

UNIVERSIDADE DE SÃO PAULO
ESCOLA DE ENGENHARIA DE SÃO CARLOS

Alex Naoki Asato Kobayashi

**Rotinas computacionais de processamento de dados de
Eddy Covariance para determinação da evapotranspiração
em fragmento de Cerrado *sensu stricto***

São Carlos

2021

Alex Naoki Asato Kobayashi

**Rotinas computacionais de processamento de dados de
Eddy Covariance para determinação da evapotranspiração
em fragmento de Cerrado *sensu stricto***

Dissertação apresentada à Escola de Engenharia de São Carlos da Universidade de São Paulo, para obtenção do título de Mestre em Ciências - Programa de Pós-Graduação em Hidráulica e Saneamento.

Área de concentração: Hidráulica e Saneamento

Orientador: Prof. Dr. Edson Cezar Wendland

VERSÃO CORRIGIDA

**São Carlos
2021**

AUTORIZO A REPRODUÇÃO TOTAL OU PARCIAL DESTE TRABALHO,
POR QUALQUER MEIO CONVENCIONAL OU ELETRÔNICO, PARA FINS
DE ESTUDO E PESQUISA, DESDE QUE CITADA A FONTE.

Ficha catalográfica elaborada pela Biblioteca Prof. Dr. Sérgio Rodrigues Fontes da
EESC/USP com os dados inseridos pelo(a) autor(a).

75r Kobayashi, Alex Naoki Asato
Rotinas computacionais de processamento de dados de
Eddy Covariance para determinação da evapotranspiração
em fragmento de Cerrado sensu stricto / Alex Naoki
Asato Kobayashi; orientador Edson Cezar Wenland. São
Carlos, 2021.

Dissertação (Mestrado) - Programa de
Pós-Graduação em Engenharia Hidráulica e Saneamento e
Área de Concentração em Hidráulica e Saneamento --
Escola de Engenharia de São Carlos da Universidade de
São Paulo, 2021.

1. Covariância de turbilhões. 2. Área de
contribuição. 3. Preenchimento de falhas. 4. Python. I.
Título.

FOLHA DE JULGAMENTO

Candidato: Bacharel **ALEX NAOKI ASATO KOBAYASHI**.

Título da dissertação: " Análise do método *Eddy Covariance* para área e cerrado *sensu stricto*".

Data da defesa: 19/05/2021.

Comissão Julgadora

Resultado

Prof. Titular **Edson Cezar Wendland**
(Orientador)
(Escola de Engenharia de São Carlos/EESC-USP)

__Aprovado__

Profa. Dra. **Débora Regina Roberti**
(Universidade Federal de Santa Maria/UFSM)

__Aprovado__

Dr. **Oswaldo Machado Rodrigues Cabral**
(Empresa Brasileira de Pesquisa Agropecuária/EMBRAPA - Jaguariúna)

__Aprovado__

Coordenador do Programa de Pós-Graduação em Engenharia Hidráulica e Saneamento:

Prof. Dr. **Eduardo Mario Mendiando**

Presidente da Comissão de Pós-Graduação:

Prof. Titular **Murilo Araujo Romero**

*Este trabalho é dedicado a todos os profissionais da linha de frente
e pesquisadores que disseminam o conhecimento científico.*

AGRADECIMENTOS

Agradeço aos meus pais, Carlinda Asato e Mário Massao Kobayashi e meu irmão Rodrigo Kenji Asato Kobayashi, por todo suporte familiar e apoio incondicional.

Agradeço aos meus amigos que me acompanharam na jornada de Campo Grande até São Carlos, Jullian e Gabriela, por estarem sempre ao meu lado.

Agradeço aos meus amigos que ficaram em São Carlos no ano de uma pandemia, Dimaghi e Yuri, pela parceria que aumentou no isolamento de um ano solitário.

Agradeço aos meus amigos que ficaram em Campo Grande que sempre me apoiaram e motivaram para continuar adiante, apesar da distância.

Agradeço ao Jamil pela mentoria e parceria desde o início do processo do mestrado.

Agradeço todos os amigos do Laboratório de Hidráulica Computacional, André, Jean, Hélio, Lívia, Alan, Alessandra, Kalyl, Lazarini, Erick, Jaqueline, Alice, Bertotto e Rubens, pelos grande momentos (Ping-Pong) e apoio na curta estadia em São Carlos.

Agradeço aos amigos de outros laboratórios que tive a oportunidade de trocar idéias e perspectivas ao longo do mestrado.

Agradeço ao Bérnago pela ajuda no transporte e por ter compartilhado um pouco da sua sabedoria e apoio.

Agradeço ao professor Dr. Edson Wendland pela oportunidade e confiança depositado ao longo do mestrado.

Agradeço ao professor Dr. Paulo Tarso por ter motivado desde a graduação no processo do mestrado e busca por ideais científicos.

Às secretárias do SHS, Sá, Priscila e Rose, pela paciência e auxílio no mestrado, sempre com ótima disposição seja qualquer o problema.

Ao setor de Transportes da EESC para o auxílio nos trabalhos de campo.

A Fundação de Ampara à Pesquisa do Estado de São Paulo (FAPESP) e ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pelo suporte financeiro ao projeto de pesquisa.

Ao programa de Pós-Graduação do Departamento de Hidráulica e Saneamento da Escola de Engenharia de São Carlos e à USP, pela excelência acadêmica e crescimento profissional.

*“Sometimes the best way to solve your
own problems is to help someone else.”*

Uncle Iroh

RESUMO

KOBAYASHI, A. N. A. **Rotinas computacionais de processamento de dados de *Eddy Covariance* para determinação da evapotranspiração em fragmento de Cerrado *sensu stricto***. 2021. 123p. Dissertação (Mestrado) - Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos, 2021.

A evapotranspiração (ET) é a componente hidrológica que representa a maior porcentagem do total da precipitação no bioma do Cerrado brasileiro. A quantificação hídrica do complexo sistema solo-planta-atmosfera conduz a necessidade de utilização de técnicas diretas na medição de ET como o método *Eddy Covariance*. Desta forma, o objetivo deste trabalho foi de analisar o emprego do método *Eddy Covariance* no monitoramento da ET de uma área de remanescente de Cerrado *sensu stricto* através de uma torre de monitoramento instalada no ano de 2018. Com esta análise, buscou-se otimizar o processamento dos dados de alta frequência, investigar a área de contribuição dos fluxos, além de realizar o preenchimento de falhas. O processamento de dados de alta frequência (20 Hz) utilizou o programa *EddyPro* para a obtenção do melhor conjunto de técnicas. Para a determinação da área de contribuição, utilizou-se um método que mistura simulações lagrangianas e soluções analíticas e o mapa de uso de solos do MapBiomas. Ainda, foram avaliados seis métodos de preenchimento de falhas para a consolidação de uma série de ET. Para isso, foram adotados procedimentos automáticos utilizando a capacidade interativa do programa desenvolvido em linguagem Python para testes de hipóteses. A aprimoração da série de ET, para os anos de 2019 e 2020, utilizou os filtros de qualidade, precipitação, força de sinal e área de contribuição, que filtraram cerca de 45% dos dados da série. O método de preenchimento de falhas por *Random Forest* obteve uma média de ET para o período chuvoso de 3,71 *mm/dia* e para o período seco de 2,54 *mm/dia* e um acumulado de 1011 *mm* para 2020, com ótimas métricas na avaliação para o período diurno ($\overline{\text{MAE}} = 0,0577$, $\overline{\text{RMSE}} = 0,0826$, $\overline{\text{MBE}} = 0,0007$ e $\rho = 0,853$). Contudo, os resultados mostraram que diversos métodos de preenchimento de falhas podem ser aplicados satisfatoriamente à série de ET, e que a filtragem dos dados melhorou significativamente a qualidade da série de ET.

Palavras-chave: Covariância de turbilhões. Área de contribuição. Preenchimento de falhas. Python.

ABSTRACT

KOBAYASHI, A. N. A. **Computational routines of data processing by Eddy Covariance to determine evapotranspiration in a Cerrado *sensu stricto* fragment.** 2021. 123p. Dissertação (Mestrado) - Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos, 2021.

Evapotranspiration (ET) is the hydrological component that represents the largest percentage of total precipitation in the Brazilian Cerrado biome. The water quantification of the complex soil-plant-atmosphere system leads to the need to use direct techniques in the measurement of ET such as the Eddy Covariance method. Thus, the objective of this work was to analyze the use of the *Eddy Covariance* method in the monitoring of the ET of a remnant area of Cerrado *sensu stricto* through a monitoring tower installed in 2018. With this analysis, we sought to optimize the processing of high-frequency data, investigate the contribution of the footprint, in addition to filling in gaps. The processing of high frequency data (20 Hz) used the program *EddyPro* to obtain the best set of techniques. To determine the contribution area, a method was used that mixes Lagrangian simulations and analytical solutions and the MapBiomas land use map. In addition, six gap filling methods were evaluated for the consolidation of a series of ET. For this, automatic procedures were adopted using the interactive capacity of the program developed in Python for hypothesis testing. The improvement of the ET series, for the years 2019 and 2020, used the filters of quality, precipitation, signal strength and contribution area, which filtered about 45% of the series data. The gap filling method by *Random Forest* obtained an average ET for the rainy period of 3.71 *mm/day* and for the dry period of 2.54 *mm/day* and a cumulative of 1011 *mm* for 2020, with great metrics in the daytime evaluation ($\overline{\text{MAE}} = 0.0577$, $\overline{\text{RMSE}} = 0.0826$, $\overline{\text{MBE}} = 0.0007$ and $\rho = 0.853$). Nevertheless, the results showed that several gap filling methods are able to satisfy the ET series, and the segmentation of the data significantly improved the quality ET series.

Keywords: Eddy Covariance. Footprint. Gapfilling. Python.

LISTA DE FIGURAS

Figura 1 – Esquema Random Forest	39
Figura 2 – Exemplo de neurônio básico	40
Figura 3 – Exemplo de rede neurais	41
Figura 4 – Exemplo de nó recorrente	42
Figura 5 – Área de estudo	45
Figura 6 – Precipitação nas estações do IAB1, IAB2 e IAB3	46
Figura 7 – Período de coleta de dados das estações	47
Figura 8 – Torre de monitoramento IAB3	47
Figura 9 – Vista da torre de monitoramento (IAB3) para área de Cerrado <i>sensu stricto</i> (a) e para área de pasto (b) e vista superficial do Cerrado <i>sensu stricto</i> (c) e da área de pasto (d)	48
Figura 10 – Etapas do trabalho	49
Figura 11 – Painel de visualização dos dados processados do balanço energético e ET, com dados de entradas e ajuste dinâmico de parâmetros dos filtros	57
Figura 12 – Boxplot do desempenho do balanço energético (inclinação) sobre a correção espectral de alta frequência	58
Figura 13 – Boxplot do desempenho do balanço energético (inclinação) sobre a correção espectral de baixa frequência	58
Figura 14 – Boxplot do desempenho do balanço energético (inclinação) sobre a rotação de coordenadas	58
Figura 15 – Desempenho do fechamento do balanço energético (inclinação) para todas as configurações testadas	59
Figura 16 – Série temporal do balanço energético de melhor desempenho (p08) sem aplicação dos filtros, sendo que (a) representa a série do fluxo de calor sensível e latente, (b) o saldo radiométrico e (c) a nuvem de pontos com as métricas de correlação e fechamento de balanço energético	60
Figura 17 – Série temporal do balanço energético de melhor desempenho (p08) com aplicação do filtro de precipitação e de qualidade, sendo que (a) representa a série do fluxo de calor sensível e latente, (b) o saldo radiométrico e (c) a nuvem de pontos com as métricas de correlação e fechamento de balanço energético	60
Figura 18 – Série temporal binária com o filtro de precipitação para a estação IAB3	61
Figura 19 – Série temporal binária com o filtro de qualidade para estação IAB3	61
Figura 20 – Interface gráfica para análise da área de contribuição, utilizando o método por Kljun et al. (2015), por intervalo de tempo de 30 minutos	62

Figura 21 – Criação do mapa de aceitação para a torre de monitoramento IAB3: (a) imagem de satélite; (b) recorte de mapa de uso do solo e; (c) mapa de aceitação	63
Figura 22 – Interface gráfica para análise da área de contribuição, utilizando o método por Kljun et al. (2015), por faixa de tempo determinada pelo usuário	63
Figura 23 – Progressão do grau de aceitação da área de contribuição, utilizando o método por Kljun et al. (2015), acumulada da torre de monitoramento IAB3 no período de 2019 até 2020	64
Figura 24 – Desempenho do fechamento do balanço energético (inclinação) e correlação de Pearson em relação ao grau de aceitação da área de contribuição para a torre de monitoramento IAB3	65
Figura 25 – Série temporal binária com o filtro de área de contribuição para a estação IAB3	65
Figura 26 – Série de Evapotranspiração com aplicação de filtros para a estação IAB3	66
Figura 27 – Esquema da progressão da série de ET na filtragem dos dados e preenchimento de falhas	67
Figura 28 – Evapotranspiração com o método de preenchimento de falhas de Variação Diurna Média	68
Figura 29 – Evapotranspiração com o método de preenchimento de falhas de Regressão Linear Múltipla	69
Figura 30 – Evapotranspiração com o método de preenchimento de falhas de <i>Random Forest</i>	69
Figura 31 – Evapotranspiração com o método de preenchimento de falhas de redes neurais densa	70
Figura 32 – Evapotranspiração com o método de preenchimento de falhas de redes neurais recorrente (LSTM)	70
Figura 33 – Condutância superficial para a vegetação Cerrado sensu stricto para o ano de 2019 por: (a) média mensal e hora do dia e (b) média do período seco e chuvoso	71
Figura 34 – Condutância superficial para a vegetação Cerrado sensu stricto para o ano de 2020 por: (a) média mensal e hora do dia e (b) média do período seco e chuvoso	72
Figura 35 – Evapotranspiração com o método de preenchimento de falhas de Penman-Monteith inverso	73
Figura 36 – Correlação da validação do método de variação diurna média	74
Figura 37 – Correlação da validação do método de regressão linear múltipla	75
Figura 38 – Correlação da validação do método de Random Forest	76
Figura 39 – Correlação da validação do método de Rede neural densa	77

Figura 40 – Correlação da validação do método de Rede neural recorrente (LSTM)	78
Figura 41 – Correlação da validação do método de Penman-Monteith	79
Figura 42 – Quantidade de dados diários da umidade relativa da estação IAB1 . . .	93
Figura 43 – Quantidade de dados meteorológicos da estação IAB2	93
Figura 44 – Quantidade de dados meteorológicos da estação IAB3	94
Figura 45 – Quantidade e tipos de falhas restantes na série de ET após o método de variação diurna média	95
Figura 46 – Quantidade e tipos de falhas restantes na série de ET após o método de regressão linear múltipla	96
Figura 47 – Quantidade e tipos de falhas restantes na série de ET após o método de random forest	96
Figura 48 – Quantidade e tipos de falhas restantes na série de ET após o método de rede neural densa	97
Figura 49 – Quantidade e tipos de falhas restantes na série de ET após o método de rede neural recorrente	97
Figura 50 – Quantidade e tipos de falhas restantes na série de ET após o método de Penman-Monteith inverso	98

LISTA DE TABELAS

Tabela 1 – Lista de principais métodos de correção espectral	34
Tabela 2 – Tabela de métodos de sinalização da qualidade dos dados	35
Tabela 3 – Instrumentos utilizados para a coleta e armazenamento de dados (IAB3)	48
Tabela 4 – Métodos para processamento de dados de alta frequência (EddyPro) .	50
Tabela 5 – Arquitetura da rede neural densa para o preenchimento de falhas . . .	53
Tabela 6 – Arquitetura da rede neural recorrente do tipo LSTM para o preenchi- mento de falhas	53
Tabela 7 – Tipos de cobertura e aceitação (MapBiomas)	62
Tabela 8 – Métricas da avaliação do método de variação diurna média	73
Tabela 9 – Métricas da avaliação do método de regressão linear múltipla	74
Tabela 10 – Métricas da avaliação do método de Random Forest	75
Tabela 11 – Métricas da avaliação do método de rede neural densa	76
Tabela 12 – Métricas da avaliação do método de rede neural recorrente	77
Tabela 13 – Métricas da avaliação do método de Penman-Monteith inverso	78
Tabela 14 – Média e desvio padrão das séries preenchidas da ET	79
Tabela 15 – Erro médio absoluto (\overline{MAE}) entre os métodos de preenchimento de falhas	80
Tabela 16 – Raiz quadrada do erro médio (\overline{RMSE}) entre os métodos de preenchi- mento de falhas	80
Tabela 17 – Erro de viés médio (\overline{MBE}) entre os métodos de preenchimento de falhas	81
Tabela 18 – Resultados do coeficiente de correlação de Pearson (ρ) entre os métodos de preenchimento de falhas	81

LISTA DE ABREVIATURAS E SIGLAS

EC	<i>Eddy Covariance</i>
IAB	Instituto Arruda Botelho
SAG	Sistema Aquífero Guarani
USP	Universidade de São Paulo
USPSC	Campus USP de São Carlos
SND	Schotanus, Nieuwstadt e De Bruin
WPL	Webb-Pearman-Leuning

LISTA DE SÍMBOLOS

ET	Evapotranspiração [mm/h]
F_s^{EC}	Covariância da velocidade vertical e massa escalar [$g\ m^3\ s^{-1}$]
R_n	Radiação líquida [$W\ m^{-2}$]
G	Fluxo de calor do solo [$W\ m^{-2}$]
e_s	Pressão de saturação de vapor d'água no ar [kPa]
e_a	Pressão de vapor d'água no ar [kPa]
ρ_a	Densidade média do ar [$kg\ m^{-3}$]
c_p	Calor específico do ar [$MJ\ kg^{-1}\ ^\circ C^{-1}$]
Δ	Relação da pressão de vapor saturada e temperatura [$kPa\ ^\circ C^{-1}$]
γ	Constante psicométrica [$kPa\ ^\circ C^{-1}$]
r_s	Resistência superficial [$s\ m^{-1}$]
r_a	Resistência aerodinâmica [$s\ m^{-1}$]
R	Constante universal do gás
p	Pressão do ambiente [kPa]
ρ_{H_2O}	Densidade da água [$kg\ m^{-3}$]
T_s	Temperatura sônica [$^\circ C^{-1}$]
T_c	Temperatura do ar real [$^\circ C^{-1}$]
TF	Função de transferência
C_{ws}	Co-espectro da velocidade vertical e gás escalar
F_a	Fator atenuação
LE	Fluxo de calor latente [$W\ m^{-2}$]
H	Fluxo de calor sensível [$W\ m^{-2}$]
k	Constante de von Kármán
g_s	Condutância superficial [$m\ s^{-1}$]

g_a	Condutância aerodinâmica [$m s^{-1}$]
λET	Fluxo de calor latente [$W m^{-2}$]
VPD	Déficit de vapor de pressão [kPa]
\overline{MAE}	Erro absoluto médio
\overline{RMSE}	Raiz do erro quadrado médio
\overline{MBE}	Erro de viés médio
DNN	Rede neural densa
LSTM	Rede de memória longa de curto prazo
PM	Penman-Monteith

SUMÁRIO

1	INTRODUÇÃO	27
2	OBJETIVOS	29
2.1	Objetivo Geral	29
2.2	Objetivos Específicos	29
3	REVISÃO BIBLIOGRÁFICA	31
3.1	Evapotranspiração	31
3.1.1	Método Eddy Covariance	32
3.1.1.1	Pré-processamento	32
3.1.1.2	Processamento	33
3.1.1.3	Controle de qualidade	35
3.2	Método de determinação da área de contribuição	36
3.3	Métodos de preenchimento de falhas	37
3.3.1	Variação diurna média	38
3.3.2	Regressão linear múltipla	38
3.3.3	Random Forest	39
3.3.4	Redes neurais	40
3.3.4.1	Redes neurais densas (DNN)	40
3.3.4.2	Redes neurais recorrentes (RNN)	41
3.3.4.3	Redes de memória longa de curto prazo (LSTM)	42
3.3.5	Equação de Penman-Monteith inversa	43
4	MATERIAL E MÉTODOS	45
4.1	Descrição da área de estudo	45
4.1.1	Torre de monitoramento (IAB3)	47
4.2	Etapas do trabalho	49
4.3	Processamento de dados de alta frequência	49
4.3.1	Escolha do software para processamento	49
4.3.2	Determinação dos processos para o processamento dos dados	50
4.4	Determinação da área de contribuição da torre de monitoramento	51
4.5	Preenchimento de falhas da série de evapotranspiração	52
4.5.1	Métodos de preenchimento de falhas	52
4.5.2	Validação das estimativas	53
4.5.2.1	Erro médio absoluto	54
4.5.2.2	Raiz do erro quadrático médio	54

4.5.2.3	Erro de viés médio	54
4.5.2.4	Distribuição das estimativas	54
4.5.2.4.1	Teste de normalidade	55
4.5.2.4.2	Análise de variância	55
5	RESULTADOS E DISCUSSÃO	57
5.1	Processamento dos dados de alta frequência	57
5.1.1	Determinação dos processos para o processamento dos dados	57
5.1.2	Filtro dos dados processados	59
5.2	Análise da área de contribuição	61
5.2.1	Mapa de classificação do uso do solo	61
5.2.2	Filtro dos dados com uso da análise de contribuição	63
5.3	Preenchimento de falhas	66
5.3.1	Métodos de preenchimento de falhas	66
5.3.1.1	Variação Diurna Média	66
5.3.1.2	Regressão Linear Múltipla	67
5.3.1.3	Método de Regressão Logística (<i>Random Forest</i>)	68
5.3.1.4	Redes neurais densa	69
5.3.1.5	Redes neurais recorrentes (LSTM)	70
5.3.1.6	Penman-Monteith inverso	71
5.3.2	Avaliação dos métodos de preenchimento de falhas	73
5.3.2.1	Variação diurna média	73
5.3.2.2	Regressão linear múltipla	74
5.3.2.3	Random Forest	75
5.3.2.4	Rede neural densa	76
5.3.2.5	Rede neural recorrente (LSTM)	77
5.3.2.6	Penman-Monteith inverso	78
5.3.3	Métricas de avaliação	79
6	CONCLUSÃO	83
	REFERÊNCIAS	85
	APÊNDICES	91
	APÊNDICE A – QUANTIDADE DE DADOS METEOROLÓGICOS NAS ESTAÇÕES DO IAB	93

APÊNDICE B – QUANTIDADE DE FALHAS NA SÉRIE DE ET APÓS APLICADO OS MÉTODOS DE PREEN- CHIMENTO DE FALHAS	95
APÊNDICE C – RESULTADOS DOS MÉTODOS DE PREENCHI- MENTO DE FALHAS	99
APÊNDICE D – CÓDIGO EM PYTHON	123

1 INTRODUÇÃO

O Cerrado brasileiro é um dos biomas mais importantes no mundo em termos de biodiversidade (MYERS *et al.*, 2000), onde as fisionomias variam de campo à floresta (Cruz Ruggiero *et al.*, 2002). Por cobrir cerca de 20% em área do Brasil, é crucial na dinâmica dos recursos hídricos, pois compõe parte das principais bacias hidrográficas do Brasil, incluindo a bacia do São Francisco, Tocantins, Paraná e Paraguai. Porém, o aumento da demanda global por água, energia e alimento (Nexo) tem provocado uma mudança no uso e cobertura do solo (GIBBS *et al.*, 2010; MACEDO *et al.*, 2012), tornando o Cerrado um dos biomas mais ameaçados na região dos trópicos (MARRIS, 2005). As alterações de uso do solo, de maneira insustentável, acabam por alterar taxas de perda de solo e os valores médios anuais dos componentes do balanço hídrico (OLIVEIRA *et al.*, 2014; ANACHE *et al.*, 2018; ANACHE *et al.*, 2019).

Em florestas tropicais, a evapotranspiração é o segundo maior componente do ciclo hidrológico, sendo responsável por retornar para a atmosfera grande parte da precipitação anual (OKI; KANAE, 2006). A evapotranspiração depende de características físicas e químicas da vegetação e do solo, além de condições climáticas regionais (MONTEITH, 1981; VERSTRAETEN; VEROUSTRATE; FEYEN, 2008). Atualmente existem diversas formas de mensurar evapotranspiração, sendo os principais métodos: lisímetro, balanço hídrico, razão de Bowen, *Eddy Covariance*, centilômetro, fluxo de seiva e o sensoriamento remoto.

Tratando-se do Cerrado, o método *Eddy Covariance* apresenta vantagens por ser um método direto e capaz de integrar a heterogeneidade espacial, em razão da sua capacidade de amostrar uma porção significativa de área. Contudo, este método impõe desafios no processamento dos dados, devido a quantidade de etapas e opções no processo dos dados de alta frequência. Outra problemática que deve ser considerada na utilização do método *Eddy Covariance* é a dinâmica espacial dos fluxos. O viés da área de estudo da torre de monitoramento é um importante aspecto para avaliar a representatividade dos fluxos (CHU *et al.*, 2021). Sendo assim, escolhas inadequadas podem impactar na qualidade dos dados, afetando negativamente nas validações de modelos hidrológicos que empreguem tais dados.

A quantificação da evapotranspiração, em áreas de Cerrado, tem sido objeto de discussão na comunidade científica (OLIVEIRA *et al.*, 2005; OLIVEIRA *et al.*, 2015; GARCIA-MONTIEL *et al.*, 2008; GIAMBELLUCA *et al.*, 2009; CABRAL *et al.*, 2015). Diversos fatores da vegetação, tais como densidade da vegetação e diversidade de espécies causam variações nas respostas hidrológicas (VILLALOBOS-VEGA *et al.*, 2014). Assim, o aprimoramento dessa componente com o viés da área de estudo é fundamental para o

entendimento dos processos hidrológicos em área de Cerrado.

Como no ano de 2018 foi instalado uma torre de monitoramento para determinação dos fluxos com o método *Eddy Covariance* de uma área de Cerrado *sensu stricto* em Itirapina (SP), faz-se necessária a otimização dos procedimentos de análise dos dados de evapotranspiração. Logo, este trabalho busca aprimorar o conhecimento quantitativo da evapotranspiração através da investigação da qualidade dos dados coletados pela torre de monitoramento nessa área de estudo e propor uma rotina automática de controle de qualidade e preenchimento de falhas.

2 OBJETIVOS

2.1 Objetivo Geral

Encontrar a melhor rotina computacional para a determinação da evapotranspiração através do método *Eddy Covariance* em área de Cerrado *sensu stricto*.

2.2 Objetivos Específicos

- i. Avaliar o desempenho dos software para o processamento de dados de fluxo de alta frequência, analisando o balanço de energia e métricas padrão de controle de qualidade.
- ii. Desenvolver uma ferramenta computacional para facilitar a implementação dos métodos de análise de área de contribuição e preenchimento de falhas.
- iii. Avaliar as técnicas de preenchimento para as diferentes janelas de falhas da série.
- iv. Obter uma série temporal consolidada para a evapotranspiração em área de Cerrado *sensu stricto*.

3 REVISÃO BIBLIOGRÁFICA

3.1 Evapotranspiração

A evapotranspiração (ET) é a soma de dois processos naturais: a evaporação e a transpiração da vegetação. A evaporação se trata de um fenômeno físico que acontece de forma gradual e em diversas condições de temperatura e pressão, porém é influenciado pelas condições da camada superficial atmosférica. A transpiração vegetal é um processo biológico, o qual depende do balanço energético ao nível da folha e da área foliar, e é controlada pela condutância estomática (ALLEN *et al.*, 1998).

A quantificação hídrica do complexo sistema solo-planta-atmosfera é dependente de diversos dados meteorológicos e da vegetação (ALLEN *et al.*, 2006; GONG *et al.*, 2006; MONSON; BALDOCCHI, 2014). O método de Penman-Monteith (Equação 3.1) é um dos métodos mais difundidos e recomendados para a estimativa de fluxo de calor latente (evapotranspiração). Trata-se de um método que combina o balanço de energia e o método de transferência de massa, o qual depende de variáveis da vegetação e dados atmosféricos. Porém, trata-se de um método dependente de variáveis que tentam reduzir a complexidade da camada vegetal em duas variáveis de resistência (ALLEN *et al.*, 1998; ALLEN *et al.*, 2006).

As nomenclaturas de resistência são divididas em resistência aerodinâmica e resistência superficial. O fluxo de vapor d'água através da abertura estomatal, área foliar e superfície do solo é descrito pela resistência superficial. Enquanto a resistência aerodinâmica é descrito pelo atrito do ar que percorre sobre a vegetação.

$$\lambda ET = \frac{\Delta(R_n - G) + \rho_a c_p \frac{(e_s - e_a)}{r_a}}{\Delta + \gamma(1 + \frac{r_s}{r_a})} \quad (3.1)$$

Em que: R_n [$W m^{-2}$] é a radiação líquida; G [$W m^{-2}$] é o fluxo de calor no solo; $(e_s - e_a)$ [kPa] representa o déficit de pressão de vapor no ar; ρ_a [$kg m^{-3}$] é a densidade média do ar; c_p [$MJ kg^{-1} °C^{-1}$] é o calor específico do ar; Δ [$kPa °C^{-1}$] representa a relação da pressão de vapor saturada e temperatura; γ [$kPa °C^{-1}$] é a constante psicométrica; r_s [$s m^{-1}$] e r_a [$s m^{-1}$] são a resistência superficial e aerodinâmica, respectivamente.

No bioma do Cerrado, os valores médios diários de ET dos estudos podem variar significativamente devido a sua heterogeneidade. Oliveira *et al.* (2005) e Giambelluca *et al.* (2009) analisaram o Cerrado denso, encontrando valores de ET de 1,7 – 2,5 mm/dia no período seco e até 4,6 mm/dia no período chuvoso. Enquanto, Cabral *et al.* (2015) obteve valores de ET para o período seco de 1,0 mm/dia e 7,1 mm/dia no período chuvoso do ano de 2010.

3.1.1 Método Eddy Covariance

Os métodos micrometeorológicos de ET são amplamente utilizados devido à capacidade de serem utilizados para vegetações heterogêneas e integrarem uma área significativa (STULL, 1988; ARYA, 2001; FOKEN, 2008). O método *Eddy Covariance*, ou covariância de turbilhões, é capaz de medir diretamente o fluxo de calor latente através dos turbilhões gerados na camada superficial. Estes turbilhões são redemoinhos sobrepostos de diferentes tamanhos com movimentos irregulares, os quais seguem o vento predominante (STULL, 1988). Para o método *Eddy Covariance* (Equação 3.2), utilizando um analisador de gás de massa específica e sem qualquer correção, o fluxo de massa é definido como (AUBINET; VESALA; PAPALE, 2012):

$$F_s^{EC} = \overline{w'\rho'_s} \quad (3.2)$$

Em que: w' [$m s^{-1}$] é a flutuação da velocidade vertical e ρ'_s [$g m^{-3}$] é a flutuação da massa específica de uma escalar (água ou dióxido de carbono). No caso especial do fluxo de vapor d'água, geralmente é definido em termos energéticos ou em fluxo de calor latente, logo:

$$\lambda ET = F_s^{EC} \lambda \quad (3.3)$$

Em que: λ [$MK kg^{-1}$] é o calor latente de vaporização.

3.1.1.1 Pré-processamento

A equação 3.2 é a formulação padrão do método *Eddy Covariance*, porém, imperfeições instrumentais e de posicionamento afetam os resultados. Desta forma, é necessário a realização de uma etapa de pré-processamento dos dados, com o objetivo de condicioná-los para o devido processamento dos fluxos. Os procedimentos incluem: remoção de valores inválidos ou imprecisos e a rotação de coordenadas.

A remoção de valores inválidos ou imprecisos inclui os valores *NaN*, os limites dos equipamentos de medição e os picos anômalos. Os picos anômalos na série temporal são identificados e removidos, de maneira iterativa, considerando os valores fora dos limites estabelecidos (SABBATINI *et al.*, 2018).

A rotação de coordenadas é essencial no tratamento dos dados, pois o método requer que as medições dos fluxos sejam perpendiculares às linhas aerodinâmicas médias do terreno (BALDOCCHI, 2019). Ou seja, uma das hipóteses do método *Eddy Covariance* é de que, em média, a velocidade do vento vertical é nula ($\bar{w} = 0$). No entanto, devido à dificuldade de se posicionar os sensores nesta posição, a disposição dos sensores é definida

na posição geopotencial criando uma média da velocidade do vento vertical diferente de zero.

A correção da orientação pode ser realizada principalmente através de duas formas: Rotação Dupla (HYSON; GARRAT; FRANCEY, 1977) e Regressão Planar (WILCZAK; ONCLEY; STAGE, 2001). A rotação dupla consiste na conversão entre dois sistemas de coordenadas utilizando matrizes de rotação (*Yam, Pitch and Roll*), a qual para cada intervalo médio (30 minutos) é realizado a rotação. Enquanto, a regressão planar é um método que requer uma série temporal sem interrupções ou mudanças da disposição do anemômetro. E através da série de velocidades, define-se uma equação de ajuste para que a velocidade vertical média se aproxime do zero.

3.1.1.2 Processamento

As limitações dos equipamentos para a amostragem dos dados torna a execução de certas correções necessárias, então, a etapa de processamento é definida pelas correções SND e espectrais, e a adição de termos WPL.

A correção SND, desenvolvida por Schotanus, Nieuwstadt e De Bruin (1983), transforma a temperatura sônica (não considera a presença de vapor d'água) em temperatura do ar real. Apesar de um termopar ser mais preciso que a temperatura sônica, a medição da temperatura sônica é extremamente útil para determinação de pequenas e rápidas variações de temperatura a qual são utilizadas na computação do fluxo de calor sensível (KAIMAL; GAYNOR, 1991; AUBINET; VESALA; PAPAIE, 2012). A equação para correção é definida por:

$$T_c = \frac{T_s}{\frac{1+0,32*\rho_{H_2O}*R*T_s}{p*18,02}} \quad (3.4)$$

Em que: T_s é a temperatura sônica, T_c é a temperatura do ar real, R é a constante universal do gases, p é a pressão do ambiente e ρ_{H_2O} é a densidade da água.

