

RELATÓRIO PARCIAL

FAPESP – Fundação de Amparo à Pesquisa do Estado de São Paulo
MODALIDADE – Iniciação Científica

**EMIÇÃO DE CO₂ DO SOLO E SUA RELAÇÃO COM CO₂ ATMOSFÉRICO
EM ÁREAS AGRÍCOLAS DO MATO GROSSO DO SUL: UMA ABORDAGEM
DE APRENDIZADO DE MÁQUINA**

Letícia Roberta de Lima

Letícia Roberta de Lima

Bolsista



Dr. Alan Rodrigo Panosso

Orientador

Jaboticabal – SP – Brasil
Julho de 2021

SUMÁRIO

	Resumo do plano inicial:.....	3
1	RESUMO DAS ATIVIDADES REALIZADAS	3
2	INTRODUÇÃO E JUSTIFICATIVA	4
3	OBJETIVOS	5
4	MATERIAL E MÉTODOS.....	5
5	FORMA DE ANÁLISE DOS RESULTADOS.....	6
6	CRONOGRAMA DE ATIVIDADES.....	7
7	RESULTADOS PRELIMINARES.....	7
8	REFERÊNCIAS BIBLIOGRÁFICAS	13

ABORDAGEM DE APRENDIZADO DE MÁQUINA SOBRE A RELAÇÃO ENTRE A EMISSÃO DE CO₂ NO SOLO E O CO₂ ATMOSFÉRICO EM ÁREAS AGRÍCOLAS NO BRASIL CENTRAL

Resumo do plano inicial:

Projeções indicam um contínuo crescimento do agronegócio brasileiro e consequente incremento nas emissões de Gases do Efeito Estufa (GEE) advindas desse setor. O dióxido de carbono (CO₂) representa cerca de 66% das emissões totais de GEE do planeta, sendo o carbono orgânico do solo um dos principais reservatórios terrestre para o armazenamento e o intercâmbio de carbono (C) atmosférico, uma vez que, dependendo do uso e manejo de solos agrícolas, podem atuar como fontes ou sumidouros desse carbono. Modelar a dinâmica do carbono em áreas agrícolas é uma ação estratégica para a diminuição das incertezas associadas aos processos de mitigação de GEE e melhorar a capacidade de análises para construção de cenários mais acurados. Nas últimas décadas, técnicas de inteligência artificial e mineração de dados têm sido aplicadas com sucesso na modelagem de inúmeros atributos em ciência do solo. Assim, o objetivo da proposta será avaliar o desempenho preditivo de algoritmos de aprendizado de máquina para o estudo da relação de emissão de dióxido de carbono no solo (FCO₂) e o CO₂ atmosférico em áreas agrícolas na região do estado do Mato Grosso do Sul (MS), a partir de uma série temporal de 2015 a 2017. As técnicas utilizadas serão: máquina de vetores de suporte (SVM) e árvores de decisão (*Random Forest* e *Gradient Boosting Machines*). De maneira geral 70-80% das observações serão utilizadas para aprendizagem (processo de treinamento) dos modelos e 30-20% para validação. Novos experimentos estão sendo realizados em campo para validação dos resultados em áreas agrícolas. A acurácia dos modelos será determinada por meio da correlação de Pearson (r), coeficiente de determinação (R²), erro quadrático médio (RMSE), erro médio (ME), índice de concordância (d), coeficiente de confiança (c) e menor erro percentual absoluto médio (MAPE). Espera-se que essa abordagem contribua para melhorar o entendimento da dinâmica do FCO₂ e CO₂ atmosférico em diferentes regiões, usos e manejos dos solos no Brasil central, produzindo cenários com menores incertezas que possam servir como sustentáculo nas tomadas de decisões focadas na mitigação das emissões de CO₂ em áreas agrícolas.

Palavras-chaves: respiração do solo, inteligência artificial, mineração de dados, mudanças climáticas.

1 RESUMO DAS ATIVIDADES REALIZADAS

A hipótese da presente proposta é que a dinâmica temporal da emissão de CO₂ do solo e do CO₂ atmosférico pode ser descrita como um fenômeno com estruturas multidimensionais que evoluem no tempo, aptas a seres modeladas por algoritmos de aprendizado de máquina. Os dados coletados foram compilados e assim

tornou-se publicitário utilizando a plataforma GitHub com a linguagem R, com o objetivo de gerar material que possa garantir a sua reprodutividade e deste modo ser disposto para outras pesquisas. Os dados de CO₂ atmosférico para o Brasil se encontram disponíveis no link https://arpanosso.github.io/fco2r/reference/oco2_br.html e os dados de emissão de CO₂ se encontram disponíveis no link https://arpanosso.github.io/fco2r/reference/data_fco2.html.

