processos físicos e biogeoquímicos	14
2.4 Modelos de aprendizado de máquina	15
2.4.1 Floresta aleatória (RF)	16
2.4.2 Gradient boosting regressor (Gboost)	17
2.4.3 MultiLayer perceptron	17
2.4.4 Sistema adaptativo de inferência neuro-fuzzy (ANFIS)	18
2.4.5 Rede neural de grafos	21
2.5 Ampliação de escala (do inglês, <i>Upscaling</i>)	24
3 DADOS	27
3.1 Área de estudo: Amazônia Legal	27
3.2 Covariância de vórtices turbulentos	30
3.3 Sensoriamento remoto (RS)	35
3.4 Meteorológicos (METEO)	41
4 MÉTODOS	47
4.1 Integração e pré processamento	47
4.2 Análise exploratória dos dados e intercomparação	49
4.2.1 Intercomparação da temperatura do ar entre o modelo ERA5 e os dados observados das torres K34 e K67	50
4.2.2 Distribuição dos dados, células K67 e K34.	52
4.2.3 Análise das séries temporais	52
4.3 Treinamento dos modelos de aprendizado de máquina	54
4.3.1 Definição dos conjuntos de preditores e de variáveis alvo	54

4.3.2	Treinamento, validação e teste	59
4.4	Ampliação de escala	62
5	RESULTADOS E DISCUSSÃO	65
5.1	Análise exploratória dos dados: intercomparação temperatura do modelo ERA5 versus temperatura das torres K67 e K34	65
5.2	Análise exploratória dos dados: distribuição dos conjuntos de dados pre- ditores	66
5.2.1	Distribuição dos dados dos atributos do conjunto METEO nas células K67 e K34	66
5.2.2	Distribuição dos dados dos atributos do conjunto RS nas células K67 e K34	70
5.3	Análise exploratória dos dados: variabilidade temporal do NEE	73
5.4	Análise exploratória dos dados: análise de correlação no conjunto de pre- ditores METEO	82
5.5	Análise exploratória dos dados: Análise de correlação no conjunto de preditores RS	88
5.6	Modelos de aprendizado de máquina: preparação dos dados	95
5.7	Modelos de aprendizado de máquina: treinamento e validação e teste . . .	96
5.8	Ampliação de escala	108
6	CONCLUSÕES	119
	REFERÊNCIAS BIBLIOGRÁFICAS	123
	APÊNDICE A	135
A.1	Figuras extras da ampliação de escala de NEE para o bioma Amazônia .	135

1 INTRODUÇÃO

A ONU reúne líderes de todo o mundo desde 1995 para discutir possíveis ações de enfrentamento às mudanças climáticas globais. Em 2023, ocorreu a COP28, que teve como um dos objetivos a busca por soluções para frear ou minimizar os impactos das mudanças climáticas na Terra, incluindo a neutralização do carbono. O Governo Brasileiro assumiu neste evento diversos compromissos para reduzir as emissões de carbono, que incluem zerar o desmatamento ilegal até 2030, restaurar milhões de hectares de pastagens degradadas e reflorestar outros milhões de hectares de áreas desmatadas, com o objetivo de promover o aumento dos estoques de carbono das florestas e compensar emissões antrópicas de dióxido de carbono, CO₂, para a atmosfera ([UNITED NATIONS FRAMEWORK CONVENTION ON CLIMATE CHANGE \(UNFCCC\), 2023](#)).

De acordo com [Sistema de estimativas de emissões e remoções de gases de efeito estufa \(2022\)](#), em 2022, cerca de 66% das emissões líquidas brasileiras de CO₂ são provenientes das alterações da vegetação, 29% de mudanças de uso da terra e floresta, 37% do setor de agropecuária e o restante são dos setores de energia, resíduos e processos industriais. Essa configuração diferencia enormemente o Brasil dos outros países desenvolvidos e mesmo de economias emergentes como China e Índia: enquanto nessas a queima de combustíveis fósseis é responsável por 60% a 80% das emissões, no Brasil, ao contrário, 66% das emissões são resultantes direta ou indiretamente da mudança do uso da terra.

Os compromissos foram escolhidos com base no que foi exposto anteriormente e no comportamento do ciclo do carbono em ecossistemas florestais, tendo como objetivo atingir o balanço de carbono neutro. Esse ciclo é composto de três processos fundamentais, a absorção do carbono por meio da fotossíntese, o estoque do carbono por meio dos processos de manutenção da vegetação e a emissão do carbono através dos processos de respiração autotrófica, heterotrófica e fluxos de emissão decorrentes de perturbações induzidas pelo homem ([ARAGÃO; POULTER, 2014](#)).

Se um ecossistema está em equilíbrio do ponto de vista do ciclo do carbono, a quantidade de carbono emitida é igual à quantidade absorvida pelo ecossistema, mantendo assim um balanço neutro. Quando um ecossistema florestal apresenta um balanço de carbono positivo, significa que ele emite mais CO₂ do que absorve. Por outro lado, quando o balanço de carbono é negativo, o sistema florestal captura mais CO₂ do que emite.

Os distúrbios antrópicos, tais como o desmatamento, queimadas e outras atividades, ocasionam um aumento significativo na emissão de CO_2 e, conseqüentemente, no acúmulo desse gás na atmosfera. A emissão advém da combustão de material orgânico, transferindo o carbono que estava estocado na vegetação e no solo para a atmosfera. Com isso, após o distúrbio, o ecossistema afetado passa a apresentar um balanço de carbono positivo. Esse acúmulo de CO_2 , juntamente com outros gases gerados a partir desses ou outros processos antrópicos, intensificam o efeito estufa, contribuindo para o aquecimento global (INTERGOVERNMENTAL PANEL ON CLIMATE CHANGE (IPCC), 2021). De acordo com Gatti et al. (), as mudanças climáticas também possuem um impacto significativo nas emissões de CO_2 na região amazônica, foi verificado que o aumento da intensidade da estação seca e a aceleração de desflorestação promovem o estresse dos ecossistemas da região, ocasionando menor absorção de CO_2 pela floresta, comportamento predominante no leste da região.

Na natureza, os sistemas tendem naturalmente ao estado de equilíbrio. Por exemplo, há sistemas com várias variáveis que se equilibram mutuamente. Quando o valor de uma dessas variáveis é aumentado, o sistema reorganiza-se (as variáveis são alteradas) para alcançar novamente o equilíbrio. O equilíbrio é, portanto, o estado natural de qualquer sistema físico. O sistema climático da Terra não é exceção a essa regra. Quando a temperatura da Terra é aumentada de forma anômala, o sistema climático é desequilibrado e as variáveis do sistema são reorganizadas para restabelecer o equilíbrio. Na prática, o rearranjo que o clima faz e que vemos são eventos extremos, tais como o aumento da frequência e intensidade de ondas de calor em várias regiões do mundo e aumento da intensidade das chuvas extremas, ocasionando maior frequência de inundações mais frequentes e severas em diversas partes do globo. (SENEVIRATNE et al., 2021)

Os compromissos firmados pelo Brasil para mitigar as mudanças climáticas é justificado pelos elementos apresentados anteriormente. A mudança climática representa um risco para a vida humana e deve ser evitada. Diversas iniciativas de pesquisas em todo o mundo foram criadas para estudar o ciclo do carbono e avaliar os impactos do seu desequilíbrio em ecossistemas florestais e no clima.

No entanto, o estudo do ciclo do carbono em ecossistemas florestais não é trivial e exige um entendimento espacial e temporal dos processos ecossistêmicos (WARING; RUNNING, 2007). Existem métodos experimentais para quantificar estoques e trocas de carbono a partir da observação direta (KONDO; PATRA, 2020). Porém, há

importantes limitações relacionadas a cobertura espacial e temporal das observações diretas. Isso implicou na necessidade de encontrar um meio alternativo para expressar quantitativamente o conhecimento adquirido sobre o ciclo de carbono no nível de ecossistemas, a fim de extrapolar para escalas de tempo e espaço maiores (WARING; RUNNING, 2007).

O método convencional utilizado por cientistas do ecossistema para medir os fluxos de gases traço entre os ecossistemas e a atmosfera é geralmente a covariância de vórtices turbulentos. Uma das principais vantagens deste método é a sua capacidade de medir fluxos diretamente (*in situ*), sem perturbar o ecossistema. A medição é realizada em torres localizadas dentro dos ecossistemas em estudo, as quais estão distribuídas em alguns pontos do globo e são operadas por diferentes centros de pesquisa (e.g., (RESTREPO-COUBE et al., 2013)). Em 2007, foi lançada a iniciativa Fluxnet, que teve como objetivo unir, documentar e padronizar os dados coletados por essas torres, tornando-os disponíveis para a comunidade científica, em geral por meio dos conjuntos de dados La Thuile e Fluxnet 2015. Essa disponibilização tem como objetivo possibilitar que diferentes grupos possam realizar estudos para entender o funcionamento dos diferentes ecossistemas e a detecção de tendências em relação ao clima, gases de efeito estufa e poluição do ar (FLUXNET, 2023).

As torres permitem o estudo local do comportamento do ecossistema em que se encontram. No entanto, para estudar o comportamento regional, é necessário utilizar dados que possibilitem a análise física do ciclo do carbono espacial e temporalmente. Modelos físicos-computacionais de superfície e de vegetação dinâmica oferecem uma alternativa válida para experimentos em larga escala de ecossistemas (FISHER; KOVEN, 2020). Isso ocorre se eles puderem representar mecanismos de acoplamento de processos biogeoquímicos de maneira realista, sem exigir uma quantidade excessiva de dados e parâmetros difíceis ou impossíveis de adquirir.

Outra possibilidade é o uso de modelos baseados em dados. Iniciativas como a Fluxcom (TRAMONTANA et al., 2016) e a desenvolvida por Zhuravlev (ZHURAVLEV et al., 2022) têm como objetivo abordar o problema da limitação de cobertura espacial e temporal de observações diretas. Essas iniciativas utilizaram estratégias que combinam dados de sensoriamento remoto e meteorológicos para medir variáveis biogeoquímicas e estimar as variáveis de balanço de carbono e energia em ecossistemas terrestres. O objetivo é ampliar as medições locais realizadas em torres de fluxo para uma escala global, usando modelos de regressão baseados em aprendizado de máquina, em um processo conhecido como ampliação de escala (do inglês, upscaling).

Na empreitada da Fluxcom, os modelos foram submetidos a uma análise rigorosa por meio de validação cruzada, resultando em uma compilação global dos fluxos baseada na mediana das distribuições dos resultados dos modelos de aprendizado de máquina (TRAMONTANA et al., 2016). Essa avaliação metódica permitiu à Fluxcom identificar a capacidade preditiva para todas as variáveis de fluxo investigadas. Contudo, a estimativa da troca líquida de carbono emergiu como aquela com a menor capacidade de predição. Enquanto a radiação líquida apresentou valores de eficiência do modelo em torno de 0,99, a troca líquida de carbono obteve apenas 0,64. Essa disparidade ressalta uma lacuna significativa na precisão das estimativas de carbono, apesar da eficácia geral dos modelos Fluxcom em reproduzir outros fluxos ambientais (TRAMONTANA et al., 2016).

A partir da ampliação de escala, a Fluxcom alcançou resultados notáveis para regiões como os Estados Unidos da América e a Europa e atingindo um coeficiente de determinação de 0,99. No entanto, para a designada "América do Sul Tropical" que abrange o Bioma Amazônia, foram obtidas as piores estimativas (TRAMONTANA et al., 2016). O maior coeficiente de determinação encontrado para a área pela Fluxcom foi, a partir dos dados de sensoriamento remoto, 0,1 e com a junção com os dados meteorológicos foi de 0,33 (JUNG et al., 2020).

O estudo da região Amazônica é severamente limitado devido à escassez de dados disponíveis em toda a área. Esta escassez é uma grande limitação para pesquisas na região devido à sua vasta extensão espacial, à falta de torres de fluxo e à alta nebulosidade presente na região. Esses fatores afetam significativamente grande parte das medições realizadas na região, principalmente por satélites. Outro fator importante é que o desenvolvimento de um modelo generalizado para todas as regiões deve favorecer outras regiões em detrimento da Amazônia, o que deixa uma lacuna a ser investigada, de modo a melhorar essa estimativa.

A heterogeneidade no uso do solo e na vegetação da região e a falta de continuidade temporal e espacial das medições de fluxo torna necessário o uso de novos métodos capazes de lidar com esses fatores limitantes dos dados sem comprometer as informações físicas do ecossistema em questão. O bioma Amazônia possui uma grande variedade de ecossistemas e para o entendimento deles é necessário o desenvolvimento de modelos computacionais avançados que permitam uma maior compreensão dos ecossistemas e conseqüente maior entendimento das emissões de carbono na região.

1.1 Objetivos e hipóteses

1.1.1 Objetivo principal

Com base na relevância de compreender os fluxos de carbono na região Amazônica e seu impacto nas mudanças climáticas, este trabalho teve como objetivo a estimativa da troca líquida de carbono para a Amazônia Legal. Esta estimativa foi elaborada a partir de diferentes modelos de aprendizado de máquina, nos quais foram utilizados como atributos preditivos dados meteorológicos, dados de sensoriamento remoto e dados de torres de fluxo. Essa abordagem ajuda a compreender melhor o comportamento dos ecossistemas amazônicos em relação ao fluxo de carbono e às suas interações com o meio ambiente.

1.1.2 Objetivos específicos

Para atingir o objetivo principal, foram realizados os objetivos específicos listados abaixo: 1. Análise estatística das variáveis preditoras meteorológicas e de sensoriamento remoto da célula onde estão as torres de fluxo de carbono.

2. Uma análise exploratória e comparativa do comportamento mensal da variável de troca líquida de carbono no ecossistema das Torres na Amazônia Legal do conjunto FLUXNET e LBA no período em questão.

3. Estudo da correlação e do comportamento das variáveis preditoras.

4. Avaliação de diferentes métodos de aprendizado de máquinas.

5. Avaliação dos melhores modelos em termos de atributos alvo e preditores.

6. Avaliação da suficiência de variáveis preditivas para a estimativa das séries temporais diárias da troca líquida de CO₂ para avaliação da capacidade preditiva da região.

7. Ampliação de escala a partir da estimativa das séries temporais diárias da troca líquida de CO₂ para toda a Amazônia Legal com resolução espacial de 0.25°.

1.1.3 Hipótese

A estimativa atual da troca líquida de carbono (do inglês, net ecosystem exchange - NEE) para a região Amazônica foi obtida através do modelo global criado pela iniciativa Fluxcom. No entanto, supõe-se que um modelo desenvolvido exclusivamente

para a região Amazônica, utilizando dados de conjuntos de dados específicos dessa área, pode resultar em estimativas mais precisas e confiáveis de NEE. A justificativa para essa suposição é que um modelo focado na Amazônia pode capturar melhor as características únicas e as variáveis ambientais específicas dessa região, que são fundamentais para uma estimativa mais precisa do NEE.

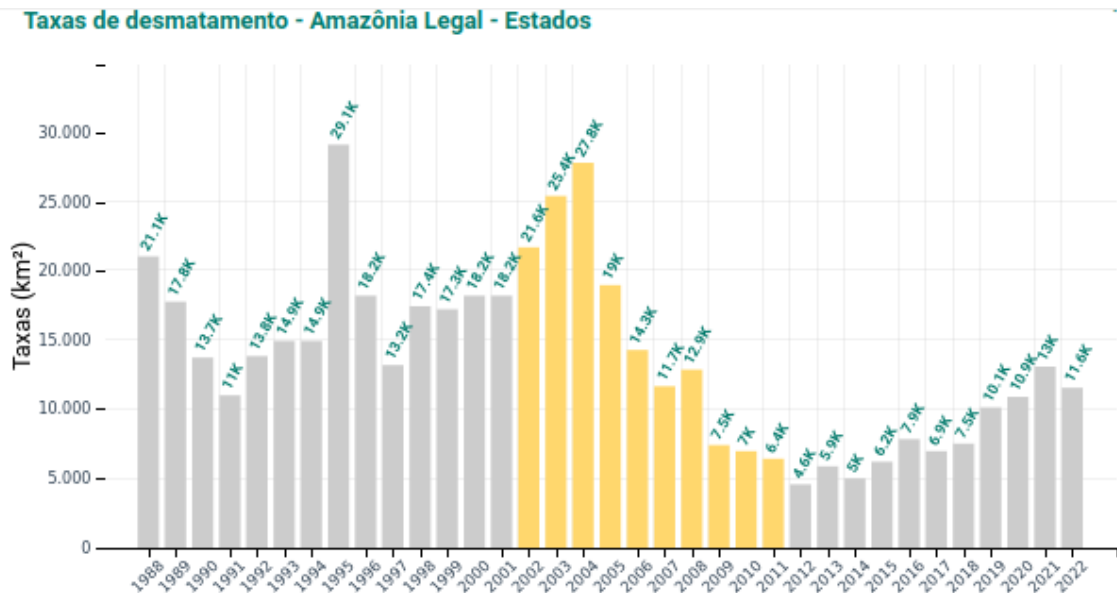
2 REFERENCIAL TEÓRICO

O bioma Amazônia abriga 1,5 milhão de espécies vegetais catalogadas (INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (IBGE), 2019), de modo que os ciclos de carbono, água e nutrientes são diversos entre si, seja pelos diferentes ecossistemas e/ou pelos diferentes microclimas presentes nesse bioma. As respostas biofísicas e bioquímicas da vegetação sobre o clima depende da sua fenologia e resposta fisiológica às condições ambientais e ao ciclo de nutrientes. Ou seja, para diferentes espécies há diferentes respostas (PAGAN et al., 2019).

No entanto, o bioma sofre com muitas atividades antrópicas, a mais recorrente é o desmatamento que é ocasionado por exploração madeireira ilegal, agricultura, pecuária, urbanização e mineração. De acordo com o projeto Assis (2019), o desmatamento ocorre em milhares de km² na Amazônia Legal, todo ano. Na Figura 2.1 é possível ver a evolução anual desse processo antrópico na região.

O comportamento sazonal do fluxo de carbono na Amazônia exhibe padrões distintos, incluindo um contraste notável entre o aumento da GPP na estação seca em

Figura 2.1 - Taxa de desmatamento anual na Amazônia Legal em milhares de quilômetros.



As barras amarelas no histograma se referem ao período do estudo desse trabalho, de 2012 a 2011.

Fonte: Assis (2019).

regiões de floresta tropical intacta da Amazônia central versus declínios sazonais em florestas convertidas, florestas do sul ou savanas. Esses padrões emergem de padrões de disponibilidade de água, distribuição de raízes, luz solar e estratégia fenológica das plantas. A compreensão de como os ritmos sazonais da biologia, clima e recursos interagem para regular os fluxos de carbono é uma parte fundamental da compreensão e previsão da resposta, resiliência e mudança futura da floresta à seca (RESTREPO-COUBE et al., 2013; ARAGÃO; POULTER, 2014).

2.1 Trocas de carbono entre a biosfera e a atmosfera

Para estudar o ciclo do carbono em um ecossistema é necessário considerar o ciclo da água, de nutrientes e elementos climáticos que o condicionam. Entender o funcionamento de cada componente do ciclo do carbono e quais elementos os afetam é essencial para conseguir encontrar as variáveis necessárias para elaboração de um modelo que consiga estimar o fluxo líquido de carbono entra a biosfera e a atmosfera.

De acordo com Waring e Running (2007), o carbono inicia seu ciclo na fotossíntese, quando os ecossistemas florestais assimilam o CO_2 atmosférico no processo de produção primária bruta (GPP, do inglês *Gross Primary Production*), sendo o carbono utilizado na síntese e manutenção de células vivas. A partir da respiração autotrófica (R_a) há a liberação de CO_2 de volta para a atmosfera. A diferença entre GPP e R_a é chamada de produção primária líquida, armazenada nas folhagens, galhos, troncos, raízes e órgãos reprodutivos das plantas.

Ao decorrer do tempo as plantas perdem folhas, raízes e morrem. Essa matéria orgânica forma detritos, o qual é um substrato que sustenta animais e micróbios, e através do seu metabolismo heterotrófico, libera CO_2 para a atmosfera. Um ecossistema não perturbado, ao decorrer do tempo, consome mais carbono do que emite, devido ao seu desenvolvimento. No entanto, se há a ocorrência de algum distúrbio que reduza a fotossíntese, o sistema irá perder carbono e ao decorrer do tempo, liberar mais CO_2 do que consome (WARING; RUNNING, 2007).

Esses distúrbios podem ocorrer de diversas maneiras, geralmente são decorrentes de ações antrópicas, seja a partir de desmatamento, queimadas, coleta seletiva de madeiras e degradação florestal. Ou a partir do efeito das mudanças climáticas, com a ocorrência de anomalias nos elementos climáticos, como precipitação e temperatura, que estão correlacionados diretamente com a realização de fotossíntese e manutenção dos ecossistemas. Ou, como no caso da floresta Amazônia, pode ser decorrente de ambos (ARAGÃO; POULTER, 2014; GATTI et al.,).

Segundo [Gatti et al. \(\)](#), em seu estudo realizado entre os anos de 2010 a 2018, partes da Amazônia podem não atuar como um sumidouro de carbono e sim como fonte de CO₂ para a atmosfera. Foi visto que as regiões nordeste e sudeste da Amazônia são as que mais sofrem com o desmatamento e são as que emitem maior quantidade de CO₂ na atmosfera, além de serem as que se mostraram com a maior diminuição de precipitação e o maior aumento da temperatura (aproximadamente 2°C) entre as quatro regiões de estudo. Sendo que a estação seca nessas regiões tem ficado cada vez mais seca, mais quente e mais longa. Ao somar as emissões naturais e as de queimadas é visto que há mais CO₂ liberado do que a capacidade que a Floresta Amazônica tem de absorver ([GATTI et al.,](#)).

Um dos motivos da diminuição da absorção de CO₂ pode ser a diminuição da capacidade fotossintética e o comportamento dos elementos ambientais do ecossistema. A diminuição da disponibilidade de água e a variação anômala da temperatura pode ocasionar a diminuição da atividade fotossintética, devido à resposta fisiológica que promove o fechamento dos estômatos sob estresse hídrico ([BARKHORDARIAN; SAATCHI, 2019](#)).

A fotossíntese é restrita por processos físicos e bioquímicos e requer algumas reações que requerem luz e outras que não. Nas reações que incluem luz, a radiação absorvida pela clorofila, também chamada de PAR (radiação fotossinteticamente ativa), causa excitação dos elétrons que são transferidos por uma cadeia de moléculas de pigmento para centros de reação onde compostos de alta energia são formados, a água é dividida e O₂ liberado. A parte inicial é limitada pela irradiância e quantidade de clorofila presente nas folhas, e a taxa de elétrons é sensível à temperatura ([WARING; RUNNING, 2007](#)).

Na presença de luz, a fotorrespiração também ocorre no processo de geração do substrato ribulose bifsfosfato (RuBP) na fotossíntese, particularmente quando a proporção de O₂ para CO₂ aumenta dentro do cloroplasto. Compostos adicionais de alta energia são necessários para criar açúcares de seis carbonos e, como resultado, CO₂ adicional é respirado. As reações escuras são limitadas por CO₂ e temperatura e dependem da disponibilidade de nitrogênio e outros substratos ([WARING; RUNNING, 2007](#)).

Outra variável que ocasiona alteração na capacidade fotossintética da vegetação é a temperatura. Florestas tropicais como a Amazônia possuem aumento da capacidade fotossintética quando a temperatura do ambiente se encontra entre 30 e 35°C ([LLOYD et al., 1995](#)). No entanto, essa faixa de temperatura ótima pode mudar com

a alteração de níveis internos de CO₂ e eficiência da maquinaria fotossintética. Altas concentrações de CO₂ permitem que a temperatura ideal seja deslocada para cima, reduzindo a fotossíntese respiratória. Em temperaturas acima de 40°C, a fotossíntese bruta diminui abruptamente devido a alterações nas atividades dos cloroplastos e enzimas. A absorção de carbono pelas folhas geralmente aumenta gradualmente até uma temperatura ideal e depois diminui mais abruptamente à medida que o limite máximo de temperatura é atingido (WARING; RUNNING, 2007).

Nas componentes de respiração autotrófica e heterotrófica, a temperatura e a umidade são elementos importantes nas variações de emissões, onde no processo de respiração de manutenção é visto um aumento exponencial com a temperatura. A atividade metabólica de micro-organismos varia com a umidade e temperatura da superfície da serrapilheira e do solo no processo heterotrófico. A resposta à temperatura é geralmente exponencial, mas altamente variável. A atividade metabólica permanece estável até que o substrato esteja quase seco (WARING; RUNNING, 2007).

De acordo com Barkhordarian e Saatchi (2019) qualquer alteração na função fisiológica da floresta depende fortemente da disponibilidade de água disponível na atmosfera, e ela é impulsionada por dois fatores: o déficit de pressão de vapor (VPD, do inglês *Vapor Pressure Deficit*) e a radiação líquida. O VPD nos fornece uma síntese do comportamento da combinação entre umidade e temperatura do ar, se mostrando fundamental para o entendimento da evaporação, transpiração e perda de água no ambiente. Ele controla a demanda evaporativa, afetando diretamente a disponibilidade de água para a vegetação.

A radiação solar é essencial para o processo de fotossíntese e o aquecimento da superfície terrestre. No contexto das florestas do bioma Amazônia, tanto a radiação direta, que incide perpendicularmente sobre a superfície, quanto a radiação difusa, que é dispersa em várias direções na atmosfera, desempenham papéis cruciais. No caso dos dosséis florestais, onde a penetração da luz solar direta é limitada devido ao arranjo multidirecional das árvores e à frequente cobertura de nuvens, a radiação difusa torna-se particularmente significativa no processo de fotossíntese das vegetações nos dosséis. A radiação difusa permite uma distribuição mais uniforme da energia solar entre as camadas inferiores da floresta, garantindo assim a disponibilidade de luz necessária para o crescimento e desenvolvimento das plantas em diferentes níveis da floresta. (RODRIGUES; CIRINO, 2024) identificou que nas áreas de florestas semidecíduas no sul da Amazônia houve um aumento da capacidade de absorção de

CO₂ pelo ecossistema que se mostrou intimamente ligado à composição das diferentes vegetações do local, que consistem em espécies adaptadas à baixa luminosidade e são mais eficientes na captura da radiação difusa durante o processo de fotossíntese.

Nota-se pelo que foi dito anteriormente que existem diversas variáveis que podem impactar diretamente o balanço de CO₂. Algumas dessas variáveis e relações físicas serão avaliadas durante o desenvolvimento dos modelos de aprendizado de máquina, visando estimar a troca líquida de CO₂.

2.2 Troca líquida de carbono (NEE)

De acordo com Restaino (2013), o NEE (*Net ecosystem exchange*) é uma medida da diferença líquida entre os fluxos turbulentos ascendentes (respiração) e descendentes (GPP) de CO₂ na interface atmosfera-dossel, na camada limite atmosférica. O NEE tem uma definição semelhante à do NPP, mas adota a atmosfera como referencial ao invés da vegetação. Se o NEE for positivo, o ecossistema atua como uma fonte de CO₂, se for negativo o ecossistema atua como um sumidouro.

A medição da troca líquida de carbono em torres de fluxo ocorre a partir da técnica de covariância de vórtices turbulentos (*Eddy Covariance* em inglês, EC) (RESTREPO-COUBE et al., 2013). É uma técnica importante para entender o papel da biosfera terrestre nas mudanças climáticas globais e para avaliar a eficácia de estratégias de mitigação de emissões de CO₂ (BALDOCCHI, 2003).

A técnica de covariância de vórtices turbulentos fornece a medição direta do fluxo vertical de um escalar como o CO₂ entre a superfície e a atmosfera. Sendo assim, as torres de fluxo, que são responsáveis por esse tipo de medição, se tornaram pontos para validação de modelos e estimativas, além de se tornar um local de calibração de técnicas de sensoriamento remoto. De um modo geral, esse tipo de medição proporcionou a melhoria dos modelos dos ecossistemas (WARING; RUNNING, 2007). A técnica é capaz de medir a troca de CO₂ após distúrbios ambientais e pode ser usada para avaliação da idade e tipo funcional da vegetação local. Ela é utilizada em diferentes ecossistemas pelo mundo, o que permite a comparação entre eles (BALDOCCHI, 2003).

A teoria física por trás dessa técnica foi formulada por Osborn Reynolds e ela consiste na utilização das leis de conservação de massa, energia e momento na camada limite planetária (JACKSON; LAUNDER, 2007). A equação que define a conservação da massa utilizada nesta técnica é dada pela Equação 2.1 (TORO, 2018;

BITENCOURT, 2008; FOKEN et al., 2012):

$$\frac{\partial(\rho)}{\partial(t)} + u * \frac{\partial(\rho)}{\partial(x)} + v * \frac{\partial(\rho)}{\partial(y)} + \omega * \frac{\partial(\rho)}{\partial(z)} = S \quad (2.1)$$

Onde ρ é a concentração de CO₂, v e u são as componentes da velocidade do vento na direção x e y, ω é a componente da velocidade na direção z (normal da superfície) e S é o termo que representa a fonte/sumidouro de carbono. As medições que são realizadas pela técnica de covariância turbulenta são decompostas através da decomposição de Reynolds, na qual há uma componente média \bar{u} e uma componente turbulenta u' . Aplicando essa decomposição às velocidades nas Equações 2.2, 2.3 e 2.4 (FOKEN et al., 2012):

$$u = \bar{u} + u' \quad (2.2)$$

$$v = \bar{v} + v' \quad (2.3)$$

$$\omega = \bar{\omega} + \omega' \quad (2.4)$$

Considerando que não há fluxos horizontais divergentes de covariância turbulenta e integrando na direção z da superfície até a altura de medição h_m na Equação 2.5 (FOKEN et al., 2012).

$$\begin{array}{ccc} \text{I} & \text{II} & \text{III} \\ \int_0^{h_m} S dz = \overline{\omega' \rho'} + \int_0^{h_m} \overline{\frac{\partial(\rho)}{\partial(t)}} dz + \\ \text{IV} & \text{V} & \text{VI} \\ \int_0^{h_m} \bar{u} \overline{\frac{\partial(\rho)}{\partial(x)}} dz + \int_0^{h_m} \bar{v} \overline{\frac{\partial(\rho)}{\partial(y)}} dz + \int_0^{h_m} \bar{\omega} \overline{\frac{\partial(\rho)}{\partial(z)}} dz \end{array} \quad (2.5)$$

A componente I da Equação 2.5 se refere à magnitude da fonte ou sumidouro do escalar ρ , a II se refere à covariância turbulenta na camada que vai da superfície à altura de medição h_m , a componente III se refere ao armazenamento do escalar abaixo do nível de medição e as componentes IV, V e VI se referem aos fluxos de advecção nas direções x, y e z, respectivamente (FOKEN et al., 2012).

Na *Eddy Covariance*, os termos de advecção (IV, V, VI) podem ser considerados nulos em um dossel homogêneo e em um terreno plano, utilizando um procedimento

de rotação de coordenadas. Com isso, o cálculo do fluxo de carbono utilizando a técnica é dada pela Equação 2.6:

$$\int_0^{h_m} S dz = \overline{\omega' \rho'} + \int_0^{h_m} \frac{\partial(\rho)}{\partial(t)} dz \quad (2.6)$$

Ou de uma forma mais próxima de como é geralmente encontrado na literatura na forma da Equação 2.7 (ARAÚJO, 2010):

$$NEE = F_C + S_{CO_2} \quad (2.7)$$

Sendo que o fluxo de CO₂ F_C é medido pelo método de covariância de vórtices turbulentos e o termo de armazenamento no dossel S_{CO_2} é medido a partir do perfil vertical de concentração de CO₂ ao longo do dossel.

Em termos dos processos do ciclo do Carbono, o NEE permite entender os balanços diurnos e noturnos do ciclo. De acordo com [Waring e Running \(2007\)](#), os processos que ocorrem durante o dia são a produtividade primária bruta (GPP), a fotorrespiração (R_p), a respiração de manutenção (R_m), a respiração de síntese (R_s) de plantas autotróficas e a respiração heterotrófica (R_h) por animais e micróbios ([WARING; RUNNING, 2007](#)). O NEE diurno pode ser representado a partir da Equação 2.8 :

$$NEE_{diurno} = -GPP + R_p + R_m + R_s + R_h \quad (2.8)$$

Durante a noite, os termos fotossintéticos GPP e R_p não compõem o fluxo, já que a vegetação necessita de radiação solar para realização da fotossíntese. O NEE noturno pode ser representado pela Equação 2.9:

$$NEE_{noturno} = R_m + R_s + R_h = R_e \quad (2.9)$$

Onde R_e é a respiração total do ecossistema. Quando integrado ao longo de um dia, a convenção de sinais adotada indica que quando $NEE > 0$ $R > GPP$, de modo que as fontes de CO₂ superam os sumidouros, de modo que a superfície atua como uma fonte de CO₂ para a atmosfera, e vice-versa.

É importante ressaltar que a técnica de *Eddy Covariance* obtém valores de NEE, que correspondem ao fluxo líquido de CO₂. Para obter os fluxos de emissão (respiração) e de sumidouro (GPP) de CO₂ na atmosfera, é necessário realizar o particionamento do fluxo líquido NEE. Existem vários métodos de particionamento, sendo que um dos mais utilizados é estimar a respiração a partir dos dados noturnos de NEE, quando $GPP = 0$ (STOY et al., 2006).

Medidas de fluxo por *Eddy Covariance* na Amazônia são essenciais para o entendimento do papel da Amazônia no ciclo global do carbono, no entanto devido à extensão da Floresta Amazônica, não há disponibilidade de torres de fluxo suficiente para estudo do comportamento de todos os ecossistemas desse Bioma. Isso dificulta o entendimento das características regionais a partir dessas medições, o que torna necessário a realização de estimativas de NEE e de outros componentes do ciclo de carbono a partir de modelos físicos e biogeoquímicos (WARING; RUNNING, 2007), modelos de superfície e de vegetação (RESTREPO-COUBE; LEVINE, 2017) e modelos de aprendizado de máquina (TRAMONTANA et al., 2016; BAUER, 2022; JUNG et al., 2020; ZHURAVLEV et al., 2022). Sendo este último o focal principal deste estudo.

2.3 Modelos baseados em processos físicos e biogeoquímicos

De acordo com Jung et al. (2020), os modelos globais de vegetação dinâmica constituem uma abordagem *bottom-up* (isto é, de baixo para cima), para representar as trocas de carbono e sua evolução com mudanças nas condições ambientais, detalhando processos que ocorrem em uma escala espacial pequena e expandindo as estimativas para uma escala espacial mais ampla. Restrepo-Coupe e Levine (2017) encontrou dificuldades em simular a sazonalidade das trocas de carbono nas diversas regiões da Floresta ao utilizar modelos globais dinâmicos de vegetação global na região Amazônica. Os modelos mostraram limitações em representar a produtividade primária bruta e a capacidade fotossintética anual, além de superestimar a troca líquida de carbono e a respiração dos ecossistemas. A sazonalidade dos fluxos de carbono em florestas tropicais exige uma incorporação dos mecanismos biofísicos internos que controlam o fluxo de carbono nos modelos para melhores representações dos elementos biogeoquímicos das florestas (RESTREPO-COUBE; LEVINE, 2017)

De acordo com Waring e Running (2007), os ecossistemas são muito complexos para serem descritos por algumas equações; os modelos atuais de ecossistemas possuem centenas de equações que representam interações de formas não contínuas e não lineares. Um modelo biogeoquímico abrangente deve tratar todos os processos ecos-

sistêmicos e é essencial que a energia, o carbono, a água e os ciclos elementares sejam representados, mesmo que de maneira simples. São as interações entre os ciclos que gerem o ecossistema e elas sempre devem ser consideradas.

A dinâmica temporal dos processos ecossistêmicos também deve ser considerada, já que o balanço de energia, a quantidade de folhas e os fluxos gasosos, entre outros, variam em diversas escalas de tempo, muitas vezes com atrasos entre o instante de estímulo e o instante de resposta a uma perturbação. A dinâmica de alocação do carbono é observada em intervalos semanais a mensais, enquanto a mobilização de nutrientes pode ser mensurável sazonalmente (WARING; RUNNING, 2007; FISHER; KOVEN, 2020).

2.4 Modelos de aprendizado de máquina

Além dos modelos físicos, modelos estatísticos baseados em técnicas de inteligência artificial (IA) podem ser utilizadas para representar os processos do balanço de carbono. Nesta seção serão mostrados alguns trabalhos que aplicaram técnicas de aprendizado de máquina com este foco. Tramontana et al. (2016) realizaram um experimento de validação cruzada para prever as componentes do ciclo do carbono, NEE, GPP e Re, utilizando 11 técnicas de aprendizado de máquina, entre elas, a rede neural *multilayer perceptron* (MLP). O experimento foi realizado com dados de torres de fluxo do conjunto Fluxnet¹, dados sazonais provenientes do sensor MODIS e dados meteorológicos baseados em saídas de modelos numéricos diários como variáveis preditoras. A validação do experimento foi feita a partir de métricas de regressão e comparação entre os resultados do modelo e os valores medidos via torres de fluxo (TRAMONTANA et al., 2016).

As variáveis preditoras utilizadas na estimativa dos fluxos de carbono são variáveis que medem o comportamento físico e biogeoquímico da floresta em diferentes escalas (JUNG et al., 2009; TRAMONTANA et al., 2016; ZHURAVLEV et al., 2022; JUNG et al., 2020). As variáveis que são tipicamente utilizadas para os estudos do ciclo do carbono em escalas regionais e globais são o índice de vegetação da diferença normalizada (NDVI), o índice de área foliar (LAI), o índice de vegetação melhorado (EVI), a refletância bi direcionada, variáveis provenientes da atividade fotossintética APAR (radiação fotossinteticamente ativa absorvida) entre outras. Estas variáveis refletem, até certo ponto, a variabilidade das componentes do ciclo do carbono em escala regional.

¹<<https://fluxnet.org/data/fluxnet2015-dataset/>>

O uso de índices de vegetação para a estimativa de fluxo de carbono tem sido amplamente explorado por diversos pesquisadores, incluindo Xiao et al. (2004), Sims et al. (2008), Wu et al. (2012), Gitelson e Viña (2006). Esses estudos demonstraram que índices de vegetação, como o NDVI, EVI e índices relacionados à saúde da vegetação, juntamente com informações como temperatura da superfície terrestre (LST) e radiação fotossinteticamente ativa (PAR), podem servir como indicadores confiáveis para indicar estágios fenológicos das plantas, estresses no dossel vegetativo e condições ambientais.

Variáveis meteorológicas provenientes de modelos ou medições, como temperatura do ar, precipitação, temperatura do solo, índice normalizado de água (NDWI), deficit de pressão de vapor, entre outras são também incorporadas nas estimativas de fluxos de carbono. Essas variáveis desempenham um papel crucial no ciclo do carbono e, portanto, são consideradas importantes para a estimativa precisa dos fluxos. Essa abordagem é adotada tanto em modelos físicos quanto em modelos de aprendizado de máquina (TRAMONTANA et al., 2016; ZHURAVLEV et al., 2022; BAUER, 2022; JUNG et al., 2020), decorrente da influência dessas variáveis no processo de trocas líquidas de carbono entre a biosfera e a atmosfera.

No trabalho de Tramontana et al. (2016) foram utilizadas as técnicas *random forest*, *support vector machine* e a rede neural *multiLayer perceptron*, e obtiveram um resultado ruim, quando comparado a outros sítios, para a América do Sul. (JUNG et al., 2020) ao realizar a ampliação de escala (do inglês, *upscaling*) para todo o globo, obteve o melhor coeficiente de determinação de 0,33 para toda a região Amazônica utilizando os modelos criados pela Fluxcom².

2.4.1 Floresta aleatória (RF)

A Floresta Aleatória, conhecida como *Random Forest* (RF), é um algoritmo de aprendizado de máquina que se baseia no conceito de aprendizagem em conjunto. Ele é uma combinação de previsões de múltiplos algoritmos de aprendizado de máquina (árvores de decisão), de forma a obter previsões mais acuradas do que qualquer modelo individual.

A RF é formada pelo crescimento de árvores de decisão a partir de um vetor aleatório que representa uma combinação de características e subdivisões aleatórias do conjunto de dados durante o treinamento. A função preditora de uma árvore de regressão é representada por $h(x, \theta)$, onde x são os atributos de entrada e θ são os

²<https://www.fluxcom.org/CF-Products/>

parâmetros aleatórios que determinam como a árvore será construída (BREIMAN, 2001).

Esses parâmetros garantem que cada árvore na floresta seja construída de forma diferente, introduzindo diversidade e reduzindo a correlação entre as árvores na RF. Ele cria uma floresta de árvores de decisão durante o treinamento do modelo e faz previsões com base na votação das previsões individuais das árvores. A previsão das árvores é então agregada e, assim, obtido o valor final da previsão.

A RF é amplamente utilizada para resolução de problemas da área de sensoriamento remoto, como para previsão de fluxo de água (ISLAM, 2023), na estimativa de biomassa (WANG et al., 2016), mapeamento de áreas de desmatamento (VIEIRA et al., 2022), além de já ter sido utilizada também para estimativas fluxos de carbono pela FLUXCOM (TRAMONTANA et al., 2016).

2.4.2 Gradient boosting regressor (Gboost)

O algoritmo *Gradient Boosting* (GBoost) possui um funcionamento parecido com a Floresta Aleatória, pois ele cria múltiplos modelos simples e os combina para garantir melhor performance na estimativa. Ele constrói um modelo preditivo por meio da combinação de várias árvores de decisão simples.

O processo de treinamento inicia com uma única árvore de decisão que estima os valores. O modelo então é ajustado aos dados de treinamento, é realizado o cálculo dos resíduos e um novo estimador é treinado para prever os resíduos do modelo anterior. A estimativa da variável alvo é então atualizada com os resultados dos estimadores anteriores e esse fluxo se repete até que o número especificado de estimadores seja alcançado.

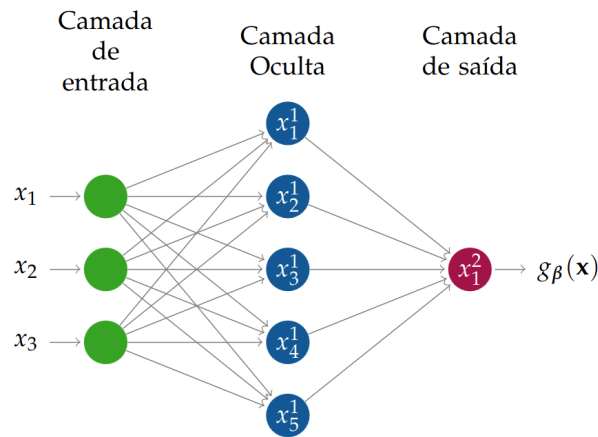
O GBoost é iterativo e sequencial, de modo com que cada novo estimador é treinado para corrigir os erros do modelo anterior, o que torna o modelo robusto e preciso. Ele já foi utilizado para estimativa da produção primária bruta por Zhou et al. (2023), para previsão de emissões de carbono (CUI et al., 2021), previsão de NEE (CAI et al., 2020), entre outros problemas de estimativa que envolvem o ciclo do carbono.

2.4.3 MultiLayer perceptron

A *multilayer perceptron* (MLP) é uma rede neural artificial que possui camadas de entrada, camadas ocultas e camadas de saída, contendo múltiplos neurônios, como pode ser visto na Figura 2.2. Cada neurônio de uma camada está conectado a um

neurônio da camada posterior e assim sucessivamente. As conexões entre os neurônios possuem pesos associadas a elas e esses são ajustados e modificados durante o processo de treinamento da rede.

Figura 2.2 - Representação da arquitetura geral de uma MLP.



Fonte: Izbicki e Santos (2020).

Durante o processo de treinamento a MLP possui hiper parâmetros que são utilizados para um melhor ajuste da rede aos dados de treinamento e para aprendizado de comportamento e parametrização desses dados. Dentre eles estão a função de ativação, a função de otimização, o número de camadas, o número de neurônios, entre outros. A rede neural MLP é amplamente utilizada na estimativa da troca líquida de carbono, seja a partir de preditores das torres de fluxo ou a partir de preditores de dados de sensoriamento remoto (TRAMONTANA et al., 2016; BAUER, 2022; ZHURAVLEV et al., 2022).

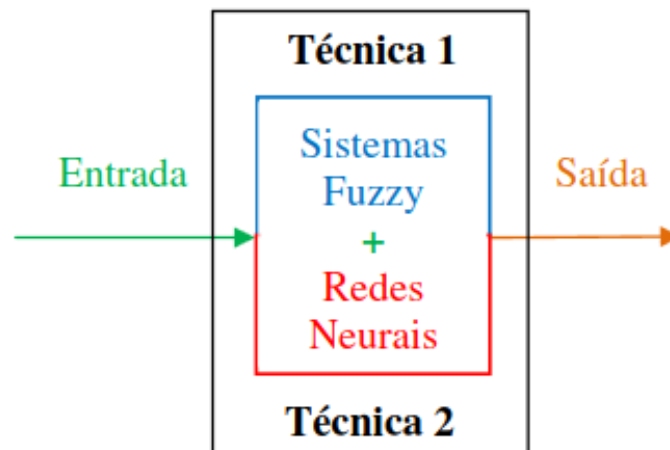
2.4.4 Sistema adaptativo de inferência neuro-fuzzy (ANFIS)

O sistema Neuro-Fuzzy ANFIS (Sistema Adaptativo de Inferência Neuro-Fuzzy) é usado para aplicações de previsão e aproximação de funções. Ela permite a extração de conhecimento na forma de regras de inferência *fuzzy* com a integração do conhecimento da experiência do especialista e o conhecimento implícito a partir do conjunto de dados. Dou et al. (2018) verificaram em seu trabalho que a rede ANFIS apresentou resultados promissores e foi considerada uma abordagem viável para estimar fluxos diários de carbono a partir de dados de covariância turbulenta. Seus

resultados são comparáveis em termos de precisão aos métodos tradicionais ANN (Redes neurais artificiais) e SVM (*Support Vector Machine*).

