

PIBIC/Reitoria/CNPq/UNESP 2021/2022
EDITAL 04/2021 – PROPe
Edital 02/2021 PROPe/COPE

Plano de Atividades do Aluno

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

Maio/2021

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

1) INTRODUÇÃO

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 é avaliar o desempenho preditivo dos principais algoritmos de aprendizado de máquina para emissão de CO₂ do solo (FCO₂) em diferentes usos e manejos no Brasil central (estados de Mato Grosso do Sul (MS) e São Paulo (SP)), a partir de uma série temporal de 2015 a 2020 para os dados de XCO₂ e SIF do Observatório do Carbono em Órbita-2 (OCO-2), índices climáticos obtidos pela NASA-POWER.

As técnicas utilizadas serão: redes neurais artificiais (RNA), máquina de vetores de suporte (SVM) sistema de inferência Neuro-Fuzzy (ANFIS) 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 (R^2), 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.

2) OBJETIVOS

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 serem modeladas por algoritmos de aprendizado de máquina. Assim, o objetivo 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: redes neurais artificiais (RNA), máquina de vetores de suporte (SVM) sistema de inferência Neuro-Fuzzy (ANFIS) 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. Espera-se que a proposta, além de impactar na formação do bolsista, traga avanços para a geração de produtos e processos com potencial de auxílio na tomada de decisão sobre mitigação das emissões de CO₂ no Brasil.

4) METODOLOGIA

A presente proposta irá compilar achados anteriores e alguns resultados ainda não publicados sobre a emissão de CO₂ do solo (FCO₂) a partir de 9 ensaios realizados em diferentes regiões do sudeste e centro-oeste do Brasil, conduzidos nos últimos 7 anos. 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.

Em todos os ensaios a emissão de CO₂ do solo foi registrada por meio do sistema LI-COR (LI-8100), que em seu modo de medição o sistema monitora as mudanças na concentração de CO₂ dentro da câmara para solos por meio de espectroscopia na região do infravermelho. 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.

Os dados referentes aos atributos físico e químicos do solo foram determinados na profundidade de 0 – 0,10 e 0,10-0,20 m. Foram realizadas as seguintes análises de rotina: pH, determinação do teor de matéria orgânica do solo (MO), fósforo disponível (P), potássio (K), Cálcio (Ca), Magnésio (Mg), soma de bases (SB), acidez potencial (H+Al). Os teores de cálcio, magnésio e potássio trocáveis e fósforo disponível serão extraídos utilizando-se o método da resina trocadora de íons (RAIJ, 2001); a capacidade de troca de cátions (CTC) e a saturação por base (V%) também serão calculados. O carbono orgânico será pelo método da combustão úmida, via colorimétrica (RAIJ et al., 1987), a determinação do teor de nitrogênio será realizada por meio de digestão sulfúrica (MALAVOLTA; VITTI; OLIVEIRA, 1997). As amostras indeformadas foram coletadas com amostrador adaptado a cilindros com dimensões médias de 0,05 m de diâmetro interno e 0,04 m de altura (EMBRAPA, 1997). A macroporosidade (Ma) e microporosidade (Mi) foram determinadas por meio de mesa de tensão com 0,60 m de altura de coluna d'água em amostras previamente saturadas. O volume de água retido na amostra nesta condição corresponde à Ma. A Mi foi determinada após a retirada dos anéis da estufa à 105 °C num período de 24 h e posterior pesagem. Já a porosidade total foi calculada pela soma dos macroporos e microporos. Também foi determinado a densidade do solo (Ds) utilizando-se metodologia de EMBRAPA (1997).

Todos os modelos de aprendizado de máquina e etapas de pré-processamento e validação dos modelos serão realizados a partir de implementações na linguagem R (R DEVELOPMENT CORE TEAM, 2020), 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.