Assim, o banco de dados está pronto para desenvolver as análises com objetivo de avaliar o desempenho preditivo dos principais algoritmos de aprendizado de máquina para FCO₂ e sua relação com o CO₂ atmosférico em áreas agrícolas na região do estado do Mato Grosso do Sul. As técnicas utilizadas serão: máquina de vetores de suporte (SVM) e árvores de decisão (*Random Forest* e *Gradient Boosting Machines*). Especificamente, pretende-se usar a abordagem de aprendizado de máquina para investigar os padrões temporais de FCO₂ e CO₂ atmosféricos no Brasil central avaliando as fontes de incerteza e de previsão em áreas agrícolas.

2 INTRODUÇÃO E JUSTIFICATIVA

Os gases de efeito estufa (GEE) permitem que a radiação solar atinja a superfície da Terra, eles absorvem e reirradiam a radiação infravermelha de saída, armazenando efetivamente parte do calor na atmosfera. À vista disso, os GEE retêm o calor na atmosfera, resultando em aquecimento (LUO; ZHOU, 2006). A emissão de GEE pelo Brasil é devido, principalmente aos setores Agropecuária e Mudança do Uso da Terra e Florestas. No ano de 2016 esses setores emitiram juntos cerca de 730,1 Mt de CO₂eq, ou seja, 56% das emissões totais de GEE do País (BRASIL, 2019).

A presente proposta compilou achados anteriores e alguns resultados ainda não publicados sobre a emissão de CO₂ do solo (FCO₂) a partir de 4 experimentos realizados durante 3 anos (Tabela 1) no Mato Grosso do Sul. Os experimentos selecionados representam usos da terra típicos da região específica onde os processos envolvidos na mudança do uso ocorreram e, em sua maioria, em áreas próximas dentro da localidade. Nesses estudos a variabilidade de FCO₂ foi determinada ao longo do tempo e do espaço, bem como de demais variáveis como temperatura, umidade do solo e atributos físicos e químicos do solo.

Tabela1. Informações geográficas, clima e histórico das áreas onde foram conduzidos os experimentos de variabilidade espaço-temporal da emissão de CO₂ e atributos físicos e químicos do solo ao longo de 3 anos no Brasil central.

Cidade	Estado	Coordenadas Geográficas	Elevação (m) ¹	Clima ²	Cultura	Manejo	TC	Ano	Solo	N
Selvíria	MS	20°20' S; 51°24' O	362	Aw	Eucalipto	Reflorestamento	29	2015-2017	Latossolo Vermelho distroférico	102
Selvíria	MS	20°20' S; 51°24' O	362	Aw	Pinus	Reflorestamento	29	2015-2017	Latossolo Vermelho distroférico	15

Selvíria	MS	20°20' S; 51°24' O	362	Aw	Silvipastoril	ILPF	29	2015- 2017	Latossolo Vermelho distroférrico	15
Selvíria	MS	20°20' S; 51°24' O	362	Aw	Nativas Cerrado	Reflorestamento	29	2015- 2016	Latossolo Vermelho distroférrico	15

TC = Tempo de conversão em anos; N = número de pontos amostrais; Ano = ano de condução do experimento.

¹ Elevação em metros acima do nível do mar.

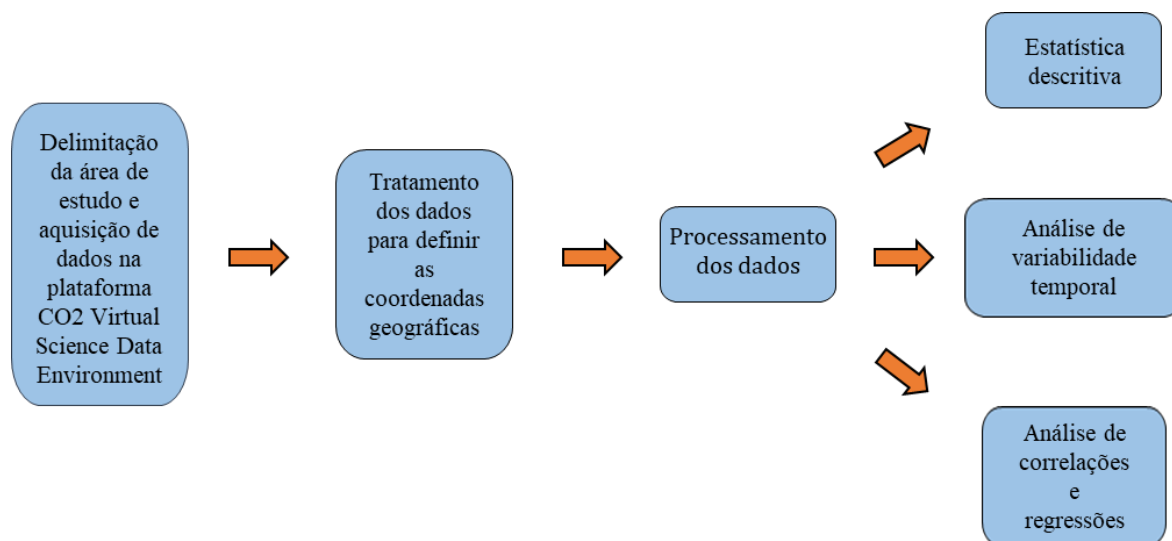
² Classificação climática de acordo com Köppen.

