LUCAS DE CASTRO MOREIRA DA SILVA

MODELING AND ASSESSING HYDRAULIC PROPERTIES OF SELECTED BRAZILIAN AND AUSTRALIAN SOILS

Thesis submitted to the Agricultural Engineering Graduate Program of the Universidade Federal de Viçosa and to the School of Agricultural Sciences of Murdoch University in partial fulfillment of the requirements for the degree of *Doctor Scientiae – Joint Doctoral Degree*.

Advisers: Ricardo Santos Silva Amorim
Richard William Bell

Co-advisers: Elpídio Inácio Fernandes Filho
Luca De Prato
Wendy Vance

VIÇOSA - MINAS GERAIS, BRAZIL
2024

Ficha catalográfica elaborada pela Biblioteca Central da Universidade Federal de Viçosa - Campus Viçosa

T

S586m
2024

Silva, Lucas de Castro Moreira da, 1991-

Modeling and assessing hydraulic properties of selected brazilian and australian soils / Lucas de Castro Moreira da Silva. – Viçosa, MG, 2024.

1 tese eletrônica (164 f.): il. (algumas color.).

Texto em inglês.

Orientador: Ricardo Santos Silva Amorim.

Tese (doutorado) - Universidade Federal de Viçosa, Departamento de Engenharia Agrícola, 2024.

Inclui bibliografia.

DOI: https://doi.org/10.47328/ufvbbt.2024.098

Modo de acesso: World Wide Web.

1. Solos - Umidade - Medição. 2. Solos - Umidade - Modelos matemáticos. 3. Solos - Brasil. 4. Solos - Austrália. 5. Aprendizado do computador. I. Amorim, Ricardo Santos Silva, 1971-. II. Universidade Federal de Viçosa. Departamento de Engenharia Agrícola. Programa de Pós-Graduação em Engenharia Agrícola. III. Título.

CDD 22. ed. 631.432

Bibliotecário(a) responsável: Euzébio Luiz Pinto CRB-6/3317

LUCAS DE CASTRO MOREIRA DA SILVA

MODELING AND ASSESSING HYDRAULIC PROPERTIES OF SELECTED BRAZILIAN AND AUSTRALIAN SOILS

Thesis submitted to the Agricultural Engineering Graduate Program of the Universidade Federal de Viçosa and to the School of Agricultural Sciences of Murdoch University in partial fulfillment of the requirements for the degree of *Doctor Scientiae – Joint Doctoral Degree*.

Advisers: Ricardo Santos Silva Amorim
Richard William Bell
Co-advisers: Elpídio Inácio Fernandes Filho
Luca De Prato
Wendy Vance

APPROVED: March 12, 2024.

Assent:

Documento assinado digitalmente
LUCAS DE CASTRO MOREIRA DA SILVA
Data: 18/03/2024 09:17:07-0300
Verifique em https://validar.iti.gov.br

Lucas de Castro Moreira da Silva
Author

Documento assinado digitalmente
RICARDO SANTOS SILVA AMORIM
Data: 22/03/2024 07:29:46-0300
Verifique em https://validar.iti.gov.br

Ricardo Santos Silva Amorim
Adviser

ACKNOWLEDGEMENTS

To the Superior Force, to the Universal Power, to all divine beings.

To my parents, Cláudio and Adriana, for their unconditional support and love.

To my supervisors at Federal University of Viçosa (UFV), Prof. Ricardo Amorim and Prof. Elpídio Fernandes Filho, for their support, teachings, encouragement, and for always being willing to help me.

To my supervisors at Murdoch University (MU), Prof. Richard Bell, Dr. Luca De Prato, and Dr. Wendy Vance, for all the opportunities, for their support, teachings, help, and for the kind welcome and hospitality in Australia.

To my dear friends who supervised my master's degree at Federal University of Lavras (UFLA), Prof. Junior Cesar Avanzi, Prof. Bruno Montoani Silva, Dr. Devison Souza Peixoto, and Prof. Nilton Curi, for their invaluable support in preparing me to pursue a doctoral degree.

To Prof. Teogenes Oliveira and Dr. Marta Ottoni for their valuable contribution and thoughtful comments.

To all colleagues, professors, and employees from UFV, UFLA, and MU, with a special thanks to those from the Center of Reference in Water Resources at UFV, Soil Science Department at UFLA, and from the Land Management Group at MU.

To all relatives Castro and Moreira.

To all good friends from BH, Viçosa, Lavras, and Perth.

In loving memory of
HELENA VIEIRA DA COSTA E CASTRO
and
ANTÔNIO CARLOS PEREIRA DE CASTRO

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001, CAPES/PRINT Nº 88887.695902/2022-00.

“A good will isn't good because of what it effects or accomplishes, it's good in itself. Even if by utmost effort the good will accomplishes nothing, it would still shine like a jewel for its own sake as something which has its full value in itself. Usefulness or fruitlessness can neither augment nor diminish this value."

