

RESSALVA

Atendendo solicitação do(a)
autor(a), o texto completo desta tese
será disponibilizado somente a partir
de 09/05/2024.

UNIVERSIDADE ESTADUAL PAULISTA – UNESP

CÂMPUS DE JABOTICABAL

**ESTIMATIVA DE INDICADORES DE QUALIDADE DO SOLO
POR MEIO DE MODELOS DE “MACHINE LEARNING”**

Nayane Jaqueline Costa Maia

Engenheira Agrônoma

Ma. em Agronomia (Ciência do Solo)

2022

UNIVERSIDADE ESTADUAL PAULISTA – UNESP
CÂMPUS DE JABOTICABAL

**ESTIMATIVA DE INDICADORES DE QUALIDADE DO SOLO
POR MEIO DE MODELOS DE “MACHINE LEARNING”**

Discente: Nayane Jaqueline Costa Maia

Orientadora: Prof. Dr. Glauco de Souza Rolim

Coorientadora: Dra. Flávia Fernanda Simili

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

M217e

Maia, Nayane Jaqueline Costa

Estimativa de indicadores de qualidade do solo por meio de modelos de "Machine Learning" / Nayane Jaqueline Costa
Maia. -- Jaboticabal, 2022

99 p. : il., tabs., 2 v. + 1 CD-ROM

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

Orientador: Glauco de Souza Rolim

Coorientadora: Flávia Fernanda Simili

1. Ciência do Solo. 2. Modelagem do Solo. 3. Inteligência Artificial. 4. Integração Lavoura-Pecuária. I. Título.

Sistema de geração automática de fichas catalográficas da Unesp. Biblioteca da Faculdade de Ciências Agrárias e Veterinárias, Jaboticabal. Dados fornecidos pelo autor(a).

Essa ficha não pode ser modificada.



UNIVERSIDADE ESTADUAL PAULISTA

Câmpus de Jaboticabal

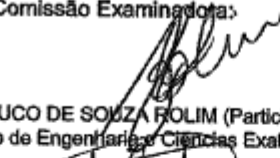



CERTIFICADO DE APROVAÇÃO


TÍTULO DA TESE: ESTIMATIVA DE INDICADORES DE QUALIDADE DO SOLO ESTIMADOS POR MEIO DE MODELOS DE MACHINE LEARNING

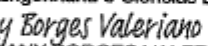
AUTORA: NAYANE JAQUELINE COSTA MAIA
ORIENTADOR: GLAUCO DE SOUZA ROLIM
COORIENTADORA: FLÁVIA FERNANDA SIMILI

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


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


Pesquisadora Dra. MARA REGINA MOITINHO (Participação Presencial)
CNPEM / Campinas/SP


Prof. Dr. ROGÉRIO TEIXEIRA DE FÁRIA (Participação Presencial)
Departamento de Engenharia e Ciências Exatas DECEEx / FCAV UNESP Jaboticabal


Dra. TAYNARA TUANY BORGES VALERIANO (Participação Presencial)
Instituição Bayer S.A. / Agronomic Predictive Modeler at Bayer CropScience


Dr. DIEGO SILVA SIQUEIRA (Participação Presencial)
Agrônomo Autônomo - Núcleo de Inovação Tecnológica (NIT) - Supera Parque / Ribeirão Preto/SP

Jaboticabal, 09 de novembro de 2022

DADOS CURRICULARES DA AUTORA

NAYANE JAQUELINE COSTA MAIA – nascida em 23 de novembro de 1993, na cidade de Castanhal – PA. Concluiu o ensino médio integrado ao curso Técnico em Agropecuária (2008-2011), pelo Instituto Federal de Educação, Ciência e Tecnologia (IFPA), Câmpus de Castanhal. É Engenheira Agrônoma formada pelo IFPA-Câmpus Castanhal (2012-2017). Durante a graduação foi integrante do Núcleo de Pesquisa e Estudos Agropecuários (NUPAGRO) e do Núcleo de Pesquisa em Ciência do Solo e Água da Amazônia (NUPECSA). Em 2012 foi bolsista de iniciação científica do Departamento de Zootecnia do IFPA-Câmpus Castanhal, com pesquisas voltadas para ganho de peso animal usando resíduos de frutas da Amazônia. Em 2013 foi bolsista da FAPESPA/CNPq, no projeto em que foram avaliados indicadores da qualidade do solo em áreas sob vegetação natural e cultivos na Amazônia. Em 2014 foi monitora do Laboratório de Solos e Plantas, do Departamento de Solos e Adubos do IFPA-Câmpus Castanhal. Em 2015 foi bolsista de iniciação científica do Departamento de Ciência do Solo, com pesquisa sobre indicadores de fertilidade do solo na cultura do açaí (*Euterpe oleracea*) em áreas de várzea e de terra firme. Nos anos de 2017 e 2018, foi bolsista de mestrado (CAPES) no Programa de Pós-Graduação em Agronomia (Ciência do Solo) da FCAV-Unesp, no departamento de Solos e Adubos e durante esse período participou do projeto da FAPESP “Impacto ambiental, produtividade e viabilidade econômica de sistemas convencional ou integrado de lavoura pecuária” no Instituto de Zootecnia de Sertãozinho-SP. Nos anos de 2019 a 2021 foi bolsista de doutorado (CAPES) no Programa de Pós-Graduação em Agronomia (Ciência do Solo) da FCAV-Unesp, no departamento de Engenharia e Ciências Exatas. No ano de 2021 passou no concurso público para professora substituta de Agronomia no IFMS-Campus Naviraí. Desde 2021 vem trabalhando na área de ciência de dados e modelagem preditiva no setor privado, com otimização da produtividade no campo sucroenergético.

