

PREDIÇÃO DA QUALIDADE DO SOLO EM ÁREAS DE CANA-DE-AÇÚCAR SOB DIFERENTES SISTEMAS DE MANEJO

**CAMILA V. V. FARHATE¹, ZIGOMAR MENEZES DE SOUZA², MAURÍCIO R.
CHERUBIN³, NEWTON LA SCALA JUNIOR⁴, LENON H. LOVERA⁵, INGRID N. DE
OLIVEIRA⁶**

¹Eng^a Agrônoma, Pós-doutoranda, Universidade Estadual Paulista "Júlio de Mesquita Filho", Faculdade de Ciências Agrárias e Veterinárias, UNESP/FCAV, Jaboticabal – SP, Fone: (019) 3521-1069, camilavianav@hotmail.com.

²Eng^o Agrônomo, Prof. Associado, Faculdade de Engenharia Agrícola, FEAGRI/UNICAMP, Campinas – SP.

³Eng^o Agrônomo, Prof. Doutor, Departamento de Ciência do Solo, ESALQ-USP, Piracicaba – SP.

⁴Físico, Prof. Titular, Universidade Estadual Paulista "Júlio de Mesquita Filho", Faculdade de Ciências Agrárias e Veterinárias, UNESP/FCAV, Jaboticabal – SP.

⁵Eng^o Agrônomo, Doutorando, Faculdade de Engenharia Agrícola, FEAGRI/UNICAMP, Campinas – SP.

⁶Eng^a Agrícola, Doutoranda, Faculdade de Engenharia Agrícola, FEAGRI/UNICAMP, Campinas – SP.

Apresentado no
XLVIII Congresso Brasileiro de Engenharia Agrícola - CONBEA 2019
17 a 19 de setembro de 2019 - Campinas - SP, Brasil

RESUMO: Preservar a qualidade do solo sempre foi um desafio para os agricultores. Nesse contexto, o objetivo desse estudo foi analisar o desempenho do algoritmo *Random Forest* para prever a qualidade do solo a partir de um conjunto mínimo de variáveis. O estudo foi realizado na usina Santa Fé, em Ibitinga, São Paulo, Brasil. Foram analisados atributos físicos e químicos e posteriormente foi elaborado um único índice de qualidade do solo por meio de três etapas: i) seleção de atributos, ii) transformação de valores e, iii) integração. Em seguida foi realizada a modelagem preditiva da qualidade do solo utilizando o algoritmo *Random Forest*. O modelo preditivo gerado a partir de variáveis físicas e químicas do solo e, do algoritmo *Random Forest* apresentou elevado potencial para prever a qualidade do solo, alcançando uma correlação entre os dados preditos e observados de 94%.

PALAVRAS-CHAVE: Plantas de cobertura, sistema de preparo do solo, degradação do solo.

PREDICTION OF SOIL QUALITY IN SUGARCANE AREAS UNDER DIFFERENT MANAGEMENT SYSTEMS

ABSTRACT: Preserving soil quality has always been a challenge for farmers. In this context, the aim of this study was to analyze the performance of the Random Forest algorithm to predict soil quality from a minimum set of variables. The study was carried out at the Santa Fé Mill, in Ibitinga, São Paulo, Brazil. Physical and chemical attributes were analyzed and subsequently a single index of soil quality was elaborated through three stages: i) selection of attributes, ii) transformation of values and iii) integration. Then the predictive modeling of soil quality was performed using the Random Forest algorithm. The predictive model generated from soil physical and chemical variables, and from the Random Forest algorithm presented high potential to predict soil quality, reaching a correlation between predicted and observed data of 94%.

KEYWORDS: Cover crops, soil tillage system, soil degradation.

INTRODUÇÃO: O termo qualidade do solo pode ser definido como a capacidade de um solo funcionar, dentro dos limites do ecossistema e do uso da terra, para manter a produtividade, a qualidade ambiental e promover a saúde das plantas e dos animais (DORAN; PARKIN, 1994). Contudo, preservar a qualidade do solo atualmente é um desafio na agricultura canavieira, visto que o uso do preparo convencional e o alto nível de mecanização da cultura resultam em degradação do solo (BORDONAL et al., 2018).