Immanuel Kant

Groundwork for the Metaphysics of Morals, 1785

ABSTRACT

SILVA, Lucas de Castro Moreira da, D.Sc., Universidade Federal de Viçosa, Murdoch University, March, 2024. **MODELING AND ASSESSING HYDRAULIC PROPERTIES OF SELECTED BRAZILIAN AND AUSTRALIAN SOILS**. Advisers: Ricardo Santos Silva Amorim and Richard William Bell. Co-advisers: Elpídio Inácio Fernandes Filho, Luca De Prato, and Wendy Vance.

Water availability is the main constraint for plants growth and current agricultural systems face the challenge of achieving high yields by optimizing the water use. The water dynamics in soils is governed by soil hydraulic properties and assessing these properties is essential for the efficient water use in agriculture. In this sense, several soil management practices have been employed to improve the soil water availability in agricultural fields, especially in water-limited regions. Thus, this study aimed to assess and predict soil hydraulic patterns in two different scenarios: Brazilian (First part) and Australian soils (Second part). In the first paper, a literature review was performed to bring to light what was done in ten years (2012-2021) regarding the prediction of soil hydraulic properties. In the second paper, machine learning models were developed to create regional pedotransfer functions for an important tropical agricultural center, the Mato Grosso state in Brazil. In the third paper, empirical models were tested for fitting water retention in Western Australian sandy soils modified by soil managements. In the fourth paper, popular practices employed to overcome sandy soils constraints were evaluated based on their effectiveness in enhance water availability. Results indicate that machine learning models are more accurate in predicting hydraulic properties compared to conventional methods. Regional-specific models were developed for soil hydraulic properties of Mato Grosso and are well calibrated for 91% of the state's territory using basic predictors. However, additional predictors reduce their applicability. Brooks and Corey model showed the best performance and a consistent negligible bias in estimating soil water retention of Western Australian sandy soils. Adding subsoil clay significantly increased total porosity and microporosity of sandy soils but did not improve water availability.

Keywords: Pedotransfer functions; Machine Learning; Soil hydraulic properties; Soil water retention; Available water

RESUMO

SILVA, Lucas de Castro Moreira da, D.Sc., Universidade Federal de Viçosa, Murdoch University, March, 2024. **MODELING AND ASSESSING HYDRAULIC PROPERTIES OF SELECTED BRAZILIAN AND AUSTRALIAN SOILS**. Orientadores: Ricardo Santos Silva Amorim and Richard William Bell. Coorientadores: Elpídio Inácio Fernandes Filho, Luca De Prato e Wendy Vance.

A disponibilidade de água limita o crescimento de plantas e sistemas agrícolas atuais têm o desafio de atingir altas produtividades enquanto otimizam o uso da água. A dinâmica da água no solo é governada por propriedades hidráulicas e a avaliação desses atributos é essencial para grarantir seu uso eficiente. Nesse sentido, diversas práticas de manejo de solo têm sido empregadas para melhorar a disponibilidade de água em campos agrícolas, especialmente em regiões onde há escassez hídrica. Nesse contexto, esse estudo objetivou predizer e avaliar propriedades hidráulicas do solo em dois cenários: solos do Brasil (Primeira parte) e Áustralia (Segunda parte). No primeiro artigo, foi realizada uma revisão de literatura para avaliar resultados de predição de propriedades hidráulicas de solos em dez anos (2012-2021). No segundo artigo, modelos de aprendizado de máquina foram usados para gerar funções de pedotransferência específicas para o estado de Mato Grosso. No terceiro artigo, modelos empíricos de ajuste da curva de retenção de água foram testados em solos arenosos da Austrália Ocidental modificados por sistemas de manejo. No quarto artigo sistemas de manejo foram avaliados de acordo com sua eficácia em aumentar a disponibilidade de água em solos arenosos da Austrália Ocidental. Os resultados indicam que modelos de aprendizado de máquina performaram melhor que técnicas estatísticas convencionais. Modelos regionais específicos foram desenvolvidos para propriedades hidráulicas de solos de Mato Grosso e têm calibração satisfatória em 91% da área do estado. No entanto, a adição de diferentes preditores reduz essa aplicabilidade. O modelo de Brooks e Corey exibiu a melhor performance e viés insignificante na estimativa da retenção de água no solo. A adição de argila de subsolo na matriz arenosa promoveu aumento significativo da porosidade total e da microporosidade, mas não demonstrou melhorias no que se refere à capacidade de água disponível.

Palavras-chave: Funções de pedotransferência; Aprendizagem de máquina; Propriedades hidráulicas do solo; Retenção de água no solo; Água disponível

SUMMARY

FIRST PART: Brazilian Soils

1. GENERAL INTRODUCTION

The world population will be approximately 8.4 billion in 2029 and Brazil is likely to continue to play a leading role in world food production, being one of the main producers and exporters of agricultural commodities (OECD-FAO, 2020). Mato Grosso State leads the Brazilian production of soybean, corn and cotton, accounting for 26.3% of the national grain harvest and has potential for further expanding its agricultural fields. However, the expansion of agricultural lands and increase in productivity has taken place through land use intensification and adoption of irrigation techniques, directly impacting water resources. In this sense, increasing food production can be threatened by water scarcity, which has been verified in different regions of the world and is projected to affect a large part of society in the upcoming decades (Schewe et al., 2014).