As correções espectrais são necessárias visto que a instrumentação utilizada para o método age como um filtro, removendo componentes de sinal de alta e baixa frequência. Este fenômeno ocorre devido à resposta de frequência dada pelo equipamento, ou seja, o sinal capturado é do tipo discreto e não contínuo. Outro fator filtrante é o próprio intervalo médio de análise, pois este é responsável por delimitar as contribuições dos turbilhões de baixa frequência (KAIMAL *et al.*, 1972; MOORE, 1986).

Assim, de forma geral a atenuação do fluxo é descrita por funções de transferência ($TF(f)$) multiplicadas pelo co-espectro ideal ($C_{ws}(f)$), definidas como:

$$F_a = \frac{\overline{w's'_{medido}}}{\overline{w's'_{ideal}}} = \frac{\int_0^\infty TF(f)C_{ws}(f)df}{\int_0^\infty C_{ws}(f)df} \quad (3.5)$$

Para a correção dessas atenuações espectrais, existem diversos métodos que são capazes de fazer as correções para baixa e alta frequência. A literatura não expõe o melhor método para utilizar, pois isso depende de diversas condições do ambiente e do equipamento. A tabela 1 mostra os principais métodos de correções espectrais, sendo que alguns são capazes de corrigir as atenuações na alta frequência e outros na baixa frequência.

Tabela 1 – Lista de principais métodos de correção espectral

Métodos	Modelo	Função de transferências
Moncrieff <i>et al.</i> (1997)	Kaimal <i>et al.</i> (1972), Moore (1986)	Resposta do sensor, intervalo médio, separação dos sensores
Massman (2000)	Kaimal e Finnigan (1994), Horst (1997)	Resposta do sensor, intervalo médio, separação dos sensores
Horst (1997) Fratini <i>et al.</i> (2012)	Kaimal e Finnigan (1994)	Resposta do sensor Resposta do sensor, intervalo médio, separação dos sensores
Ibrom <i>et al.</i> (2007)	Moore (1986)	Resposta do sensor (recomendado para sistemas <i>closed-path</i>), intervalo médio

Fonte: LI-COR¹

Os termos WPL, ou termos Webb-Pearman-Leuning (WEBB; PEARMAN; LEUNING, 1980), referem-se aos autores que primeiro discutiram sobre a flutuação de densidade de vapor d'água. Uma forma de visualizar este fenômeno é imaginar uma superfície quente com zero de fluxo ($\overline{w'\rho_s}$) com o ar aquecendo gradualmente, resultando em uma expansão do ar próximo da superfície que será detectado pelo equipamento de medição como fluxo ascendente do gás medido. Portanto, são adicionados aos valores de fluxo de água dois termos: (i) o termo de diluição da água e (ii) o termo de expansão termal (Equação 3.6). Estes termos devem ser aplicados para analisadores de gases do tipo *open-path*, pois estes medem a densidade dos gases (BURBA, 2013).

$$F_{q,WPL} = \underbrace{\overline{w'\rho'_q}}_{\text{fluxo } H_2O} + \underbrace{\left(\frac{M_a \overline{\rho_q}}{M_w \rho_a}\right) \overline{w'\rho'_q}}_{\text{Termo de diluição do } H_2O} + \underbrace{\left(1 + \frac{M_a \overline{\rho_q}}{M_w \rho_a}\right) \frac{\overline{\rho_q} \overline{w'T'}}{\overline{T}}}_{\text{Termo devido à expansão térmica}} \quad (3.6)$$

¹ Disponível em: < https://www.licor.com/env/support/EddyPro/topics/advanced-settings-spectral_corrections.html>. Acesso em: 23 mar. 2021.

3.1.1.3 Controle de qualidade

A medição de fluxos não é perfeita por diversos fatores como: limitações instrumentais, fenômenos físicos e particularidades do local de estudo (BURBA, 2013). A aplicação das correções impacta de maneira significativa os resultados finais, porém, isso não significa que todos os intervalos de 30 minutos terão valores de ótima qualidade.

Então, o controle de qualidade dos dados é a etapa responsável pela classificação de qualidade de cada intervalo de medição (30 minutos), que possibilita separar os dados considerados insatisfatórios, aceitáveis e ótimos. Para isso, os primeiros testes aplicados são o teste de estado estacionário e o teste das características integrais de turbulência (FOKEN; WICHURA, 1996; MAUDER; FOKEN, 2011; BURBA, 2013).

O teste de estado estacionário investiga se, dentro do intervalo médio de análise (30 minutos), os parâmetros estatísticos não variaram significativamente. Para isto, Vickers e Mahrt (1997) faz uso da regressão da série temporal do intervalo médio, em que se verificou a estacionariedade relativa do vento predominante e do vento cruzado. Mahrt (1998) utiliza a razão entre o desvio padrão de segmentos e a covariância do intervalo médio de análise, enquanto que Foken e Wichura (1996) verificam a qualidade utilizando segmentos do intervalo médio e a covariância do período total.

O segundo teste é a verificação das características integrais de turbulência, cuja hipótese da metodologia *Eddy Covariance* requer que os fluxos sejam turbulentos para que ocorra a troca de gases entre a atmosfera e a superfície. Para isto, utilizam-se funções modelo dependentes da estabilidade atmosférica, desvio padrão normalizado das componentes de velocidade de vento e dos escalares para medição (KAIMAL; FINNIGAN, 1994; FOKEN; WICHURA, 1996; AUBINET; VESALA; PAPALE, 2012). E a partir das funções modelo e as medidas, define-se um limite de aceitação para considerar um intervalo médio suficiente para um bom desenvolvimento do fluxo turbulento.

A partir do desempenho dos testes, o fluxo é classificado de acordo com o grau de qualidade do intervalo de medição (Tabela 2).

Tabela 2 – Tabela de métodos de sinalização da qualidade dos dados

Métodos	Tipo	Descrição
Foken e Wichura (1996)	0 - 9	0 = Qualidade boa 9 = Qualidade ruim
Mauder e Foken (2011)	0 - 2	0 = Qualidade boa 2 = Qualidade ruim

Fonte: LI-COR¹

¹ Disponível em: < https://www.licor.com/env/support/EddyPro/topics/advanced-settings-spectral_corrections.html>. Acesso em: 23 mar. 2021.

Considerando que os fluxos de calor sensível e latente são calculados, a determinação do fechamento de balanço de energia é uma importante ferramenta para garantir a qualidade dos dados. Diversos estudos que avaliam o fechamento do balanço energético descobrem uma subestimação de 10 a 30% (TWINE *et al.*, 2000; ONCLEY *et al.*, 2007). Contudo, essa subestimação é atrelada ao desafio na determinação da área de contribuição e na amostragem espacial dos sensores de radiação líquida e fluxo de calor do solo (BALDOCCHI, 2008). A equação de fechamento de balanço de energia é definida como:

$$LE + H = R_n - G \quad (3.7)$$

Em que: $LE [W m^{-2}]$ é o fluxo de calor latente, $H [W m^{-2}]$ é o fluxo de calor sensível, $R_n [W m^{-2}]$ é a radiação líquida e $G [W m^{-2}]$ é o fluxo de calor do solo.

3.2 Método de determinação da área de contribuição

A determinação da área de contribuição ou *Footprint* é utilizada para caracterizar a área responsável por gerar os fluxos registrados pelos equipamentos de medição. O entendimento e verificação da origem dos fluxos são essenciais para a execução do método de *Eddy Covariance*, pois este método micrometeorológico é dependente de apenas um ponto de medição. Diversos fatores influenciam a determinação do *Footprint*, sendo que os principais fatores influenciadores são a altura de medição, direção e velocidade do vento, rugosidade da superfície e estabilidade atmosférica (LECLERC; THURTELL, 1990; KLJUN; ROTACH; SCHMID, 2002).

Segundo Leclerc e Foken (2014), existem cinco principais tipos de abordagens utilizadas pela comunidade científica: (i) soluções analíticas com diversos graus de complexidade, (ii) simulações lagrangianas baseados em modelos estocásticos, (iii) modelos de fechamento de ordem superior, (iv) simulação de grandes turbilhões e (v) combinações entre abordagens. Dentre os diversos métodos utilizados, Kljun *et al.* (2015) desenvolveram um método que mistura as simulações lagrangianas e as soluções analíticas, criando assim uma abordagem que é compatível, em termos de velocidade de processamento, para longas séries temporais e para diversas condições atmosféricas.

O método por Kljun *et al.* (2015) é uma versão aprimorada de sua versão anterior (KLJUN *et al.*, 2004), sendo que a nova versão utiliza dados de vento na direção predominante e a componente perpendicular, gerando assim estimativas em duas dimensões. O cálculo é realizado para cada intervalo médio de análise (30 minutos) e utiliza diversos parâmetros de escala. As etapas de cálculo são definidas por:

1. Cálculo da distância não dimensional contra o vento (X^*) em função de x .

$$X^* = \frac{x}{z_m} \left(1 - \frac{z_m}{h} \right) \left(\frac{\bar{u}(z_m)}{u_*} k \right)^{-1} \quad (3.8)$$

Sendo que: x [m] é a distância no eixo x da torre; z_m [m] é a altura de medição da torre; h [m] é altura da camada limite do planeta; $\bar{u}(z_m)$ [m/s] é a velocidade média na altura de medição; u_* [m/s] é a velocidade de fricção e k é a constante de von Karman.

2. Cálculo da parametrização do *Footprint* não dimensional.

$$F^{y*} = 1.452(X^* - 0.136)^{1.991} \exp\left(\frac{-1.462}{X^* - 0.136}\right) \quad (3.9)$$

$$\sigma_y^* = 2.17 \left(\frac{1.66(X^*)^2}{1 + 20.0X^*} \right)^{1/2} \quad (3.10)$$

3. Cálculo da função de *Footprint* integrada ao vento cruzado e o desvio padrão.

$$\bar{f}^y = F^{y*} \left(1 - \frac{z_m}{h}\right) \frac{u_*}{\bar{u} z_m k} \quad (3.11)$$

$$\sigma_y = \sigma_y^* \frac{z_m}{p_{s1}} \frac{\sigma_v}{u_*} \quad (3.12)$$

Sendo que: σ_y é o desvio padrão da velocidade perpendicular do vento predominante e p_{s1} é um fator de proporcionalidade que depende da estabilidade atmosférica.

4. Por fim, calcular o fluxo de *Footprint*.

$$f(x, y) = \bar{f}^y(x) \frac{1}{\sqrt{2\pi}\sigma_y} \exp\left(-\frac{y^2}{2\sigma_y^2}\right) \quad (3.13)$$

Dentre os métodos rápidos de determinação do *Footprint*, este é um novo método que consegue abranger uma diversidade de situações e condições atmosféricas. Pode ser aplicado tanto durante o dia quanto no período noturno, no ano inteiro, para uma diversidade de tipos de vegetação, de campos a florestas (KLJUN *et al.*, 2015).

3.3 Métodos de preenchimento de falhas

O aumento de estudos com a utilização do método *Eddy Covariance* a partir dos anos 80 para a medição direta de fluxos de carbono e água requisitou a utilização de equipamentos com uma alta taxa de aquisição de dados (10 Hz - 20 Hz), permitindo a agregação dos dados colhidos em uma alta resolução temporal (30 minutos). Porém, diversos critérios de qualidade relacionados à condição atmosférica, como a velocidade do vento mínima de fricção (u_*), limites e falhas do equipamento acabam por resultar em uma alta perda de dados (FALGE *et al.*, 2001; REICHSTEIN *et al.*, 2005; MOFFAT *et al.*, 2007).

A fragmentação da série faz com que a série se torne incompleta, resultando na dificuldade de análise quando a escala é diária, sazonal ou anual. Então, os métodos de preenchimento de falhas têm como objetivo reduzir as falhas e não modificar as características da série temporal.

3.3.1 Variação diurna média

O método de preenchimento de falhas pela variação média utiliza apenas valores da variável objetivo (ET). Este método tenta substituir os valores de 30 minutos com falha pela média dos valores dos dias adjacentes nos mesmos horários da falha.

$$y = \frac{1}{2n} \left(\sum_{i=1}^n x_i + \sum_{j=-1}^{-n} x_j \right) \quad (3.14)$$

Sendo que, n é a quantidade de dias antes ou depois da falha.

A janela de dias adjacentes a serem escolhidos pode variar de acordo com as configurações, porém, o tamanho da janela é usualmente de 3 - 15 dias (FALGE *et al.*, 2001). Essa janela de dias pode ser realizada de maneira progressiva, ou seja, caso não haja dados suficientes para preencher certa falha, um aumento da janela de dias pode ser efetuado, assim, aumentando a capacidade de preenchimento das falhas na série. Contudo, esse aumento da janela causa uma redução da qualidade do preenchimento.

Em outros trabalhos, esse método de interpolação apresenta um desempenho moderado, porém consistente (MOFFAT *et al.*, 2007). Contudo, a dependência de apenas uma variável (ET), sem a utilização de variáveis meteorológicas auxiliares, faz com que possíveis variações intermediárias do fluxo na falha não seja capaz de prever.

3.3.2 Regressão linear múltipla

O método de regressão linear múltipla utiliza diversas variáveis meteorológicas ou qualquer outros tipos de variáveis para estimar a variável objetivo (ET). A utilização desta técnica examina a relação linear entre as variáveis independentes e ajusta um modelo linear com uma função objetivo de minimizar a soma residual dos quadrados entre os alvos observados no conjunto de dados e os alvos previstos pela aproximação linear.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \epsilon, \text{ onde } k \in Z \quad (3.15)$$

Sendo que, x são variáveis explicativas e β são os coeficientes de regressão.

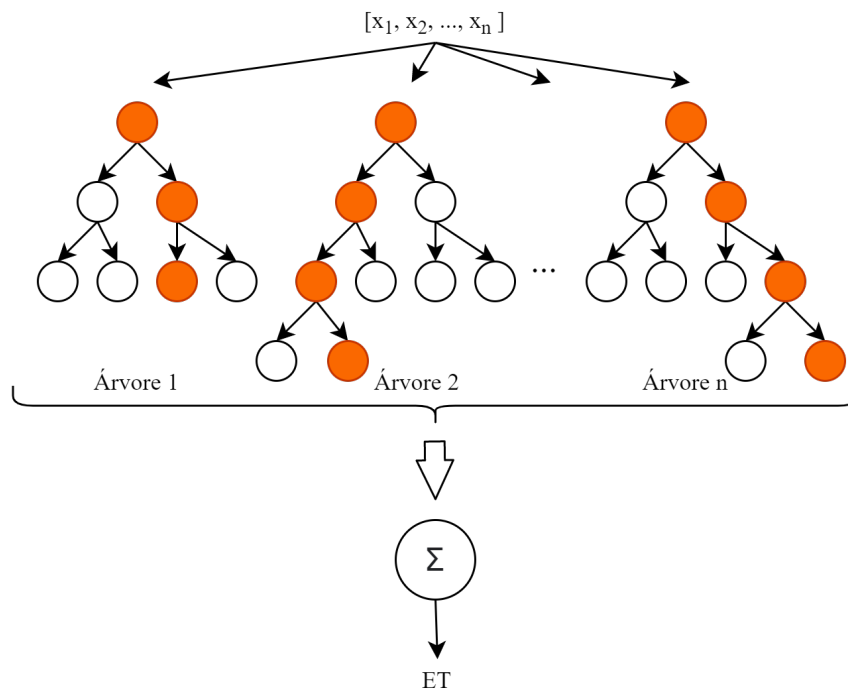
Após definido quais variáveis a serem utilizadas, o método é capaz de preencher somente os dias que todas as variáveis de entradas estão presentes.

Este método de regressão linear múltipla, apesar de agregar diversas variáveis meteorológicas, possui a desvantagem do comportamento linear. Entende-se que apesar de haver correlações entre variáveis meteorológicas e a ET, o comportamento é não linear e há variação de comportamento devido a própria variação da sazonalidade. Assim, o ideal seria a utilização de uma série de equações não lineares e a separação por diversos períodos do ano, gerando um melhor ajuste das equações.

3.3.3 Random Forest

O método de regressão logística, *Random Forest*, utiliza como dados de entrada diversas variáveis meteorológicas ou outros tipos de variáveis para determinar a variável objetivo (ET).

Figura 1 – Esquema Random Forest



Fonte: Autor

O resultado final é obtido através da média do conjunto de modelos de árvores (*Trees*) (BREIMAN, 2001). Os modelos de árvores são formados por folhas (*leaves*) e nós de decisões, que são as responsáveis por separar os dados até que a criação de novas folhas não seja capaz de gerar possíveis valores de se prever (Figura 1). A separação dos dados ocorre em nós de decisões que realizam perguntas, por exemplo: "A radiação líquida é maior que 500 W m^{-2} ?", e de maneira recursiva os dados são divididos em diversas folhas após passado por diversos nós de decisões.

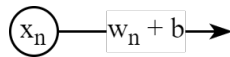
Alguns dos parâmetros importantes para criação das árvores são o número de árvores a ser determinadas, profundidade máximas das árvores e a quantidade máximas de

folhas. Estes três parâmetros são os principais parâmetros que determinam a rapidez do processo e a qualidade dos resultados finais.

3.3.4 Redes neurais

A estrutura das redes neurais é baseada em unidades básicas de um neurônio, sendo dividida em duas partes: nós e redes. Nessas unidades básicas, cada neurônio é composto por pesos (w_n) e bias (b) (Figura 2) que recebem e transmitem sinais para os nós adjacentes. Formam-se uma rede de nós com diversas camadas sendo caracterizadas por camadas de entrada, escondidas e de saída (Figura 3).

Figura 2 – Exemplo de neurônio básico



Fonte: Autor

A forma com que uma rede neural aprende a solucionar um problema se inicia com os pesos e bias aleatórios da rede. Neste estágio, a rede neural não possui nenhuma informação, pois necessita que esta rede seja treinada com dados. O treinamento da rede consiste nos ajustes dos pesos entre nós para que os dados de entrada melhor consigam representar os dados de saída. Na etapa de treinamento, dois aspectos são essenciais para que este treinamento seja eficiente: uma função de perda e um otimizador.

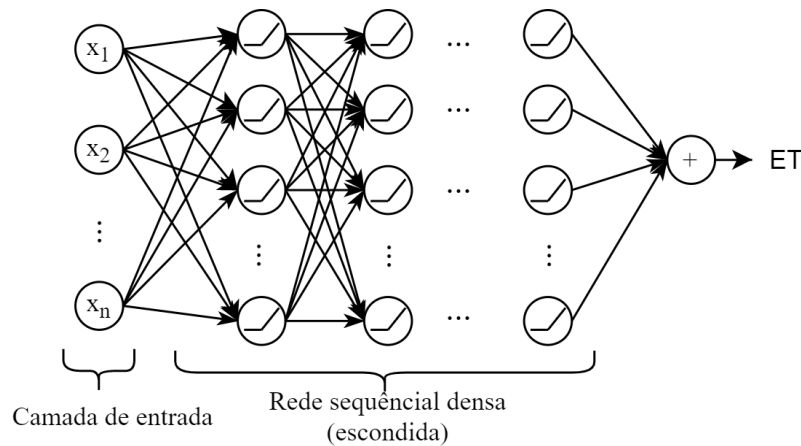
A função de perda tem como objetivo medir qual é o desempenho da rede. Uma função de perda comum para problemas de regressão é o erro absoluto médio, que é o responsável por medir a diferença absoluta do resultado real e o calculado. Enquanto o otimizador é o responsável por ajustar os pesos das ligações entre nós com o objetivo de diminuir a função de perda. Geralmente, os algoritmos de otimização são da família da descida gradiente estocástica. Estes algoritmos ajustam os pesos de maneira iterativa por pacotes (*batch*) e os dados de treinamento são rodados diversas vezes (cada rodada de treinamento é chamada de *epochs*) até que o número de rodadas de dados seja atingido, ou um valor mínimo da métrica seja atingido (SHERSTINSKY, 2020).

3.3.4.1 Redes neurais densas (DNN)

A capacidade de formar e resolver relações complexas é a marca de uma rede neural densa.

A complexidade das relações pode ser gerada pela criação de camadas internas, também chamadas de camadas escondidas, as quais são as camadas que o usuário não consegue interagir diretamente. Contudo, as camadas escondidas por si só não são capazes de aprender relações não lineares; desta forma, funções de ativação são utilizadas para serem aplicadas na saída destas camadas. Por exemplo, a função de ativação mais simples

Figura 3 – Exemplo de rede neurais



Fonte: Autor

é aquela que retifica a parte negativa para zero (equação 3.16), também chamada de ReLU (*rectified linear unit*).

$$f(x) = \begin{cases} 0 & , \text{ se } wx + b \leq 0 \\ wx + b & , \text{ se } wx + b > 0 \end{cases} \quad (3.16)$$

Após empilhar camadas escondidas com funções de ativações, a rede passa a ter características não lineares. A figura 3 é um exemplo de rede neural densa, com múltiplas camadas escondidas, com a forma de ativação ReLU nas camadas escondidas. Porém, sua última camada é uma unidade linear (sem função de ativação), a qual é a responsável por agregar todos os pesos e viés da camada anterior. Esta característica torna a rede compatível para um problema não linear capaz de resolver tarefas de regressão, com o objetivo de prever valores numéricos.

Uma simples alteração na arquitetura da rede é capaz de produzir uma rede compatível para solucionar tarefas de classificação. Para isto, a principal alteração seria na utilização de uma função de ativação na última camada do tipo sigmoide (para classificação de problemas binários) ao invés de não usar função alguma. A função de ativação sigmoide tem a forma de um "S", assim, os valores próximos de zero poderiam ser representados por um tipo "A" enquanto valores próximos de 1 poderiam ser representados por um tipo "B".

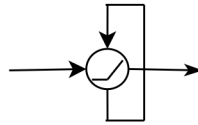
Estes são dois exemplos simples que utilizam redes neurais densas com nós simples (camadas escondidas). No entanto, existem outros tipos de nós que conseguem armazenar informações, criando redes denominadas recorrentes.

3.3.4.2 Redes neurais recorrentes (RNN)

As redes neurais recorrentes são caracterizadas pelas camadas escondidas onde os nós são capazes de armazenar as saídas ativadas e converter como um dado de entrada

para o próprio nó (Figura 4).

Figura 4 – Exemplo de nó recorrente



Fonte: Autor

Esta capacidade de armazenar informação acaba sendo limitada devido ao problema de gradiente de desaparecimento (GRAVES; SCHMIDHUBER, 2005), a qual limita a quantidade de informações de longo termo. Assim, uma variante das redes neurais recorrentes é o uso de memória longa de curto prazo (LSTM).

3.3.4.3 Redes de memória longa de curto prazo (LSTM)

A rede de memória longa de curto prazo foi desenvolvida para mitigar as limitações do RNN (HOCHREITER; SCHMIDHUBER, 1997) e para tarefas a qual contexto é de importância vital. Este tipo de arquitetura de rede neural têm sido utilizado de maneira abrangente quando o objetivo é dependente de algum tipo de sequência. Assim, problemas que utilizam dados sequenciais como séries temporais são capazes de convergir de maneira eficiente (GRAVES; SCHMIDHUBER, 2005). E para uma profunda análise das equações envolvidas no método, Sherstinsky (2020) equaciona e prova todas as etapas do processo dessa arquitetura de rede neural.

Contudo de maneira geral, o funcionamento da rede é composta a célula de uma rede de memória longa de curto prazo é composta por um vetor de memória, diversos portões lógicos e 4 redes neurais.

O vetor de memória é ligado ao longo da cadeia iterativa, somente alterado em operações pontuais. Os portões lógicos são compostos por uma rede neural sigmoidal e uma operação pontual, sendo a responsável por determinar qual informação é passada. A saída da rede neural sigmoidal é binária (zero ou um), determinando qual parte do vetor sofrerá modificação.

A primeira etapa do processo do LSTM é a decisão de qual informação será esquecida no vetor de memória. Esta decisão é realizada com o *Forget Gate*, que possui como dados de entrada o vetor de dados e o vetor de memória (do ciclo anterior) e como saída da rede neural sigmoidal tem um vetor com valores binários.

Na próxima etapa é decidida qual nova informação será armazenada no vetor de memória. Então, divide-se em duas partes: portão de entrada (*input gate layer*) e um vetor

com novos valores candidatos. Sendo que o portão de entrada é o responsável por decidir quais valores entrarão e em quais posições.

Atualiza-se o vetor de memória para um novo estado, baseado nas etapas anteriores. Primeiramente esquecendo as posições calculadas pela rede sigmoideal e implementando as novas informações nas devidas posições. Estes processos ocorrem via operação pontual de multiplicação e adição. Finalmente, o dado de saída calculado é baseado no estado do vetor de memória, o qual trata-se de uma versão filtrada. Então, o portão lógico decide qual parte do vetor de memória será o dado de saída.

Para o preenchimento de falha, pode-se utilizar da capacidade de memória do LSTM para criar camadas de células de LSTM capazes de prever, a partir de dados meteorológicos, os valores de ET.

3.3.5 Equação de Penman-Monteith inversa

Este método utiliza a equação de Penman-Monteith (Equação 3.1) para o preenchimento de falhas. Primeiramente, necessita-se determinar uma curva mensal de condutância superficial e para isto, a inversão da equação de Penman-Monteith (Equação 3.17) pode ser computada para sua determinação (JARVIS; MCNAUGHTON, 1986; GRACE *et al.*, 1995; LU *et al.*, 2003; VOURLITIS *et al.*, 2002; MALLICK *et al.*, 2018). Deve-se lembrar que a condutância superficial é o inverso da resistência superficial.

$$g_s = \frac{\gamma \lambda ET g_a}{\Delta(R_n - G) + K_{time} \rho_a c_p V PD g_a - \lambda(\Delta + \gamma) ET} \quad (3.17)$$

Sendo que, K_{time} é a unidade de conversão igual a $3600 \text{ s } h^{-1}$, $g_s [m \text{ s}^{-1}]$ é a condutância superficial e $g_a [m \text{ s}^{-1}]$ é a condutância aerodinâmica.

Desta forma, a condutância superficial fica em função das variáveis meteorológicas e da condutância aerodinâmica. O cálculo da condutância aerodinâmica é definido segundo Vourlitis *et al.* (2002), através da equação 3.18.

$$g_a = \left[\frac{u}{u^{*2}} \right]^{-1} \quad (3.18)$$

Sendo que, $u [m/s]$ é a velocidade do vento e $u^* [m/s]$ é a velocidade de fricção.

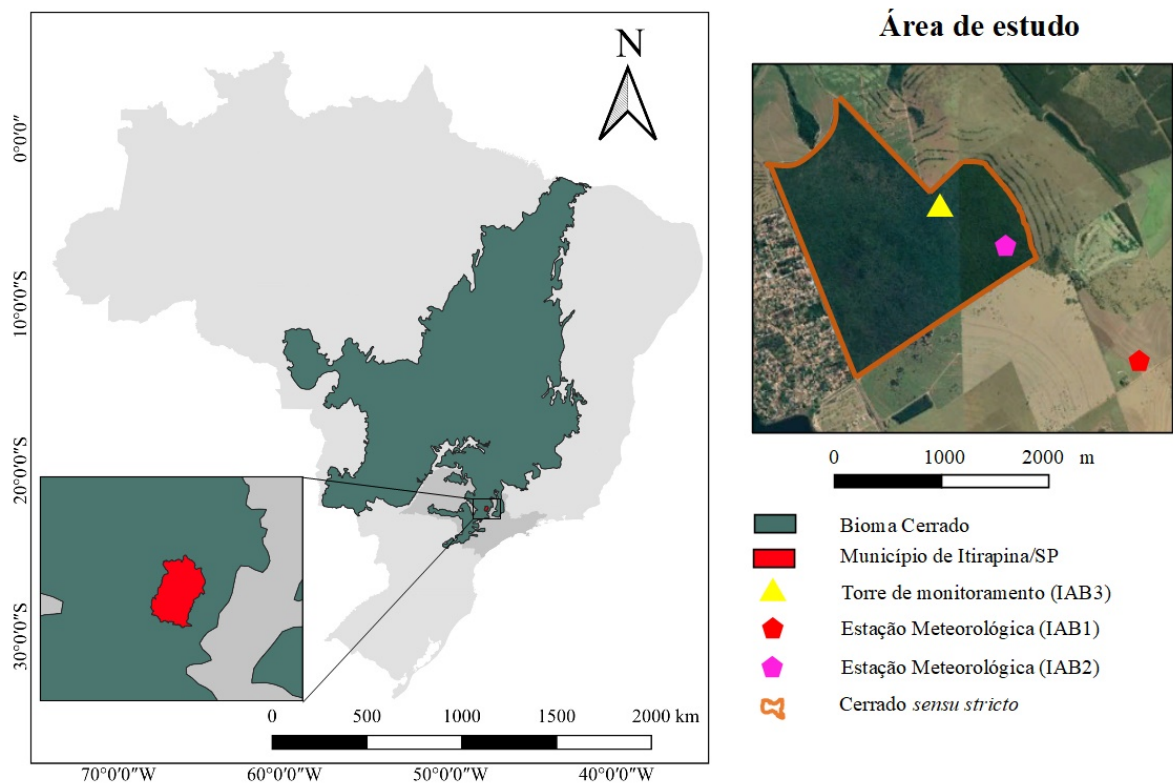
Após determinados os valores médios mensais da condutância superficial e aerodinâmica, utiliza-se essas variáveis e os dados meteorológicos na equação 3.1 para o preenchimento de falhas.

4 MATERIAL E MÉTODOS

4.1 Descrição da área de estudo

A área de estudo está localizada no Instituto Arruda Botelho (IAB), no município de Itirapina, Estado de São Paulo (latitude 22°11'5"S, longitude 47°51'11"O) com elevação de 790 metros, inserida no bioma Cerrado e em região de afloramento do Sistema Aquífero Guarani (SAG). O IAB é localizado próximo ao município de Brotas/SP e Represa do Lobo (Figura 5), sendo dividido em duas áreas: IAB1, que contém diferentes usos do solo; IAB2 e IAB3, que estão dentro do remanescente de Cerrado *sensu stricto* de cerca de 330 hectares.

Figura 5 – Área de estudo

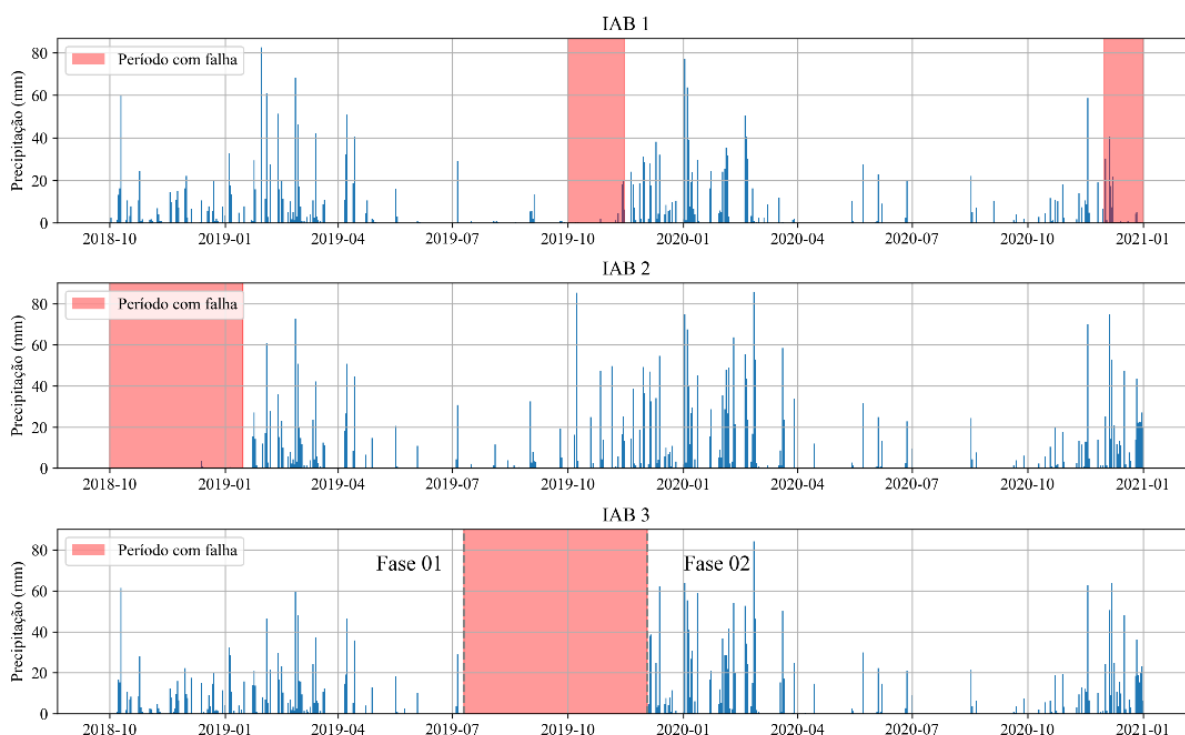


Fonte: Autor

O estudo é focado na torre de monitoramento (IAB3) instalada na área de Cerrado *sensu stricto* (Figura 5). Entretanto, foram também utilizados dados das estações meteorológicas próximas: (i) estação meteorológica (IAB1) e (ii) estação meteorológica (IAB2). No IAB1 estão instaladas parcelas experimentais de 3 tratamentos distintos (cana-de-açúcar, pastagem, soja e solo exposto), desde de 2011. Enquanto que a estação de monitoramento IAB2 está localizada dentro do remanescente de Cerrado *sensu stricto* desde 2012.