A ANFIS é uma técnica híbrida que combina o aprendizado de máquina com a lógica *fuzzy*, como pode ser visto na Figura 2.3, para modelagem de sistemas complexos.

Figura 2.3 - Demonstração da técnica híbrida: Lógica Fuzzy + Redes Neurais.



Representação da organização híbrida da rede neural ANFIS.

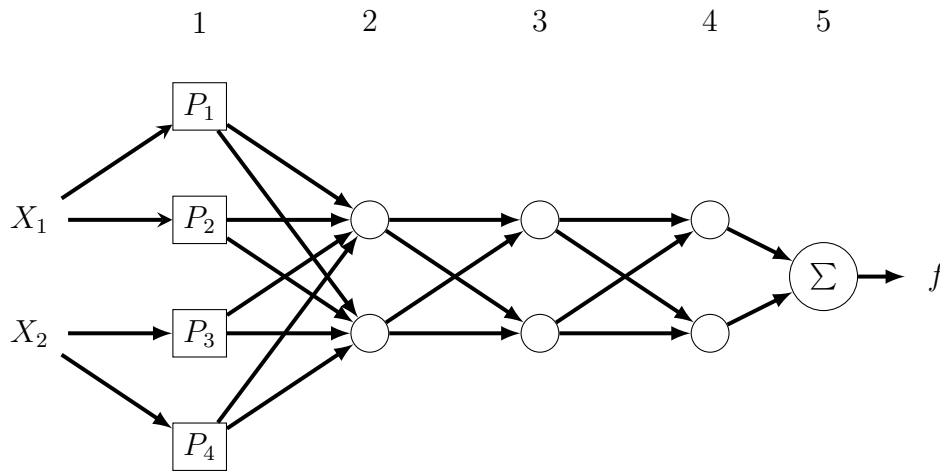
Fonte: Sanchez (2009).

A lógica Fuzzy é uma extensão da lógica booleana. Na lógica booleana, um elemento pertence ou não a um conjunto de forma binária, ou seja, ou está verdadeiramente dentro ou fora do conjunto. Já na lógica fuzzy, os elementos podem ter grau de pertinência em relação a um determinado conjunto, ou seja, há uma medida contínua do quanto um elemento pertence a um conjunto (SANCHEZ, 2009). Por exemplo, pode-se definir um conjunto de valores de um determinado preditor 'X alto', sendo que a função de pertinência para este conjunto pode ser uma curva triangular, onde os valores de X próximos ao pico têm uma pertinência alta, enquanto os valores mais distantes do pico têm uma pertinência menor.

A ANFIS é composta por cinco camadas (SANCHEZ, 2009; JANG, 1993):

1. Camada de entrada: Recebe as variáveis de entrada e as encaminha às funções de pertinência.
2. Camada de pertinência: atribui graus de pertinência a cada um dos valores de

Figura 2.4 - Arquitetura da Rede Neural ANFIS.



entrada, convertendo-os em valores fuzzy.

3. Camada de regras: Combina as pertinências das entradas para gerar as regras *if-then* do modelo.
4. Camada de consequência: Calcula as contribuições da camada anterior para a saída do modelo.
5. Camada de saída: Soma todas as contribuições da camada anterior para gerar a saída final do modelo.

Cada camada tem um conjunto específico de nós e funções que realizam diferentes cálculos durante o processo de inferência fuzzy. Durante o treinamento da rede, há o ajuste dos parâmetros para minimizar a diferença entre as saídas do modelo e as saídas desejadas. Na Figura 2.4 se encontra um exemplo da arquitetura da rede ANFIS.

O treinamento da ANFIS possui duas etapas: a primeira inicializa a partir de um conhecimento prévio sobre o conjunto de padrões e depois é refinada por aprendizado acrescentado, diminuindo regras. Também pode ser inicializada com um conjunto vazio de regras e acrescida por aprendizado incremental (SANCHEZ, 2009). A segunda etapa ocorre após a criação da base de regras, onde se executa o algoritmo de aprendizado baseado em gradiente descendente para ajustar os parâmetros das funções de pertinências dos antecedentes das regras (SANCHEZ, 2009).

A ANFIS é uma técnica que já foi utilizada dentro da área de pesquisa desta proposta (DOU et al., 2018; WOOD, 2021). No entanto, como foi dito, é preciso inovar para investigar outras técnicas que possam melhor representar os processos físicos a partir das relações não lineares entre as variáveis a serem estudadas. Neste sentido, a seguir é apresentada uma rede neural mais atual e com diversos usos em ciências de dados, a GNN (Rede Neural de Grafos).

2.4.5 Rede neural de grafos

De acordo com Giovanni et al. (2023), as redes neurais de grafos (GNN, do inglês *graph neural network*) são uma classe de modelos de aprendizado de máquina profundo projetados para lidar com dados estruturados em forma de grafo. Eles são capazes de modelar interações complexas entre elementos do grafo, como nós e arestas, e podem ser usados para prever propriedades ou comportamentos dos elementos do grafo (JIANG, 2022).

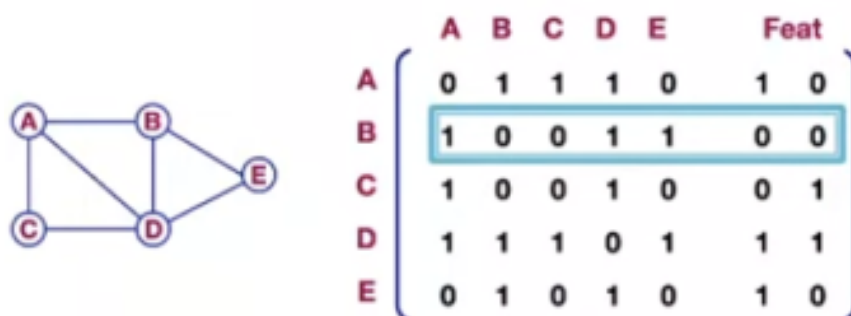
Segundo Sanchez-Gonzalez et al. (2020) os grafos conseguem representar o estado de um sistema físico, eles permitem uma maneira natural e eficiente de modelar as interações complexas entre as partículas que compõem o sistema. De acordo ele, cada partícula pode ser representada como um nó no grafo, e as arestas do grafo representam as interações entre as partículas. Isso permite a captura da estrutura complexa do sistema físico e modelar sua dinâmica com alta precisão.

Os componentes de um grafo podem ser representados como uma função de vértices e arestas. Onde G é a representação do grafo, V são os vértices e E são as arestas. Os vértices e as arestas possuem características, X_v e X_e respectivamente, e essas podem ser incluídas na representação do grafo com variação temporal em 2.10.

$$G = (V, E, X_{v(t)}, X_{e(t)}) \quad (2.10)$$

Essa é uma das estruturas utilizadas como entrada em redes neurais de grafos que utilizam dados temporais (WU et al., 2021). O grafo pode ser representado por nós e arestas, contudo a representação computacional da estrutura de dados para grafos pode ser uma matriz ou lista de adjacência, uma matriz de incidência, uma matriz de peso e outras representações mais complexas para definir as relações dos nós e arestas (LESKOVEC, 2021). Essas representações se encontram na Figura 2.5 .

Figura 2.5 - Representação do grafo e sua matriz de adjacência.



À esquerda há a representação do grafo a partir de seus nós e arestas. À direita há a representação da sua matriz de adjacência com a sua matriz de características dos nós.

Fonte: Leskovec (2021).

A matriz de adjacência do grafo $G(V,E)$ é construída pela seguinte regra:

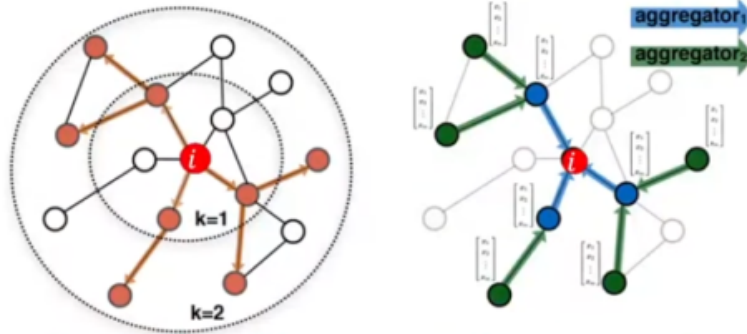
$$a_{ij} = \begin{cases} 1, & \text{se existe } (i,j) \in E \\ 0, & \text{caso contrário} \end{cases} \quad (2.11)$$

Ou seja, o valor 1 na matriz representa a relação entre os nós, ou seja, representam a presença de arestas. Quando não há relação entre os nós, o valor do elemento da matriz é zero.

A aprendizagem da GNN ocorre a partir da propagação de informações ao decorrer do grafo para calcular as características dos nós. Sendo que quando utilizada para predição permite a predição de nós, arestas ou outros grafos. Por exemplo, na Figura 2.6, é apresentada a representação do grafo, incluindo seus nós e arestas, juntamente com o nó a ser predito e as características dos nós.

A *Graph Convolutional Networks* (GCN) (KIPF; WELLING, 2017) e a *Graph Attention Networks* (GAT) (VELIČKOVIĆ et al., 2018) são duas redes neurais de grafos que são projetadas especificamente para processar e aprender a partir de grafos. A primeira considera a estrutura do grafo para realizar operações de convolução e a segunda operações de atenção nos vértices e arestas do grafo. Elas são capazes de capturar e aprender representações e informações de nós em um grafo, e as relações desses nós com os outros nós do grafo. A GAT é potencialmente mais expressiva em

Figura 2.6 - Representação do grafo de entrada de uma rede neural à esquerda e representação da propagação e transformação de informação no grafo à direita.



À esquerda, temos a representação gráfica do grafo, exibindo seus nós e arestas. À direita, ilustra-se a propagação e transformação de informações no grafo. O nó com o valor 'i' representa o nó a ser predito, enquanto os vetores adjacentes a cada nó representam suas características. A propagação e transformação de informações são realizadas por meio da agregação das características dos nós e de seus vizinhos.

Fonte: Leskovec (2021).

capturar relações complexas nos grafos do que a GCN

A GCN (Graph Convolutional Network) utiliza uma técnica de representação de nós enriquecida, combinando características do nó com as de seus vizinhos através de uma operação de convolução. Essa convolução é realizada utilizando uma matriz de peso única compartilhada entre os nós, e o resultado é normalizado com base no grau dos nós para garantir uma faixa de valores semelhante e comparabilidade entre eles. Cada camada GCN na arquitetura da rede neural de grafos realiza uma convolução em um grau mais alto, considerando vizinhos de primeiro grau na primeira camada e, ao adicionar camadas GCN, incluindo vizinhos de graus subsequentes. Esse empilhamento de camadas pode levar ao fenômeno de suavização excessiva, onde todas as incorporações se tornam semelhantes, se forem usadas muitas camadas.

Durante a propagação de informações, as representações dos nós são atualizadas com informações agregadas dos vértices vizinhos, geralmente por meio de média ou soma ponderada das representações dos vértices vizinhos. A matriz de adjacência é utilizada para determinar quais vértices estão conectados aos vértices de destino e seus pesos de conexão, permitindo que o modelo leve em consideração a estrutura do grafo ao propagar informações. Essas operações são fundamentais para a capacidade da GCN de processar dados estruturados em forma de grafo

A GAT (Graph Attention Network) difere da GCN ao atribuir pesos dinâmicos às características dos nós por meio de um processo de auto atenção. Durante esse processo, a GAT calcula a importância das características dos nós vizinhos para um nó específico, permitindo que alguns vizinhos sejam considerados mais importantes do que outros, independentemente de seus graus de nó. Essa abordagem é realizada através do cálculo de coeficientes de atenção, que são então utilizados para ponderar a incorporação do nó alvo.

Os coeficientes de atenção são aprendidos durante o treinamento da rede neural, e a técnica de multi-head attention é empregada para permitir que o modelo aprenda múltiplas representações ponderadas dos vizinhos de cada nó em um grafo. Em vez de depender de uma única função de atenção, o GAT utiliza várias cabeças de atenção para calcular a importância dos nós vizinhos de maneiras diferentes, permitindo ao modelo capturar nuances e relações complexas nos dados. As saídas das múltiplas cabeças de atenção são então combinadas através de concatenação ou média, dependendo da camada em questão (BRODY et al., 2022).

2.5 Ampliação de escala (do inglês, *Upscaling*)

Geralmente aplica-se a ampliação de escala para ampliar a representatividade espacial de observações pontuais de fluxo de carbono, estimando valores em locais que não possuem medições. Jung et al. (2020) utilizou, na iniciativa Fluxcom, a ampliação de escala estatística. Essa técnica utiliza dados observacionais para estimar valores em locais onde não há medições diretas. A abordagem é baseada na suposição de que os padrões observados em um local podem ser extrapolados para outros locais com características ou processos físicos semelhantes. A estimativa de fluxos de carbono em escala global é uma poderosa fonte de dados que possibilita o estudo da variabilidade do ciclo global de carbono. Essas estimativas podem ajudar a entender como os ecossistemas terrestres respondem às mudanças climáticas e como isso afeta o clima global. Essas informações são importantes no desenvolvimento de políticas públicas relacionadas à mitigação de mudanças climáticas (JUNG et al., 2009).

O aumento da escala das medições locais de covariância turbulenta para escalas regionais e globais utiliza variáveis preditoras disponíveis para estas escalas, como dados de satélites de sensoriamento remoto e dados meteorológicos provenientes de modelos numéricos (WARING; RUNNING, 2007). O conjunto de variáveis preditoras é utilizado para estimar fluxos em grade, aplicando modelos de aprendizado de máquina em cada ponto de grade. Com o treinamento e validação de uma ANN que utiliza dados observacionais para estimativas, é possível a partir da mesma,

realizar estimativas em locais onde não há disponibilidade de dados de covariância turbulenta (JUNG et al., 2020).

Uma vez que o ciclo do carbono na Amazônia é complexo e envolve a entrada e saída de carbono em diferentes formas e processos e pode não ser o mesmo para toda a área de cobertura da floresta, é necessário a utilização de técnicas como a utilizada por Jung et al. (2020) para o entendimento espacial do ciclo na floresta como um todo.

A Amazônia é um ecossistema complexo e diverso, com ambientes aquáticos e terrestres interconectados, o que torna importante entender como esses ciclos funcionam na floresta em diferentes escalas espaciais e temporais. Compreender esses processos é fundamental para a conservação da biodiversidade e dos serviços ecossistêmicos da Amazônia, bem como para a mitigação das mudanças climáticas globais. Logo, espera-se com esse estudo trazer mais informação sobre a espacialidade do comportamento de NEE para toda a região amazônica a partir da sua ampliação de escala utilizando técnicas de aprendizado de máquina e dados satelitais e meteorológicos.

3 DADOS

Os dados descritos a seguir foram utilizados e co-localizados em uma grade regular de resolução espacial de $0,25^\circ$, delimitada pelas extremidades da borda georreferenciada do Bioma Amazônia (INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (IBGE), 2019) no Brasil. A seleção dos dados foi baseada nos experimentos da FLUXCOM, que estimativa de NEE na escala global (TRAMONTANA et al., 2016; JUNG et al., 2020). Além das variáveis preditoras utilizadas no FLUXCOM, foram escolhidas outras variáveis incluídas no referencial teórico deste trabalho, que se mostraram importantes no comportamento biogeoquímico dos ecossistemas. A variável alvo da estimativa é a troca líquida de carbono (NEE), enquanto as variáveis meteorológicas e de superfície são utilizadas como preditoras de NEE, separadas em três conjuntos de dados: o RS (dados de sensoriamento remoto), o METEO (dados meteorológicos) e o METEORS (junção dos dois conjuntos anteriores), conforme Tramontana et al. (2016).

3.1 Área de estudo: Amazônia Legal

Segundo o INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (IBGE) (2019) a Amazônia ocupa cerca de 49,3% do território brasileiro, uma área aproximada de 4.196.943 km², abrangendo os Estados do Acre, Amapá, Amazonas, Pará, Rondônia, Roraima e parte dos estados do Maranhão, Tocantins e Mato Grosso, na Figura 3.1 há a delimitação do Bioma.

A Amazônia possui a maior floresta tropical do mundo, equivalente a 1/3 das reservas de florestas tropicais úmidas que abrigam a maior quantidade de espécies da flora e da fauna, além de conter 20% da disponibilidade mundial de água doce e grandes reservas naturais (INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (IBGE), 2019).

Figura 3.1 - Delimitação do Bioma Amazônia.

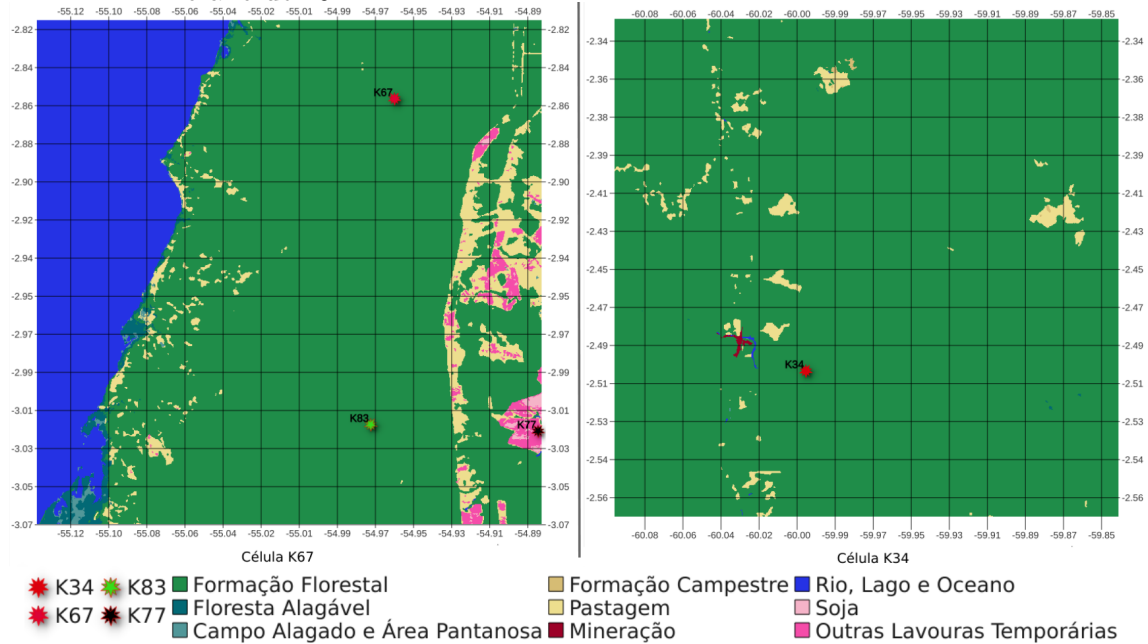


(em branco) . no quadrado vermelho há a localização da torre de fluxo utilizada como validação neste trabalho, a Sa1 e a Sa3 do conjunto Fluxnet.

Fonte: Adaptada de [INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA \(IBGE\) \(2019\)](#).

A Amazônia apresenta uma significativa heterogeneidade espacial e sazonal da pluviosidade, sendo a região brasileira com maior total pluviométrico anual. Os máximos de chuva na região ocorrem no verão austral. No litoral do Amapá, na foz do rio Amazonas e no setor ocidental da região, a precipitação excede 5.000 mm/ano. No extremo norte da Amazônia, o período mais seco ocorre no trimestre setembro-outubro-novembro, à oeste da região nordeste ocorre em agosto-setembro-outubro, no Vale da Bacia Amazônica em julho-agosto-setembro e à oeste no trimestre junho-julho-agosto ([CAVALCANTI, 2009](#)).

Figura 3.2 - Mapeamento Mapbiomas para as células K67 (à esquerda) e K34 (à direita) no ano de 2011.



Para compreender as disparidades entre as células e o comportamento de suas variáveis, foi utilizado o mapeamento de uso e cobertura da terra fornecido pelo Mapbiomas (8ª Coleção) (SOUZA et al., 2020). Esse mapeamento revelou as classes presentes em cada célula, oferecendo informações sobre a composição da paisagem nas áreas utilizadas no processo de modelagem da estimativa de NEE. Os mapeamentos individuais de cada célula podem ser observados na Figura 3.2 .

Na Célula K67, destacada à esquerda na Figura 3.2, observa-se uma predominância das classes de formação florestal, com a presença também da classe de corpos d'água, como rios, lagos e oceano. Nas proximidades do rio Tapajós, identificam-se áreas de pastagem e outras culturas temporárias. As torres K67 e K83 estão situadas em uma região de formação florestal, enquanto a torre K77 está em uma área classificada como outras lavouras temporárias. Já na Célula K34, à direita na Figura 3.2, a formação florestal predomina, com algumas áreas pequenas e dispersas de pastagem. A torre K34 está localizada em uma região de formação florestal, próxima a uma área de mineração, pastagem e à margem de um rio.

Ao comparar os mapeamentos entre os anos de 2002 e 2011, não foram observadas grandes mudanças na cobertura de uso da terra em áreas extensas das células. Identificou-se apenas um aumento em pequenas áreas de pastagem em ambas as

regiões. Essas informações são cruciais no entendimento do comportamento das variáveis preditoras das células. No entanto, é importante lembrar que o mapeamento do Mapbiomas é realizado utilizando técnicas de aprendizado de máquina, o que pode resultar em classificações incorretas em áreas pequenas (SOUZA et al., 2020).

3.2 Covariância de vórtices turbulentos

Os dados de covariância de vórtices turbulentos utilizados neste trabalho se encontram disponíveis em duas bases de dados: a Fluxnet¹ e o LBA². Esses dois conjuntos estão inter-relacionados, pois os dados do Brasil na base global do Fluxnet originam-se da iniciativa brasileira LBA. No entanto, devido aos diferentes objetivos das duas iniciativas em termos de escala, cada uma aplicou diferentes padronizações de correção e ajustes em seus respectivos dados. Para compreender as diferenças entre esses conjuntos de dados, foram calculadas métricas como média, mínimo, máximo, mediana e desvio padrão para todas as variáveis de NEE do Fluxnet e do LBA. Isso permitiu identificar as variáveis mais semelhantes entre os conjuntos, facilitando a comparação entre eles, além de oferecer uma perspectiva sobre quais variáveis são mais recomendadas para uso neste trabalho.

Na Tabela 3.1 há a descrição geral das variáveis de NEE que foram utilizadas neste estudo. Os atributos escolhidos como variável alvo foram o NEE_VUT_REF e o NEEenogap_5day_sco2_ust. O NEE_VUT_REF é recomendado para uso pela Fluxnet como variável referência, por exibir a melhor estimativa para o valor de NEE entre todos os estimados pela iniciativa (PASTORELLO et al.,). Seu indicador de qualidade, o NEE_VUT_REF_QC, varia de 0 a 1 e representa a fração de qualidade dos dados. Quanto mais próximo de 1, melhor a qualidade; quanto mais próximo de 0, pior a qualidade.. Esse indicador de qualidade é essencial para avaliar a confiabilidade dos dados de fluxo de carbono estimados através de preenchimento de lacunas, fornecendo uma medida da integridade dos dados de preenchimento de lacunas usados para estimar o NEE entre o ecossistema e a atmosfera.

O atributo NEEenogap_5day_sco2_ust foi selecionado com base nas recomendações do estudo de Hutya et al. (2008), que sugerem filtrar períodos de baixa turbulência e considerar o armazenamento de CO₂ abaixo do dossel. Além disso, essa variável tem sido empregada em trabalhos recentes de estimativa de NEE (BAUER, 2022). Para o entendimento do impacto das escolhas das variáveis alvo neste trabalho, foram calculados os bias, o coeficiente de determinação, o erro absoluto e erro quadrado

¹<<https://fluxnet.org/data/fluxnet2015-dataset/>>

²<https://daac.ornl.gov/LBA/guides/CD32_Fluxes_Brazil.html>

médio para cada uma das séries temporais das variáveis dos dois conjuntos.

Tabela 3.1 - Dados de Fluxos Turbulentos de CO₂.

Dataset	Fluxnet	LBA			
Atributo	NEE_VUT_REF	NEEnogap_5day_sco2_ust			
Unidade	$gC.m^{-2}.d^{-1}$	$\mu molCO_2.m^{-2}.s^{-1}$			
Sítio	BrSa1	K67	K77	K83	K34
Lat	-2,86	-2,86	-3,02	-3,02	-2,50
Lon	-54,96	-54,96	-54,89	-54,97	-60,00
Série Temporal	2002-2011	2002-2006	2002-2005	2002-2004	2002-2006
Resolução Temporal	diária				
Estado	Pará				Amazonas
Cidade	Santarém				Manaus
Local	Floresta Nacional do Tapajós				Reserva Biológica do Cuieiras
Resolução Espacial	in situ				

Fonte: Fluxnet (2023), Restrepo-Coupe e Rocha (2021a).

Na Tabela 3.1, observa-se que a unidade de NEE no conjunto de dados Fluxnet difere daquela no conjunto de dados LBA. Para garantir consistência, foi realizada a conversão para a unidade utilizada no conjunto de dados LBA, que é mais comum em estudos de fluxo no Brasil. A documentação fornecida por Fluxnet (2023) explica detalhadamente como realizar essa conversão, quando necessário. Assim, seguindo as regras especificadas na documentação, a conversão foi realizada utilizando a seguinte Equação 3.1:

$$\frac{gC}{m^2.d} = \frac{gC}{m^2.d} \times \left(\frac{1}{3600} \text{ hr/sec} \right) \times \left(\frac{1}{24} \text{ day/hr} \right) \times \left(\frac{1}{12.011} \text{ mol/gC} \right) \times \left(\frac{1}{10^{-6}} \mu\text{mol/mol} \right) = 1.0377504 \frac{\mu\text{molC}}{m^2.s} \quad (3.1)$$

Além disso, foram calculadas outras métricas, como a correlação de Pearson (R), erro quadrático médio (RMSE), erro médio absoluto (MAE) e viés (BIAS), para complementar a análise e identificar possíveis associações entre as variáveis. Também

foram avaliadas a correlação, o RMSE, o MAE e o viés entre as séries temporais, proporcionando uma visão abrangente da relação entre os conjuntos de dados. As equações das métricas ditas aqui se encontram na Seção 4.

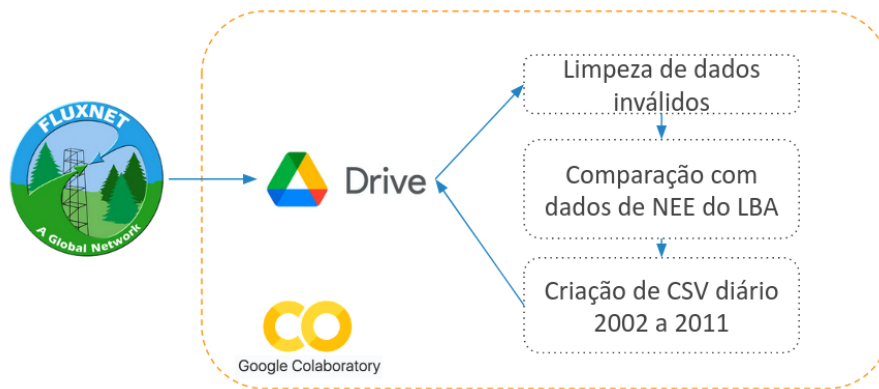
O Fluxnet é uma rede global de torres de fluxo que medem as trocas de energia, água e gases entre a superfície da Terra e a atmosfera. Essas torres são equipadas com sensores que medem continuamente as trocas de CO₂, vapor d'água, calor e outros gases entre o ecossistema abaixo da torre e a atmosfera acima dela (FLUXNET, 2023; JUNG et al., 2009). O objetivo dessa iniciativa é fornecer dados sobre as trocas entre a superfície da Terra e atmosfera em diferentes tipos de ecossistemas em todo o mundo. Os dados coletados pelas torres são usados para entender melhor os processos biogeoquímicos que ocorrem nos ecossistemas terrestres e como eles respondem às mudanças ambientais, como o aumento das concentrações de CO₂ na atmosfera e as mudanças no clima (FLUXNET, 2023; JUNG et al., 2009).

Os dados brutos coletados pelas torres são enviados para o ONEFlux, onde são processados e disponibilizados para pesquisadores de todo o mundo, permitindo o uso para análises detalhadas dos processos biogeoquímicos em diferentes tipos de ecossistemas globalmente. As etapas realizadas no ONEFlux são (PASTORELLO et al.,):

- a) Controle de qualidade dos dados brutos.
- b) Preenchimento de lacunas nos dados.
- c) Cálculo dos fluxos de CO₂, água e energia.
- d) Particionamento dos fluxos de CO₂ e componentes do ciclo do carbono.
- e) Cálculo das incertezas.
- f) Padronização dos dados.

Como mencionado anteriormente, os dados do Brasil incluídos no Fluxnet provêm das torres de medição da iniciativa *Large Scale Biosphere-Atmosphere Experiment in Amazonia* (LBA), que possui um total de nove torres de fluxo no território brasileiro. O Fluxnet incorporou em seu conjunto os dados das torres K67 e K83 da base de dados LBA. Os dados da Torre K67 foram renomeadas para BrSa1 e da torre K83 para BRSa3 no conjunto Fluxnet. Neste trabalho foram utilizados os dados da torre BrSa1, escolhida devido ao grande período de dados disponíveis, totalizando 9 anos. As características da torre BrSa1 podem ser vistas na Tabela 3.1:

Figura 3.3 - Fluxograma Fluxnet.

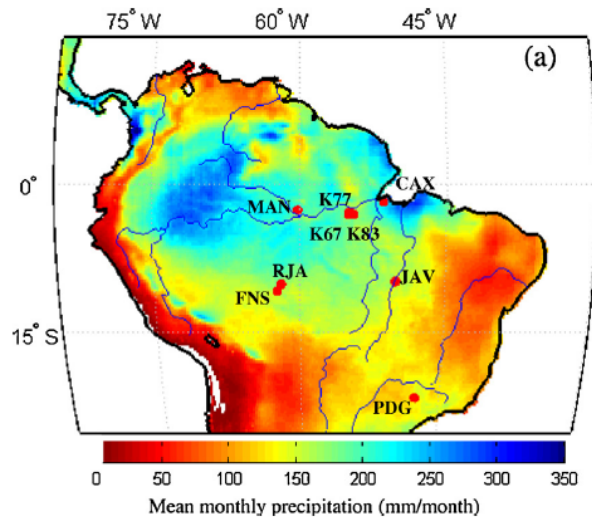


A extração dos dados Fluxnet foi realizada a partir do site Fluxnet (2023). Foi feita a limpeza de valores inválidos e a comparação com os dados de torres de fluxo no LBA, de acordo com a Figura 3.3. O atributo NEE_VUT_REF foi utilizado como variável alvo no treinamento, validação e teste nas estimativas pelos algoritmos de aprendizado de máquina usados neste trabalho.

O experimento de grande escala da biosfera-atmosfera na Amazônia (Large Scale Biosphere-Atmosphere Experiment in Amazonia, LBA) é uma iniciativa internacional de pesquisa que possui como objetivo gerar novos conhecimentos sobre o funcionamento climatológico, ecológico, biogeoquímico e hidrológico da Amazônia (LARGE SCALE BIOSPHERE ATMOSPHERE EXPERIMENT IN AMAZONIA LBA, 2006). O LBA utilizou técnicas intensivas de sensoriamento remoto e experimentos de medições terrestres para investigar a dinâmica atmosfera-biosfera-hidrosfera. Com isso, houve a criação de diversos conjuntos de dados do LBA disponibilizados publicamente que possibilitam o estudo dos ciclos biogeoquímicos da Floresta (LARGE SCALE BIOSPHERE ATMOSPHERE EXPERIMENT IN AMAZONIA LBA, 2006).

O conjunto de dados 'LBA-ECO CD-32 Flux Tower Network Data Compilation, Brazilian Amazon: 1999-2006, V2' possui uma compilação de dados de fluxos de covariância turbulenta de carbono e energia, de meteorologia, radiação, temperatura do dossel, umidade do solo e perfis de CO₂ que são coletados em nove torres da Região Amazônica. Os dados estão disponíveis no período de 1999 a 2006, sendo que a cobertura de dados nesse período varia entre as torres (RESTREPO-COUCPE; ROCHA, 2021b). A localização e nomenclatura de cada uma das torres se encontram na Figura 3.4. A resolução temporal disponibilizada por esse conjunto de dados é

Figura 3.4 - Sítios de torres de fluxo da iniciativa LBA.



Na figura há parte da América do Sul, o mapa é de precipitação média, e os pontos vermelhos são a indicações dos locais que há cada uma ds torres do projeto LBA, as torres utilizadas neste trabalho são as K67, K83, K77 e MAN, elas se encontram no Bioma Amazônia.

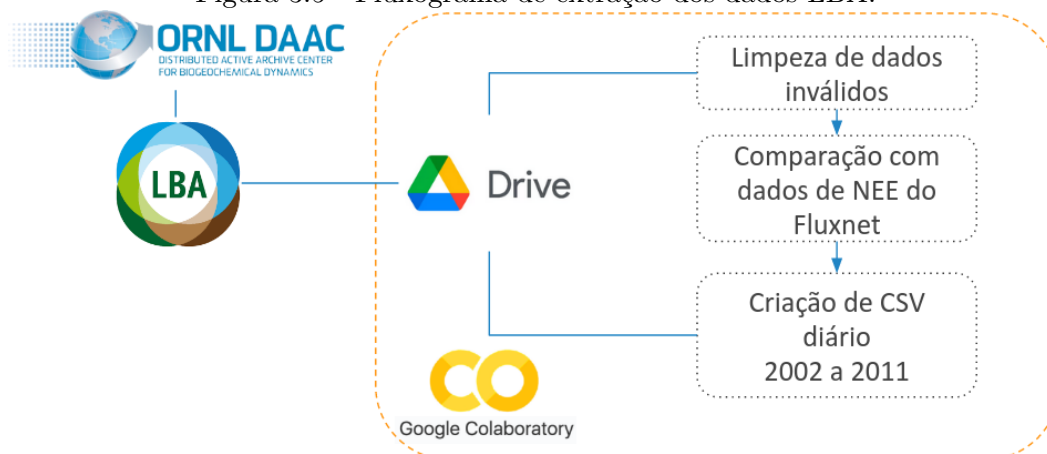
Fonte: Adaptado de Restrepo-Coupe e Rocha (2021b).

horária, diária, a cada 16 dias e mensais. Neste trabalho foram usados os dados diários.

Os dados do LBA que são descritos na Tabela 3.1 foram utilizados nas etapas de treinamento, validação e teste dos algoritmos de aprendizado de máquina e no teste da ampliação de escala a partir da comparação entre dados observados e preditos para as células que possuem torres. Na etapa de intercomparação dos dados meteorológicos foram utilizados também dados das torres K34 e K67, especificamente o atributo de temperatura do ar (t_a) das torres. A extração dos dados LBA foi realizada de forma semelhante à extração de dados do Fluxnet, de acordo com a Figura 3.5

De acordo com Restrepo-Coupe et al. (2013), os dados de NEE do conjunto LBA diários foram criados a partir de dados fornecidos a cada meia hora, foram calculados então o NEE a partir da soma do fluxo turbulento acima da floresta com a taxa de variação temporal de CO_2 armazenado dentro da copa (SCO_2). Em alguns locais, devido à ausência de medidas de perfis verticais de CO_2 , não foi possível a medição de SCO_2 e a partir disso foram utilizados métodos para a estimativa desse fluxo

Figura 3.5 - Fluxograma de extração dos dados LBA.



de armazenamento . No preenchimento de dados faltantes, foram calculadas os ciclos anuais médios e adicionados à série temporal (RESTREPO-COUCPE; LEVINE, 2017).

As quatro torres do LBA utilizadas neste estudo se encontram em áreas de floresta no bioma Amazônia. As de Santarém se encontram na Unidade de Conservação Federal, na Floresta Nacional do Tapajós. Contudo, como foi visto na Figura 3.2 em seus arredores há diferentes distúrbios antrópicos. A Torre K77 se encontra em um campo agrícola, que une a atividade de pasto e agricultura (SAKAI et al., 2004), a torre K83 está em uma área de exploração seletiva de madeira (ROCHA, 2004; MILLER et al., 2004; GOULDEN et al., 2004) e a Torre K67 está situada em uma área de floresta primária (HUTYRA et al., 2007; SALESKA et al., 2003). A Torre K34 está situada na Reserva Biológica do Cuieiras do Instituto Nacional de Pesquisas da Amazônia e está localizada a aproximadamente 60km a noroeste da capital do Amazonas, Manaus, em uma área de floresta primária (ARAÚJO et al., 2002).

3.3 Sensoriamento remoto (RS)

Os atributos de índice de vegetação, que se relacionam com o comportamento fotosintético da vegetação, advieram de produtos dos sensores MODIS (Espectrorradiômetro imageador de resolução moderada) que se encontram a bordo dos satélites Terra e Aqua. Índices de vegetação geralmente são utilizados para a compreensão da dinâmica global da vegetação, sendo em escala local ou regional. Também são utilizados no desenvolvimento de modelos físicos e de aprendizado de máquina globais para fornecer aos modelos características da superfície, principalmente os diferentes

tipos funcionais de plantas (JUNG et al., 2009). Os atributos MODIS utilizados neste trabalho e suas características se encontram na Tabela 3.2.

Tabela 3.2 - Produtos do MODIS e suas variáveis.

MODIS	MOD11A2.061	MOD13A2.061					MCD15A3H.061	
Atributo	LST_day	NDVI	EVI	b02	b03	b07	Fpar	Lai
λ (nm)				841 a 876	459 a 479	2.105 a 2.155	400 a 700	
QA Flag	QC_Day	DetailedQA, SummaryQA					FparLaiQC	
Unidade	K	1						
Resolução Espacial	1km						500m	
Resolução Temporal	8 dias	16 dias					4 dias	
Série Temporal	01/01/2002 a 31/12/2011						04/07/2002 a 31/12/2011	

Fonte: Wan et al. (2021), Didan (2021), Myneni e Knyazikhin (2021).

Tabela 3.3 - Produtos do CERES e suas variáveis.

CERES	SYN1deg						
Atributo	sw				par		
λ	300 a 5000nm				357.50 a 689.7nm		
Filtro	all	clr	pri	naer	clr	all	pri
Tipo de filtro	Céu total	Céu limpo	Céu limpo e sem aerossóis	Céu total e e sem aerossóis	Céu limpo	Céu total	Céu limpo e sem aerossóis
Unidade	Wm^{-2}						
Resolução Espacial	1°						
Resolução Temporal	diária						
Série Temporal	2002/01/01 a 2011/12/31						

Fonte: Team (2021).

Os atributos de Sensoriamento Remoto do MODIS escolhidos para uso neste estudo se encontram na Tabela 3.2 e os mesmos são definidos brevemente abaixo:

NDVI: O *normalized difference vegetation index* é um índice que é calculado com base no comportamento espectral da refletância dos corpos na banda do infravermelho próximo e na banda do vermelho de imagens de satélites. O NDVI varia de -1 a +1, onde valores próximos a -1 indicam superfícies que não possuem vegetação, como água e áreas urbanas, e valores próximos a +1 indicam áreas densamente cobertas por vegetação saudável. Ele possibilita a identificação da quantidade de vegetação em uma área, a detecção de mudanças na cobertura vegetal e da saúde da vegetação. **EVI:** O índice de vegetação melhorado (do inglês *enhanced vegetation index*) é semelhante ao NDVI, mas com algumas melhorias para áreas com densa cobertura vegetal e em regiões montanhosas. Ele foi projetado para minimizar distorções causadas por efeitos atmosféricos e propriedades de superfície do terreno. O cálculo do EVI considera além das bandas já utilizadas pelo NDVI, também a refletância na banda azul, um fator de correção do solo e coeficientes de ajustes. Ele é utilizado principalmente em regiões onde o NDVI não é muito eficaz devido às condições específicas do terreno e da vegetação.

Lai: O índice de área foliar é definido como a área de folhas verdes unilaterais por

unidade de área terrestre em copas de folhas largas. O Lai desempenha um papel importante na modelagem dos processos de troca de energia e de gases entre a superfície da vegetação e a atmosfera, influenciando diretamente a fotossíntese, a evaporação da água, a condensação da água na folha, a interceptação da luz solar e outros processos biogeoquímicos.

Fpar: O Fpar é a fração de radiação fotossinteticamente ativa (do inglês *fraction of photosynthetically active radiation*) absorvida pela vegetação, essa fração é a que é utilizada pela clorofila para a fotossíntese. O Fpar está diretamente relacionado com a capacidade da vegetação em absorver luz solar para a fotossíntese.

LST_Day: A temperatura da superfície da Terra medida pelo MODIS possibilita o entendimento e monitoramento de mudanças na cobertura terrestre, de variações sazonais e impactos das atividades humanas no ambiente, a partir de distúrbios.

b02: Refletância na banda do infravermelho próximo (NIR - Near Infrared), é uma medida da quantidade de radiação solar refletida pela superfície da Terra na faixa espectral do infravermelho próximo. Ela é útil para avaliar saúde e cobertura vegetal, pois a clorofila da vegetação reflete a luz no infravermelho próximo, com isso, uma vegetação mais saudável tende a ter alta refletâncias nessa banda, enquanto vegetações secas ou doentes possui menor refletância.

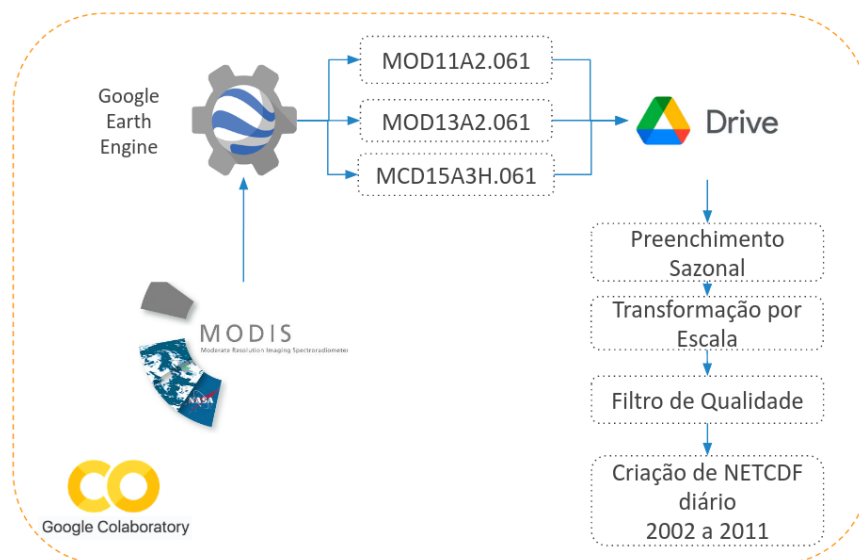
b03: É a fração de quantidade de radiação solar refletida pela superfície da Terra na faixa espectral do azul. A banda azul é ótima para distinguir diferentes tipos de cobertura terrestre, além de detectar presença de aerossóis ou outros poluentes na atmosfera.

b07: É a fração de quantidade de radiação solar refletida pela superfície da Terra na faixa do infravermelho médio. Essa refletância possui uma forte relação com a quantidade de água presente nas folhas de um dossel vegetal. A água nas plantas absorve energia no comprimento de onda de 2100 a 2300 nm, de tal forma que quanto maior a turgidez das folhas, menor a refletância no infravermelho médio.

Esses atributos foram extraídos dos seguintes produtos escolhidos com base do que já havia sido usado pela FLUXCOM (TRAMONTANA et al., 2016).

O produto MCD15A3H foi criado a partir da combinação de dados de Fpar e Lai. O algoritmo que gerou esse produto selecionou o melhor pixel disponível de todas as aquisições dos sensores MODIS em um período de 4 dias, sendo essa a resolução temporal do produto (MYNENI; KNYAZIKHIN, 2021).

Figura 3.6 - Fluxograma de extração dos produtos MODIS.



O produto MOD11A2 fornece uma média de temperatura e emissividade da superfície terrestre por pixel ao longo de 8 dias. Cada valor do pixel é uma média simples de todos os pixels correspondentes de temperatura da superfície terrestre coletados dentro do período de 8 dias (WAN et al., 2021).

O produto MOD13A2 fornece duas camadas de índices de vegetação, a primeira é o NDVI, que é referido como o índice de continuidade ao NDVI derivado do NOAA-AVHRR. A segunda camada é o EVI que possui uma sensibilidade aprimorada sobre regiões de alta biomassa, como é o caso da Floresta Amazônica. O algoritmo deste produto agiu escolhendo o melhor valor de pixel disponível de todas as aquisições do período de 16 dias (DIDAN, 2021).

Os índices de qualidade dos produtos MODIS nos fornecem a informação sobre como o dado foi processado e a qualidade dele, considerando que a quantidade de aerossol na atmosfera e a quantidade e presença de nuvens podem afetar as estimativas.

Como foi visto anteriormente, os produtos MODIS possuem muitas divergências em relação à sua resolução temporal e resolução espacial. Devido a isso, para o uso dos dados foi necessário fazer a extração dos dados utilizando a plataforma de análise *Google Earth Engine*. Nela foram feitas as etapas de extração e pré-processamento dos dados exibidas na Figura 3.6. Os atributos MODIS foram utilizados como preditores nos conjuntos de dados RS e METEORS com o objetivo de prover o comportamento espacial da vegetação dos ecossistemas.