Water deficit is a major constraint for agricultural crop growth (Silva et al., 2019; Srayeddin and Doussan, 2009) thereby impacting humanity's food security. For that reason, Visser et al. (2019) state that the soil-water system is one of the bases to be studied in order to achieve the United Nations Sustainable Development Goals (SDGs) and Land Degradation Neutrality (LDN) by 2030 (Keesstra et al., 2018, 2016). Water scarcity imposes on production systems the challenge of achieving high yields by optimizing the use of water by crops (Whalley et al., 2017). Nevertheless, land use intensification can directly impact ecosystem functions, altering the water dynamics in soil, which is conditioned by the soil hydraulic properties (e.g., field capacity, permanent wilting point, plant-available water capacity and hydraulic conductivity).

Understanding soil hydraulic properties is essential for assessing available water for plants, consisting of important information for the efficient management of agricultural crops (Qiao et al., 2019; Santra et al., 2018; Silva et al., 2014). Moreover, these properties are required as input data in many study topics, including hydrology, agronomy, meteorology, ecology, and environment (Gupta et al., 2021a; Lehmann et al., 2020). Nonetheless, soil hydraulic properties are not easily found and made available in public databases, mainly because they are laborious and expensive to determine through conventional methods (Botula et al., 2014; Romano and Palladino, 2002).

Data scarcity can be minimized by alternative methods for indirectly estimating soil hydraulic properties (Botula et al., 2014), such as through pedotransfer functions (PTFs). PTFs were introduced by Bouma (1989) as an alternative method to predict complex soil attributes from more easy-to-measure attributes, e.g., particle size and organic carbon. The use of PTFs is an applicable and widely-used approach for estimating properties related to soil water retention processes (Botula et al., 2014, 2012; Cueff et al., 2021; Dobarco et al., 2019; Nguyen et al., 2015; Quentin and Philippe, 2021) and soil water movement (Araya and Ghezzehei, 2019; Gupta et al., 2021b; Ottoni et al., 2019; Zhang and Schaap, 2019) However, despite having several advantages, PTFs carry sources of error and uncertainty (Chirico et al., 2010; Liao et al., 2014; Mcbratney et al., 2002), and their general reliability is limited. As an example, the direct application of PTFs in different edaphic conditions from which they were developed can lead to accuracy loss due to regional soil particularities (Minasny et al., 1999).

Considering the specificities of tropical soils such as structure and mineralogy (Ferreira et al., 1999; Ker, 1997; Martinez and Souza, 2020; Schaefer et al., 2008; Schaefer, 2001; Silva et al., 2022), several studies have been carried out in order to apply and validate PTFs for hydraulic properties of Brazilian soils (Botula et al., 2014; Silva et al., 2017; Hodnett and Tomasella, 2002; Kotlar et al., 2020, 2019; Medrado and Lima, 2014; Ottoni et al., 2019, 2018; Tomasella et al., 2000; Tomasella and Hodnett, 2004, 1998). Ottoni et al. (2018) propose a database of hydrophysical properties of Brazilian soils and summarizes important information regarding PTFs application in highly weathered tropical soils. The authors emphasize the ineffectiveness of PTFs developed under temperate soils in predicting Brazilian soil features. This generates a demand for developing region-specific PTFs, which is likely to continue until generic PTFs are generated and validated in large representative databases (Patil and Singh, 2016).

Despite efforts devoted to establishing a hydrophysical database of Brazilian soils (HYBRAS; Ottoni et al., 2018), this library does not contain data of soils from Mato Grosso State. The work by Medrado and Lima (2014) includes 25 samples from Mato Grosso, with limited coverage to the central region of the state and considering soils from one biome (Cerrado) of the three biomes found in the state (i.e., Amazon, Cerrado and Pantanal). The absence of available data and specific models for the Mato Grosso state limits the implementation of agricultural and environmental studies in the region. The Mato Grosso State has a large variation of soil hydraulic properties resulting from the great pedological and land use diversity in its vast extension (~90 million hectares). Hence, it is crucial to develop and validate specific PTFs for this important tropical agricultural center.

Robust prediction techniques are required to ensure models reliability (Patil and Singh, 2016), and improving the performance of PTF predictions is still a challenge (Benke et al., 2020; Kotlar et al., 2019; Liao et al., 2014; Mcbratney et al., 2002; Pham et al., 2019). In this sense, machine learning techniques have been widely used in recent studies for generating PTFs (Araya and Ghezzehei, 2019; Guio Blanco et al., 2018; Gunarathna et al., 2019; Ramcharan et al., 2017; Van Looy et al., 2017; Yamaç et al., 2020). Machine learning algorithms have been employed to improve the calibration performance of PTFs, historically done by conventional statistical and regression methods (Patil and Singh, 2016; Van Looy et al., 2017). Machine learning-based PTFs may have significant advantages compared to conventional ones, since machine learning algorithms in general do not require strict assumptions about data distribution and are able to promote good fits and generalizations in large databases, which is a drawback of traditional regression methods (Pham et al., 2019). Furthermore, machine learning techniques are convenient for fitting complex and non-linear functions (Witten et al., 2011), commonly found in the relationship among soil properties (Gebauer et al., 2020).