3 OBJETIVOS

O objetivo do projeto será avaliar o desempenho preditivo dos principais algoritmos de aprendizado de máquina para FCO₂ e sua relação com o CO₂ atmosférico em diferentes regiões do Brasil. As técnicas utilizadas serão: máquina de vetores de suporte (SVM) e árvores de decisão (*Random Forest* e *Gradient Boosting Machines*). Especificamente, pretende-se usar a abordagem de aprendizado de máquina para investigar os padrões temporais de FCO₂ e CO₂ atmosféricos no Brasil central avaliando as fontes de incerteza e de previsão em diferentes usos e manejos do solo.

4 MATERIAL E MÉTODOS

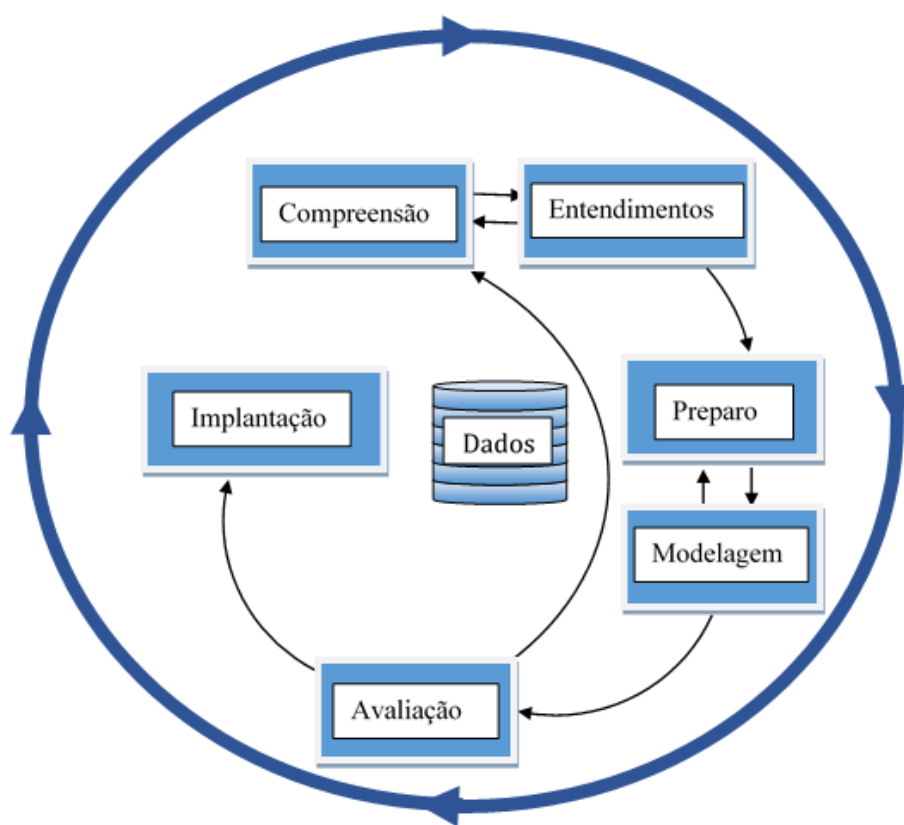
Afim de uma determinação acurada da concentração de CO₂ atmosférico, o presente estudo visa coligir dados de sensores orbitais com os dados coletados em campo, empregando um processo, descrito no Fluxograma 1, que estabelecerá um padrão entre os dados dos sensores e os dados coletados em campo a partir das coordenadas obtidas na plataforma CO₂ Virtual Science Data Environment (<https://arpanosso.github.io/oco2/index.html>).