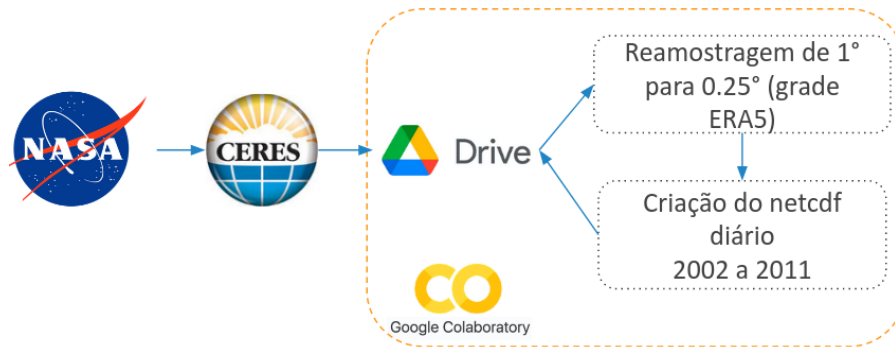
Para complementar as informações do comportamento da vegetação, foram utilizados também os atributos de radiação total difusa e da radiação PAR difusa fornecida pelo produto CERES SYN1deg Ed4A do sensor CERES (NATIONAL AERONAUTICS AND SPACE ADMINISTRATION (NASA), 2017; TEAM, 2021). Este é um diferencial deste trabalho em relação à iniciativa FLUXCOM, que não considera dados de radiação difusa entre os seus preditores no conjunto RS (TRAMONTANA et al., 2016). A atividade fotossintética das plantas depende da disponibilidade de radiação para realizar fotossíntese. A radiação solar, tanto direta quanto difusa, fornece a energia necessária para impulsionar a fotossíntese. A radiação difusa, em particular, é essencial em condições parcialmente sombreadas, onde a radiação direta é limitada. Com isso, a presença da radiação difusa pode influenciar positivamente a atividade fotossintética das plantas, especialmente em locais onde a luz direta é reduzida, como em dias nublados e florestas densas. Devido a esse comportamento foram escolhidos os produtos de radiação difusa dos sensores CERES.

Os produtos CERES são derivados de medições feitas por instrumentos CERES a bordo dos satélites Aqua, Terra, S-NPP e NOAA-20, combinados com dados de imagens de alta resolução de satélites em órbita polar e geoestacionária. Esses produtos fornecem uma representação da variação na distribuição de radiação na parte superior da atmosfera, na atmosfera e na superfície terrestre. Neste trabalho foi utilizado o produto CERES_SYN1deg_Ed4.1, que possui uma resolução espacial de 1° e uma resolução temporal diária. Os dados utilizados são provenientes dos sensores CERES a bordo dos satélites Terra + Aqua e NOAA20.

Na Tabela 3.3 são listados os quatro produtos distintos dos fluxos atmosféricos do CERES, o *Clear Sky*, o *All Sky*, o *Pristine* e o *No Aerosol* e abaixo são descritas as diferenças entre eles (WIELICKI et al., 1996):

- a) **All-sky (todo o céu, isto é, sob condições nubladas e limpas)**: Os fluxos sob todas as condições de céu são calculados a cada hora para cada ponto de grade de $1^\circ \times 1^\circ$. Os fluxos médios mensais de curto e longo prazo para superfície são calculados pela média dos fluxos em todo o céu.
- b) **Clear-sky (céu limpo, isto é, sem nuvens)**: Os fluxos de céu limpo também são calculados a cada hora para cada ponto de grade de $1^\circ \times 1^\circ$, removendo-se os pixels afetados por nuvens.
- c) **Pristine (prístino)**: Os fluxos prístinos são calculados removendo aerossóis dos cálculos de céu limpo, ou seja, incluindo apenas o espalhamento

Figura 3.7 - Fluxograma de extração dos produtos CERES.



molecular e a absorção e excluindo o efeito das nuvens.

- d) **All sky with no aerosol (céu sob todas as condições sem aerossol):**
Os fluxos sem aerossol são calculados removendo aerossóis dos cálculos em todo o céu.

A extração dos dados CERES para a grade do ERA5 foi realizada partir da aquisição dos dados na plataforma da NASA ([NATIONAL AERONAUTICS AND SPACE ADMINISTRATION \(NASA\), 2017](#)), logo após os dados foram armazenados na nuvem e re-amostrados da grade de 1° para a grade de 0,25° do ERA5, esse processamento foi exemplificado na Figura 3.7.

3.4 Meteorológicos (METEO)

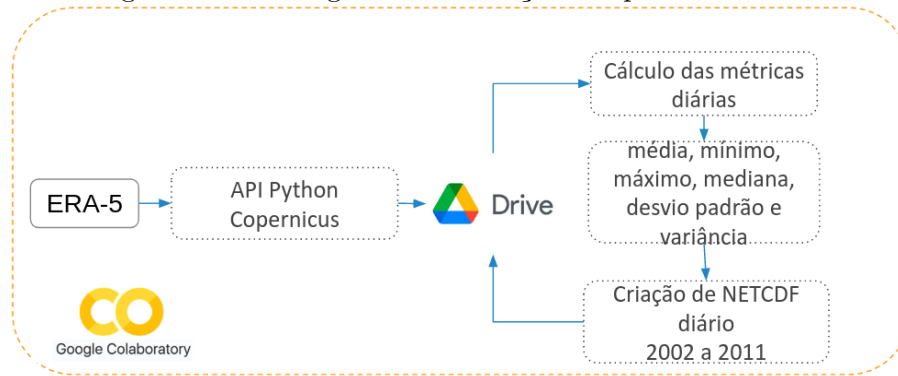
Os dados meteorológicos deste trabalho são provenientes, em maior parte, do produto meteorológico de reanálises "*ERA5 hourly data on single levels from 1940 to present*" criado pelo Centro Europeu de Previsão de Tempo a Médio Prazo (ECMWF, European Centre for Medium-Range Weather Forecasts). Os dados estão disponíveis a partir de 1940 e combinam dados de modelos e observações meteorológicas em um conjunto de dados global completo e consistente ([HERSBACH; BELL, 2023](#)).

Tabela 3.4 - Dados Meteorológicos do Modelo Era5 do ECMWF Climate Reanalysis e Precipitação do Produto MERGE.

Atributo	Definição	Unidade	Resolução espacial	Resolução Temporal	Série Temporal
ERA5					
t2m	Temperatura do ar	K	0,25°	diária	2002/01/01 a 2011/12/31
d2m	Temperatura do ponto de orvalho	K	0,25°	diária	2002/01/01 a 2011/12/31
aluvd	Albedo UV visível para radiação difusa	1	0,25°	diária	2002/01/01 a 2011/12/31
alnid	Albedo no infravermelho próximo para radiação difusa	1	0,25°	diária	2002/01/01 a 2011/12/31
e	Evaporação	kg m ⁻²	0,25°	diária	2002/01/01 a 2011/12/31
stl1	Temperatura do Solo - 0 a 7cm	K	0,25°	diária	2002/01/01 a 2011/12/31
ro	Escoamento superficial	m	0,25°	diária	2002/01/01 a 2011/12/31
VPD	Déficit de Pressão de Vapor	hPa	0,25°	diária	2002/01/01 a 2011/12/31
MERGE					
prec	Precipitação Acumulada em 24h	kg m ⁻²	0,1°	diária	2002/01/01 a 2011/12/31

Fonte: [Hersbach e Bell \(2023\)](#), [Rozante \(2000\)](#).

Figura 3.8 - Fluxograma de extração dos produtos ERA5.



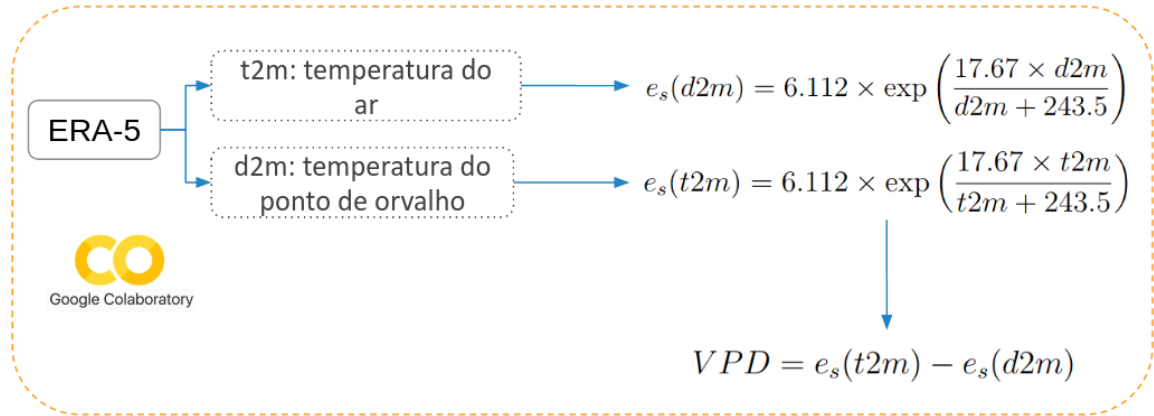
Este produto fornece estimativas horárias, atualizadas com uma latência de aproximadamente 5 dias. Foram utilizados os dados horários na superfície (em nível único) neste trabalho. Na Tabela 3.4, encontra-se a descrição dos atributos meteorológicos utilizados, os quais constituem as componentes meteorológicas predictoras dos conjuntos METEO e METEORS.

Na Figura 3.8, foi ilustrado o processo de extração dos dados ERA5, o qual envolveu o uso da API Python disponibilizada pela Copernicus para download dos dados. Os dados, em formato NETCDF, foram então armazenados na nuvem. Após a conclusão da extração, foram calculadas métricas diárias, incluindo média, variância, desvio padrão, mediana, mínimo e máximo de cada atributo do ERA5. Em seguida, um arquivo único diário contendo esses atributos foi criado.

O deficit de pressão de vapor (VPD) representa a diferença entre a quantidade máxima de vapor de água que o ar pode conter quando saturado e a quantidade real de vapor de água presente no ar. Nos ecossistemas, o VPD está intimamente ligado à transpiração das plantas e à evaporação do solo, ambos influenciados por condições de alta temperatura e baixa umidade, e vice-versa. Essa medida é crucial para entender as condições de umidade do ar e os processos de troca de água entre plantas e atmosfera, desempenhando um papel fundamental na regulação do ciclo hidrológico dos ecossistemas. Um VPD elevado indica que o ar contém menos umidade relativa em comparação com o máximo possível na mesma temperatura, geralmente ocorrendo em condições de ar seco, onde a evaporação é alta e a umidade relativa do solo é baixa.

Um deficit de vapor pronunciado pode acelerar a evaporação da umidade do solo e das superfícies das plantas, resultando em condições de seca e estresse hídrico

Figura 3.9 - Cálculo do Deficit de Pressão de Vapor.



para as plantas, especialmente se a taxa de transpiração não for compensada pela disponibilidade de água no solo (PALACIOS; PAULO, 2023). Neste trabalho, a equação de Magnus (MURRAY, 1967) foi empregada para calcular o Deficit de Pressão de Vapor (VPD), utilizando as variáveis de temperatura do ar e temperatura do ponto de orvalho provenientes do conjunto ERA5 (PALACIOS; PAULO, 2023; FOREST,).

A equação de Magnus para o cálculo do pressão de vapor saturado é dada pela Equação 3.2:

$$e_s(T) = 6.112 \times \exp\left(\frac{17.67 \times T}{T + 243.5}\right) \quad (3.2)$$

onde:

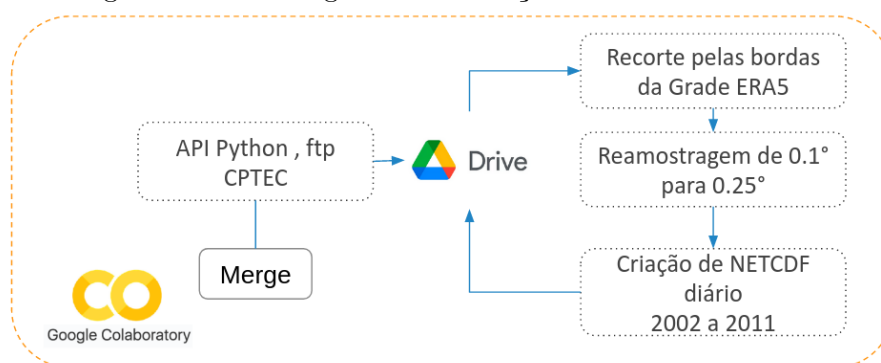
- $e_s(T)$ é a pressão de vapor saturado à temperatura T (em hPa).
- T é a temperatura do ar em graus Celsius.

A mesma equação pode ser utilizada para obter a pressão de vapor e_v a partir da temperatura do ponto de orvalho ($d2m$). A partir da pressão de vapor saturado $e_s(T)$ e da pressão de vapor e_v , o VPD pode ser calculado como 3.3:

$$VPD = e_s(T) - e_v \quad (3.3)$$

Usando as variáveis de temperatura do ar ($t2m$) e temperatura do ponto de orvalho ($d2m$) do ERA5, a Equação 3.2 foi utilizada para o cálculo do VPD para os dois níveis

Figura 3.10 - Fluxograma de extração dos dados MERGE.



de referencia do calculo, na Figura 3.9 há a exemplificação de como foi realizado esse cálculo. Também foram calculados os valores de mínimo, máximo e médio de VPD, com as respectivas métricas de t2m e d2m, que estão representados na Tabela 3.4. Após o cálculo das séries temporais de VPD, esse atributo foi incluído ao conjunto de dados do ERA5.

Para complementar os dados do produto ERA5 nos conjuntos METEO e METEORS, foram incorporados os dados de precipitação do MERGE. Esse produto foi selecionado devido à sua origem e validação específicas para a região da América do Sul pelo Instituto Nacional de Pesquisas Espaciais (INPE) (ROZANTE; MOREIRA, 2010). O MERGE é gerado combinando observações de precipitação com estimativas de precipitação por satélite (ROZANTE, 2000). Utilizamos os dados diários do MERGE, cujos atributos são descritos na Tabela 3.4.

O processo de preparação dos dados do MERGE é detalhado na Figura 3.10. Os dados foram obtidos através do endereço ftp disponibilizado pelo INPE, apresentando-se no formato ".grib", posteriormente convertidos para ".NetCDF" para facilitar operações e integração com outros conjuntos de dados. Em seguida, os dados foram armazenados na nuvem, recortados de acordo com os limites da grade ERA5 e reamostrados de 0,1° para 0,25°. Esse procedimento permitiu o cálculo dos valores médios das células do MERGE para cada célula do ERA5.

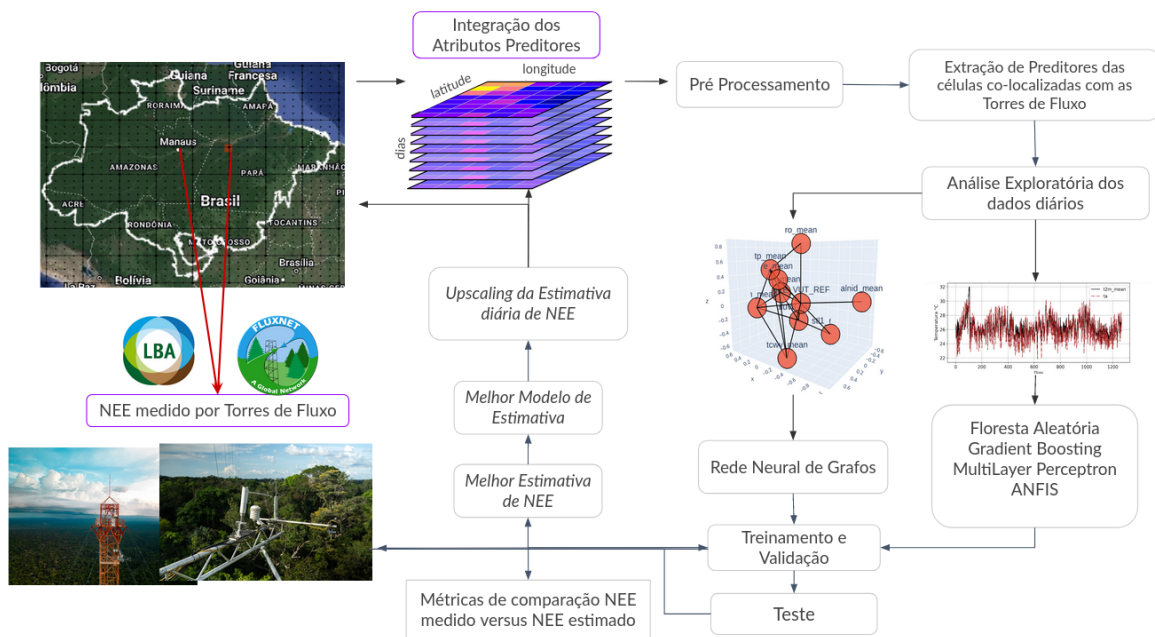
É importante ressaltar que neste trabalho foram utilizados mais atributos meteorológicos em comparação com a iniciativa FLUXCOM. O FLUXCOM considerou no conjunto METEO apenas as variáveis temperatura do ar, radiação global, VPD e precipitação (TRAMONTANA et al., 2016). Neste trabalho, foram incluídos atributos adicionais 3.4 que possuem uma relação física, direta ou indireta, com as trocas

de carbono entre a biosfera e a atmosfera, como a temperatura do solo, a evaporação, o escoamento superficial e os albedos de superfície.

4 MÉTODOS

O fluxo geral da metodologia foi delineado na Figura 4.1. Em resumo, o método envolveu a utilização dos atributos citados anteriormente na aprendizagem de cinco diferentes modelos de aprendizado de máquina para estimar a série temporal de NEE de 2002 a 2011 em uma célula de grade co-localizada com a torre de fluxo BrSa1. Após a conclusão do processamento e verificação de qualidade dos modelos, foi iniciado o processo de treinamento dos modelos e logo após os processos de validação e teste de estimativa para as torres LBA e Fluxnet. Com base no desempenho, o melhor modelo, devidamente treinado e validado, foi selecionado para realizar a ampliação de escala das séries temporais diárias no Bioma Amazônia em cada célula de grade.

Figura 4.1 - Fluxograma geral da metodologia utilizada neste trabalho.



4.1 Integração e pré processamento

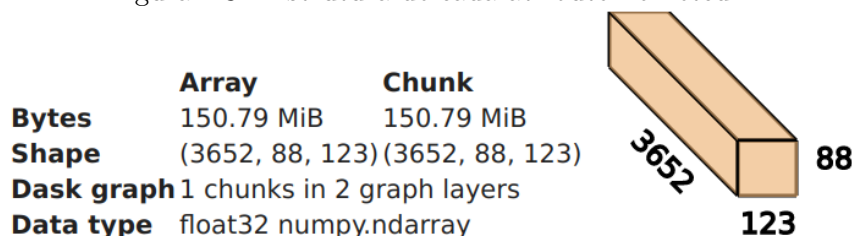
A grade de resolução espacial de $0,25^\circ$ do produto ERA5 foi escolhida como a base para a ampliação de escala. Todos os outros atributos foram re-amostrados para essa mesma resolução e co-localizados, garantindo uniformidade nos dados e facilitando sua integração e comparação.

Figura 4.2 - Arquivo netcdf.

```
xarray.Dataset
  Dimensions:      (date: 3652, latitude: 88, longitude: 123)
  Coordinates:
    date           (date)      object '2002-01-01' ... '2011-12-31'
    longitude      (longitude) float32 -73.98 -73.73 ... -43.73 -43.48
    latitude       (latitude)  float32  5.09 4.84 4.59 ... -16.41 -16.66
  Data variables:
    (76)
  Indexes: (3)
  Attributes: (0)
```

Exemplo da estrutura do arquivo netcdf integrado do Bioma Amazônia ao ser aberto com a biblioteca Xarray. Esse conjunto de dados possui 76 variáveis e cada variável tem comportamento tridimensional, ela varia por 88 latitude, 123 longitudes e 3652 dias.

Figura 4.3 - Estrutura de cada atributo no netcdf.



Estrutura de cada uma das variáveis, exemplificado o cubo de dados e o tamanho dos dados.

Os dados integrados deste trabalho foram armazenados em um arquivo netcdf, exemplificado nas Figuras 4.2 e 4.3, os diferentes atributos foram combinados para formar os três conjuntos de dados preditores: METEO, RS e METEO + RS. Uma descrição geral desses conjuntos de dados está presente na Tabela 4.1.

Tabela 4.1 - Conjuntos de Dados Preditores.

Conjuntos de Dados	Conjunto de Atributos	Número de Preditores	Série Temporal
METEO	ERA5+MERGE	49	01/01/2002 31/12/2011
RS	MODIS+CERES	15	04/07/2002 31/12/2011
METEORS	ERA5+MERGE +MODIS+CERES	64	04/07/2002 31/12/2011

Como pode ser visto na Tabela 4.1, a série temporal dos conjuntos RS e METE-

ORS teve início em 04/07/2002, devido à indisponibilidade de dados do produto MCD15A3H.061 antes deste período.

$$\text{Valor Normalizado} = \frac{x - \text{Mínimo}}{\text{Máximo} - \text{Mínimo}} \quad (4.1)$$

Algumas análises estatísticas, como a intercomparação por meio de boxplot, se beneficiam da normalização de dados, especialmente quando as escalas variam consideravelmente. Essa prática é útil até mesmo para o uso dos atributos em redes neurais. Portanto, os conjuntos preditores METEO, RS e METEORS foram normalizados. Essa transformação foi realizada de acordo com a regra de normalização de min-max, conforme descrito na Equação 4.1.

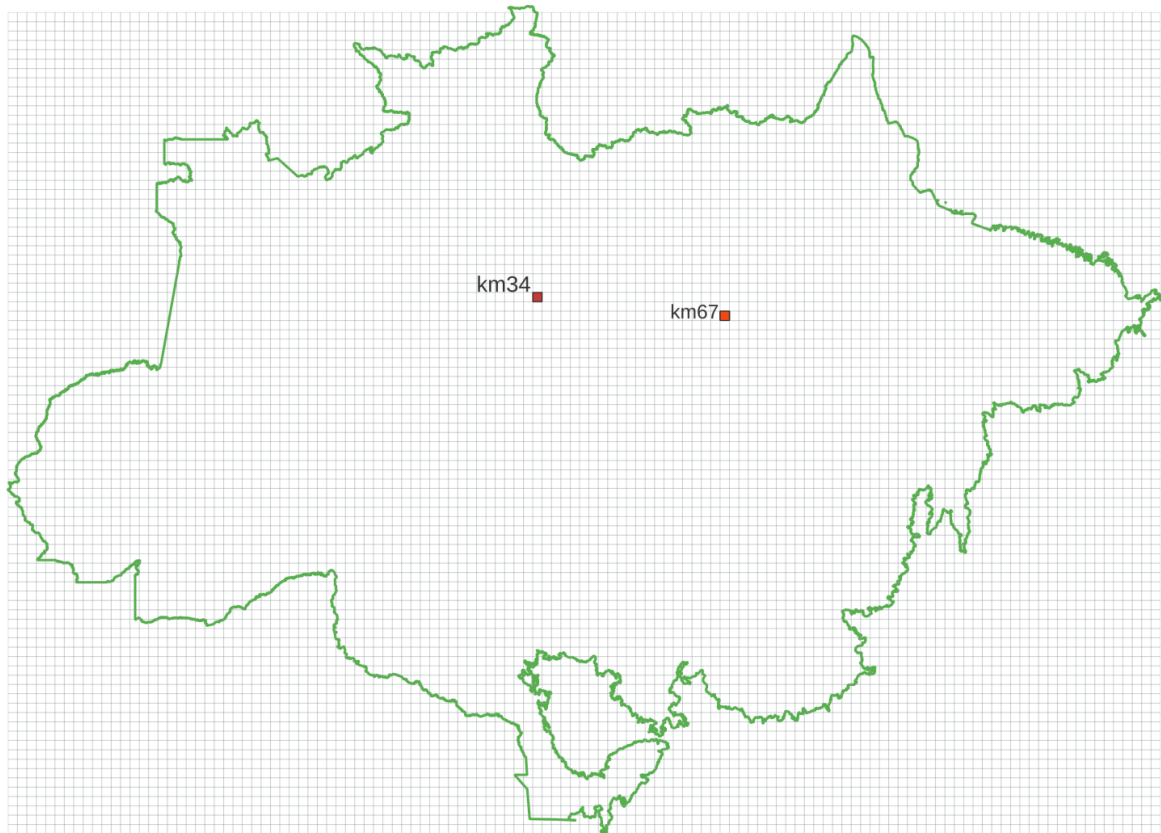
4.2 Análise exploratória dos dados e intercomparação

A grade utilizada neste estudo está representada na Figura 4.4, a mesma possui 88x123 células e as torres de fluxo que foram utilizadas neste trabalho estão contidas em apenas 2 células. Assim, as células que possuem torres de fluxo foram chamadas pelo nome de sua respectiva torre referência na célula, K67 e K34. A **Célula K34** possui em seu interior a torre K34 do LBA, e a **Célula K67** possui em seu interior as torres K67, a K77 e a K83 do LBA. Logo, os dados da torre BrSa1 do Fluxnet também são atribuídos a essa última célula.

A primeira parte da análise exploratória de dados (EDA) neste estudo foi realizada para investigar o comportamento estatístico dos atributos preditores nas **Células K34 e K67**, no período de 2002 a 2011. O objetivo foi compreender as semelhanças e diferenças entre esses preditores, bem como identificar possíveis vieses, padrões e tendências. Essa análise foi fundamental para entender como essas variáveis podem influenciar as estimativas de NEE por meio de algoritmos de aprendizado de máquina.

A segunda parte deste estudo visou compreender o comportamento das séries temporais de NEE observadas nas Torres K77, K67, K83, BrSa1 e K34. O objetivo foi identificar padrões nas séries temporais, avaliar a disponibilidade e integridade dos dados, identificar locais de preenchimento de dados ausentes e examinar possíveis situações que possam influenciar as estimativas de NEE. Essa análise forneceu detalhes essenciais para embasar as decisões tomadas ao longo do estudo.

Figura 4.4 - Grade Bioma Amazônia , Célula K34 e Célula K67, em vermelho e laranja, respectivamente.



4.2.1 Intercomparação da temperatura do ar entre o modelo ERA5 e os dados observados das torres K34 e K67

Para avaliar a representatividade dos atributos meteorológicos das células do ERA5 em relação aos locais de estudo, foram calculadas métricas estatísticas de intercomparação. Estas métricas incluem o coeficiente de determinação, o viés, o erro quadrático médio e o erro médio absoluto para as séries temporais de temperatura do ar das Células K67 e K34, em comparação com as medições de temperatura do ar observadas nas Torres K34 e K67. A seguir, são apresentadas as definições das métricas utilizadas:

Correlação de Pearson: É uma medida de quão bem os valores modelados e observados estão relacionados linearmente. Ele varia de -1 a 1, onde -1 indica uma correlação negativa (uma diminuição em um valor está associada a um aumento no outro), 0 indica nenhuma correlação e 1 indica uma correlação positiva perfeita (um aumento

em um valor está associado a um aumento no outro). Seu cálculo é realizado a partir da Equação 4.2.

$$R = \frac{\sigma_{xy}}{\sigma_x \sigma_y} \quad (4.2)$$

Onde σ_{xy} é a covariância entre os valores modelados e observados, σ_x é o desvio-padrão dos valores modelados e σ_y é o desvio-padrão dos valores observados.

Coefficiente de Determinação: Mede a proporção da variabilidade total dos valores observados que é explicada pelo modelo. Varia de 0 a 1, onde 0 indica que o modelo não explica a variabilidade dos dados e 1 indica que o modelo explica totalmente a variabilidade dos dados. Quando R^2 é alto, é um bom indicativo de que o modelo se ajusta bem aos dados observados. Seu cálculo é realizado a partir da Equação 4.3.

$$R^2 = \left(\frac{\sigma_{xy}}{\sigma_x \sigma_y} \right)^2 \quad (4.3)$$

Bias: O bias ou viés é uma medida de quão bem o modelo é capaz de prever a média dos valores observados. É a diferença média entre os valores observados e previstos. Quando o valor do BIAS é zero, o modelo não apresenta nenhum viés. Quando o valor do BIAS é positivo, o modelo tende a superestimar os valores observados e, quando negativo, o modelo tende a subestimar os valores observados. Seu cálculo é realizado a partir da Equação 4.4.

$$BIAS = \frac{1}{n} \sum_{i=1}^n (x_i - y_i) \quad (4.4)$$

Onde y_i é o valor observado, x_i é o valor previsto e n é o número de observações.

MAE e RMSE: O erro quadrático médio (RMSE) é uma medida da diferença média entre os valores observados e previstos. É uma medida de dispersão, pois mede a raiz quadrada da média dos erros quadrados. Quanto menor o valor do RMSE, melhor é o desempenho do modelo. Uma diferença com relação ao Erro Médio Absoluto (MAE) está relacionado ao fato que os *outliers* são melhor representados no RMSE, contudo o MAE dá o erro mais próximo da magnitude observada. O MAE e o RMSE são calculados a partir das Equações (4.5) e (4.6).

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - x_i| \quad (4.5)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - x_i)^2} \quad (4.6)$$

onde n é o número de observações e y_i é o valor observado e x_i é o valor estimado e $|\cdot|$ é o valor absoluto.

4.2.2 Distribuição dos dados, células K67 e K34.

Após a intercomparação, foi verificado se a distribuição de dados de cada um dos preditores das células K67 e K34 é normal utilizando o teste de Shapiro-Wilk. Se o valor do teste for próximo de 1 e seu *valor p* de significância for maior que 0,05, a hipótese nula não pode ser rejeitada e os dados podem ser considerados normalmente distribuídos. No entanto, se o valor p de significância for menor que 0,05, os dados podem ser considerados não normalmente distribuídos. Em caso de desvio da normalidade, os coeficientes de assimetria (Sk) e de variação (CV) foram utilizados para compreender os padrões de distribuição dos atributos.

Coefficiente de Variação: É uma medida que expressa a variabilidade dos dados em relação à média, representando a variação relativa dos dados em porcentagem. Ele é calculado pela razão entre o desvio padrão e a média dos dados, multiplicado por 100. A fórmula para calcular o coeficiente de variação é expressa pela Equação 4.7.

$$CV = \frac{s}{\bar{x}} \times 100 \quad (4.7)$$

Onde s é o desvio-padrão e \bar{x} é a média da variável.

Coefficiente de Assimetria : Indica se a distribuição dos dados é simétrica ou assimétrica em relação à média. Um valor próximo a zero indica uma distribuição simétrica, enquanto valores positivos ou negativos indicam assimetrias e suas respectivas direções. Essa análise, combinada com os histogramas dos preditores, permite identificar distorções nas estimativas estatísticas, como na média e na variância, uma vez que a assimetria na distribuição dos dados pode levar a erros na interpretação dos resultados. A Equação (4.8) representa o cálculo do coeficiente de assimetria.

$$Sk = \frac{3(\bar{x} - M_o)}{s} \quad (4.8)$$

Onde \bar{x} é a média dos dados, s é o desvio padrão e M_o é a mediana dos dados.

4.2.3 Análise das séries temporais

A maioria dos atributos preditores das Células K34 e K67, utilizados neste trabalho, exibiu comportamento de não normalidade. Devido a isso, para entender as asso-

ciações entre tais atributos, foi utilizada a correlação de Spearman. Essa técnica é não paramétrica e não exige a suposição de normalidade dos dados (CONOVER, 1999). O coeficiente de correlação de Spearman varia de -1 a 1 e segue as seguintes regras: 1 indica uma correlação positiva perfeita monotônica, onde à medida que uma variável aumenta, a outra também aumenta, seguindo uma relação monotonicamente crescente; -1 indica uma correlação negativa perfeita monotônica, onde à medida que uma variável aumenta, a outra diminui, seguindo uma relação monotonicamente decrescente; e 0 indica ausência de correlação monotônica entre as variáveis. A correlação de Spearman foi calculada, juntamente com o nível de significância, valor p , da métrica, para cada par de atributos, e foi gerado matrizes de correlação para cada um dos conjuntos (METEO, RS, METEORS) das Células K34 e K67.

Os atributos preditores deste estudo foram selecionados por sua relação intrínseca com os processos que regulam o ciclo do carbono e, conseqüentemente, influenciam o comportamento da NEE. É crucial determinar se houve um aumento geral em alguma das variáveis ao longo da série temporal deste estudo e compreender quais componentes dos processos ecossistêmicos podem ter sido afetados por essa tendência identificada.

Para esta análise, utilizou-se o teste de Mann-Kendall, uma técnica estatística não paramétrica usada para detectar tendências em séries temporais. O teste de Mann-Kendall avalia se há uma tendência monotônica crescente ou decrescente ao longo do tempo em um conjunto de dados. Ele se baseia nas diferenças entre pares de observações em uma série temporal e compara as direções das mudanças entre esses pares de dados, determinando se há uma tendência significativa em uma direção específica. Em resumo, o teste de Mann-Kendall verifica se há uma tendência monotônica crescente ou decrescente nos dados ao longo do tempo (KENDALL, 1975). A hipótese nula do teste é que não há tendência na série temporal dos atributos, enquanto a hipótese alternativa é que há uma tendência temporal monotônica crescente ou decrescente (positiva ou negativa). Os resultados do teste de Mann-Kendall fornecem a direção da tendência, o resultado de significância do teste, *valor p* correspondente e a intensidade da tendência, medida pelo coeficiente *tau* (HUSSAIN; MAHMUD, 2019).

Como dito anteriormente na Seção 3, as séries temporais de NEE dos conjuntos Fluxnet e LBA possuem muitos dados preenchidos. Por isso, foi realizada uma análise da quantidade de dados preenchidos por torres e identificação de possíveis limitações

que essas situações poderiam ocasionar nos algoritmos de aprendizado de máquina e na estimativa de NEE. Para a compreensão das semelhanças e diferenças entre os valores da troca líquida de carbono nas quatro torres utilizadas neste estudo, foram calculadas estatísticas básicas (média, máximo, mínimo, desvio padrão) para cada variável alvo de NEE.

Adicionalmente, foram criados gráficos que demonstram o comportamento médio mensal por ano de cada torre. Essa análise permite uma compreensão mais detalhada dos padrões interanuais e das discrepâncias entre as bases de dados. O objetivo foi identificar as variáveis alvo mais adequadas para as estimativas de NEE, uma vez que os dados preditores possuem menor escala espacial e devem refletir o comportamento do ecossistema de cada torre. Vale ressaltar que a exclusão de dados preenchidos das séries temporais de NEE foi necessária para garantir uma representatividade mais precisa das variáveis alvo. No entanto, essa prática resultou em lacunas de dados, limitando assim a quantidade de informações disponíveis para serem utilizadas nos modelos de aprendizado de máquina.

4.3 Treinamento dos modelos de aprendizado de máquina

4.3.1 Definição dos conjuntos de preditores e de variáveis alvo

Com base na análise exploratória realizada para os preditores das Células K34 e K67 e para as variáveis alvo de cada torre, foram criadas algumas variações dos conjuntos para o treinamento dos modelos. Isso foi feito com o objetivo de evitar viés no modelo, aumentar a capacidade de generalização e melhor representar o comportamento dos preditores e das variáveis alvo.

Considerando o resultado advindo da análise do comportamento dos preditores das células com o coeficiente de correlação de spearman, foram criadas três variações de cada conjunto de preditores, os quais podem ser vistos na Tabela 4.2. O termo ' - ' indica retirada de dados, as mudanças realizadas e suas indicações seguem abaixo:

- **corr** : Foram retirados os atributos altamente correlacionados positivamente e negativamente, pois os mesmos contam a mesma história para o algoritmo de aprendizado de máquina

- max - min - var** : Foram retirados os atributos meteorológicos que possuíam esses prefixos.

Devido à diversidade dos conjuntos de dados e das Torres do LBA e do Fluxnet,

Tabela 4.2 - Conjuntos de Atributos utilizados no treinamento.

Conjunto de Dados	Conjunto Treinamento
METEO	METEO METEO - corr METEO - mean - max - min - var
METEORS	METEORS METEORS - corr METEORS - mean - max - min - var
RS	RS RS - corr

foram criados diferentes conjuntos de treinamento, conforme representado na Tabela 4.3. Esses conjuntos foram elaborados com o intuito de identificar e avaliar qual variável alvo é melhor prevista pelos modelos. Levando em consideração que os dados provenientes do Fluxnet diferem dos dados do LBA, e considerando também a presença de três torres (K67, K77 e K83) dentro da célula K67, busca-se determinar qual medida de NEE é mais facilmente estimada pelos preditores deste estudo.

Para compreender os diferentes padrões de comportamento dos preditores em relação à variável alvo durante o treinamento pelos modelos, optou-se por embaralhar os dados de entrada ao dividir o conjunto em treinamento, validação e teste. Uma vez que os dados de entrada constituem uma série temporal e o objetivo principal não era replicar exatamente essa série, mas sim entender os padrões diários para estimar a série de NEE com dados não conhecidos e generalizar o modelo, essa decisão de embaralhamento foi tomada. Isso evita que os dados sejam tendenciosos em relação ao comportamento sazonal da série temporal de NEE usada como variável alvo de treinamento.

A integração dos conjuntos para diferentes variáveis alvo teve como objetivo reduzir a quantidade de lacunas na representação de NEE na Célula K67. Para alcançar isso, apenas dados de boa qualidade das torres foram utilizados nos modelos de aprendizado de máquina. Na união dos conjuntos de todas as torres, seguimos uma sequência específica: primeiro, combinamos os conjuntos da K77 e K83 nos locais onde faltavam dados da torre K67. Em seguida, os conjuntos da Fluxnet foram usados para preencher os dados ausentes desse conjunto resultante dos dados da K67+K77+K83. Essa abordagem de integração ajudou a maximizar a disponibilidade de dados e a melhorar a representatividade da variável alvo NEE na Célula K67.

Tabela 4.3 - Conjuntos de treinamento.

Conjuntos	Torre alvo	Série Tempo- ral
RS e METEORS	K67	04/07/2002 a 31/12/2005
METEO	K67	01/01/2002 a 31/12/2005
RS e METEORS	K77	04/07/2002 a 31/12/2005
METEO	K77	01/01/2002 a 31/12/2005
RS e METEORS	K83	04/07/2002 a 31/12/2004
METEO	K83	01/01/2002 a 31/12/2004
RS e METEORS	K83	04/07/2002 a 31/12/2004
METEO	K83	04/07/2002 a 31/12/2004
RS e METEORS	BrSa1	04/07/2002 a 31/12/2011
METEO	BrSa1	01/01/2002 a 31/12/2011
RS e METEORS	K67+ BrSa1	04/07/2002 a 31/12/2011
METEO	K67+ BrSa1	01/01/2002 a 31/12/2011
RS e METEORS	K77 + BrSa1	04/07/2002 a 31/12/2011
METEO	K77 + BrSa1	01/01/2002 a 31/12/2011
RS e METEORS	Todas as torres	04/07/2002 a 31/12/2011
METEO	Todas as torres	01/01/2002 a 31/12/2011

Neste trabalho foram utilizados cinco modelos de aprendizado de máquina, sendo que cada um deles possuem diferentes arquiteturas e diferentes entradas. São eles a Floresta Aleatória (RF), o Gradient Boost (GBoost), a rede neural Multilayer Perceptron, a rede neural ANFIS e a Rede Neural de Grafos (GNN). Durante o processo de treinamento dos modelos, foram realizadas várias rodadas de testes iniciais utilizando hiperparâmetros aleatórios e amplamente variados. Isso foi feito com o objetivo de compreender o comportamento e o desempenho dos modelos em diferentes configurações. A partir dos resultados desses testes foi identificado as combinações de hiperparâmetros que levaram a melhorias significativas no desempenho do modelo. Essas observações foram utilizadas para definir os intervalos e configurações de busca na técnica de busca bayesiana com o Optuna (AKIBA et al., 2019). Dessa forma, pude direcionar a busca para regiões do espaço de hiperparâmetros onde havia maior probabilidade de encontrar configurações com desempenho otimizado.

Os conjuntos de treinamento, teste e validação foram separados de forma a 60% dados no primeiro e outros 20% nos dois últimos. A única exceção aconteceu com a separação dos conjuntos da Rede Neural de Grafos, pois essa deve ter o mesmo número de dados de treinamento, validação e teste, portanto os conjuntos foram divididos em três partes iguais.

A estrutura de entrada da GNN difere em relação aos outros algoritmos utilizados, já que ela está organizada em grafos. Eles foram construídos a partir das relações entre NEE e as variáveis preditoras deste estudo, considerando o comportamento qualitativo e quantitativo encontrados na análise exploratória dos dados. O grafo foi criado de modo que as variáveis preditoras são os vértices, as arestas são as relações entre tais variáveis e as séries temporais de cada variável preditora será a componente de característica do seu vértice. Nas Figuras 4.5 e 4.6 há a representação dos grafos base do conjunto METEO e RS, respectivamente. Essas são apenas representações dos grafos bases, pois como há muitas variáveis, os grafos resultantes de cada conjunto de treinamento possuem muitos nós e arestas para uma representação clara. Na Tabela 4.4 está apresentada a arquitetura dos grafos criados a partir do conjunto de dados para a Rede Neural de Grafos. Os grafos que possuem atributos das arestas iguais ao coeficiente de correlação de Spearman foram utilizados como entrada da GAT. Na GAT, a correlação de Spearman é útil para calcular a atenção entre nós, levando em consideração a relação entre os atributos. Já na GCN, que depende da convolução nos grafos, essa correlação não é utilizada explicitamente para calcular a importância entre os nós.

Figura 4.5 - Grafo base METEO.

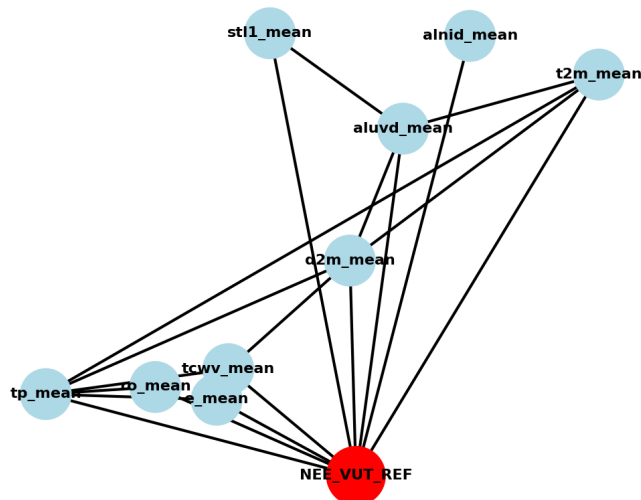


Figura 4.6 - Grafo base RS.

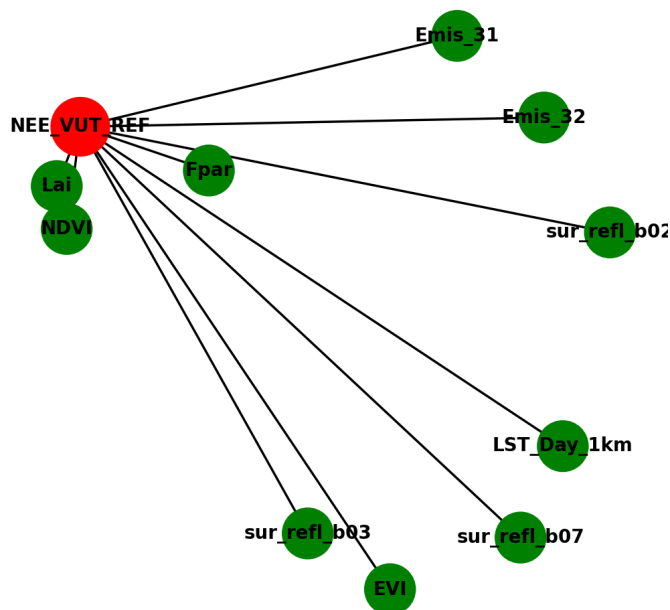


Tabela 4.4 - Características dos grafos de entradas na GNN.

Conjunto	Nós	Arestas	Grau	Características por nó	Atributos das arestas
METEO	48	1128	47	1012	não
RS	17	136	16	963	não
METEORS	63	1953	62	963	não
METEO	48	1128	47	1012	Correlação de Spearman
RS	17	136	16	963	Correlação de Spearman
METEORS	63	1953	62	963	Correlação de Spearman

4.3.2 Treinamento, validação e teste

O treinamento, validação e teste dos modelos Random Forest, Gradient Boost, Multilayer Perceptron e ANFIS foram efetuados com todos os conjuntos de dados das Tabelas 4.3 e 4.2. Os hiperparâmetros de cada modelo foram ajustados dentro dos limites especificados na Tabela 4.5.

Os modelos foram inicialmente treinados e validados com os dados das torres de Santarém (K67, K77, K83 e BrSa1). Posteriormente, os modelos que demonstraram melhor desempenho foram selecionados para os testes. Esses testes foram conduzidos tanto com as torres de Santarém (K67, K77, K83) quanto com a Torre K34, localizada na Célula K34. Essa abordagem permitiu avaliar a capacidade de generalização dos modelos para diferentes conjuntos de dados. É importante ressaltar que cada modelo foi treinado individualmente para cada conjunto de dados, garantindo uma análise específica e precisa do desempenho de cada modelo em relação aos diferentes conjuntos de atributos.

O treinamento, validação e teste dos modelos das variações da Rede Neural de Grafos (GCN e GAT) foram efetuados apenas com o conjunto de dados do Fluxnet. Não foram utilizados nessas redes os conjuntos de dados que utilizam como variável alvo o NEE do LBA. Isso foi feito devido ao fato de que a GNN é uma rede neural profunda e a quantidade de dados para seu treinamento diminuiu drasticamente quando utilizados os dados do LBA.

Os modelos RF, GBoost e MLP foram criados, treinados e validados utilizando a biblioteca do *python Scikit Learn* (PEDREGOSA et al., 2011), sendo que para mudanças dos hiperparâmetros de forma automática foi utilizado a biblioteca *Optuna*. Já os modelos ANFIS, GAT e GCN foram criados, treinados e validados utilizando a biblioteca *Pytorch*. Em específico, para criação dos grafos, criação das matrizes de adjacências dos grafos e da arquitetura das GNN foi utilizada a biblioteca *Pytorch Geometric* (FEY; LENSSEN, 2019).

Para todos os modelos de RNAs, a função de perda escolhida foi o erro quadrático médio (MSE), enquanto o método de otimização utilizado foi o Adam. O processo de validação das redes neurais foi feito durante o treinamento, de modo a obter as curvas do MSE nos dados de treinamento e nos dados de validação. O objetivo era identificar se as redes neurais, ao mesmo tempo que estavam aprendendo, estavam generalizando a aprendizagem. Devido a isso, após o treinamento, na etapa de validação foram calculadas as métricas BIAS (Eq. 4.4), R (Eq. 4.2), R^2 (Eq. 4.3),

Tabela 4.5 - Hiperparâmetros Ajustados dos modelos de aprendizado de máquina.