O clima da região é subtropical úmido (Cwa, Köppen), com verão quente e chuvoso e inverno seco. Os dados climatológicos da estação climatológica do CRHEA-USP de 1979 até 2014 indicam precipitação média anual observada de 1486 mm, temperatura média de 21,6 °C e umidade relativa de 71% (CABRERA *et al.*, 2016). Os dados de chuva das estações meteorológicas do IAB mostram precipitação de 1600 mm para o ano de 2019 (IAB 1) e uma precipitação de 1721 mm (IAB 2) e 1891 mm (IAB 3) para o ano de 2020 (Figura 6). O período seco compreende os meses de abril a setembro; e o período chuvoso corresponde ao intervalo de outubro a março.

Figura 6 – Precipitação nas estações do IAB1, IAB2 e IAB3

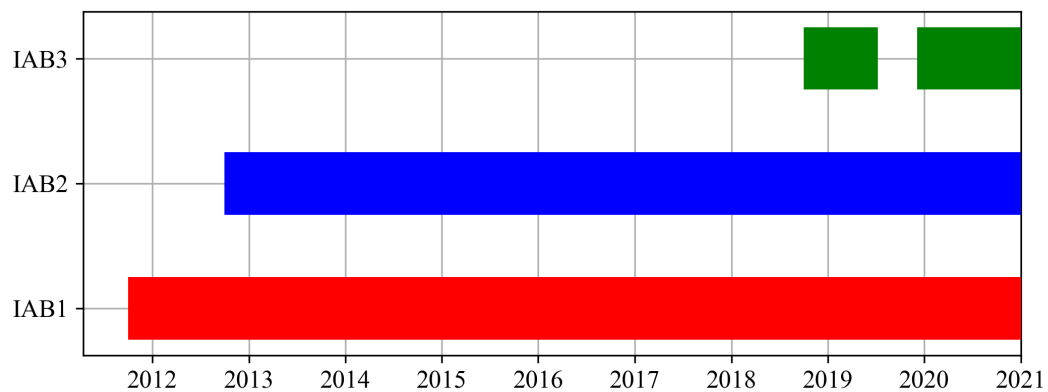


Fonte: Autor

A extensão dos dados meteorológicos das estações IAB1, IAB2 e IAB3 possui uma grande sobreposição entre as duas primeiras estações e cerca de 2 anos de dados para o IAB3 (Figura 7). Porém, no IAB3 existe uma lacuna de 5 meses, aproximadamente, devido a problemas técnicos no armazenamento de dados. Sendo assim, dividiu-se a análise dos dados em 2 fases: (i) dados até a grande falha (Fase 1) e (ii) a partir da retomada de dados até atualmente (Fase 2).

Em relação ao IAB1 e IAB2, apesar de haver uma constante manutenção nos equipamentos, problemas como entupimento do pluviômetro acabam sendo recorrentes. Além disso, um evento de descarga elétrica atingiu o sensor de umidade relativa do IAB2 no final do ano de 2015 desativando o sensor permanentemente, contudo, outros equipamentos de medição (como o radiômetro, velocidade do vento, e outros) continuaram funcionando normalmente.

Figura 7 – Período de coleta de dados das estações



Fonte: Autor

4.1.1 Torre de monitoramento (IAB3)

A metodologia do *Eddy Covariance* requer instrumentos que trabalham em alta frequência como o anemômetro sônico e o analisador de gases. A torre de monitoramento (Figura 8) possui os instrumentos (Tabela 3) necessários para a aplicação da metodologia do *Eddy Covariance* e aplicação da correção SND (Equação 3.4), correções espectrais e termos WPL (Equação 3.6).

Figura 8 – Torre de monitoramento IAB3



Fonte: Autor

A torre de monitoramento (IAB3) está localizada no limiar da área de vegetação Cerrado *sensu stricto* (Figura 5 e 9). Para a determinação de sua localização, considerou-se a direção predominante dos ventos e a possibilidade de acesso à torre. Porém, a proximidade

Tabela 3 – Instrumentos utilizados para a coleta e armazenamento de dados (IAB3)

Instrumentos	Modelo	Altura
Analisador de Gás (<i>Open-Path</i>) e Anemômetro Sônico 3D	IRGASON	16 m
Sonda de Temperatura/Umidade Relativa	HMP155A	16 m
Sensor de Radiação Líquida	CNR4	10 m
Placa de Fluxo de Calor no Solo	HFP01	-0.1 m
Pluviômetro	TB4	16 m
Sonda de Temperatura	108-U	16 m
Datalogger	CR3000	-
Módulo de Cartão de Memória	CFM100	-
Cartões de memória	CFMC2G	-

da torre a uma das bordas da área faz com que existam contribuições da vegetação adjacente dependendo, principalmente, da direção do vento.

Figura 9 – Vista da torre de monitoramento (IAB3) para área de Cerrado *sensu stricto* (a) e para área de pasto (b) e vista superficial do Cerrado *sensu stricto* (c) e da área de pasto (d)



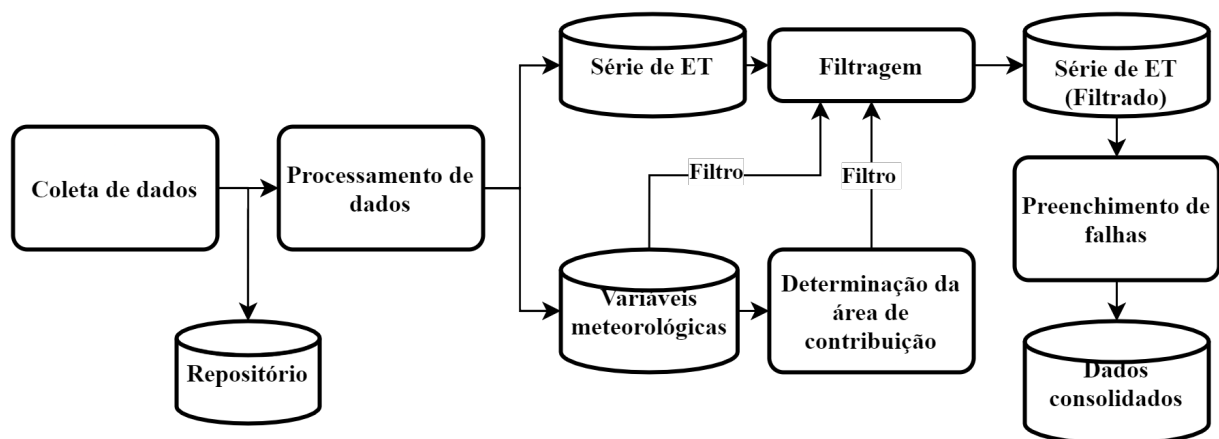
Fonte: Autor

4.2 Etapas do trabalho

As etapas do trabalho foram sequenciais (Figura 10), ou seja, necessitava-se dos dados da etapa anterior para prosseguir. A primeira etapa foi a coleta de dados, isto é, coletas a cada três semanas aproximadamente do cartão de memória na torre de monitoramento (IAB3). Os dados coletados eram armazenados em um repositório antes da realização do processamento.

A etapa de processamento de dados gera a série de dados de ET, assim como os dados meteorológicos. Os dados meteorológicos são utilizados para a determinação da área de contribuição, gerando um filtro que foi usado posteriormente. Enquanto a série de ET precisava passar por um aprimoramento, era realizada uma etapa de filtragem dos dados. A filtragem é responsável por retirar dados que não satisfazem a qualidade mínima, resultando em perda de dados na série de ET. Por fim, a etapa de preenchimento das falhas, assim, gerando uma série de ET consolidada.

Figura 10 – Etapas do trabalho



Fonte: Autor

4.3 Processamento de dados de alta frequência

4.3.1 Escolha do software para processamento

A escolha do software para o processamento dos dados de alta frequência considerou dois principais pontos: uso dentro da comunidade científica e flexibilidade na escolha de parâmetros. Desta forma, o software *EddyPro-7.0.6* foi o escolhido para processar os dados de alta frequência (20 Hz). A escolha do *EddyPro* não necessariamente significa que os outros software não são capazes de processar os dados, porém, algumas restrições do *TK3* e *EasyFluxPC v.1.0* impõe desafios que o *EddyPro* não possui.

A citar como exemplo, o software *TK3*, apesar de ser um dos software mais antigos para o processamento de dados e produzir resultados compatíveis com o *EddyPro* (FRATINI;

Tabela 4 – Métodos para processamento de dados de alta frequência (EddyPro)

Código	Rotação de Coordenadas	Baixa Frequência	Alta Frequência
p00	Rotação 2D	-	-
p01	Rotação 2D	Moncrieff <i>et al.</i> (1997)	Moncrieff <i>et al.</i> (1997)
p02	Rotação 2D	Moncrieff <i>et al.</i> (1997)	Massman (2000)
p03	Rotação 2D	Moncrieff <i>et al.</i> (1997)	-
p04	Rotação 2D	-	Moncrieff <i>et al.</i> (1997)
p05	Rotação 2D	-	Massman (2000)
p06	Planar-Fit	-	-
p07	Planar-Fit	Moncrieff <i>et al.</i> (1997)	Moncrieff <i>et al.</i> (1997)
p08	Planar-Fit	Moncrieff <i>et al.</i> (1997)	Massman (2000)
p09	Planar-Fit	Moncrieff <i>et al.</i> (1997)	-
p10	Planar-Fit	-	Moncrieff <i>et al.</i> (1997)
p11	Planar-Fit	-	Massman (2000)
p12	Rotação 2D	-	Horst (1997)
p13	Planar-Fit	-	Horst (1997)
p14	Rotação 2D	Moncrieff <i>et al.</i> (1997)	Horst (1997)
p15	Planar-Fit	Moncrieff <i>et al.</i> (1997)	Horst (1997)
p16	Rotação 2D	-	Fratini e Mauder (2014)
p17	Planar-Fit	-	Fratini e Mauder (2014)
p18	Rotação 2D	Moncrieff <i>et al.</i> (1997)	Fratini e Mauder (2014)
p19	Planar-Fit	Moncrieff <i>et al.</i> (1997)	Fratini e Mauder (2014)

MAUDER, 2014), não permite o processamento de dados quando existe mudança de ano (exemplo: 01/07/2019 até 01/07/2020). Apesar da facilidade na utilização do software *EasyFluxPC v.1.0*, a configuração de parâmetros para o processamento dos dados não é flexível e, por trás de sua interface gráfica, utiliza-se uma a versão antiga *EddyPro-5.6* para seu processamento. Contudo, a configuração do arquivo de metadados no *EasyFluxPC* com os equipamentos da *Campbell Scientific* é mais simples, uma vez que o software e o equipamento utilizado serem da mesma empresa.

Desta forma, optou-se pela utilização do arquivo metadados produzido pelo *EasyFluxPC* na configuração do metadados no programa *EddyPro-7.0.6*.

4.3.2 Determinação dos processos para o processamento dos dados

A escolha da utilização do programa *EddyPro-7.0.6* gerou o desafio na determinação de quais operações seriam aplicadas para os dados. Para determinar a forma ideal do processamento dos dados, realizou-se permutações dos diversos métodos listados no programa *EddyPro*, resultando em um total de 20 permutações (Tabela 4). O processo de permutação de processos foi aplicado somente para os dados de alta frequência na Fase 1 (Outubro/2018 até Julho/2019).

A determinação da melhor permutação a ser utilizada para toda série de dados de alta frequência foi baseada no desempenho das permutações. A principal métrica de

desempenho utilizada foi o fechamento do balanço de energia (Equação 3.7), a qual utiliza o fluxo de calor sensível e latente (*EddyPro*) e radiação líquida (CNR4) e fluxo de calor do solo (HFP01). Para isso, utilizou-se a regressão linear, a qual, o intercepto da ordenada é igual a zero e a qualidade do fechamento do balanço de energia é a inclinação desta reta (o fechamento ideal é a reta 1:1). Ainda, foi realizada uma análise de correlação de Pearson utilizando os resultados de fluxo de calor latente e sensível, dados de radiação líquida e fluxo de calor do solo.

4.4 Determinação da área de contribuição da torre de monitoramento

A análise da área de contribuição na torre é essencial para a determinação qualitativa dos fluxos nos intervalos de medição (30 minutos). Para isso, a análise foi realizada com o objetivo de segmentar a série temporal em duas partes: (i) fluxos aceitáveis (vegetação Cerrado *sensu stricto*) e (ii) fluxos rejeitados (outras vegetações). Esta segmentação não necessariamente significa uma contribuição de 100%, porém, uma maioria significativa arbitrária ao usuário.

A análise da área de contribuição da torre (IAB3) foi discretizada em intervalos de 30 minutos, que são os resultados do processamento do programa *EddyPro*. O método por Kljun *et al.* (2015) é utilizado para a determinação da área de contribuição. Para o processamento deste método, utilizou-se o código¹ na linguagem *Python* disponibilizado pelos autores.

Os dados de entrada para o código são: altura de medição (z_m), velocidade do vento médio (u_{mean}), altura da camada limite (h), comprimento de Obukhov (ol), desvio padrão da velocidade do vento (σ_v), velocidade de fricção (u^*) e direção do vento, devendo ser definida qual é a porcentagem da área de footprint a ser calculada. Os principais dados de saída são as coordenadas do limite da linha de contorno da área da respectiva porcentagem definida, sendo que a linha de contorno neste estágio não está inserida em nenhum mapa ou sistema de coordenadas.

O próximo passo então foi de juntar e geolocalizar a estação e as linhas de contorno. Utilizou-se a coleção 5 do mapa de uso do solo do MapBiomas a qual utiliza rasters do Landsat 5, 7 e 8 com auxílio do *Google Earth Engine*, com uma resolução de 30 metros e acurácia de até 82% para o bioma do Cerrado (SOUZA *et al.*, 2020). Adicionalmente, foi verificada a classificação do uso do solo para a área de estudo e foi confirmado que, para o recorte utilizado, a acurácia dos tipos de uso de solo está correta.

Após inseridas as linhas de contorno no mapa de classificação, pôde-se calcular a porcentagem da área dentro do limite do contorno de contribuição e classificar o fluxo

¹ Disponível em <<http://footprint.kljun.net/>>. Acesso em: 19 fev. 2021.

para aquele período.

$$\text{Porcentagem do fluxo aceito (\%)} = \frac{n_{\text{aceitáveis}}}{n_{\text{total}}} \times 100 \quad (4.1)$$

Sendo que, $n_{\text{aceitáveis}}$ é o número de pixels da vegetação Cerrado *sensu stricto* dentro da linha limite e n_{total} é o número total de pixels da linha limite do contorno. A classificação do fluxo entre aceitável ou rejeitado é definido através do grau de aceitação definido pelo usuário. Esta classificação é determinada para toda a série temporal, assim, criando um filtro de área de contribuição.

4.5 Preenchimento de falhas da série de evapotranspiração

As falhas na série de ET podem ser atribuídas a diversos fatores: (i) fora do limite de aceitação; (ii) erro na instrumentação; (iii) filtro de precipitação; (iv) filtro de força do sinal; (v) pouca turbulência (filtro de qualidade); (vi) não estacionariedade (filtro de qualidade) e/ou (vii) devido a área de contribuição (filtro de área de contribuição).

Esses filtros têm como objetivo a eliminação de dados de baixa confiabilidade, e conseqüentemente um aumento significativo de falhas da série é criado. Então, com o objetivo de obter uma série temporal de ET, necessita-se de métodos de preenchimento de falhas que sejam capazes de reproduzir as mesmas características da série original de ET com a menor quantidade de distorções.

4.5.1 Métodos de preenchimento de falhas

A partir dos dados dos sensores instalados na torre de monitoramento (IAB3), foi possível realizar as etapas de processamento dos dados. Na série de ET, os registros considerados inadequados foram selecionados e eliminados, restando somente os dados de melhor qualidade. Além da série de ET de alta qualidade, dados meteorológicos da torre (saldo radiométrico, temperatura, velocidade do vento, umidade relativa, etc) e dados meteorológicos das estações próximas (IAB1 e IAB2) foram incluídos no banco de dados deste trabalho.

Para o método de variação diurna média, utilizou-se três tamanhos de janelas de média móvel diferentes, 7, 11 e 15 dias.

As variáveis utilizadas para a regressão linear múltipla e *Random Forest* foram o saldo de radiação, umidade relativa, temperatura, pressão e vapor de pressão. Nesse caso, utilizou-se apenas dados da torre de monitoramento (IAB3).

Foram utilizadas dois tipos de redes neurais: (i) uma rede neural densa e (ii) uma rede neural recorrente do tipo LSTM. A arquitetura da rede neural densa é de 5 camadas escondidas (Tabela 5) e os dados de entrada foram os mesmos da regressão linear múltipla

e *Random Forest*. Porém, realizou-se um escalonamento dos valores das variáveis para evitar a influência de escala das variáveis nos cálculos dos pesos e viés.

Tabela 5 – Arquitetura da rede neural densa para o preenchimento de falhas

Camada	Nós	Função de ativação
1	1000	ReLU
2	800	ReLU
3	600	ReLU
4	150	ReLU
5	1	Linear

Para a rede de LSTM, os dados de entrada são os mesmos da rede neural densa, porém, existe a possibilidade de haver mais de um intervalo de medição nos dados de entrada. Ou seja, os dados de entrada podem ser entendidos como uma janela em movimento de tamanho especificado pelo usuário. A arquitetura da rede LSTM foi de apenas 3 camadas escondidas, sendo a última camada um nó comum para que houvesse apenas um valor (Tabela 6). Assim como na rede neural densa, os dados de entradas foram escalonados.

Tabela 6 – Arquitetura da rede neural recorrente do tipo LSTM para o preenchimento de falhas

Camada	Células/nós	Função de ativação
1	32	ReLU
2	32	ReLU
3	1	Linear

Por fim, o último método utilizado foi o Penman-Monteith inverso. Foram utilizados dados das torres meteorológicas próximas para o preenchimento de falhas entre a fase 01 e fase 02. Porém, devido às falhas das estações IAB1 e IAB2, ainda existe um período sem preenchimento.

Deve-se observar que os métodos de preenchimentos de falhas não são capazes de preencher 100% as falhas da série de ET. As falhas que não são preenchidas indicam as lacunas para as quais as suposições ou interpolações com dados não foram possíveis. Os métodos de preenchimento de falhas utilizados realizam, apenas, o preenchimento com os dados que foram coletados em campo, com exceção dos métodos que possuem como característica fundamental a interpolação.

4.5.2 Validação das estimativas

Para a validação dos métodos utilizados para o preenchimento de falhas, dividiram-se os dados em treinamento e validação de forma randômica, sendo que 75% dos dados foram para o treinamento e 25% foram para a validação. As métricas utilizadas na validação foram o erro absoluto médio ($\overline{\text{MAE}}$), o erro quadrado médio ($\overline{\text{RMSE}}$), o erro de viés médio

($\overline{\text{MBE}}$) e a correlação de Pearson (ρ). Ainda, foi dividido a análise das métricas para todo o período e somente para o período diurno devido à natureza do ciclo diário de ET.

Após a avaliação das métricas de erros, foram aplicados os métodos de preenchimento de falhas para todos os dados disponíveis, a fim de verificar a normalidade das estimativas. Em seguida, realizou-se a comparação das distribuições dos dados reais com os dados estimados pelos métodos de preenchimento de falhas.

4.5.2.1 Erro médio absoluto

O erro médio absoluto ou $\overline{\text{MAE}}$ (Equação 4.2), avalia a diferença entre os valores originais e os valores estimados em termos absolutos. O x_i representa o valor real e y_i representa o valor estimado pelo método.

$$\overline{\text{MAE}} = \frac{1}{n} \sum_{i=1}^n |x_i - y_i| \quad (4.2)$$

4.5.2.2 Raiz do erro quadrático médio

A raiz do erro quadrático médio ou $\overline{\text{RMSE}}$ (Equação 4.3), avalia a raiz do quadrado da diferença entre os valores originais e os valores estimados. O x_i representa o valor real e y_i representa o valor estimado pelo método.

$$\overline{\text{RMSE}} = \left(\frac{1}{n} \sum_{i=1}^n (x_i - y_i)^2 \right)^{1/2} \quad (4.3)$$

4.5.2.3 Erro de viés médio

O erro de viés médio ou $\overline{\text{MBE}}$ (Equação 4.4), mede o grau de viés médio entre os valores originais e os valores estimados. O x_i representa o valor real e y_i representa o valor estimado pelo método.

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

4.5.2.4 Distribuição das estimativas

Além das métricas $\overline{\text{MAE}}$, $\overline{\text{RMSE}}$ e $\overline{\text{MBE}}$, foi realizada a comparação entre as distribuições dos valores estimados e a distribuição dos valores da série original. Porém, alguns métodos de comparação de distribuições seguem algumas hipóteses que não devem ser quebradas, como a normalidade dos dados. Assim, realizou-se o teste de normalidade para os dados.

4.5.2.4.1 Teste de normalidade

A verificação da distribuição dos dados estimados foi feita através de um teste de normalidade. Para isto, optou-se pelo teste de normalidade de D'Agostino-Pearson (D'AGOSTINO; PEARSON, 1973) capaz de lidar com uma amostragem numerosa ($n > 5000$).

Foi realizada uma divisão da série temporal de ET, partes diurna e noturna, para o teste de normalidade. O objetivo da separação foi incluir no teste apenas os valores que impactariam significativamente na série temporal de ET.

4.5.2.4.2 Análise de variância

Para a comparação da análise de variância do ET original e ET estimados, optou-se por utilizar o teste de Kruskal-Wallis (KRUSKAL; WALLIS, 1952), pois a hipótese de normalidade dos dados foi rejeitada. Trata-se de um teste não paramétrico, o qual não assume que as amostras são distribuídas normalmente. O teste paramétrico equivalente seria o *one-way analysis of variance* ou ANOVA.

A rejeição da hipótese nula significa que existe uma divergência entre as características da distribuição dos grupos.

5 RESULTADOS E DISCUSSÃO

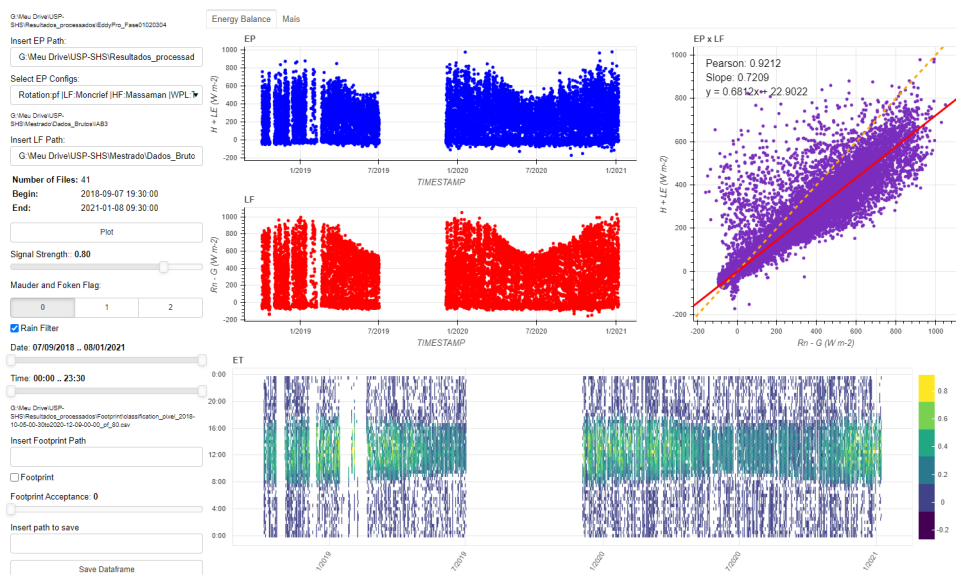
No processamento de dados de alta frequência (20 Hz) foram utilizados dois software: *EasyFlux* e *EddyPro*. Como descrito no capítulo anterior, foi necessário utilizar o arquivo de metadata do *EasyFlux* para utilização do processamento no *EddyPro*. Para a transformação dos dados de baixa frequência para os pós processados foram desenvolvidos códigos computacionais na linguagem *Python* (Apêndice D). Nessa etapa de processamento de dados processados, priorizou-se a maximização na capacidade de visualização dos dados, assim, facilitando a compreensão dos resultados. E alinhado com a visualização, a interatividade com os parâmetros mostrou-se essencial para a compreensão dos resultados e facilitação no processo de verificação de hipóteses.

5.1 Processamento dos dados de alta frequência

5.1.1 Determinação dos processos para o processamento dos dados

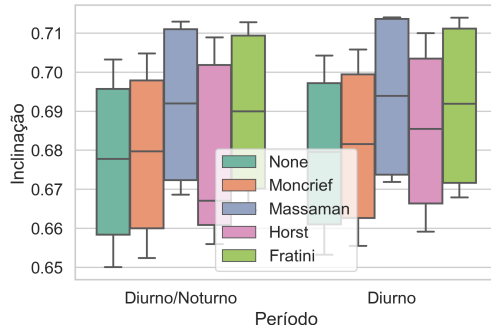
Visando a alta capacidade de visualização, desenvolveu-se um painel de visualização dos dados processados do *EddyPro* (pontos em azul) e dos dados de radiação (pontos em vermelho), sendo possível a interação com os dados e aplicação de filtros de controle de qualidade, precipitação, força de sinal e isolar horários do dia e períodos do ano (Figura 11). Ainda, é possível a visualização do fechamento do balanço energético e correlação (pontos em roxo) e a série de ET.

Figura 11 – Painel de visualização dos dados processados do balanço energético e ET, com dados de entradas e ajuste dinâmico de parâmetros dos filtros



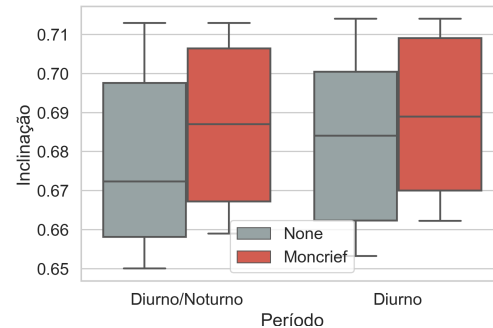
Fonte: Autor

Figura 12 – Boxplot do desempenho do balanço energético (inclinação) sobre a correção espectral de alta frequência



Fonte: Autor

Figura 13 – Boxplot do desempenho do balanço energético (inclinação) sobre a correção espectral de baixa frequência

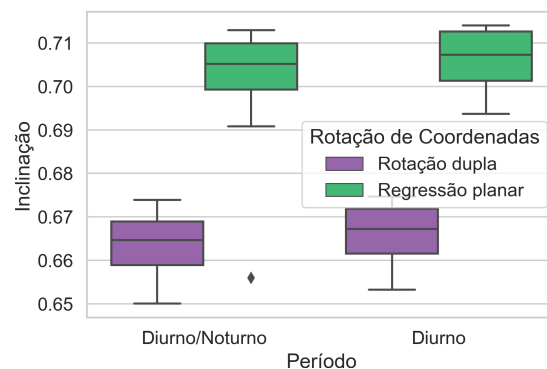


Fonte: Autor

Dentre 20 permutações (Tabela 4), foi possível observar o efeito das correções espectrais de alta e baixa frequência. As correções por Massman (2000) e Fratini e Mauder (2014) foram as correções de alta frequência que mais aprimoraram o balanço energético (Figura 12). Em relação à utilização da correção de baixa frequência, o uso da correção por Moncrieff *et al.* (1997) sempre aprimora o fechamento do balanço (Figura 13).

Do mesmo modo, a regressão planar na rotação de coordenadas apresentou melhores resultados quando comparado com a rotação 2D (Figura 14).

Figura 14 – Boxplot do desempenho do balanço energético (inclinação) sobre a rotação de coordenadas

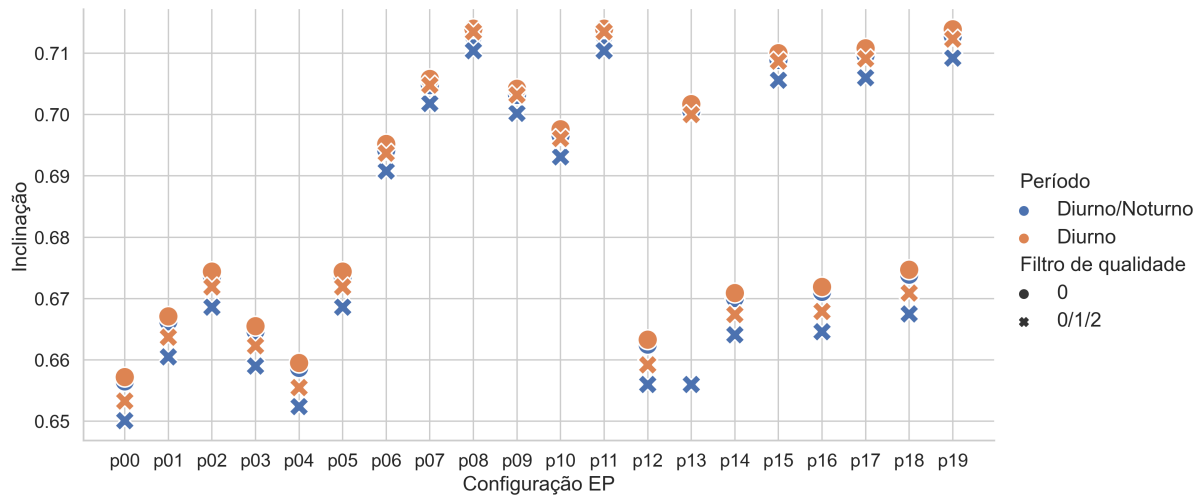


Fonte: Autor

A distinção de horários mostrou ter relevância, pois grande parte dos dados de radiação possui valores próximo de zero, devido ao ciclo diurno e noturno. Estes valores acabam afetando a reta de inclinação do balanço de energia de maneira significativa mesmo não contribuindo significativamente para o saldo total de energia ou evapotranspiração. No fim, a inclinação do balanço de energia sempre melhorou quando se analisa somente no período diurno (Figura 12, 13, 14 e 15).

Ainda, pode-se verificar que a aplicação do filtro Mauder e Foken (2011), o qual se inclui apenas o fluxo considerado ótimo, melhorou o fechamento energético para todas as permutações (Figura 15).

Figura 15 – Desempenho do fechamento do balanço energético (inclinação) para todas as configurações testadas



Fonte: Autor

A permutação $p08$ e $p11$ (Tabela 4) geraram os melhores resultados em relação a inclinação (Figura 15), a qual consistem na utilização da regressão planar como forma de rotação de coordenadas; Massman (2000) para correção espectral de alta frequência, e no caso do $p08$, Moncrieff *et al.* (1997) para correção espectral de baixa frequência.

5.1.2 Filtro dos dados processados

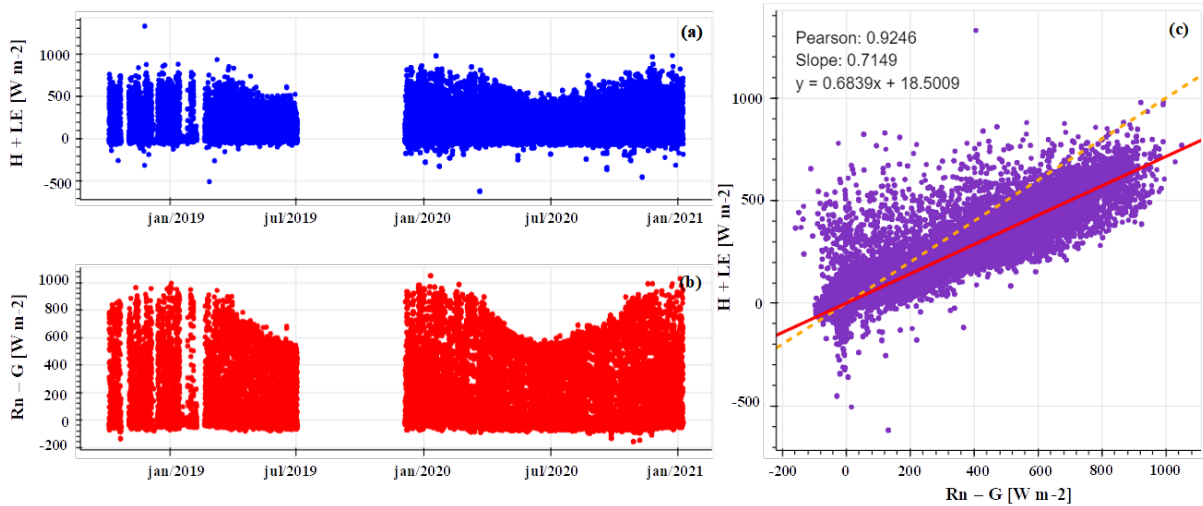
A figura 16a é toda a série temporal do balanço energético com os dados de saída do *EddyPro*, enquanto a figura 16b é toda a série temporal dos dados de radiação do CNR4 e HFP01. Por fim, a figura 16c é a nuvem de pontos, sendo que a reta tracejada amarela é a reta 1:1 e a reta vermelha é a reta de melhor ajuste para a inclinação. A ausência dos filtros revela que a série contém diversos *outliers*.

Após a aplicação dos filtros de precipitação e de qualidade por Mauder e Foken (2011) é observado que os pontos *outliers* são retirados e a inclinação aumenta para 0,7208 (Figura 17).

A aplicação desse filtro de precipitação resulta em uma perda de dados de aproximadamente 5% (Figura 18). O critério utilizado foi a presença de precipitação ou não.