“Nós somos, de fato, aquilo que escolhemos e as consequências que assumimos.”

Mario Sergio Cortella

Aos meus pais Nelma Maia e Jesus Maia pela coragem de mudar de cidade, que mudaria tudo para nós a partir de então. Ao meu querido amigo Washington Pereira – *In memoriam*, por me ensinar que a vida é uma só, e que todo dia ela precisa ser vivida intensamente (“*Abre a porta do teu ser, sinta o vento te soprar*”).

DEDICO

AGRADECIMENTOS

À Deus, por proporcionar esta oportunidade e me entregar tanta coragem para chegar até aqui sozinha.

À minha família, em especial aos meus pais, Jesus Maia e Nelma Maia, e aos meus irmãos, especialmente a Nayana Maia, que não mediram esforços para me ajudar; devo a eles tudo o que tenho e o que sou.

Ao professor Dr. Glauco de Souza Rolim, por me apresentar um novo mundo (ciência de dados), por me apoiar na execução desse trabalho e na minha carreira fora da academia.

À Pesquisadora Dra. Flávia Fernanda Simili, por me apoiar com a continuação do projeto e o uso dos dados do seu projeto, e na grande parceria de redação científica que construímos nos últimos anos.

À Professora Dra. Mara Cristina Pessôa da Cruz, pelo carinho e todo aprendizado repassado durante o mestrado no laboratório de fertilidade do solo da FCAV/Unesp.

Ao Conselho do Programa de Pós-Graduação em Agronomia (Ciência do Solo). O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de financiamento 001.

Ao Instituto de Zootecnia de Sertãozinho-SP, pela instalação do projeto. À Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), pelos recursos e financiamento do projeto (Processo Nº 2014/24514-6).

Ao Laboratório de Fertilidade do Solo da FCAV/Unesp – Campus de Jaboticabal onde foram realizadas as análises. E ao Group of Agrometeorological Studies (GAS). A todos os amigos que fiz nesses grupos, por me auxiliarem e compartilharem o aprendizado comigo.

Aos amigos que fiz na cidade de Jaboticabal-SP e São José do Rio Preto-SP, pelo carinho, paciência e toda ajuda.

A todos que contribuíram para a realização desse trabalho.

Muito obrigada!

ESTIMATIVA DE INDICADORES DE QUALIDADE DO SOLO POR MEIO DE MODELOS DE “MACHINE LEARNING”

RESUMO – O monitoramento de indicadores de qualidade do solo contribui para mitigar as mudanças climáticas globais sem diminuir a produtividade dos sistemas, além de avaliar a eficiência de sistemas sustentáveis e complexos. O uso de modelos de “machine learning” no atual cenário do aumento dos dados é necessário para que as informações dos sistemas agrícolas sejam mais colaborativas, de maneira que, diminua custos e se traduza em otimização dos sistemas, tornando-os mais sustentáveis. Com este trabalho objetivou-se testar modelos de machine learning para estimar indicadores de qualidade do solo, a fim de melhorar o monitoramento dos solos agrícolas em sistema de monocultivo de milho, de pecuária e de Integração Lavoura-Pecuária, avaliando: 1) as combinações de variáveis para estimar o carbono e nitrogênio orgânico solo; e 2) variáveis que impactam a disponibilidade do nitrogênio potencialmente mineralizável no solo. Os modelos de machine learning testados foram Multilayer Perceptron Regressor (MLP), Random Forest Regressor (RF), K Neighbors Regressor (KNN), Support Vector Regression (SVR), Multiple Linear Regression (MLR), Adaptive Boosting Regressor (AdaBoost) e eXtreme Gradient Boosting Regressor (XGBoost), usando a linguagem de programação Python. Os modelos utilizados se mostraram eficientes para estimar carbono e nitrogênio do solo. As variáveis combinadas de plantas-animais-solo-clima estimaram com boa acurácia e precisão o carbono e nitrogênio orgânico do solo. No sistema de monocultivo de Milho, os modelos AdaBoost e MLP obtiveram alta precisão e $MAPE < 1\%$. No sistema monocultivo de Pecuária, os modelos MLP, SVR e Adaboost foram mais precisos e acurados ($MAPE < 3\%$). No sistema de integração Lavoura-Pecuária, todos os modelos tiveram elevada acurácia ($MAPE < 5\%$), no entanto, o modelo SVR obteve a menor precisão para estimar carbono orgânico do solo ($R^2 < 2\%$), e o SVR e KNN obtiveram a menor precisão para estimar o nitrogênio orgânico do solo ($R^2 < 50\%$). Para estimar o nitrogênio potencialmente mineralizável no solo, o modelo XGBoost foi o mais preciso, com o menor acurácia e viés ($R^2 = 0.97$, $MAPE = 3\%$ e $MBE = 0.10 \text{ mg kg}^{-1}$), superando o AdaBoost, RF e MLR, nessa ordem. As diferentes combinações de variáveis estimadoras podem indicar processos importantes que influenciam na liberação de nitrogênio potencialmente mineralizável no solo, por meio de métodos de explicações aditivas de Shapley (SHAP). No geral, esses resultados fornecem uma nova perspectiva com o uso de aplicação de machine learning para estimar importantes nutrientes do solo, aproveitando os mais diversos históricos de dados de sistemas agrícolas, que podem ser úteis para os tomadores de decisão na produção de alimentos.