Algoritmo	Hiperparâmetros	Limites
RF	n_estimators max_depth min_samples_split min_samples_leaf min_impurity_decrease min_weight_fraction_leaf ccp_alpha random_state	range(5, 200, 20) range(5, 30, 5) range(2, 10) range(1, 4) range(0.0, 0.2) range(0.0, 0.2) range(0.0, 0.2) 42
GBoost	n_estimators learning_rate max_depth min_samples_split min_samples_leaf random_state	range(5, 200) range(0.01, 0.5) range(3, 7) range(2, 10) range(1, 4) 42
MLP	learning_rate dropout weight_decay hidden_layers activation_function interactions_number random_state	range(0.0001, 0.5) range(0.02, 0.2) range(0.001, 0.2) range(2, 300) logistic, tanh, relu range(10, 5000) 42
ANFIS	learning_rate dropout weight_decay hidden_layers activation_function epochs fuzzy_function random_state	range(0.0001, 0.5) range(0.02, 0.2) range(0.001, 0.2) range(2, 300) logistic, tanh, relu range(10, 5000) gaussiana 42
GAT	learning_rate dropout weight_decay hidden_layers activation_function GATConv num_heads epochs random_state	range(0.0001, 0.5) range(0.02, 0.2) range(0.001, 0.2) range(5, 400) logistic, tanh, relu range(2, 64) range(2, 50) range(10, 2000) 42
GCN	learning_rate dropout weight_decay hidden_layers activation_function GCNConv epochs random_state	range(0.0001, 0.5) range(0.02, 0.2) range(0.001, 0.2) range(5, 500) logistic,tanh,relu range(2, 64) range(10, 2000) 42

MEF (Eq. 4.9) e ROV (Eq. 4.11) e MSE (Eq. 4.10) para os dados de validação, de modo a obter um resultado preliminar da qualidade do modelo na generalização. O cálculo de métricas na etapa de validação foi feito para todos os modelos de aprendizado de máquina utilizados neste estudo. Essas métricas já foram utilizadas na literatura para estimativas de fluxos de NEE utilizando aprendizado de máquina (TRAMONTANA et al., 2016; BAUER, 2022; ZHURAVLEV et al., 2022; JUNG et al., 2020).

Eficiência do Modelo (MEF): De acordo com Tramontana et al. (2016), o MEF é uma medida da capacidade de um modelos estimar uma variável alvo melhor do que uma referência, que é o valor médio das observações. Quando $MEF > 0$, a capacidade preditiva do modelo foi melhor do que a média ($MEF = 1$ para o modelo ideal); se $MEF = 0$, a capacidade preditiva do modelo foi equivalente à média e finalmente, se $MEF < 0$, a capacidade preditiva do valor médio da variável alvo foi melhor do que a do modelo . Seu cálculo é dado por 4.9.

$$MEF = 1 - \frac{\sum_{i=1}^n (x_i - y_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (4.9)$$

onde y_i é o valor observado, x_i é o valor previsto, \bar{y} é a média dos valores observados e n é o número de observações.

Erro Quadrático Médio (MSE): É uma medida comum de precisão para avaliar a qualidade das previsões de um modelo em relação aos valores reais. Ele é calculado como a média dos quadrados das diferenças entre os valores previstos e os valores observados. O MSE é uma métrica que fornece uma medida do quanto os valores previstos estão distantes dos valores reais. Quanto menor o valor do MSE, melhor é a capacidade do modelo em fazer previsões precisas.

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - x_i)^2 \quad (4.10)$$

onde y_i é o valor observado e x_i é o valor previsto e n é o número total de observações.

Variabilidade Observada Relativa (ROV): O ROV é uma medida da proporção da variância dos valores previstos em relação à variância dos valores observados. Ele indica o quão bem o modelo é capaz de capturar a variação dos dados observados. Quanto mais próximo de 1 for o valor do ROV, melhor será o desempenho do modelo em relação à variação dos dados observados.

$$ROV = \frac{\sigma_x}{\sigma_y} \quad (4.11)$$

onde σ_x é o desvio padrão dos valores previstos e σ_y é o desvio padrão dos valores observados.

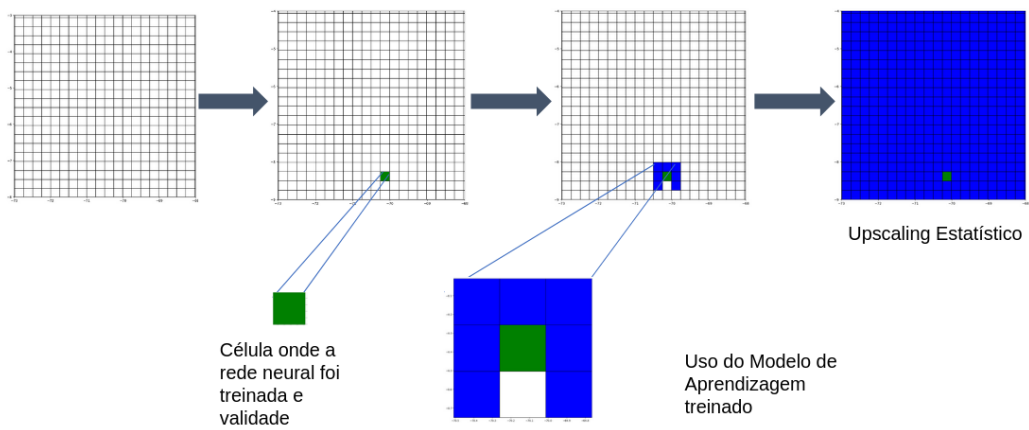
Após identificar os modelos com melhor desempenho durante as etapas de treinamento e validação, eles foram selecionados para a fase de teste. Essa etapa envolveu o uso de dados não vistos durante o treinamento e validação, especificamente na Célula K67, onde o modelo foi treinado e validado, e na Torre K34, localizada na Célula K34. As mesmas métricas utilizadas na validação foram calculadas para avaliar o desempenho dos modelos no conjunto de teste.

4.4 Ampliação de escala

Após a seleção dos modelos mais promissores, optou-se pelo modelo que demonstrou as melhores métricas de teste para o processo de ampliação de escala. Este modelo foi então aplicado às células K34 e K67 para estimar as séries temporais de NEE. Posteriormente, após identificar o modelo mais adequado para a ampliação de escala, foi feita a sua aplicação, como exemplificado na Figura 4.7. Utilizando este modelo, foi realizada a estimativa do NEE para cada célula de grade no bioma Amazônia.

Para identificação do comportamento de NEE no Bioma Amazônia foi calculado os valores médios, mínimos e máximos, desvio padrão e variância das séries temporais de NEE, além da tendência de Mann Kendall. Esses foram utilizados para entender o comportamento geral estatístico e temporal do NEE, como foi feito anteriormente nas etapas anteriores deste trabalho. O mapa de NEE criado foi então comparado com os dados mensais de NEE gerados por Jung et al. (2020) para os produtos globais da iniciativa FLUXCOM com resolução espacial de 0,08°.

Figura 4.7 - Método de ampliação de escala.



5 RESULTADOS E DISCUSSÃO

Uma vez que os dados do modelo de reanálise substituíram os dados meteorológicos medidos pelas torres para cobrir toda área de estudo, se fez necessário a avaliação da similaridade dentre os dados do modelo e os dados medidos. Na seção a seguir foram feitas as análises intercomparativas para certas variáveis comuns em cada conjunto de dados.

5.1 Análise exploratória dos dados: intercomparação temperatura do modelo ERA5 versus temperatura das torres K67 e K34

O conjunto METEO foi criado predominantemente a partir dos atributos do produto ERA5, por isso, para seu uso, foi feita a comparação entre as séries temporais de temperatura do ar do produto ERA5 das células K34 e K67 com suas respectivas torres da iniciativa LBA, K34 e K67. A comparação foi feita a partir das métricas estatísticas MAE, RMSE, BIAS e R da Tabela 5.1, onde foi visto uma alta correlação entre as temperaturas medidas nas torres e suas respectivas células do ERA5, sendo que a série temporal de temperatura da torre K34 se apresenta com menor BIAS e, assim, maior proximidade dos valores de temperatura do produto ERA5, como pode ser visto nas séries temporais da Figura 5.2.

Tabela 5.1 - Métricas de Intercomparação entre as temperaturas 't2m' do modelo ERA5 e 'ta' das Torres K67 e K34.

Célula/Torre	MAE	RMSE	BIAS	R	R ²
K67	1,74	1,88	1,73	0,82	0,67
K34	0,76	0,95	0,38	0,83	0,69

Notou-se, tanto pela Figura 5.1 como pela Tabela 5.1, que os valores de temperatura do ar do modelo ERA5 são superestimados em comparação aos dados medidos nas duas torres, principalmente na K67. Tal comportamento pode se repetir em outras células da grade do Bioma Amazônia, o que pode causar uma certa incerteza nas estimativas de NEE com base nesses dados, além de possivelmente dificultar a generalização dos modelos.

Figura 5.1 - Comparação da temperatura do ar a 2m (t_{2m}) do ERA5 e temperatura do ar (t_a) da Torre K67.

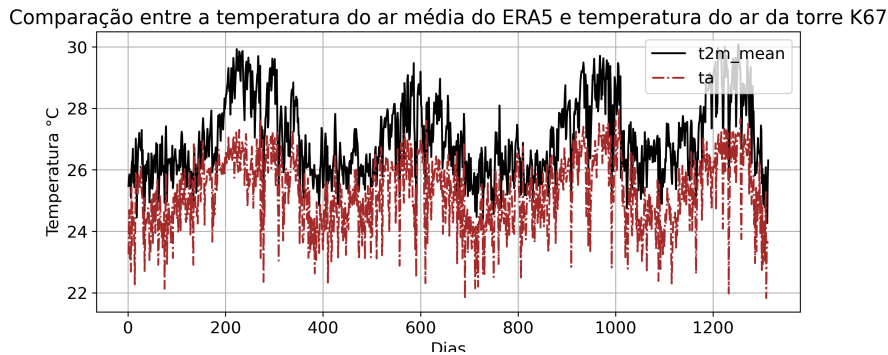
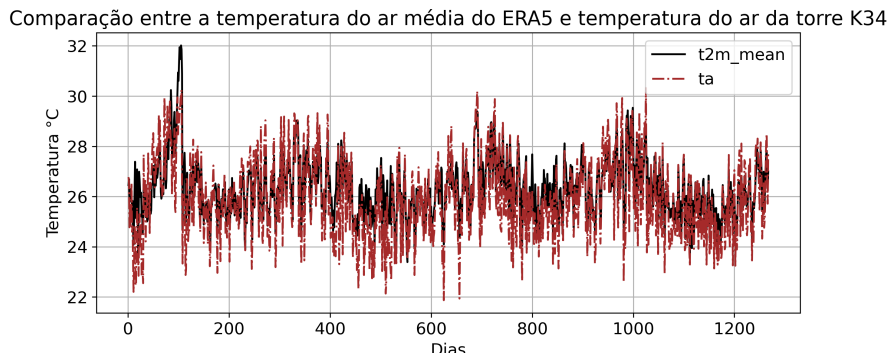


Figura 5.2 - Comparação da temperatura do ar a 2m (t_{2m}) do modelo ERA5 com a temperatura do ar (t_a) da Torre K34.



5.2 Análise exploratória dos dados: distribuição dos conjuntos de dados preditores

5.2.1 Distribuição dos dados dos atributos do conjunto METEO nas células K67 e K34

Foi identificado a partir do Teste Shapiro Wilk que os conjuntos METEO das células K34 e K67 possuem atributos com comportamento não normal, a Tabela 5.2 mostra o resultado do teste para cada um dos atributos e suas respectivas células. A maioria dos resultados do teste aproximou-se do valor 1, no entanto seus valores p foram menores que o nível de significância 0,05, o que causa a rejeição da hipótese de normalidade.

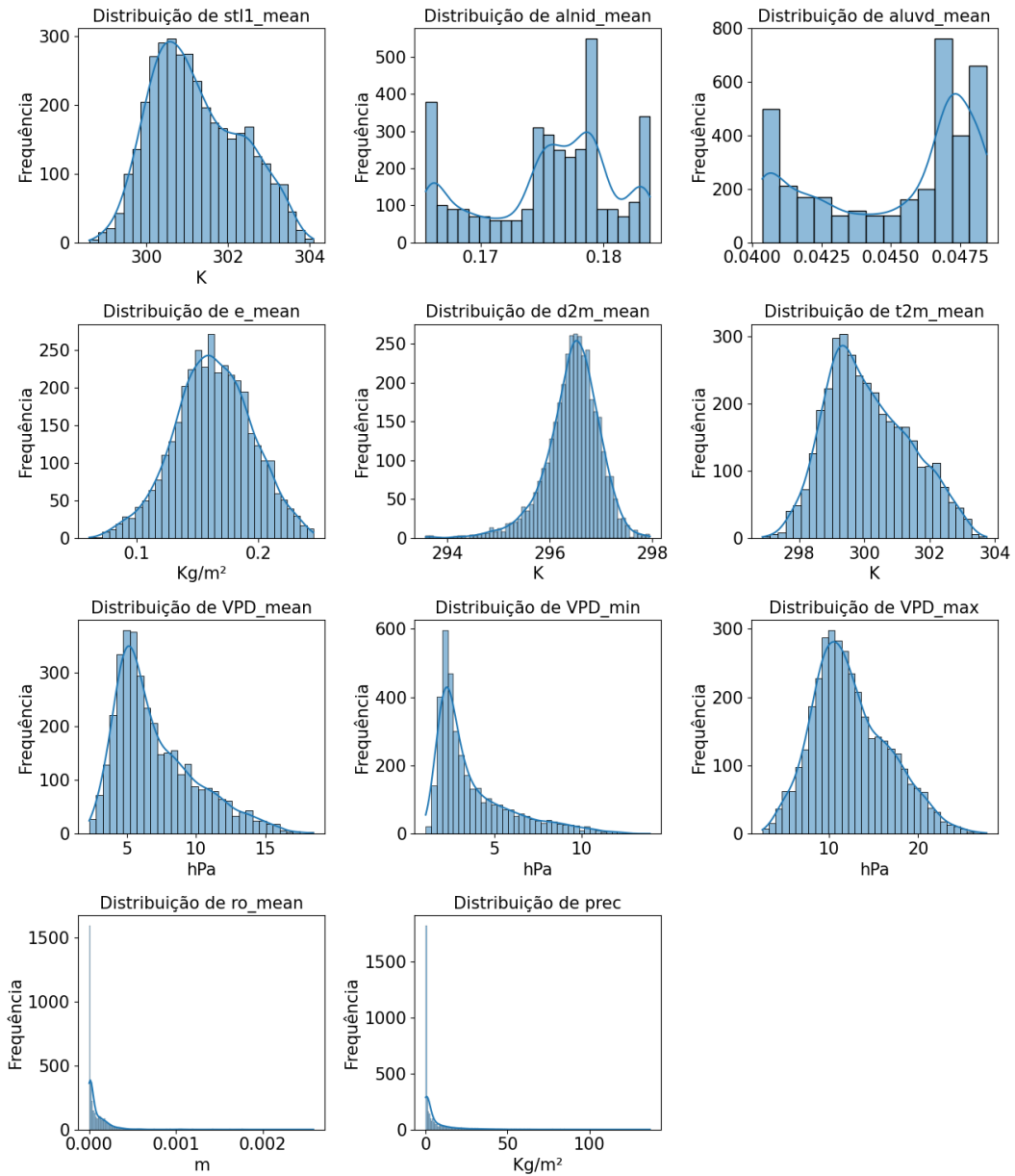
Os atributos VPD , t_{2m} , $stl1$, $prec$ e ro , nos histogramas 5.3 e 5.4, mostraram deslocamento à esquerda e valores de simetria positivos na Tabela 5.2, $prec$ e ro obtiveram os maiores valores de assimetria nos dois conjuntos. Esse comportamento indica que a maioria desses dados possuem valores baixo de magnitude ao se comparar à toda

a distribuição. O comportamento inverso foi identificado para os atributos *aluvd*, *alnid*, *d2m* e *e*, os quais possuem valores negativos de simetria e deslocamento à direita, indicando a maior quantidade de valores com mais alta magnitude.

Tabela 5.2 - Métricas estatísticas dos atributos do conjunto METEO para as células K67 e K34. A descrição dos atributos pode ser encontrada na Tabela 3.4.

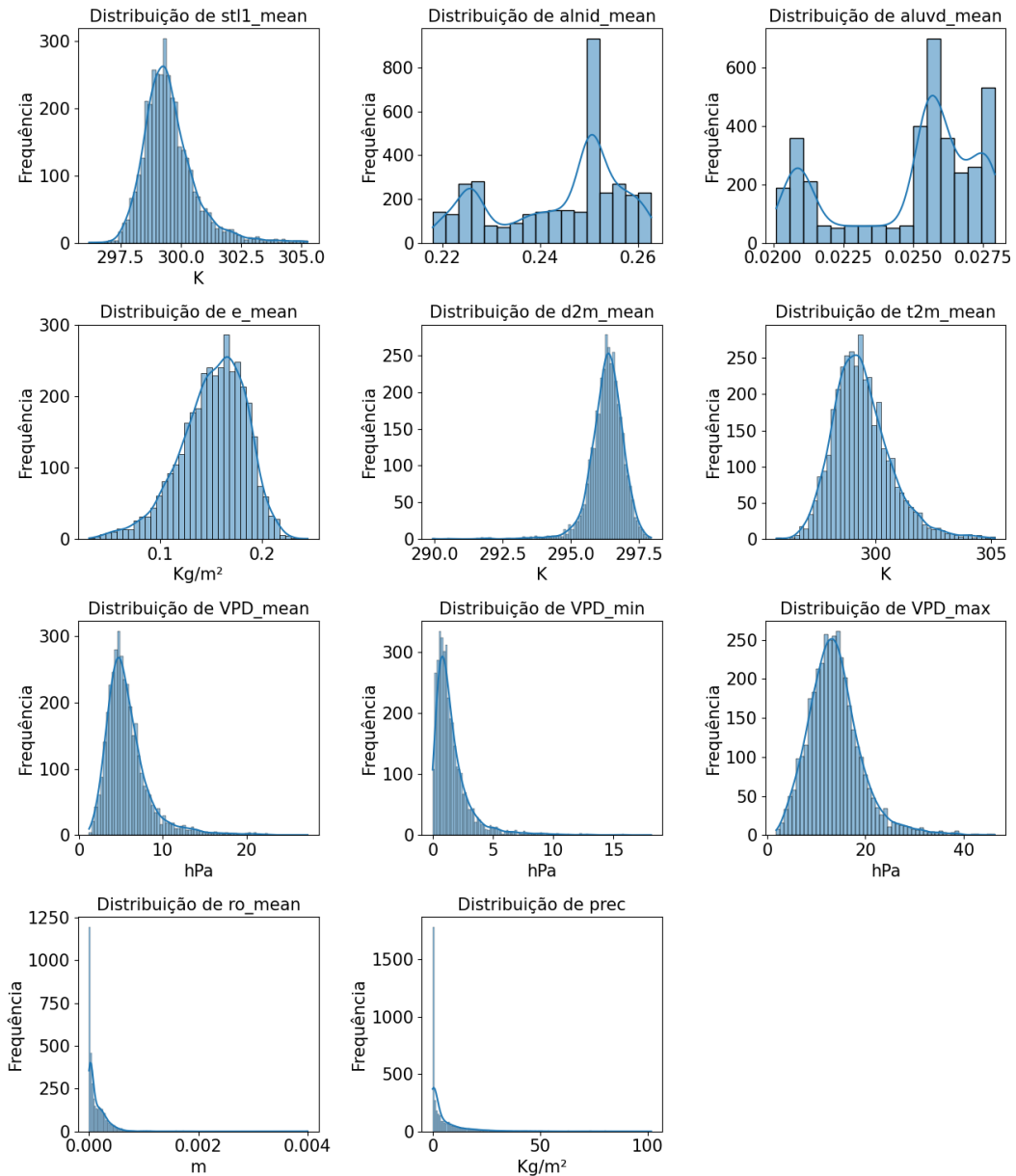
Atributo	Shapiro Wilk		<i>valor p</i>		Sk		CV	
	K67	K34	K67	K34	K67	K34	K67	K34
Torre								
t2m_mean	0.98	0.96	1.22E-22	1.58E-29	0.43	0.32	39.03	34.13
stl1_mean	0.97	0.92	1.12E-24	8.22E-41	0.50	0.46	41.37	32.86
aluvd_mean	0.85	0.86	0	0	-1.44	-0,88	58.68	50.93
alnid_mean	0.93	0.91	1.99E-38	9.25E-43	-0,52	-1.44	52.71	49.51
e_mean	0.99	0.98	9.77E-05	2.53E-23	0,02	-0.31	31,09	26,37
ro_mean	0.47	0.52	0	0	1.10	1.00	199.71	165.92
d2m_mean	0.96	0.90	7.83E-31	1.68E-43	-0,27	-0,23	18.64	9.86
vpd_mean	0.92	0.81	9.99E-41	0	0.89	0.65	60.74	61.50
prec	0.51	0.57	0	0	1.20	1.26	214.63	189.18

Figura 5.3 - Gráficos de distribuições das médias diárias dos atributos do conjunto METEO para a Célula K67.



Histogramas das variáveis do conjunto METEO da Célula K67.

Figura 5.4 - Gráficos de distribuições das médias diárias dos atributos do conjunto METEO para a Célula K34.



Histogramas das variáveis do conjunto METEO da Célula K34.

5.2.2 Distribuição dos dados dos atributos do conjunto RS nas células K67 e K34

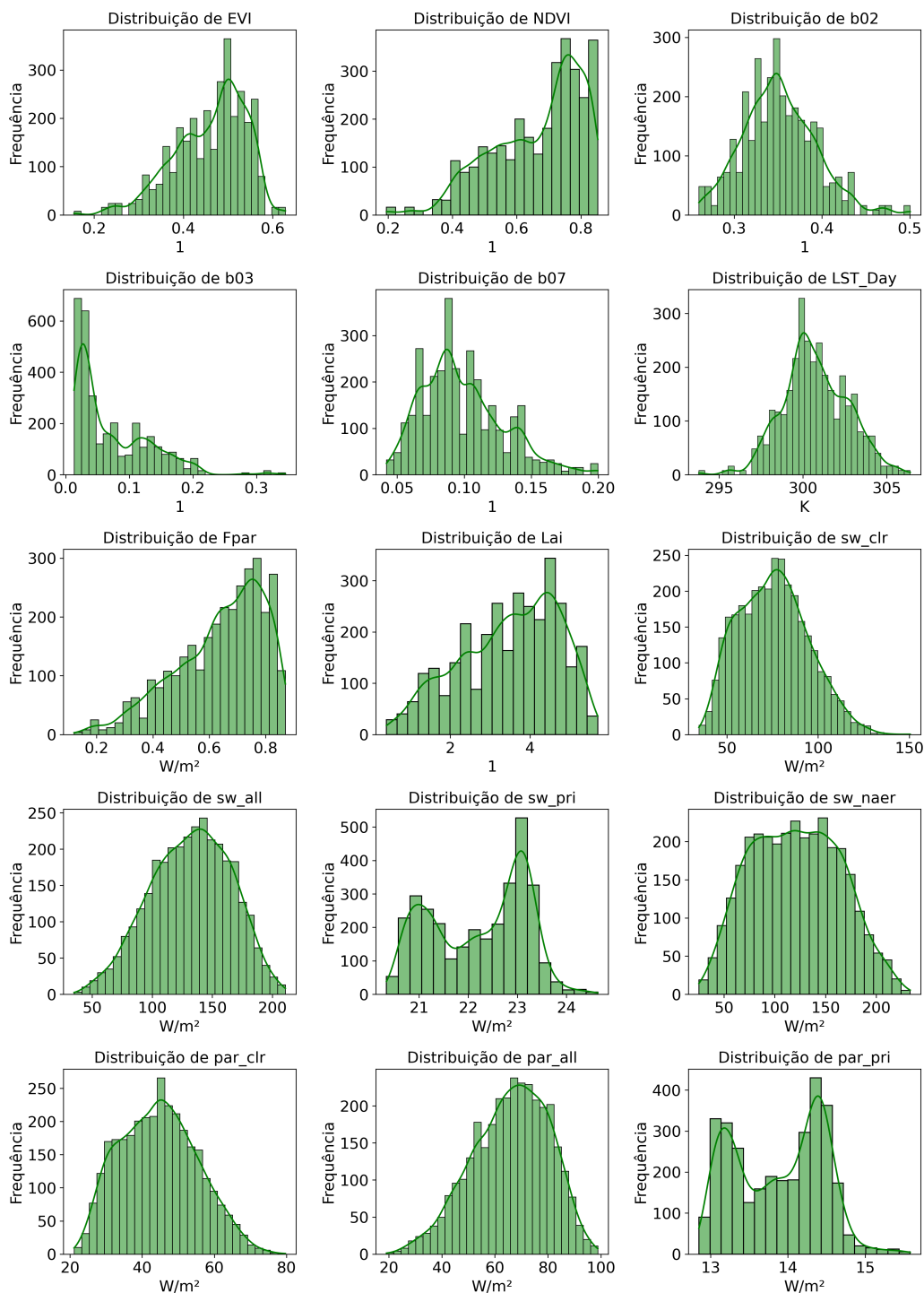
Foi visto na Tabela 5.3 que os atributos de RS das duas células no Teste de Shapiro Wilk obtiveram valores bem próximos de 1 e significância menor que 0,05, rejeitando a hipótese de normalidade.

As variáveis que se referem à cobertura da vegetação e atividade fotossintética, o EVI, NDVI, Lai e Fpar mostraram, nos histogramas das Figuras 5.5 e 5.6, deslocamento à direita e coeficientes de simetria negativos, indicando que a distribuição possui valores de maior magnitude mais frequentes. Com exceção do EVI, as outras variáveis citadas mostraram maior assimetria e conseqüente maior frequência em valores de maior magnitude na célula K34 do que na célula K67, o que possivelmente ocorre devido à maior presença da classe de Formação Florestal na mesma. Esses desbalanceamentos de dados pode ser um problema em modelos de aprendizado de máquina, onde o treinamento pode ficar enviesado, o que pode trazer desempenho insatisfatório para a ampliação de escala

Tabela 5.3 - Métricas estatísticas dos atributos do conjunto RS para a célula K67. A descrição dos atributos pode ser encontrada nas Tabelas 3.2 e 3.3.

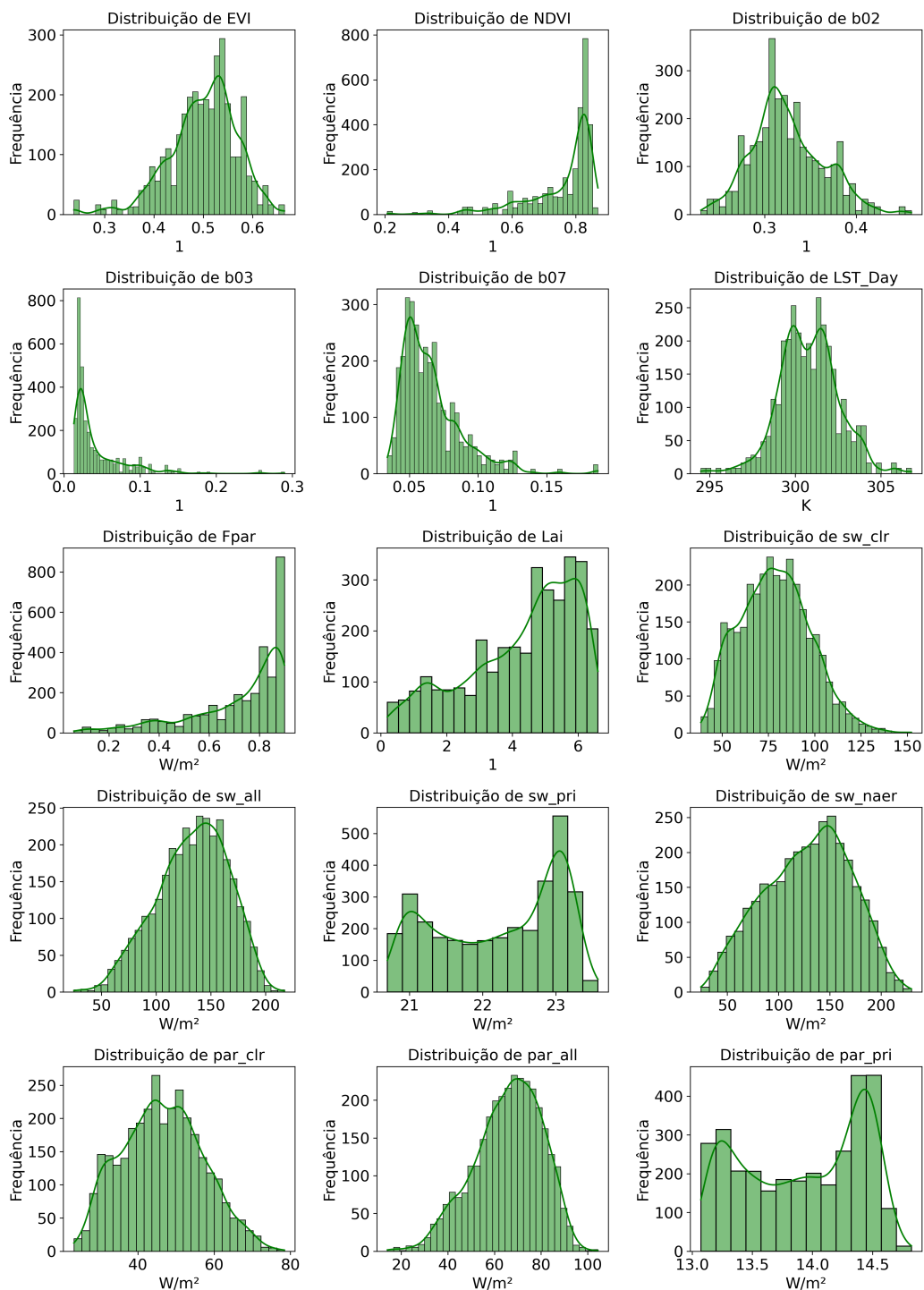
Atributo	Shapiro Wilk		<i>valor p</i>		Sk		CV %	
	K67	K34	K67	K34	K67	K34	K67	K34
Torre	K67	K34	K67	K34	K67	K34	K67	K34
EVI	0.96	0.97	1.25E-28	3.85E-25	-0,76	-0,42	26.29	26.38
NDVI	0.93	0.77	3.50E-37	0	-0,86	-1.43	29.76	19.97
b02	0.99	0.98	6.61E-17	1.92E-18	0.18	0.36	45.73	41.63
b03	0.85	0.69	0,0	0	1.43	1.34	99.06	125.74
b07	0.96	0.86	7.71E-28	0	0.71	0.70	54.00	69.66
LST_Day	0.99	0.99	2.18E-09	2.43E-13	0.24	0,02	27.91	27.47
Fpar	0.94	0.83	9.29E-34	0	-0,59	-1.13	30,07	29.58
Lai	0.96	0.92	1.66E-27	3.64E-38	-0,29	-0,71	40,05	40.37
sw_clr	0.99	0.99	2.81E-16	1.59E-14	0,07	0,09	46.41	46.92
sw_all	0.99	0.99	2.47E-12	8.20E-16	-0,15	-0,26	34.32	30.16
sw_pri	0.93	0.90	1.35E-36	6.10E-41	-0,65	-0,62	51.34	54.64
sw_naer	0.99	0.99	1.61E-17	2.39E-17	0,02	-0,23	45.44	40,05
par_clr	0.99	0.99	1.68E-15	1.13E-13	0,04	0,09	45.42	45.39
par_all	0.99	0.99	1.24E-16	1.29E-18	-0,24	-0,28	30.89	27.83
par_pri	0.94	0.91	5.17E-34	1.49E-39	-0,40	-0,40	55.69	58.00

Figura 5.5 - Gráficos de distribuições das médias diárias dos atributos do conjunto RS para a Célula K67.



Histogramas das variáveis do conjunto RS da Célula K34.

Figura 5.6 - Gráficos de Distribuições das médias diárias dos atributos do conjunto RS para a Célula K34.



Histogramas das variáveis do conjunto RS da Célula K34.

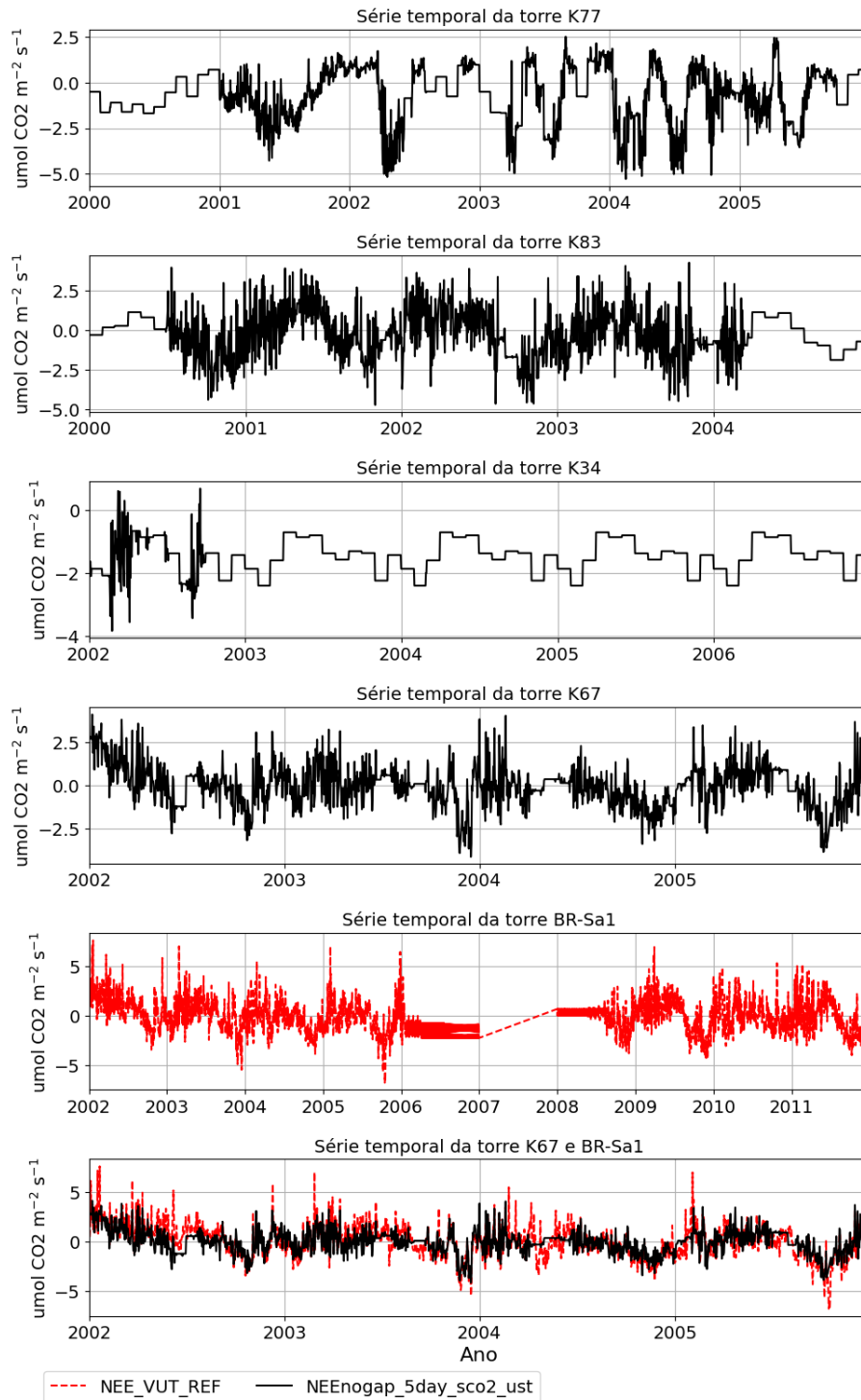
O comportamento das variáveis que se referem ao comportamento de refletância, o *b02*, *b03* e *b07* obtiveram coeficientes de simetria positivos e deslocamento à esquerda, com baixos valores predominantes para as duas células. As variáveis de radiação difusa obtiveram comportamentos semelhantes nas duas células em relação às métricas utilizadas nesta etapa.

De um modo geral o comportamento dos atributos das duas células foram muito semelhantes, a célula K67 mostrou-se com maior assimetria e variação em relação à média em todos os atributos, exibindo possivelmente maior variação em seu comportamento ambiental do que na célula K34.

5.3 Análise exploratória dos dados: variabilidade temporal do NEE

Nesta seção são apresentadas as tendências de longo prazo, as séries temporais e os ciclos anuais dos atributos de NEE provenientes das bases Fluxnet e LBA para as torres K67, K83, K77 e K34. Foram discutidas semelhanças e diferenças entre os ciclos anuais, além de ressaltar períodos de longa duração (vários meses) em que houve preenchimento de dados faltantes. Esta análise foi importante para caracterizar as diferentes versões da variável NEE, que foi utilizada como variável alvo nos modelos de aprendizado de máquina.

Figura 5.7 - Séries Temporais de NEE para as Torres K77, K67, K83 e BrSa1. E Gráfico de Comparação entre as séries K67 e BrSa1 ao fim da Figura.



Em vermelho há a variável NEE_VUT_REF do dados Fluxnet e em preto a variável NEEenogap_5day_sco2_ust das Torres do LBA.

Como dito anteriormente na Seção 3.2 as variáveis alvo *NEE_VUT_REF* e *NEEnogap_5day_sco2_ust* são provenientes de duas bases de dados que passaram por diferentes processamentos e ajustes, e isso inclui diferentes métodos de preenchimento de dados para melhora de estimativas de dados de baixa qualidade ou substituição de dados ausentes. Na Figura 5.7 há as séries temporais diárias de cada uma das torres utilizadas neste trabalho e é possível verificar graficamente o preenchimento de dados aplicados e quais anos foram impactados com o mesmo.

A maior série temporal é a da Torre BrSa1 da base do Fluxnet e nos anos de 2006 e 2008 as séries temporais diárias de *NEE_VUT_REF* (em vermelho) foram preenchidas pela técnica de Amostragem da Distribuição Marginal (PASTORELLO et al.,), tal preenchimento pôde ser identificado pelo comportamento de oscilação constante na Figura 5.7. Já no ano de 2007 não houve dados disponíveis para a série. Os valores ausentes de *NEEnogap_5day_sco2_ust* na base de dados LBA foram preenchidos pelo comportamento do ciclo anual médio da torre e o mesmo foi identificado nos gráficos da Figura 5.7 pelo comportamento temporal constante que pode ser visto em trechos das séries temporais. A torre K34 mostrou mais de 91% dos dados preenchidos, a torre K83 mostrou 27% de preenchimento a partir de março de 2004 a 2005 e a torre K77 mostrou 17% de preenchimento em diferentes períodos da série. A torre que mostrou a menor quantidade de dados preenchidos é a K67, com cerca de 7%. Esses preenchimentos podem afetar significativamente os processos de aprendizagem dos modelos, gerando um viés na aprendizagem em que os algoritmos se adaptam ao comportamento do ciclo médio anual e se especializam nele. Ocasionalmente assim dificuldades nas generalizações dos modelos em locais de maior variabilidade temporal.

Tabela 5.4 - Tendência de Mann Kendall de cada série temporal das variáveis NEE de suas respectivas Torres.

Atributo	Torre	Tendência MK	valor p	Tau
<i>NEE_VUT_REF</i>	BrSa1	decrecente	5.11E-15	-0,09
<i>NEEnogap_5day_sco2_ust</i>	K34	não	0.25	0,02
<i>NEEnogap_5day_sco2_ust</i>	K77	crescente	7.13E-06	0,06
<i>NEEnogap_5day_sco2_ust</i>	K67	decrecente	2.35E-11	-0,12
<i>NEEnogap_5day_sco2_ust</i>	K83	decrecente	8.13E-11	-0,10

Foi aplicado o teste de Mann Kendall nas séries temporais de NEE dos conjuntos LBA e Fluxnet. Na Tabela 5.4 foi identificada uma tendência significativa de dimi-

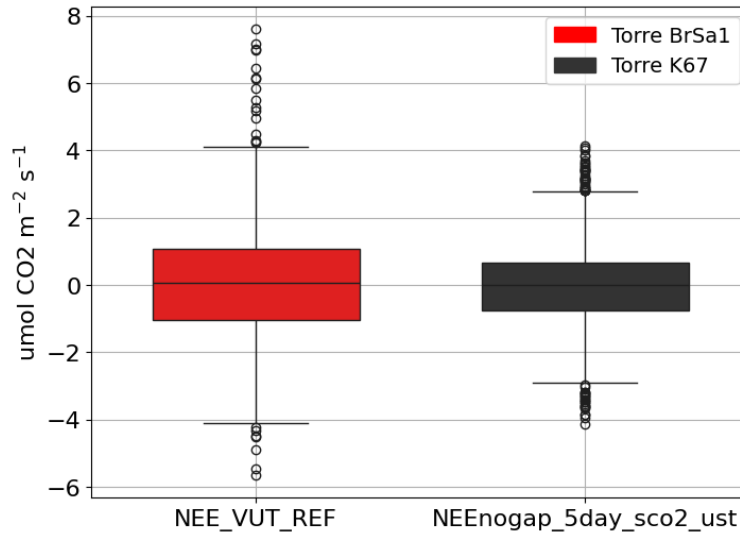
nuição do NEE nas torres K83 e K67 entre 2002 e 2005, o que pode ser interpretado como um aumento na absorção de CO₂ pela floresta. As duas torres se encontram em regiões de floresta, sendo que a primeira está situada em um local onde há exploração seletiva de madeira e a segunda está em uma área de conservação (SAKAI et al., 2004). Essa tendência também foi encontrada para a variável NEE da Fluxnet da torre BrSa1 de 2002 a 2011. Por outro lado, para a torre K77, localizada em uma área de campo agrícola, foi encontrada uma tendência crescente, ou seja, uma tendência de aumento de emissão de CO₂. No caso da torre K34, não houve tendência estatisticamente significativa, provavelmente devido a suavização dos dados decorrente do preenchimento existente nos mesmos. As tendências encontradas para as quatro séries temporais são fracas, como pode ser visto pelo valor de *Tau* na Tabela 5.4 e a mesma também pode ser afetada pelos preenchimentos citados anteriormente.

As séries temporais BrSa1 e K67 possuem a maior quantidade de dados em boa qualidade. Considerando que elas possuem diferentes padronizações e foram originadas da mesma torre de medição, foi visto que as mesmas obtiveram a mesma tendência e valores de *Tau* próximos, mostrando uma semelhança no comportamento temporal das duas séries. Essa situação pode indicar uma consistência temporal da tendência, de modo com que ela se mantém ao decorrer do tempo, já que a série de BrSa1 é de 2002 a 2011 e a de K67 é de 2002 a 2005.

Tabela 5.5 - Estatísticas dos Atributos de NEE da Fluxnet (BrSa1) e LBA (K67) e de comparação entre as séries temporais do Fluxnet BrSa1 e LBA K67.

LBA	<i>NEEnogap_5day_sco2_ust</i>		
Fluxnet	<i>NEE_VUT_REF</i>		
<i>NEE_VUT_REF_QC</i>	> 0	>0,7	= 1
R	0,73	0,77	0,81
RMSE	1,12	1,09	1,09
MAE	0,82	0,79	0,76
BIAS	0,09	0,15	0,30
tamanho (dias)	3287	2256	867

Figura 5.8 - Boxplots mostrando a distribuição dos atributos de NEE das bases Fluxnet BrSa1 e LBA K67.



Para o melhor entendimento do comportamento e diferenças entre os dados das Torres BrSa1 e K67, foram calculadas as estatísticas para as séries temporais no período de 2002 a 2006, representadas nas Tabelas 5.5 e a partir da análise gráfica da Figura 5.8.

Foi visto na Figura 5.7 que as séries temporais das variáveis escolhidas de K67 (LBA) e BrSa1 (Fluxnet) não são idênticas. Como esperado, os processamentos e ajustes realizados durante a padronização dos dois conjuntos de dados geraram diferenças entre as variáveis. A variável *NEE_VUT_REF* possui maior desvio-padrão e seus valores mínimos e máximos diferem da variável do LBA (Tabela 5.6), tal comportamento pode impactar na criação dos modelos e suas intercomparações. Para entender se mesmo com essas diferenças o comportamento geral delas é semelhante, foram calculadas métricas de comparação na Tabela 5.5. Notou-se que os comportamentos gerais das duas séries divergem entre si e foi possível afirmar estatisticamente que toda a série temporal possui comportamento geral discrepante.

Considerando que os atributos *NEE_VUT_REF* e *NEEnogap_5day_sco2_ust* são estatisticamente diferentes. Quando utilizados como variável alvo nos modelos de aprendizado de máquina, os mesmos tentariam estimar esse comportamento divergente. Com isso, provavelmente, um modelo treinado com dados de BrSa1 terá dificuldades em reproduzir os valores da K67 na etapa de generalização. No entanto, ao utilizar dados com qualidade acima de 0,7, a correlação e as métricas de comparação

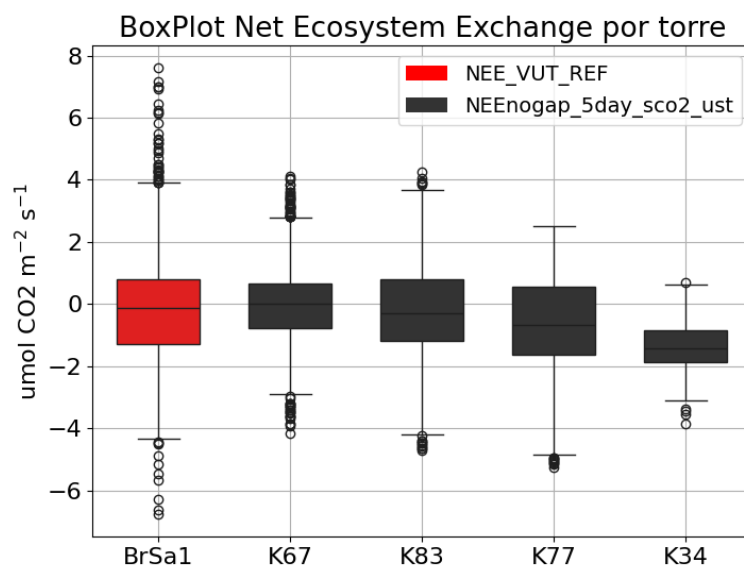
entre as séries melhoram. Desse modo, quanto melhor a qualidade do dado de BrSa1, melhor seria a generalização dos modelos.