In this context, this first part aimed to: (i) perform a literature review to bring to light what was done in the period 2012-2021 regarding the modeling of soil hydraulic properties by PTFs, raising many aspects of PTF development, such as coverage/domain, predominant climate, countries, target variable, methods to derive PTF, algorithms (in case of machine learning), predictors, model sensitivity and uncertainty assessment, number of samples for models training and testing, error metrics and goodness-of-fit. (ii) Develop PTFs for predicting water retention at matric potentials of 0, -4, -6, -10, -33, -100, -500 and -1500 kPa and hydraulic conductivity of soils from Mato Grosso state using machine learning techniques, testing different predictor sets, and establishing the PTFs domain of application.

References

Araya, S.N., Ghezzehei, T.A., 2019. Using Machine Learning for Prediction of Saturated Hydraulic Conductivity and Its Sensitivity to Soil Structural Perturbations. Water Resour Res 55, 5715–5737. https://doi.org/10.1029/2018WR024357

Benke, K.K., Norng, S., Robinson, N.J., Chia, K., Rees, D.B., Hopley, J., 2020. Development of pedotransfer functions by machine learning for prediction of soil electrical conductivity and organic carbon content. Geoderma 366. https://doi.org/10.1016/j.geoderma.2020.114210

Botula, Y.D., Cornelis, W.M., Baert, G., Van Ranst, E., 2012. Evaluation of pedotransfer functions for predicting water retention of soils in Lower Congo (D.R. Congo). Agric Water Manag 111, 1–10. https://doi.org/10.1016/j.agwat.2012.04.006

Botula, Y.-D., Ranst, E. Van, Cornelis, W.M., 2014. Pedotransfer functions to predict water retention for soils of the humid tropics: a review. Rev Bras Cienc Solo 38, 679–698.

Bouma, J., 1989. Using Soil Survey Data for Quantitative Land Evaluation, in: Advances in Soil Science. pp. 177–211.

Chirico, G.B., Medina, H., Romano, N., 2010. Functional evaluation of PTF prediction uncertainty: An application at hillslope scale. Geoderma 155, 193–202. https://doi.org/10.1016/j.geoderma.2009.06.008

Cueff, S., Coquet, Y., Aubertot, J.N., Bel, L., Pot, V., Alletto, L., 2021. Estimation of soil water retention in conservation agriculture using published and new pedotransfer functions. Soil Tillage Res 209. https://doi.org/10.1016/j.still.2021.104967

Dobarco, M.R., Cousin, I., Le Bas, C., Martin, M.P., 2019. Pedotransfer functions for predicting available water capacity in French soils, their applicability domain and associated uncertainty. Geoderma 336, 81–95. https://doi.org/10.1016/j.geoderma.2018.08.022

Ferreira, M.M., Fernandes, B., Curi, N., 1999. Mineralogia da fração argila e estrutura de Latossolos da região sudeste do Brasil. Rev Bras Cienc Solo 23, 507–514.

Gebauer, A., Ellinger, M., M. Brito Gomez, V., Ließ, M., 2020. Development of pedotransfer functions for water retention in tropical mountain soil landscapes: Spotlight on parameter tuning in machine learning. SOIL 6, 215–229. https://doi.org/10.5194/soil-6-215-2020

Guio Blanco, C.M., Brito Gomez, V.M., Crespo, P., Ließ, M., 2018. Spatial prediction of soil water retention in a Páramo landscape: Methodological insight into machine learning using random forest. Geoderma 316, 100–114. https://doi.org/10.1016/j.geoderma.2017.12.002

Gunarathna, M.H.J.P., Sakai, K., Nakandakari, T., Momii, K., Kumari, M.K.N., 2019. Machine learning approaches to develop pedotransfer functions for tropical Sri Lankan soils. Water (Switzerland) 11. https://doi.org/10.3390/w11091940

Gupta, S., Hengl, T., Lehmann, P., Bonetti, S., Or, D., 2021a. SoilKsatDB: global database of soil saturated hydraulic conductivity measurements for geoscience applications. Earth Syst Sci Data 13, 1593–1612. https://doi.org/10.5281/zenodo.3752721

Gupta, S., Hengl, T., Lehmann, P., Bonetti, S., Papritz, A., Or, D., 2021b. Global Prediction of Soil Saturated Hydraulic Conductivity Using Random Forest in a Covariate-Based GeoTransfer Function (CoGTF) Framework. Journal of Advances in Modeling Earth Systems (JAMES) 13, e2020MS002242.

Hodnett, M.G., Tomasella, J., 2002. Marked differences between van Genuchten soil water-retention parameters for temperate and tropical soils: a new water-retention pedo-transfer functions developed for tropical soils. Geoderma 108, 155–180.

Keesstra, S., Mol, G., de Leeuw, J., Okx, J., Molenaar, C., de Cleen, M., Visser, S., 2018. Soil-related sustainable development goals: Four concepts to make land degradation neutrality and restoration work. Land (Basel) 7. https://doi.org/10.3390/land7040133

Keesstra, S.D., Bouma, J., Wallinga, J., Tittonell, P., Smith, P., Cerdà, A., Montanarella, L., Quinton, J.N., Pachepsky, Y., Van Der Putten, W.H., Bardgett, R.D., Moolenaar, S., Mol, G., Jansen, B., Fresco, L.O., 2016. The significance of soils and soil science towards realization of the United Nations sustainable development goals. SOIL 2, 111–128. https://doi.org/10.5194/soil-2-111-2016

Ker, J.C., 1997. Latossolos do Brasil: uma revisão. GEONOMOS 5, 17–40.