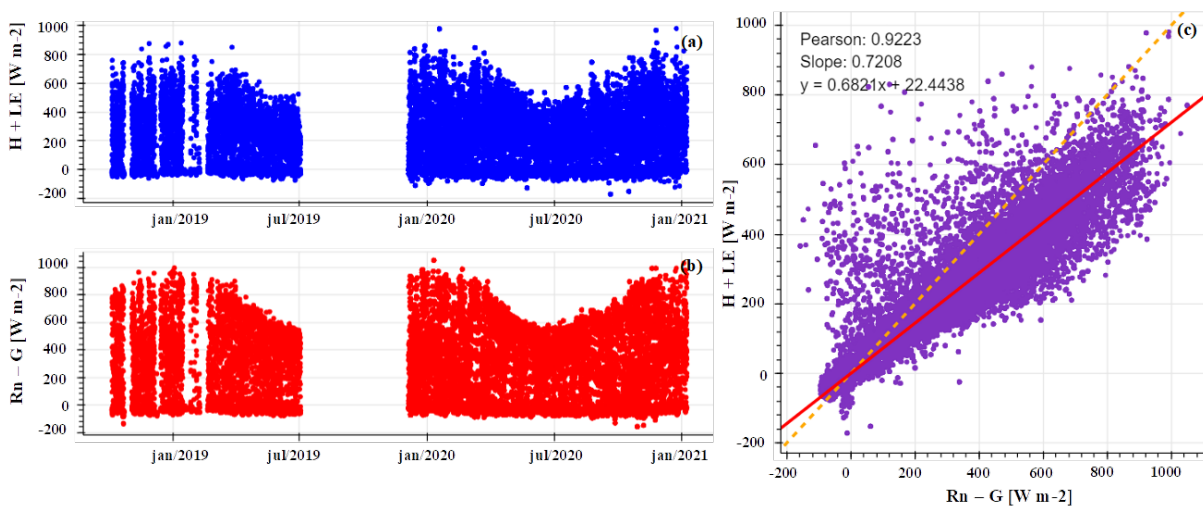
O filtro de qualidade por Mauder e Foken (2011) resultou em uma perda de dados de 40% (Figura 19). Nota-se que existe um padrão de perda de dados a qual é relacionada

Figura 16 – Série temporal do balanço energético de melhor desempenho (p08), sendo que (a) representa a série do fluxo de calor sensível e latente, (b) o saldo radiométrico e (c) a nuvem de pontos com as métricas de correlação e fechamento de balanço energético



Fonte: Autor

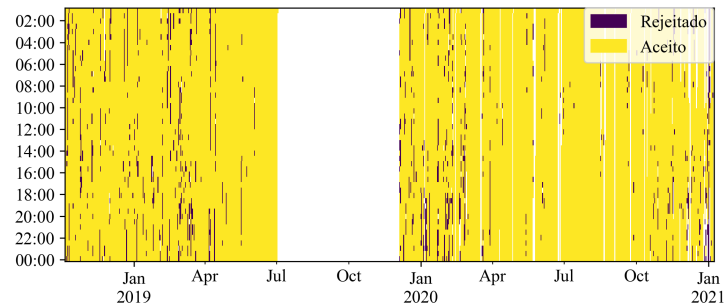
Figura 17 – Série temporal do balanço energético de melhor desempenho (p08) com aplicação do filtro de precipitação e de qualidade, sendo que (a) representa a série do fluxo de calor sensível e latente, (b) o saldo radiométrico e (c) a nuvem de pontos com as métricas de correlação e fechamento de balanço energético



Fonte: Autor

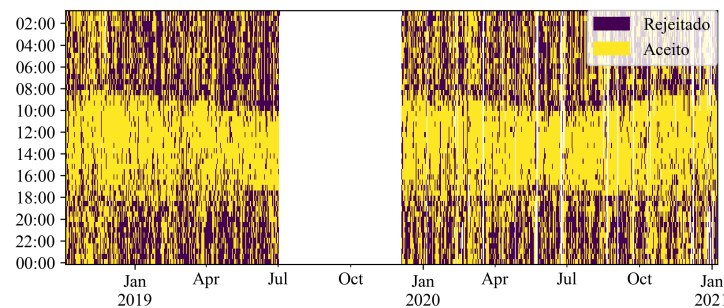
com o horário do dia, ou seja, a verificação das características integrais de turbulência rejeita grande parte dos dados noturnos devido a um menor fluxo turbulento.

Figura 18 – Série temporal binária com o filtro de precipitação para a estação IAB3



Fonte: Autor

Figura 19 – Série temporal binária com o filtro de qualidade para estação IAB3



Fonte: Autor

5.2 Análise da área de contribuição

A etapa de análise da área de contribuição envolveu duas partes principais: o mapa de classificação de uso do solo e o cálculo da área de contribuição. Desta forma, pôde-se classificar se determinado fluxo é correspondente ou não à área de interesse.

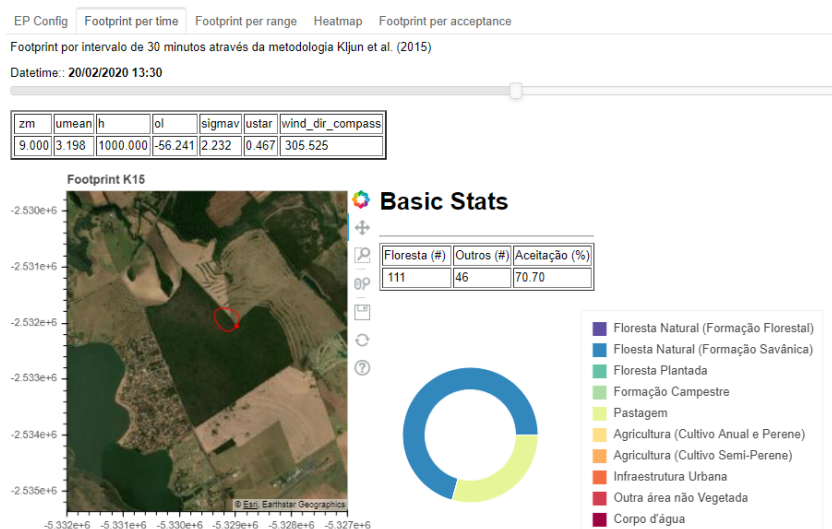
Para análise da área de contribuição, criou-se uma interface gráfica na linguagem *Python* (Apêndice D) para visualizar o *Footprint* por intervalo de tempo (Figura 20) e por período do tempo (Figura 22).

5.2.1 Mapa de classificação do uso do solo

A classificação dos diferentes tipos de uso do solo é essencial para que se realize o filtro dos dados dos fluxos. O mapa de aceitação é o responsável para dar o contexto da linha de contorno da área de contribuição. Para o recorte utilizado, o mapa de uso do solo identificou 10 diferentes usos do solo (Tabela 7), a qual o tipo de cobertura de Floresta Natural é a tipo de cobertura de interesse (Figura 21b).

Após identificados o tipo de cobertura de interesse, criou-se um filtro binário para que fossem separados os pixels em "aceito" e "rejeitado". E por fim, foi criado um mapa de aceitação do uso do solo para a área do entorno do IAB 3 (Figura 21c).

Figura 20 – Interface gráfica para análise da área de contribuição, utilizando o método por Kljun et al. (2015), por intervalo de tempo de 30 minutos



Fonte: Autor

Tabela 7 – Tipos de cobertura e aceitação (MapBiomass)

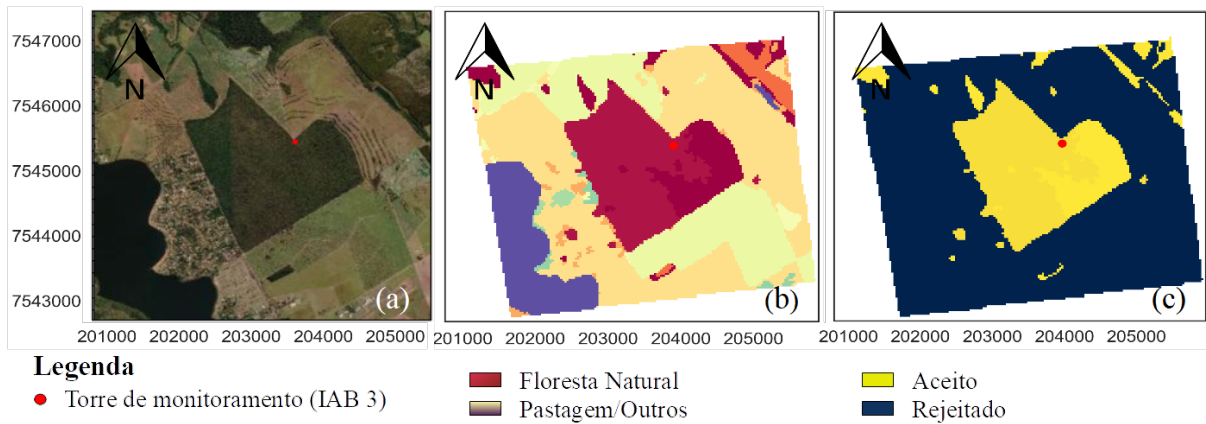
Tipo de cobertura	Aceitação
Floresta Natural/Formação Florestal	Sim
Floresta Natural/Formação Savânica	Sim
Floresta Plantada	Não
Formação Campestra/Outros	Não
Pastagem	Não
Agricultura/Cultivo Anual e Perene	Não
Agricultura/Cultivo Semi-Perene	Não
Infraestrutura Urbana	Não
Outra área não Vegetada	Não
Corpo d'água	Não

O programa desenvolvido utilizou o código¹ desenvolvido em *Python* a qual foi integrado ao código da interface gráfica (Figura 22). A imagem do mapa gerado possui a projeção chamada *Web Mercator*, porém, ela só foi utilizada para a visualização do mapa. O cálculo dos pixels é realizado no sistema de coordenadas geográfica SIRGAS 2000 UTM 23 S. A interface gráfica tem a capacidade de apresentar de uma maneira rápida a situação da área de contribuição.

O programa possibilita selecionar uma faixa de dias do período disponível e até uma faixa de horas do dia. Pode-se visualizar a área de contribuição, a direção do vento e a quantidade de pixels para cada tipo de vegetação.

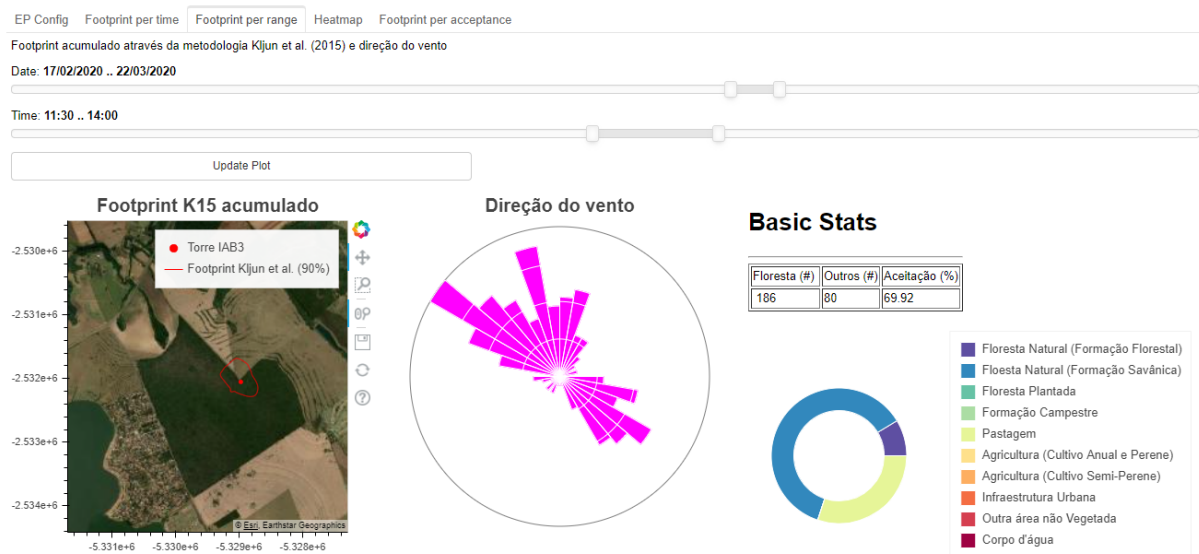
¹ Disponível em <<http://footprint.kljun.net/>>. Acesso em: 19 fev. 2021.

Figura 21 – Criação do mapa de aceitação para a torre de monitoramento IAB3: (a) imagem de satélite; (b) recorte de mapa de uso do solo e; (c) mapa de aceitação



Fonte: Autor

Figura 22 – Interface gráfica para análise da área de contribuição, utilizando o método por Kljun et al. (2015), por faixa de tempo determinada pelo usuário



Fonte: Autor

5.2.2 Filtro dos dados com uso da análise de contribuição

Esta característica heterogênea do IAB3 é encontrada em outros campos da rede global (FLUXNET e AmeriFlux) e, a filtragem por área de contribuição pode beneficiar na representatividade dos campos (CHU *et al.*, 2021).

O filtro dos dados tem como característica a redução de dados para o aperfeiçoamento da série, então como forma de visualização dos dados aceitos, seis graus de aceitação do fluxo acumulado foram plotados (Figura 23). As linhas de contorno acumuladas são criadas a partir dos acumulados de pesos de toda a série temporal, ou seja, a densidade do

grau de influência é significativamente maior nas regiões mais próximas quando comparado com as regiões mais distantes.

Figura 23 – Progressão do grau de aceitação da área de contribuição, utilizando o método por Kljun et al. (2015), acumulada da torre de monitoramento IAB3 no período de 2019 até 2020



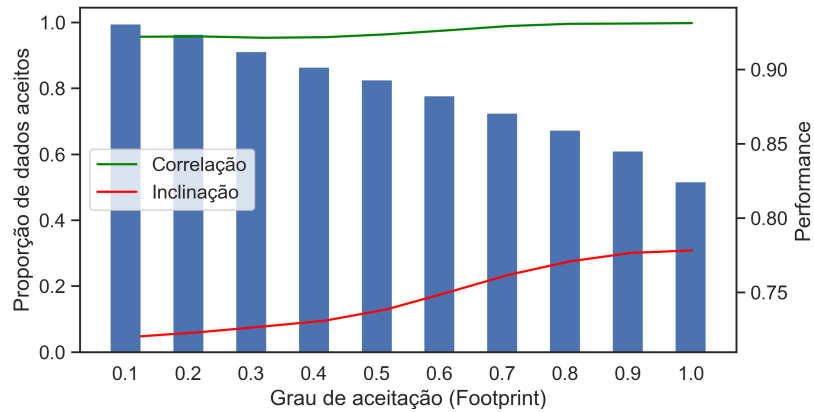
Fonte: Autor

Uma das hipóteses simplificadoras na utilização da área de contribuição é de que após determinada a linha de contorno da contribuição (90%), supõe-se que as contribuições dos pixels são iguais em termos de peso. Assim, pode-se calcular mais facilmente a porcentagem de aceitação de um fluxo (Equação 4.1). Esta simplificação acaba superestimando a porcentagem de fluxos rejeitados, ou seja, fluxos que seriam aceitos acabam sendo desconsiderados. Para minimizar a rejeição indevida de fluxos, o grau de aceitação deve ser ajustado para que haja um equilíbrio entre fluxos da vegetação vizinha e fluxos que acabam sendo marginalizados pela simplificação.

O parâmetro que foi utilizado para ajustar o grau de aceitação foi a métrica de inclinação, sendo que se considerou a porcentagem de dados rejeitados (Figura 24). Observa-se que com o aumento do grau de aceitação existe uma redução de até 50% dos dados aceitos. A redução da quantidade de dados aceitos aprimorou a métrica de inclinação (fechamento do balanço energético) para até 0,7781, sugerindo que essa filtragem impacta

positivamente para o fechamento do balanço energético. Enquanto a correlação de Pearson sugere um leve aumento com o aumento do grau de aceitação.

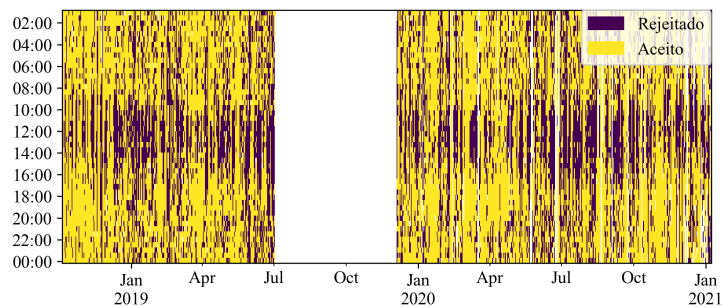
Figura 24 – Desempenho do fechamento do balanço energético (inclinação) e correlação de Pearson em relação ao grau de aceitação da área de contribuição para a torre de monitoramento IAB3



Fonte: Autor

Então, apesar da melhor métrica de inclinação indicar que um grau de aceitação de 100% produz o melhor resultado em termos de métrica, o ganho de 0,0074 de inclinação em relação a um grau de aceitação de 80% não justifica a quantidade de dados perdidos. Logo, uma filtragem com um grau de aceitação de 80% foi considerado como uma taxa ótima em termos de métrica e em termos de retenção da quantidade de dados. Nessas configurações, a filtragem por área de contribuição resulta em uma perda de dados de aproximadamente 35% (Figura 25). Comparando com outros trabalhos realizados no Cerrado *sensu stricto*, a capacidade do fechamento do balanço energético para essa configuração é compatível (GIAMBELLUCA *et al.*, 2009; Da Rocha *et al.*, 2009; CABRAL *et al.*, 2015).

Figura 25 – Série temporal binária com o filtro de área de contribuição para a estação IAB3

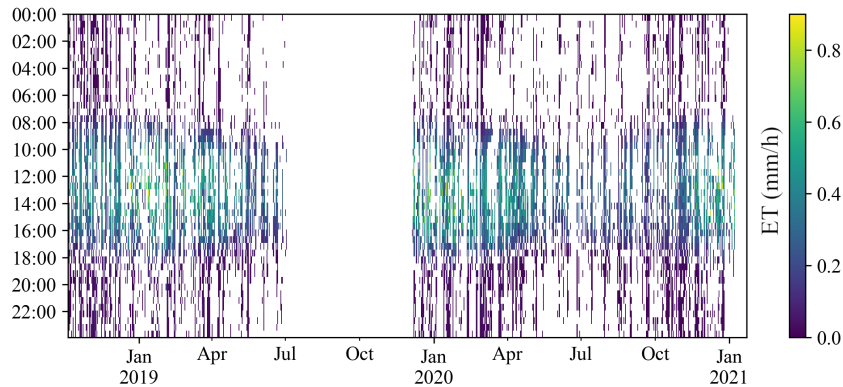


Fonte: Autor

5.3 Preenchimento de falhas

O processo de filtragem dos dados por precipitação, qualidade, força de sinal e área de contribuição resultou em uma perda de dados de 45% (Figura 26).

Figura 26 – Série de Evapotranspiração com aplicação de filtros para a estação IAB3



Fonte: Autor

A figura 27 mostra a progressão dos dados da série da ET. Os dados, após processados pelo programa *EddyPro*, estão misturados em dados de baixa confiabilidade e boa confiabilidade, então a filtragem deve ser realizada. O exemplo retrata um recorte de um mês da série de ET, que depois da filtragem dos dados se torna necessário a utilização de métodos de preenchimento de falhas para a consolidação de uma série. Foram utilizados seis métodos de preenchimento de falhas: (i) variação diurna média (MDV), (ii) regressão linear múltipla, (iii) *Random Forest*, (iv) rede neural densa, (v) rede neural recorrente (LSTM) e (vi) Penman-Monteith inversa.

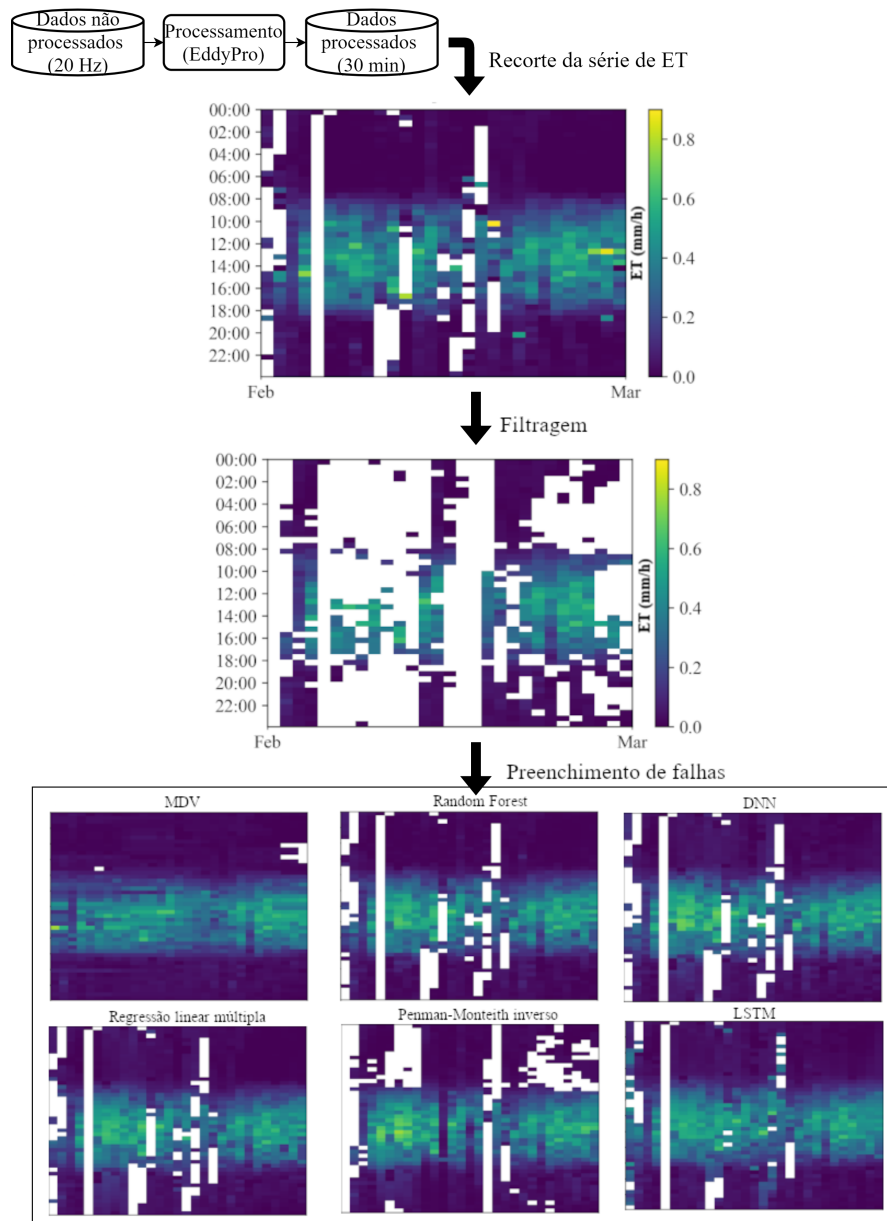
Os métodos de preenchimento de falhas foram somente utilizados nos casos em que houvesse informação mínima para o seu preenchimento. Caso não existisse o patamar mínimo, o valor não era preenchido. Informações em relação a quantidade de falhas após o preenchimento encontram-se no apêndice B e a tabela final de dados da ET com os métodos de preenchimento de falhas no apêndice C.

5.3.1 Métodos de preenchimento de falhas

5.3.1.1 Variação Diurna Média

O método de variação diurna média foi utilizado para três janelas diferentes (7, 11 e 15 dias). Por este método utilizar a média da ET aceitos, só foi possível o preenchimento dos dados nos dias que houve dados adjacentes no mesmo horário. Após o preenchimento da série da ET, a média da série para o ano de 2019 para o período chuvoso foi de $4,29 \text{ mm/dia}$ e para o período seco foi de $3,09 \text{ mm/dia}$ e com um acumulado de 711 mm , sendo que não existe valores de ET em parte do ano. Para o ano de 2020, a média da série

Figura 27 – Esquema da progressão da série de ET na filtragem dos dados e preenchimento de falhas



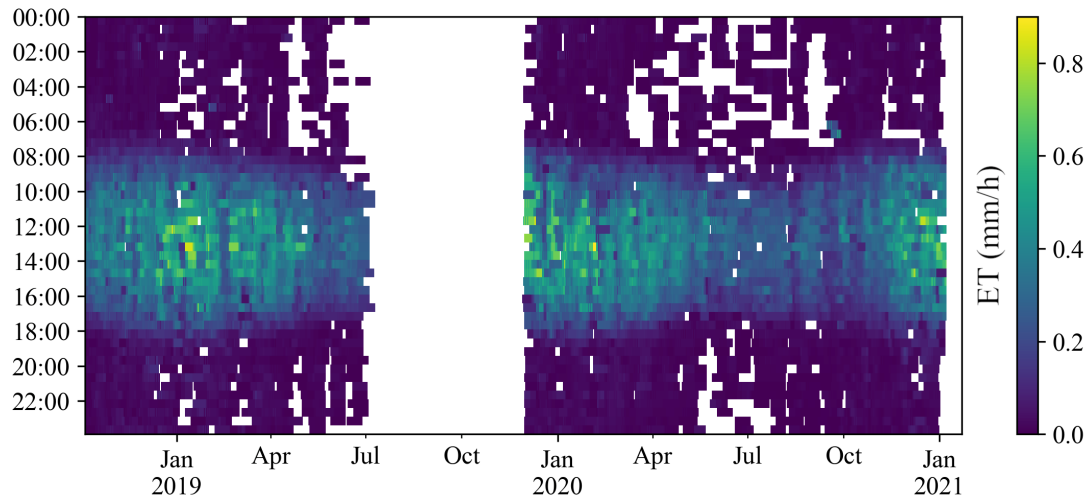
Fonte: Autor

do período chuvoso foi de 3,58 mm/dia , enquanto para o período seco foi de 2,56 mm/dia e um acumulado de 1027 mm .

5.3.1.2 Regressão Linear Múltipla

O método de regressão linear possuiu oito parâmetros: radiação líquida ($R_n [W m^2]$), umidade relativa ($RH [\%]$), déficit de vapor de pressão ($VPD [Pa]$), temperatura ($T [^{\circ}C]$), pressão ($P [Pa]$), fluxo de calor no solo ($G [W m^2]$), vapor de pressão ($e [Pa]$) e velocidade do vento ($u [m/s]$). A equação 5.1 é a regressão linear múltipla ajustada usando todos os

Figura 28 – Evapotranspiração com o método de preenchimento de falhas de Variação Diurna Média



Fonte: Autor

dados.

$$ET = 4,68 \times 10^{-4}R_n - 5,01 \times 10^{-3}RH + 1,77 \times 10^{-4}VPD - 3,93 \times 10^{-2}T - 1,60 \times 10^{-5}P - 7,46 \times 10^{-3}G + 3,81 \times 10^{-4}e + 2,92 \times 10^{-3}u + 12,67 \quad (5.1)$$

Após o preenchimento da série da ET, a média da série para o ano de 2019 para o período chuvoso foi de $3,90 \text{ mm/dia}$ e para o período seco foi de $2,67 \text{ mm/dia}$. O acumulado para o ano de 2019 foi de 697 mm . Para o ano de 2020, a média da série do período chuvoso foi de $3,86 \text{ mm/dia}$ e para o período seco foi de $2,56 \text{ mm/dia}$ e o acumulado de 2020 foi de 1058 mm .

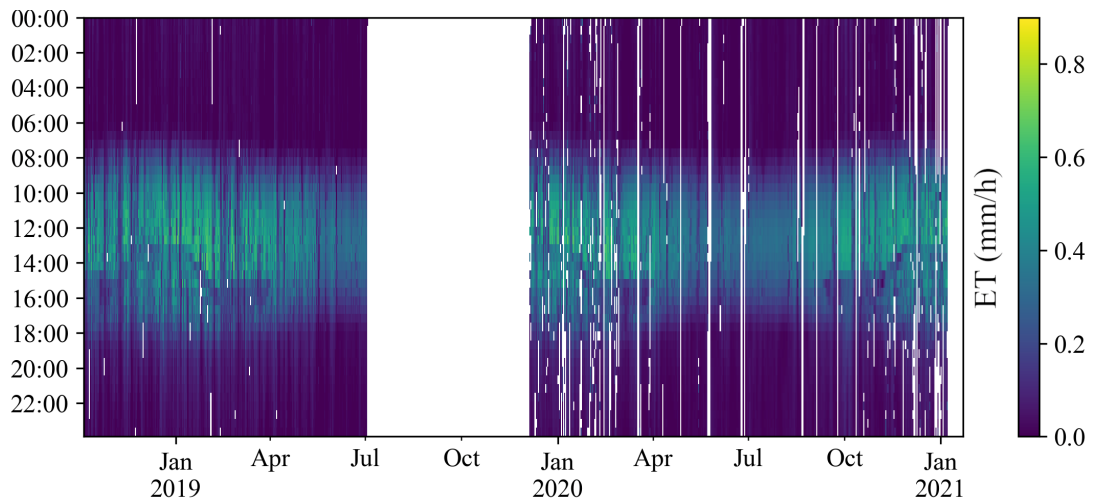
5.3.1.3 Método de Regressão Logística (*Random Forest*)

Assim como o método de regressão linear múltipla, foram utilizadas as mesmas variáveis para a criação das árvores de decisão.

A média da série da ET para o período chuvoso do ano de 2019 foi de $3,75 \text{ mm/dia}$ e para o período seco foi de $2,63 \text{ mm/dia}$ e o acumulado foi de 675 mm . Enquanto para o ano de 2020, o período chuvoso foi de $3,67 \text{ mm/dia}$ e para o período seco $2,46 \text{ mm/dia}$ e o acumulado foi de 1011 mm .

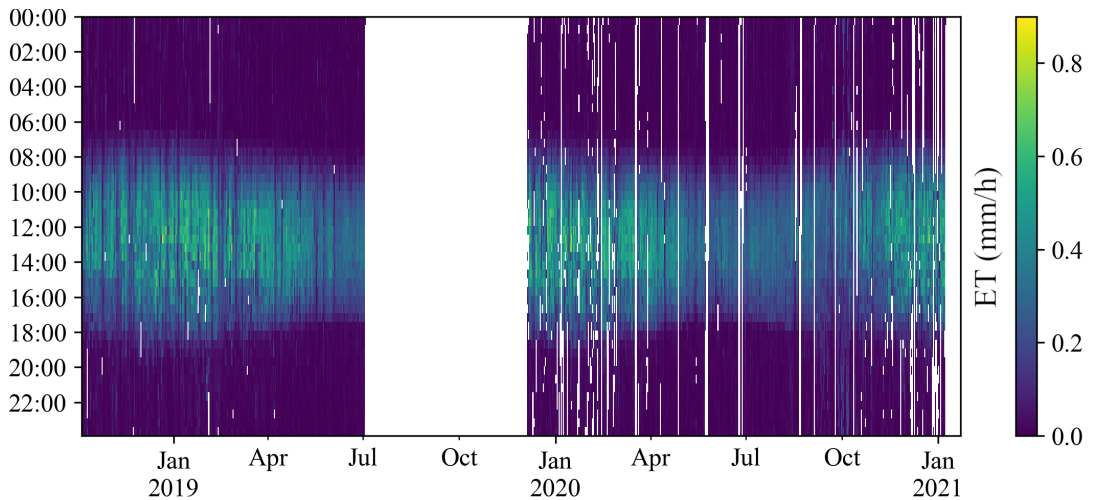
E a ordem de importância das variáveis, segundo o *Random Forest*, foi em primeiro a radiação líquida, em seguida temperatura, velocidade do vento, pressão do ar, vapor de pressão, fluxo de calor do solo, umidade relativa e déficit de vapor de pressão.

Figura 29 – Evapotranspiração com o método de preenchimento de falhas de Regressão Linear Múltipla



Fonte: Autor

Figura 30 – Evapotranspiração com o método de preenchimento de falhas de *Random Forest*

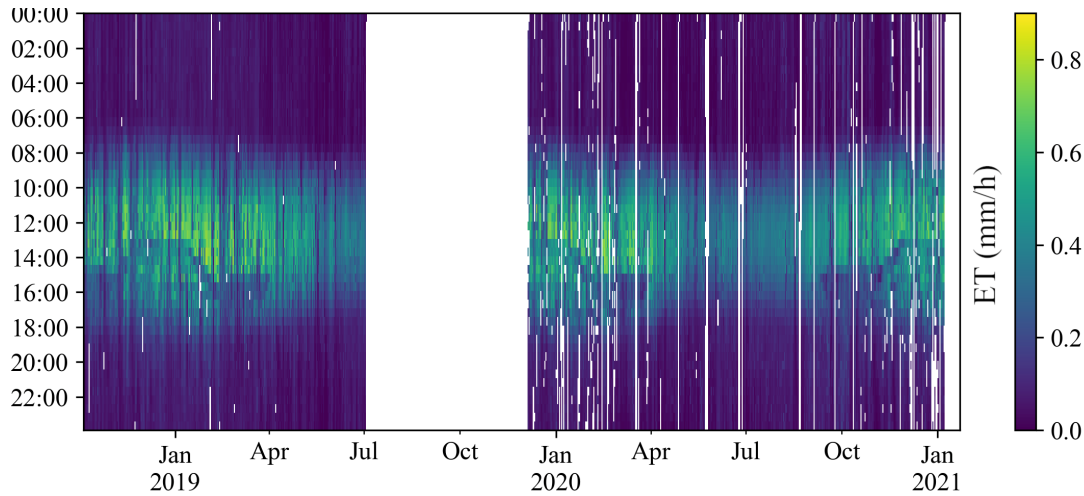


Fonte: Autor

5.3.1.4 Redes neurais densa

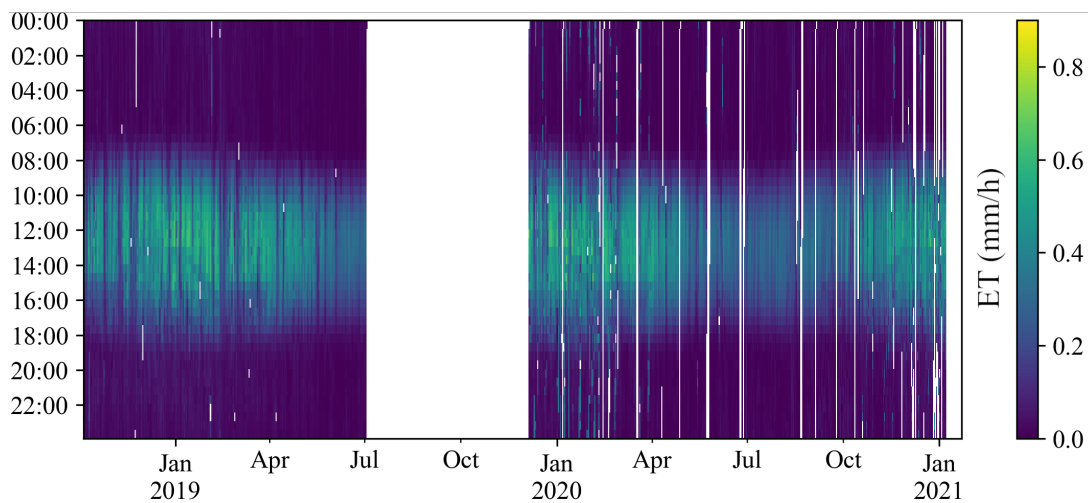
O método de redes neurais densa utilizou as mesmas variáveis que foram utilizadas para o regressão linear múltipla e possui uma arquitetura de rede conforme a tabela 5. Após o preenchimento de falhas da ET, a média da série de ET para o período chuvoso do ano de 2019 foi de $4,47 \text{ mm}/\text{dia}$ e para o período seco foi de $3,24 \text{ mm}/\text{dia}$ e um acumulado foi de 815 mm . Enquanto para o ano de 2020, para o período chuvoso foi de $4,23 \text{ mm}/\text{dia}$ e para o período seco foi de $3,02 \text{ mm}/\text{dia}$ e um acumulado de 1198 mm .