“Palavras-chave:” combinação de variáveis, carbono orgânico, nitrogênio orgânico, nitrogênio potencialmente mineralizável, inteligência artificial, agricultura 4.0.

ESTIMATION OF SOIL QUALITY INDICATORS WITH MACHINE LEARNING MODELS

ABSTRACT - Monitoring soil quality indicators contribute to mitigating global climate change without reducing the influence of systems, in addition to measuring the efficiency of systems and complexes. The use of "machine learning" models in the current scenario of the data age is necessary for information from agricultural systems to be more collaborative, in a way that reduces costs and translates into the optimization of systems, making them more intelligent. The objective of this work was to test machine learning models to estimate soil quality indicators, to improve the monitoring of agricultural soils in a corn and pasture monoculture system and Crop-Livestock Integration system, evaluating: 1) a combination of variables to estimate carbon and organic only; and 2) variables that impact the availability of mineralizable power in the soil. The machine learning models tested were Multilayer Perceptron Regressor (MLP), Random Forest Regressor (RF), K Neighbors Regressor (KNN), Support Vector Regression (SVR), Multiple Linear Regression (MLR), Adaptive Boosting Regressor (AdaBoost) and eXtreme Gradient Boosting Regressor (XGBoost), using the Python programming language. The models used are efficient to estimate soil carbon and tolerance. The combined plant-animal-soil-climate variables accurately and accurately estimated soil carbon and organics. In the corn monoculture system, the AdaBoost and MLP models achieved high accuracy and MAPE<1%. In the livestock monoculture system, the MLP, SVR, and Adaboost models were more precise and accurate (MAPE<3%). In the Crop-Livestock integration system, all models had high accuracy (MAPE<5%), however, the SVR model obtained a lower accuracy for estimating soil organic carbon ($R^2 < 2\%$), and the SVR and KNN obtained the lowest precision to estimate the soil organic ($R^2 < 50\%$). To estimate the electrically mineralizable in the soil, the XGBoost model was the most accurate, with the lowest accuracy and bias ($R^2 = 0.97$, MAPE = 3% and MBE = 0.10 mg kg⁻¹), surpassing AdaBoost, RF, and MLR, in that order. The different combinations of estimated variables may indicate important processes that influenced the release of potentially mineralizable material in the soil, using Shapley's induced additive methods (SHAP). Overall, these results provide a new perspective with the use of machine learning applications to estimate important soil nutrients, taking advantage of the most diverse historical data of agricultural systems, which can be useful for decision-makers in food production.

"Keywords:" combination of variables, organic carbon, organic nitrogen, potentially mineralizable nitrogen, artificial intelligence, agriculture 4.0.

2.1 INTRODUÇÃO

A recarbonização dos solos é uma solução global e vem demandando grandes pesquisas sendo de interesse mundial, no que tange o âmbito da agricultura sustentável, qualidade dos solos agrícolas, agricultura regenerativa, tecnologia aliada a agricultura sustentável, e tudo isso junto para diminuir o impacto da agricultura no ambiente, mitigar as mudanças climáticas globais e otimizam a produtividade dos sistemas. A Organização das Nações Unidas para Agricultura e Alimentação (FAO) lançou um programa de recarbonização de solos (RECSOIL), no qual sua mensagem principal identifica o aumento de carbono orgânico no solo (TOC) como uma das opções mais econômicas para a mitigação das mudanças climáticas, bem como para combater a desertificação, a degradação da terra e a insegurança alimentar (Campbell et al., 2018; FAO, 2019; Amelung et al., 2020).

A melhoria do manejo do solo pode resultar em grandes impactos a longo prazo, todos positivamente ligados aos objetivos de desenvolvimento sustentável (ODS). Entre esses manejos do solo, a reutilização de resíduos orgânicos e sua transformação em aditivos orgânicos é uma das estratégias de ODS, resultados positivos dessas estratégias ocorrem com eficiência em sistemas combinados de pecuária e produção de grãos (Oliveira et al., 2022). De acordo com McGuire et al. (2022), o aumento dos níveis de TOC no solo tem o potencial global de sequestro equivalentes a 5 a 10% das emissões globais de gases de efeito estufa. A fim de aproveitar totalmente o potencial do carbono dos solos, que dependendo do sistema de manejo o solo, pode armazenar (sumidouro) ou emitir (fonte) carbono na forma de CO_2 (C- CO_2).

Para avaliar e monitorar a qualidade dos sistemas agrícolas e pecuários, são necessárias análises laboratoriais dispendiosas e demoradas, e que geram grandes

quantidades de resíduos tóxicos no meio ambiente. O método mais consolidado atualmente para a determinação de TOC do solo é o de Walkley e Black (1934), que envolve análise com resíduos ricos em ácidos puros (H_2SO_4 , H_3PO_4) e dicromato de potássio ($K_2Cr_2O_7$). Esses resíduos quando descartados incorretamente, podem causar riscos ao meio ambiente (Ontañon et al., 2018) e são cancerígenos aos seres humanos (Wu et al., 2019). Com o intuito de reduzir esses resíduos, alguns pesquisadores desenvolveram estimativas de TOC a partir do uso de ondas magnéticas (Wang et al., 2016). Segundo Mahmoud et al. (2017), a principal desvantagem desse método é a necessidade de vários ajustes de refletância dos raios gama e a compra de equipamentos caros. Para a estimativa de nitrogênio orgânico total do solo (TON), é realizada usando sensores do solo, entretanto, é um processo oneroso devido ao alto custo desses sensores. (Zhang et al., 2019; Lin et al., 2020; Xu et al., 2021).