Kotlar, A.M., de Jong van Lier, Q., Barros, A.H.C., Iversen, B. V., Vereecken, H., 2019. Development and Uncertainty Assessment of Pedotransfer Functions for Predicting Water Contents at Specific Pressure Heads. Vadose Zone Journal 18. https://doi.org/10.2136/vzj2019.06.0063

Kotlar, A.M., de Jong van Lier, Q., de Souza Brito, E., 2020. Pedotransfer functions for water contents at specific pressure heads of silty soils from Amazon rainforest. Geoderma 361. https://doi.org/10.1016/j.geoderma.2019.114098

Lehmann, P., Bickel, S., Wei, Z., Or, D., 2020. Physical Constraints for Improved Soil Hydraulic Parameter Estimation by Pedotransfer Functions. Water Resour Res 56. https://doi.org/10.1029/2019WR025963

Liao, K., Xu, S., Wu, J., Zhu, Q., 2014. Uncertainty analysis for large-scale prediction of the van Genuchten soil-water retention parameters with pedotransfer functions. Soil Research 52, 431–442. https://doi.org/10.1071/SR13230

Martinez, P., Souza, I.F., 2020. Genesis of pseudo-sand structure in Oxisols from Brazil – A review. Geoderma Regional. https://doi.org/10.1016/j.geodrs.2020.e00292

Mcbratney, A.B., Minasny, B., Cattle, S.R., Vervoort, R.W., 2002. From pedotransfer functions to soil inference systems, geoderma.

Medrado, E., Lima, J.E.F.W., 2014. Development of pedotransfer functions for estimating water retention curve for tropical soils of the Brazilian savanna. Geoderma Regional 1, 59–66. https://doi.org/10.1016/j.geodrs.2014.08.003

Minasny, B., Mcbratney, A.B., Bristow, K.L., 1999. Comparison of different approaches to the development of pedotransfer functions for water-retention curves, Geoderma.

Nguyen, P.M., Van Le, K., Botula, Y.D., Cornelis, W.M., 2015. Evaluation of soil water retention pedotransfer functions for Vietnamese Mekong Delta soils. Agric Water Manag 158, 126–138. https://doi.org/10.1016/j.agwat.2015.04.011

OECD-FAO Agricultural Outlook 2020-2029, 2020. , OECD-FAO Agricultural Outlook. OECD. https://doi.org/10.1787/1112c23b-en

Ottoni, M.V., Ottoni Filho, T.B., Lopes-Assad, M.L.R.C., Rotunno Filho, O.C., 2019. Pedotransfer functions for saturated hydraulic conductivity using a database with temperate and tropical climate soils. J Hydrol (Amst) 575, 1345–1358. https://doi.org/10.1016/j.jhydrol.2019.05.050

Ottoni, M.V., Ottoni Filho, T.B., Schaap, M.G., Lopes-Assad, M.L.R.C., Rotunno Filho, O.C., 2018. Hydrophysical Database for Brazilian Soils (HYBRAS) and Pedotransfer Functions for Water Retention. Vadose Zone Journal 17, 1–17. https://doi.org/10.2136/vzj2017.05.0095

Patil, N.G., Singh, S.K., 2016. Pedotransfer Functions for Estimating Soil Hydraulic Properties: A Review. Pedosphere 26, 417–430. https://doi.org/10.1016/S1002-0160(15)60054-6

Pham, K., Kim, D., Yoon, Y., Choi, H., 2019. Analysis of neural network based pedotransfer function for predicting soil water characteristic curve. Geoderma 351, 92–102. https://doi.org/10.1016/j.geoderma.2019.05.013

Qiao, J., Zhu, Y., Jia, X., Huang, L., Shao, M., 2019. Pedotransfer functions for estimating the field capacity and permanent wilting point in the critical zone of the Loess Plateau, China. J Soils Sediments 19, 140–147. https://doi.org/10.1007/s11368-018-2036-x

Quentin, S., Philippe, L., 2021. Uncertainty assessment of soil available water capacity using error propagation: A test in Languedoc-Roussillon. Geoderma 391. https://doi.org/10.1016/j.geoderma.2021.114968

Ramcharan, A., Hengl, T., Beaudette, D., Wills, S., 2017. A Soil Bulk Density Pedotransfer Function Based on Machine Learning: A Case Study with the NCSS Soil Characterization Database. Soil Science Society of America Journal 81, 1279–1287. https://doi.org/10.2136/sssaj2016.12.0421

Romano, N., Palladino, M., 2002. Prediction of soil water retention using soil physical data and terrain attributes. J Hydrol (Amst) 265, 56–75.