Fluxograma 1. Fluxograma de aquisição, processamento e análise dos dados.

5 FORMA DE ANÁLISE DOS RESULTADOS

Todos os modelos de aprendizado de máquina e etapas de pré-processamento e validação dos modelos serão realizado a partir de implementações na linguagem R (R DEVELOPMENT CORE TEAM, 2021), cuja maior vantagem é ser livre, de código aberto, amplamente utilizada e de alta longevidade, o que garantirá a inspeção e reprodutibilidade dos resultados, facilitando a divulgação para a comunidade científica. As interfaces de aplicação das técnicas serão utilizados os seguintes pacotes (coleções de funções implementadas dentro da linguagem R): *Caret* e *Keras*, para aprendizado de máquina e aprendizado profundo respectivamente. Esses pacotes são conjuntos de funções implementadas para simplificação de todo o processo de criação e avaliação dos modelos preditivos além de ser recomendado para a visualização dos resultados. A Figura 5 apresenta as etapas do processo de data mining, o qual será utilizado para o processo de aprendizado de máquina. O protocolo adotado será o CRISP-DM (*Cross Industry Standard Process fo Data Mining*) o qual adaptado de Chapman et al. (2000).



a)

Figura 1. Etapas do modelo de referência adaptado CRISP-DM, onde as sequencias das fases não serão rígidas uma vez que o movimento entre etapas é requerido (CHAPMAN et al., 2000).

6 CRONOGRAMA DE ATIVIDADES

As atividades em negrito e sublinhado já foram executadas ou estão em fase de finalização.

Tabela 1 Cronograma de atividades previstas para o período de vigência da proposta atual, 2021 a 2022, período total de 1 ano (4 trimestres).

ATIVIDADES	PERÍODO: (ANO) / - TRIMESTRE -			
	2021		2022	
	- 3º -	- 4º -	- 1º -	- 2º -
Revisão de literatura	<u>X</u>			X
Avaliação da emissão de CO ₂ , captura de O ₂ temperatura e umidade do solo	<u>X</u>	<u>X</u>		
Preparo do Banco de dados para mineração e incorporação de variáveis ambientais	<u>X</u>	<u>X</u>		
Aplicação dos algoritmos de aprendizado de máquina			X	X
Confecção de dissertações, teses; Publicação de trabalho científico e programa computacional para previsão de FCO ₂ .				X
Divulgação de Resultados e preparo de relatório científico		<u>X</u>		X
Atualização de disponibilização da base de dados	<u>X</u>	<u>X</u>		

7 RESULTADOS PRELIMINARES

A presente proposta apresenta alguns resultados promissores a respeito da utilização das técnicas de aprendizado de máquina para dados de emissão de CO₂ do solo (FCO₂) e a captura de oxigênio pelo solo (FO₂) em áreas de floresta plantada no Cerrado sul-mato-grossense, município de Selvíria em Mato Grosso do Sul (MS) (Figura 2). As áreas foram: Pinus (**pi**), Eucalipto (**eu**) e área reflorestada com espécies nativas (**ns**), todas com histórico de 30 anos de conversão do solo de culturas anuais (soja, milho e algodão). As técnicas utilizadas foram redes neurais artificiais (RNA), máquina de vetores de suporte (SVM) e árvore de decisão (utilizando algoritmo *random forest* - RP). Na etapa de modelagem foram utilizados fatores ambientais e do solo, totalizando 23 covariáveis. A avaliação do fluxo de gases foi realizada entre os anos de 2015 e 2019 (proposta atual em andamento 301606/2017-0).