Pode ser visto na Tabela 5.5 os valores estatísticos de comparação entre as séries, como a correlação Pearson (R), RMSE, MAE, BIAS, além do tamanho da população analisada. Foi identificado que à medida que há a diminuição da quantidade de dados preenchidos, ou seja, há a melhoria da qualidade dos dados a partir do filtro de NEE_VUT_REF_QC, há também uma melhoria geral no desempenho das séries temporais. Tal comportamento foi evidenciado pelo aumento da correlação entre as séries e a redução dos erros (RMSE e MAE). Foi também observado um aumento do BIAS à medida que a qualidade dos dados aumenta, sugerindo uma tendência de superestimação dos valores de NEE com o aumento da qualidade dos dados. O uso de dados de boa qualidade é importante para uma boa estimativa pelos modelos de aprendizado de máquina. Contudo, para melhorar a qualidade, é necessário filtrar os dados, o que causa grande perda de informação, o que também impacta o modelo.

Tabela 5.6 - Estatísticas dos Atributos de NEE das Torres do LBA e Fluxnet.

Torre	Atributo	média	min	max	med	std
K67	NEEnogap_5day_sco2_ust	-0,03	-4.15	4.13	0,01	1.24
K77	NEEnogap_5day_sco2_ust	-0,71	-5.27	2.49	-0,62	1.45
K83	NEEnogap_5day_sco2_ust	-0,25	-4.72	4.25	-0,31	1.43
K34	NEEnogap_5day_sco2_ust	-1.46	-3.84	0.69	-1.42	0.54
BrSa1	<i>NEE_VUT_REF</i>	-0,16	-6.77	7.60	-0,11	1.59

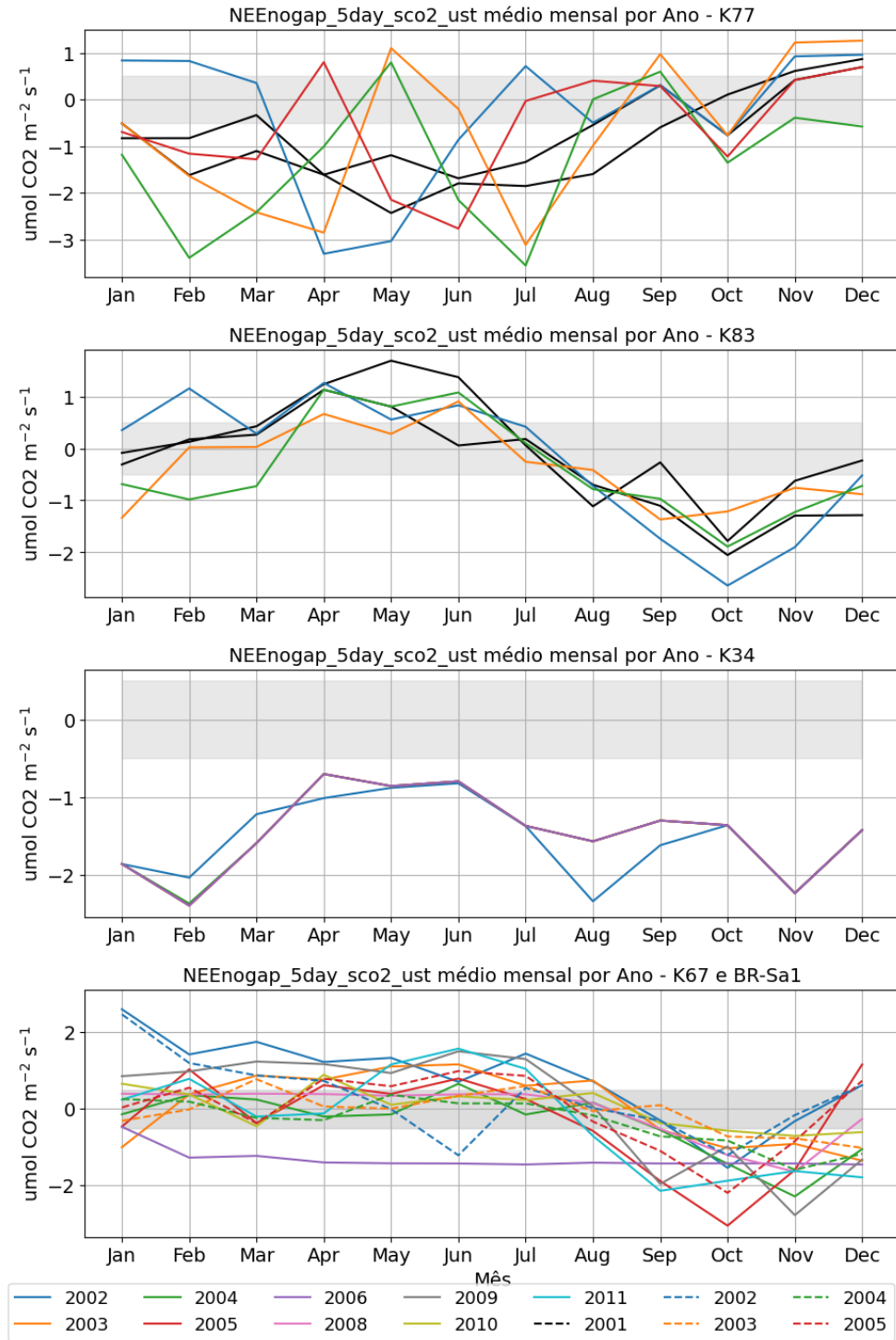
Figura 5.9 - BoxPlot das variáveis NEE de cada uma das torres.



Nesta figura há a análise de boxplot dos dados da série temporal de cada uma das torres. Em vermelho está os dados da torre BrSa1 do conjunto Fluxnet e em preto há os dados das Torres K6, K83, K77 e K34 do conjunto LBA.

Foram realizadas análises dos comportamentos estatísticos e mensais das torres de fluxo utilizada neste estudo. Essas análises foram feitas com base na Figura 5.10, na Tabela 5.6 e no boxplot da Figura 5.9. Dentre as torres que se encontram na célula K67 (K67, K77, K83 e BrSa1), todas mostraram o comportamento de sumidouro como predominante, isso pôde ser verificado a partir da constância dos valores abaixo de zero nos gráficos mensais, pelos valores negativos dos valores médios e medianos na Tabela 5.6 e pelo comportamento exibido do boxplot na Figura 5.9. A torre K34 mostrou um comportamento semelhante, no entanto, devido ao preenchimento contínuo predominante em sua série temporal, não é possível afirmar que esse comportamento é verdadeiro.

Figura 5.10 - Gráfico dos comportamento mensal anual de NEE nas Torres K77, K67, K34, K83 e BrSa1.



A partir das métricas estatísticas de cada NEE na Tabela 5.6 foi possível identificar que a série temporal que possui maior variância em relação ao valor médio é o da torre BrSa1, com desvio-padrão igual a 1,59, o que é o esperado devido às maiores amplitudes (valores máximos e mínimos) de NEE encontradas nos dados do Fluxnet. O segundo maior é a série temporal K77, que possui também maior variação mensal, como é visto no gráfico 5.10. O que possui menor desvio padrão, bem abaixo das outras torres, é a série temporal da K34 com valor de 0,54, esse desvio padrão baixo é devido ao fato de que grande parte da série foi criada a partir do preenchimento de dados pelo ciclo anual médio.

O Gráfico 5.10 mostra a média do comportamento mensal por ano de todas as torres utilizadas neste trabalho. As séries temporais de BrSa1 e K67 no período de fevereiro à julho, estação úmida na Amazônia mostrou valores de NEE predominantemente positivos, indicando maior emissão de CO₂. Já na estação seca, no período de agosto à setembro, foi identificado valores predominante negativos, indicando maior absorção CO₂ pela floresta. Esse comportamento na estação seca pode ser decorrente de maior quantidade de radiação no período, o que resulta em uma maior atividade fotossintética. Restrepo-Coupe et al. (2013) comenta que essa parte da Amazônia não sofre grande estresse hídrico, por isso a absorção de carbono aumenta na estação seca da Amazônia. Outra possibilidade é que no período de ago-nov há maior presença de queimadas na região Sudeste da Amazônia, além de estresse hídrico na vegetação, ocasionando o aumento da quantidade de CO₂ no local da Torre K67. Uma vez que ocorre o aumento de CO₂ na atmosfera, tal aumento pode ocasionar aumento da atividade fotossintética a partir da maior absorção de CO₂ pelos estômatos no processo de fotossíntese (MARENCO et al., 2014).

A Torre K83 (atributo NEEenogap_5day_sco2_ust da base LBA) mostrou uma variabilidade sazonal semelhante ao da torre K67. O comportamento do NEE, no período de 2002 a 2004, da Torre K83 no gráfico da Figura 5.10 durante os meses de abril a julho foi de emissor de CO₂, havendo um decaimento da emissão após o mês de junho. Já durante os meses de julho a dezembro o comportamento foi predominantemente de sumidouro de CO₂.

Já a Torre K77 foi a que mostrou o comportamento mais distinto em relação às outras torres e em relação ao ciclo anual do NEE, como pode ser visto na Figura 5.10. Tal comportamento é determinado pelos diferentes tipos de cultura que são cultivadas em sua área, o que difere do que foi visto para as torres em áreas florestais. Durante os meses de julho a dezembro há um aparente aumento do NEE, sendo

que esse mostra a diminuição de absorção e aumento da emissão, onde a partir de novembro, com exceção do ano de 2004, o local funciona como emissor de CO₂. O comportamento da K77 é decorrente de atividade agrícola no local dessa torre, a preparação da terra, plantio, crescimento e colheita influenciam diretamente no NEE. Logo, o ciclo deixa de ser natural e passa a ser governado pela prática agrícola (SAKAI et al., 2004). O comportamento de emissão visto na análise mensal pode ser afirmado também devido à análise de tendência mostrada anteriormente, em que a K77 foi a única torre que obteve tendência de aumento de CO₂ em sua série temporal diária.

Durante os anos de 2003 e 2006, o comportamento médio mensal da torre K34 foi como sumidouro, como pode ser visto no gráfico 5.10 , onde a maior absorção de CO₂ ocorre entre os meses de julho à fevereiro e a menor absorção de CO₂ ocorre durante os meses de março e julho. Entretanto, não é possível afirmar que o comportamento é verdadeiro ou um efeito dos dados faltantes. Os anos de 2003 a 2006 possuem o mesmo comportamento mensal, isso ocorre devido ao preenchimento pelo ciclo anual médio a partir, principalmente, do ano de 2003 até 2006, por isso não é possível visualizar todas as curvas, pois elas estão sobrepostas no gráfico da K34 na Figura 5.10 , onde a partir de 2003 os valores diários anuais são os mesmos.

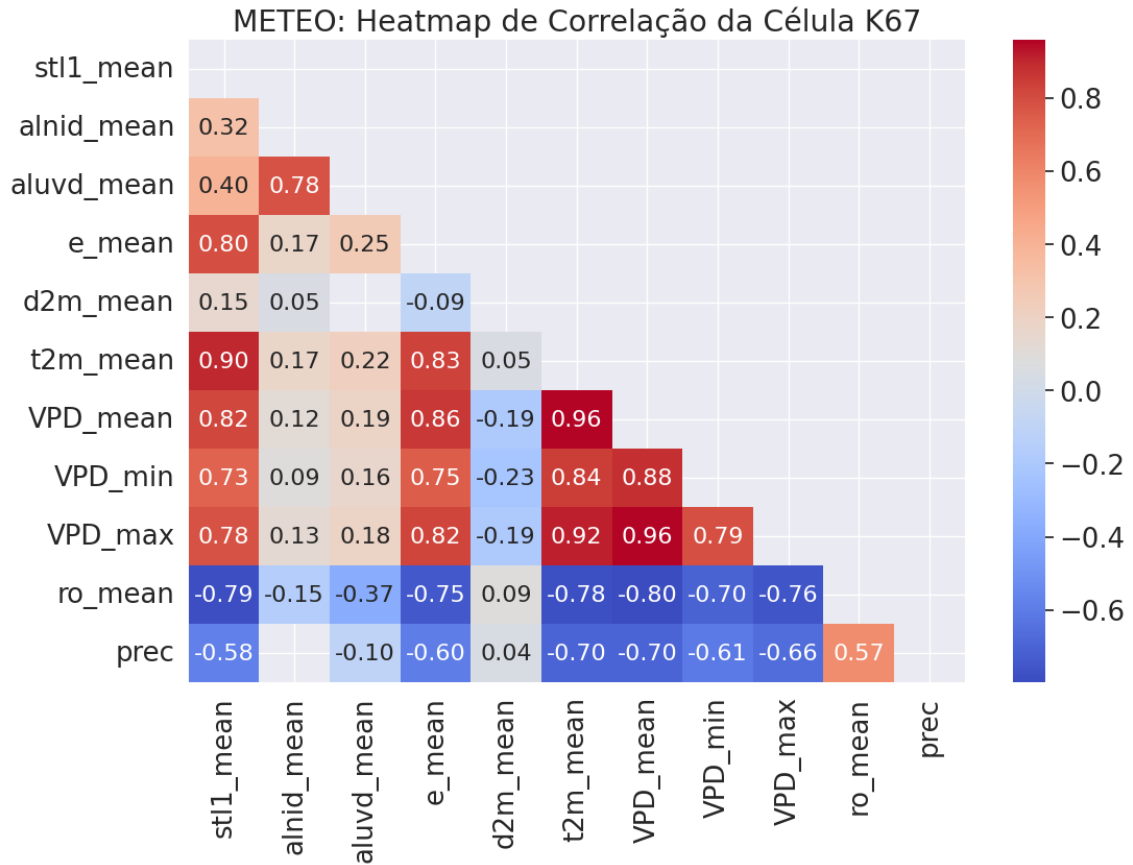
5.4 Análise exploratória dos dados: análise de correlação no conjunto de preditores METEO

A análise de correlação de preditores é importante para o entendimento das relações entre as variáveis independentes em um conjunto de dados. Ela permite a identificação de padrões de associação entre os preditores, destacando possíveis redundâncias ou interdependências. Além disso, a análise de correlação ajuda a selecionar os preditores mais relevantes para o modelo final, garantindo que apenas as variáveis mais informativas sejam incluídas, o que pode melhorar a eficiência e a precisão das previsões (DANASINGH et al.,). Contudo, cabe lembrar que essa correlação é síncrona, ou seja, no mesmo instante temporal. Análises de causa e efeito fogem ao escopo deste estudo, devido a sua complexidade e, conseqüentemente, custosa em termos de desenvolvimento, o que impacta no cronograma de entrega do trabalho.

Na seção anterior foi visto que todos os atributos preditores que foram utilizados neste trabalho possuem distribuição não normal. Diante disso, optou-se por empregar a correlação de Spearman para avaliar a relação entre esses atributos, uma vez que a mesma é não paramétrica (CONOVER, 1999). Essa análise fornece suporte na identificação e remoção de atributos altamente correlacionados, esses poderiam

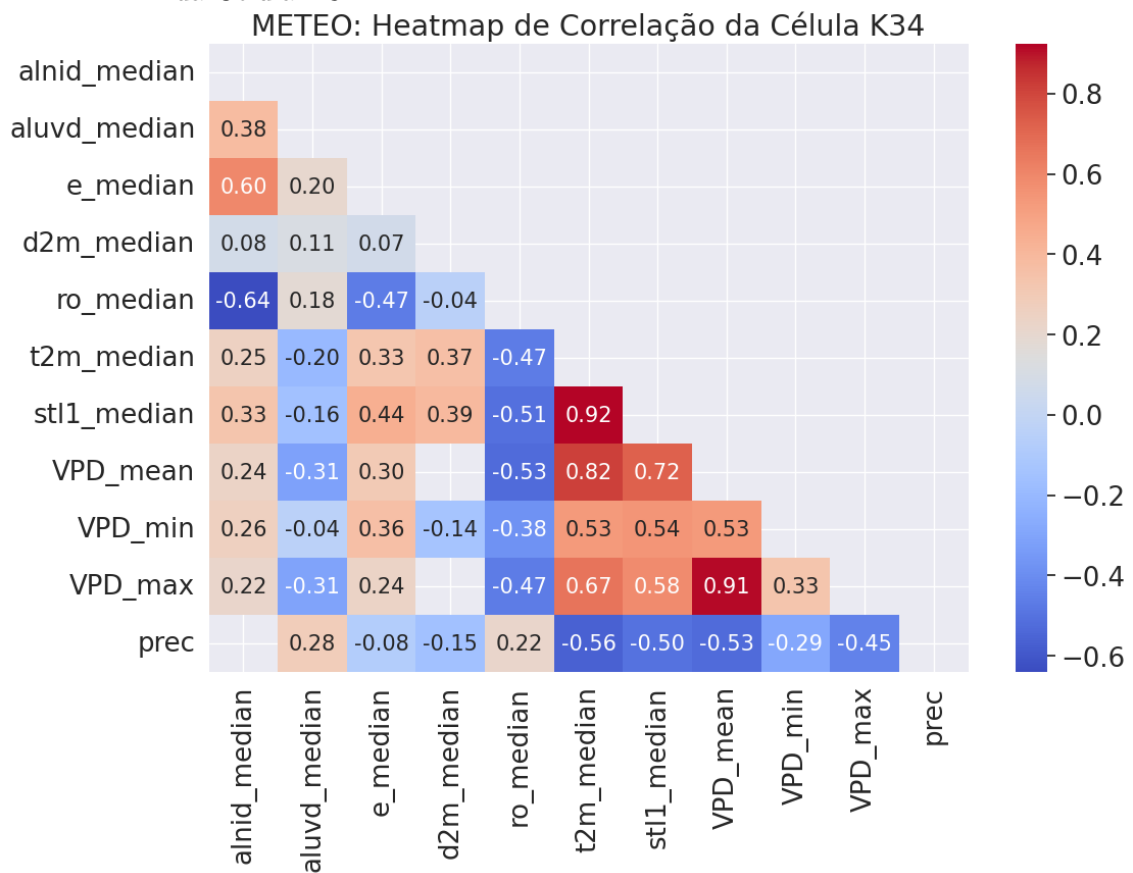
prejudicar o desempenho dos algoritmos de aprendizado de máquina deste trabalho (DANASINGH et al.,).

Figura 5.11 - Gráfico da matriz diagonal da correlação de Spearman dos atributos METEIO da Célula K67.



Matriz diagonal de correlação de Spearman entre as variáveis Predictoras do Conjunto METEIO para a Célula K67. As correlações exibidas possuem significância valor-p<0,05.

Figura 5.12 - Gráfico da matriz diagonal da correlação de Spearman dos atributos METEO da Célula K34.



Matriz diagonal de correlação de Spearman entre as variáveis preditoras do conjunto METEO para a Célula K34. As correlações exibidas possuem significância valor- $p < 0,05$.

As Figuras 5.11 e 5.12 exibem as matrizes de correlação diagonal para o conjunto de dados METEO nas células correspondentes. A partir dessas matrizes observou-se que os atributos que se remetem à disponibilidade e transporte de água no ambiente, a precipitação (prec) e o escoamento líquido (ro), possuem correlação positiva moderada entre si. Essa tendência é mais evidente na célula K67. Além disso, esses atributos mostram uma forte correlação negativa com a evaporação 'e'. Isso ocorre devido à interdependência dessas três variáveis no ciclo da água. De acordo com Nobre et al. (2021), cerca de 50% da precipitação na região Amazônica retorna para a atmosfera por meio da evapotranspiração (evaporação + transpiração da vegetação) e o restante contribui no escoamento líquido que abastece lençóis freáticos e contribui nas formações de rios e córregos da Bacia Amazônica.

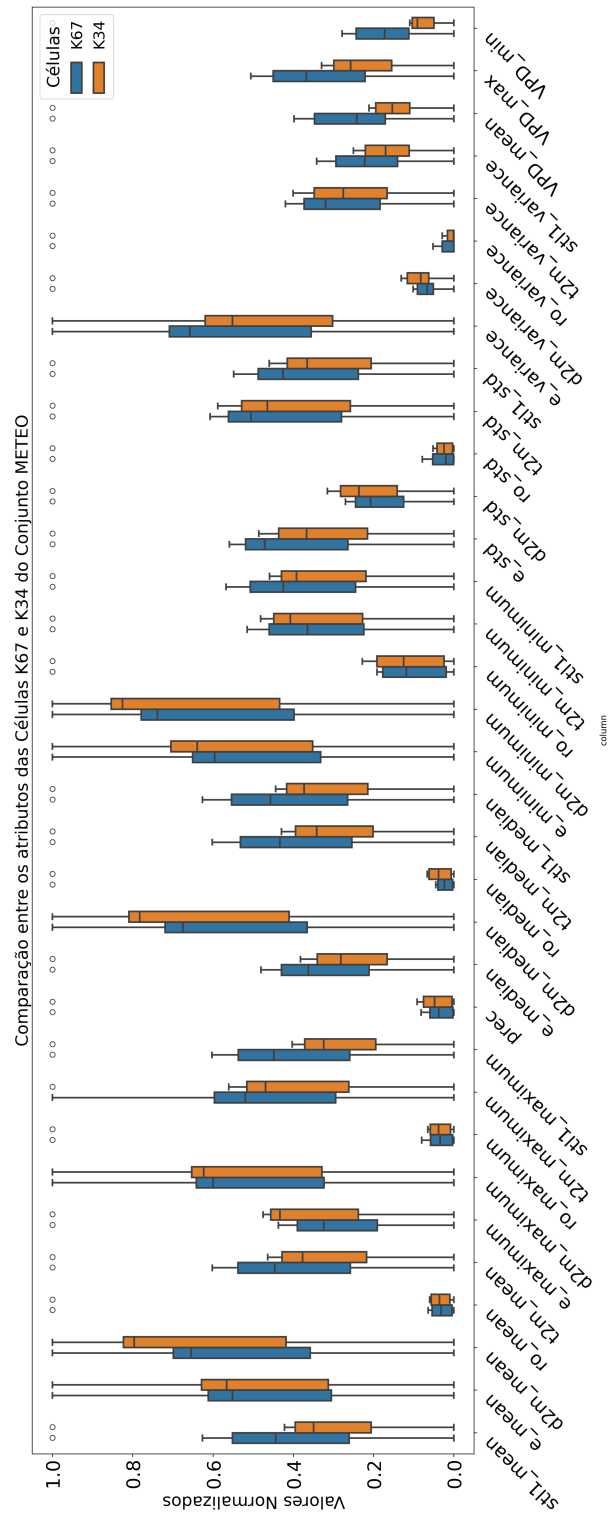
A evaporação, 'e' mostrou uma correlação positiva forte com as variáveis que remetem ao comportamento térmico, como a temperatura do ar (t2m), temperatura do solo (stl1) e ao Deficit de Pressão de Vapor (VPD), com maior intensidade na célula K67. A temperatura no ambiente exerce uma influência direta no ciclo da água em um ecossistema, regendo suas fases físicas. A quantidade de vapor d'água possível na atmosfera a uma determinada temperatura até chegar ao ponto de saturação, é determinada pelas temperatura do ar, temperatura do ponto de orvalho e a disponibilidade de água disponível no ambiente para a ocorrência de evaporação. O VPD nos fornece a capacidade de mudança de fase da água para vapor. E foi observado, em todas as análises realizadas o quão intrínseco é essa relação. Na célula K67 foi identificado uma tendência de aumento de todas as variáveis relacionadas ao VPD, indicando um aumento da evaporação na região.

Tabela 5.7 - Tendências avaliadas com o teste de Mann Kendall dos preditores do conjunto METEO nas Células K67 e K34.

Atributo	Tendência		valor-p		Tau	
	K67	K34	K67	K34	K67	K34
Célula						
stl1_mean	+	+	4.63E-02	1.13E-03	0,02	0,04
e_mean	+	não	6.50E-04	3.06E-01	0,04	-0,01
d2m_mean	não	-	2.43E-01	0,00	-0,01	-0,15
ro_mean	não	-	7.94E-01	1.01E-07	0,03E-01	-0,06
t2m_mean	+	não	2.72E-05	1.03E-01	0,05	0,02
e_max	não	+	6.81E-01	1.38E-03	-0,04E-01	0,03
d2m_max	+	não	4.56E-02	7.99E-01	0,02	-0,02E-01
ro_max	não	-	4.89E-01	1.36E-10	-0,08E-01	-0,07
t2m_max	+	não	1.97E-04	5.00E-02	0,04	0,02
stl1_max	+	+	9.92E-04	3.78E-10	0,04	0,07
prec	não	+	7.68E-01	1.13E-12	0,03E-01	0,08
e_med	não	+	6.68E-01	1.63E-04	0,05E-01	0,04
d2m_med	não	-	6.52E-01	0,00	0,05E-01	-0,16
ro_med	não	-	6.62E-01	8.72E-08	0,05E-01	-0,06
t2m_med	+	não	5.81E-06	5.49E-02	0,05	0,02
stl1_med	+	+	3.09E-02	5.57E-03	0,02	0,03
e_min	+	-	8.06E-08	3.87E-03	0,06	-0,04
d2m_min	-	-	8.10E-06	0,00	-0,05	-0,20
ro_min	não	-	4.42E-01	1.62E-08	0,01	-0,06
t2m_min	não	não	2.44E-01	2.02E-01	0,01	-0,01
stl1_min	-	-	4.31E-03	3.09E-02	-0,03	-0,02
e_std	-	+	7.63E-08	3.17E-02	-0,06	0,02
d2m_std	+	+	5.51E-14	0,00	0,08	0,15
ro_std	não	-	2.87E-01	2.60E-08	-0,01	-0,06
t2m_std	+	+	1.61E-05	2.13E-04	0,05	0,04
stl1_std	+	+	7.03E-08	0,00	0,06	0,11
e_var	-	+	7.63E-08	3.16E-02	-0,06	0,02
d2m_var	+	+	5.51E-14	0,00	0,08	0,15
ro_var	não	-	2.87E-01	2.60E-08	-0,01	-0,06
t2m_var	+	+	1.61E-05	2.13E-04	0,05	0,04
stl1_var	+	+	7.03E-08	0,00	0,06	0,11
VPD_mean	+	+	1.73E-04	1.12E-09	0,04	0,067
VPD_min	+	+	7.98E-04	0	0,04	0,017
VPD_max	+	não	6.85E-03	0,09	0,03	0,018

Tabela sobre a tendência de Mann Kendall calculada para os preditores METEO. O valor-p é a significância do teste, para valor-p<0,05 há tendência na série temporal do atributo, o valor Tau é a intensidade da tendência e varia de -1 a 1. A cor vermelha e símbolo '+' se refere à tendência de aumento, a cor azul e símbolo '-' se refere à tendência de diminuição.

Figura 5.13 - Comparação da Distribuição de Dados Meteorológicos entre a Célula K67 e a Célula K34: BoxPlot.



Um boxplot é um gráfico que resume a distribuição de um conjunto de dados, destacando a mediana, os quartis e os valores extremos de forma visualmente concisa. Em azul está as caixas referentes ao dados da K67 e em laranja os referente à K34.

No entanto, não foram identificadas mudanças significativas de tendências nas variáveis de ‘prec’ e ‘ro’. É possível que o aumento na evaporação seja decorrente de mudanças no uso do solo ocorridas entre 2002 e 2011, ainda que tenham sido pequenas em relação à área total da célula. Essas alterações podem ter contribuído para o leve aumento observado na tendência da evaporação e de outras variáveis relacionadas. De acordo com Nobre et al. (2021) e Restrepo-Coupe et al. (2013), a evaporação na região Amazônica segue a sazonalidade do comportamento da radiação líquida, o que sugere a possibilidade de um leve aumento ao longo do período analisado. O comportamento da tendência dos atributos METEO no período de 2002 a 2011 se encontra na Tabela 5.7 e o gráfico boxplot comparativo para identificação das diferentes características dos atributos das Células K34 e K67 está na Figura 5.13.

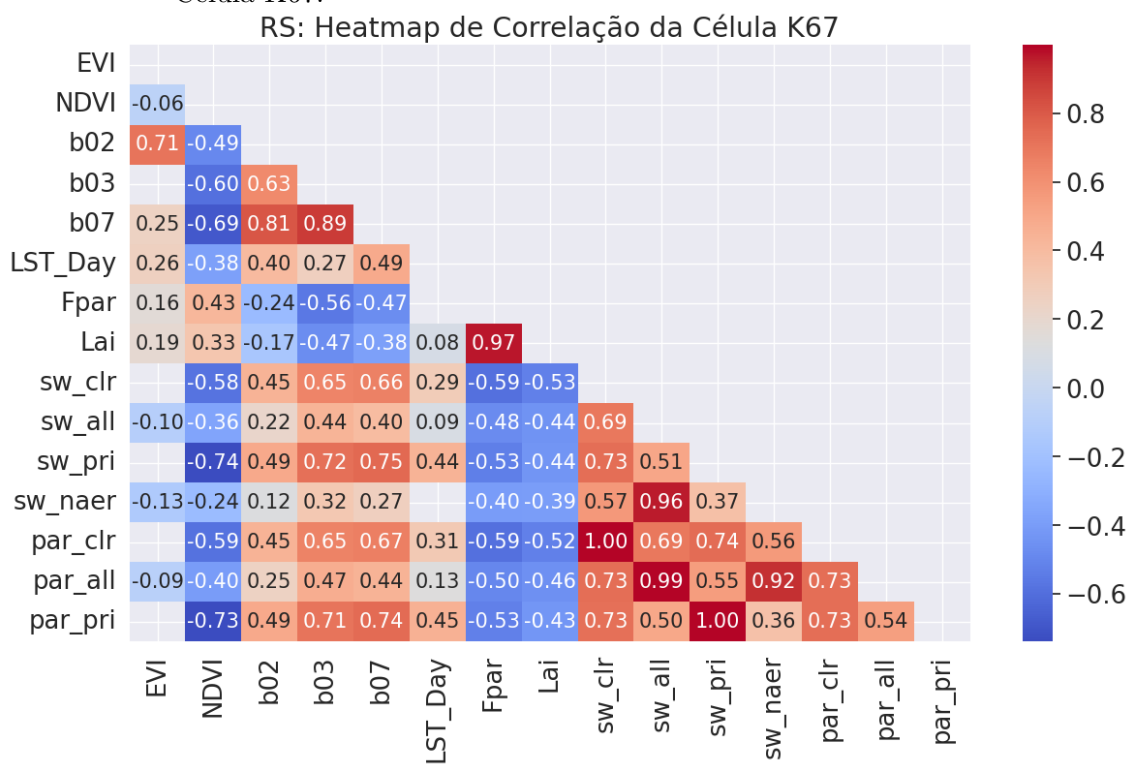
Na célula K34, foi observado uma ‘e’ menor em comparação com a K67, resultando em valores mais baixos dos atributos mencionados acima. No entanto foi notado uma maior variabilidade e valores mais altos de ‘d2m’, ‘prec’ e do ‘ro’, sugerindo um maior potencial de disponibilidade de água nessa célula. Essa dinâmica pode parecer contraditória, uma vez que na célula K67 há a presença do Rio Tapajós, o que normalmente indicaria uma maior disponibilidade de água.

A utilização dos preditores do conjunto METEO da célula K67 durante o processo de treinamento e validação pode ocasionar estimativas superestimadas de NEE no teste de estimativa para a célula K34. Isso pode ocorrer devido ao comportamento observado na validação dos dados de temperatura do ar do modelo, conforme discutido na Seção 5, e devido a própria dinâmica da célula K67, que é caracterizada por uma aumento térmico e evaporativo distinto da célula K34.

5.5 Análise exploratória dos dados: Análise de correlação no conjunto de preditores RS

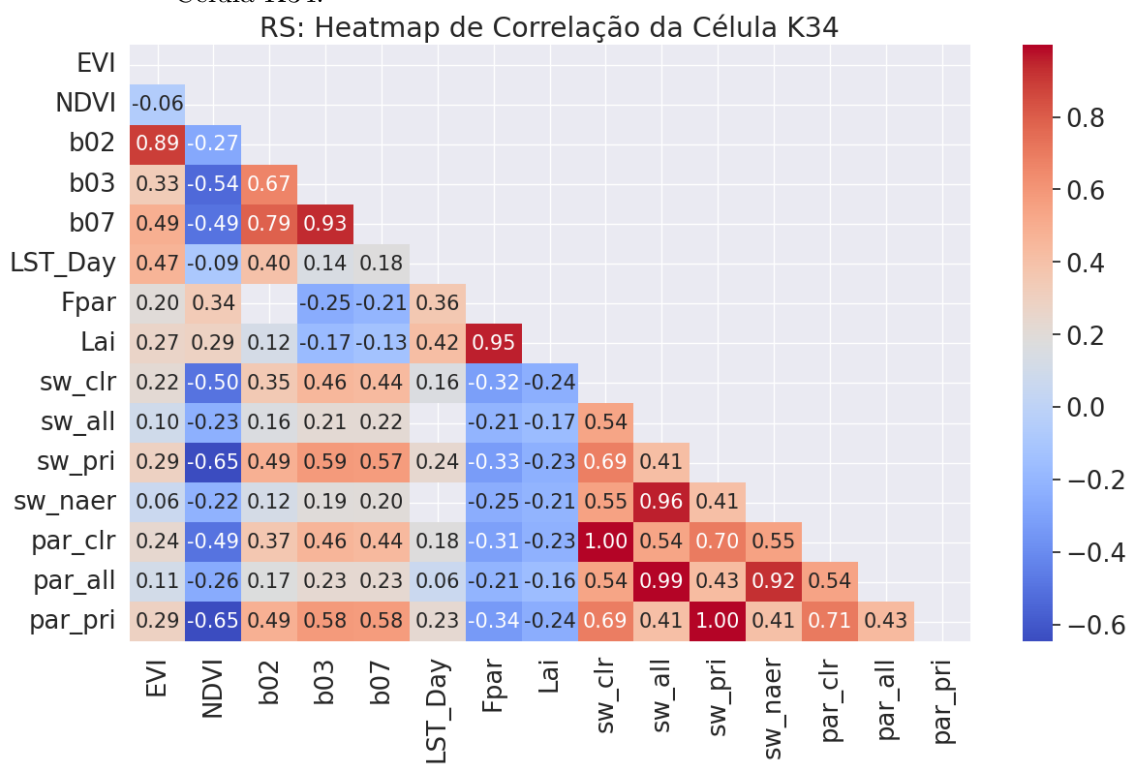
A atividade fotossintética do dossel florestal é influenciada pela saúde e extensão da vegetação (representadas pelas variáveis Lai, EVI, Fpar, NDVI, b02 e b07), pelas condições físicas do ambiente (METEO e b03), e pela disponibilidade de energia disponível para o processo (sw e par).

Figura 5.14 - Gráfico da matriz diagonal da Correlação de Spearman dos atributos RS da Célula K67.



Matriz Diagonal de Correlação de Spearman entre as variáveis Preditoras do Conjunto RS para a Célula K67. As correlações exibidas possuem significância valor-p<0,05.

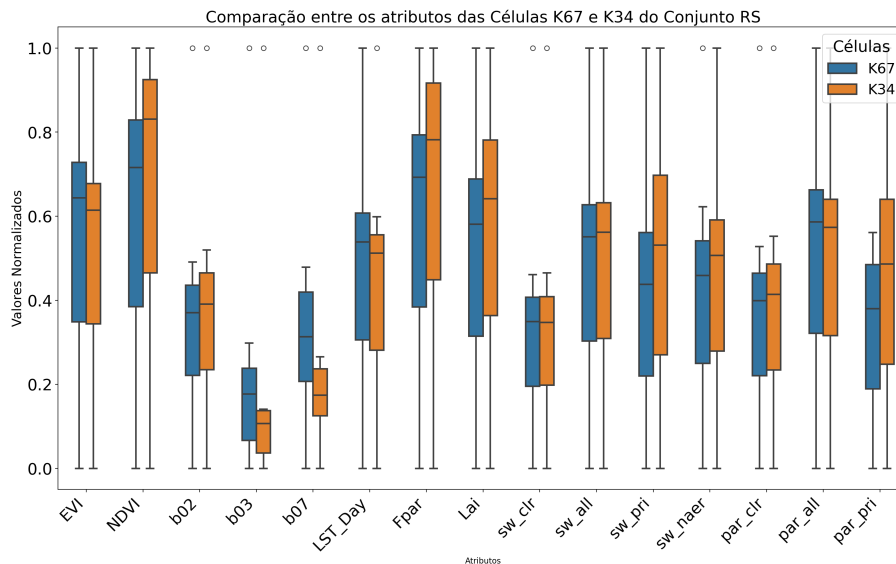
Figura 5.15 - Gráfico da matriz diagonal da Correlação de Spearman dos atributos RS da Célula K34.



Matriz Diagonal de Correlação de Spearman entre as variáveis Predictoras do Conjunto RS para a Célula K34. As correlações exibidas possuem significância valor-p<0,05.

Nas Figuras 5.15 e 5.14, foi observado uma correlação positiva de fraca a moderada entre os atributos Lai, EVI, Fpar e NDVI para as células K67 e K34, sendo mais intensa na primeira. Os mesmos atributos mostraram correlação negativa com todos os atributos referentes à radiação difusa, indicando que, à medida que a atividade fotossintética e saúde da vegetação aumentam, maior quantidade de radiação é absorvida na fotossíntese, resultando em uma diminuição na quantidade de radiação difusa no ambiente. A radiação difusa possui maior penetração ao decorrer do dossel, aumentando a eficiência da fotossíntese. Como resultado, a atividade fotossintética e o crescimento da vegetação são estimulados, levando ao aumento de biomassa e densidade vegetal (RODRIGUES; CIRINO, 2024).

Figura 5.16 - BoxPlot RS K34 x BrSa1.



Esta figura contém a análise boxplot dos dados dos conjuntos RS das células K67 e K34. Em azul são os dados da célula K67 e a laranja os dados da célula K34, no período de 2002 a 2011.

A partir do mapeamento de Fitofisionomias da Vegetação fornecido pelo IBGE (INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (IBGE), 2012), foi identificado que ambas as células estão localizadas em regiões de Floresta Ombrófila Densa de Terras Baixas, porém parte da célula K67 apresenta vegetação secundária e atividades agrícolas, o que foi visto também na Figura 3.2. Esse cenário pode ter ocasionado a maior atividade fotossintética e índices de vegetação mais elevados, vistos na Figura 5.16, na célula K34. No entanto, os dados revelam que os valores de radiação difusa disponível na célula K67 são superiores aos da K34, o que não é

esperado já que os indicadores de atividade fotossintética são maiores na segunda. Isso sugere que, apesar da área de Floresta na célula K34 ser maior, ela absorve menor quantidade de radiação difusa na fotossíntese, o que pode explicar os valores mais baixos de EVI quando comparados aos valores da K67.

Tabela 5.8 - Tendências avaliadas com o teste de Mann Kendall dos preditores do conjunto RS nas Células K67 e K34.

Atributo	tendência		p		Tau	
	K67	K34	K67	K34	K67	K34
EVI	+	não	1.62E-08	2.48E-01	0,07	-0,01
NDVI	não	-	2.12E-01	1.75E-07	0,01	-0,06
b02	+	+	3.22E-13	5.43E-04	0,09	0,04
b03	não	+	9.50E-02	2.07E-08	-0,02	0,07
b07	não	+	6.35E-01	1.11E-03	0,01	0,04
LST_Day	+	não	1.53E-06	5.83E-02	0,06	0,02
Fpar	não	não	1.25E-01	9.81E-02	0,02	-0,02
FparStdDev	não	+	3.18E-01	8.00E-03	-0,01	0,03
Lai	não	-	3.78E-01	3.59E-02	0,01	-0,02
sw_clr	não	não	2.34E-01	2.37E-01	0,01	-0,01
sw_all	-	-	2.80E-04	1.84E-10	-0,04	-0,07
sw_pri	não	não	7.71E-02	3.65E-01	0,02	-0,01
sw_naer	-	-	2.55E-04	8.69E-12	-0,04	-0,08
par_clr	não	não	3.561E-01	7.32E-02	0,01	-0,02
par_all	-	-	1.74E-03	4.62E-09	-0,04	-0,07
par_pri	+	não	4.85E-02	5.35E-01	0,02	-0,01

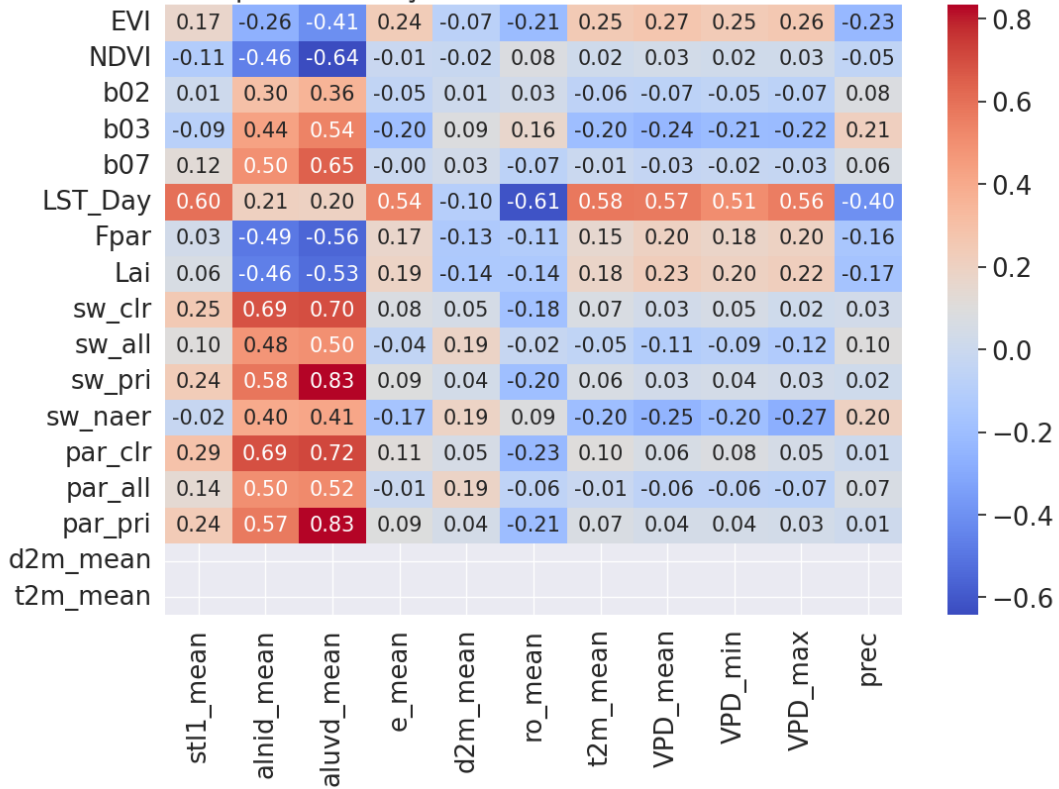
Tabela sobre a Tendência de Mann Kendall calculada para os preditores RS. O valor-p é a significância do teste, para valor-p<0,05 há tendência na série temporal do atributo, o valor Tau é a intensidade da Tendência e varia de -1 a 1. A cor vermelha e símbolo '+' se refere à tendência de aumento, a cor azul e símbolo '-' se refere à tendência de diminuição.

A análise da Tendência de Mann Kendall, na Tabela 5.8, revelou um aumento contínuo ao longo dos anos do EVI, de b02 e em LST_Day, a última compatível com o aumento de t2m vista no conjunto METEO para a célula K67 na Tabela 5.7. Em ambas as células, no período entre 2002 e 2011, observou-se uma fraca diminuição nos atributos de radiação difusa (sw e par) que possui o valor para o céu total (all e pri). Possivelmente a diminuição pode ser atribuída à natureza dos dados e à falta de tendência nas outras variáveis de radiação difusa. Uma possível explicação para essa redução na radiação difusa poderia ser o aumento da cobertura de nuvens na região durante esse período, no entanto, para o entendimento completo dessa situação, seria necessário a análise de cobertura de nuvens na região, o que no

momento não se encontra no escopo deste trabalho.

Figura 5.17 - Gráfico da matriz diagonal da Correlação de Spearman dos atributos METEORS da Célula K67.

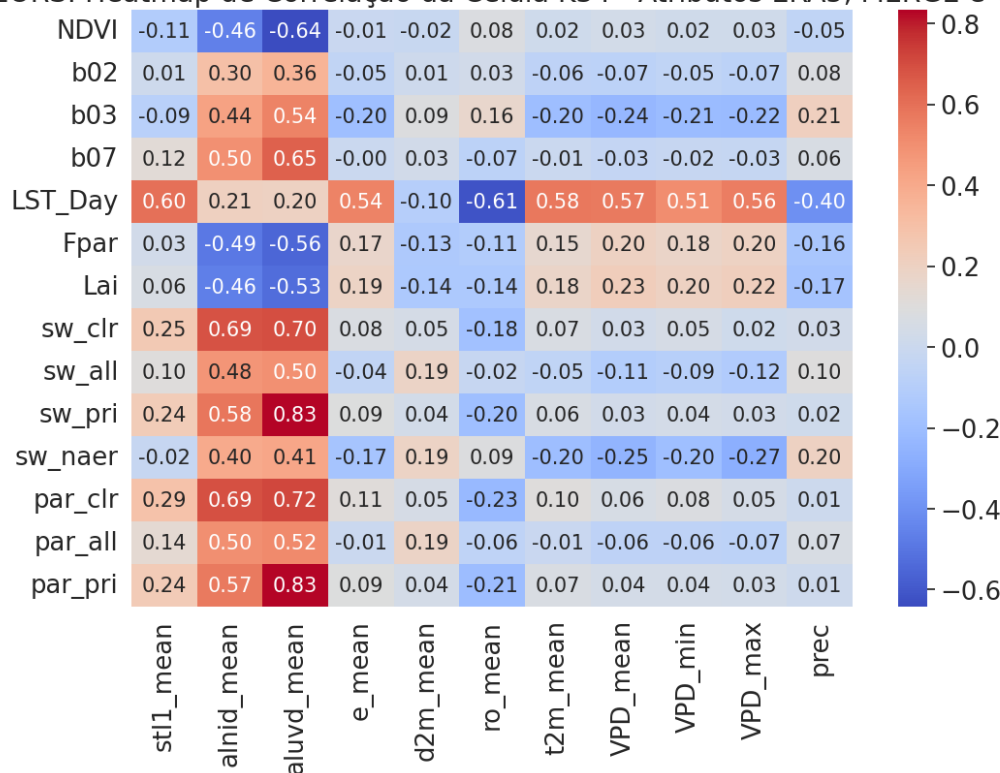
METEORS: Heatmap de Correlação da Célula K67 - Atributos ERA5, MERGE e VPD



Matriz de Correlação de Spearman entre as variáveis Preditoras do Conjunto METEORS para a Célula K67. As correlações exibidas possuem significância valor- $p < 0,05$.