Santra, P., Kumar, M., Kumawat, R.N., Painuli, D.K., Hati, K.M., Heuvelink, G.B.M., Batjes, N.H., 2018. Pedotransfer functions to estimate soil water content at field capacity and permanent wilting point in hot Arid Western India. Journal of Earth System Science 127. https://doi.org/10.1007/s12040-018-0937-0

Schaefer, C.E.G.R., Fabris, J.D., Ker, J.C., 2008. Minerals in the clay fraction of Brazilian Latosols (Oxisols): a review. Clay Miner 43, 137–154. https://doi.org/10.1180/claymin.2008.043.1.11

Schaefer, C.E.R., 2001. Brazilian latosols and their B horizon microstructure as long-term biotic constructs. Australian Journal of Soil Research 39, 909–926. https://doi.org/10.1071/SR00093

Schewe, J., Heinke, J., Gerten, D., Haddeland, I., Arnell, N.W., Clark, D.B., Dankers, R., Eisner, S., Fekete, B.M., Colón-González, F.J., Gosling, S.N., Kim, H., Liu, X., Masaki, Y., Portmann, F.T., Satoh, Y., Stacke, T., Tang, Q., Wada, Y., Wisser, D., Albrecht, T., Frieler, K., Piontek, F., Warszawski, L., Kabat, P., 2014. Multimodel assessment of water scarcity under climate change. Proc Natl Acad Sci U S A 111, 3245–3250. https://doi.org/10.1073/pnas.1222460110

Silva, A.C., Armindo, R.A., Brito, A. dos S., Schaap, M.G., 2017. An Assessment of Pedotransfer Function Performance for the Estimation of Spatial Variability of Key Soil Hydraulic Properties. Vadose Zone Journal 16, 1–10. https://doi.org/10.2136/vzj2016.12.0139

Silva, B.M., Oliveira, G.C., Serafim, M.E., Silva, É.A., Guimarães, P.T.G., Melo, L.B.B., Norton, L.D., Curi, N., 2019. Soil moisture associated with least limiting water range, leaf water potential, initial growth and yield of coffee as affected by soil management system. Soil Tillage Res 189, 36–43. https://doi.org/10.1016/j.still.2018.12.016

Silva, B.M., Silva, É.A. da, Oliveira, G.C. de, Ferreira, M.M., Serafim, M.E., 2014. Plant-available soil water capacity: estimation methods and implications. Rev. Bras. Ci. Solo 38, 464–475.

Silva, L. de C.M. da, Peixoto, D.S., Gomes, J.B.V., Avanzi, J.C., Amorim, R.S.S., Borghi, E., Resende, Á.V. de, Silva, B.M., Mancini, M., Curi, N., 2022. Mineralogy and pore size distribution of clayey Oxisols with granular structure and the effect of management systems. Soil Tillage Res 223. https://doi.org/10.1016/j.still.2022.105479

Srayeddin, I., Doussan, C., 2009. Estimation of the spatial variability of root water uptake of maize and sorghum at the field scale by electrical resistivity tomography. Plant Soil 319, 185–207. https://doi.org/10.1007/s11104-008-9860-5

Tomasella, J., Hodnett, M., 2004. Pedotransfer functions for tropical soils. Developments in Soil Science. https://doi.org/10.1016/S0166-2481(04)30021-8

Tomasella, J., Hodnett, M.G., 1998. Estimating soil water retention characteristics from limited data in Brazilian Amazonia. Soil Sci 163, 190–202. https://doi.org/10.1097/00010694-199803000-00003

Tomasella, J., Hodnett, M.G., Rossato, L., 2000. Pedotransfer Functions for the Estimation of Soil Water Retention in Brazilian Soils. Soil Science Society of America Journal 64, 327–338. https://doi.org/10.2136/sssaj2000.641327x

Van Looy, K., Bouma, J., Herbst, M., Koestel, J., Minasny, B., Mishra, U., Montzka, C., Nemes, A., Pachepsky, Y.A., Padarian, J., Schaap, M.G., Tóth, B., Verhoef, A., Vanderborght, J., van der Ploeg, M.J., Weihermüller, L., Zacharias, S., Zhang, Y., Vereecken, H., 2017. Pedotransfer Functions in Earth System Science: Challenges and Perspectives. Reviews of Geophysics. https://doi.org/10.1002/2017RG000581

Visser, S., Keesstra, S., Maas, G., de Cleen, M., Molenaar, C., 2019. Soil as a basis to create enabling conditions for transitions towards sustainable land management as a key to achieve the SDGs by 2030. Sustainability (Switzerland) 11. https://doi.org/10.3390/su11236792

Whalley, W.R., Binley, A., Watts, C.W., Shanahan, P., Dodd, I.C., Ober, E.S., Ashton, R.W., Webster, C.P., White, R.P., Hawkesford, M.J., 2017. Methods to estimate changes in soil water for phenotyping root activity in the field. Plant Soil 415, 407–422. https://doi.org/10.1007/s11104-016-3161-1

Witten, I.H., Frank, E., Hall, M.A., 2011. Data Mining: Practical Machine Learning Tools and Techniques.