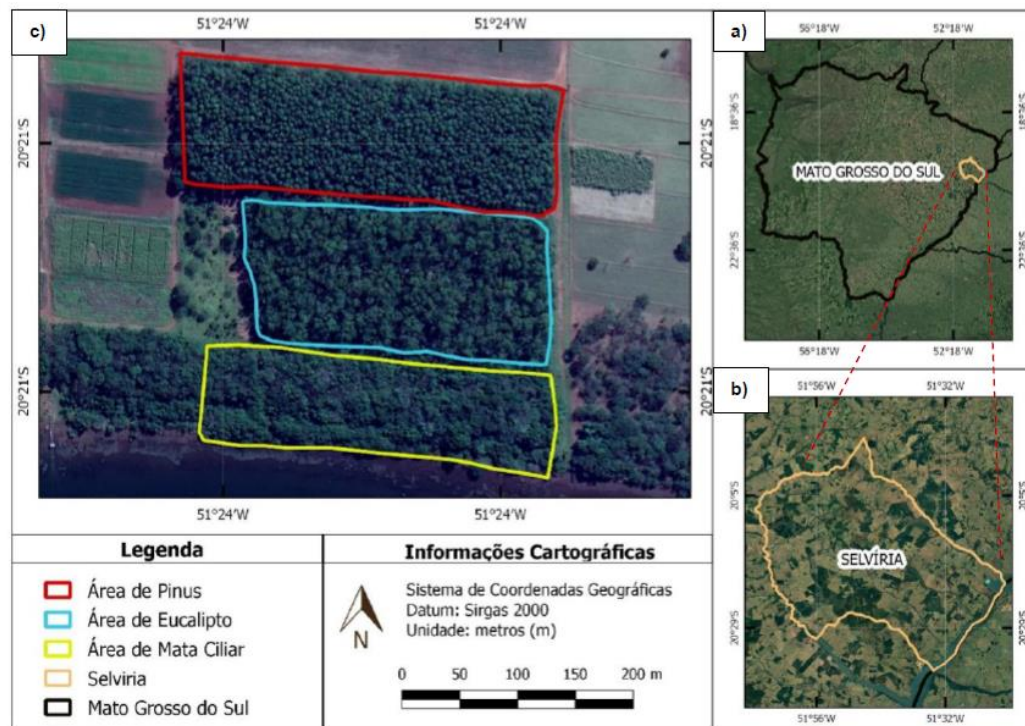


Figura 2. Mapa das áreas de estudo localizadas no município de Selvíria, MS. a) Estado do Mato Grosso do Sul, Brasil Central; b) Município de Selvíria; c) Áreas de florestas plantadas: pinus, eucalipto e mata ciliar, respectivamente.

A Tabela 2 apresenta informações compiladas e organizadas referentes aos experimentos que foram utilizados para fazer a coleta dos dados de emissão de CO₂ do solo.

Tabela 2 Dados sobre o experimento.

Experimento	Data do experimento	Preparo	Revol.	Data de preparo	Data de conversão	Cobertura	Cultura	Longitude	Latitude	Município/Estado	Prof. (m)	n
Espacial	02/10 a 25/11/2015	Reflorestamento	Não	01/03/1986	01/03/1986	Coberto	Eucalipto	458447.46	7749821.85	Selvíria/MS	0-0.1	102
Temporal	07/11 a 18/12/2015	Reflorestamento	Não	01/03/1986	01/03/1986	Coberto	Eucalipto	458447.46	7749821.85	Selvíria/MS	0-0.1	25
							Pinus					
							Mata ciliar					
	20/01 a 06/05/2016	Reflorestamento	Não	01/03/1986	01/03/1986	Coberto	Eucalipto	458447.46	7749821.85	Selvíria/MS	0-0.1	25
							Pinus					
							Mata ciliar					
	29/03 a 24/10/2017	Natural	Não	01/03/1986	01/03/1986	Coberto	Cerrado	458447.46	7749821.85	Selvíria/MS	0-0.1	15
		Reflorestamento	Não	01/03/1986	01/03/1986	Coberto	Eucalipto	458447.46	7749821.85	Selvíria/MS	0-0.1	20
							Pinus					
Espacial	03/02 a 17/06/2017	Reflorestamento	Não	01/03/1986	01/03/1986	Coberto	Eucalipto	458447.46	7749821.85	Selvíria/MS	0-0.1	102
							Silvipastoril					86
Temporal	29/03 a 24/10/2017	Natural	Não	01/03/1986	01/03/1986	Coberto	Cerrado	458447.46	7749821.85	Selvíria/MS	0-0.1	15
		Reflorestamento	Não	01/03/1986	01/03/1986	Coberto	Eucalipto	458447.46	7749821.85	Selvíria/MS	0-0.1	20
							Pinus					
							Silvipastoril					

A imagem a seguir foi gerada usando a linguagem R e o banco de dados (https://arpanosso.github.io/fco2r/reference/oco2_br.html), demonstra com os pontos na cor azul os dados de CO₂ atmosférico, medidas pela taxa fotossintética nas frequências de 771 nm e 757 nm.