Em contrapartida, sistemas de produção conservacionistas fundamentados na mínima ou ausência de mobilização do solo e no uso de plantas de cobertura, apresentam potencial para reverter o processo de degradação inerente a lavouras de cana-de-açúcar brasileiras.

Em adição, o aprendizado de máquina é um subcampo da ciência da computação, em que algoritmos inteligentes são capazes de aprender padrões a partir de um conjunto de dados treinamento e, por meio dos modelos preditivos gerados é possível classificar ou estimar novos casos (VOYANT et al., 2017). Nesse contexto, a elaboração de uma ferramenta de apoio que classifique a qualidade do solo e possibilite a tomada de decisão de forma rápida e precisa em relação às práticas de manejo que estão sendo adotadas torna-se essencial.

O algoritmo *Random Forest* é citado na literatura como um dos melhores métodos de aprendizagem na atualidade (WANG et al., 2016). Dessa forma, o objetivo desse estudo foi analisar o desempenho do algoritmo *Random Forest* para prever a qualidade do solo a partir de um conjunto mínimo de variáveis.

MATERIAL E MÉTODOS: O estudo foi realizado na usina Santa Fé, em Ibitinga, São Paulo, Brasil (21°83'43"S, 48°87'50 O a 455 metros acima do nível do mar). O solo foi classificado como Argissolo Vermelho distrófico típico de acordo com o Sistema Brasileiro de Classificação de Solos (SANTOS et al., 2018).

O delineamento experimental utilizado foi em faixas, onde as plantas de cobertura (crotalária, amendoim, milho e sorgo biomassa) foram implantadas nas faixas horizontais e, os sistemas de preparo do solo (sem preparo, subsolagem a 0,40 m e, subsolagem a 0,70 m de profundidade) nas faixas verticais. Para fins de comparação, foi instalado também um tratamento testemunha, sem plantas de cobertura e com preparo convencional do solo. Cada parcela foi repetida três vezes.

Foram realizadas amostragens de solo ao final dos ciclos de produção da cana planta e cana soca. A densidade do solo (D_s - kg dm^{-3}) e a macroporosidade (Macro - $\text{m}^3 \text{m}^{-3}$) foram determinadas conforme as metodologias estabelecidas por Teixeira et al. (2017). O diâmetro médio ponderado dos agregados (DMP; mm), por sua vez, foi determinado por agitação via úmida de acordo com o método proposto por Kemper e Chepil (1965). A resistência do solo à penetração (RP) foi obtida por meio de um Penetrômetro eletrônico de bancada, Modelo MA 933.

Para a análise dos atributos químicos do solo as amostras, foram secas ao ar e em seguida passadas em peneira de dois milímetros para análise dos macronutrientes contidos no solo, no qual foram determinados o teor de fósforo (P), cálcio (Ca), magnésio (Mg), potássio (K) e potencial hidrogeniônico (pH) de acordo com metodologia proposta por Raij et al. (2001). O teor de carbono total foi obtido por combustão via seca em um analisador elementar (modelo Truspec) (NELSON; SOMMERS, 1996).

A análise do índice de qualidade do solo envolveu três etapas: i) seleção de atributos, ii) transformação de valores e, iii) integração em um único índice de qualidade do solo conforme descrito por Andrews et al. (2002). Para seleção de indicadores foi utilizada uma análise de componentes principais. A transformação ocorreu considerando os valores obtidos pelo tratamento testemunha como referência. Por fim, as pontuações dos indicadores foram integradas em um único índice considerando uma adição ponderada.

Em seguida, a modelagem do índice de qualidade do solo foi realizada utilizando o algoritmo *Random Forest* implementado no programa R, sendo a validação do modelo realizada por meio do método *hold-out*, em que 70% dos dados foram utilizados para treinamento e 30% para teste. Por fim, os resultados foram expressos graficamente por meio de uma regressão, onde o resultado final consiste na média dos resultados de todas as árvores de regressão que compõe o algoritmo *Random Forest* (BREIMAN, 2001).