Para superar as restrições e limitações de custo e ambientais envolvidas nas estimativas de TOC e TON, uma alternativa viável para a estimativa são os algoritmos de machine learning (ML), que podem ser utilizados como uma ferramenta de estimativa ou como um complemento às abordagens químicas. Esse método possibilita simular e prever o teor de TOC e TON no solo a partir da combinação de variáveis oriundas de diferentes fontes em modelos para estimar as propriedades do solo. Modelos de redes neurais artificiais (Multilayer Perceptron-MLP), Multiple Linear Regression (MLR), e também Random Forest (RF) junto com Support Vector Regression (SVR) são os modelos mais utilizados para estimar o teor de TOC e TON do solo (Mahmoud et al., 2017; Emamgholizadeh et al., 2018; Reda et al., 2019; Lin et al., 2020; Mahmoud et al. 2020). Entretanto, o modelo K-Nearest Neighbor (KNN) também já foi utilizado para prever o teor de nitrogênio do solo (Grell et al., 2021). E o algoritmo Adaptive Boosting (AdaBoost) obteve alta acurácia para classificar solos (Pham et al., 2021) e alta precisão ($R^2=0.91$) para estimar o teor de matéria orgânica do solo (Wei et al., 2020).

Na Alemanha, Wiesmeier et al. (2014) combinou atributos do solo, precipitação e temperatura do ar anual para estimar estoque de carbono no solo, obtendo uma boa precisão usando o modelo RF. Em outro estudo, foram combinados atributos do solo e variáveis climáticas usando índices de vegetação e temperatura do ar anual, e a precisão da estimativa de TOC foi melhorada usando o RF (John et al., 2020). Os

estudos de Wiesmeier et al. (2014) e John et al. (2020) são as únicas tentativas de combinar variáveis ambientais com indicadores de qualidade do solo até agora, e indicam a importância de incluir preditores auxiliares, para representar a importância do clima e propriedades físicas e químicas do solo nos sistemas.

Apesar da aceitabilidade do ML para estimar TOC e TON, poucos estudos consideram a incorporação conjunta de indicadores do solo, análises de plantas, influência dos animais nos sistemas e características climáticas na sua estimação, especialmente comparando monocultivos e sistemas de integração Lavoura-Pecuária (ICLS). Nos sistemas de ICLS os animais são muito importantes, pois grande parte dos nutrientes ingeridos retorna ao solo via fezes e urina, os quais são liberados em um curto intervalo de tempo em formas prontamente disponíveis no solo (Vilela et al., 2011). Essa demanda continuada entre entradas e saídas de nutrientes no sistema pode aumentar a reserva de TOC e TON no solo quando comparados aos monocultivos (Bieluczyk et al., 2017; Liebig et al., 2017).

Diante deste contexto, objetivou-se (1) correlacionar o TOC e TON com variáveis de solos, plantas e animais considerando as condições climáticas e balanço hídrico de cada sistema avaliado, para avaliar e comparar o potencial de diferentes combinações dessas variáveis; visando (2) calibrar e testar seis algoritmos de machine learning (MLP, RF, MLR, SVR, KNN, AdaBoost) para estimar o teor de TOC e TON do solo em sistemas de monocultivos de milho, pecuária e sistema de integração Lavoura-Pecuária.

3.4 CONCLUSÕES

O carbono lábil, carbono da liteira, porosidade total do solo, capacidade de troca de cátions e estoque de nitrogênio no solo associado ao extreme gradient boosting (XGBoost) é eficiente na estimativa de PMN ($R^2=0.97$, RMSE de 7.22 mg kg^{-1} e MBE de 0.10 mg kg^{-1}), indicando que pode ser usado para estimar e monitorar indicadores importantes de qualidade do solo.

Ao interpretar os modelos mais acurados com o método SHAP, constatamos que 1) as variáveis mais importantes que impactam no PMN do solo são carbono lábil, carbono da liteira, porosidade total do solo, capacidade de troca de cátions e estoque de nitrogênio no solo, nessa ordem de importância; 2) a diminuição de carbono lábil no solo impacta positivamente na liberação de PMN no solo a curto prazo; 3) solos de sistemas integrados liberam mais carbono oriundo dos restos vegetais (carbono da liteira); 4) o balanceamento entre as estruturas físicas (porosidade total) e de fertilidade do solo (troca catiônica) influenciam na liberação de PMN do solo; 5) estoque de nitrogênio no solo resultam em menor probabilidade de estimar PMN do solo.

Esse estudo demonstra que modelos de machine learning podem ser usados para estimar o nitrogênio potencial mineralizável no solo, e através disso, monitorar a qualidade dos solos. No entanto, limitado pela disponibilidade de dados, nosso estudo não considerou o nitrogênio mineral ou o retorno desse nutriente para os sistemas de cultivo e colheitas. Assim, estudos futuros são necessários para melhorar ainda mais a acurácia dos modelos e auxiliar na interpretação das causas de explicabilidade de liberação de nitrogênio dos solos.