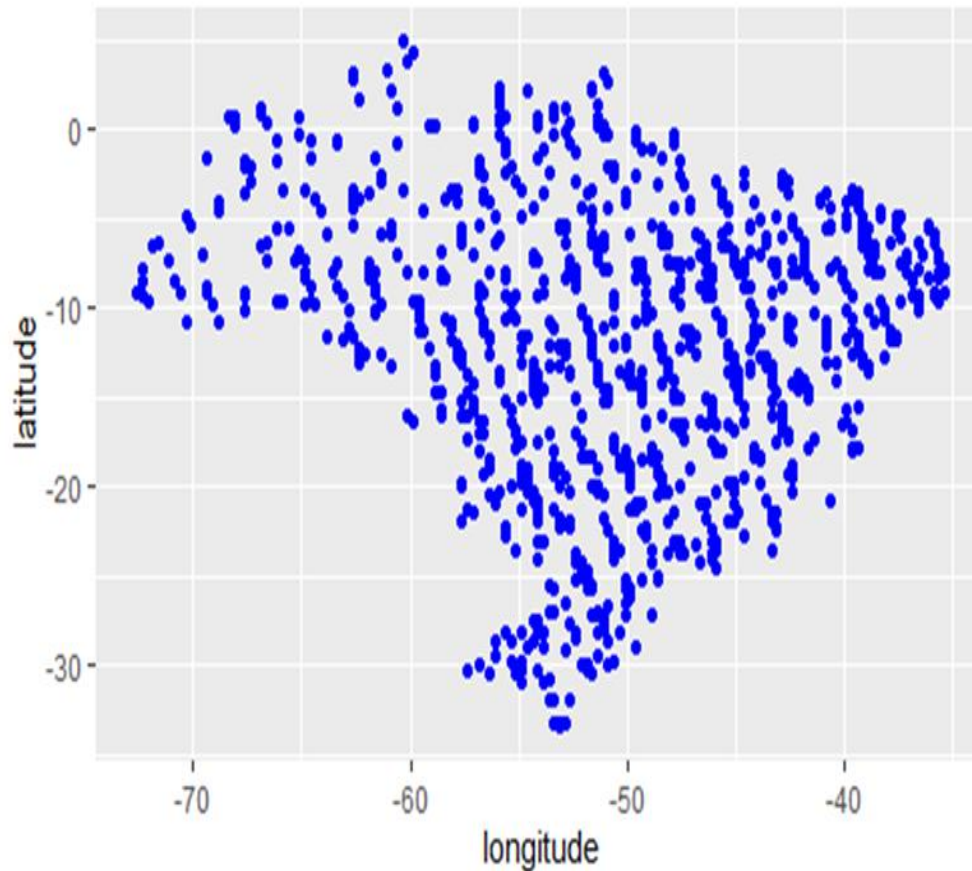


Figura 3. Dados de CO₂ atmosférico no Brasil.

Com base nos dados obtidos e compilados de CO₂ atmosférico, a Figura 4 mostra os pontos que corresponde aos dados limitando-os dentro do território brasileiro, tornando a visualização gráfica mais agradável e didática.