Yamaç, S.S., Şeker, C., Negiş, H., 2020. Evaluation of machine learning methods to predict soil moisture constants with different combinations of soil input data for calcareous soils in a semi arid area. Agric Water Manag 234. https://doi.org/10.1016/j.agwat.2020.106121

Zhang, Y., Schaap, M.G., 2019. Estimation of saturated hydraulic conductivity with pedotransfer functions: A review. J Hydrol (Amst) 575, 1011–1030. https://doi.org/10.1016/j.jhydrol.2019.05.058

2. Paper I: Pedotransfer functions and machine learning: Advancements and challenges in tropical soils

This paper is published at Geoderma Regional and available online since October 11th, 2023: https://doi.org/10.1016/j.geodrs.2023.e00720

Silva, L.C.M., Amorim, R.S.S., Fernandes Filho, E.I., Bocuti, E.D., Silva, D.D., 2023. Pedotransfer functions and machine learning: Advancements and challenges in tropical soils. Geoderma Regional, p.e00720.

Abstract

Soil hydraulic properties (SHP) are crucial information for several Earth system applications and their availability strongly relies on pedotransfer functions (PTFs). In this review, we aimed to bring to light what was done in the period 2012-2021 regarding the modeling of SHP by PTFs. Data were collected from 101 publications, yielding information from 1844 different PTFs across the world. PTF development was analyzed regarding coverage/domain, predominant climate, countries, target variable, methods to derive PTF, algorithms used (in case of machine learning), predictors, model sensitivity and uncertainty assessment, number of samples for models training and testing, error metrics and goodness-of-fit criteria, among more. A significant contribution from tropical regions was observed, which was seldom in past studies. The increased use of machine learning techniques to build PTFs was also a remarkable result, although conventional regression still prevailed. Given the growth of machine learning techniques in recent years, we delve deeper into this discussion. Generally, machine learning-based PTFs provided lower errors than conventional PTFs and should be considered for future studies. As a case study, we evaluated existing PTFs to predict water retention and hydraulic conductivity of soils from an important tropical agricultural center, the Mato Grosso state in Brazil. Results reinforce the low ability of PTFs based on temperate soils data to predict SHP of tropical soils, especially concerning fine-textured soils. Nonetheless, PTFs based on tropical soils also showed modest performances, indicating that the seek for greater reliability of PTFs is still pursued. The most outstanding results were observed for fine-textured soils, for which all PTFs had poor performance. Overall, results corroborate that there is a special demand for developing PTFs capable of well recognizing the hydraulic patterns of fine-textured tropical soils.

Keywords: hydraulic conductivity; soil water retention; available water capacity; Oxisols; Ferralsols

2.1 Introduction

The soil-water system is a key study topic aiming at achieving the United Nations Sustainable Development Goals and Land Degradation Neutrality by 2030 (Keesstra et al., 2018, 2016; Visser et al., 2019). Soil hydraulic properties (SHP) are required as input data in agronomic, climatic and hydrological models (Dobarco et al., 2019; Gupta et al., 2020; Lehmann et al., 2020), being an essential information for the efficient use of water in agriculture, soil and water conservation, and water resources management. Knowledge of soil water retention and soil hydraulic conductivity is crucial to support actions regarding agricultural drought susceptibility, agricultural zoning, definition of sowing dates, land use planning (Quentin and Philippe, 2021; Silva et al., 2014), runoff generation, recharge potential, flood risk, among others (Araya and Ghezzehei, 2019; Lim et al., 2020; Ottoni et al., 2019; Zhang et al., 2020). SHP are also sensitive indicators of environmental quality and are currently used to assess climate and land use changes (Méndez-Toribio et al., 2020; Poggio et al., 2010; Scheffler et al., 2011). Nevertheless, obtaining these data is far from trivial.

Collecting SHP information by conventional methods such as field and laboratory works is laborious and expensive (Romano and Palladino, 2002; Santra et al., 2018). In addition, SHP commonly have a high spatial variation (Minasny and McBratney, 2002a; Santos et al., 2021), requiring a large number of observations to represent this variability in simulations of field conditions. This difficulty in collecting and determining SHP leads to an absence of representative SHP information in soil surveys and the consequent lack of SHP data in large databases. In this regard, an alternative to obtain SHP information is by indirectly estimating these soil properties (Botula et al., 2014), such as via pedotransfer functions (PTFs). Although real field and laboratory observations are still essential for the development of PTFs, once a reliable model is developed, SHP can be predicted, overcoming SHP information scarcity. PTFs were introduced by Bouma (1989) as an alternative method for predicting complex soil attributes (e.g., SHP) from easy-to-measure attributes (e.g., particle size distribution and organic carbon). Overall, PTFs can be subdivided in two main groups considering the input variables, class and continuous type. Class-PTFs predict the target based on soil textural classes or intervals of predictors, whereas continuous-PTFs use the predictor actual values to derive equations (Dobarco et al., 2019; Wösten et al., 1999). PTFs are widely used for estimating soil water retention (Botula et al., 2014; Cueff et al., 2021; Dobarco et al., 2019; Quentin and Philippe, 2021; Rab et al., 2011) and soil hydraulic conductivity (Araya and Ghezzehei, 2019; Gupta et al., 2021; Ottoni et al., 2019; Zhang and Schaap, 2019). However, despite having several advantages, PTFs usually carry errors and sources of uncertainty (Chirico et al., 2010; Liao et al., 2014), which can be accumulated during the modeling processes (Minasny et al., 1999; Minasny and McBratney, 2002b) and propagated in the final prediction, compromising their reliability.