Figura 31 – Evapotranspiração com o método de preenchimento de falhas de redes neurais densa



Fonte: Autor

Figura 32 – Evapotranspiração com o método de preenchimento de falhas de redes neurais recorrente (LSTM)



Fonte: Autor

5.3.1.5 Redes neurais recorrentes (LSTM)

Assim como o método utilizado nas redes neurais densa, foram utilizadas as mesmas variáveis de entrada e com uma arquitetura de rede descrito na tabela 6. A diferença dos dados de entrada foi de que ao invés de um único valor de cada variável, utilizou-se um vetor de valores para cada variável. Este vetor possui um comprimento de quatro horas, ou seja, oito intervalos de medição.

A média da série da ET para o período chuvoso do ano de 2019 foi de $4,13 \text{ mm/dia}$ e para o período seco foi de $2,85 \text{ mm/dia}$ e um acumulado de 639 mm . Para o período chuvoso para o ano de 2020 foi de $4,08 \text{ mm/dia}$ e para o período seco foi de $2,61 \text{ mm/dia}$

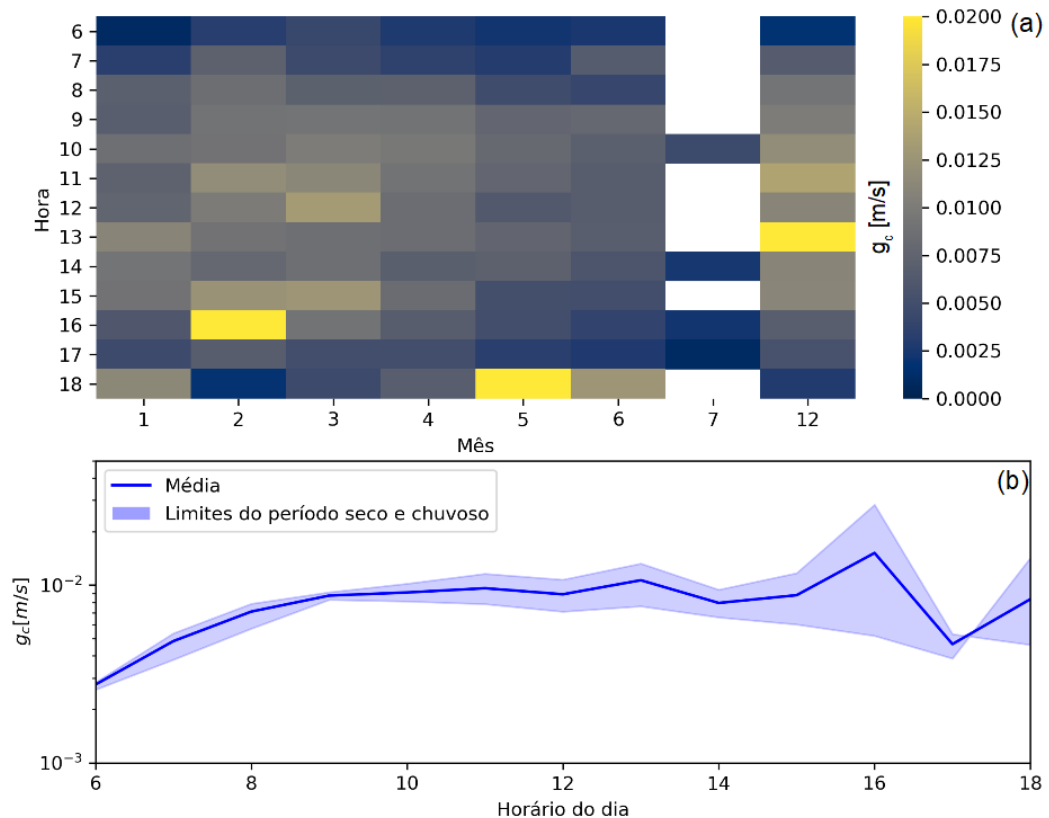
e um acumulado de 992 *mm*.

5.3.1.6 Penman-Monteith inverso

O método de preenchimento de falhas baseado na equação de Penman-Monteith é um método que requer o cálculo da condutância superficial e aerodinâmica. O valor utilizado para o cálculo da condutância aerodinâmica foi o valor médio de 0,05 *m/s*, baseado no cálculo da equação 3.18.

O cálculo da condutância superficial utilizou a equação de Penman-Monteith inversa (Equação 3.17). O ano de 2019 foi um ano anômalo que houve a grande falha entre a fase 01 e fase 02, então, perdeu-se grande parte do período seco e o começo do período chuvoso. Apesar disso, a condutância média da vegetação Cerrado *sensu stricto* se manteve em torno de 2×10^{-3} *m/s* até 2×10^{-2} *m/s* (Figura 33). Deve-se destacar que o horário do dia em que ocorre o pico da ET ocorre entre as 10:00 até as 15:00. Nos horários de menor ET existe um aumento da variabilidade da condutância superficial, o efeito da menor quantidade de dados aceitos pelos filtros faz com que valores mais erráticos apareçam.

Figura 33 – Condutância superficial para a vegetação Cerrado *sensu stricto* para o ano de 2019 por: (a) média mensal e hora do dia e (b) média do período seco e chuvoso

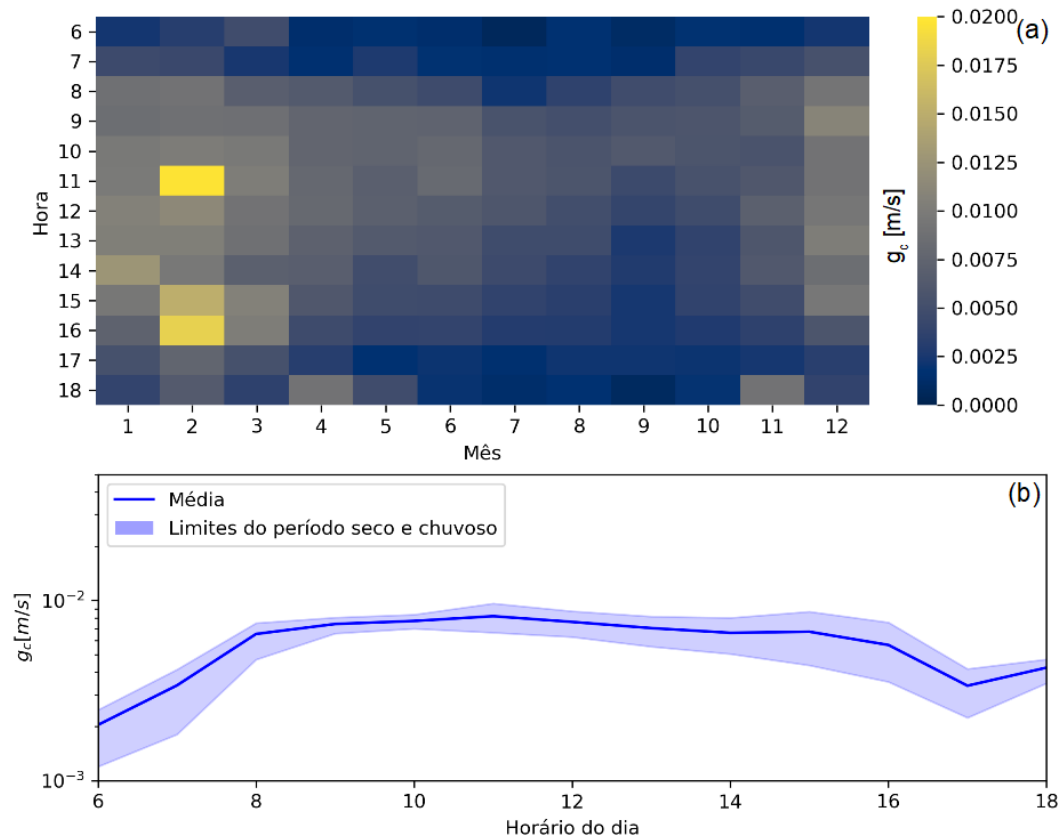


Fonte: Autor

O ano de 2020 foi um ano completo em relação a quantidade de dados, havendo

apenas falhas pontuais. A condutância média da vegetação Cerrado *sensu stricto* variou entre $2 \times 10^{-3} \text{ m/s}$ até $1 \times 10^{-2} \text{ m/s}$ (Figura 34).

Figura 34 – Condutância superficial para a vegetação Cerrado *sensu stricto* para o ano de 2020 por: (a) média mensal e hora do dia e (b) média do período seco e chuvoso



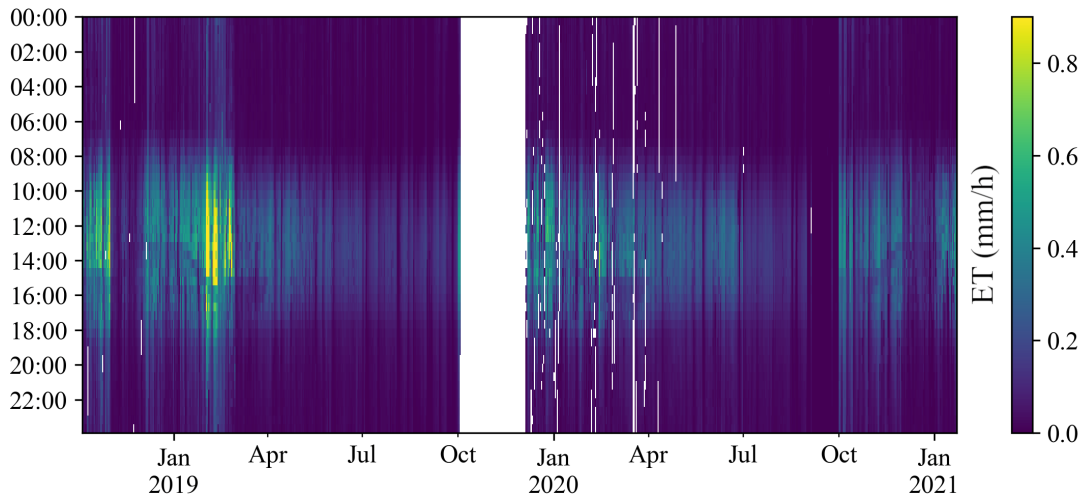
Fonte: Autor

Resultados similares, em termo de média da condutância superficial, foram encontrados em estudos no Cerrado *sensu stricto* (GIAMBELLUCA *et al.*, 2009; CABRAL *et al.*, 2015). O padrão diário da condutância superficial é semelhante com o observado no estudo realizado por Cabral *et al.* (2015), para o qual existe uma variação dos valores da condutância ao longo do dia e por período seco e chuvoso.

Após definidas as condutâncias, a equação de Penman-Monteith fica em função de variáveis meteorológicas, que foram preenchidas com os dados da estação IAB3. Em casos em que não foi possível o preenchimento dos dados do IAB3, foram utilizados os dados das estações meteorológicas IAB1 e IAB2. A utilização dos dados de duas estações próximas fez com que fosse possível o preenchimento de falhas para os meses de julho, agosto e setembro de 2019. A quantidade de dados meteorológicos nas estações do IAB se encontra no apêndice A.

A média da série da ET para o período chuvoso do ano de 2019 foi de 3.98 mm/dia

Figura 35 – Evapotranspiração com o método de preenchimento de falhas de Penman-Monteith inverso



Fonte: Autor

e para o período seco foi de $1.58 \text{ mm}/\text{dia}$ e um acumulado de 805 mm . Para o ano de 2020, o período chuvoso teve média de $2.98 \text{ mm}/\text{dia}$ e para o período seco foi de $1.87 \text{ mm}/\text{dia}$ e um acumulado de 990 mm .

5.3.2 Avaliação dos métodos de preenchimento de falhas

5.3.2.1 Variação diurna média

Os erros entre as estimativas de todo o período e de somente o período diurno apresentaram a influência dos dados noturnos (Tabela 8). A escala dos valores diurnos é significativamente superior em relação ao dos dados noturnos. Ainda, o coeficiente de correlação de Pearson apresentou uma piora em relação aos dados do dia inteiro (Figura 36).

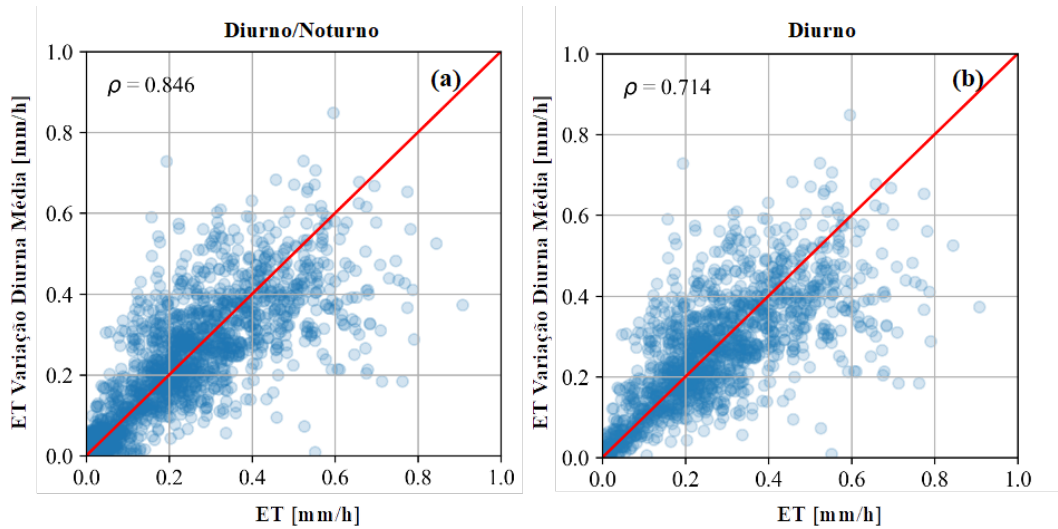
Tabela 8 – Métricas da avaliação do método de variação diurna média

Métrica	Diurno/Noturno <i>mm/h</i>	Diurno <i>mm/h</i>
$\overline{\text{MAE}}$	0,0585	0,0816
$\overline{\text{RMSE}}$	0,0923	0,1165
$\overline{\text{MBE}}$	-0,0013	0,0079

O alto valor de $\overline{\text{RMSE}}$ em relação ao $\overline{\text{MAE}}$ mostrou que existem valores com significativa diferença entre os valores estimados e os reais. Contudo, não existe uma significativa tendência de subestimação ou sobrestimação.

O teste de normalidade aplicado para todos os dados preenchidos para o período diurno foi rejeitado, ou seja, a distribuição não é normal. Então, realizou-se um teste não-paramétrico de Kruskal-Wallis para a análise de variância entre a série da ET (real)

Figura 36 – Correlação da validação do método de variação diurna média



Fonte: Autor

e a série da ET preenchida (estimada). Para um intervalo de confiança de 99%, o teste aceitou a hipótese nula ($p_{value} = 0,03$) a qual considera os grupos de mesma população.

Este método de preenchimento de falhas, essencialmente, é um método de interpolação a qual depende exclusivamente dos valores da ET. O método é dependente da janela dos dias adjacentes, sendo que por muitas vezes só existem valores aceitáveis quando a janela é expandida. Outros estudos que utilizaram este método apresentaram resultados aceitáveis, contudo, são apenas utilizados quando não existem dados meteorológicos aceitáveis (FALGE *et al.*, 2001; PAPALE; VALENTINI, 2003; MOFFAT *et al.*, 2007; AUBINET; VESALA; PAPALE, 2012).

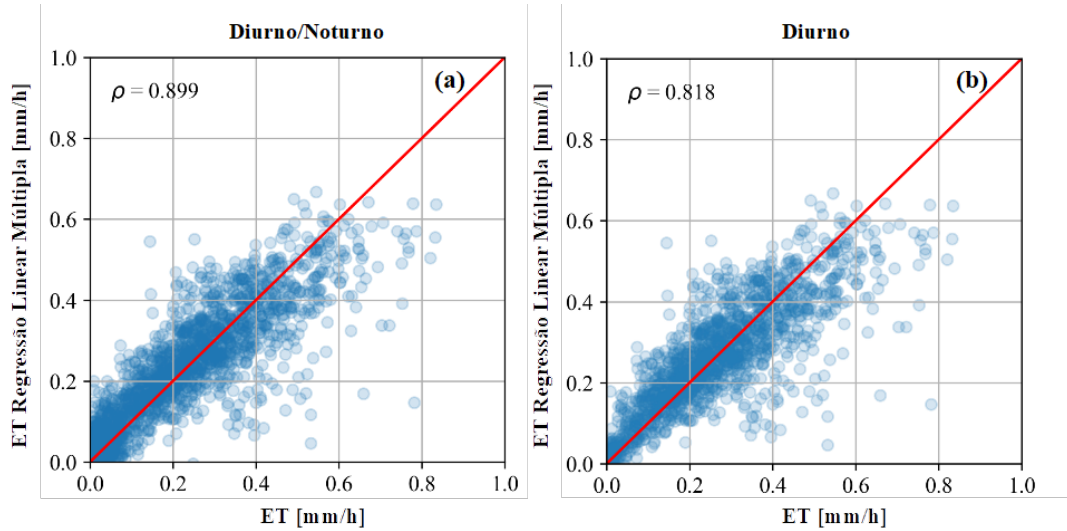
5.3.2.2 Regressão linear múltipla

A correlação de Pearson pode ser visualizada para todo o período ou somente o período diurno (Figura 37). Os erros entre as estimativas com todo o dia e com somente o período diurna apresentaram a mesma influência dos dados noturnos (Tabela 9). O valor inferior do \overline{MBE} indica a presença de alguns valores subestimados por este método. Contudo, esse método é um dos três melhores métodos de acordo com as métricas de \overline{MAE} e \overline{RMSE} .

Tabela 9 – Métricas da avaliação do método de regressão linear múltipla

Métrica	Diurno/Noturno <i>mm/h</i>	Diurno <i>mm/h</i>
\overline{MAE}	0,0494	0,0641
\overline{RMSE}	0,0750	0,0914
\overline{MBE}	-0,0006	-0,0033

Figura 37 – Correlação da validação do método de regressão linear múltipla



Fonte: Autor

O teste de normalidade rejeitou a hipótese nula. Assim, aplicou-se um teste não paramétrico de Kruskal-Wallis e foi determinado que a hipótese nula foi aceita ($p_{values} = 0,39$). Este método de preenchimento de falhas é o melhor capacitado para solução de problemas de natureza linear. Desta forma, este método é melhor representada quando as características meteorológicas são lineares aos valores da ET.

5.3.2.3 Random Forest

O desempenho dos erros (Tabela 10) e a correlação de Pearson (Figura 38) entre as estimativas para todo o período e somente para o período diurno são consistentes como os outros métodos.

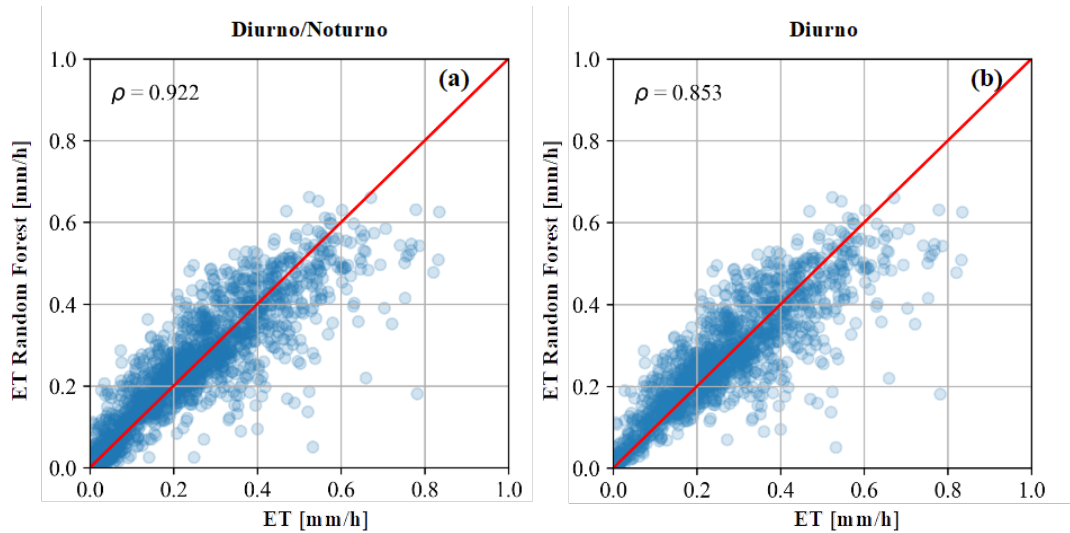
Tabela 10 – Métricas da avaliação do método de Random Forest

Métrica	Diurno/Noturno <i>mm/h</i>	Diurno <i>mm/h</i>
\overline{MAE}	0,0401	0,0577
\overline{RMSE}	0,0663	0,0826
\overline{MBE}	-0,0011	0,0007

Apesar da semelhança com o método de regressão linear múltipla, a densidade de pontos do método de *Random Forest* próxima a reta 1:1 é superior ao método anterior. Porém, existem alguns *outlier* subestimando os valores reais. O teste de normalidade foi aplicado para todos os dados disponíveis e foi rejeitada a hipótese nula para o período diurno. Contudo, no teste de Kruskal-Wallis a hipótese nula foi aceita ($p_{value} = 0,15$).

O desempenho das métricas deste método vai de acordo com outro estudo realizado para fluxo de metano, para o qual considerou-se esse como o melhor método em relação

Figura 38 – Correlação da validação do método de Random Forest



Fonte: Autor

a acurácia, independentemente do tamanho da falha (KIM *et al.*, 2020). A criação de diversas árvores de decisão torna este método resiliente para o preenchimento de falhas. Diferentemente da regressão linear múltipla, este método é melhor adaptado quando existe um aumento significativo na quantidade de dados. Porém, o desafio deste método é a incapacidade de extrapolação dos dados, atividade para qual ele não foi treinado.

5.3.2.4 Rede neural densa

Os erros entre as estimativas para todo o período e para o período diurno é observado na tabela 11. Enquanto a correlação de Pearson é observada na figura 39.

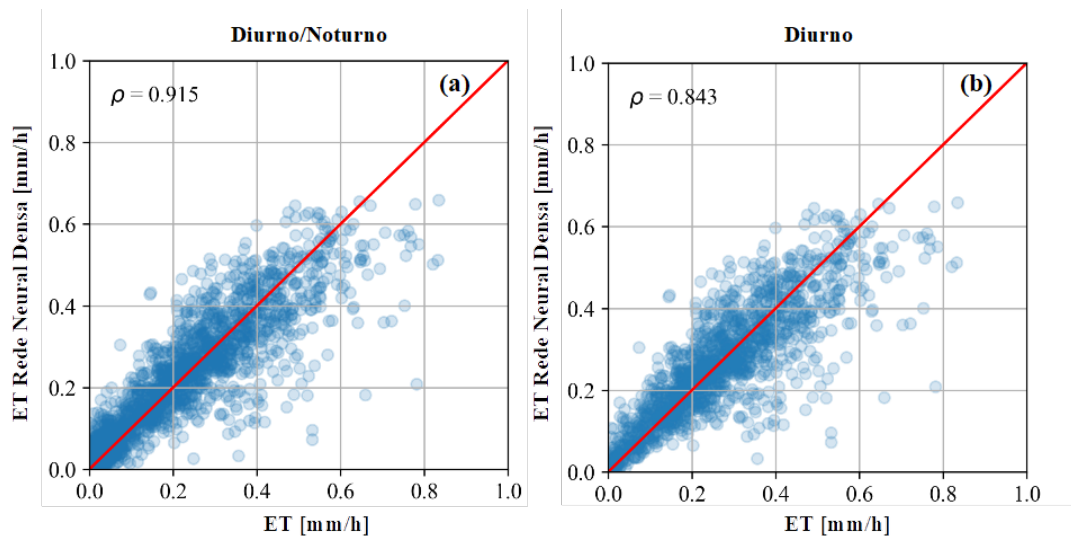
Tabela 11 – Métricas da avaliação do método de rede neural densa

Métrica	Diurno/Noturno <i>mm/h</i>	Diurno <i>mm/h</i>
$\overline{\text{MAE}}$	0,0433	0,0589
$\overline{\text{RMSE}}$	0,0693	0,0856
$\overline{\text{MBE}}$	0,0005	0,0056

Muito semelhante aos dois métodos anteriores, existe alguns *outlier* subestimando os valores reais. Contudo, este método é um dos melhores métodos em relação as métricas $\overline{\text{MAE}}$ e $\overline{\text{RMSE}}$. A hipótese nula foi rejeitada no teste de normalidade. E no teste de Kruskal-Wallis a hipótese nula também foi rejeitada.

Os métodos de redes neurais têm a capacidade de modificar a arquitetura da rede para um melhor encaixe, assim como a quantidade de repetições (*epochs*). Porém, este melhor encaixe pode resultar em um possível *overfit* da rede neural. Este método é considerado ideal na resolução de problemas onde existam relações não lineares, sendo que

Figura 39 – Correlação da validação do método de Rede neural densa



Fonte: Autor

a não linearidade das redes é atingida através das funções de ativação. Considerando que a relação da ET e variáveis meteorológicas nem sempre são lineares, este método é capaz de realizar um preenchimento de falhas desde que os dados sejam de boa qualidade.

Existem diversos estudos que utilizam este método devido ao seu desempenho nas métricas de erro. Desta forma, esse método é considerado padrão para o preenchimento de falhas do fluxo de carbono e energia no banco de dados do FLUXNET (PAPALE; VALENTINI, 2003; MOFFAT *et al.*, 2007; AUBINET; VESALA; PAPALE, 2012).

5.3.2.5 Rede neural recorrente (LSTM)

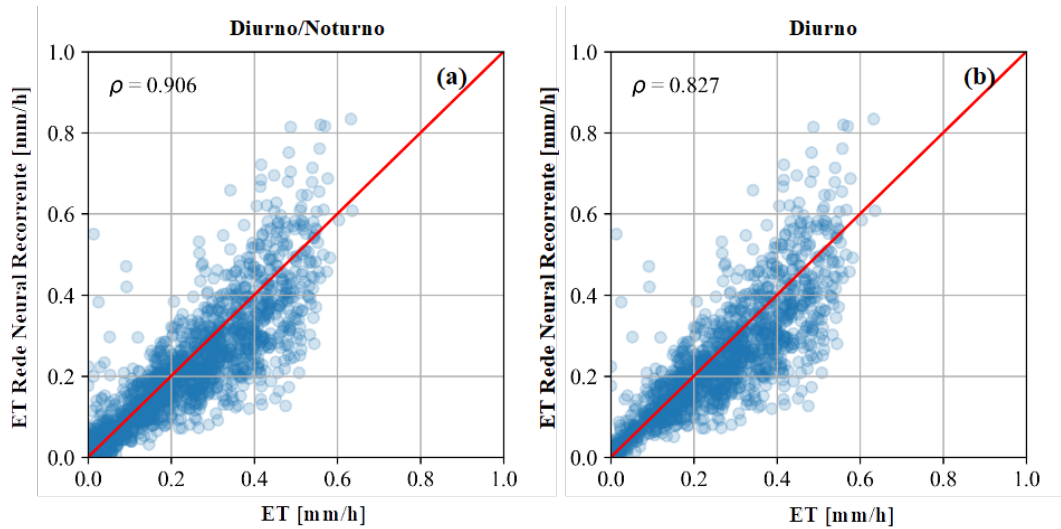
A correlação de Pearson (Figura 40) e as métricas de erros (Tabela 12) entre as estimativas de todo o período e somente para o período diurno são consistentes com os métodos anteriores.

Tabela 12 – Métricas da avaliação do método de rede neural recorrente

Métrica	Diurno/Noturno <i>mm/h</i>	Diurno <i>mm/h</i>
MAE	0,0494	0,0641
$\overline{\text{RMSE}}$	0,0750	0,0914
MBE	-0,0006	-0,0033

Esta rede neural com a célula LSTM teve performance do $\overline{\text{RMSE}}$ e $\overline{\text{MAE}}$ parecidas com as anteriores, porém, existe uma leve tendência a subestimação dos valores estimados. A hipótese nula foi rejeitada após utilizado o teste de normalidade e o teste de Kruskal-Wallis.

Figura 40 – Correlação da validação do método de Rede neural recorrente (LSTM)



Fonte: Autor

Definitivamente a configuração da rede para este teste não foi a ideal, contudo, com uma simples rede neural com nós LSTM demonstrou uma boa capacidade de preenchimento de falhas. Outra alteração que poderia ser realizada é no tamanho da janela escolhida, porém, deve-se atentar que a janela escolhida não poderá conter falhas dentro dela. A janela escolhida para este estudo foi de quatro horas, mas um maior intervalo de tempo poderia ser utilizado, ao custo de tempo de processamento e quantidade de dados de entrada aceitável.

5.3.2.6 Penman-Monteith inverso

Os erros entre estimativas são observados na tabela 13. O $\overline{\text{MAE}}$ e $\overline{\text{RMSE}}$ para a série considerando o período diurno e noturno foi considerado aceitável.

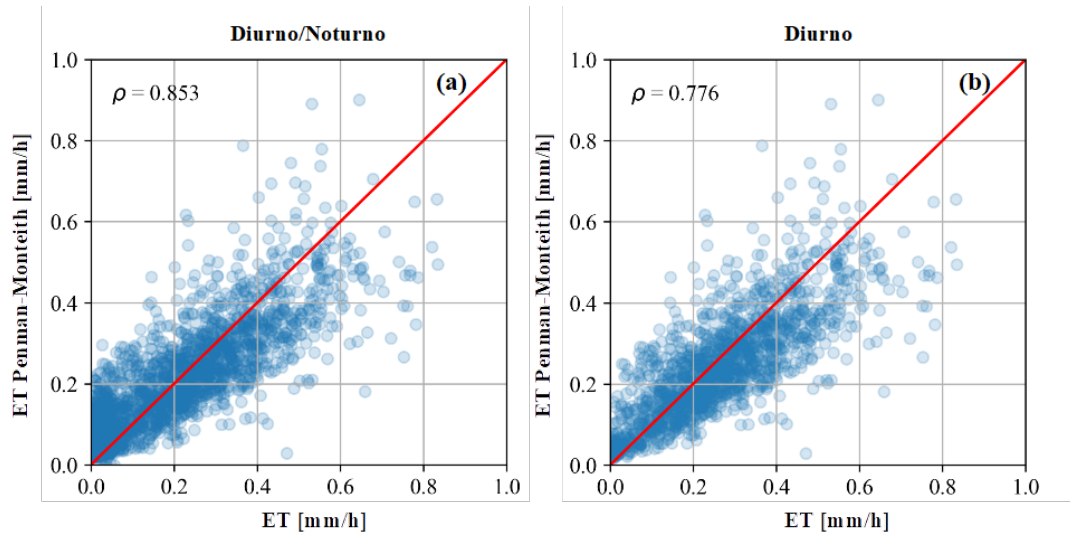
Tabela 13 – Métricas da avaliação do método de Penman-Monteith inverso

Métrica	Diurno/Noturno <i>mm/h</i>	Diurno <i>mm/h</i>
$\overline{\text{MAE}}$	0,0657	0,0753
$\overline{\text{RMSE}}$	0,0899	0,1024
$\overline{\text{MBE}}$	-0,0055	0,0203

Assim como os demais métodos, a hipótese nula do teste de normalidade foi rejeitada. E no teste de Kruskal-Wallis a hipótese nula foi rejeitada.

Na avaliação dessas métricas, utilizou-se apenas os dados da torre de monitoramento do IAB3. Porém, o objetivo principal deste método foi de utilizar um método físico capaz de utilizar dados meteorológicos de estações próximas.

Figura 41 – Correlação da validação do método de Penman-Monteith)



Fonte: Autor

5.3.3 Métricas de avaliação

Os resultados de cada métrica de avaliação de todos os métodos de preenchimento de falhas utilizados foram organizados em forma de tabela, para uma melhor comparação do desempenho de cada um dos métodos (Tabela 14). A tabela foi dividida em dados do dia inteiro (Diurno/Noturno) e dados somente diurnos. A partir desses resultados, observa-se uma significativa diferença das médias quando considerada a parte noturna, sendo que esta parte possui pouca influência no acumulado de ET.