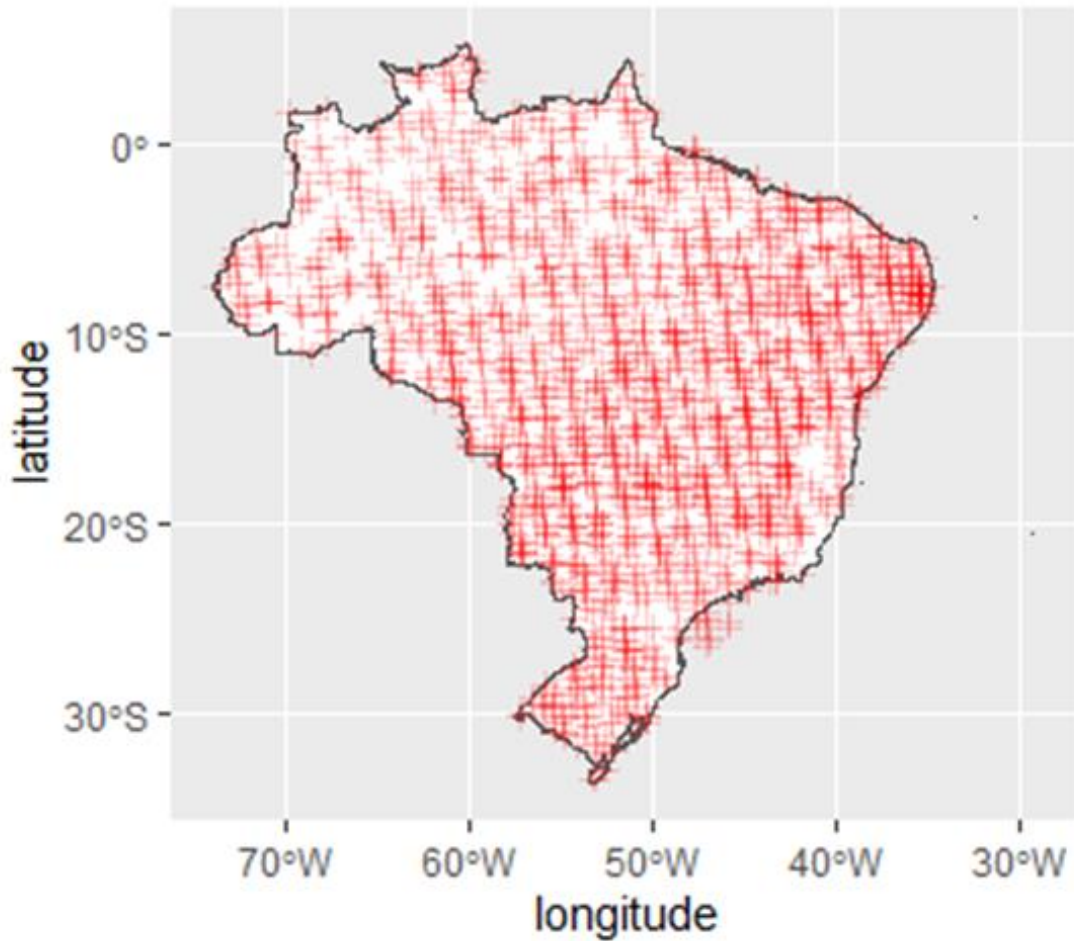


Figura 4. Dados de CO₂ atmosférico dentro do território brasileiro.

Os dados compilados de emissão de CO₂ do solo entre 2001 a 2020, variam de acordo com as datas e culturas, e está em exibição na Figura 5.