3.5 REFERÊNCIAS

- Abalos D, Jeffery S, Sanz-Cobena A, Guardia G, Vallejo A (2014) Meta-analysis of the effect of urease and nitrification inhibitors on crop productivity and nitrogen use efficiency. **Agriculture, Ecosystems & Environment** 189: 136-144. <https://doi.org/10.1016/j.agee.2014.03.036>
- Abbruzzini TF, Davies CA, Toledo FH, Cerri CEP (2019) Dynamic biochar effects on nitrogen use efficiency, crop yield and soil nitrous oxide emissions during a tropical wheat-growing season. **Journal of environmental management** 252: 109638. <https://doi.org/10.1016/j.jenvman.2019.109638>
- Bansal S, Chakraborty P, Kumar S (2022) Crop–livestock integration enhanced soil aggregate-associated carbon and nitrogen, and phospholipid fatty acid. **Scientific Reports** 12(1): 1-13. <https://doi.org/10.1038/s41598-022-06560-6>
- Burger M, Jackson LE (2003) Microbial immobilization of ammonium and nitrate in relation to ammonification and nitrification rates in organic and conventional cropping systems. **Soil Biology and Biochemistry** 35(1): 29-36. [https://doi.org/10.1016/S0038-0717\(02\)00233-X](https://doi.org/10.1016/S0038-0717(02)00233-X)
- Blair GJ, Lefroy RDB, Lisle L (1995) Soil carbon fractions based on their degree of oxidation, and the development of a carbon management index for agricultural systems. **Australian Journal of Agricultural Research** 46: 1459-1466. <https://doi.org/10.1071/AR9951459>
- Bockheim JG, Gennadiyev AN, Hartemink AE, Brevik EC (2014) Soil-forming factors and Soil Taxonomy. **Geoderma** 231-237. <https://doi.org/10.1016/j.geoderma.2014.02.016>
- Carvalho PCDF, Peterson CA, Nunes PADA, Martins AP, Souza Filho, W, Bertolazi VT, Kunrath TR, Moraes A, Anghinoni I (2018) Animal production and soil characteristics from integrated crop-livestock systems: toward sustainable intensification. **Journal of animal science** 96(8): 3513-3525. <https://doi.org/10.1093/jas/sky085>
- Chatzisyneon E, Foteinis S, Borthwick AGL (2017) Life cycle assessment of the environmental performance of conventional and organic methods of open field pepper cultivation system. **The International Journal of Life Cycle Assessment** 22: 896–908. <https://doi.org/10.1007/s11367-016-1204-8>
- Chirinda N, Olesen JE, Porter JR, Schjønning P (2010). Soil properties, crop production and greenhouse gas emissions from organic and inorganic fertilizer-based arable cropping systems. **Agriculture, Ecosystems & Environment** 139(4): 584-594. <https://doi.org/10.1016/j.agee.2010.10.001>
- Cookson WR, Abaye DA, Marschner P, Murphy DV, Stockdale EA, Goulding KW, (2005) The contribution of soil organic matter fractions to carbon and nitrogen mineralization and microbial community size and structure. **Soil Biology and Biochemistry** 37(9): 1726-1737. <https://doi.org/10.1016/j.soilbio.2005.02.007>

Córdova SC, Olk DC, Dietzel RN, Mueller KE, Archontoulis SV, Castellano MJ (2018) Plant litter quality affects the accumulation rate, composition, and stability of mineral-associated soil organic matter. **Soil Biology and Biochemistry** 125: 115-124. <https://doi.org/10.1016/j.soilbio.2018.07.010>

EMBRAPA (1997) Manual de métodos de análise de solo. Rio de Janeiro: Centro Nacional de Pesquisa de Solos.

Filippi P, Whelan BM, Vervoort RW, Bishop TF (2020). Mid-season empirical cotton yield forecasts at fine resolutions using large yield mapping datasets and diverse spatial covariates. **Agricultural Systems** 184: 102894. <https://doi.org/10.1016/j.agsy.2020.102894>

Franzluebbers AJ (2016) Should soil testing services measure soil biological activity? **Agricultural & Environmental Letters** 1(1). <https://doi.org/10.2134/ael2015.11.0009>

Friedman JH (2001) Greedy function approximation: a gradient boosting machine. **Annals of statistics** 1189-1232. <http://www.jstor.org/stable/2699986>.

Galindo FS, Delate K, Heins B, Phillips H, Smith A, Pagliari PH (2020) Cropping System and Rotational Grazing Effects on Soil Fertility and Enzymatic Activity in an Integrated Organic Crop-Livestock System. **Agronomy** 10: 803. <https://doi.org/10.3390/agronomy10060803>

Genuer R, Poggi JM, Tuleau-Malot C (2010) Variable selection using random forests. **Pattern recognition letters** 31(14): 2225-2236. <https://doi.org/10.1016/j.patrec.2010.03.014>

Guo L, Ma Y, Cukic B, Singh H (2004) Robust prediction of fault-proneness by random forests. In: 15th international symposium on software reliability engineering. **IEEE** 417-428. <https://doi.org/10.1109/ISSRE.2004.35>

Ghimire R, Thapa VR, Cano A, Acosta-Martinez V (2019) Soil organic matter and microbial community responses to semiarid croplands and grasslands management. **Applied Soil Ecology** 141: 30-37. <https://doi.org/10.1016/j.apsoil.2019.05.002>

Grell M, Barandun G, Asfour T, Kasimatis M, Collins ASP, Wang J, Güder F (2021) Point-of-use sensors and machine learning enable low-cost determination of soil nitrogen. **Nature Food** 2(12): 981-989. <https://doi.org/10.1038/s43016-021-00416-4>

Hassen TB, El Bilali H (2022) Impacts of the Russia-Ukraine War on Global Food Security: Towards More Sustainable and Resilient Food Systems?. **Foods** 11(15): 2301. <https://doi.org/10.3390/foods11152301>

Hebebrand C, Laborde D (2022) High Fertilizer Prices Contribute to Rising Global Food Security Concerns. Disponível em: <<https://www.ifpri.org/blog/high-fertilizer-prices-contribute-rising-global-food-security-concerns>>. Acesso em: 22 set. 2022.