3) CRONOGRAMA DE EXECUÇÃO

O cronograma de execução (Tabela 1) juntamente com o plano de atividade do acadêmico de iniciação científica será composto das seguintes fases: **Revisão de literatura**). Ficará a cargo do bolsista, com supervisão do Orientador, realizar a revisão bibliográfica sobre os principais temas abordados no projeto referente ao aprendizado de máquina, a saber: redes neurais e "deep learning". Será utilizado o acervo de periódicos disponível na web por meio de ferramentas de buscas na internet (Portal de Periódicos CAPES, *Web of Knowledge* e *Scopus*); **Análise dos dados experimentais**). Atualmente o aluno já está recebendo treinamentos específicos no programa R e pacotes computacionais. O candidato, demonstra conhecimento na área de Estatística e Experimentação, e com auxílio do nosso grupo de pesquisa, o aluno vem demonstrando excelente desempenho na disciplina e nas práticas de programação. **Produção de Relatório Científico**). Ficará sobre a responsabilidade do bolsista a produção e entrega dos relatórios científicos: parcial e final; além da construção da base de dados e disponibilização pública de *web app* para previsão de FCO₂ em áreas agrícolas no Brasil. **Publicação de trabalho científico**). Ficará sobre a responsabilidade do bolsista a apresentação dos resultados em congresso ou qualquer outro evento científico da área e, com o auxílio do orientador, espera-se a publicação em periódicos científicos. Gostaríamos de salientar que as práticas de programação das análises no R serão utilizadas como ferramenta didática para desenvolver a estruturação do pensamento lógico e da linguagem computacional no aluno bolsista, complementando assim, a sua formação acadêmica e profissional.

Tabela 1. Cronograma de execução das atividades do bolsista de iniciação científica, período de 12 meses, com implementação da bolsa a partir de setembro de 2021 e entrega do relatório final em agosto de 2022.

ATIVIDADES	PERÍODO: (ANO) / - TRIMESTRE -			
	2021		2022	
	- 3º -	- 4º -	- 1º -	- 2º -
Revisão de literatura				X
Avaliação da emissão de CO ₂ , captura de O ₂ temperatura e umidade do solo	X	X	X	
Preparo do Banco de dados para mineração e incorporação de variáveis ambientais	X	X		
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		X		X
Atualização de disponibilização da base de dados			X	X

6) REFERÊNCIAS

BESALATPOUR, A. A.; AYOUBI, S.; HAJABBASI, M. A.; MOSADDEGHI, M. R.; SCHULIN, R. Estimating wet soil aggregate stability from easily available properties in a highly mountainous watershed. *Catena*, v. 111, n., p. 72-79, 2013.