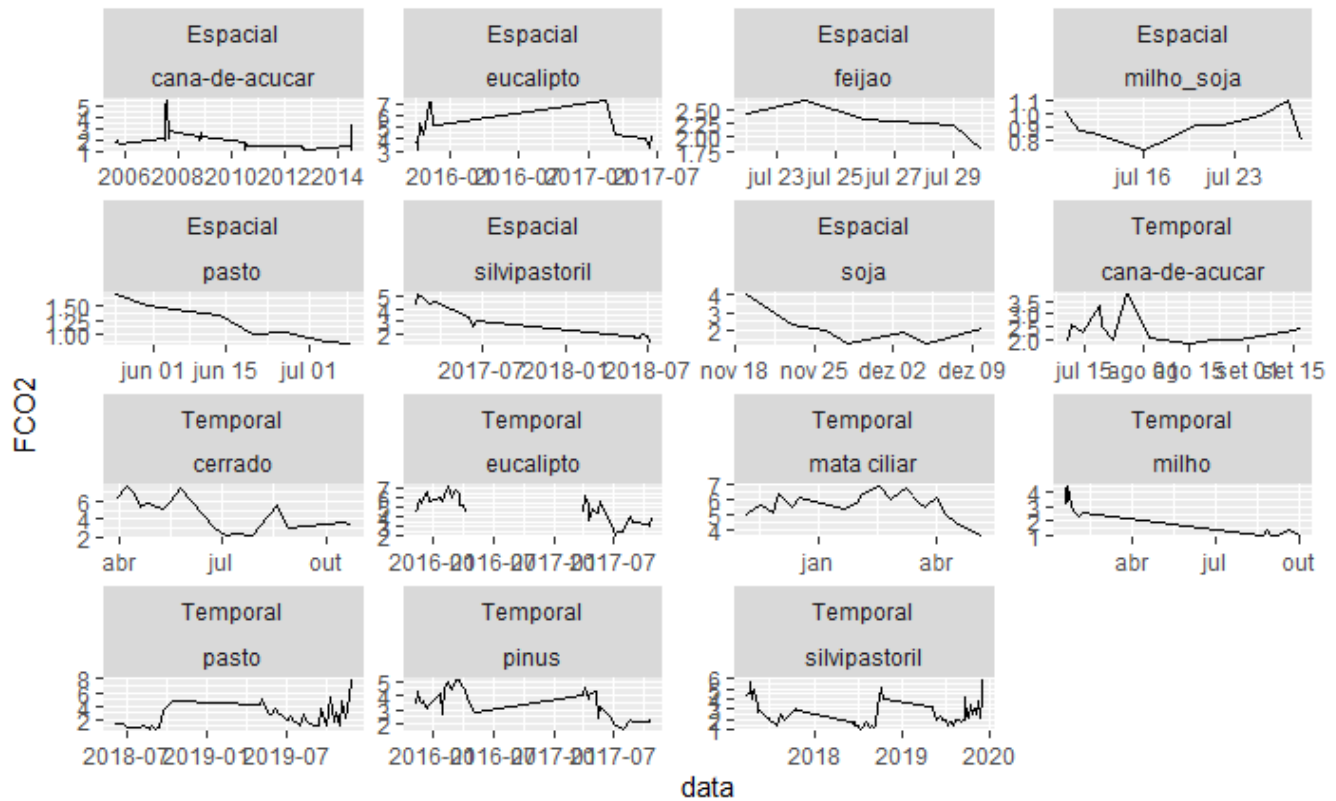


Figura 5. Emissão de CO₂ do solo entre 2001 e 2020.

A Figura 6 mostra o gráfico que apresenta os dados referentes ao CO₂ do solo, tendo a limitação do território brasileiro.

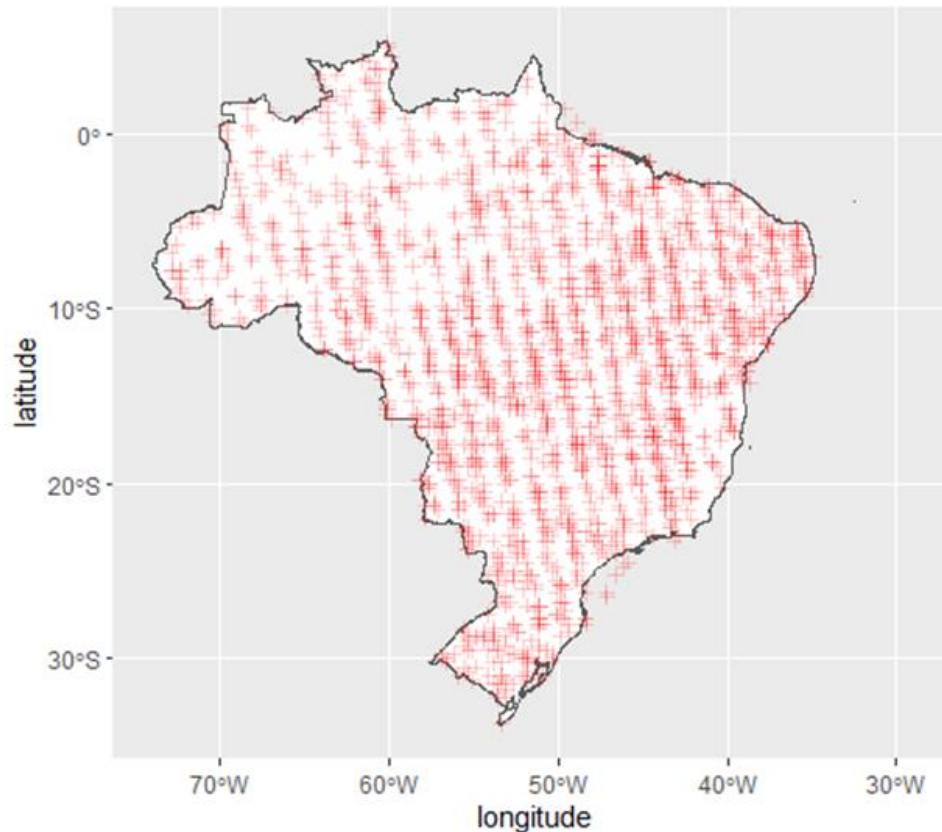


Figura 6. Emissão de CO₂ do solo.

8 REFERÊNCIAS BIBLIOGRÁFICAS

BRASIL. (2019). ESTIMATIVAS ANUAIS DE EMISSÕES DE GASES DE EFEITO ESTUFA NO BRASIL. Fonte: gov.br: https://www.gov.br/mcti/pt-br/acompanhe-o-mcti/sirene/publicacoes/estimativas-anuais-de-emissoes-gee/arquivos/livro_digital_5ed_estimativas_anuais.pdf

CHAPMAN, P.; CLINTON, J.; KERBER, R.; KHABAZA, T.; REINARTZ, T.; SHEARER, C. R.; WIRTH, R., 2000. CRISP-DM 1.0: Step-by-step data mining guide.

LUO, Y.; ZHOU, X., 2006. Soil respiration and the environment.

R DEVELOPMENT CORE TEAM. **A language and environment for statistical computing**. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org/>.