Tabela 14 – Média e desvio padrão das séries preenchidas da ET

Método de preenchimento	Diurno/Noturno		Diurno	
	Média (mm/h)	Desvio padrão	Média (mm/h)	Desvio padrão
Variação diurna média	0.1423	0.1630	0.2464	0.1613
Regressão linear múltipla	0.1384	0.1560	0.2431	0.1583
<i>Random Forest</i>	0.1336	0.1587	0.2421	0.1595
Rede neural densa	0.1587	0.1602	0.2655	0.1631
Rede neural recorrente	0.1272	0.1597	0.2338	0.1612
Penman-Monteith inverso	0.1254	0.1474	0.2104	0.1615

Na tabela 15 são apresentados os resultados de erro médio absoluto (\overline{MAE}) para todos os métodos de preenchimento de falhas. Foi apresentado o dado em mm/h e em termos de porcentagem em relação a média do período. Deve-se observar a diferença que faz a segmentação das métricas, a qual os dados do período noturno impactam significativamente na análise do erro. Em termos de mm/h, o método de *Random Forest* foi o melhor. Contudo, a porcentagem indica que a rede neural densa apresentou o menor erro.

Tabela 15 – Erro médio absoluto ($\overline{\text{MAE}}$) entre os métodos de preenchimento de falhas

Métrica	Método de preenchimento	Diurno/Noturno		Diurno	
		<i>mm/h</i>	Porcentagem da média	<i>mm/h</i>	Porcentagem da média
Erro médio absoluto ($\overline{\text{MAE}}$)	Variação diurna média	0.0585	41%	0.0816	33%
	Regressão linear múltipla	0.0494	35%	0.0641	26%
	<i>Random Forest</i>	0.0401	30%	0.0577	23%
	Rede neural densa	0.0433	27%	0.0589	22%
	Rede neural recorrente	0.0449	35%	0.0635	27%
	Penman-Monteith inverso	0.0657	52%	0.0753	32%

A tabela 16 apresenta a raiz quadrada do erro médio ($\overline{\text{RMSE}}$) para todos os métodos de preenchimento de falhas. Igualmente o $\overline{\text{MAE}}$, o impacto do período noturno é significativo, sendo que os dois melhores métodos são a rede neural densa e o *Random Forest*.

Tabela 16 – Raiz quadrada do erro médio ($\overline{\text{RMSE}}$) entre os métodos de preenchimento de falhas

Métrica	Método de preenchimento	Diurno/Noturno		Diurno	
		<i>mm/h</i>	Porcentagem da média	<i>mm/h</i>	Porcentagem da média
Raiz quadrada do erro médio ($\overline{\text{RMSE}}$)	Variação diurna média	0.0923	64%	0.1165	47%
	Regressão linear múltipla	0.0750	54%	0.0914	37%
	<i>Random Forest</i>	0.0663	49%	0.0826	34%
	Rede neural densa	0.0693	43%	0.0856	32%
	Rede neural recorrente	0.0704	55%	0.0887	37%
	Penman-Monteith inverso	0.0899	71%	0.1024	48%

A tabela 17 apresenta o erro de viés médio ($\overline{\text{MBE}}$) para todos os métodos de preenchimento de falhas. Os valores de erros desta métrica mostram que existe um viés baixo para os métodos de variação diurna média, regressão linear múltipla, *Random Forest* e rede neural densa. Existe uma subestimação do método de rede neural recorrente e uma sobrestimação do método de Penman-Monteith inverso, quando analisado somente para o período diurno.

A tabela 18 resume as métricas de correlação de Pearson para todos os métodos de preenchimento de falhas. Observa-se o impacto dos baixos valores da porção noturna da série da ET para todos os métodos. Para o período diurno, o método de preenchimento de falhas *Random Forest* apresentou o melhor resultado.

Tabela 17 – Erro de viés médio ($\overline{\text{MBE}}$) entre os métodos de preenchimento de falhas

Métrica	Método de preenchimento	Diurno/Noturno		Diurno	
		<i>mm/h</i>	Porcentagem da média	<i>mm/h</i>	Porcentagem da média
Erro de viés médio ($\overline{\text{MBE}}$)	Variação diurna média	-0.0013	-0,91%	0.0079	3,20%
	Regressão linear múltipla	-0.0006	-0,43%	-0.0033	-1,35%
	<i>Random Forest</i>	-0.0011	-0,82%	0.0007	0,28%
	Rede neural densa	0.0005	0,31%	0.0056	2,10%
	Rede neural recorrente	-0.0093	-7,31%	-0.0021	-9,06%
	Penman-Monteith inverso	-0.0055	-4,38%	0.0203	9,64%

Tabela 18 – Resultados do coeficiente de correlação de Pearson (ρ) entre os métodos de preenchimento de falhas

Métrica	Método de preenchimento	Diurno/Noturno	Diurno
Correlação de Pearson (ρ)	Variação diurna média	0.846	0.714
	Regressão linear múltipla	0.899	0.818
	<i>Random Forest</i>	0.922	0.853
	Rede neural densa	0.915	0.843
	Rede neural recorrente	0.906	0.827
	Penman-Monteith inverso	0.853	0.776

6 CONCLUSÃO

Este trabalho teve como objetivo aprimorar o conhecimento quantitativo da evapotranspiração através da investigação da qualidade dos dados coletados pela torre de monitoramento no IAB e propor uma rotina automática de controle de qualidade desses dados.

O processamento de dados de alta frequência (20 Hz) foi o primeiro desafio para a determinação da série, devido aos diversos programas estabelecidos e diversos conjuntos de etapas procedimentais e correções que afetam o resultado final. Para o presente trabalho, determinou-se como ótimo a utilização do programa *EddyPro*, desenvolvido pela LICOR, sendo considerado o melhor em termos de desempenho e utilização dentro do meio acadêmico.

O critério escolhido para definição do conjunto de técnicas a serem utilizadas no processamento dos dados, utilizando o *EddyPro*, foi o fechamento do balanço energético. Determinou-se que para a área de estudo IAB3, o método de regressão planar como forma de rotação de coordenadas do anemômetro. Além disso, o método por Moncrieff *et al.* (1997) para correção espectral de baixa frequência e Massman (2000) para correção espectral de alta frequência, e a utilização dos termos por WPL, são o conjunto de métodos que devem ser aplicados para a determinação da série de evapotranspiração.

Para a determinação da série de evapotranspiração, houve uma ênfase na qualidade dos dados utilizados, sendo que os critérios de qualidade dos dados foram definidos e utilizados como filtros. Os filtros utilizados mais relevantes foram o de chuva, teste de estado estacionário, verificação das características integrais de turbulência e a criação do filtro por área de contribuição. O filtro de área de contribuição serviu para dar contexto na localidade da torre do IAB3, sendo que a torre do IAB3 está localizada no em um remanescente da vegetação Cerrado *sensu stricto* próximo a outros tipos de vegetações. Através do filtro da área de contribuição, a série de evapotranspiração foi aprimorado em termos do fechamento do balanço de energia.

A utilização de diversos filtros na série de evapotranspiração reduziu a quantidade de dados de forma significativa em troca do aumento da qualidade da série. Então, foram utilizadas seis técnicas de preenchimento de falhas para que fosse possível usar a série completa sem a perda de características fundamentais. A série preenchida com o método *Random Forest* forneceu o melhor resultado em termos das métricas com um período chuvoso variando de 3,67–3,75 *mm/dia* e um período seco variando de 2,46–2,63 *mm/dia* e um acumulado de 1011 *mm* para o ano de 2020. Em valores diários e anuais, os resultados apresentaram um intervalo de evapotranspiração compatível com estudos utilizando o

método *Eddy Covariance* para a vegetação Cerrado *sensu stricto*.

Todos os procedimentos de análise dos dados gerados pelo programa *EddyPro*, análise da área de contribuição e preenchimento de falhas foram realizados utilizando ferramentas computacionais a qual permitiram a fácil interação com os dados. Esta interatividade teve como objetivo o aprimoramento pessoal em relação aos comportamentos de cada processo, assim como um legado para a utilização e manutenção da torre de monitoramento do IAB3.

Em futuros projetos, a integração dos dados da torre do IAB3 com dados de sensoriamento remoto auxiliará no aprimoramento dos modelos hidrológicos, sendo que a continuidade da coleta de dados é um importante passo para a contribuição na rede global de fluxos (FLUXNET).

REFERÊNCIAS

ALLEN, R. G. *et al.* **Crop evapotranspiration: guidelines for computing crop water requirements**. Rome: FAO Irrigation and Drainage Paper 56, 1998. 300 p. ISBN 978-92-5-104219-9.

ALLEN, R. G. *et al.* A recommendation on standardized surface resistance for hourly calculation of reference ETo by the FAO56 Penman-Monteith method. **Agricultural Water Management**, v. 81, n. 1-2, p. 1–22, mar 2006. ISSN 03783774. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S037837740500154X>.

ANACHE, J. A. *et al.* Land use and climate change impacts on runoff and soil erosion at the hillslope scale in the Brazilian Cerrado. **Science of the Total Environment**, Elsevier B.V., v. 622-623, p. 140–151, 2018. ISSN 18791026. Disponível em: <https://doi.org/10.1016/j.scitotenv.2017.11.257>.

ANACHE, J. A. *et al.* Hydrological trade-offs due to different land covers and land uses in the Brazilian Cerrado. **Hydrology and Earth System Sciences**, v. 23, n. 3, p. 1263–1279, 2019. ISSN 16077938.

ARYA, S. **Introduction to Micrometeorology**. Second edi. [*S.l.: s.n.*]: Academic Press, 2001. 420 p. ISBN 0120593548.

AUBINET, M.; VESALA, T.; PAPALE, D. **Eddy Covariance**. 1º. ed. Dordrecht: Springer Netherlands, 2012. 438 p. ISBN 978-94-007-2350-4. Disponível em: <http://link.springer.com/10.1007/978-94-007-2351-1>.

BALDOCCHI, D. TURNER REVIEW No. 15. 'Breathing' of the terrestrial biosphere: Lessons learned from a global network of carbon dioxide flux measurement systems. **Australian Journal of Botany**, v. 56, n. 1, p. 1–26, 2008. ISSN 00671924.

BALDOCCHI, D. **ESPM 228 - Lecture 4 Micrometeorological Flux Measurements, Eddy Covariance, Application, Part 2**. 2019. 50 p.

BREIMAN, L. Random forests. **Machine Learning**, v. 45, p. 5–32, 2001.

BURBA, G. **Eddy Covariance Method for Scientific, Industrial, Agricultural, and Regulatory Applications**. Lincoln: LI-COR Biosciences, 2013. 331 p. ISBN 9780615768274.

CABRAL, O. M. *et al.* Water and energy fluxes from a woodland savanna (cerrado) in southeast Brazil. **Journal of Hydrology: Regional Studies**, v. 4, p. 22–40, sep 2015. ISSN 22145818. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S2214581815000440>.

CABRERA, M. C. M. *et al.* Performance of evaporation estimation methods compared with standard 20 m² tank. **Revista Brasileira de Engenharia Agrícola e Ambiental**, v. 20, n. 10, p. 874–879, oct 2016. ISSN 1415-4366. Disponível em: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1415-43662016001000874&lng=en&tlng=en.

CHU, H. *et al.* Representativeness of Eddy-Covariance flux footprints for areas surrounding AmeriFlux sites. **Agricultural and Forest Meteorology**, v. 301-302, n. February, 2021. ISSN 01681923.

Cruz Ruggiero, P. G. *et al.* Soil-vegetation relationships in cerrado (Brazilian savanna) and semideciduous forest, Southeastern Brazil. **Plant Ecology**, v. 160, n. 1, p. 1–16, 2002. ISSN 13850237.

Da Rocha, H. R. *et al.* Patterns of water and heat flux across a biome gradient from tropical forest to savanna in Brazil. **Journal of Geophysical Research: Biogeosciences**, v. 114, n. 1, p. 1–8, 2009. ISSN 01480227.

D'AGOSTINO, R.; PEARSON, E. S. Tests for Departure from Normality. Empirical Results for the Distributions of b_2 and b_1 . **Biometrika**, v. 60, n. 3, p. 613, dec 1973. ISSN 00063444. Disponível em: <https://www.jstor.org/stable/2335012?origin=crossref>.

FALGE, E. *et al.* Gap filling strategies for defensible annual sums of net ecosystem exchange. **Agricultural and Forest Meteorology**, v. 107, n. 1, p. 43–69, mar 2001. ISSN 01681923. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0168192300002252>.

FOKEN, T. **Micrometeorology**. Second. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008. 362 p. ISSN 1871756X. ISBN 978-3-540-74665-2. Disponível em: <http://link.springer.com/10.1007/978-3-540-74666-9>.

FOKEN, T.; WICHURA, B. Tools for quality assessment of surface-based flux measurements. **Agricultural and Forest Meteorology**, v. 78, n. 1-2, p. 83–105, jan 1996. ISSN 01681923. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0168192395022481>.

FRATINI, G. *et al.* Relative humidity effects on water vapour fluxes measured with closed-path eddy-covariance systems with short sampling lines. **Agricultural and Forest Meteorology**, Elsevier, v. 165, p. 53–63, nov 2012. ISSN 01681923. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0168192312001955?via%3Dihubhttps://linkinghub.elsevier.com/retrieve/pii/S0168192312001955>.

FRATINI, G.; MAUDER, M. Towards a consistent eddy-covariance processing: an intercomparison of EddyPro and TK3. **Atmospheric Measurement Techniques**, v. 7, n. 7, p. 2273–2281, jul 2014. ISSN 1867-8548. Disponível em: <https://amt.copernicus.org/articles/7/2273/2014/>.

GARCIA-MONTIEL, D. C. *et al.* Estimating Seasonal Changes in Volumetric Soil Water Content at Landscape Scales in a Savanna Ecosystem Using Two-Dimensional Resistivity Profiling. **Earth Interactions**, v. 12, n. 2, p. 1–25, mar 2008. ISSN 1087-3562. Disponível em: <https://journals.ametsoc.org/doi/10.1175/2007EI238.1>.

GIAMBELLUCA, T. W. *et al.* Evapotranspiration and energy balance of Brazilian savannas with contrasting tree density. **Agricultural and Forest Meteorology**, v. 149, n. 8, p. 1365–1376, aug 2009. ISSN 01681923. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0168192309000604>.

GIBBS, H. K. *et al.* Tropical forests were the primary sources of new agricultural land in the 1980s and 1990s. **Proceedings of the National Academy of Sciences of the United States of America**, v. 107, n. 38, p. 16732–16737, 2010. ISSN 10916490.

- GONG, L. *et al.* Sensitivity of the Penman–Monteith reference evapotranspiration to key climatic variables in the Changjiang (Yangtze River) basin. **Journal of Hydrology**, v. 329, n. 3-4, p. 620–629, oct 2006. ISSN 00221694. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0022169406001399>.
- GRACE, J. *et al.* Fluxes of carbon dioxide and water vapour over an undisturbed tropical forest in south-west Amazonia. **Global Change Biology**, v. 1, n. 1, p. 1–12, feb 1995. ISSN 1354-1013. Disponível em: <http://doi.wiley.com/10.1111/j.1365-2486.1995.tb00001.x>.
- GRAVES, A.; SCHMIDHUBER, J. Framewise phoneme classification with bidirectional LSTM and other neural network architectures. **Neural Networks**, v. 18, n. 5-6, p. 602–610, 2005. ISSN 08936080.
- HOCHREITER, S.; SCHMIDHUBER, J. Long Short-Term Memory. **Neural Computation**, v. 9, n. 8, p. 1735–1780, nov 1997. ISSN 0899-7667. Disponível em: <https://www.mitpressjournals.org/doi/abs/10.1162/neco.1997.9.8.1735>.
- HORST, T. W. A SIMPLE FORMULA FOR ATTENUATION OF EDDY FLUXES MEASURED WITH FIRST-ORDER-RESPONSE SCALAR SENSORS. **Boundary-Layer Meteorology**, v. 82, n. 2, p. 219–233, feb 1997. ISSN 0006-8314. Disponível em: <http://link.springer.com/10.1023/A:1000229130034>.
- HYSON, P.; GARRAT, J.; FRANCEY, R. Algebraic and Electronic Corrections of Measured uw Covariance in the Lower Atmosphere. **Journal of Applied Meteorology**, v. 16, p. 43–47, 1977. Disponível em: <http://www.jstor.org/stable/26177591>.
- IBROM, A. *et al.* Strong low-pass filtering effects on water vapour flux measurements with closed-path eddy correlation systems. **Agricultural and Forest Meteorology**, v. 147, n. 3-4, p. 140–156, 2007. ISSN 01681923.
- JARVIS, P. G.; MCNAUGHTON, K. G. Stomatal Control of Transpiration: Scaling Up from Leaf to Region. **Advances in Ecological Research**, v. 15, n. C, p. 1–49, 1986. ISSN 00652504.
- KAIMAL, J. C.; FINNIGAN, J. J. **Atmospheric boundary layer flows: Their structure and measurement**. First edit. New York: Oxford University Press, Inc, 1994. 289 p. ISBN 0195062396.
- KAIMAL, J. C.; GAYNOR, J. E. Another look at sonic thermometry. **Boundary-Layer Meteorology**, v. 56, n. 4, p. 401–410, sep 1991. ISSN 0006-8314. Disponível em: <http://link.springer.com/10.1007/BF00119215>.
- KAIMAL, J. C. *et al.* Spectral characteristics of surface-layer turbulence. **Quarterly Journal of the Royal Meteorological Society**, v. 98, n. 417, p. 563–589, 1972. ISSN 1477870X.
- KIM, Y. *et al.* Gap-filling approaches for eddy covariance methane fluxes: A comparison of three machine learning algorithms and a traditional method with principal component analysis. **Global Change Biology**, Blackwell Publishing Ltd, v. 26, n. 3, p. 1499–1518, mar 2020. ISSN 13652486.
- KLJUN, N. *et al.* A simple parameterisation for flux footprint predictions. **Boundary-Layer Meteorology**, v. 112, n. 3, p. 503–523, 2004. ISSN 00068314.

KLJUN, N. *et al.* A simple two-dimensional parameterisation for Flux Footprint Prediction (FFP). **Geoscientific Model Development**, v. 8, n. 11, p. 3695–3713, 2015. ISSN 19919603.

KLJUN, N.; ROTACH, M. W.; SCHMID, H. P. A three-dimensional backward Lagrangian footprint model for a wide range of boundary-layer stratifications. **Boundary-Layer Meteorology**, v. 103, n. 2, p. 205–226, 2002. ISSN 00068314.

KRUSKAL, W. H.; WALLIS, W. A. Use of Ranks in One-Criterion Variance Analysis. **Journal of the American Statistical Association**, v. 47, n. 260, p. 583, dec 1952. ISSN 01621459. Disponível em: <https://www.jstor.org/stable/2280779?origin=crossref>.

LECLERC, M. Y.; FOKEN, T. **Footprints in Micrometeorology and Ecology**. Berlin, Heidelberg: Springer Berlin Heidelberg, 2014. 1–239 p. ISBN 978-3-642-54544-3. Disponível em: <http://link.springer.com/10.1007/978-3-642-54545-0>.

LECLERC, M. Y.; THURTELL, G. W. Footprint prediction of scalar fluxes using a Markovian analysis. **Boundary-Layer Meteorology**, v. 52, n. 3, p. 247–258, 1990. ISSN 00068314.

LU, P. *et al.* Regulation of canopy conductance and transpiration and their modelling in irrigated grapevines. **Functional Plant Biology**, v. 30, n. 6, p. 689, 2003. ISSN 1445-4408. Disponível em: <http://www.publish.csiro.au/?paper=FP02181>.

MACEDO, M. N. *et al.* Decoupling of deforestation and soy production in the southern Amazon during the late 2000s. **Proceedings of the National Academy of Sciences of the United States of America**, v. 109, n. 4, p. 1341–1346, 2012. ISSN 00278424.

MAHRT, L. Flux sampling errors for aircraft and towers. **Journal of Atmospheric and Oceanic Technology**, v. 15, n. 2, p. 416–429, 1998. ISSN 07390572.

MALLICK, K. *et al.* A critical evaluation on the role of aerodynamic and canopy-surface conductance parameterization in SEB and SVAT models for simulating evapotranspiration: A case study in the Upper Biebrza National ParkWetland in Poland. **Water (Switzerland)**, v. 10, n. 12, 2018. ISSN 20734441.

MARRIS, E. The forgotten ecosystem. **Nature**, v. 437, n. 7061, p. 944–945, oct 2005. ISSN 0028-0836. Disponível em: <http://www.nature.com/articles/437944a>.

MASSMAN, W. A simple method for estimating frequency response corrections for eddy covariance systems. **Agricultural and Forest Meteorology**, v. 104, n. 3, p. 185–198, aug 2000. ISSN 01681923. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0168192300001647>.

MAUDER, M.; FOKEN, T. Documentation and Instruction Manual of the Eddy-Covariance Software Package TK3. **Arbeitsergebnisse**, v. 3, n. 46, p. 60, ISSN 1614–8916, 2011. ISSN 1614-8916.

MOFFAT, A. M. *et al.* Comprehensive comparison of gap-filling techniques for eddy covariance net carbon fluxes. **Agricultural and Forest Meteorology**, v. 147, n. 3-4, p. 209–232, 2007. ISSN 01681923.

- MONCRIEFF, J. B. *et al.* A system to measure surface fluxes of momentum, sensible heat, water vapour and carbon dioxide. **Journal of Hydrology**, v. 188-189, n. 1-4, p. 589–611, 1997. ISSN 00221694.
- MONSON, R.; BALDOCCHI, D. **Terrestrial Biosphere-Atmosphere Fluxes**. Cambridge: Cambridge University Press, 2014. ISBN 9781139629218. Disponível em: <http://ebooks.cambridge.org/ref/id/CBO9781139629218>.
- MONTEITH, J. L. Evaporation and surface temperature. **Quarterly Journal of the Royal Meteorological Society**, v. 107, n. 451, p. 1–27, 1981. ISSN 1477870X.
- MOORE, C. J. Frequency response corrections for eddy correlation systems. **Boundary-Layer Meteorology**, v. 37, n. 1-2, p. 17–35, oct 1986. ISSN 0006-8314. Disponível em: <http://link.springer.com/10.1007/BF00122754>.
- MYERS, N. *et al.* Biodiversity hotspots for conservation priorities. **Nature**, v. 403, n. 6772, p. 853–858, feb 2000. ISSN 0028-0836. Disponível em: <http://www.nature.com/articles/35002501>.
- OKI, T.; KANAE, S. Global hydrological cycles and world water resources. **Science**, v. 313, n. 5790, p. 1068–1072, 2006. ISSN 00368075.
- OLIVEIRA, P. T. *et al.* The water balance components of undisturbed tropical woodlands in the Brazilian cerrado. **Hydrology and Earth System Sciences**, v. 19, n. 6, p. 2899–2910, 2015. ISSN 16077938.
- OLIVEIRA, P. T. S. *et al.* Trends in water balance components across the Brazilian Cerrado. **Water Resources Research**, v. 50, n. 9, p. 7100–7114, sep 2014. ISSN 00431397. Disponível em: <http://doi.wiley.com/10.1002/2013WR015202>.
- OLIVEIRA, R. S. *et al.* Deep root function in soil water dynamics in cerrado savannas of central Brazil. **Functional Ecology**, v. 19, n. 4, p. 574–581, 2005. ISSN 02698463.
- ONCLEY, S. P. *et al.* The energy balance experiment EBEX-2000. Part I: Overview and energy balance. **Boundary-Layer Meteorology**, v. 123, n. 1, p. 1–28, 2007. ISSN 00068314.
- PAPALE, D.; VALENTINI, R. A new assessment of European forests carbon exchanges by eddy fluxes and artificial neural network spatialization. **Global Change Biology**, v. 9, n. 4, p. 525–535, 2003. ISSN 13541013.
- REICHSTEIN, M. *et al.* On the separation of net ecosystem exchange into assimilation and ecosystem respiration: Review and improved algorithm. **Global Change Biology**, v. 11, n. 9, p. 1424–1439, 2005. ISSN 13541013.
- SABBATINI, S. *et al.* Eddy covariance raw data processing for CO₂ and energy fluxes calculation at ICOS ecosystem stations. **International Agrophysics**, v. 32, n. 4, p. 495–515, dec 2018. ISSN 2300-8725. Disponível em: <http://archive.sciendo.com/INTAG/intag.2017.32.issue-4/intag-2017-0043/intag-2017-0043.pdf>.
- SCHOTANUS, P.; NIEUWSTADT, F. T.; De Bruin, H. A. Temperature measurement with a sonic anemometer and its application to heat and moisture fluxes. **Boundary-Layer Meteorology**, v. 26, n. 1, p. 81–93, 1983. ISSN 00068314.

SHERSTINSKY, A. Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) network. **Physica D: Nonlinear Phenomena**, Elsevier B.V., v. 404, p. 132306, 2020. ISSN 01672789. Disponível em: <https://doi.org/10.1016/j.physd.2019.132306>.

SOUZA, C. M. *et al.* 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, p. 2735, 2020. ISSN 2072-4292. Disponível em: <https://www.mdpi.com/2072-4292/12/17/2735>.

STULL, R. B. **An Introduction to Boundary Layer Meteorology**. 1st. ed. Dordrecht: Springer Netherlands, 1988. 670 p. ISBN 978-90-277-2769-5. Disponível em: <http://link.springer.com/10.1007/978-94-009-3027-8>.

TWINE, T. E. *et al.* Correcting eddy-covariance flux underestimates over a grassland. **Agricultural and Forest Meteorology**, v. 103, n. 3, p. 279–300, 2000. ISSN 01681923.

VERSTRAETEN, W. W.; VEROUSTRAETE, F.; FEYEN, J. Assessment of evapotranspiration and soil moisture content across different scales of observation. **Sensors**, v. 8, n. 1, p. 70–117, 2008. ISSN 14248220.

VICKERS, D.; MAHRT, L. Quality Control and Flux Sampling Problems for Tower and Aircraft Data. **Journal of Atmospheric and Oceanic Technology**, v. 14, n. 3, p. 512–526, jun 1997. ISSN 0739-0572. Disponível em: [http://journals.ametsoc.org/doi/10.1175/1520-0426\(1997\)014%3C0512:QCAFSP%3E2.0.CO;2](http://journals.ametsoc.org/doi/10.1175/1520-0426(1997)014%3C0512:QCAFSP%3E2.0.CO;2).

VILLALOBOS-VEGA, R. *et al.* Do groundwater dynamics drive spatial patterns of tree density and diversity in Neotropical savannas? **Journal of Vegetation Science**, v. 25, n. 6, p. 1465–1473, 2014. ISSN 16541103.

VOURLITIS, G. L. *et al.* Seasonal variations in the evapotranspiration of a transitional tropical forest of Mato Grosso, Brazil. **Water Resources Research**, v. 38, n. 6, p. 30–1–30–11, 2002. ISSN 00431397.

WEBB, E.; PEARMAN, G.; LEUNING, R. Correction of flux measurements for density effects due to heat and water vapour transfer. **J. R. Met. Soc.**, p. 85–100, 1980. ISSN 00359009.

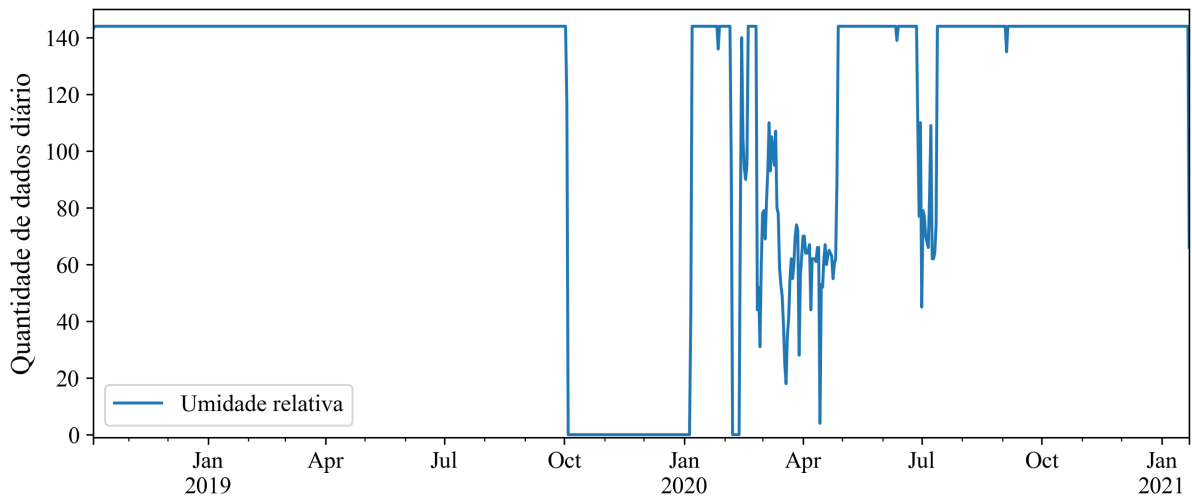
WILCZAK, J. M.; ONCLEY, S. P.; STAGE, S. A. Sonic Anemometer Tilt Correction Algorithms. **Boundary-Layer Meteorology**, v. 99, n. 1, p. 127–150, apr 2001. ISSN 0006-8314. Disponível em: <http://link.springer.com/10.1023/A:1018966204465>.

APÊNDICES

APÊNDICE A – QUANTIDADE DE DADOS METEOROLÓGICOS NAS ESTAÇÕES DO IAB

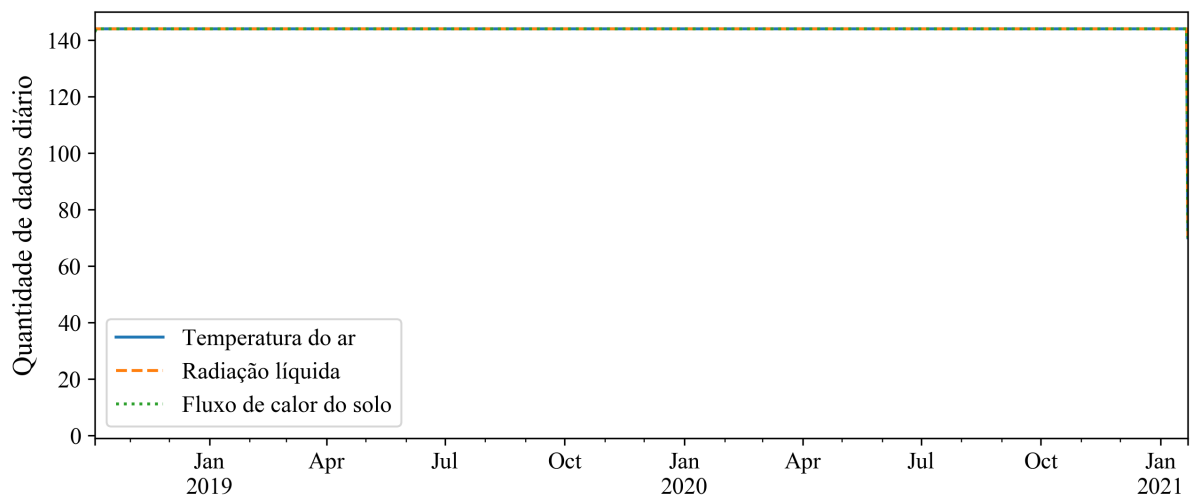
A discretização dos dados meteorológicos para a estação IAB1 e IAB2 é de a cada 10 minutos, enquanto para a estação IAB3 é de 30 minutos.

Figura 42 – Quantidade de dados diários da umidade relativa da estação IAB1



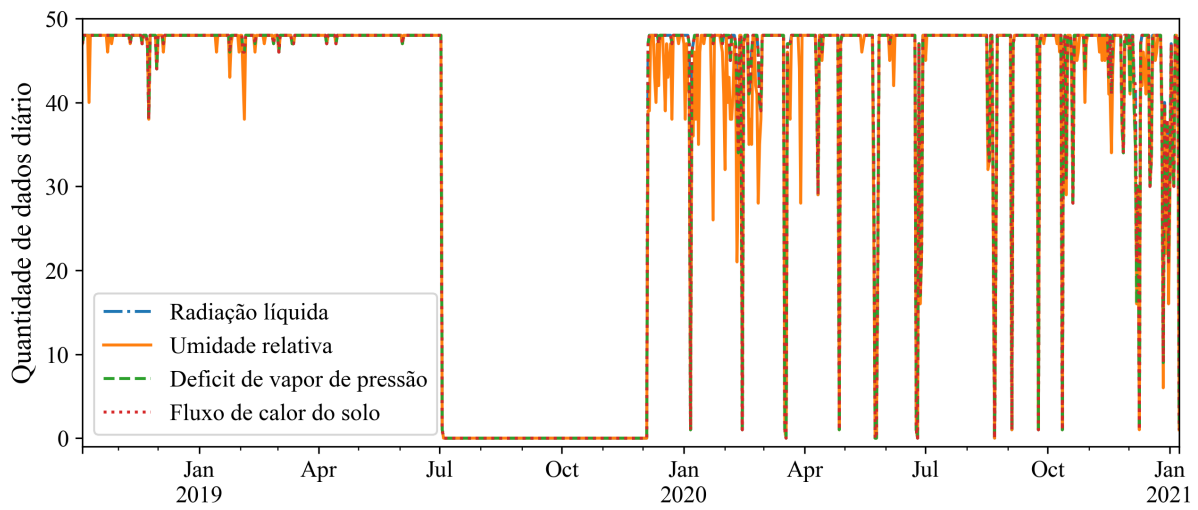
Fonte: Autor

Figura 43 – Quantidade de dados meteorológicos da estação IAB2



Fonte: Autor

Figura 44 – Quantidade de dados meteorológicos da estação IAB3



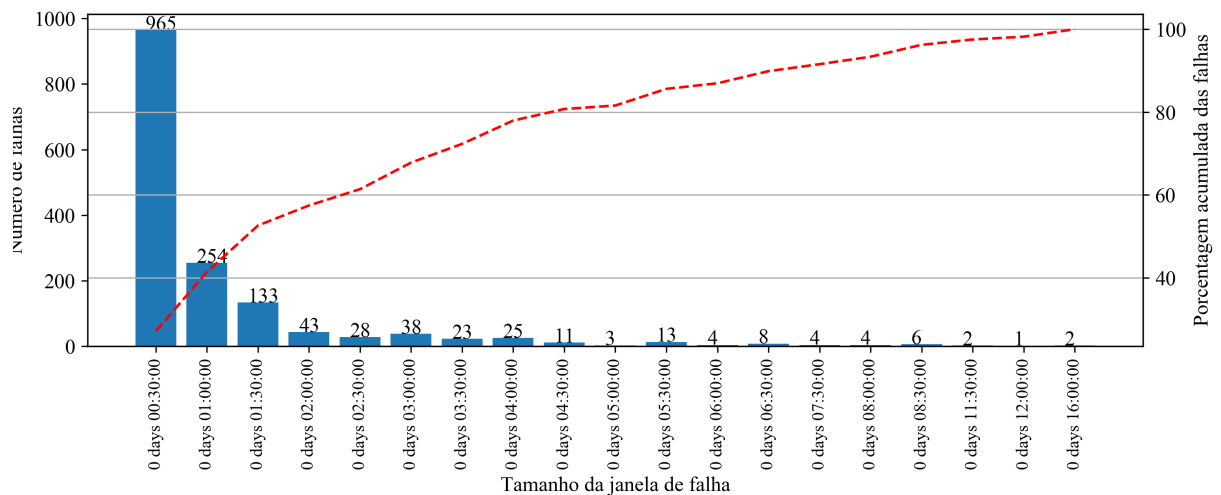
Fonte: Autor

APÊNDICE B – QUANTIDADE DE FALHAS NA SÉRIE DE ET APÓS APLICADO OS MÉTODOS DE PREENCHIMENTO DE FALHAS

A barra azul representa a quantidade de falhas para cada tipo de janela, enquanto a reta tracejada vermelha representa o acumulado de todos os tipos de falhas levando, sendo ponderado o tamanho da falha.