- BICALHO, E. S.; PANOSSO, A. R.; TEIXEIRA, D. D. B.; MIRANDA, J. G. V.; PEREIRA, G. T.; LA SCALA, N. Spatial variability structure of soil CO₂ emission and soil attributes in a sugarcane area. **Agriculture Ecosystems & Environment**, v. 189, n., p. 206-215, 2014.
- BRASIL. **Brasil Projeções do Agronegócio 2017/2018 a 2027/2028**. Brasília: Ministério da Agricultura, Pecuária e Abastecimento, 2018. p.
- BRASIL. **Estimativas anuais de emissões de gases do efeito estufa no Brasil**. Brasília: Ministro da Ciência, Tecnologia, Inovações e Comunicações, 2019. p.
- BRUCE, P.; BRUCE, A. **Practical Statistics for Data Scientists**. United States of America: O'Reilly Media, Inc., 2017. 562 p.
- CARVALHO, J. L. N.; CERRI, C. E. P.; FEIGL, B. J.; PICCOLO, M. C.; GODINHO, V. P.; CERRI, C. C. Carbon sequestration in agricultural soils in the Cerrado region of the Brazilian Amazon. **Soil & Tillage Research**, v. 103, n. 2, p. 342-349, 2009.
- 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.
- CHEN, G. Q.; CHEN, Z. M. Carbon emissions and resources use by Chinese economy 2007: A 135-sector inventory and input-output embodiment. **Communications in Nonlinear Science and Numerical Simulation**, v. 15, n. 11, p. 3647-3732, 2010.
- CHIAVEGATTO FILHO, A. D. P.; DOS SANTOS, H. G.; DO NASCIMENTO, C. F.; MASSA, K.; KAWACHI, I. Overachieving Municipalities in Public Health: A Machine-learning Approach. **Epidemiology**, v. 29, n. 6, p. 836-840, 2018.
- CHOLLET, F.; ALLAIRE, J. J. **Deep Learning with R**. 2017. 341 p.
- DI MININ, E.; FINK, C.; TENKANEN, H.; HIIPALA, T. Machine learning for tracking illegal wildlife trade on social media. **Nature Ecology & Evolution**, v. 2, n. 3, p. 406-407, 2018.
- ELLERT, B. H.; BETTANY, J. R. Calculation of organic matter and nutrients stored in soils under contrasting management regimes. **Canadian Journal of Soil Science**, v. 75, n. 4, p. 529-538, 1995.
- EMBRAPA. **Manual de métodos de análise de solo**. 2 ed. Brasília: Ministério da Agricultura e do Abastecimento / EMBRAPA-CNPq, 1997. 212 p.
- FARHATE, C. V. V.; DE SOUZA, Z. M.; OLIVEIRA, S. R. D.; TAVARES, R. L. M.; CARVALHO, J. L. N. Use of data mining techniques to classify soil CO₂ emission induced by crop management in sugarcane field. **Plos One**, v. 13, n. 3, p., 2018.
- GE, Z.; SONG, Z.; DING, S. X.; HUANG, B. Data Mining and Analytics in the Process Industry: The Role of Machine Learning. **IEEE Access**, v. 5, n., p. 20590-20616, 2017.
- GRAF, A.; HERBST, M.; WEIHERMULLER, L.; HUISMAN, J. A.; PROLINGHEUER, N.; BORNEMANN, L.; VEREECKEN, H. Analyzing spatiotemporal variability of heterotrophic soil respiration at the field scale using orthogonal functions. **Geoderma**, v. 181, n., p. 91-101, 2012.
- KHAN, M. Z.; KHAN, M. F. Application of ANFIS, ANN and fuzzy time series models to CO₂ emission from the energy sector and global temperature increase. **International Journal of Climate Change Strategies and Management**, v. 11, n. 5, p. 622-642, 2019.
- KUHN, M.; JOHNSON, K., 2013. Applied predictive modeling.
- LAGANIERE, J.; ANGERS, D. A.; PARE, D. Carbon accumulation in agricultural soils after afforestation: a meta-analysis. **Global Change Biology**, v. 16, n. 1, p. 439-453, 2010.
- LEON, E.; VARGAS, R.; BULLOCK, S.; LOPEZ, E.; PANOSSO, A. R.; LA SCALA JR, N. Hot spots, hot moments, and spatio-temporal controls on soil CO₂ efflux in a water-limited ecosystem. **Soil Biology and Biochemistry**, v. 77, n. 0, p. 12-21, 2014.
- MALAVOLTA, E.; VITTI, G. C.; OLIVEIRA, A. S. **Avaliação do estado nutricional das plantas: princípios e aplicações**. Piracicaba: Potafós, 1997. p.
- MCBRATNEY, A. B.; ODEH, I. O. A.; BISHOP, T. F. A.; DUNBAR, M. S.; SHATAR, T. M. An overview of pedometric techniques for use in soil survey. **Geoderma**, v. 97, n. 3-4, p. 293-327, 2000.
- MORAES, J. F. L.; VOLKOFF, B.; CERRI, C. C.; BERNOUX, M. Soil properties under Amazon forest and changes due to pasture installation in Rondonia, Brazil. **Geoderma**, v. 70, n. 1, p. 63-81, 1996.
- 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/>.