Figura 5.18 - Gráfico da matriz diagonal da Correlação de Spearman dos atributos METEORS da Célula K34.

METEORS: Heatmap de Correlação da Célula K34 - Atributos ERA5, MERGE e VPD



Matriz de Correlação de Spearman entre as variáveis Predictoras do Conjunto METEORS para a Célula K34. As correlações exibidas possuem significância valor- $p < 0,05$.

Foi observado também a correlação entre os atributos METEORS, nas Figuras 5.5 e 5.5 , foi visto para a célula K67 uma correlação positiva fraca a moderada entre as radiações difusas de céu limpo e as refletâncias b02, b03 e b07, juntamente com os atributos stl1, t2m, 'e', d2m e VPD. Esse comportamento condiz com o que foi visto por (NOBRE et al., 2021) e (RESTREPO-COUBE et al., 2013), que indicam que o processo de evaporação segue a sazonalidade da radiação líquida em um ecossistema, e a radiação difusa nada mais é do que parte dessa radiação líquida.

No entanto, essa relação não foi observado na célula K34, onde os atributos acima que remetem ao processo de evaporação mostraram uma correlação positiva fraca, não observada na K67, com os atributos que remetem à atividade fotossintética do sistema Fpar, Lai, EVI e b02. Essas diferenças podem ser atribuídas aos diferentes elementos das células pertencentes ao ciclo d'água. Na célula K67, a evaporação ocorre a partir de várias fontes, incluindo a evaporação de água do Rio Tapajós, da irrigação agrícola, do solo e da transpiração da Floresta. Esses diferentes tipos de

evaporação estão intrinsecamente ligados ao comportamento térmico do ambiente, à disponibilidade de vapor d'água e, conseqüentemente, aos maiores valores de VPD, o que foi visto na análise do conjunto METEO da célula K67 e é visto na conexão com o conjunto RS. Por outro lado, na célula K34, o processo de evaporação é possivelmente mais influenciado pela floresta, especialmente pelos processos de evapotranspiração do dossel florestal e do rio próximo à torre de medição.

Após essas análises de correlação entre os atributos, foram retirados do conjunto de dados de treinamento os atributos que possuíram correlação alta com outros, os atributos retirados dos conjuntos foram o stl1, VPD_mean, sw_naer, sw_clr, sw_pri, b03, NDVI, alnid e aluvd.

5.6 Modelos de aprendizado de máquina: preparação dos dados

Os modelos de aprendizado de máquina neste estudo foram alimentados com uma variedade de entradas, determinadas pelo tipo de algoritmo empregado, pelas relações de correlação entre as variáveis previamente identificadas e pela importância das variáveis na análise dos ecossistemas, conforme indicado pela literatura.

Optou-se por tratar cada observação diária como uma amostra de comportamento, o que nos levou a embaralhar as entradas durante o treinamento. Isso significa que não foi considerado o comportamento temporal da série durante o treinamento dos modelos, mas sim a relação entre os preditores e a variável alvo. Essa abordagem foi adotada para melhorar a capacidade de generalização dos modelos, tendo em vista que as séries temporais das variáveis alvos continham muitos dados preenchidos pelo ciclo anual médio e foi necessário retirá-los, criando assim lacunas na série temporal. Portanto, foi utilizado os dados diários embaralhados, o que reduz a especialização dos algoritmos de aprendizado de máquina e permite uma melhor generalização.

Dito isso, adotou-se um padrão de separação de amostras, reservando 60% para a etapa de treinamento, 20% para validação e 20% para teste (validação fora do processo de treinamento). Antes de serem utilizados nos modelos de RNAs, os dados passaram por normalização através do algoritmo de mínimos e máximos, garantindo que os preditores tivessem seus valores reescalados para o intervalo de 0 a 1. Para os modelos de Floresta Aleatória e Gboost, não foi realizada a normalização dos dados, uma vez que os valores limítrofes são importantes para entender as escolhas aplicadas pelo algoritmo.

Um dos objetivos desta pesquisa está relacionado ao uso de multi-técnicas de apren-

dizado de máquina. Embora algumas técnicas já tenham sido aplicadas em estudos anteriores (TRAMONTANA et al., 2016; JUNG et al., 2020; BAUER, 2022; ZHU-RAVLEV et al., 2022), é necessário investigar novas abordagens para determinar se é possível o aprimoramento das estimativas. Neste trabalho, foram conduzidos experimentos com diferentes arquiteturas de Redes Neurais de Grafos (por exemplo, GAT e GCN simples), utilizando os dados dos conjuntos Fluxnet e LBA como variável alvo. Dado que essas redes são recentes e até então poucos aplicadas em problemas como este. Sendo assim, seus resultados precisam ser comparados com técnicas mais estabelecidas para se verificar sua contribuição a temática deste estudo.

Neste estudo optou-se pela inclusão de algoritmos mais simples, que não demandam um grande número de amostras, aumentando a interpretabilidade dos resultados. Assim sendo, foram selecionados os algoritmos RF, MLP e GB, sendo que os dois primeiros já foram anteriormente utilizados pela Fluxcom na estimativa do NEE (TRAMONTANA et al., 2016). Contudo, com resultados limitados a região da Floresta Amazônica.

A partir disso, foram feitos uma série de experimentos envolvendo diferentes configurações de parâmetros e hiperparâmetros e com diferentes conjuntos de preditores, com o objetivo de identificar e melhorar as arquiteturas dos modelos e o melhor conjunto de dados de entrada. Dentre os experimentos realizados, os 3 melhores modelos foram escolhidos para a análise intercomparativa. Os experimentos com diferentes conjuntos de preditores foram realizados em cada um dos algoritmos considerados. Além disso, foi feito testes com diferentes variáveis alvo, utilizando tanto a série temporal do conjunto de dados Fluxnet quanto outras séries temporais do conjunto de dados LBA. Os preditores utilizados com suas respectivas variáveis alvos se encontram na Tabela 4.3 na Subseção 4.3.1.

5.7 Modelos de aprendizado de máquina: treinamento e validação e teste

As etapas de treinamento e validação são fundamentais no desenvolvimento de modelos de aprendizado de máquina, elas garantem que o modelo tenha a capacidade de aprender o comportamento dos dados de treinamento e ao mesmo tempo possa generalizar para os dados de validação. Durante o processo de treinamento, os modelos ajustam seus parâmetros para minimizar a discrepância entre os valores estimados de NEE e os valores observados. A métrica utilizada para avaliar o desempenho de todos os modelos nesta etapa foi o MSE.

Foram empregados dois métodos para ajuste de hiperparâmetros, visando identificar

a melhor arquitetura para os modelos de aprendizado de máquina, mais detalhes podem ser encontrados na Seção 4.3.2. O primeiro método consistiu em uma abordagem aleatória, na qual uma ampla gama de hiperparâmetros com diferentes magnitudes foi explorada. Essa abordagem permitiu entender o comportamento da rede e determinar os melhores hiperparâmetros e seus respectivos limites. Umz vez filtrados em parâmetros, os mesmos foram utilizados como ponto de partida para a busca otimizada realizada pelo Optuna, uma busca dos melhores conjuntos de hiperparâmetros utilizando uma abordagem probabilística bayesiana (AKIBA et al., 2019). Essa técnica buscou encontrar os melhores conjuntos de hiperparâmetros para cada algoritmo, visando alcançar o melhor desempenho do modelo.

A validação de cada modelo foi realizada ao final de cada etapa de treinamento, durante a qual foram armazenadas as configurações dos melhores modelos e suas respectivas métricas de validação (MSE, MEF, R^2 , BIAS, ROV e R) que definem a concordância entre os valores estimados pelo modelo e os valores observados. As arquiteturas dos modelos com melhores métricas de desempenho se encontram na Tabela 5.9. No total, foram realizadas 5.000 alterações nos hiperparâmetros para os modelos RF, Gboost e MLP; 1000 alterações para o GAT e GCN; e 500 alterações para o ANFIS. Essas alterações foram aplicadas a cada conjunto de dados das Tabelas 4.3 e 4.2.

Os modelos GCN e GAT mostraram os piores resultados nas etapas de treinamento e validação, durante o treinamento eles rapidamente realizavam o ajuste entre os dados preditos e observados, mostravam o MSE de 0,000005. No entanto, durante a validação se mostrou com comportamento contrário, com altos valores de MSE ao se comparar aos valores de treinamento. Aumentar a complexidade da arquitetura levou a uma especialização mais rápida da rede e piorou a capacidade de generalização da rede. Já com a diminuição da complexidade, ocorria a especialização e a redução do erro da validação, mas ainda não o suficiente para generalização dos dados. As melhores métricas de validação foram encontradas com a GAT e, mesmo assim, se mostrou com desempenho insatisfatório. A melhor métrica de validação exibiu $MEF=-0,14$, $MSE=3,09$, $R=-0,05$ e $BIAS=0,11$ e tais foram obtidas com a variável alvo de BrSa1 utilizando o conjunto de preditores RS. O MEF negativo indica que o modelo não se ajusta aos dados, sendo pior do que uma simples média, o MSE é consideravelmente alto, indicando que as estimativas do modelo estão distantes dos valores reais, além disso, o R próximo de zero sugere uma correlação muito fraca entre as estimativas do modelo e os valores reais. Embora o BIAS seja próximo de zero, indicando ausência de tendência de superestimar ou subestimar, os resultados gerais

Tabela 5.9 - Hiperparâmetros dos Modelos com melhor desempenho.

Modelos	Hiperparâmetros
RF	ccp_alpha: 0.0 criterion: squared_error max_depth: 25 min_samples_leaf: 1 min_samples_split: 2 min_weight_fraction_leaf: 0.0 n_estimators: 65 random_state: 42
Gboost	alpha: 0.9 criterion: friedman_mse learning_rate: 0.05 loss: squared_error max_depth: 5 min_samples_leaf: 1 min_samples_split: 2 n_estimators: 145 n_iter_no_change: None
MLP	activation_function: tanh alpha: 0.0022 batch_size: auto epsilon: 1e-08 hidden_layer_sizes: (10, 10) learning_rate: constant learning_rate_init: 0.0047 max_fun: 15000 max_iter: 1123 momentum: 0.9 n_iter_no_change: 10 nesterovs_momentum: True power_t: 0.5 random_state: 42 solver: adam
ANFIS	epochs: 300 learning_rate: 0,01 hidden_layer: 150 activation_function:: relu fuzzy_rule: gaussian random_state: 42

não são promissores. Contudo, o BIAS (ou viés) não é uma métrica confiável, uma vez que variações negativas e positivas dos erros encontrados podem ser anuladas na média.

Durante os processos de treinamento e validação, as redes neurais MLP e ANFIS exibiram um comportamento de especialização dos dados de treinamento, especialmente com o aumento de hiperparâmetros, como o número de neurônios e camadas, o que resultou em uma rápida especialização, semelhante à GNN. Contudo, um desafio enfrentado em ambas as redes neurais foi a presença constante de sobreajuste durante essas etapas. À medida que a complexidade das redes aumentava, com mais neurônios e camadas, o sobreajuste também aumentava. No entanto, ao longo dos experimentos, observou-se que, para a MLP, a redução do número de camadas e neurônios levou a uma melhoria nas métricas de validação, apesar da persistência do sobreajuste. Para a ANFIS, o aumento da quantidade de neurônios resultou em uma maior especialização e redução dos erros de validação. Foram realizados diversos experimentos para mitigar o sobreajuste e melhorar a generalização dos dados, embora o processamento fosse demorado, limitando a realização de mais experimentos com a rede ANFIS. Apesar disso, a ANFIS mostrou-se promissora para a Amazônia, devido à sua convergência lenta e capacidade de resposta às mudanças nos hiperparâmetros.

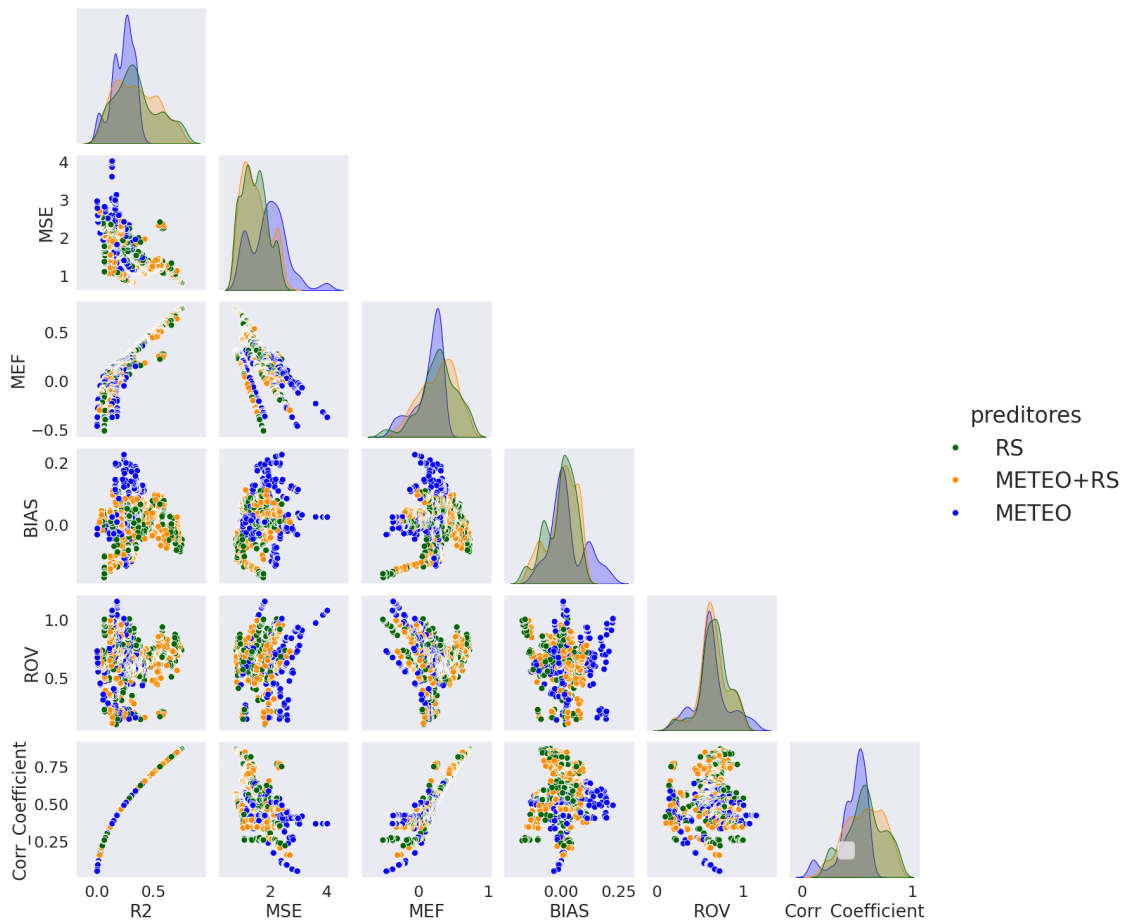
Os modelos Gboost e RF destacaram-se pelo comportamento mais robusto durante o treinamento e validação. Desde o início do treinamento, com configurações iniciais simples, já foi possível visualizar uma melhora no aprendizado dos modelos e na capacidade de generalização. Com isso, os dois algoritmos conseguiram ajustar os preditores na variável alvo com facilidade, obtendo resultados próximos aos dos melhores modelos na validação, tanto com configurações mais complexas quanto com configurações mais simples.

De um modo geral, o RF quanto o Gboost demonstraram bom desempenho em relação à estimativa de NEE. A Floresta Aleatória mostrou-se mais robusta em termos de ajuste aos dados, exibindo valores mais altos de métricas como R, R^2 , ROV e MEF. Por outro lado, o Gboost destacou-se por ter menores valores de MSE e BIAS, indicando uma melhor precisão nas previsões.

Ao acompanhar o desempenho dos modelos RS, Gboost, MLP e ANFIS, observou-se na Figura 5.19 que os preditores RS e METEORS apresentaram as métricas mais elevadas e melhor se ajustaram à variável alvo. Esse resultado sugere um forte impacto dos atributos de sensoriamento remoto, fornecendo considerações valiosas

sobre a importância do comportamento da vegetação na previsão de NEE. No entanto, devido às diferenças de comportamento dos preditores nas células K34 e K67 em relação ao comportamento da vegetação e das condições meteorológicas, conforme discutido no Capítulo 5, é possível que o modelo treinado na célula K67 não generalize bem para os dados da célula K34.

Figura 5.19 - Gráficos de Pares Comparativos das Métricas de Validação dos Preditores separados por cores para todos os modelos utilizados neste trabalho.



Os modelos que incorporaram os preditores METEO na etapa de validação demonstraram valores semelhantes de métricas como ROV e BIAS em comparação com os modelos que utilizaram outros preditores. No entanto, durante o treinamento e a validação dos modelos, os preditores METEO exibiram os maiores erros e as métricas mais baixas de R, R^2 e MEF. Entre os modelos que empregaram os preditores METEO, o GB e o RF foram os que alcançaram as melhores métricas, com a variável

alvo sendo a da Torre K83.

Um dos principais desafios enfrentados na estimativa de NEE neste trabalho foi incorporar o comportamento físico por meio de preditores. Os preditores utilizados neste estudo têm uma resolução espacial e temporal mais baixa do que os dados observados das torres, o que pode resultar em comportamentos ambientais distintos daqueles observados *in loco*. Para compreender essa discrepância, foram realizados experimentos em várias torres, com o objetivo de determinar se algum modelo se ajustava melhor a uma torre específica, se havia um desempenho superior nas torres com séries temporais mais longas, ou se a inclusão de dados de todas as torres resultaria no melhor ajuste.

A iniciativa FLUXCOM utilizou dados do conjunto global de torres de fluxo da FLUXNET para realizar estimativas de NEE em todo o mundo. Foram empregados onze modelos distintos com preditores RS e mais quatro modelos com preditores METEORS. Entretanto, durante a estimativa de NEE, a FLUXCOM identificou desafios significativos ao estimar essa variável em regiões tropicais e em áreas de Floresta Perene de Folhas Largas (EBF), e a célula deste estudo se encontra exatamente nessa área. De acordo com [Tramontana et al. \(2016\)](#), essas áreas demonstraram um desempenho relativamente inferior na previsão dos fluxos de carbono, atribuído a desafios específicos relacionados à dinâmica sazonal dessas florestas, bem como à ausência de um ciclo sazonal claro nessas regiões. Ao calcular a capacidade de estimativa usando a mediana dos modelos globais, o NEE teve a pior performance, com $MEF=0.46$, $RMSE=1.24$, $R=0.68$, $ROV=0.39$, $BIAS=-0.07$ ([TRAMONTANA et al., 2016](#)). Contudo, [Jung et al. \(2020\)](#) atribuem a dificuldade da FLUXCOM em prever a estimativa de NEE nas regiões tropicais e em EBF à escassez de dados de satélites, contaminação por nuvens e limitações na detecções de processos climático complexos. Ele fornece os resultados da FLUXCOM para a região da América do Sul Tropical, onde é localizada o Bioma Amazônia e o coeficiente de determinação para região a partir dos modelos com preditores METEORS é 0,33 e para os modelos com preditores RS é 0,1, sendo o coeficiente de correlação negativo.

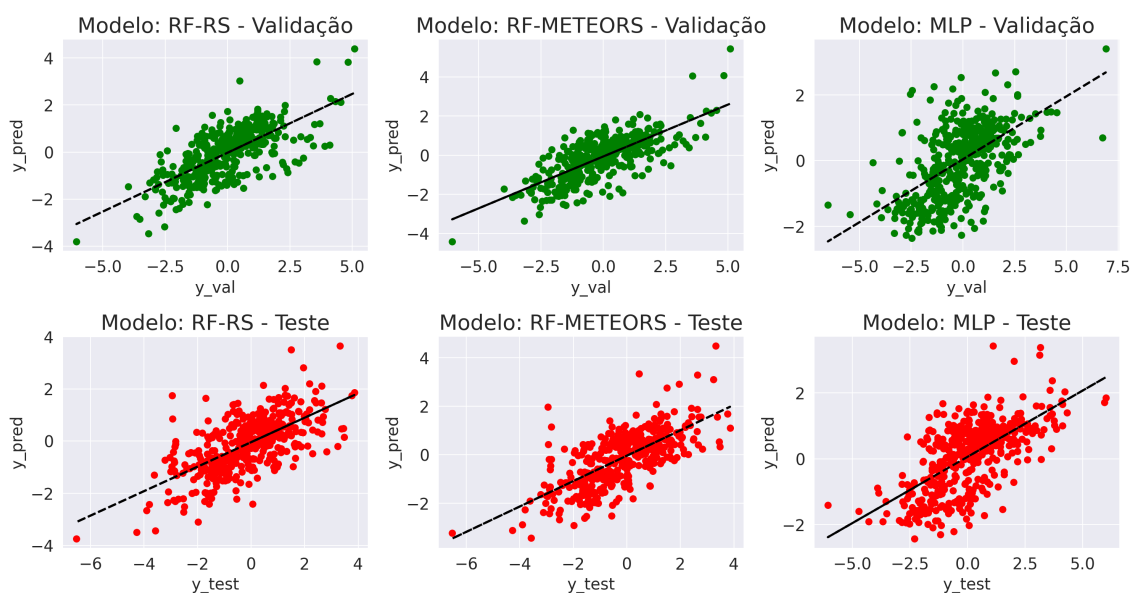
Na Tabela 5.10 e na Figura 5.20, são apresentadas as métricas e gráfico de dispersão dos melhores modelos na estimativa da variável NEE_VUT_REF da FLUXNET. Estes resultados foram comparados com os valores de capacidade preditiva obtidos pela FLUXCOM, os quais foram gerados utilizando a RF e a MLP em suas previsões.

Notavelmente, os resultados da RF, com preditores METEORS e RS, superam as expectativas em relação à capacidade preditiva global do modelo para o NEE e

Tabela 5.10 - Métricas de validação e teste do modelo com melhor desempenho utilizando dados do conjunto Fluxnet - Torre BrSa1 com filtro de qualidade de $>0,7$.

Melhor Modelo	RF		RF		MLP	
Torre da Variável Alvo	Torre BrSa1		Torre BrSa1		Torre BrSa1	
Variável Alvo	NEE_VUT_REF		NEE_VUT_REF		NEE_VUT_REF	
Preditor	RS		METEORS		METEORS	
Etapa	Val	Teste	Val	Teste	Val	Teste
MEF	0,50	0,42	0,53	0,49	0,31	0,39
RMSE	1,10	1,16	1,06	1,08	1,35	1,28
ROV	0,71	0,72	0,72	0,75	0,66	0,64
R	0,70	0,65	0,73	0,70	0,57	0,63
R²	0,50	0,42	0,54	0,49	0,33	0,33
BIAS	0,006	0,06	0,02	0,06	0,14	0,03

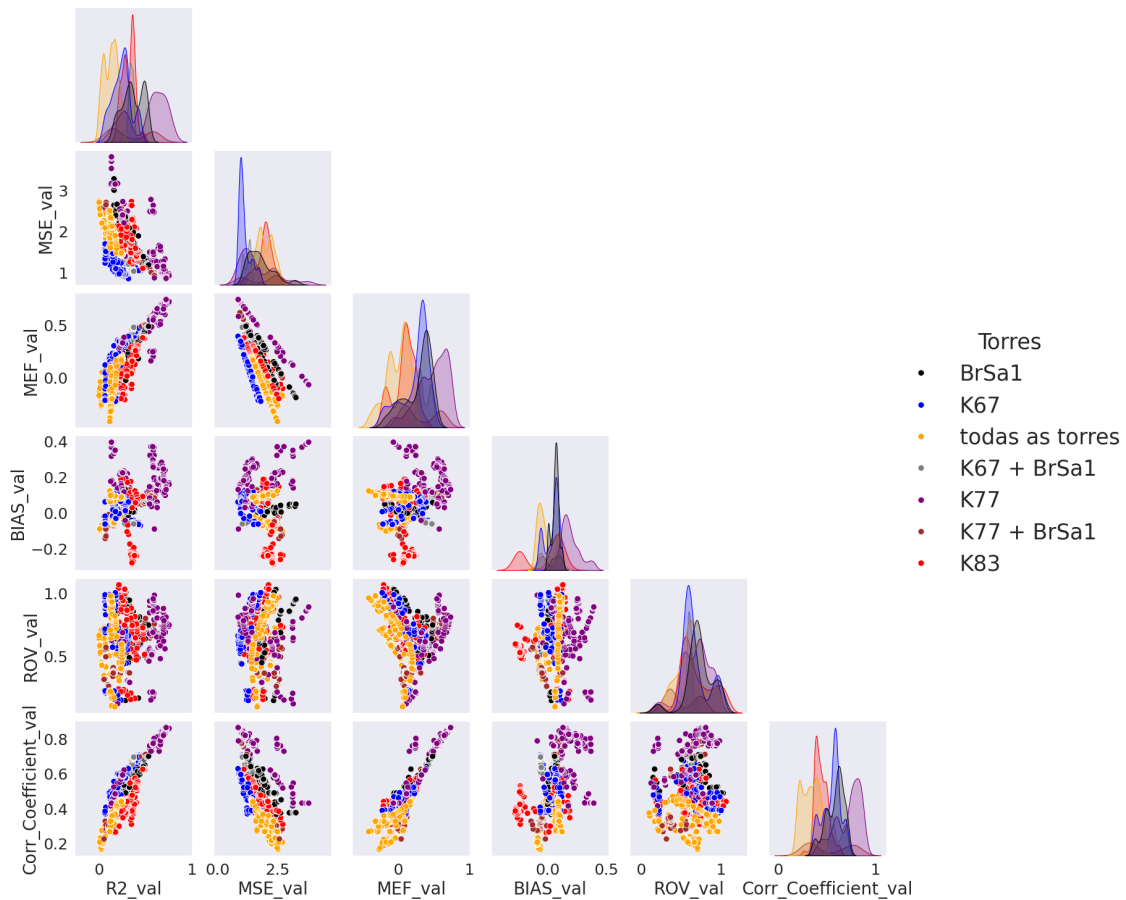
Figura 5.20 - Gráficos de Dispersão Comparativos dos Dados Observados e Simulados para as Etapas de Validação e Testes na estimativa de NEE para a Torre BrSa1 do conjunto FLUXNET. Referente aos modelos da Tabela 5.10.



Esta figura apresenta gráficos de dispersão que comparam os resultados da validação (pontos verdes) e teste (pontos vermelhos) dos modelos. As linhas de tendência indicam a relação geral entre os resultados previstos e observados em ambas as etapas, destacando a consistência da performance dos modelos em diferentes conjuntos de dados.

aos resultados encontrados para a região da América do Sul Tropical. Este modelo apresenta valores mais elevados de MEF, R e ROV, além de menores valores de erro RMSE. Por outro lado, o conjunto MLP com preditores METEORS demonstra uma aparente semelhança com os resultados encontrados para a região Tropical da América do Sul. No entanto, ele obteve um coeficiente de correlação positivo, indicando um ajuste do modelo superior.

Figura 5.21 - Gráficos de Pares Comparativos das Métricas de Validação dos Modelos RF e Gboost separados por cores por Variável Alvo do Modelo.



Gráficos de dispersão que comparam os resultados das métricas encontradas na etapa de validação.

Ao longo do processo de treinamento e validação, os modelos que tinham como alvo apenas os dados de uma única torre (K67, K77, K83 e BrSa1), apresentaram as melhores métricas nos dados preditos em comparação com aqueles que utilizaram a combinação de dados de duas ou mais torres, tal comportamento pode ser visto na Figura 5.21 . Como destacado na seção sobre a variabilidade de NEE, as torres apre-

sentam comportamentos distintos entre si, influenciados pela localização geográfica e pelo uso do solo. No entanto, os preditores empregados neste estudo possuem uma resolução espacial menor e capturam a variabilidade dos diferentes usos do solo e dos elementos presentes nas três torres, conforme demonstrado no Capítulo 5.

Essas diferenças foram refletidas nos resultados de validação dos modelos, já que o aumento da diversidade dos valores observados de NEE ocasionou uma aparente "confusão" e diminuição dos desempenhos dos modelos.

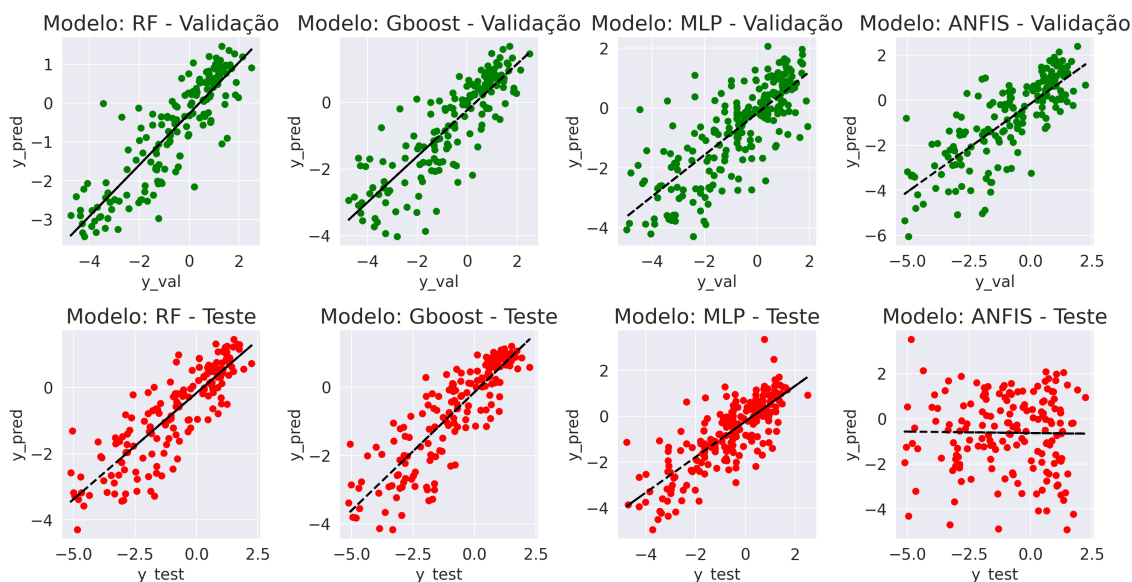
Na Figura 5.21 são apresentadas as métricas de desempenho dos modelos em relação a cada uma das torres, nas quais o NEE foi utilizado como variável alvo. Ao analisar essa figura, observa-se que os modelos que empregaram os dados de todas as torres como variável alvo apresentaram os menores valores nas métricas de ajuste (R , R^2 , MEF, ROV). Dentre esses modelos, aquele que obteve o melhor ajuste foi o modelo RF, que estimou os valores de NEE para a torre K77 + BrSa1 (MEF=0,62, $R=0,79$, ROV=0,75, MSE=0,91). As melhores métricas foram alcançadas com a estimativa de NEE da torre K77, sendo que o modelo anteriormente mencionado inclui dados dessa torre. Essas melhores estimativas foram obtidas através do uso dos preditores RS e METEORS nos modelos.

Como observado na análise da variabilidade do NEE na Torre K77, essa torre demonstrou ter o comportamento mais previsível em relação ao ciclo anual do NEE, e tal comportamento pode estar relacionado à atividade agrícola na área onde a torre está localizada. Ao analisar o mapeamento do Mapbiomas e do IBGE, juntamente com detalhes das áreas das torres encontradas na literatura, constatou-se que a célula K67 (na qual se encontra a Torre K77) engloba áreas agrícolas, o Rio Tapajós e florestas primárias e de manejo. Portanto, é provável que o comportamento meteorológico e da vegetação, ou seja, dos preditores utilizados neste estudo, reflitam os padrões desses diferentes usos do solo. Assim, isso pode ser um indicativo do por que foram obtidas melhores estimativas para a Torre K77.

Tabela 5.11 - Métricas de validação e teste dos modelos com melhor desempenho em relação a Torre K77.

Melhor Modelo	Gboost		RF		RF		MLP		ANFIS	
Torre da Variável Alvo	Torre K77		Torre K77		Torre K77		Torre K77		Torre K77	
Preditor	METEORS		METEORS		RS		RS		METEORS	
Etapa	Val	Teste	Val	Teste	Val	Teste	Val	Teste	Val	Teste
MEF	0,72	0,74	0,72	0,72	0,75	0,71	0,58	0,58	0,57	-0,90
MSE	0,86	0,87	0,92	0,95	0,79	0,95	1,15	0,87	1,43	6,27
ROV	0,81	0,80	0,72	0,72	0,76	0,75	0,90	0,98	0,99	0,92
R	0,85	0,86	0,85	0,86	0,87	0,85	0,77	0,79	0,78	-0,01
R²	0,72	0,72	0,72	0,72	0,76	0,76	0,60	0,62	0,61	0
BIAS	-0,05	0,13	-0,003	0,15	-0,06	0,16	0,02	-0,09	0,04	0,28

Figura 5.22 - Gráficos de dispersão comparativos dos dados observados e estimados para as etapas de validação e testes na estimativa de NEE para a Torre K77.



Esta figura apresenta gráficos de dispersão que comparam os resultados da validação (pontos verdes) e teste (pontos vermelhos) dos modelos. As linhas de tendência indicam a relação geral entre os resultados previstos e observados em ambas as etapas, destacando a consistência da performance dos modelos em diferentes conjuntos de dados.

Na Tabela 5.11 e na Figura 5.22 estão os modelos que apresentaram as melhores métricas nos experimentos realizados. Como mencionado anteriormente, esses modelos conseguiram ajustar os preditores aos dados da Torre K77 de forma muito eficaz. A RF com os preditores RS foi o que obteve o melhor ajuste. Ele exibiu os melhores resultados tanto na etapa de validação quanto na de teste, com os valores mais altos para as métricas de ajuste em conjunto (MEF, R, R^2 e ROV) com o menor valor de MSE .

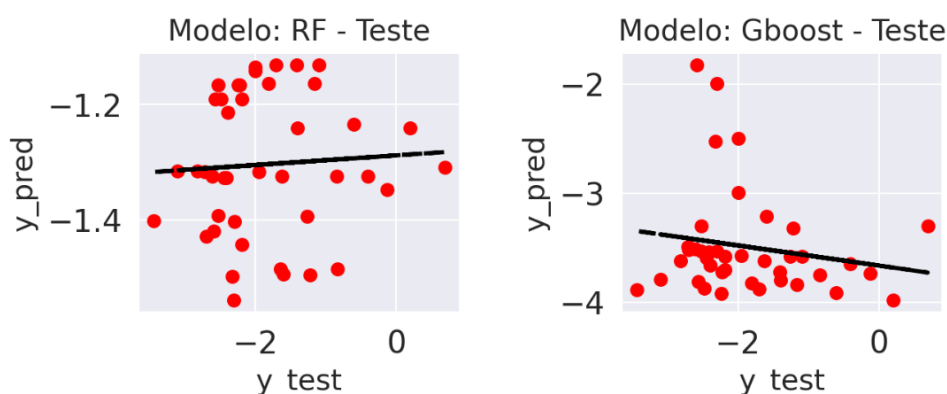
O ANFIS obteve um desempenho moderado na etapa de validação, mas obteve uma piora significativa nas estimativas durante a etapa de teste. Isso pode ser atribuído ao fato de que o modelo foi ajustado durante o treinamento e a validação para se adaptar especificamente aos dados de validação. No entanto, durante esse processo, ocorreu overfitting, o que significa que o modelo se ajustou muito bem aos dados de treinamento e validação, mas teve dificuldades em generalizar para novos dados, como os dados de teste. Isso é claramente evidenciado na Figura 5.22, onde é visível um ajuste aos dados de validação, mas uma falta de adequação aos dados de teste. Isso sugere que o modelo ANFIS não conseguiu generalizar adequadamente além dos dados utilizados no treinamento e na validação.

O escopo da generalização do modelo é a sua capacidade de lidar com dados não vistos durante o treinamento. Isso é avaliado através das métricas de comparação entre os dados previstos e observados, determinando se o modelo conseguiu generalizar eficazmente para esses novos dados. Na Tabela 5.11 há os testes realizados com dados da Célula K67. Na Tabela 5.12 e Figura 5.23, há os testes de generalização dos modelos com os valores observados de NEE na torre K34, que se encontra na Célula K34, ou seja, fora do alcance geográfico dos dados de treinamento.

Tabela 5.12 - Métricas de teste dos modelos com melhor desempenho - Teste com a Torre K34.

Melhor Modelo	Gboost	RF	RF	MLP
Torre da Variável Alvo	Torre K34	Torre K34	Torre K34	Torre K34
Preditor	METEORS	METEORS	RS	RS
Etapa	Teste	Teste	Teste	Teste
MEF	-2,87	-0,34	-0,35	-0,57
MSE	3,11	1,08	1,07	1,24
ROV	1,01	0,41	0,13	0,86
R	-0,29	-0,21	0,06	0,15
R²	0,08	0,04	0,003	0,02
BIAS	-1,00	0,09	0,53	0,29

Figura 5.23 - Gráficos de Dispersão Comparativos dos Dados Observados e Simulados para a Etapa de Testes na estimativa de NEE para a Torre K34.



Esta figura apresenta gráficos de dispersão que comparam os resultados do teste (pontos vermelhos) dos modelos. As linhas de tendência indicam a relação geral entre os resultados previstos e observados, destacando a consistência da performance dos modelos em diferentes conjuntos de dados.

Foi notado que nenhum dos modelos foi capaz de se ajustar aos dados da torre K34. Isso pode ser atribuído, em parte, à disponibilidade limitada de dados para esta torre, representando apenas 9% do total, com os demais dados preenchidos pela média anual do ciclo. A incerteza quanto à representatividade desses dados para o comportamento real da NEE na Torre K34 pode ter impactado negativamente nos resultados dos testes. Além disso, a discrepância nos padrões de comportamento dos preditores entre as duas torres, como evidenciado nas análises de correlação, pode ter contribuído para a inabilidade do modelo em generalizar efetivamente para diferentes ambientes.

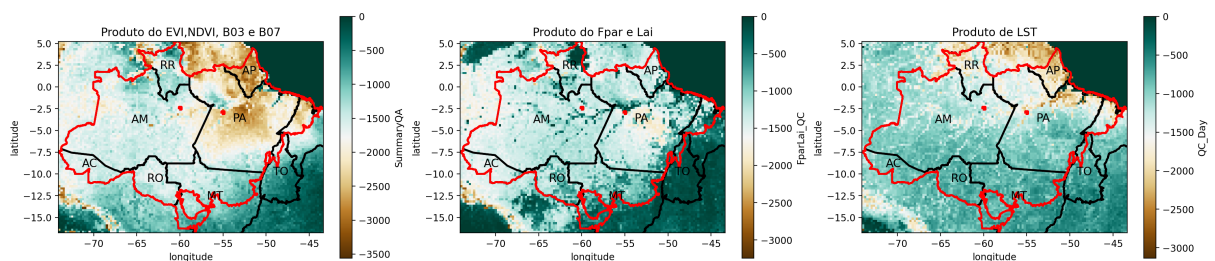
A Floresta Aleatória se destacou como o modelo com as melhores métricas de desempenho desta série de experimentos, o que o torna a escolha preferencial para ampliar a escala da estimativa de NEE na Amazônia Legal.

5.8 Ampliação de escala

A ampliação de escala é uma técnica que permite a extensão de conclusões e previsões de uma escala local para uma regional ou mais ampla. O realizado neste trabalho possui o objetivo de fornecer o comportamento espacial do NEE para a Amazônia Legal a partir do modelo RF com preditores RS e METEORS com variável alvo da torre K77. Esse foi o que obteve o melhor desempenho dentre todos os modelos analisados nos experimentos anteriores. Ele estimou a série temporal de 2002 a 2011 de cada célula presente na grade do Bioma Amazônia.

De acordo com [Tramontana et al. \(2016\)](#), a qualidade dos dados de sensoriamento remoto é afetada por fatores externos, como condições atmosféricas, cobertura de nuvens e estado da superfície terrestre e esses problemas podem ser minimizados com o uso do ciclo sazonal médio suavizado dos dados satelitais, que contém muito menos ruído para a estimativa de NEE. Neste trabalho foi utilizado este ciclo sazonal a partir do preenchimento diário dos dados de RS. No entanto, é importante entender como é a qualidade desses dados, pois por mais que haja a suavização dos elementos acima, os mesmos ainda são presentes na Região Amazônica, em grande escala e ocorrência. Além disso, é preciso analisar se esses dados impactam diretamente no comportamento da distribuição espacial das estimativas de NEE feitas pela RF. Uma vez que, neste trabalho, não foi utilizado nenhuma outra fonte de dados que mostrem o padrão de NEE nestes pontos distantes das torres de treinamento.

Figura 5.24 - Valores de baixa qualidade por célula para cada produto MODIS.

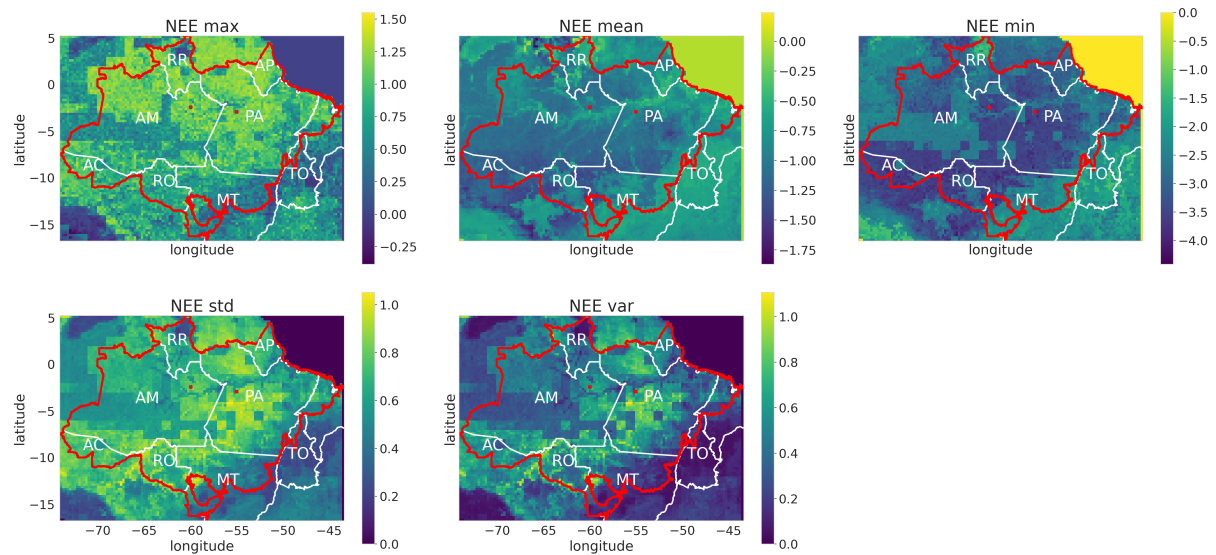


Na figura acima foi realizada a contabilização da quantidade de indicadores de baixa qualidade por célula. Foi contabilizado a quantidade e multiplicado por -1 para melhora da visualização da quantidade de indicadores.

A partir disso e dos indicadores de qualidade dos produtos de Sensoriamento Remoto utilizado neste trabalho, na Figura 5.8 foi identificado por célula a quantidade de dados que os produtos indicam como de baixa qualidade. A partir dessas informações é possível entender quais as regiões da Amazônia terão a estimativa de NEE realizada com dados de baixa qualidade.

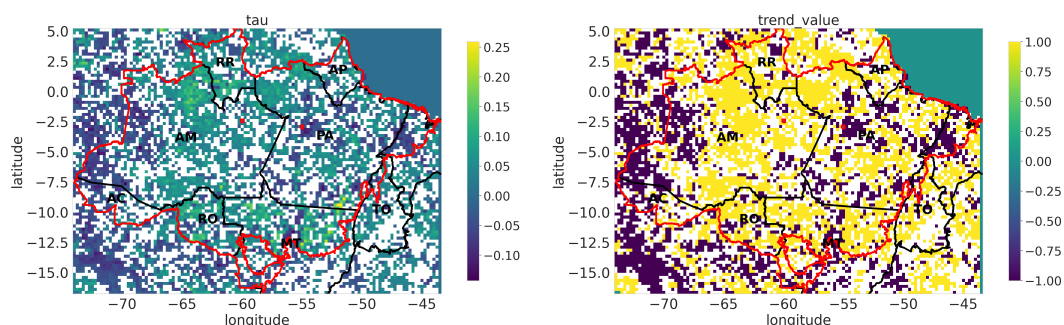
Observou-se na, Figura 5.8 , o indicador SummaryQA que a região mais impactada por dados de baixa qualidade está localizada na área central do Pará e em todo o estado do Amapá, destacada em bege na figura, com variações de 200 a 3000 dias (negativos para visualização) aproximadamente. Esse indicador afeta diretamente os valores de EVI, NDVI, B03 e B07. Já o indicador FparLaiQC do produto de LAI e FPAR do MODIS revelou (em branco) que em grande parte das células, em todo a Amazônia Legal, há aproximadamente 1500 dias de baixa qualidade para suas variáveis. A análise das tonalidades brancas e bege mostra que, para o produto de temperatura da superfície, a maior concentração de dados de baixa qualidade está na região ao norte do Pará e no estado do Amapá.