Haynes RJ, Williams PH (1993) Nutrient cycling and soil fertility in the grazed pasture ecosystem. **Advances in agronomy** 49: 119-199. [https://doi.org/10.1016/S0065-2113\(08\)60794-4](https://doi.org/10.1016/S0065-2113(08)60794-4)

- Hikouei IS, Kim SS, Mishra DR (2021) Machine-learning classification of soil bulk density in salt marsh environments. **Sensors** 21(13): 4408. <https://doi.org/10.3390/s21134408>
- Ho LS, Tran VQ (2022) Machine learning approach for predicting and evaluating California bearing ratio of stabilized soil containing industrial waste. **Journal of Cleaner Production** 133587. <https://doi.org/10.1016/j.jclepro.2022.133587>
- Jarray N, Abbes AB, Farah IR (2021) An evaluation of soil moisture retrieval using machine learning methods: Application in arid regions of Tunisia. **IEEE** 6331-6334. <https://doi.org/10.1109/IGARSS47720.2021.9554585>
- Jones EJ, Bishop TF, Malone BP, Hulme PJ, Whelan BM, Filippi P (2022) Identifying causes of crop yield variability with interpretive machine learning. **Computers and Electronics in Agriculture** 192: 106632. <https://doi.org/10.1016/j.compag.2021.106632>
- Kastner T, Chaudhary A, Gingrich S, Marques A, Persson UM, Bidoglio G, Provost GL, Schwarzmüller F (2021) Global agricultural trade and land system sustainability: Implications for ecosystem carbon storage, biodiversity, and human nutrition. **One Earth** 4(10): 1425-1443. <https://doi.org/10.1016/j.oneear.2021.09.006>
- Lantinga EA, Boele E, Rabbinge R (2013) Maximizing the nitrogen efficiency of a prototype mixed crop-livestock farm in The Netherlands. **NJAS: Wageningen Journal of Life Sciences** 66(1): 15-22. <https://doi.org/10.1016/j.njas.2013.07.001>
- Liebig MA, Tanaka DL, Kronberg SL, Scholljegerdes EJ, Karn JF (2012) Integrated crops and livestock in central North Dakota, USA: Agroecosystem management to buffer soil change. **Renewable Agriculture and Food Systems** 27(2): 115-124. <https://doi.org/10.1017/S1742170511000172>
- Lin L, Gao Z, Liu X (2020) Estimation of soil total nitrogen using the synthetic color learning machine (SCLM) method and hyperspectral data. **Geoderma** 380: 114664. <https://doi.org/10.1016/j.geoderma.2020.114664>
- Linquist BA, Liu L, Van Kessel C, Van Groenigen KJ (2013) Enhanced efficiency nitrogen fertilizers for rice systems: Meta-analysis of yield and nitrogen uptake. **Field Crops Research** 154: 246-254. <https://doi.org/10.1016/j.fcr.2013.08.014>
- Liu L, King JS, Booker FL, Giardina CP, Lee Allen H, Hu S (2009) Enhanced litter input rather than changes in litter chemistry drive soil carbon and nitrogen cycles under elevated CO₂: a microcosm study. **Global Change Biology** 15(2): 441-453. <https://doi.org/10.1111/j.1365-2486.2008.01747.x>
- Lundberg SM, Lee SI (2017) A unified approach to interpreting model predictions. **Advances in neural information processing systems** 30.
- Lundberg SM, Erion G, Chen H, DeGrave A, Prutkin JM, Nair B, Katz R, Himmelfarb J, Bansal N, Lee SI (2020) From local explanations to global understanding with explainable AI for trees. **Nature machine intelligence** 2(1): 56-67. <https://doi.org/10.1038/s42256-019-0138-9>

Maia NJC, Cruz MCPD, Dubeux Jr JCB, Menegatto LS, Augusto JG, Mendonça GG, Terçariol MC, Oliveira JG, Simili FF (2021) Integrated crop-livestock versus conventional systems: use of soil indicators to detect short-term changes during seasonal variation. **Bragantia** 80. <https://doi.org/10.1590/1678-4499.20210127>

Mahmoud AA, Elkatatny S, Mahmoud M, Abouelresh M, Abdulraheem A, Ali A (2017) Determination of the total organic carbon (TOC) based on conventional well logs using artificial neural network. **International Journal of Coal Geology** 179: 72-80. <https://doi.org/10.1016/j.coal.2017.05.012>

Mendonça GG, Simili FF, Augusto JG, Bonacim PM, Menegatto LS, Gameiro AH, (2020) Economic gains from crop-livestock integration in relation to conventional systems. *Revista Brasileira de Zootecnia* 49: 1-11. <https://doi.org/10.37496/rbz4920190029>

Monz CA, Reuss DE, Elliott ET (1991) Soil microbial biomass carbon and nitrogen estimates using 2450 MHz microwave irradiation or chloroform fumigation followed by direct extraction. *Agriculture, Ecosystems & Environment* 34: 55-63. [https://doi.org/10.1016/0167-8809\(91\)90093-D](https://doi.org/10.1016/0167-8809(91)90093-D)