- RAICH, J. W.; SCHLESINGER, W. H. The global carbon-dioxide flux in soil respiration and its relationship to vegetation and climate. **Tellus Series B-Chemical and Physical Meteorology**, v. 44, n. 2, p. 81-99, 1992.
- RAIJ, B. V. **Análise química para avaliação da fertilidade de solos tropicais**. Campinas: Instituto Agrônomo, 2001. 285 p.
- RAIJ, B. V.; DE ANDRADE, J. C.; CANTARELLA, H.; QUAGGIO, J. A. **Análise química do solo para fins de fertilidade**. Campinas: Fundação Cargill, 1987. 170 p.
- REICHSTEIN, M.; CAMPS-VALLS, G.; STEVENS, B.; JUNG, M.; DENZLER, J.; CARVALHAIS, N.; PRABHAT Deep learning and process understanding for data-driven Earth system science. **Nature**, v. 566, n. 7743, p. 195-204, 2019.
- SANTOS, G. A. D.; MOITINHO, M. R.; SILVA, B. D.; XAVIER, C. V.; TEIXEIRA, D. D.; CORA, J. E.; LA SCALA, N. Effects of long-term no-tillage systems with different succession cropping strategies on the variation of soil CO₂ emission. **Science of the Total Environment**, v. 686, n., p. 413-424, 2019a.
- SANTOS, H. G. D.; NASCIMENTO, C. F. D.; IZBICKI, R.; DUARTE, Y. A. D. O.; PORTO CHIAVEGATTO FILHO, A. D. Machine learning para análises preditivas em saúde: exemplo de aplicação para prever óbito em idosos de São Paulo, Brasil. **Cadernos de Saúde Pública**, v. 35, n., p., 2019b.
- SEEG. **ANÁLISE DAS EMISSÕES BRASILEIRAS DE GASES DE EFEITO ESTUFA E SUAS IMPLICAÇÕES PARA AS METAS DE CLIMA DO BRASIL 1970-2019**. Observatório do Clima, 2020. 40 p.
- SEGNINI, A.; CARVALHO, J. L. N.; BOLONHEZI, D.; MILORI, D.; DA SILVA, W. T. L.; SIMOES, M. L.; CANTARELLA, H.; DE MARIA, I. C.; MARTIN-NETO, L. Carbon stock and humification index of organic matter affected by sugarcane straw and soil management. **Scientia Agricola**, v. 70, n. 5, p. 321-326, 2013.
- SHADRIN, D.; PUKALCHIK, M.; KOVALEVA, E.; FEDOROV, M. Artificial intelligence models to predict acute phytotoxicity in petroleum contaminated soils. **Ecotoxicology and Environmental Safety**, v. 194, n., p., 2020.
- SIERRA, C. A.; TRUMBORE, S. E.; DAVIDSON, E. A.; VICCA, S.; JANSSENS, I. Sensitivity of decomposition rates of soil organic matter with respect to simultaneous changes in temperature and moisture. **Journal of Advances in Modeling Earth Systems**, v. 7, n. 1, p. 335-356, 2015.
- THANGAVEL, R.; KANCHIKERIMATH, M.; SUDHARSANAM, A.; AYYANADAR, A.; KARUNANITHI, R.; DESHMUKH, N. A.; VANAIO, N. S. Evaluating organic carbon fractions, temperature sensitivity and artificial neural network modeling of CO₂ efflux in soils: Impact of land use change in subtropical India (Meghalaya). **Ecological Indicators**, v. 93, n., p. 129-141, 2018.
- TRAMONTANA, G.; JUNG, M.; SCHWALM, C. R.; ICHII, K.; CAMPS-VALLS, G.; RADULY, B.; REICHSTEIN, M.; ARAIN, M. A.; CESCATTI, A.; KIELY, G.; MERBOLD, L.; SERRANO-ORTIZ, P.; SICKERT, S.; WOLF, S.; PAPALE, D. Predicting carbon dioxide and energy fluxes across global FLUXNET sites with regression algorithms. **Biogeosciences**, v. 13, n. 14, p. 4291-4313, 2016.
- VICENTINI, M. E.; PINOTTI, C. R.; HIRAI, W. Y.; DE MORAES, M. L. T.; MONTANARI, R.; FILHO, M. C. M. T.; MILORI, D. M. B. P.; JÚNIOR, N. L. S.; PANOSSO, A. R. CO₂ emission and its relation to soil temperature, moisture, and O₂ absorption in the reforested areas of Cerrado biome, Central Brazil. **Plant and Soil**, v., n., p., 2019.
- WARNER, D. L.; BOND-LAMBERTY, B.; JIAN, J.; STELL, E.; VARGAS, R. Spatial Predictions and Associated Uncertainty of Annual Soil Respiration at the Global Scale. **Global Biogeochemical Cycles**, v. 33, n. 12, p. 1733-1745, 2019.
- WARNER, D. L.; VARGAS, R.; SEYFFERTH, A.; INAMDAR, S. Transitional slopes act as hotspots of both soil CO₂ emission and CH₄ uptake in a temperate forest landscape. **Biogeochemistry**, v. 138, n. 2, p. 121-135, 2018.
- WMO. **WMO Greenhouse Gas Bulletin (GHG Bulletin) - No. 16: The State of Greenhouse Gases in the Atmosphere Based on Global Observations through 2019**. World Meteorological Organization, World Data Centre for Greenhouse Gases, Tokyo, Japan. https://library.wmo.int/doc_num.php?explnum_id=10437.
- YILMAZ, I.; KAYNAR, O. Multiple regression, ANN (RBF, MLP) and ANFIS models for prediction of swell potential of clayey soils. **Expert Systems with Applications**, v. 38, n. 5, p. 5958-5966, 2011.