Figura 5.25 - Valores máximos, mínimos e médios de NEE (ampliação de escala) no bioma Amazônia. Estimativa realizada com preditores RS e modelo RF.



Na figura acima, os valores em tons de azul representam o comportamento neutro a sumidouro de carbono, enquanto os tons de verde até amarelo indicam comportamento de emissão de carbono.

Figura 5.26 - Tendência de Mann Kendall da estimativa de NEE para toda a Amazônia Legal com valor $p < 0,05$ para o modelo RF com preditores RS.



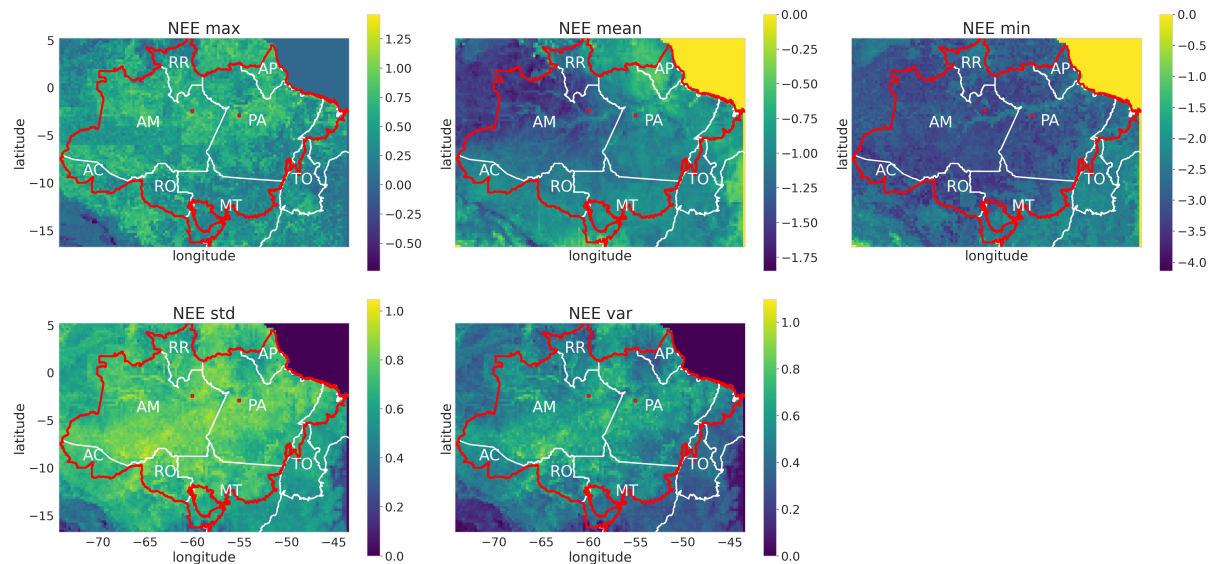
Resultado do Teste de Mann Kendall, com valor $p < 0,05$, do resultado de NEE a partir do modelo RF com preditores RS para a grade do Bioma Amazônia utilizada neste trabalho. Em vermelho há a delimitação do Bioma Amazônia e as células K34 e K67. Em preto há a delimitação de estados e suas siglas da Região Norte do Brasil.

Na Figura 5.25, as métricas estatísticas da ampliação de escala criada a partir do uso do modelo RF a partir de preditores RS são apresentadas. Observa-se que a variação de NEE e seus limites foram aparentemente influenciados pelos dados de radiação difusa do conjunto METEO, como pode ser visto pelo tamanho *pixels*, já que os dados de radiação difusa são provenientes de uma escala de menor resolução espacial, 1° .

O comportamento médio do NEE revela um padrão de sumidouro de carbono para todo o Bioma Amazônia, com comportamento mais intensificado e homogêneo nos estados do Amazonas, Acre, parte de Roraima e Pará. As regiões com maior emissão, evidenciadas pelos valores máximos, incluem o Norte do Pará, Amapá, Roraima e Amazonas. O estado do Pará apresenta a maior variação de NEE em relação aos valores médios. O que pode ser uma consequência da má qualidade dos dados empregados.

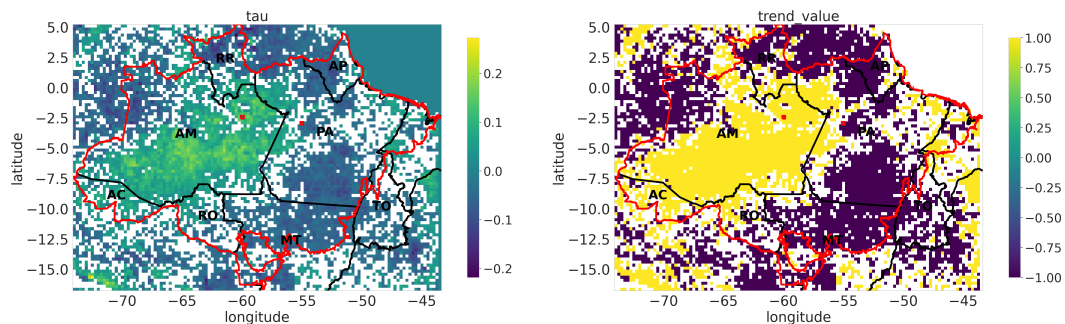
Utilizando a tendência de Mann Kendall para analisar o comportamento de NEE ao longo dos 10 anos, observou-se, na Figura 5.26 um aumento fraco no fluxo de carbono para grande parte do Bioma Amazônia. As áreas com maior aumento foram identificadas pelo comportamento de tau, destacando-se o Norte da Amazônia e o leste de Roraima, evidenciados por regiões mais amarelas. No Mato Grosso, também foi identificado um aumento significativo de NEE.

Figura 5.27 - Valores máximos, mínimos e médios de NEE (Ampliação de escala) no Bioma Amazônia. Estimativa realizada com Preditores METEORS e modelo RF.



Na figura acima, os valores em tons de azul representam o comportamento neutro a sumidouro de carbono, enquanto os tons de verde até amarelo indicam comportamento de emissão de carbono.

Figura 5.28 - Tendência de Mann Kendall da estimativa de NEE para todo o Bioma Amazônia com valor $p < 0,05$ para o modelo RF com preditores METEORS.

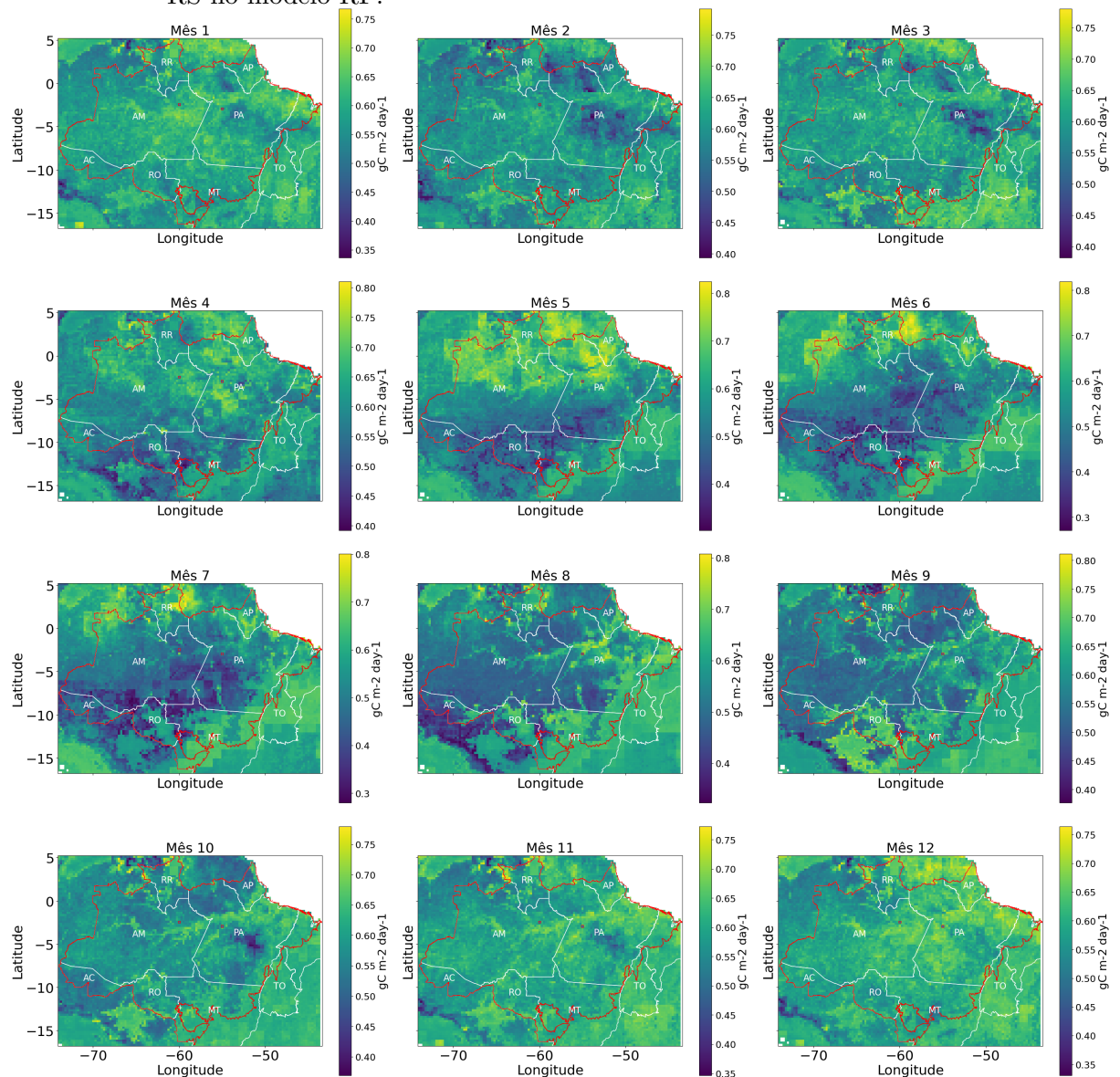


Resultado da tendência de Mann Kendall, com valor $p < 0,05$, do resultado de NEE a partir do modelo RF com preditores METEORS para a grade do Bioma Amazônia utilizada neste trabalho. Em vermelho há a delimitação do Bioma Amazônia e as células K34 e K67. Em preto há a delimitação de estados e suas siglas da Região Norte do Brasil.

Nas Figuras 5.27 e 5.28 há o resultado das estatísticas de NEE estimadas a partir do modelo RF com o uso de preditores METEORS. Este modelo demonstrou métricas de validação e teste muito similares ao modelo RF utilizando dados de RS. Foi observado que as métricas de máximo, mínimo, desvio padrão e variância exibiram comportamentos análogos, contudo, houve uma suavização aparente dos dados devido à impossibilidade de visualizar os *pixels* do dado de 1° dos dados de radiação difusa.

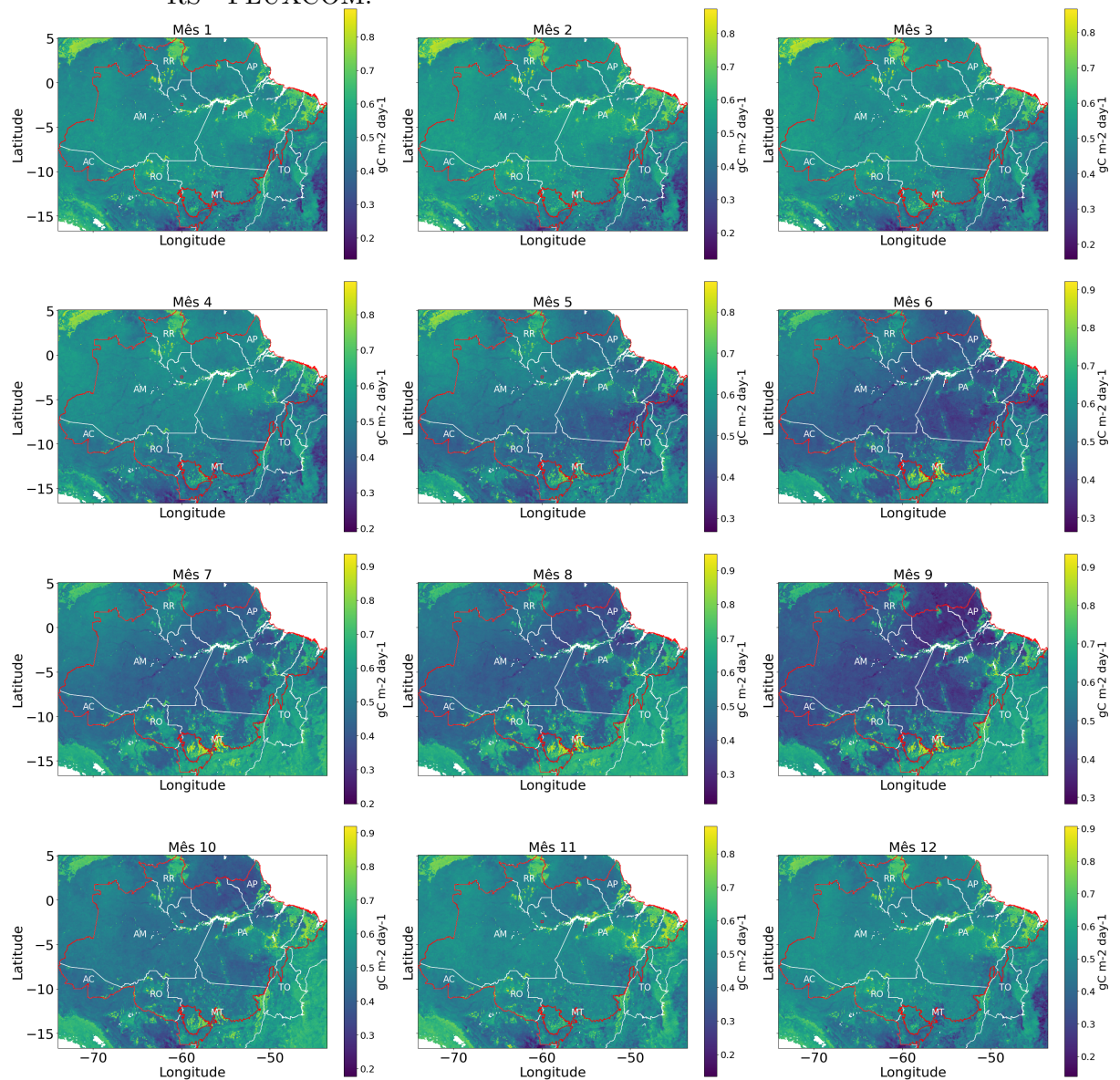
Quanto à tendência, identificou-se um aumento na NEE no Estado do Amazonas e uma tendência de diminuição para a maior parte do restante do Bioma. Essa tendência contrasta com o comportamento observado na Figura 5.26.

Figura 5.29 - Comportamento mensal médio de NEE criado a partir de dados preditores RS no modelo RF.



Na figura acima, os valores em tons de azul representam o comportamento neutro a sumidouro de carbono, enquanto os tons de verde até amarelo indicam comportamento de emissão de carbono.

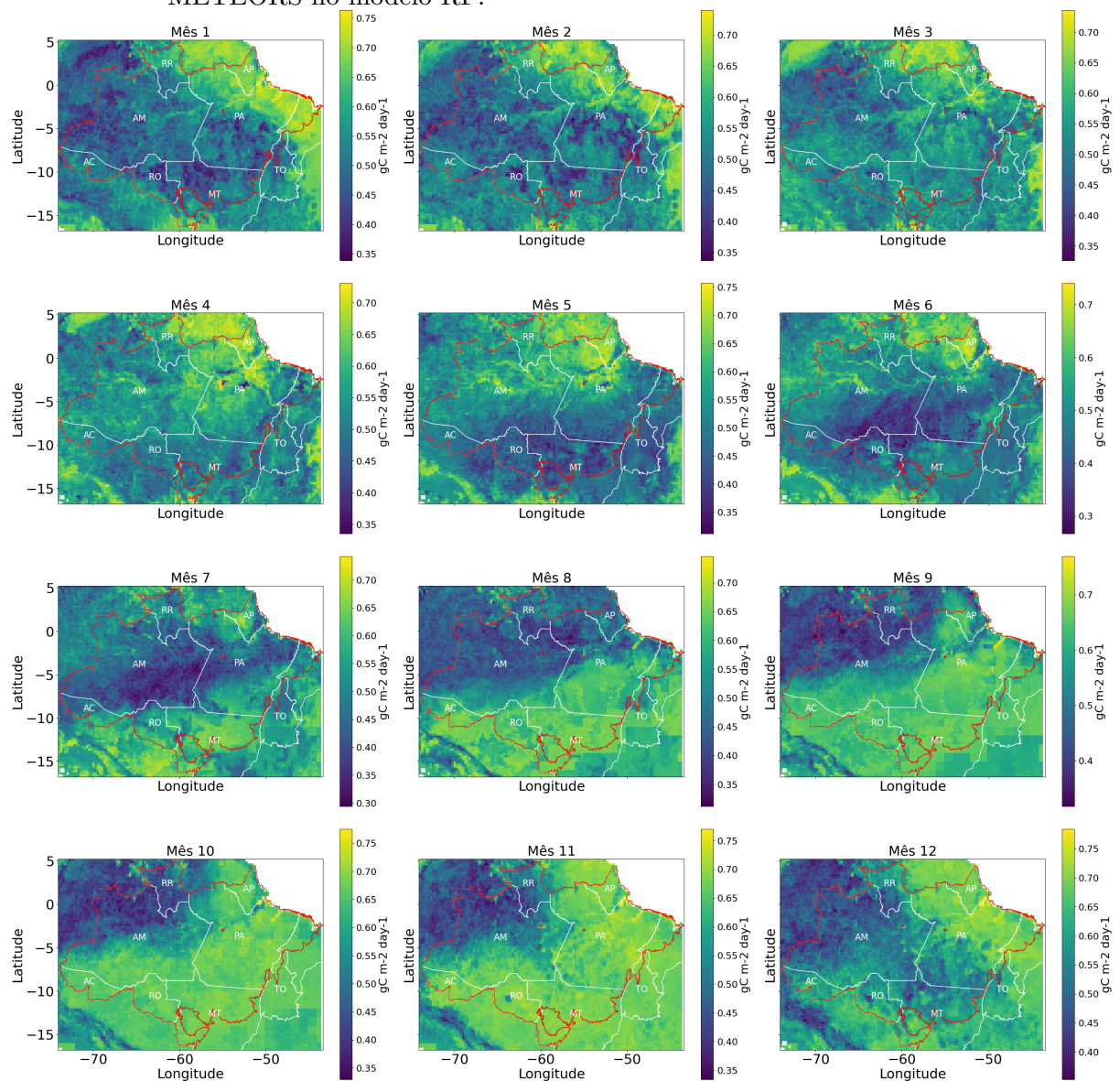
Figura 5.30 - Comportamento mensal médio de NEE criado a partir de dados preditores RS - FLUXCOM.



Na figura acima, os valores em tons de azul representam o comportamento neutro a sumidouro de carbono, enquanto os tons de verde até amarelo indicam comportamento de emissão de carbono.

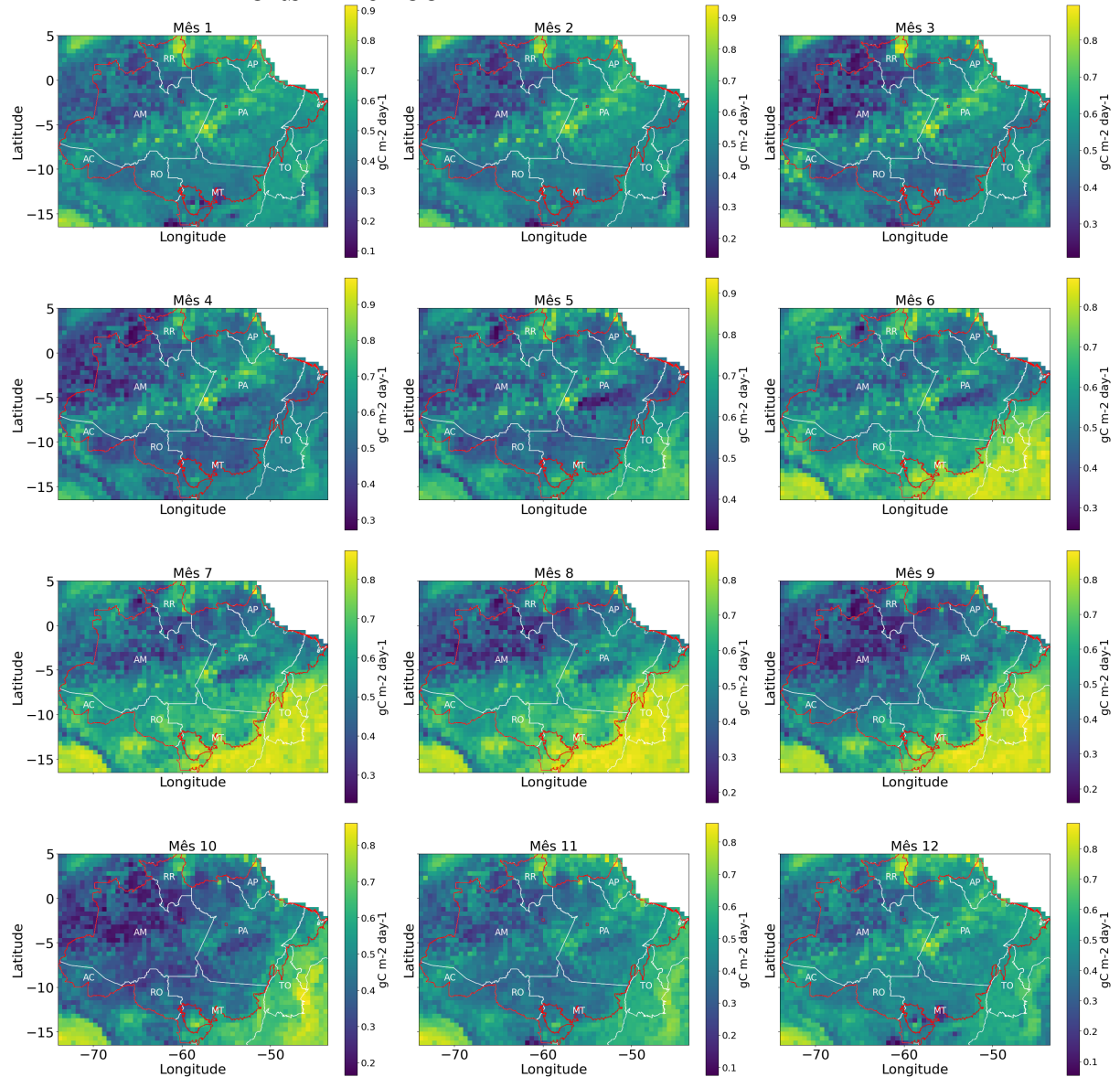
A análise do comportamento médio mensal e anual de NEE para o Bioma Amazônia, nas Figuras 5.29 e 5.30, entre 2002 e 2011, em comparação com os resultados da Fluxcom, revelou uma grande diferença de comportamento visual entre os dados dos modelos RF e Fluxcom com preditores RS. Ao examinar o comportamento médio mensal, observou-se que o modelo RF-RS possui maior variabilidade espacial de NEE do que o modelo Fluxcom. Essa variabilidade é tanto ao decorrer da Amazônia Legal e ao decorrer das médias mensais.

Figura 5.31 - Comportamento mensal médio de NEE criado a partir de dados preditores METEORS no modelo RF.



Na figura acima, os valores em tons de azul representam o comportamento neutro de sumidouro de carbono, enquanto os tons de verde até amarelo indicam comportamento de emissão de carbono.

Figura 5.32 - Comportamento mensal médio de NEE do criado a partir de dados preditores METEORS - FLUXCOM.



Na figura acima, os valores em tons de azul representam o comportamento neutro a sumidouro de carbono, enquanto os tons de verde até amarelo indicam comportamento de emissão de carbono.

A ampliação de escala realizada a partir do modelo RF com preditores METEORS, que se encontram nas Figuras 5.32 e 5.31 exibiu um comportamento distinto em relação ao modelo Fluxcom com preditores METEORS.

Tabela 5.13 - Métricas de comparação dos resultados dos modelos RF e Fluxcom com a Torre K77.

Melhor Modelo	FLUXCOM	RF	FLUXCOM	RF
Torre da Variável Alvo	Torre K77	Torre K77	Torre K77	Torre K77
Preditor	METEORS	METEORS	RS	RS
MEF	-0,44	0,84	-0,33	0,93
MSE	3,12	0,34	2,88	0,15
ROV	0,19	0,70	0,21	0,81
R	-0,46	0,95	-0,17	0,98
R²	0,21	0,91	0,03	0,96
BIAS	-0,71	-0,07	-0,68	0,02

De um modo geral, o comportamento de NEE para a Amazônia Legal estimado pelos modelos RF-RS e RF-METEORS são discrepantes entre si e entre os resultados da Fluxcom. Com isso para o melhor entendimento do quanto cada um desses modelos se aproximam da realidade foi calculado métricas estatísticas para comparação entre esses modelos e os valores médios mensais medidos pela torre K77, na área co-localizada com a mesma, essas métricas se encontram na Tabela 5.13 . A partir disso, foi visto que os modelos da Fluxcom mostraram os maiores erros (MSE e RMSE), os piores coeficientes de ajuste e eficiência (MEF, R, R²), a menor variabilidade dos dados (ROV) e maior viés negativo em relação ao dados medidos pela torre K77.

Já os modelos RF-RS e RF-METEORS exibiram melhor estimativa e aproximação em relação aos dados medidos pela torre K77, com menores erros, bons coeficientes de ajuste e eficiência e maior variabilidade dos dados. Tal comportamento é esperado devido ao fato de que os modelos foram criados para ajuste aos dados dessa torre. No entanto, isso também evidencia que foram encontrados melhores resultados na ampliação da escala para a Amazônia Legal a partir dos modelos deste trabalho em relação ao modelo global Fluxcom.

Figura 5.33 - Gráficos de Dispersão Comparativos dos Dados Observados da Torre K77 e Simulados pelos modelos Fluxcom e RF, com preditores METEORS, para a Etapa de Testes na estimativa de NEE.

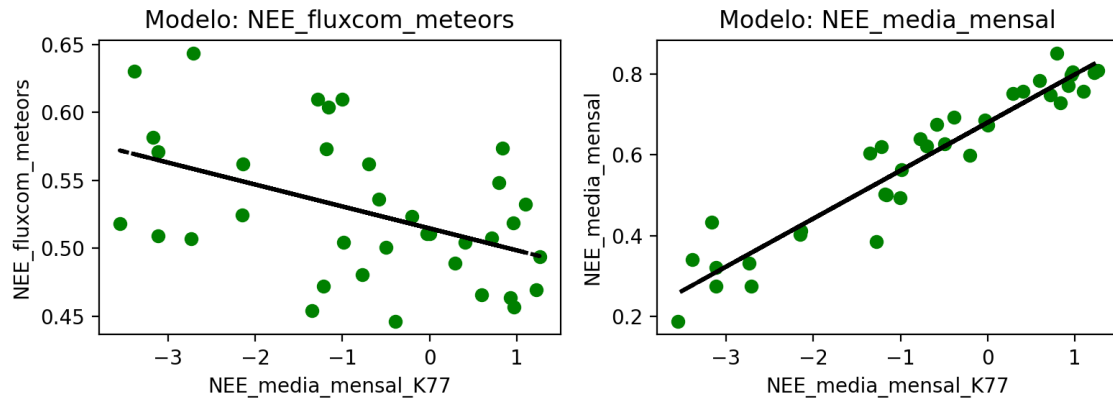
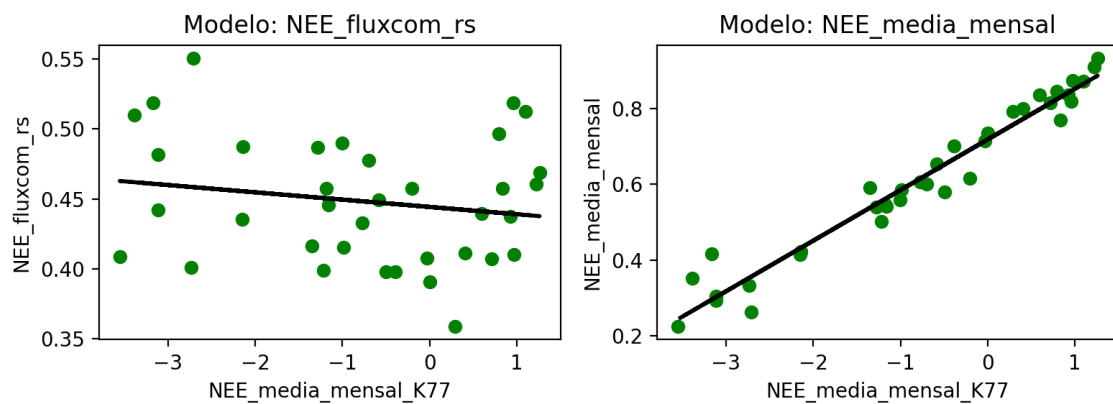


Figura 5.34 - Gráficos de Dispersão Comparativos dos Dados Observados da Torre K77 e Simulados pelos modelos Fluxcom e RF, com preditores RS, para a Etapa de Testes na estimativa de NEE.



Os melhores ajustes mostrados nas métricas da Tabela 5.13 podem ser vistos nas Figuras 5.33 e 5.34, onde ficou evidenciado que os modelos RF RS e RF METEORS mostraram melhores estimativas que os modelos globais da Fluxcom para a região da Amazônia Legal.

6 CONCLUSÕES

Essa pesquisa teve como objetivo principal a estimativa de NEE para a Amazônia Legal, utilizando uma abordagem baseada em modelos de aprendizado de máquina e preditores provenientes de produtos meteorológicos e de sensoriamento remoto. Para alcançar esse objetivo principal foram delineados objetivos específicos que envolveram a análise estatística e exploratória das variáveis alvo e preditoras, a avaliação de diferentes algoritmos de aprendizado de máquina e a análise da capacidade preditiva das variáveis de entrada utilizadas.

A análise estatística dos preditores mostrou que todos possuem distribuição não normal. Na análise exploratória foi visto as principais diferenças entre cada uma das células, sendo que a célula K67 mostrou uma dinâmica mais intensa de evaporação e interações térmicas, enquanto a célula K34 demonstrou uma maior disponibilidade de água. A célula K67 mostrou maior complexidade em relação aos elementos de uso do solo que afetam sua disponibilidade de água e absorção de radiação, enquanto a célula K34 mostrou que seu comportamento é influenciado pela Floresta.

A partir da análise exploratória das variáveis alvo foi visto que há muitos dados faltantes e dados provenientes de preenchimentos médios nos conjuntos de dados LBA e Fluxnet, o que gera uma limitação dos dados de boa qualidade e uma análise profunda de como é o comportamento de NEE para cada uma das torres e o quanto esse comportamento é intrínseco aos ecossistemas de cada uma delas. No entanto, ainda foi possível ver que o comportamento mensal do NEE medido pelas torres K83 e K67 são semelhantes, principalmente devido ao fato de que ambas estão em uma região de floresta. Foi visto que o comportamento do NEE na torre K77 é governado pelo cultivo existente na área da torre. Já para a Torre K34 não foi possível se aproximar de algum padrão de comportamento, pois a mesma possui mais de 90% de dados preenchidos pelo ciclo anual médio.

Com relação aos modelos de aprendizado de máquina, os resultados indicaram que o Random Forest e o Gradient Boosting foram os algoritmos que apresentaram as melhores métricas de desempenho na estimativa da NEE. Além disso, os dados de sensoriamento remoto provenientes dos sensores MODIS e CERES demonstraram um bom ajuste aos dados da troca líquida de carbono, contribuindo para uma melhor capacidade preditiva do modelo.

Os resultados também mostraram que o uso de dados específicos que representam o funcionamento dos ecossistemas amazônicos resultou em uma melhora significa-

tiva na capacidade preditiva da troca líquida de carbono em comparação com os resultados encontrados no FLUXCOM. Isso foi evidenciado pelo desempenho do modelo, que superou os valores de capacidade preditiva de NEE global e os valores encontrados anteriormente para a região de estudo.

Destaca-se ainda que os melhores resultados foram obtidos a partir de dados da iniciativa LBA. O que ressalta a importância desses dados para a validação e atuação como variável alvo nos modelos. No entanto, a falta de dados representativos de outras torres na região amazônica (como a K34) limitou a diversidade dos dados de treinamento e teste, prejudicando o ajuste dos modelos e a validação da ampliação de escala.

Para trabalhos futuros, sugere-se a expansão da análise para incluir dados de outras torres na região amazônica, a fim de melhorar a representatividade dos modelos e validar a ampliação de escala em uma variedade de condições ambientais. No entanto para isso talvez seja necessário a inclusão de dados de torres de fluxo a partir de 1999, se estes não tiverem muitos dados preenchidos, o que pode ser valioso para a melhoria da estimativa. Com isso, será necessário a mudança da fonte dos preditores de sensoriamento remoto, pois o produto utilizado neste trabalho só é disponibilizado a partir de 2002.

É essencial utilizar outros produtos do ciclo de carbono, como a produtividade primária bruta (GPP) e a respiração (Re), para compreender a dinâmica interanual ou anual do NEE na Amazônia Legal em relação ao mapa de ampliação de escala encontrado neste estudo e em outros. Foram obtidas estimativas de NEE satisfatórias neste trabalho, destacando-se a significativa influência da interação entre a radiação e o dossel florestal, como evidenciado pelos melhores resultados com preditores RS. Seria interessante explorar outras variáveis da interação radiativa para essa estimativa, tais como a radiação líquida, SIF (fluorescência induzida por radiação solar), entre outros produtos que estão sendo desenvolvidos para uma melhor compreensão do comportamento do ciclo do carbono na floresta Amazônica.

Ainda, como sugestão de trabalhos futuros, análises de seleção de atributos de maneira assíncrona devem ser realizadas para se verificar como algumas variáveis preditoras se comportam com relação ao NEE. Além disso, entender a distribuição das variáveis METEO que mais se adequam às previsões de NEE, o qual permita a sua modelagem estatística para uso em modelos de mudanças climáticas.

É importante também entender como o uso de solo afeta cada variável dos ecos-

sistemas e mais especificamente o NEE. Realizar trabalhos parecidos com este com uma maior resolução espacial do conjunto de dados de modo a obter uma análise mais detalhada para diferentes usos de solo. Pode ser verificado a possibilidade de implementar multi-modelos na escolha de regiões específicas para que seja possível a identificação de regiões que sofrem com o regime local de NEE.

Algumas técnicas de clusterização podem ser usadas com esse intuito. É possível experimentar o uso de dados de outras regiões do Brasil ou do continente da América do Sul para identificar tele-conexões climáticas que ajudem a entender a mudança de NEE em escalas temporais mais longas. De modo geral, testar novos modelos e preditores que venham a ser desenvolvidos no futuro é um processo contínuo para obter a melhor estimativa.

REFERÊNCIAS BIBLIOGRÁFICAS

- AKIBA, T.; SANO, S.; YANASE, T.; OHTA, T.; KOYAMA, M. **Optuna: A Next-generation Hyperparameter Optimization Framework**. 2019. Disponível em: <<<https://arxiv.org/abs/1907.10902>>>. 57, 97
- ARAGÃO, L. E. O. C.; POULTER. Environmental change and the carbon balance of Amazonian forests. **Biological Reviews**, v. 89, n. 4, 2014. Disponível em: <<<https://onlinelibrary.wiley.com/doi/abs/10.1111/brv.12088>>>. 1, 8
- ARAÚJO, A. C.; NOBRE, A. D.; KRUIJT, B.; ELBERS, J. A.; DALLAROSA, R.; STEFANI, P.; RANDOW, C. von; MANZI, A. O.; CULF, A. D.; GASH, J. H. C.; VALENTINI, R.; KABAT, P. Comparative measurements of carbon dioxide fluxes from two nearby towers in a central Amazonian rainforest: the Manaus LBA site. **Journal of Geophysical Research: Atmospheres**, v. 107, n. D20, p. LBA 58–1–LBA 58–20, 2002. Disponível em: <<<https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2001JD000676>>>. 35
- ARAÚJO, A. D. A.C de. The spatial variability of co2 storage and the interpretation of eddy covariance fluxes in central Amazonia. **Agricultural and Forest Meteorology**, v. 150, n. 2, 2010. ISSN 0168-1923. Disponível em: <<<https://www.sciencedirect.com/science/article/pii/S0168192309002706>>>. 13
- ASSIS, F. Terrabrasilis: a spatial data analytics infrastructure for large-scale thematic mapping. **ISPRS International Journal of Geo-Information**, v. 8, n. 513, 2019. 7
- BALDOCCHI, D. D. Assessing the eddy covariance technique for evaluating carbon dioxide exchange rates of ecosystems: past, present and future. **Global Change Biology**, v. 9, n. 4, 2003. Disponível em: <<<https://onlinelibrary.wiley.com/doi/abs/10.1046/j.1365-2486.2003.00629.x>>>. 11
- BARKHORDARIAN, A.; SAATCHI. A recent systematic increase in vapor pressure deficit over tropical South America. **Scientific Reports**, v. 9, n. 1, 2019. ISSN 2045-2322. Disponível em: <<<https://doi.org/10.1038/s41598-019-51857-8>>>. 9, 10
- BAUER, L. d. O. **Predição de trocas de carbono entre a biosfera e a atmosfera na Flona-Tapajós a partir de variáveis ambientais**. Dissertação (Mestrado em Análise Ambiental Integrada) — Universidade Federal de São Paulo, Diadema, 2022. 14, 16, 18, 30, 61, 96
- BITENCOURT, D. P. **Estudo dos fluxos turbulentos de calor sensível e latente no fundo do vale do Rio da Prata**. Tese (Doutorado em Física) — Universidade Federal de Santa Maria, Santa Maria, RS, 2008. 11, 12

BREIMAN, L. **Random forests**. New York, NY: Springer, 2001. Disponível em: <<<https://www.stat.berkeley.edu/~breiman/randomforest2001.pdf>>>. 17

BRODY, S.; ALON, U.; YAHAV, E. **How attentive are graph attention networks?** 2022. Disponível em: <<<https://arxiv.org/abs/2105.14491>>>. 24

CAI, J.; XU, K.; ZHU, Y.; HU, F.; LI, L. Prediction and analysis of net ecosystem carbon exchange based on gradient boosting regression and random forest. **Applied Energy**, v. 262, p. 114566, 2020. ISSN 0306-2619. Disponível em: <<<https://www.sciencedirect.com/science/article/pii/S0306261920300787>>>. 17

CAVALCANTI, N. F. I. **Tempo e clima no Brasil**. [S.l.]: Oficina de Textos, 2009. ISBN 978-85-86238-92-5. 28

CONOVER, W. J. **Practical nonparametric statistics**. 2. ed. New York: John Wiley & Sons, 1999. 53, 82

CUI, X.; E, S.; NIU, D.; CHEN, B.; FENG, J. Forecasting of carbon emission in china based on gradient boosting decision tree optimized by modified whale optimization algorithm. **Sustainability**, v. 13, n. 21, 2021. ISSN 2071-1050. Disponível em: <<<https://www.mdpi.com/2071-1050/13/21/12302>>>. 17

DANASINGH, A. A. G. S.; SUBRAMANIAN, A. a. B.; EPIPHANY, J. L. Identifying redundant features using unsupervised learning for high-dimensional data. **SN Applied Sciences**, v. 2, n. 8, p. 1367. ISSN 2523-3971. Disponível em: <<<https://doi.org/10.1007/s42452-020-3157-6>>>. 82, 83

DIDAN, K. **MODIS/Terra vegetation indices 16-Day L3 global 1km SIN Grid V061**. Washington: NASA: NASA EOSDIS land processes distributed active archive center, 2021. Disponível em: <<<https://doi.org/10.5067/MODIS/MOD13A2.061>>>. Acesso em: 25 dez 2023. 36, 39

DOU, X.; YANG, Y.; LUO, J. Estimating forest carbon fluxes using machine learning techniques based on eddy covariance measurements. **Sustainability**, v. 10, n. 1, 2018. ISSN 2071-1050. Disponível em: <<<https://www.mdpi.com/2071-1050/10/1/203>>>. 18, 21

FEY, M.; LENSSSEN, J. E. **Fast Graph Representation Learning with PyTorch Geometric**. 2019. Disponível em: <<<https://arxiv.org/abs/1903.02428>>>. 59

FISHER, R. A.; KOVEN, C. D. Perspectives on the future of land surface models and the challenges of representing complex terrestrial systems. **Journal of Advances in Modeling Earth Systems**, v. 12, n. 4, 2020. Disponível em: <<<https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2018MS001453>>>. 3, 15

FLUXNET. **About the FLUXNET network**. 2023. Disponível em: <<<https://fluxnet.org/about/>>>. Acesso em: 15 fev. 2023. 3, 31, 32, 33

FOKEN, T.; AUBINET, M.; LEUNING, R. The eddy covariance method. In: _____. **Eddy covariance: a practical guide to measurement and data analysis**. Dordrecht: Springer Netherlands, 2012. p. 1–19. ISBN 978-94-007-2351-1. Disponível em: <<https://doi.org/10.1007/978-94-007-2351-1_1>>. 11, 12

FOREST, A. **Dewpoint, vapor pressure deficit VPD and relative humidity calculations in the Andrews Forest weather database**. Disponível em: <<https://andrewsforest.oregonstate.edu/sites/default/files/lter/data/studies/ms01/dewpt_vpd_calculations.pdf>>. 44

GATTI, L. V.; BASSO, L. S.; MILLER, J. B.; GLOOR, M.; DOMINGUES, L. G.; CASSOL, H. L. G.; TEJADA, G.; ARAGÃO, L. E. O. C.; NOBRE, C.; PETERS, W.; MARANI, L.; ARAI, E.; SANCHES, A. H.; CORRÊA, S. M.; ANDERSON, L.; RANDOW, C. V.; CORREIA, C. S. C.; CRISPIM, S. P.; NEVES, R. A. L. Amazonia as a carbon source linked to deforestation and climate change. **Nature**, v. 595, n. 7867, p. 388–393. ISSN 1476-4687. Disponível em: <<<https://doi.org/10.1038/s41586-021-03629-6>>>. 2, 8, 9

GIOVANNI, F. D.; ROWBOTTOM, J.; CHAMBERLAIN, B. P.; MARKOVICH, T.; BRONSTEIN, M. M. **Understanding convolution on graphs via energies**. 2023. Disponível em: <<<https://arxiv.org/abs/2206.10991>>>. 21

GITELSON, A. A.; VIÑA. Relationship between gross primary production and chlorophyll content in crops: implications for the synoptic monitoring of vegetation productivity. **Journal of Geophysical Research: Atmospheres**, v. 111, n. D8, 2006. Disponível em: <<<https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2005JD006017>>>. 16

GOULDEN, M. L.; MILLER, S. D.; ROCHA, H. R. D.; MENTON, M. C.; FREITAS, H. C.; FIGUEIRA, A. M. S.; SOUSA, C. A. D. Diel and seasonal patterns of tropical forest co2 exchange. **Ecological Applications**, v. 14, n. sp4, p. 42–54, 2004. 35

HERSBACH, H.; BELL. **ERA5 hourly data on single levels from 1940 to present**. 2023. Disponível em: <<<https://doi.org/10.24381/cds.adbb2d47>>>. Acesso em: 25 dez. 2023. 41, 42

HUSSAIN, M.; MAHMUD, I. pymannkendall: a python package for non parametric mann kendall family of trend tests. **Journal of Open Source Software**, v. 4, n. 39, p. 1556, 2019. 53

HUTYRA, L. R.; MUNGER, J. W.; HAMMOND-PYLE, E.; SALESKA, S. R.; RESTREPO-COUBE, N.; DAUBE, B. C.; CAMARGO, P. B. d.; WOFYSY, S. C. Resolving systematic errors in estimates of net ecosystem exchange of co2 and ecosystem respiration in a tropical forest biome. **Agricultural and Forest Meteorology**, 2008. 30

HUTYRA, L. R.; MUNGER, J. W.; SALESKA, S. R.; GOTTLIEB, E.; DAUBE, B. C.; DUNN, A. L.; AMARAL, D. F.; CAMARGO, P. B.; WOFSY, S. C. Seasonal controls on the exchange of carbon and water in an Amazonian rain forest. **Journal of Geophysical Research**, v. 112, n. G3, p. 1–16, 2007. 35

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (IBGE). **Manual técnico da vegetação brasileira: sistema fitogeográfico: inventário das formações florestais e campestres: técnicas e manejo de coleções botânicas: procedimentos para mapeamentos**. 2. ed. Rio de Janeiro: IBGE, 2012. 91

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (IBGE). **Biomass e sistema costeiro-marinho do Brasil : compatível com a escala 1:250 000**. Rio de Janeiro: IBGE, 2019. ISBN 978-8-52-404510-3. Disponível em: <<<https://biblioteca.ibge.gov.br/index.php/biblioteca-catalogo?view=detalhes&id=2101676>>>. 7, 27, 28

INTERGOVERNMENTAL PANEL ON CLIMATE CHANGE (IPCC). Summary for policymakers. In: MASSON-DELMOTTE, V.; ZHAI, P.; PIRANI, A.; CONNORS, S.; PéAN, C.; BERGER, S.; CAUD, N.; CHEN, Y.; GOLDFARB, L.; GOMIS, M.; HUANG, M.; LEITZELL, K.; LONNOY, E.; MATTHEWS, J.; MAYCOCK, T.; WATERFIELD, T.; YELEKçİ, O.; YU, R.; ZHOU, B. e. (Ed.). **Climate change 2021: the physical science basis. Contribution of working group I to the sixth assessment report of the intergovernmental panel on climate change**. [S.l.]: IPCC, 2021. 2