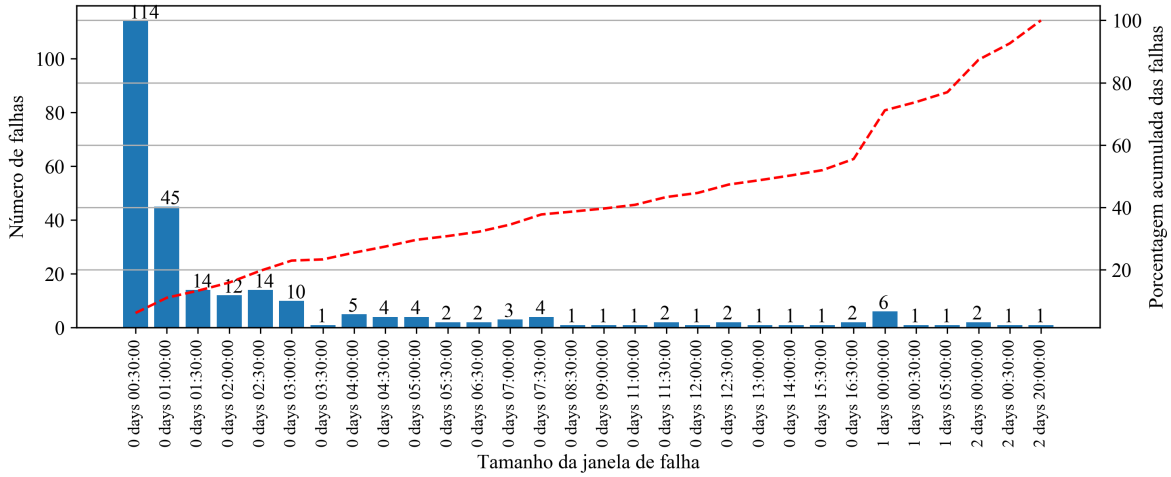
As figuras abaixo desconsideraram a grande falha entre a fase 01 e fase 02, somente o último método de Penman-Monteith, tem o registro da grande falha (62 dias). Sendo que o tamanho original da falha era de aproximadamente 120 dias.

Figura 45 – Quantidade e tipos de falhas restantes na série de ET após o método de variação diurna média



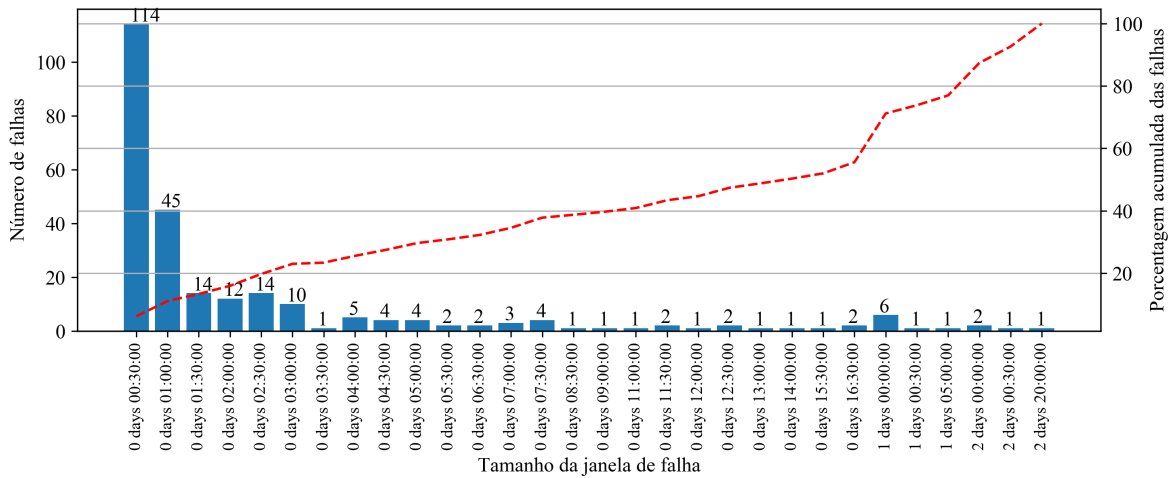
Fonte: Autor

Figura 46 – Quantidade e tipos de falhas restantes na série de ET após o método de regressão linear múltipla



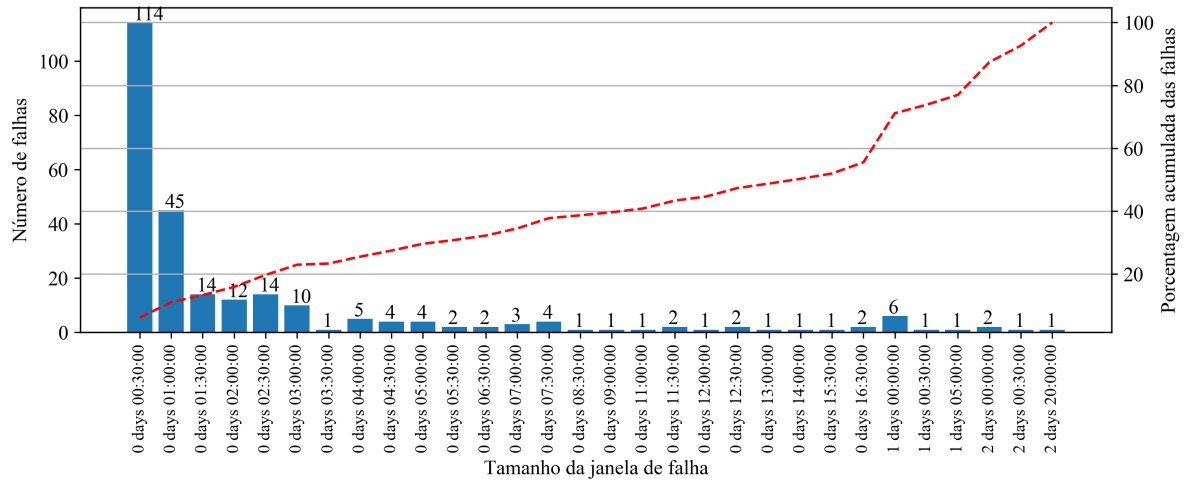
Fonte: Autor

Figura 47 – Quantidade e tipos de falhas restantes na série de ET após o método de random forest



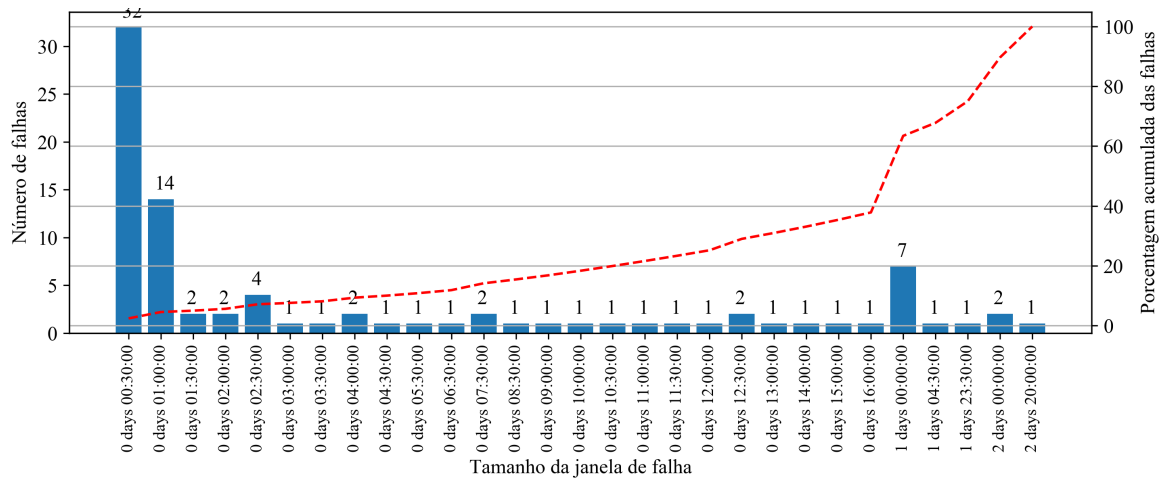
Fonte: Autor

Figura 48 – Quantidade e tipos de falhas restantes na série de ET após o método de rede neural densa



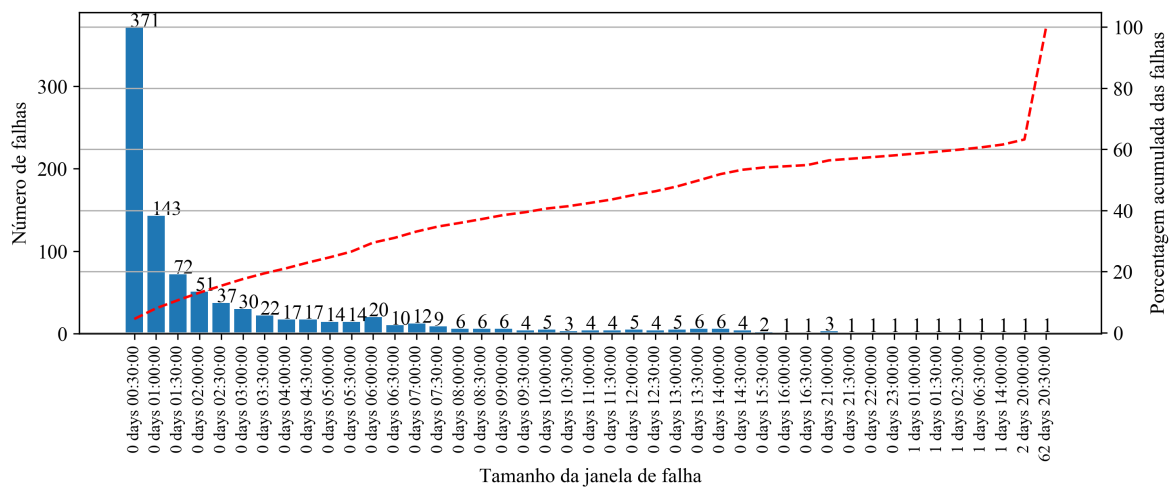
Fonte: Autor

Figura 49 – Quantidade e tipos de falhas restantes na série de ET após o método de rede neural recorrente



Fonte: Autor

Figura 50 – Quantidade e tipos de falhas restantes na série de ET após o método de Penman-Monteith inverso



Fonte: Autor

APÊNDICE C – RESULTADOS DOS MÉTODOS DE PREENCHIMENTO DE FALHAS

O método de preenchimento de falhas utilizados são: variação diurna média (ET_{MDV}), regressão linear múltipla (ET_{LR}), *random forest* (ET_{RFR}), Penman-Monteith inverso (ET_{PM}), rede neural densa (ET_{DNN}) e rede neural recorrente do tipo LSTM (ET_{LSTM}).

A tabela é o acumulado dos métodos na escala temporal diária em *mm*.

TIMESTAMP	ET_{MDV}	ET_{LR}	ET_{RFR}	ET_{PM}	ET_{DNN}	ET_{LSTM}
2018-10-05	2.54	3.47	3.18	2.64	4.59	3.15
2018-10-06	2.64	2.72	2.6	1.88	3.66	2.34
2018-10-07	2.75	1.72	1.83		2.67	1.25
2018-10-08	2.81	1.92	1.82		2.82	1.24
2018-10-09	2.86	4.33	3.73		4.99	3.29
2018-10-10	2.92	3.54	3.5		4.24	3.69
2018-10-11	3	2.61	2.86	1.51	3.96	2.18
2018-10-12	3.03	4.74	3.8		5.38	3.49
2018-10-13	3.04	3.95	3.29		4.61	2.51
2018-10-14	3.02	3.47	3.25	1.87	4.21	2.62
2018-10-15	3.09	2.94	3.02		3.55	2.42
2018-10-16	3.28	3.93	3.46		4.69	2.89
2018-10-17	3.59	3.72	3.36		4.35	2.72
2018-10-18	3.7	4.82	4.56		5.72	3.85
2018-10-19	3.71	3.29	3.74	2.37	4.61	3.18
2018-10-20	3.75	3.06	3.82	2.87	4.03	2.91
2018-10-21	3.64	3.27	3.62	3.14	4.11	3.15
2018-10-22	3.57	3.59	3.76		4.56	3.23
2018-10-23	3.41	3.95	3.96		5	3.23
2018-10-24	3.11	1.47	1.57		2.15	1.25
2018-10-25	3.17	1.51	1.46		2.04	0.96
2018-10-26	3.21	3.06	3.01		3.89	2.51
2018-10-27	3.27	2.49	2.73		3.21	1.73
2018-10-28	3.37	3.21	3.6	3.15	4.05	2.63
2018-10-29	3.5	3.1	3.33	3.09	4	2.61
2018-10-30	3.54	4	3.86		5.01	3.51
2018-10-31	3.59	3.99	3.58		4.88	3.05
2018-11-01	3.79	3.46	3.65		4.53	2.26

TIMESTAMP	ET_{MDV}	ET_{LR}	ET_{RFR}	ET_{PM}	ET_{DNN}	ET_{LSTM}
2018-11-02	3.6	4.26	4.12		5.12	2.94
2018-11-03	3.56	4.36	4.24		5.32	3.57
2018-11-04	3.5	3.58	4.01	0	4.78	3.18
2018-11-05	3.43	2.9	2.93	0	4.13	2.61
2018-11-06	3.32	3.7	3.71		4.72	3.15
2018-11-07	3.3	1.94	2.35	0	2.88	1.64
2018-11-08	2.89	1.22	2.24		2.64	1.36
2018-11-09	3.06	2.14	2.89		3.3	2.05
2018-11-10	3.08	2.58	2.9		3.49	2.11
2018-11-11	3.19	3.46	3.26		4.21	2.6
2018-11-12	3.52	4.94	4.65		5.77	4.13
2018-11-13	3.62	5.11	4.75		6	4.3
2018-11-14	3.6	4.54	4.05		5.3	3.3
2018-11-15	3.71	5.37	4.42	0	6.11	4.23
2018-11-16	3.83	5.04	4.61	0	6.19	4.43
2018-11-17	3.56	3.76	3.36		4.85	3.17
2018-11-18	3.4	3.36	3.22		4.21	2.39
2018-11-19	3.2	1.34	1.8		2.44	1.13
2018-11-20	2.89	2.39	2.87	0	3.58	2.2
2018-11-21	2.94	2.52	2.49	0	3.45	2.05
2018-11-22	2.98	2.88	3		3.82	2.37
2018-11-23	2.98	3.05	2.8		3.92	2.25
2018-11-24	3.23	3.03	3.31		4.11	2.31
2018-11-25	3.43	2.99	3.3		3.94	2.38
2018-11-26	3.37	3.13	3.52	0	4.33	2.86
2018-11-27	3.77	4.23	4.33	0	5.1	4.12
2018-11-28	3.62	4.12	4.18		4.94	3.78
2018-11-29	3.81	4.75	4.74		5.36	4.1
2018-11-30	3.81	4.07	3.8		5.02	3.36
2018-12-01	3.74	2.95	2.63		4.28	2.1
2018-12-02	3.83	3.77	4.13	2.78	4.94	3.17
2018-12-03	3.7	3.95	4.02	3.36	5.01	3.73
2018-12-04	3.65	4.38	4.27		4.82	4.05
2018-12-05	3.79	3.25	3.65		4.72	3.3
2018-12-06	3.69	3.5	3.78		4.62	2.59
2018-12-07	3.65	3.63	3.8	6.29	4.44	2.56
2018-12-08	3.63	3.38	3.72	3.71	4.43	2.95
2018-12-09	3.55	3.6	3.78		4.64	3.55

TIMESTAMP	ET_{MDV}	ET_{LR}	ET_{RFR}	ET_{PM}	ET_{DNN}	ET_{LSTM}
2018-12-10	3.53	3.63	3.88		5	3.36
2018-12-11	3.9	3.84	3.95		5.34	3.26
2018-12-12	4.21	4.55	4.31		5.73	4.1
2018-12-13	4.39	4.36	4.39		5.45	3.56
2018-12-14	4.4	4.49	4.77		5.67	4.36
2018-12-15	4.41	4.34	4.52		5.51	4.06
2018-12-16	4.43	4.27	4.17		5.41	3.76
2018-12-17	4.62	5.16	4.77		5.87	4.05
2018-12-18	4.54	4.36	3.97		5.07	3.39
2018-12-19	4.61	3.75	3.63		4.72	3.05
2018-12-20	4.47	5.43	4.99		6.33	4.57
2018-12-21	4.52	4.82	4.53		5.78	3.88
2018-12-22	4.46	5.13	4.92		5.96	4.24
2018-12-23	4.38	3.95	3.63		4.9	3.24
2018-12-24	4.13	3.41	3.27		4.45	2.76
2018-12-25	3.97	2.69	2.68		3.64	2.18
2018-12-26	3.75	2.5	2.49		3.36	1.92
2018-12-27	3.87	3.93	3.64		4.97	3.44
2018-12-28	3.85	4.22	3.95		5.05	3.76
2018-12-29	4.1	5.24	4.67		5.84	4.48
2018-12-30	3.99	5.18	4.85		6.04	4.59
2018-12-31	4.13	4.9	4.93		6.09	4.71
2019-01-01	4.07	3.3	3.4	2.64	4.28	2.8
2019-01-02	3.84	4.66	3.89	3.74	5.15	3.8
2019-01-03	3.64	5.31	4.41	4.62	5.83	4.28
2019-01-04	3.39	3.52	3.11	2.56	4.2	2.75
2019-01-05	3.5	2.05	1.98	1.07	2.91	1.4
2019-01-06	3.56	2.87	2.63	1.96	3.56	2.08
2019-01-07	3.39	5.38	5.06	5.69	6.2	4.62
2019-01-08	3.69	4.63	3.74	4.4	5.22	3.69
2019-01-09	3.85	5.32	4.94	5.82	6.18	4.53
2019-01-10	4.41	5.14	5.01	6.61	6.08	3.85
2019-01-11	4.53	4.41	4.05	4.82	5.18	3.7
2019-01-12	4.96	3.87	3.98	3.48	4.94	3.76
2019-01-13	4.79	5.07	5.44	4.99	6.17	4.65
2019-01-14	4.75	4.8	4.64	5.64	5.76	4.81
2019-01-15	4.85	4.7	4.62	5.74	5.61	3.9
2019-01-16	4.84	4.02	3.95	4.72	4.94	3.28

TIMESTAMP	ET_{MDV}	ET_{LR}	ET_{RFR}	ET_{PM}	ET_{DNN}	ET_{LSTM}
2019-01-17	4.74	3.38	3.68	4.33	4.6	2.7
2019-01-18	4.66	4.82	4.47	5.08	5.63	4.06
2019-01-19	4.58	4.91	4.41	5.32	5.55	4.21
2019-01-20	4.84	4.77	4.43	5.13	5.49	4.04
2019-01-21	4.55	5.49	4.85	5.85	6.37	4.43
2019-01-22	4.34	5.13	4.59	5.84	5.79	4.3
2019-01-23	4.33	4.27	4.38	4.2	5.3	4.04
2019-01-24	4.24	2.75	2.94	3.81	3.71	3.26
2019-01-25	4.27	2.99	3.11	4.34	3.7	2.01
2019-01-26	4.56	3.7	4.22	4.21	4.82	3.87
2019-01-27	4.45	3.91	4.06	4.53	5.01	3.39
2019-01-28	4.84	4.47	4.29	6.55	5.76	3.97
2019-01-29	5.12	4.67	4.65	7.11	6.14	3.89
2019-01-30	4.93	5.21	4.7	6.33	6.12	4.16
2019-01-31	5.14	4.83	4.54	6.12	5.83	4.79
2019-02-01	4.63	4.78	4.34	8.52	5.84	3.76
2019-02-02	4.29	5.01	4.62	8.11	6.03	3.61
2019-02-03	4.24	5.25	5.37	6.87	6.02	4.73
2019-02-04	4.19	3.13	3.69	4.34	4.2	4.05
2019-02-05	4.21	2.99	3.4	4.78	4.24	2.74
2019-02-06	4.32	1.64	1.91	1.79	2.66	1.49
2019-02-07	4.04	3.64	3.96	4.62	4.72	3.49
2019-02-08	4.1	4.4	4.78	6.3	5.65	4.3
2019-02-09	4.13	4.54	4.73	7.38	5.75	4.03
2019-02-10	3.74	4.65	4.48	6.55	5.79	3.82
2019-02-11	3.35	5.27	4.7	6.11	6.19	3.91
2019-02-12	3.08	2.68	2.51	2.9	3.73	1.99
2019-02-13	2.82	2.03	2.69	2.96	3.11	1.77
2019-02-14	2.69	3.27	3.44	5.3	4.53	2.87
2019-02-15	2.7	2.66	2.46	2.57	3.59	1.7
2019-02-16	2.64	2.31	2.24	1.77	3.03	1.39
2019-02-17	2.92	2.43	2.03	2.51	3	1.57
2019-02-18	2.8	3.1	3.02	3.29	4	2.19
2019-02-19	3.07	3.17	2.98	4.16	3.82	2.37
2019-02-20	3.18	3.06	2.96	3.05	3.91	2.67
2019-02-21	3.56	3.02	2.79	2.97	3.62	2.37
2019-02-22	3.64	3.16	3.12	3.01	4.08	2.82
2019-02-23	3.68	4.51	4	4.69	5.03	3.84

TIMESTAMP	ET_{MDV}	ET_{LR}	ET_{RFR}	ET_{PM}	ET_{DNN}	ET_{LSTM}
2019-02-24	3.57	3.97	3.52	4.04	4.55	3.2
2019-02-25	3.58	4.76	4.27	4.57	5.47	3.77
2019-02-26	3.72	2.29	2.18	1.97	3.08	1.61
2019-02-27	3.75	3.4	3.51	2.99	4.31	2.72
2019-02-28	3.67	2.17	2.07	1.41	3.06	1.65
2019-03-01	3.38	1.53	1.57	0.95	2.46	1.03
2019-03-02	3.45	2.49	2.32	1.5	3.21	1.88
2019-03-03	3.5	2.85	2.81	1.72	3.75	2.42
2019-03-04	3.68	3.1	3.27	3.24	4.1	2.65
2019-03-05	3.81	3.25	3.03	2.86	4.14	2.71
2019-03-06	3.8	3.97	3.83	3.57	4.93	3.62
2019-03-07	3.54	2.65	2.61	2.35	3.45	2.16
2019-03-08	3.71	4.14	3.66	4	4.98	3.39
2019-03-09	3.71	4.58	4.04	4.19	5.41	3.8
2019-03-10	3.71	4.05	3.6	3.23	4.97	3.43
2019-03-11	3.76	2.83	3.12	2.3	3.72	2.42
2019-03-12	3.67	3.58	3.42	3.56	4.48	2.99
2019-03-13	3.43	4.15	3.73	3.46	4.85	3.39
2019-03-14	3.47	3.45	3.16	2.44	4.32	2.9
2019-03-15	3.63	3.55	3.45	2.55	4.5	3.09
2019-03-16	3.98	3.19	2.82	2.28	3.97	2.4
2019-03-17	4.06	4.36	3.82	3.42	5.05	3.46
2019-03-18	3.68	4.47	3.77	3.3	5.19	3.62
2019-03-19	3.56	4.67	4.53	3.36	5.52	3.74
2019-03-20	3.74	3.69	3.3	2.65	4.78	2.88
2019-03-21	3.69	2.3	2.79	1.87	3.54	2.11
2019-03-22	3.61	2.54	2.76	2.49	3.43	2.41
2019-03-23	3.31	3.37	3.62	3.35	4.45	3.26
2019-03-24	3.29	3.53	3.53	3.3	4.66	3.43
2019-03-25	3.37	3.98	3.83	4.8	4.89	3.54
2019-03-26	3.55	4.34	4.08	5.02	5.27	3.85
2019-03-27	3.55	3.69	3.69	3.65	4.73	3.33
2019-03-28	3.74	3.23	3.69	3.27	4.41	3.24
2019-03-29	3.75	3.01	3.36	3.17	4.11	3.14
2019-03-30	3.72	3.47	3.77	3.95	4.59	3.6
2019-03-31	3.71	3.7	3.99	4.27	4.73	3.65
2019-04-01	3.68	3.37	3.77	3.54	4.63	3.34
2019-04-02	3.78	3.67	3.63	3.66	4.82	3.52

TIMESTAMP	ET_{MDV}	ET_{LR}	ET_{RFR}	ET_{PM}	ET_{DNN}	ET_{LSTM}
2019-04-03	3.76	3.41	3.72	4.07	4.35	2.62
2019-04-04	3.75	4.39	3.89	4.67	5.09	3.49
2019-04-05	3.51	4.55	3.69	3.21	5.01	3.16
2019-04-06	3.07	4.95	4.46	3.35	5.67	4.02
2019-04-07	2.96	3.37	2.72	1.92	3.78	2.33
2019-04-08	2.84	1.74	1.99	0.84	2.81	1.34
2019-04-09	2.86	2.02	2.47	1.67	3.07	1.91
2019-04-10	2.86	2.73	3.11	2.41	3.94	2.71
2019-04-11	2.86	3.27	3.31	2.86	4.29	3.1
2019-04-12	2.83	3.55	3.6	3.22	4.65	3.28
2019-04-13	3.09	2.79	2.99	2.28	3.77	2.26
2019-04-14	2.95	1.1	1.47	0.68	1.95	0.84
2019-04-15	3.22	3.29	3.25	2.27	4.29	2.94
2019-04-16	3.25	3.83	3.34	2.62	4.5	3.22
2019-04-17	3.22	3.14	3.02	2.35	4.27	2.85
2019-04-18	3.23	2.99	3.87	3.39	3.89	2.91
2019-04-19	3.2	2.94	3.55	3.23	3.93	2.99
2019-04-20	3.21	3.04	3.39	3.58	4.08	2.96
2019-04-21	3.4	3.45	3.81	4.17	4.56	2.92
2019-04-22	3.09	3.58	3.38	3.72	4.38	2.93
2019-04-23		2.89	2.82	2.05	3.79	2.52
2019-04-24		3.32	2.97	2.34	4.01	2.69
2019-04-25		3.9	3.6	3.09	4.89	3.3
2019-04-26		3.51	2.95	2.67	4.18	3.01
2019-04-27		3.83	3.32	3.13	4.7	3.22
2019-04-28		3.48	3.19	2.9	4.42	2.86
2019-04-29		2.55	2.62	2.01	3.45	2.41
2019-04-30		2.82	3.2	2.72	3.79	2.96
2019-05-01	2.32	2.5	2.51	2.44	3.29	2.19
2019-05-02	2.34	2.21	2.15	1.89	2.89	1.58
2019-05-03	2.34	3.55	3.19	2.68	4.33	2.79
2019-05-04	2.31	3.25	2.97	2.48	4.07	2.62
2019-05-05	2.26	3.57	3.17	2.63	4.31	3.1
2019-05-06	2.47	3.59	2.9	2.73	4.27	3.05
2019-05-07	2.43	2.83	2.65	2.07	3.56	2.45
2019-05-08	2.54	2.56	2.6	2.18	3.58	2.56
2019-05-09	2.55	3.04	3.09	2.17	3.93	2.76
2019-05-10	2.55	3.21	2.79	1.93	3.78	2.54

TIMESTAMP	ET_{MDV}	ET_{LR}	ET_{RFR}	ET_{PM}	ET_{DNN}	ET_{LSTM}
2019-05-11	2.53	2.87	2.35	1.66	3.27	1.83
2019-05-12	2.41	3.68	2.95	2.52	4.24	3
2019-05-13	2.27	2.9	2.68	2.39	3.8	2.6
2019-05-14	2.26	2.8	2.54	1.98	3.47	2.59
2019-05-15	2.2	1.75	1.99	1.83	2.58	1.63
2019-05-16	2.17	1.75	1.91	1.7	2.57	1.57
2019-05-17	2.13	1.26	1.45	1.04	2.07	1.11
2019-05-18	2.06	1.41	1.81	0.89	2.03	1.2
2019-05-19	2.08	2.31	2.28	1.82	3.04	2.32
2019-05-20	2.11	2.35	2.49	2.09	3.12	2.38
2019-05-21	2	2.29	2.42	2.37	2.9	2.07
2019-05-22	2.01	2.42	2.66	2.83	3	2.11
2019-05-23	2.11	2.41	2.24	2.91	3.06	1.78
2019-05-24	1.99	1.51	1.63	1.28	2.05	0.88
2019-05-25	1.93	2.15	1.69	1.51	2.15	1.53
2019-05-26		2.38	2.45	2.24	2.73	1.75
2019-05-27	1.81	2.57	2.48	2.98	2.77	1.84
2019-05-28	2.03	2.87	2.74	2.79	3.68	1.82
2019-05-29	1.98	2.97	2.82	2.45	3.9	2.37
2019-05-30	2.02	3.18	2.96	2.59	4.1	2.48
2019-05-31	1.86	3.12	2.96	2.98	3.97	1.83
2019-06-01	1.98	2.86	2.52	2.32	3.7	2.18
2019-06-02	2.04	1.98	1.98	1.53	2.86	1.8
2019-06-03	2.04	0.71	0.9	0.54	1.52	0.67
2019-06-04	2.04	1.75	2.05	1.32	2.22	1.37
2019-06-05	2.04	2.04	2.14	1.57	2.45	1.52
2019-06-06	2.09	2.17	2.1	1.54	2.57	1.6
2019-06-07	2.18	2.19	2.18	1.94	2.92	1.53
2019-06-08	2.13	1.82	1.99	1.8	2.43	1.49
2019-06-09	2.14	2.13	2.18	1.78	2.67	1.83
2019-06-10	2.15	2.11	2.01	1.72	2.65	1.57
2019-06-11	2.24	2.32	2.35	2.26	3.21	1.78
2019-06-12		2.19	2.19	2.2	2.95	1.76
2019-06-13		2.53	2.51	2.28	3.5	1.79
2019-06-14		2.48	2.66	2.63	3.52	1.53
2019-06-15		2.25	2.38	2.4	3.2	1.77
2019-06-16		2.2	2.42	2.44	3.24	2.02
2019-06-17		1.76	1.99	1.82	2.67	1.57

TIMESTAMP	ET_{MDV}	ET_{LR}	ET_{RFR}	ET_{PM}	ET_{DNN}	ET_{LSTM}
2019-06-18		2.18	2.14	2.09	2.92	1.63
2019-06-19		1.98	2.04	2.16	2.64	1.6
2019-06-20		2.2	2.16	2.28	3.01	1.91
2019-06-21	2.01	1.89	2.15	1.82	2.79	1.77
2019-06-22	2.03	1.97	2.14	1.65	2.9	1.92
2019-06-23	2.05	1.99	2.13	1.98	2.84	1.53
2019-06-24	1.9	2.28	2.24	2.13	2.85	1.75
2019-06-25	1.84	2.58	2.59	2.56	3.37	1.49
2019-06-26	1.82	2.78	2.73	2.84	3.46	1.49
2019-06-27	1.83	2.3	2.12	2.13	3.3	1.81
2019-06-28		2.45	2.53	2.39	3.17	1.72
2019-06-29		2.43	2.54	2.48	3.23	1.56
2019-06-30		2.53	2.58	2.28	3.3	1.78
2019-07-01		2.37	2.42	2.1	3.08	1.62
2019-07-02		2.4	2.33	2.07	2.77	1.4
2019-07-03				2.31		
2019-07-04				1.32		
2019-07-05				0.22		
2019-07-06				0.81		
2019-07-07				0.84		
2019-07-08				1.23		
2019-07-09				1.62		
2019-07-10				1.8		
2019-07-11				1.82		
2019-07-12				2.14		
2019-07-13				2.22		
2019-07-14				2.51		
2019-07-15				2.37		
2019-07-16				0.94		
2019-07-17				1.07		
2019-07-18				1.33		
2019-07-19				1.57		
2019-07-20				1.83		
2019-07-21				1.79		
2019-07-22				1.58		
2019-07-23				2.05		
2019-07-24				2.23		
2019-07-25				2.09		