Moraes A, Carvalho PCF, Anghinoni I, Lustosa SBC, Andrade SEVG, Kunrath TR, (2014) Integrated crop–livestock systems in the Brazilian subtropics. **European Journal of Agronomy** 57: 4-9. <https://doi.org/10.1016/j.eja.2013.10.004>

Nguyen TT, Pham TD, Nguyen CT, Delfos J, Archibald R, Dang KB, Hoang NB, Guo W, Ngo HH (2022) A novel intelligence approach based active and ensemble learning for agricultural soil organic carbon prediction using multispectral and SAR data fusion. **Science of the Total Environment** 804: 150187. <https://doi.org/10.1016/j.scitotenv.2021.150187>

Oliveira JG, Santana Júnior ML, Maia NJC, Dubeux Jr JCB, Gameiro AH, Kunrath TR, Mendonça GG, Simili FF (2022) Nitrogen balance and efficiency as indicators for monitoring the proper use of fertilizers in agricultural and livestock systems. **Scientific Reports** 12(1): 1-10. <https://doi.org/10.1038/s41598-022-15615-7>

Osterholz WR, Rinot O, Shaviv A, Linker R, Liebman M, Sanford G, Strock J, Castellano MJ (2017) Predicting gross nitrogen mineralization and potentially mineralizable nitrogen using soil organic matter properties. **Soil Science Society of America Journal** 81(5): 1115-1126. <https://doi.org/10.2136/sssaj2017.02.0055>

Osterholz WR, Liebman M, Castellano MJ (2018) Can soil nitrogen dynamics explain the yield benefit of crop diversification?. **Field crops research** 219: 33-42. <https://doi.org/10.1016/j.fcr.2018.01.026>

Pathy A, Meher S, Balasubramanian P (2020) Predicting algal biochar yield using eXtreme Gradient Boosting (XGB) algorithm of machine learning methods. **Algal Research** 50: 102006. <https://doi.org/10.1016/j.algal.2020.102006>

Raj BV, Andrade JC, Cantarella H, Quaggio JA (Eds.) (2001) Análise química para avaliação da fertilidade de solos tropicais. Campinas: Instituto Agrônomo, 235p.

Rakkar MK, Blanco-Canqui H, Drijber RA, Drewnoski ME, MacDonald JC, Klopfenstein T (2017) Impacts of cattle grazing of corn residues on soil properties after 16 years. **Soil Science Society of America Journal** 81(2): 414-424. <https://doi.org/10.2136/sssaj2016.07.0227>

Reda R, Saffaj T, Ilham B, Saidi O, Issam K, Brahim L (2019) A comparative study between a new method and other machine learning algorithms for soil organic carbon and total nitrogen prediction using near infrared spectroscopy. **Chemometrics and Intelligent Laboratory Systems**. 195: 103873. <https://doi.org/10.1016/j.chemolab.2019.103873>

Rezende CP, Cantarutti RB, Braga JM, Gomide JA, Pereira JM, Ferreira E, Tarré RM, Macedo R, Alves BJR, Urquiaga S, Cadisch G, Giller KE, Boddey RM (1999) Litter deposition and disappearance in *Brachiaria* pastures in the Atlantic Forest region of the south of Bahia, Brazil. **Nutrient Cycling in Agroecosystems** 54: 99-112.

Rütting T, Aronsson H, Delin S (2018) Efficient use of nitrogen in agriculture. **Nutrient cycling in Agroecosystems** 110(1): 1-5. <https://doi.org/10.1007/s10705-017-9900-8>

Roberts TL, Norman RJ, Slaton NA, Wilson CE, Ross WJ, Bushong JT (2009) Direct steam distillation as an alternative to the Illinois soil nitrogen test. **Soil Science Society of America Journal** 73: 1268-1275. <https://doi.org/10.2136/sssaj2008.0165>

Romanyà J, Arco N, Solà-Morales I, Armengot L, Sans FX (2012) Carbon and nitrogen stocks and nitrogen mineralization in organically managed soils amended with composted manures. **Journal of Environmental Quality** 41(4): 1337-1347. <https://doi.org/10.2134/jeq2011.0456>

Ryschawy J, Liebigh MA, Kronberg SL, Archer DW, Hendrickson JR (2017) Integrated crop-livestock management effects on soil quality dynamics in a semiarid region: a typology of soil change over time. **Applied and Environmental Soil Science** 2017. <https://doi.org/10.1155/2017/3597416>

Scavuzzo CM, Scavuzzo JM, Campero MN, Anegagrie M, Aramendia AA, Benito A, Periago V (2022) Feature importance: Opening a soil-transmitted helminth machine learning model via SHAP. **Infectious Disease Modelling** 7(1): 262-276. <https://doi.org/10.1016/j.idm.2022.01.004>

Silva EE, Azevedo PHS, De-Polli H (2007) Determinação do Nitrogênio da Biomassa Microbiana do Solo (BMS-N). *Seropédica – RJ: EMBRAPA (EMBRAPA. Comunicado técnico, 96).*

Sinclair TR, Rufty TW (2012) Nitrogen and water resources commonly limit crop yield increases, not necessarily plant genetics. **Global Food Security** 1(2): 94-98. <https://doi.org/10.1016/j.gfs.2012.07.001>

Smith P, House JI, Bustamante M, Sobocká J, Harper R, Pan G, West PC, Clark JM, Adhya T, Rumpel C, Paustian K, Kuikman P, Cotrufo MF, Elliott JA, McDowell R, Griffiths RI, Asakawa S, Bondeau A, Jain AK, Meersmans J, Pugh TAM (2016) Global change pressures on soils from land use and management. **Global change biology** 22(3): 1008-1028. <https://doi.org/10.1111/gcb.13068>