O índice de qualidade do solo foi submetido a análise de variância (ANOVA) e, ao apresentarem significância foram submetidos a um teste de Tukey ao nível de 5% de probabilidade. Além disso, foi realizada uma análise em faixas visando avaliar as interações obtidas entre as fontes de variação, ou seja, plantas de cobertura e sistema de preparo do solo.

RESULTADOS E DISCUSSÃO: Das 11 variáveis contidas no banco de dados original apenas três foram selecionadas por meio da análise de componentes principais para compor o índice de qualidade do solo, sendo duas variáveis químicas (pH e carbono) e uma física a (densidade do solo).

A variável pH foi a que apresentou o maior score no índice de qualidade. Verificou-se que a utilização de plantas de cobertura e sistemas de preparo de solo com menor grau de revolvimento proporcionaram maior qualidade do solo que o sistema convencional, com destaque para o uso de crotalária e milho e do sistema de preparo do solo com subsolagem a 0,70 m de profundidade.

O modelo preditivo da qualidade do solo gerado por meio do algoritmo *Random Forest* apresentou elevada taxa de acerto (Figura 1). Estudos utilizando técnicas de mineração de dados para prever especificamente a qualidade do solo em áreas de cana-de-açúcar são escassos. Entretanto, outros estudos presentes na literatura concordam com os resultados deste estudo em demonstrar elevada capacidade preditiva do modelo *Random forest*. Por exemplo, Tavares *et al.* (2018) em análise preditiva do fluxo de CO₂ em sistemas de manejo da cana-de-açúcar, observaram que uso do algoritmo *Random Forest* originou um modelo com ajuste de 80% entre os dados preditos e observados.

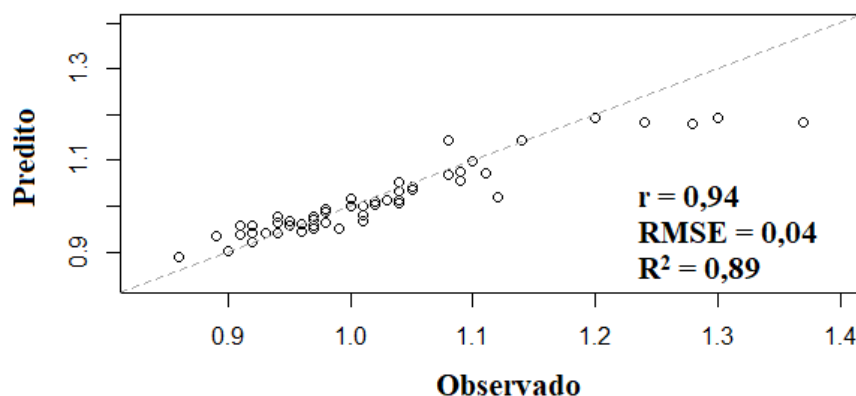


FIGURA 1. Validação do modelo gerado para predição da qualidade do solo em áreas de cana-de-açúcar.

Atualmente o setor agrícola tem passado por modificações e o termo Agricultura 4.0 tem obtido destaque, pois engloba a agricultura e pecuária de precisão, a automação e a robótica agrícola, além de técnicas de Big Data e a Internet das Coisas e, caracteriza-se por envolver alto nível tecnológico para o aumento dos índices de produtividade, eficiência do uso de insumos, da redução de custos com mão de obra, melhoria da qualidade do trabalho e da segurança dos trabalhadores, além de diminuir os impactos ao meio ambiente (MASSRUHÁ; LEITE, 2017). Logo, os resultados desse estudo poderão servir de base para avanços na

modelagem computacional na área agrícola, no qual a popularização de modelos de qualidade do solo em áreas de cultivo de cana-de-açúcar, revolucionariam o sistema atual de produção.