TIMESTAMP	ET_{MDV}	ET_{LR}	ET_{RFR}	ET_{PM}	ET_{DNN}	ET_{LSTM}
2019-07-26				2.01		
2019-07-27				2.44		
2019-07-28				1.76		
2019-07-29				1.69		
2019-07-30				2.05		
2019-07-31				2.06		
2019-08-01				2.38		
2019-08-02				2.68		
2019-08-03				0.37		
2019-08-04				0.17		
2019-08-05				0.35		
2019-08-06				0.95		
2019-08-07				1.58		
2019-08-08				2.1		
2019-08-09				2.32		
2019-08-10				2.81		
2019-08-11				2.67		
2019-08-12				2.62		
2019-08-13				2.34		
2019-08-14				0.96		
2019-08-15				1.4		
2019-08-16				1.89		
2019-08-17				2.27		
2019-08-18				2.71		
2019-08-19				0.97		
2019-08-20				0.96		
2019-08-21				1.28		
2019-08-22				1.35		
2019-08-23				1.11		
2019-08-24				1.39		
2019-08-25				1.58		
2019-08-26				2		
2019-08-27				1.95		
2019-08-28				1.85		
2019-08-29				2.31		
2019-08-30				2.88		
2019-08-31				3.08		
2019-09-01				0.64		

TIMESTAMP	ET_{MDV}	ET_{LR}	ET_{RFR}	ET_{PM}	ET_{DNN}	ET_{LSTM}
2019-09-02				0.99		
2019-09-03				0.63		
2019-09-04				1.05		
2019-09-05				0.79		
2019-09-06				1.03		
2019-09-07				1.9		
2019-09-08				2.45		
2019-09-09				2.52		
2019-09-10				2.91		
2019-09-11				3.18		
2019-09-12				3.46		
2019-09-13				1.79		
2019-09-14				1.82		
2019-09-15				2.32		
2019-09-16				2.92		
2019-09-17				3.25		
2019-09-18				3.33		
2019-09-19				2.4		
2019-09-20				1.4		
2019-09-21				1.25		
2019-09-22				1.16		
2019-09-23				1.33		
2019-09-24				1.59		
2019-09-25				0.22		
2019-09-26				0.3		
2019-09-27				1.09		
2019-09-28				1.31		
2019-09-29				1.64		
2019-09-30				2.15		
2019-10-01				5.16		
2019-10-02				6.13		
2019-10-03				5.01		
2019-10-04						
2019-10-05						
2019-10-06						
2019-10-07						
2019-10-08						
2019-10-09						

TIMESTAMP	ET_{MDV}	ET_{LR}	ET_{RFR}	ET_{PM}	ET_{DNN}	ET_{LSTM}
2019-10-10						
2019-10-11						
2019-10-12						
2019-10-13						
2019-10-14						
2019-10-15						
2019-10-16						
2019-10-17						
2019-10-18						
2019-10-19						
2019-10-20						
2019-10-21						
2019-10-22						
2019-10-23						
2019-10-24						
2019-10-25						
2019-10-26						
2019-10-27						
2019-10-28						
2019-10-29						
2019-10-30						
2019-10-31						
2019-11-01						
2019-11-02						
2019-11-03						
2019-11-04						
2019-11-05						
2019-11-06						
2019-11-07						
2019-11-08						
2019-11-09						
2019-11-10						
2019-11-11						
2019-11-12						
2019-11-13						
2019-11-14						
2019-11-15						
2019-11-16						

TIMESTAMP	ET_{MDV}	ET_{LR}	ET_{RFR}	ET_{PM}	ET_{DNN}	ET_{LSTM}
2019-11-17						
2019-11-18						
2019-11-19						
2019-11-20						
2019-11-21						
2019-11-22						
2019-11-23						
2019-11-24						
2019-11-25						
2019-11-26						
2019-11-27						
2019-11-28						
2019-11-29						
2019-11-30						
2019-12-01	4.26					
2019-12-02	4.43					
2019-12-03	4.3					
2019-12-04	4.53					
2019-12-05	4.61	2.24	1.69		2.99	3.26
2019-12-06	4.62	2.4	2.19		3.19	3.83
2019-12-07	4.71	4.29	5.21	5.02	5.44	4.11
2019-12-08	4.33	4.35	4.58		5.47	4.03
2019-12-09	4.22	4	3.94		4.82	3.14
2019-12-10	4.46	2.78	2.28		3.48	2.6
2019-12-11	4.21	1.93	1.69		2.58	2.92
2019-12-12	4.07	3.43	3.11		4.15	2.86
2019-12-13	4.03	4.51	4.84		5.55	4.66
2019-12-14	4.2	3.98	4.17	2.34	4.79	3.22
2019-12-15	3.96	4.22	3.75		5.1	3.39
2019-12-16	3.88	5.06	4.66		5.64	3.47
2019-12-17	3.39	3.69	3.6		4.55	3.97
2019-12-18	3.21	2.45	2.32	1.8	2.59	3.03
2019-12-19	3.04	3.04	2.99		3.84	2.89
2019-12-20	3.01	2.88	2.76		3.34	3.16
2019-12-21	3.58	3.88	3.94		4.75	3.63
2019-12-22	4.46	2.69	2.41		3.26	2.2
2019-12-23	4.53	1.49	1.19		1.88	2.59
2019-12-24	4.73	4.39	5.03	4.38	5.39	4.24

TIMESTAMP	ET_{MDV}	ET_{LR}	ET_{RFR}	ET_{PM}	ET_{DNN}	ET_{LSTM}
2019-12-25	4.67	4.9	5.33		6.04	4.75
2019-12-26	5.08	4.39	4.29		5.27	4.08
2019-12-27	5.02	5.1	5.15	4.64	6.26	4.68
2019-12-28	5.01	4.56	4.42		5.23	4.01
2019-12-29	4.82	4.94	4.87		5.83	4.38
2019-12-30	4.43	4.73	4.39		5.5	3.97
2019-12-31	4.08	4.36	3.84		5.07	3.61
2020-01-01	3.89	4.73	4.02		5.49	4.15
2020-01-02	3.95	3.37	3.04		4.23	4.26
2020-01-03	3.93	3.07	3.11		4	2.86
2020-01-04	3.94	3.72	3.91	2.59	4.51	4.2
2020-01-05	4.06	3.36	2.87		3.7	3.58
2020-01-06	4.22					
2020-01-07	4.41	4.79	4.33	3.94	5.34	4.46
2020-01-08	4.16	4.61	4.24	3.51	5.27	5.74
2020-01-09	3.87	3.82	3.15	2.34	4.31	3.32
2020-01-10	3.81	4.2	3.15	2.75	4.21	5.33
2020-01-11	3.7	5.08	4.23	3.64	5.55	3.89
2020-01-12	3.79	3.48	3.06	2.76	4.15	4.64
2020-01-13	3.88	3.52	3.13	2.71	3.93	2.95
2020-01-14	3.61	3.54	3.16	2.95	4.16	2.76
2020-01-15	3.77	5.3	4.93	5.38	6.02	4.5
2020-01-16	3.76	5.44	4.99	5.16	6.3	4.63
2020-01-17	3.92	3.11	2.79	2.81	3.98	2.66
2020-01-18	4.02	4.37	4.51	4.35	5.38	4.05
2020-01-19	3.92	4.01	3.88	3.79	4.83	3.47
2020-01-20	3.82	3.98	3.82	3.58	4.91	3.36
2020-01-21	4.14	4.26	4.13	4.28	5.12	3.94
2020-01-22	4.05	2.84	2.86	2.5	3.58	4.04
2020-01-23	4.21			0.82		4.17
2020-01-24	4.31	3.57	3.85	3.03	4.7	3.04
2020-01-25	4.34	3.8	4.34	4.42	4.54	3.87
2020-01-26	4.34	3.9	4.73	5.81	4.84	3.79
2020-01-27	4.39	4.49	4.3	6.82	4.88	3.77
2020-01-28	4.14	4.65	4.45	5.44	5.4	3.96
2020-01-29	3.93	4.2	3.78	3.76	4.87	3.55
2020-01-30	3.33	3.25	2.62	2	3.49	2.25
2020-01-31	3.32	4.26	3.29	3.14	4.53	4.66

TIMESTAMP	ET _{MDV}	ET _{LR}	ET _{RFR}	ET _{PM}	ET _{DNN}	ET _{LSTM}
2020-02-01	3.29	2.86	2.18	1.78	2.79	5.04
2020-02-02	3.45	3.71	3.14	2.96	4.4	2.93
2020-02-03	3.41	2.49	2.21	2	2.9	3.57
2020-02-04	3.53	3.35	3.3	3.09	4.01	3.94
2020-02-05	3.62	3.46	3.33	2.69	3.99	3.96
2020-02-06	3.64	3.23	3.02	2.09	3.82	4.44
2020-02-07	3.23	3.31	3.29		3.99	4.07
2020-02-08	3.41	4.09	3.83		4.72	3.58
2020-02-09	2.57	3.36	3.26		4.17	3.44
2020-02-10	2.82					4.15
2020-02-11	2.7	1.45	1.75		1.95	1.82
2020-02-12	2.99	1.73	1.83	1.74	2.53	1.41
2020-02-13	3.05	3.11	3.51	3.48	4.18	3.06
2020-02-14	3.14			5.14		
2020-02-15	3.21	4.63	4.25	4.47	5.38	4.09
2020-02-16	3.57	5.3	4.75	5.13	5.92	4.35
2020-02-17	3.59	5.25	4.69	5.3	5.9	4.46
2020-02-18	3.64	5.04	4.38	4.1	5.5	4.28
2020-02-19	3.7	4.02	3.38	3.54	4.46	4.32
2020-02-20	3.79	4.27	4.12	3.8	4.95	4.34
2020-02-21	3.69	1.2	0.65	0.77	1.43	2.77
2020-02-22	3.68	3.06	3.42	2.6	3.89	2.76
2020-02-23	3.71	3.36	3.79	3.7	4.25	3.15
2020-02-24	3.6	2.43	2.02	2.02	2.8	3.03
2020-02-25	3.4	3.69	3.08	3.3	4.12	4.04
2020-02-26	3.17					3.56
2020-02-27	2.98	2.45	2.69		2.96	2.93
2020-02-28	3.34	1.18	1.57	1.25	1.8	1.4
2020-02-29	3.16	2.67	2.78	2.74	3.44	2.45
2020-03-01	3.32	2.88	3.25	2.71	3.85	2.54
2020-03-02	3.52	3.05	3.76	3.21	4.06	2.75
2020-03-03	3.68	2.74	2.68	2.66	3.51	2.64
2020-03-04	3.77	3.46	3.93	3.53	4.4	3.55
2020-03-05	3.83	3.17	3.97	3.76	4.01	3.15
2020-03-06	3.78	3.11	3.65	3.56	4.06	3.24
2020-03-07	4.01	3.28	3.64	4.08	4.09	3.3
2020-03-08	3.97	3.3	3.85	4.06	4	3.17
2020-03-09	3.84	3.45	3.58	3.83	4.41	3.32

TIMESTAMP	ET_{MDV}	ET_{LR}	ET_{RFR}	ET_{PM}	ET_{DNN}	ET_{LSTM}
2020-03-10	3.92	3.54	3.75	4.34	4.53	3.25
2020-03-11	3.87	4.13	4.09	4.95	4.93	3.72
2020-03-12	3.69	4.05	4.1	4.55	4.84	3.63
2020-03-13	3.36	3.98	4.34	4.64	4.78	3.58
2020-03-14	3.19	4.31	4.23		5.22	3.79
2020-03-15	3.72	4.45	3.87		5.01	3.68
2020-03-16	3.57	4.28	4.02		4.95	3.65
2020-03-17	3.38					
2020-03-18	3.29					
2020-03-19	3.55	2.92	2.37		3.46	2.05
2020-03-20	3.58	3.26	2.55		3.99	3.42
2020-03-21	3.65	2.5	2.74		3.24	3.34
2020-03-22	3.67	3.05	4.04	3.59	3.99	3.34
2020-03-23	3.74	2.94	3.71	3.55	3.89	3.17
2020-03-24	3.81	2.86	3.6	3.9	3.72	2.93
2020-03-25	4.02	2.76	3.42		3.52	2.66
2020-03-26	3.84	2.81	3.49	4.08	3.84	2.77
2020-03-27	3.78	3.28	3.71		4.1	3.01
2020-03-28	3.82	3.26	3.3		4.2	3.55
2020-03-29	3.78					4.55
2020-03-30	3.69	3.36	3.56		4.38	3.33
2020-03-31	3.47	4.03	4.01		4.95	3.79
2020-04-01	3.47	4.07	4.01	3.42	5.15	3.88
2020-04-02	3.46	4.53	4.01		5.23	3.93
2020-04-03	3.43	3.93	3.44		4.9	3.18
2020-04-04	3.28	3.08	3.64		3.91	2.82
2020-04-05	3.18	3.37	3.63		4.23	2.97
2020-04-06	3.16	4.18	3.98		5.09	3.5
2020-04-07	3.27	2.28	2.22	1.51	2.98	1.82
2020-04-08	3.19	2.82	2.94	2.49	3.89	2.79
2020-04-09	3.21	2.85	3.09		3.56	2.51
2020-04-10	3.21	2.88	3.27		3.61	2.65
2020-04-11	3.19					
2020-04-12	3.03	2.94	3.38	2.94	3.93	2.89
2020-04-13	2.94	3.26	3.43	2.9	4.46	3.21
2020-04-14	2.92	0.8	0.8		1.2	0.71
2020-04-15	2.84	2.25	2.04		2.95	2.07
2020-04-16	2.62	2.46	2.44		3.31	2.32

TIMESTAMP	ET _{MDV}	ET _{LR}	ET _{RFR}	ET _{PM}	ET _{DNN}	ET _{LSTM}
2020-04-17	2.49	2.85	3.25	2.64	3.95	2.89
2020-04-18	2.64	2.74	2.82	2.45	3.58	2.68
2020-04-19	2.91	2.47	2.57	2.36	3.34	2.44
2020-04-20	3	2.7	2.91		3.67	2.79
2020-04-21	2.97	2.8	3.14		3.71	2.78
2020-04-22	3.05	2.85	3.32		3.66	2.68
2020-04-23	3.07	2.81	3.16		3.59	2.59
2020-04-24	3.16	2.97	2.92		4	2.63
2020-04-25	3.14	2.93	3.16		3.98	2.49
2020-04-26	3.07	2.74	3.06		3.77	2.55
2020-04-27	3					
2020-04-28	2.9	2.56	2.64	3.2	3.27	2.23
2020-04-29	2.95	2.57	2.72	3.49	3.47	2.21
2020-04-30	2.84	2.54	2.78	3.35	3.44	2.16
2020-05-01	2.81	2.89	3.05	3.27	3.78	2.55
2020-05-02	2.72	2.75	2.98	3.33	3.43	2.11
2020-05-03	2.72	2.56	2.57	2.45	3.27	1.93
2020-05-04	2.57	2.58	2.6	2.48	2.97	2.09
2020-05-05	2.53	2.57	2.26	2.85	2.78	1.87
2020-05-06	2.44	2.49	2.42	2.93	3.06	1.41
2020-05-07	2.42	1.89	1.86	1.65	2.24	1.24
2020-05-08	2.33	2.38	2.38	2.16	2.8	1.81
2020-05-09	2.28	2.45	2.56	2.47	3.02	1.89
2020-05-10	2.23	2.51	2.43	2.81	3.11	1.68
2020-05-11	2.23	2.43	2.32	2.83	3.24	1.63
2020-05-12	2.14	2.42	2.29	2.86	2.98	1.6
2020-05-13	2.28	1.82	1.7	1.97	2.35	0.91
2020-05-14	2.34	1.71	1.64	1.2	2.24	1.37
2020-05-15	2.38	1.82	2.09	1.34	2.56	1.59
2020-05-16	2.42	2.18	2.56	2	2.91	2.15
2020-05-17	2.48	2.38	2.63	2.59	2.89	2.02
2020-05-18	2.55	2.55	2.57	3.16	3.21	2.01
2020-05-19	2.65	2.39	2.22	2.68	3.14	1.9
2020-05-20	2.62	2.37	2.39	2.75	3.38	1.64
2020-05-21	2.55	2.29	2.39	2.73	3.27	1.61
2020-05-22	2.33	2.65	2.79	2.7	3.55	1.6
2020-05-23				0.41		1.6
2020-05-24				0.61		

TIMESTAMP	ET_{MDV}	ET_{LR}	ET_{RFR}	ET_{PM}	ET_{DNN}	ET_{LSTM}
2020-05-25				1.54		
2020-05-26				1.81		
2020-05-27		2.88	1.95	2.36	2.36	1.59
2020-05-28	1.89	2.81	2.08	2.38	2.53	1.51
2020-05-29	1.91	2.58	2.15	2.88	2.67	1.6
2020-05-30	1.85	2.46	1.97	2.88	2.63	1.45
2020-05-31	1.92	2.43	1.83	2.78	2.58	1.52
2020-06-01	1.93	2.47	2.17	2.27	2.79	1.64
2020-06-02	1.87	1.8	1.55	1.5	2.23	1.15
2020-06-03	1.89	1.68	1.88	1.18	2.33	1.42
2020-06-04	1.84	1.95	2.11	1.41	3.18	1.95
2020-06-05	1.9	1.99	2.16	1.25	3.26	1.51
2020-06-06	1.97	2.27	2.23	1.61	3.16	1.2
2020-06-07	2.04	1.91	1.81	1.08	2.81	2.44
2020-06-08	1.98	2.43	2.37	1.64	3.31	2.23
2020-06-09	2.06	2.47	2.54	1.6	3.3	1.99
2020-06-10	2.09	2.81	2.59	2.15	3.67	1.79
2020-06-11	2.14	2.72	2.5	2.53	3.68	1.73
2020-06-12	2.06	2.5	2.5	2.41	3.25	1.74
2020-06-13	2.04	2.55	2.59	2.85	3.47	1.92
2020-06-14	1.99	1.86	1.99	1.55	2.77	1.83
2020-06-15	2.01	1.97	2.11	1.68	2.79	1.79
2020-06-16	2.01	1.99	2.23	1.77	2.93	1.82
2020-06-17		2.1	2.2	1.68	3.1	1.86
2020-06-18		2.02	2.03	1.86	2.71	1.33
2020-06-19	1.92	2.07	2.12	1.93	2.83	1.64
2020-06-20	2.02	2.33	2.41	2.31	3.14	1.88
2020-06-21	1.85	2.32	2.53	2.34	3	1.59
2020-06-22	1.56	2.33	2.47	2.51	3.23	1.54
2020-06-23	1.45	1.91	1.82	1.93	2.51	1.15
2020-06-24	1.52			2.09		
2020-06-25	1.68			2.41		
2020-06-26	1.7	1.35	1.46	1	2.15	1.32
2020-06-27	1.7			0.28		
2020-06-28	1.63			0.94		
2020-06-29	1.62	1.94	1.95	1.41	2.56	1.41
2020-06-30	1.65	2.39	2.59	1.65	3.06	1.53
2020-07-01	1.64	1.96	1.69		1.7	1.72

TIMESTAMP	ET_{MDV}	ET_{LR}	ET_{RFR}	ET_{PM}	ET_{DNN}	ET_{LSTM}
2020-07-02	1.64	2.96	1.9		1.99	1.54
2020-07-03	1.58	2.56	1.8		2.26	1.46
2020-07-04	1.59	2.56	2.16		2.6	1.69
2020-07-05	1.62	2.57	2.49		3.32	1.42
2020-07-06	1.74	2.62	2.63		3.59	1.57
2020-07-07	1.74	2.45	2.53		3.06	1.67
2020-07-08	1.87	2.57	2.71	1.99	2.9	1.59
2020-07-09	1.85	2.11	2.1		2.83	1.82
2020-07-10	1.87	2.51	2.3		2.86	1.52
2020-07-11	1.81	2.36	2.39		3.1	1.36
2020-07-12	1.84	2.51	2.52		3.2	1.39
2020-07-13	1.79	2.6	2.37	2.43	3.31	1.49
2020-07-14	1.78	2.27	2.1	1.72	3.11	1.94
2020-07-15	1.85	1.99	1.91	1.35	2.94	1.72
2020-07-16	1.86	2.26	2.2	1.6	3.03	1.74
2020-07-17	1.87	2.34	2.44	2.15	3.32	1.74
2020-07-18	1.88	2.32	2.44	2.11	3.2	1.65
2020-07-19	1.99	2.28	2.25	2.03	2.91	1.44
2020-07-20	2.02	2.42	2.32	2.27	3.16	1.53
2020-07-21	2.07	2.5	2.5	2.29	3.25	1.51
2020-07-22	2	2.54	2.52	2.46	3.14	1.75
2020-07-23	1.99	2.48	2.35	2.31	3.17	1.61
2020-07-24	2.05	2.47	2.45	2.35	3.08	1.46
2020-07-25	1.93	2.57	2.57	2.68	3.4	1.45
2020-07-26	1.85	2.34	2.35	2.03	3.07	1.67
2020-07-27	1.76	2.63	2.61	2.22	3.29	1.66
2020-07-28	1.8	2.71	2.51	2.58	3.5	1.57
2020-07-29	1.87	2.46	2.31	2.01	3.19	1.95
2020-07-30	1.84	2.13	1.85	1.52	2.6	1.43
2020-07-31	1.84	2.18	2.11	1.89	2.98	1.67
2020-08-01	1.87	2.25	2.28	1.59	3.02	1.96
2020-08-02	1.9	2.26	2.27	1.71	3.08	1.73
2020-08-03	1.96	2.24	2.3	1.8	3.13	1.61
2020-08-04	1.98	2.36	2.29	1.88	3.23	1.65
2020-08-05	1.97	2.39	2.35	1.93	3.28	1.68
2020-08-06	1.96	2.54	2.32	1.98	3.28	1.62
2020-08-07	1.95	2.63	2.29	1.97	3.2	1.53
2020-08-08		2.1	1.99	1.8	2.83	1.25

TIMESTAMP	ET_{MDV}	ET_{LR}	ET_{RFR}	ET_{PM}	ET_{DNN}	ET_{LSTM}
2020-08-09	1.65	2.31	2.5	2.28	3.35	1.41
2020-08-10	1.57	2.78	2.58	2.42	3.35	1.46
2020-08-11	1.43	2.41	2.47	2.27	3.27	1.25
2020-08-12	1.39	2.97	2.86	2.71	3.57	1.58
2020-08-13	1.23	3.29	3.24	2.89	3.94	1.91
2020-08-14	1.23	3.23	2.89	2.87	4.02	1.46
2020-08-15	1.4	3.37	3.1	2.64	4.01	1.91
2020-08-16	1.44	1.62	1.28	1.09	2.14	0.75
2020-08-17	1.52	0.35	0.38	0.3	0.82	0.44
2020-08-18	1.51	1.74	1.7	1.04	2.19	1.31
2020-08-19	1.54	3.18	2.9	1.87	3.84	1.99
2020-08-20	2.07	1.99	1.78	1.12	2.46	1.17
2020-08-21	2.19			0.23		
2020-08-22	2.23			0.62		
2020-08-23	2.34			1.5		
2020-08-24	2.35	2.93	2.41	1.68	3.16	1.89
2020-08-25	2.39	2.99	2.25	1.71	3.14	1.86
2020-08-26	2.34	3.07	2.53	2.09	3.15	1.72
2020-08-27	2.21	3.07	2.67	2.38	3.54	1.92
2020-08-28	2.2	3.28	3.02	2.77	3.99	1.86
2020-08-29	2.28	3.31	2.96	2.89	4.07	1.79
2020-08-30	2.25	3.42	2.93	3.27	4.18	1.65
2020-08-31	2.26	3.26	2.68	2.49	4.32	2.09
2020-09-01	2.2	3.04	2.58	1.98	4.12	2.18
2020-09-02	2.23	2.97	2.57	1.91	3.84	2.04
2020-09-03	2.24	3.21	2.93	2.37	4.09	1.74
2020-09-04	1.98			2.35		
2020-09-05	1.98	2.97	2.59	2.56	3.86	1.71
2020-09-06	1.89	3	2.67	2.85	3.94	1.52
2020-09-07	1.95	3.23	2.72	2.65	3.75	1.76
2020-09-08	1.89	3.21	2.39	2.43	3.64	1.97
2020-09-09	1.85	3.24	2.45	2.36	3.94	1.81
2020-09-10	1.71	3.32	2.59	2.76	3.88	1.58
2020-09-11	1.68	3.47	2.69	3.16	3.91	1.39
2020-09-12	1.71	3.38	2.64	3.29	3.93	1.38
2020-09-13	1.75	3.2	2.68	3.08	4.19	1.37
2020-09-14	1.81	2.82	2.72	2.92	4.07	1.16
2020-09-15	1.7	3	2.34	2.04	3.89	1.94

TIMESTAMP	ET_{MDV}	ET_{LR}	ET_{RFR}	ET_{PM}	ET_{DNN}	ET_{LSTM}
2020-09-16	1.66	3.05	2.81	2.28	3.72	1.81
2020-09-17	1.67	3.5	2.7	3.07	4.04	1.26
2020-09-18	1.67	3.06	2.35	2.26	3.78	1.58
2020-09-19	1.72	3.05	2.23	2.36	3.66	1.49
2020-09-20	1.73	1.87	1.69	1.07	2.7	1.32
2020-09-21	1.75	1.92	1.69	0.96	2.63	1.44
2020-09-22	1.79	1.57	2.02	0.77	2.18	1.25
2020-09-23	1.79	2.35	2.13	1.49	3.31	1.9
2020-09-24	1.81			1.84		
2020-09-25	2.02	3.34	2.65	2.38	4.12	2.51
2020-09-26	2.33	3.56	2.9	3.05	3.86	1.96
2020-09-27	2.52	4.01	3.17	3.55	3.88	1.51
2020-09-28	2.53	4.14	2.86	2.55	4.29	2.02
2020-09-29	2.62	4.41	3.2	2.7	5.05	2.68
2020-09-30	2.58	4.53	3.24	3.27	4.37	1.88
2020-10-01	2.55	4.92	3.4	7.28	4.99	1.7
2020-10-02	2.5	4.97	4.08	7.43	4.98	1.4
2020-10-03	2.44	4.81	3.31	6.01	5.56	2.29
2020-10-04	2.44	3.86	2.94	3.91	4.83	2.71
2020-10-05	2.18	4.72	3.52	5.91	4.66	2.6
2020-10-06	2.08	4.97	3.86	7.32	4.98	1.76
2020-10-07	1.92	4.83	3.59	7.69	4.76	1.65
2020-10-08	1.99	2.43	2.65	2.89	3.67	1.54
2020-10-09	2.04	2.86	2.58	2.63	3.69	2.1
2020-10-10	2.06	1.61	1.76	1.96	2.29	1.69
2020-10-11	2.07	2.88	2.69	4.24	3.28	1.91
2020-10-12	2.19			4.67		
2020-10-13	2.09	3.89	2.9	5.59	4.27	2.37
2020-10-14	2.09	4.1	2.93	6.12	4.6	1.82
2020-10-15	1.96			1.14		0.5
2020-10-16	1.93	1.39	1.63	1.72	2.23	0.91
2020-10-17	1.94	3.35	2.86	3.46	4.18	2.79
2020-10-18	2.11	3.57	3.41	3.84	4.54	2.71
2020-10-19	2.24	2.62	2.39	2.9	3.3	1.62
2020-10-20	2.57			1.74		
2020-10-21	2.82	3.61	3.79	3.34	4.62	3.12
2020-10-22	3.18	3.76	3.55	4.17	4.68	2.86
2020-10-23	3.27	3.51	3.59	3.88	4.12	2.96

TIMESTAMP	ET_{MDV}	ET_{LR}	ET_{RFR}	ET_{PM}	ET_{DNN}	ET_{LSTM}
2020-10-24	3.12	2.31	2.33	1.52	3.04	2.12
2020-10-25	3.1	3.74	3.51	3.03	4.42	3.43
2020-10-26	3.16	3.83	3.48	3.78	4.62	3.16
2020-10-27	2.88	4.13	3.3	4.42	4.89	3.12
2020-10-28	2.85	4.2	3.49	5.5	4.57	2.69
2020-10-29	2.67	3.32	3.09	3.4	3.89	3.21
2020-10-30	2.55	1.62	1.73	1.18	2.49	1.38
2020-10-31	2.51	3.34	2.99	2.75	3.96	2.45
2020-11-01	2.41	3.19	2.46	2.18	3.46	2.28
2020-11-02	2.43	3.59	2.8	2.79	3.75	2.55
2020-11-03	2.57	3.19	2.74	2.83	3.61	2.26
2020-11-04	2.56	3.18	2.85	3.25	3.85	2.3
2020-11-05	2.71	3.28	2.94	3.18	3.72	2.45
2020-11-06	2.69	3.45	2.94	3.31	4	2.4
2020-11-07	2.71	3.74	3.06	4.14	4.14	2.6
2020-11-08	2.75	4	3.25	5.07	4.41	2.24
2020-11-09	2.94	4.71	3.51	5.33	4.94	2.71
2020-11-10	3.02	3.61	3.18	4.05	4.48	2.6
2020-11-11	3.12	3.53	3.26	0.92	4.38	3.31
2020-11-12	3.29	4.45	4.03	1.87	5.25	3.99
2020-11-13	3.6	3.48	3.33	0.98	4.32	3.45
2020-11-14	3.64	4.21	4.09	0.98	5.15	3.72
2020-11-15	3.76	4.81	4.34	0.83	5.78	3.98
2020-11-16	3.44	3.42	2.79	0.81	3.93	2.85
2020-11-17	3.55	3.44	2.76	0.83	4.34	3.97
2020-11-18	3.61	3.67	3.68	0.28	4.36	4.47
2020-11-19	3.77	2.19	2.53	0.78	3.07	2.3
2020-11-20	3.75	3.51	3.7	2.3	4.19	3.21
2020-11-21	3.77	3.54	3.7	2.6	4.33	3.05
2020-11-22	3.78	3.79	4.04	3.01	4.8	3.45
2020-11-23	3.96	3.68	3.79	3.17	4.7	3.18
2020-11-24	4.11	4.04	4.18	1.75	5.15	3.37
2020-11-25	4.18	4.37	3.98	0.52	5.16	3.18
2020-11-26	4.27	3.98	4.03	3.25	4.7	3.22
2020-11-27	4.41	4.05	4.39	3.06	4.96	3.94
2020-11-28	4.3	4.21	4.31	4.04	5.2	3.19
2020-11-29	4.37	4.58	4.59	4.89	5.34	3.19
2020-11-30	4.08	4.25	4.11	3.83	4.91	3.06

TIMESTAMP	ET_{MDV}	ET_{LR}	ET_{RFR}	ET_{PM}	ET_{DNN}	ET_{LSTM}
2020-12-01	3.83	4.58	4.12	5.13	5.33	3.56
2020-12-02	3.74	3.52	3.45	3.77	4.26	3.38
2020-12-03	3.78	3.72	3.62	4.03	4.67	3.22
2020-12-04	3.62	5.14	4.86	5.86	6.03	4.29
2020-12-05	3.8	3.64	3.39	3.72	4.56	3.62
2020-12-06	3.79	2.67	2.87	2.41	3.42	2.98
2020-12-07	4.05			1.75		
2020-12-08	3.85	3.75	3.7	3.4	4.3	3.31
2020-12-09	3.65			2.83		
2020-12-10	3.57	3.62	3.73	3.86	4.71	3.45
2020-12-11	3.74	2.14	2.06	2.32	2.7	2.54
2020-12-12	3.68	3.19	3.12	2.61	3.81	2.76
2020-12-13	3.97	4.16	3.86	3.77	4.91	4.09
2020-12-14	3.92	3.53	3.3	3.19	3.96	4.02
2020-12-15	4.17	4.33	3.94	3.82	5.28	3.28
2020-12-16	4.32	4.1	3.87	3.73	4.8	3.54
2020-12-17	4.32	3.42	3.46	3.68	4.16	2.88
2020-12-18	4.17	4.74	4.7	4.64	5.19	4.27
2020-12-19	4.35	5.31	4.77	5.19	6.02	4.58
2020-12-20	4.53	3.99	3.64	3.66	4.8	2.9
2020-12-21	4.52	4.85	4.16	5.15	5.58	4.25
2020-12-22	4.21	3.79	3.83	2.93	4.72	3.78
2020-12-23	4.11	4.03	4.2	3.71	5.11	3.77
2020-12-24	4.14	3.92	4.14	4.2	5.16	3.74
2020-12-25	4.21	3.92	3.99	4.03	5.01	3.67
2020-12-26	4.14	3.72	3.76	3.89	4.35	3.52
2020-12-27	4.13			0.15		
2020-12-28	3.85	2.98	3.14	0.18	3.56	2.94
2020-12-29	3.74			0.71		
2020-12-30	3.65	3.8	3.97	0.09	4.49	3.65
2020-12-31	3.94			0.15		
2021-01-01	3.96			2.26		2.72
2021-01-02		3.01	2.85	2.41	3.71	3.12
2021-01-03		3.37	3.82	3.76	4.17	3.62
2021-01-04		4.04	3.94	4.65	4.39	3.47
2021-01-05		4.62	4.5	4.96	5.32	3.68
2021-01-06		4.56	3.83	4.61	5.03	3.48
2021-01-07		4.73	4.47	4.81	5.38	4.37

TIMESTAMP	ET_{MDV}	ET_{LR}	ET_{RFR}	ET_{PM}	ET_{DNN}	ET_{LSTM}
2021-01-08				3.53		
2021-01-09				3.76		
2021-01-10				2.78		
2021-01-11				3.58		
2021-01-12				2.92		
2021-01-13				1.98		
2021-01-14				2.49		
2021-01-15				2.23		
2021-01-16				3.15		
2021-01-17				2.18		
2021-01-18				3.31		
2021-01-19				3.01		
2021-01-20				4.47		
2021-01-21				5.53		
2021-01-22						

APÊNDICE D – CÓDIGO EM PYTHON

Para visualizar os códigos escritos em Python foi criado um repositório no site <https://github.com/alexnaoki/EC-LHC>.