ISLAM, K. I. Exploring random forest machine learning and remote sensing data for streamflow prediction: an alternative approach to a process-based hydrologic modeling in a snowmelt-driven watershed. **Remote Sensing**, v. 15, n. 16, 2023. ISSN 2072-4292. Disponível em: <<<https://www.mdpi.com/2072-4292/15/16/3999>>>. 17

IZBICKI, R.; SANTOS, T. M. **Aprendizado de máquina: uma abordagem estatística**. [S.l.]: São Carlos - SP - Brazil, 2020. ISBN 9786500024104. 18

JACKSON, D.; LAUNDER, B. Osborne reynolds and the publication of his papers on turbulent flow. **Annual Review of Fluid Mechanics**, v. 39, n. 1, 2007. Disponível em: <<<https://doi.org/10.1146/annurev.fluid.39.050905.110241>>>. 11

JANG, J.-S. Anfis: adaptive-network-based fuzzy inference system. **IEEE Transactions on Systems, Man, and Cybernetics**, v. 23, n. 3, 1993. 19

JIANG, J. L. W. Graph neural network for traffic forecasting: a survey. **Expert Systems with Applications**, v. 207, nov 2022. Disponível em: <<<https://doi.org/10.10162Fj.eswa.2022.117921>>>. 21

JUNG, M.; REICHSTEIN, M.; BONDEAU, A. Towards global empirical upscaling of FLUXNET eddy covariance observations: validation of a model tree ensemble approach using a biosphere model. **Biogeosciences**, v. 6, n. 10, 2009. Disponível em: <<<https://bg.copernicus.org/articles/6/2001/2009/>>>. 15, 24, 32, 36

JUNG, M.; SCHWALM, C.; MIGLIAVACCA. Scaling carbon fluxes from eddy covariance sites to globe: synthesis and evaluation of the FLUXCOM approach. **Biogeosciences**, v. 17, n. 5, 2020. Disponível em: <<<https://bg.copernicus.org/articles/17/1343/2020/>>>. 4, 14, 15, 16, 24, 25, 27, 61, 62, 96, 101

KENDALL, M. G. **Rank correlation methods**. 4. ed. London: Charles Griffin, 1975. 53

KIPF, T. N.; WELLING, M. Semi-supervised classification with graph convolutional networks. 2017. Disponível em: <<<https://arxiv.org/abs/1609.02907>>>. 22

KONDO, M.; PATRA. State of the science in reconciling top-down and bottom-up approaches for terrestrial co2 budget. **Global Change Biology**, v. 26, n. 3, p. 1068–1084, 2020. Disponível em: <<<https://onlinelibrary.wiley.com/doi/abs/10.1111/gcb.14917>>>. 2

LARGE SCALE BIOSPHERE ATMOSPHERE EXPERIMENT IN AMAZONIA LBA. **LBA**. 2006. Disponível em: <<<http://lba.cptec.inpe.br/lba>>>. Acesso em: 27 fev. 2023. 33

LESKOVEC, J. **Stanford CS224W: machine learning with graphs**. Stanford University, 2021. Disponível em: <<<https://cs224w.stanford.edu>>>. 21, 22, 23

LLOYD, J.; GRACE, J.; MIRANDA, A. C.; MEIR, P.; WONG, S. C.; MIRANDA, H. S.; WRIGHT, I. R.; GASH, J. H. C.; MCINTYRE, J. A simple calibrated model of Amazon rainforest productivity based on leaf biochemical properties. **Plant Cell and Environment**, v. 18, p. 1129–1145, 1995. Disponível em: <<<https://api.semanticscholar.org/CorpusID:85986452>>>. 9

MARENCO, R. A.; ANTEZANA-VERA, S. A.; GOUVEA, P. R. d. S.; CAMARGO, M. A. B.; OLIVEIRA, M. F. d.; SANTOS, J. K. d. S. Fisiologia de espécies florestais da Amazônia: fotossíntese, respiração e relações hídricas. **Revista Ceres**, v. 61, p. 786–799, Nov 2014. ISSN 0034-737X. Disponível em: <<<https://doi.org/10.1590/0034-737x201461000004>>>. 81

MILLER, S. D.; GOULDEN, M. L.; MENTON, M. C.; ROCHA, H. R. da; FREITAS, H. C. de; FIGUEIRA, A. M. e. S.; DIAS, C. A. Biometric and micrometeorological measurements of tropical forest carbon balance. **Ecological Applications**, v. 14, n. sp4, p. 114–126, 2004. Disponível em: <<<https://esajournals.onlinelibrary.wiley.com/doi/abs/10.1890/02-6005>>>. 35

MURRAY, F. W. On the computation of saturation vapor pressure. **Journal of Applied Meteorology and Climatology**, v. 6, n. 1, 1967. Disponível em: <<https://journals.ametsoc.org/view/journals/apme/6/1/1520-0450_1967_006_0203_otcosv_2_0_co_2.xml>>. 44

MYNENI, R.; KNYAZIKHIN. **MODIS/Terra+Aqua leaf area index/FPAR 4-Day L4 global 500 m SIN Grid V061 – MCD15A3H**. Washington: NASA: [s.n.], 2021. Disponível em: <<<https://doi.org/10.5067/MODIS/MCD15A3H.061>>>. Acesso em: 25 dez. 2023. 36, 38

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION (NASA). **CERES and GEO-Enhanced TOA, within-atmosphere and surface fluxes, clouds and aerosols daily Terra-Aqua edition4A**. 2017. Disponível em: <<https://doi.org/10.5067/Terra+Aqua/CERES/SYN1degDay_L3.004A>>. 40, 41

NOBRE, C.; ENCALADA, A.; ANDERSON, E.; NEVES, E. G. **Science panel for the Amazon: Amazon assessment report 2021: executive summary**. [S.l.]: Science panel for the Amazon, 2021. 85, 88, 94

PAGAN, B. R.; MAES, W. H.; GENTINE, P.; MARTENS, B.; MIRALLES, D. G. Exploring the potential of satellite solar-induced fluorescence to constrain global transpiration estimates. **Remote Sensing**, v. 11, n. 4, 2019. ISSN 2072-4292. Disponível em: <<<https://www.mdpi.com/2072-4292/11/4/413>>>. 7

PALACIOS, R. d. S.; PAULO de. Temporal evolution of vapor pressure deficit observed in six locations of different brazilian ecosystems and its relationship with micrometeorological variables. **Forests**, v. 14, n. 8, 2023. ISSN 1999-4907. Disponível em: <<<https://www.mdpi.com/1999-4907/14/8/1543>>>. 44

PASTORELLO, G.; TROTTA, C.; CANFORA, E.; CHU, H.; CHRISTIANSON, D.; CHEAH, Y.-W.; POINDEXTER, C.; CHEN, J.; ELBASHANDY, A.; HUMPHREY, M.; ISAAC, P.; POLIDORI, D.; REICHSTEIN, M.; RIBECA, A.; INGEN, C. van; VUICHARD, N.; ZHANG, L.; AMIRO, B.; AMMANN, C.; ARAIN, M. A.; ARDö, J.; ARKEBAUER, T.; ARNDT, S. K.; ARRIGA, N.; AUBINET, M.; AURELA, M.; BALDOCCHI, D.; BARR, A.; BEAMESDERFER, E.; MARCHESINI, L. B.; BERGERON, O.; BERINGER, J.; BERNHOFER, C.; BERVEILLER, D.; BILLESBACH, D.; BLACK, T. A.; BLANKEN, P. D.; BOHRER, G.; BOIKE, J.; BOLSTAD, P. V.; BONAL, D.; BONNEFOND, J.-M.; BOWLING, D. R.; BRACHO, R.; BRODEUR, J.; BRÜMMER, C.; BUCHMANN, N.; BURBAN, B.; BURNS, S. P.; BUYASSE, P.; CALE, P.; CAVAGNA, M.; CELLIER, P.; CHEN, S.; CHINI, I.; CHRISTENSEN, T. R.; CLEVERLY, J.; COLLALTI, A.; CONSALVO, C.; COOK, B. D.; COOK, D.; COURSOLE, C.; CREMONESE, E.; CURTIS, P. S.; D'ANDREA, E.; ROCHA, H. da; DAI, X.; DAVIS, K. J.; CINTI, B. D.; GRANDCOURT, A. d.; LIGNE, A. D.; OLIVEIRA, R. C. D.; DELPIERRE, N.; DESAI, A. R.; BELLA, C. M. D.; TOMMASI, P. d.; DOLMAN, H.; DOMINGO, F.; DONG, G.; DORE, S.; DUCE, P.; DUFRÊNE, E.; DUNN, A.; DUŠEK, J.; EAMUS, D.; EICHELMANN, U.; ElKhidir, H. A. M.; EUGSTER, W.; EWENZ, C. M.; EWERS, B.; FAMULARI, D.; FARES, S.; FEIGENWINTER, I.; FEITZ, A.; FENSHOLT, R.; FILIPPA, G.; FISCHER, M.; FRANK, J.; GALVAGNO, M.; GHARUN, M.; GIANELLE, D.; GIELEN, B.; GIOLI, B.; GITELSON, A.; GODED, I.; GOECKEDE, M.; GOLDSTEIN, A. H.;

GOUGH, C. M.; GOULDEN, M. L.; GRAF, A.; GRIEBEL, A.; GRUENING, C.; GRÜNWARD, T.; HAMMERLE, A.; HAN, S.; HAN, X.; HANSEN, B. U.; HANSON, C.; HATAKKA, J.; HE, Y.; HEHN, M.; HEINESCH, B.; HINKO-NAJERA, N.; HÖRTNAGL, L.; HUTLEY, L.; IBROM, A.; IKAWA, H.; JACKOWICZ-KORCZYNSKI, M.; JANOUŠ, D.; JANS, W.; JASSAL, R.; JIANG, S.; KATO, T.; KHOMIK, M.; KLATT, J.; KNOHL, A.; KNOX, S.; KOBAYASHI, H.; KOERBER, G.; KOLLE, O.; KOSUGI, Y.; KOTANI, A.; KOWALSKI, A.; KRUIJT, B.; KURBATOVA, J.; KUTSCH, W. L.; KWON, H.; LAUNIAINEN, S.; LAURILA, T.; LAW, B.; LEUNING, R.; LI, Y.; LIDDELL, M.; LIMOUSIN, J.-M.; LION, M.; LISKA, A. J.; LOHILA, A.; LÓPEZ-BALLESTEROS, A.; LÓPEZ-BLANCO, E.; LOUBET, B.; LOUSTAU, D.; LUCAS-MOFFAT, A.; LÜERS, J.; MA, S.; MACFARLANE, C.; MAGLIULO, V.; MAIER, R.; MAMMARELLA, I.; MANCA, G.; MARCOLLA, B.; MARGOLIS, H. A.; MARRAS, S.; MASSMAN, W.; MASTEPANOV, M.; MATAMALA, R.; MATTHES, J. H.; MAZZENGA, F.; McCaughey, H.; McHugh, I.; McMillan, A. M. S.; MERBOLD, L.; MEYER, W.; MEYERS, T.; MILLER, S. D.; MINERBI, S.; MODEROW, U.; MONSON, R. K.; MONTAGNANI, L.; MOORE, C. E.; MOORS, E.; MOREAUX, V.; MOUREAUX, C.; MUNGER, J. W.; NAKAI, T.; NEIRYNCK, J.; NESIC, Z.; NICOLINI, G.; NOORMETS, A.; NORTHWOOD, M.; NOSETTO, M.; NOUVELLON, Y.; NOVICK, K.; OECHEL, W.; OLESEN, J. E.; OURCIVAL, J.-M.; PAPUGA, S. A.; PARMENTIER, F.-J.; PAUL-LIMOGE, E.; PAVELKA, M.; PEICHL, M.; PENDALL, E.; PHILLIPS, R. P.; PILEGAARD, K.; PIRK, N.; POSSE, G.; POWELL, T.; PRASSE, H.; PROBER, S. M.; RAMBAL, S.; RANNIK, ; RAZ-YASEEF, N.; REBMANN, C.; REED, D.; DIOS, V. R. d.; RESTREPO-COUBE, N.; REVERTER, B. R.; ROLAND, M.; SABBATINI, S.; SACHS, T.; SALESKA, S. R.; SÁNCHEZ-CAÑETE, E. P.; SANCHEZ-MEJIA, Z. M.; SCHMID, H. P.; SCHMIDT, M.; SCHNEIDER, K.; SCHRADER, F.; SCHRODER, I.; SCOTT, R. L.; SEDLÁK, P.; SERRANO-ORTÍZ, P.; SHAO, C.; SHI, P.; SHIRONYA, I.; SIEBICKE, L.; ŠIGUT, L.; SILBERSTEIN, R.; SIRCA, C.; SPANO, D.; STEINBRECHER, R.; STEVENS, R. M.; STURTEVANT, C.; SUYKER, A.; TAGESSON, T.; TAKANASHI, S.; TANG, Y.; TAPPER, N.; THOM, J.; TOMASSUCCI, M.; TUOVINEN, J.-P.; URBANSKI, S.; VALENTINI, R.; MOLEN, M. van der; GORSEL, E. van; HUISSTEDEN, K. van; VARLAGIN, A.; VERFAILLIE, J.; VESALA, T.; VINCKE, C.; VITALE, D.; VYGODSKAYA, N.; WALKER, J. P.; WALTER-SHEA, E.; WANG, H.; WEBER, R.; WESTERMANN, S.; WILLE, C.; WOFYSY, S.; WOHLFAHRT, G.; WOLF, S.; WOODGATE, W.; LI, Y.; ZAMPEDRI, R.; ZHANG, J.; ZHOU, G.; ZONA, D.; AGARWAL, D.; BIRAUD, S.; TORN, M.; PAPALE, D. The FLUXNET2015 dataset and the ONEFlux processing pipeline for eddy covariance data. **Scientific Data**, v. 7, p. 225. ISSN 2052-4463. Disponível em: <<<https://doi.org/10.1038/s41597-020-0534-3>>>. 30, 32, 75

PEDREGOSA, F.; VAROQUAUX, G.; GRAMFORT, A.; MICHEL, V.; THIRION, B.; GRISEL, O.; BLONDEL, M.; PRETTENHOFER, P.; WEISS, R.; DUBOURG, V. et al. Scikit-learn: machine learning in python. **Journal of Machine Learning Research**, v. 12, n. Oct, p. 2825–2830, 2011. 59

RESTAINO, D. L. P. J. C. Wildfire and fuel treatment effects on forest carbon dynamics in the western united states. **Forest Ecology and Management**, v. 303, 2013. ISSN 0378-1127. Disponível em: <<<https://www.sciencedirect.com/science/article/pii/S0378112713001904>>>. 11

RESTREPO-COUBE, N.; da Rocha, H. R.; HUTYRA, L. R.; da Araujo, A. C.; BORMA, L. S.; CHRISTOFFERSEN, B.; CABRAL, O. M.; de Camargo, P. B.; CARDOSO, F. L.; da Costa, A. C. L.; FITZJARRALD, D. R.; GOULDEN, M. L.; KRUIJT, B.; MAIA, J. M.; MALHI, Y. S.; MANZI, A. O.; MILLER, S. D.; NOBRE, A. D.; von Randow, C.; Sá, L. D. A.; SAKAI, R. K.; TOTA, J.; WOFSY, S. C.; ZANCHI, F. B.; SALESKA, S. R. What drives the seasonality of photosynthesis across the amazon basin? a cross-site analysis of eddy flux tower measurements from the brasil flux network. **Agricultural and Forest Meteorology**, v. 182-183, p. 128–144, 2013. ISSN 0168-1923. Disponível em: <<<https://www.sciencedirect.com/science/article/pii/S0168192313001184>>>. 3, 8, 11, 34, 81, 88, 94

RESTREPO-COUBE, N.; LEVINE. Do dynamic global vegetation models capture the seasonality of carbon fluxes in the Amazon basin? a data-model intercomparison. **Global Change Biology**, v. 23, n. 1, 2017. Disponível em: <<<https://onlinelibrary.wiley.com/doi/abs/10.1111/gcb.13442>>>. 14, 35

RESTREPO-COUBE, N.; ROCHA da. **LBA-ECO CD-32 flux tower network data compilation, brazilian Amazon: 1999-2006, V2**. ORNL Distributed Active Archive Center, 2021. Disponível em: <<https://daac.ornl.gov/cgi-bin/dsviewer.pl?ds_id=1842>>. 31

_____. _____. ORNL Distributed Active Archive Center, 2021. Disponível em: <<https://daac.ornl.gov/cgi-bin/dsviewer.pl?ds_id=1842>>. 33, 34

ROCHA, H. R. D. Seasonality of water and heat fluxes over a tropical forest in eastern Amazonia. **Ecological Applications**, v. 14, n. sp4, p. 22–32, 2004. 35

RODRIGUES, S.; CIRINO. Enhanced net co2 exchange of a semideciduous forest in the southern Amazon due to diffuse radiation from biomass burning. **Biogeosciences**, v. 21, n. 3, 2024. Disponível em: <<<https://bg.copernicus.org/articles/21/843/2024/>>>. 10, 91

ROZANTE, J. R. **Produto de precipitação MERGE**. 2000. Disponível em: <<http://ftp1.cptec.inpe.br/modelos/io/produtos/MERGE/read_me_MERGE.pdf>>. 42, 45

ROZANTE, J. R.; MOREIRA. **Combining TRMM and surface observations precipitation: technique and validation over South America**. [s.n.], 2010. Disponível em: <<http://ftp1.cptec.inpe.br/modelos/io/produtos/MERGE/rozante_et.al.2010.pdf>>. 45

SAKAI, R.; FITZJARRALD, D.; MORAES, O.; STAEBLER, R.; ACEVEDO, O.; CZIKOWSKY, M.; SILVA, R.; BRAIT, E.; MIRANDA, V. Land-use change effects on local energy, water, and carbon balances in an Amazonian agricultural field. **Global Change Biology**, v. 10, p. 895 – 907, 05 2004. 35, 76, 82

SALESKA, S. R.; MILLER, S. D.; MATROSS, D. B. Carbon in Amazon forests: unexpected seasonal fluxes and disturbance-induced losses. **Science**, v. 302, n. 5650, p. 1554–1557, 2003. 35

SANCHEZ, E. C. M. S. **Controle por aprendizado acelerado e neuro-fuzzy de sistemas servo-hidráulicos de alta frequência**. Dissertação (Mestrado em Engenharia Mecânica) — Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, RJ, 2009. 19, 20

SANCHEZ-GONZALEZ, A.; GODWIN, J.; PFAFF, T.; YING, R.; LESKOVEC, J.; BATTAGLIA, P. W. **Learning to simulate complex physics with graph networks**. 2020. Disponível em: <<<https://arxiv.org/abs/2002.09405>>>. 21

SENEVIRATNE, S. I.; ZHANG, X.; ADNAN, M.; BADI, W.; DERECZYNSKI, C.; LUCA, A. D.; GHOSH, S.; ISKANDAR, I.; KOSSIN, J.; LEWIS, S.; OTTO, F.; PINTO, I.; SATOH, M.; VICENTE-SERRANO, S. M.; WEHNER, M.; ZHOU, B. Weather and climate extreme events in a changing climate. In: **Climate change 2021: the physical science basis. Contribution of working group I to the sixth assessment report of the Intergovernmental Panel on Climate Change**. Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press, 2021. p. 1513–1766. 2

SIMS, D. A.; RAHMAN, A. F.; CORDOVA, V. D.; EL-MASRI, B. Z.; BALDOCCHI, D. D.; BOLSTAD, P. V.; FLANAGAN, L. B.; GOLDSTEIN, A. H.; HOLLINGER, D. Y.; MISSON, L.; MONSON, R. K.; OECHEL, W. C.; SCHMID, H. P.; WOFSY, S. C.; XU, L. A new model of gross primary productivity for north american ecosystems based solely on the enhanced vegetation index and land surface temperature from modis. **Remote Sensing of Environment**, v. 112, n. 4, p. 1633–1646, 2008. ISSN 0034-4257. Disponível em: <<<https://www.sciencedirect.com/science/article/pii/S0034425707004105>>>. 16

SISTEMA DE ESTIMATIVAS DE EMISSÕES E REMOÇÕES DE GASES DE EFEITO ESTUFA. **Emissões totais**. 2022. Disponível em: <<https://plataforma.seeg.eco.br/total_emission>>. Acesso em: 24 mar. 2024. 1

SOUZA, C. M.; SHIMBO, J. Z.; ROSA, M. R.; PARENTE, L. L.; ALENCAR, A. A.; RUDORFF, B. F. T.; HASENACK, H.; MATSUMOTO, M.; FERREIRA, L. G.; SOUZA-FILHO, P. W. M.; OLIVEIRA, S. W. de; ROCHA, W. F.; FONSECA, A. V.; MARQUES, C. B.; DINIZ, C. G.; COSTA, D.; MONTEIRO, D.; ROSA, E. R.; VÉLEZ-MARTIN, E.; WEBER, E. J.; LENTI, F. E. B.; PATERNOST, F. F.; PAREYN, F. G. C.; SIQUEIRA, J. V.; VIERA, J. L.; NETO, L. C. F.; SARAIVA, M. M.; SALES, M. H.; SALGADO, M. P. G.; VASCONCELOS, R.; GALANO, S.; MESQUITA, V. V.; AZEVEDO, T. Reconstructing three decades of land use and land cover changes in brazilian biomes with landsat archive and earth

engine. **Remote Sensing**, v. 12, n. 17, 2020. ISSN 2072-4292. Disponível em: <<<https://www.mdpi.com/2072-4292/12/17/2735>>>. 29, 30

STOY, P. C.; KATUL, G. G.; SIQUEIRA, M. B.; JUANG, J.-Y.; NOVICK, K. A.; UEBELHERR, J. M.; OREN, R. An evaluation of models for partitioning eddy covariance-measured net ecosystem exchange into photosynthesis and respiration. **Agricultural and Forest Meteorology**, v. 141, n. 1, p. 2–18, 2006. ISSN 0168-1923. Disponível em: <<<https://www.sciencedirect.com/science/article/pii/S0168192306002413>>>. 14

TEAM, C. S. **NASA Atmospheric science data center ASDC**. 2021. Accessed <16-11-2023> at. 37, 40

TORO, I. M. C. T. **Dinâmica de fluxos de energia e CO₂ em sistemas agrícolas com rotação de culturas no sul do Brasil utilizando o método eddy covariance**. Dissertação (Mestrado em Física) — Universidade Federal de Santa Maria, Santa Maria, RS, 2018. 11, 12

TRAMONTANA, G.; JUNG, M.; SCHWALM, C. R.; ICHII, K.; CAMPS-VALLS, G.; RÁDULY, B.; REICHSTEIN, M.; ARAIN, M. A.; CESCATTI, A.; KIELY, G.; MERBOLD, L.; SERRANO-ORTIZ, P.; SICKERT, S.; WOLF, S.; PAPAIE, D. Predicting carbon dioxide and energy fluxes across global fluxnet sites with regression algorithms. **Biogeosciences**, v. 13, n. 14, p. 4291–4313, 2016. Disponível em: <<<https://bg.copernicus.org/articles/13/4291/2016/>>>. 3, 4, 14, 15, 16, 17, 18, 27, 38, 40, 45, 61, 96, 101, 108

UNITED NATIONS FRAMEWORK CONVENTION ON CLIMATE CHANGE (UNFCCC). **Federative Republic of Brazil nationally determined contribution NDC to the Paris agreement under the UNFCCC**. 2023. Disponível em: <<<https://unfccc.int/sites>>>. Acesso em: 25 mar. 2024. 1

VELIČKOVIĆ, P.; CUCURULL, G.; CASANOVA, A.; ROMERO, A.; LIÒ, P.; BENGIO, Y. **Graph attention networks**. 2018. Disponível em: <<<https://arxiv.org/abs/1710.10903>>>. 22

VIEIRA, L. S.; QUEIROZ, G. R.; SHIGUEMORI, E. H. An analysis of the influence of the number of observations in a random forest time series classification to map the forest and deforestation in the Brazilian Amazon. **The international archives of the photogrammetry, remote sensing and spatial information sciences**, v.43, n. B3, 2022. Disponível em: <<<https://isprs-archives.copernicus.org/articles/XLIII-B3-2022/721/2022/>>>. 17

WAN, Z.; HOOK, S.; HULLEY, G. **MODIS/Terra land surface temperature/emissivity 8-Day L3 global 1 km SIN Grid V061 - MOD11A2**. Washington: NASA: NASA EOSDIS land processes distributed active archive center, 2021. Disponível em: <<<https://doi.org/10.5067/MODIS/MOD11A2.061>>>. Acesso em: 25 dez 2023. 36, 39

WANG, L.; ZHOU, X.; ZHU, X.; DONG, Z.; GUO, W. Estimation of biomass in wheat using random forest regression algorithm and remote sensing data. **The Crop Journal**, v. 4, n. 3, p. 212–219, 2016. ISSN 2214-5141. Disponível em: <<<https://www.sciencedirect.com/science/article/pii/S2214514116300162>>>. 17

WARING, R. H.; RUNNING, S. W. **Forest ecosystems**. 3. ed. San Diego: Academic Press, 2007. ISBN 978-0-12-370605-8. Disponível em: <<<https://www.sciencedirect.com/science/article/pii/B9780123706058500050>>>. 2, 3, 8, 9, 10, 11, 13, 14, 15, 24

WIELICKI, B. A.; BARKSTROM, B. R.; HARRISON, E. F.; III, R. B. L.; SMITH, G. L.; COOPER, J. E. Clouds and the earth's radiant energy system CERES: an earth observing system experiment. **Bulletin of the American Meteorological Society**, v. 77, p. 853–868, 1996. 40

WOOD, D. A. Net ecosystem carbon exchange prediction and insightful data mining with an optimized data-matching algorithm. **Ecological Indicators**, v. 124, p. 107426, 2021. ISSN 1470-160X. Disponível em: <<<https://www.sciencedirect.com/science/article/pii/S1470160X21000911>>>. 21

WU, C.; CHEN, J.; DESAI, A.; HOLLINGER, D.; ARAIN, M.; MARGOLIS, H.; GOUGH, C.; STAEBLER, R. Remote sensing of canopy light use efficiency in temperate and boreal forests of north america using modis imagery. **Remote Sensing of Environment**, v. 118, p. 60–72, 03 2012. 16

WU, Z.; PAN, S.; CHEN, F.; LONG, G.; ZHANG, C.; YU, P. S. A comprehensive survey on graph neural networks. **IEEE Transactions on Neural Networks and Learning Systems**, v. 32, n. 1, p. 4–24, 2021. ISSN 2162-2388. Disponível em: <<<http://dx.doi.org/10.1109/TNNLS.2020.2978386>>>. 21

XIAO, X.; ZHANG, Q.; BRASWELL, B.; URBANSKI, S.; BOLES, S.; WOFSY, S.; MOORE, B.; OJIMA, D. Modeling gross primary production of temperate deciduous broadleaf forest using satellite images and climate data. **Remote Sensing of Environment**, v. 91, n. 2, p. 256–270, 2004. ISSN 0034-4257. Disponível em: <<<https://www.sciencedirect.com/science/article/pii/S0034425704000975>>>. 16

ZHOU, H.; YUE, X.; WANG, B.; TIAN, C.; LU, X.; ZHU, J.; CAO, Y. Distinguishing the main climatic drivers to the variability of gross primary productivity at global fluxnet sites. **Environmental Research Letters**, v. 18, n. 12, p. 124007, nov 2023. Disponível em: <<<https://dx.doi.org/10.1088/1748-9326/ad059c>>>. 17

ZHURAVLEV, R.; DARA, A.; SANTOS, A. L. D. d.; DEMIDOV, O.; BURBA, G. Globally scalable approach to estimate net ecosystem exchange based on remote sensing, meteorological data, and direct measurements of eddy covariance sites. **Remote Sensing**, v. 14, n. 21, 2022. ISSN 2072-4292. Disponível em: <<<https://www.mdpi.com/2072-4292/14/21/5529>>>. 3, 14, 15, 16, 18, 61, 96

APÊNDICE A

A.1 Figuras extras da ampliação de escala de NEE para o bioma Amazônia

Figura A.1 - Comportamento Anual médio de NEE com o modelo RF e preditores RS.

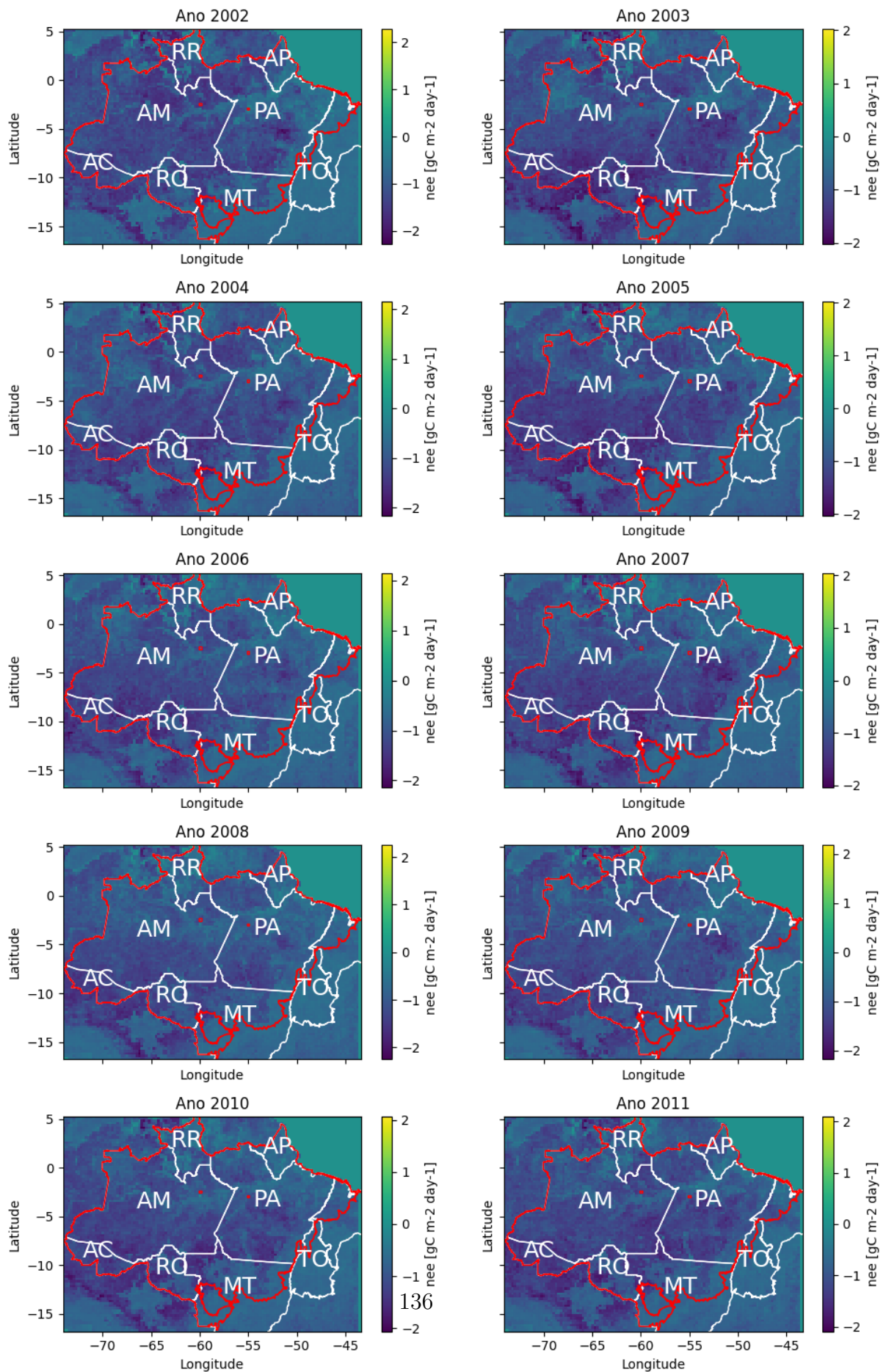


Figura A.2 - Comportamento Anual da ampliação de escala realizada pelo Projeto Fluxcom com preditores RS.

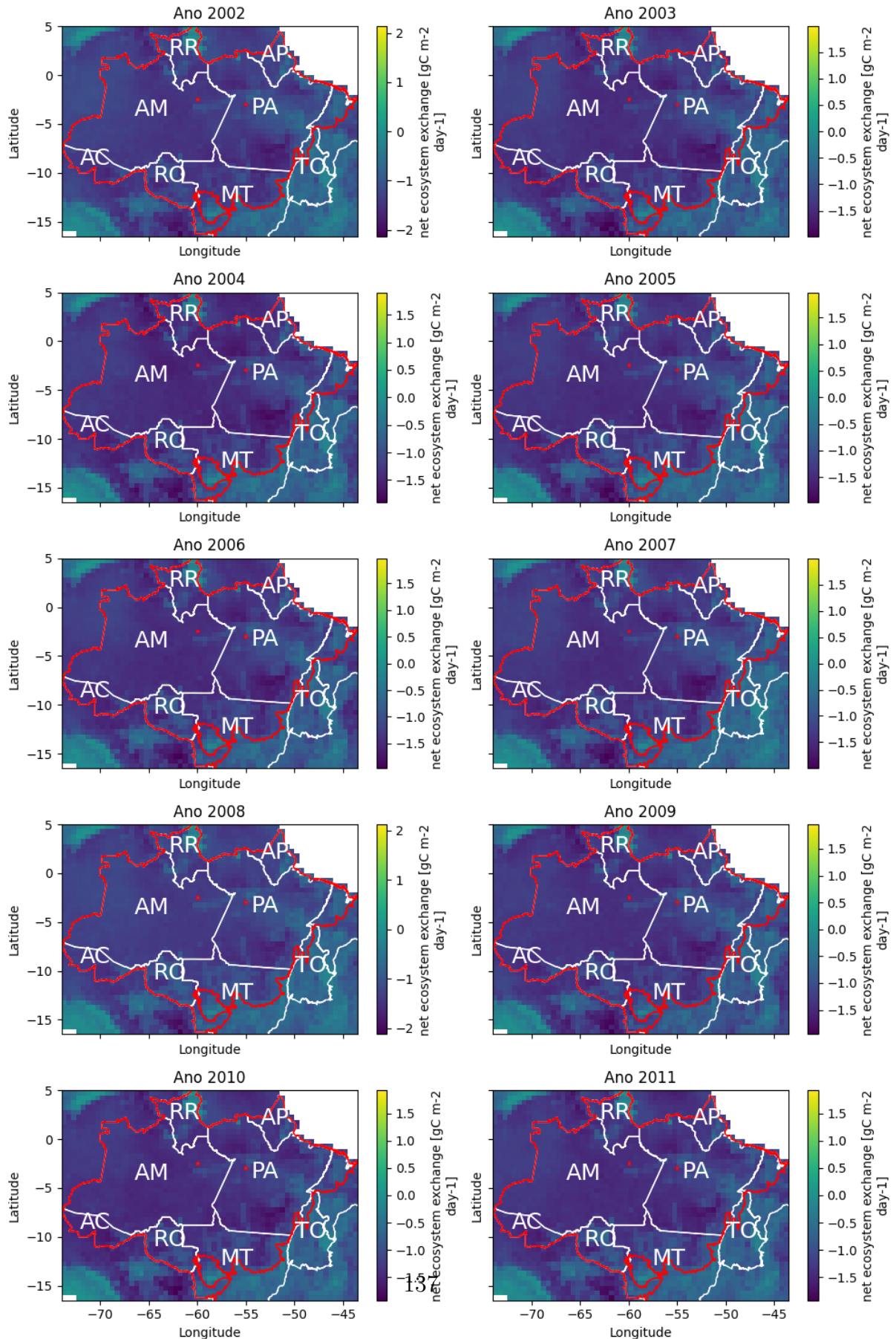
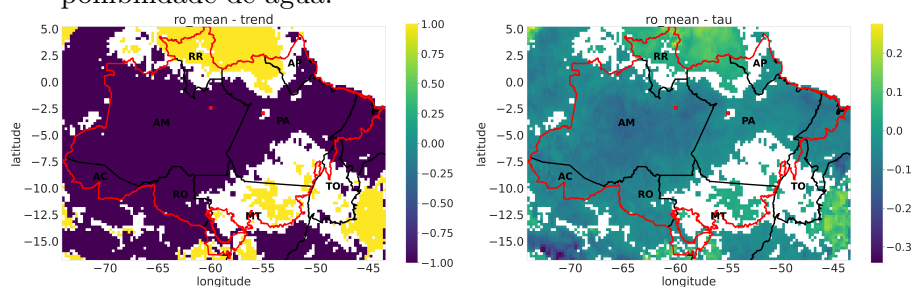
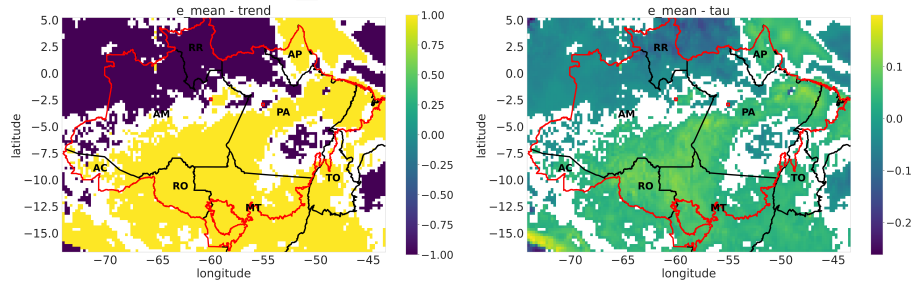


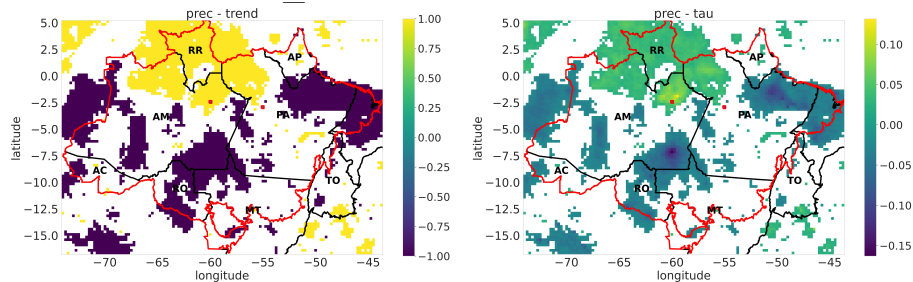
Figura A.3 - Resultados da Tendência de Mann Kendall para atributos referentes à disponibilidade de água.



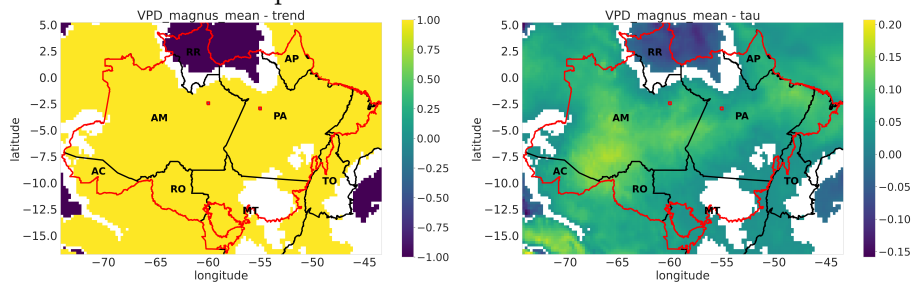
Tendência de Mann Kendall: ro_mean.



Tendência de Mann Kendall: e_mean.



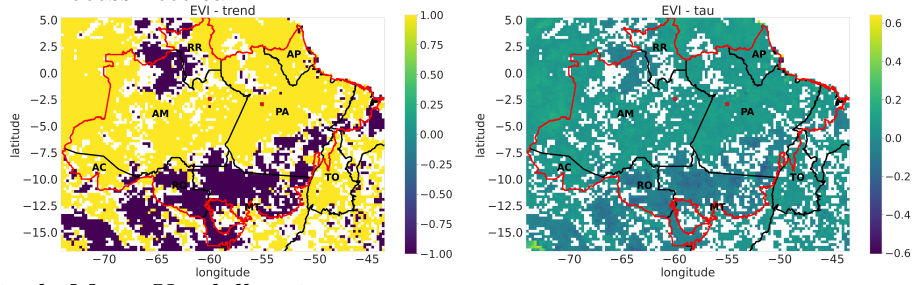
Tendência de Mann Kendall: prec.



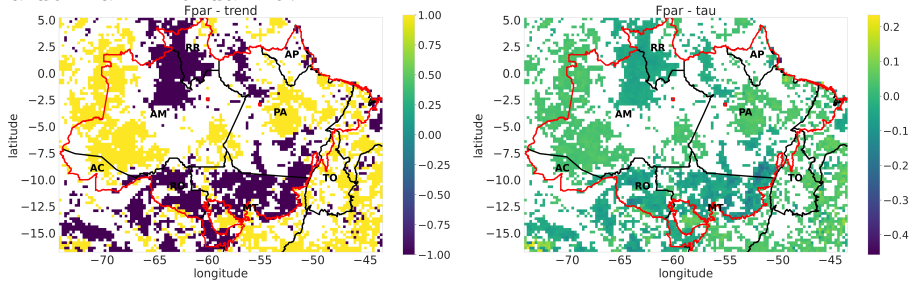
Tendência de Mann Kendall: vpd_mean.

Resultados do Teste de Mann Kendall, com $p_valor < 0,05$, das variáveis de disponibilidade de água para a grade do Bioma Amazônia utilizada neste trabalho. Em vermelho há a delimitação do Bioma Amazônia e as células K34 e K67. Em preto há a delimitação de estados e suas siglas da Região Norte do Brasil.

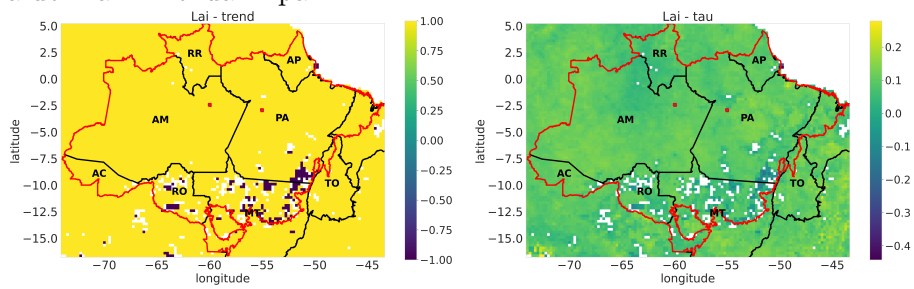
Figura A.4 - Resultados dos Testes de Mann Kendall para atributos referentes à atividade fotossintética.



Tendência de Mann Kendall: evi.



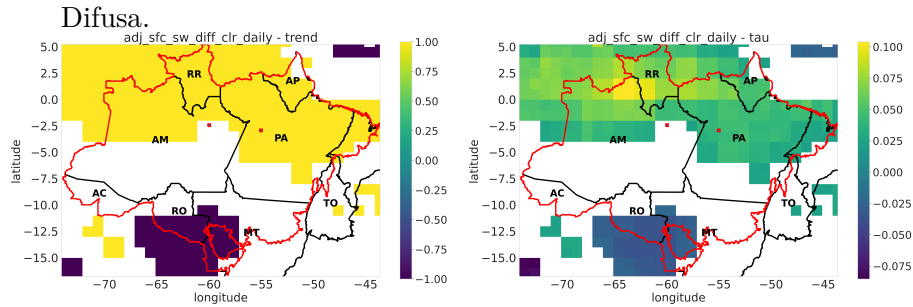
Tendência de Mann Kendall: fpar.



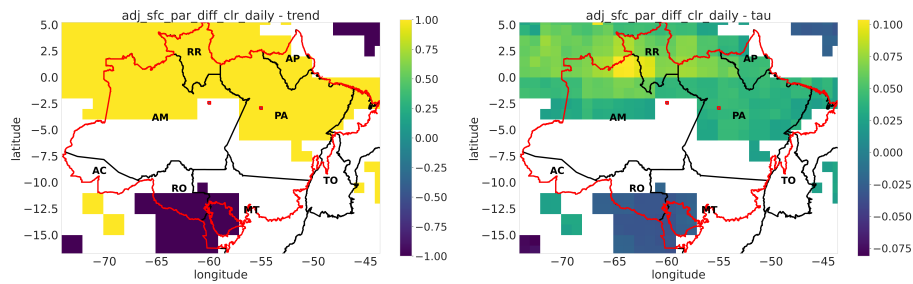
Tendência de Mann Kendall: lai.

Resultado do Teste de Mann Kendall, com $p_valor < 0,05$, das variáveis do comportamento da vegetação para a grade do Bioma Amazônia utilizada neste trabalho. Em vermelho há a delimitação do Bioma Amazônia e as células K34 e K67. Em preto há a delimitação de estados e suas siglas da Região Norte do Brasil.

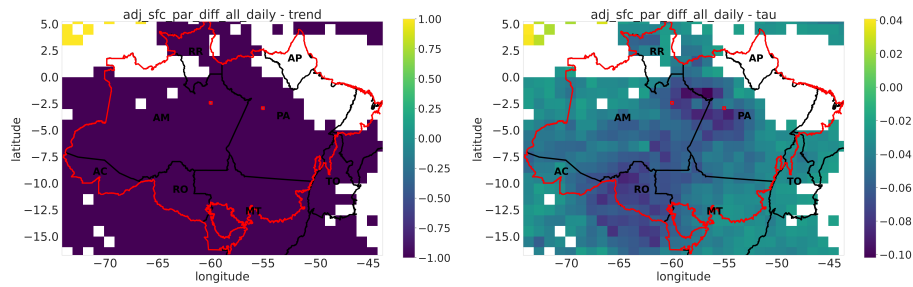
Figura A.5 - Resultados dos Testes de Mann-Kendall para atributos referentes à Radiação Difusa.



Tendência de Mann Kendall: diffuse_clr.



Tendência de Mann Kendall: par.



Tendência de Mann Kendall: par_all.

Resultado do Teste de Mann Kendall, com $p_valor < 0,05$, da radiação difusa para a grade do Bioma Amazônia utilizada neste trabalho. Em vermelho há a delimitação do Bioma Amazônia e as células K34 e K67. Em preto há a delimitação de estados e suas siglas da Região Norte do Brasil.