Schjønning P, Thomsen IK, Moldrup P, Christensen BT (2003) Linking soil microbial activity to water-and air-phase contents and diffusivities. **Soil Science Society of America Journal** 67(1): 156-165. <https://doi.org/10.2136/sssaj2003.1560>

Souza EDD, Costa SEVGDA, Anghinoni I, Lima CVSD, Carvalho PCDF, Martins AP (2010) Biomassa microbiana do solo em sistema de integração lavoura-pecuária em plantio direto, submetido a intensidades de pastejo. **Revista Brasileira de Ciência do solo** 34: 79-88. <https://doi.org/10.1590/S0100-06832010000100008>

Stevenson FJ, Braids OC (1968) Variation in the relative distribution of amino sugar with depth in some soil profiles. **Soil Science Society of America Journal** 32: 590-598. <https://doi.org/10.2136/sssaj1968.03615995003200040049x>

Strickland MS, Thomason WE, Avera B, Franklin J, Minick K, Yamada S, Badgley BD, (2019) Short-Term effects of cover crops on soil microbial characteristics and biogeochemical processes across actively managed farms. **Agrosystems, Geosciences & Environment** 2, 1-9. <https://doi.org/10.2134/age2018.12.0064>

Strobl C, Boulesteix AL, Kneib T, Augustin T, Zeileis A (2008) Conditional variable importance for random forests. **BMC bioinformatics** 9(1): 1-11. <https://doi.org/10.1186/1471-2105-9-307>

Taylor P, Mansfield ER, Helms BP, Mansfield ER, Helms BP (1981) Detecting Multicollinearity. **The American Statistician** 36: 1-4. <https://doi.org/10.1080/00031305.1982.10482818>.

Thakur AK, Rath S, Mandal KG (2013) Differential responses of system of rice intensification (SRI) and conventional flooded-rice management methods to applications of nitrogen fertilizer. **Plant Soil** 370: 59-71. <https://doi.org/10.1007/s11104-013-1612-5>

Tran VQ (2022) Machine learning approach for investigating chloride diffusion coefficient of concrete containing supplementary cementitious materials. **Construction and Building Materials** 328: 127103. <https://doi.org/10.1016/j.conbuildmat.2022.127103>

Veldkamp E (1994) Organic Carbon Turnover in Three Tropical Soils under Pasture after Deforestation. *Soil Science Society of America Journal* 58: 175-180.

Xuejun L, Fusuo Z (2011) Nitrogen fertilizer induced greenhouse gas emissions in China. **Current Opinion in Environmental Sustainability** 3(5): 407-413. <https://doi.org/10.1016/j.cosust.2011.08.006>

Walkley A, Black IA (1934) An examination of the Degtjareff method for determining soil organic matter, and a proposed modification of the chromic soil titration method. *Soil Science* 37: -38.

Wei L, Yuan Z, Wang Z, Zhao L, Zhang Y, Lu X, Cao L (2020) Hyperspectral Inversion of Soil Organic Matter Content Based on a Combined Spectral Index Model. **Sensors** 20(10): 2777. <https://doi.org/10.3390/s20102777>

Weil RR, Islam KR, Stine MA, Gruver JB, Samson-Liebig SE (2003) Estimating active carbon for soil quality assessment: a simplified method for laboratory and field use. *American Journal of Alternative Agriculture* 18: 3-17. <https://doi.org/10.1079/AJAA200228>

Willmott CJ (1981) On the validation of models. *Physical geography* 2(2): 184-194.

Yan F, Song K, Liu Y, Chen S, Chen J (2020) Predictions and mechanism analyses of the fatigue strength of steel based on machine learning. **Journal of Materials Science** 55(31): 15334-15349. <https://doi.org/10.1007/s10853-020-05091-7>

Yang H, Huang K, Zhang K, Weng Q, Zhang H, Wang F (2021) Predicting heavy metal adsorption on soil with machine learning and mapping global distribution of soil adsorption capacities. **Environmental Science & Technology** 55(20): 14316-14328. <https://doi.org/10.1021/acs.est.1c02479>

Zhang T, Luo Y, Chen HY, Ruan H (2018) Responses of litter decomposition and nutrient release to N addition: A meta-analysis of terrestrial ecosystems. **Applied Soil Ecology** 128: 35-42. <https://doi.org/10.1016/j.apsoil.2018.04.004>

Zhang S, Zheng Q, Noll L, Hu Y, Wanek W (2019) Environmental effects on soil microbial nitrogen use efficiency are controlled by allocation of organic nitrogen to microbial growth and regulate gross N mineralization. **Soil Biology and Biochemistry** 135: 304-315. <https://doi.org/10.1016/j.soilbio.2019.05.019>

Zhao Y, Gao G, Ding G, Wang L, Chen Y, Zhao Y, Yu M, Zhang Y (2022) Assessing the influencing factors of soil susceptibility to wind erosion: A wind tunnel experiment with a machine learning and model-agnostic interpretation approach. **Catena** 215: 106324. <https://doi.org/10.1016/j.catena.2022.106324>

Zhou W, Li H, Wen S, Xie L, Wang T, Tian Y, Yu W (2022) Simulation of Soil Organic Carbon Content Based on Laboratory Spectrum in the Three-Rivers Source Region of China. **Remote Sensing** 14(6): 1521. <https://doi.org/10.3390/rs14061521>