CONCLUSÕES: O modelo preditivo gerado a partir de variáveis físicas e químicas do solo e, do algoritmo *Random Forest* apresentou elevado potencial para prever a qualidade do solo, alcançando uma correlação entre os dados preditos e observados de 94%. Assim, em consequência das vantagens oferecidas pela predição da qualidade do solo em áreas de cana-de-açúcar, estudos mais detalhados e de longa duração devem ser encorajados, visando testar a reprodução desta técnica sob diferentes condições de solo e clima.

AGRADECIMENTOS: Os autores agradecem o apoio financeiro da Fundação Agrisus (Processo: 1439/15 e 2662/19) e da Fundação de Amparo à Pesquisa do Estado de São Paulo - FAPESP (Processo: 2018/14958-5) e, da Usina Santa Fé pelo fornecimento da área de estudos.

REFERÊNCIAS:

- ANDREWS, S.S.; KARLEN, D.L.; MITCHELL, J.P. A comparison of soil quality indexing methods for vegetable production systems in Northern California. **Agriculture, Ecosystems and Environment**, Beijing, v.90, n.1, p. 25-45, 2002.
- BORDONAL, R.O.; CARVALHO, J.L.N.; LAL, R.; FIGUEIREDO, E.B.; OLIVEIRA, B. G.; LA SCALA JÚNIOR, N. Sustainability of sugarcane production in Brazil. A review. **Agronomy for Sustainable Development**, Berlin, v.38, n.13, p.1-23, 2018.
- BREIMAN, L. Random forests. **Machine Learning**, Bristol, v.45, n.1, p.5-32, 2001.
- DORAN, J.W.; PARKIN, T.B. **Defining and assessing soil quality**. In: DORAN, J.W.; PARKIN, T.B. (Eds.), *Defining Soil Quality for Sustainable Environment*, SSSA and ASA, Madison, WI (1994), pp.3-21.
- KEMPER, W.D.; CHEPIL, W.S. **Size distribution of aggregates**. In: BLACK, C.A. (Ed.). *Methods of soil analysis*. Madison: American Society Agronomy, 1965. p.499-510.
- MASSRUHÁ, S, S. F. M.; LEITE, A. A. M. M. **Agro 4.0 – Rumo à agricultura digital**. JC na Escola Ciência, Tecnologia e Sociedade: Mobilizar o Conhecimento para Alimentar o Brasil. Campinas: Embrapa Informática Agropecuária, 2017. 35p.
- NELSON, D.W.; SOMMERS, L.E. **Total carbon, organic carbon and organic matter**. In: *Methods of Soil Analysis. Part 3. Chemical Methods-SSSA Book Series n° 5*. Madison: Soil Science Society of America and American Society of Agronomy, 1996. p.963-1010.
- RAIJ, B. V; ANDRADE, J.C.; CANTARELLA, H.; QUAGGIO, J.A. **Análise química para avaliação da fertilidade de solos tropicais**. Campinas: Instituto Agronômico, 2001, 285 p.
- SANTOS, H.G.; JACOMINE, P.K.T.; ANJOS, L.H.C.; OLIVEIRA, V.A.; LUMBRERAS, J. F.; COELHO, M.R.; ALMEIDA, J.A.; ARAUJO FILHO, J.C.; OLIVEIRA, J.B.; CUNHA, T.J.F. **Sistema Brasileiro de Classificação de Solos**. 5ª Ed. Revista e Ampliada, Brasília: Embrapa, 2018. 353 p.
- TEIXEIRA, P.C.; DONAGEMMA, G.K.; FONTANA, A.; TEIXEIRA, W.G. **Manual de Métodos de Análise de Solos**. 3ª Edição Revista e Ampliada, Brasília: Embrapa, 2017. 573p.
- VOYANT, C.; NOTTON, G.; KALOGIROU, S.; NIVET, M-L.; PAOLI, C.; MOTTE, F.; FOUILLOY, A. Machine learning methods for solar radiation forecasting: A review. **Renewable Energy**, Amsterdam, v.105, p.569-582, 2017.
- WANG, L.; ZHOU, X.; ZHU, X.; DONG, Z.; GUO, W. Estimation of biomass in wheat using Random Forest regression algorithm and remote sensing data. **The Crop Journal**, Beijing, v.4, n.1, p.212-219, 2016.