The greater reliability of PTFs for SHP estimation is still pursued (Van Looy et al., 2017) and improving prediction performance is a current challenge (Benke et al., 2020; Kotlar et al., 2019; Minasny and McBratney, 2002; Pham et al., 2019). According to a review by Patil and Singh (2016), robust prediction techniques should be used to improve the reliability of information generated by PTFs. In this sense, machine learning techniques, which have been widely used in the recent past years for pattern recognition in the agricultural and environmental fields, have also been used to generate recent PTFs (Araya and Ghezzehei, 2019; Benke et al., 2020; Guio Blanco et al., 2018; Gunarathna et al., 2019a; Pham et al., 2019; Ramcharan et al., 2017; Yamaç et al., 2020). As an example of the growth and importance of machine learning techniques, Nearing et al. (2020) called the present time as the machine learning age.

Machine learning algorithms have been adopted as a tool to improve PTF calibration, historically performed by conventional statistical and regression techniques (Patil and Singh, 2016; Van Looy et al., 2017). PTFs based on machine learning may show advantages when compared to conventional statistical methods. In general, machine learning algorithms do not require assumptions about data distribution and are able to promote good fits and generalizations with large databases, which is a limiting factor in traditional regression methods (Pham et al., 2019). Furthermore, machine learning techniques are convenient for fitting complex and non-linear functions (Witten et al., 2011), as can be found in soil properties interplay (Gebauer et al., 2020). Nevertheless, even with the recent advances provided by machine learning, there are gaps in PTFs development and reliability. As a classical but current example, applying PTFs in different edaphic conditions than from which they were developed may have limited reliability due to regional soil singularities (Minasny et al., 1999; Tomasella and Hodnett, 2004). This situation is especially relevant to tropical regions, since most of the available PTFs were developed in temperate climates (Botula et al., 2013; Gunarathna et al., 2019b; Ottoni et al., 2019, 2018)

Tropical regions are commonly covered by highly-weathered soils with structural and mineralogical singularities (Goedert, 1983; Lal, 1979; Schaefer et al., 2008) as a consequence of high weathering and leaching processes (Dick et al., 1986; Silva et al., 2022). Tropical soils may have a very simple clay mineralogy which may form highly-stable microaggregates made of flocculated clay (Galvão and Schulze, 1996; Donagemma et al., 2003). The arrangement of these microaggregates - which are commonly referred as pseudo-sand (Martinez and Souza, 2020) or pseudo-silt (Vitorino et al., 2003) - generates a structure with distinct hydraulic patterns (Ferreira et al., 1999; Silva et al., 2022), partially similar to sandy soils due to the high macroporosity (Pessoa and Libardi 2022; Tomassella et al., 2000), although they typically have high total porosities, in contrast with sands. For that reason, several studies have been performed to fill this gap developing and validating PTFs for hydraulic properties of tropical soils (Botula et al., 2013; Kotlar et al., 2020, 2019a; Medrado and Lima, 2014; Nguyen et al., 2015b, 2014; Ottoni et al., 2018; Tomasella et al., 2000; Tomasella and Hodnett, 2004, 1998). For instance, the work by Ottoni et al. (2018) gathered hydrophysical data of Brazilian tropical soils aiming at developing pedotransfer functions at a national scale. The authors synthesize information about the use of PTFs in tropical soils and emphasize the ineffectiveness of temperate-climate PTFs in predicting Brazilian SHP, proving that efforts still need to be made towards developing reliable PTFs for tropical soils.

In this context, our review aimed to bring to light what was done in ten years (2012-2021) regarding the prediction of SHP by PTF. We systematically reviewed several details of PTF development and validation, providing quantitative results of (i) PTFs areal coverage proposal; (ii) methods to derive PTFs; (iii) number of samples used for model training and testing; (iv) variables used as predictors and (v) PTF evaluation criteria. We emphasize the growing use of machine learning algorithms as a technique to derive PTFs and delve deeper into the discussion of this aspect. Our work also highlighted the challenges concerning the prediction of SHP in tropical regions. As a case study to address such challenges, we used existing temperate and tropical PTFs to predict soil water retention and soil hydraulic conductivity of tropical soils from an important tropical agricultural region, the Mato Grosso state bordering Bolivia in western Brazil.

2.2 Material and Methods

2.2.1 Data collection

An extensive literature review was carried on peer-reviewed articles available on Elsevier’s Scopus bibliographic database. It was assumed that articles would have at least the word “pedotransfer” or “pedo-transfer” in the title or keywords, along with “hydraulic conductivity” or words related to soil water retention (e.g., “field capacity”, “available water”) in the title or keywords. The document type was limited to scientific articles written exclusively in English. To meet the abovementioned criteria, the following search code was used:

(TITLE ("soil hydr*" OR "water retention" OR "field capacity" OR "available water" OR "permanent wilting point" OR "hydraulic conductivity") AND KEY (pedotransfer OR "pedo transfer")) OR (TITLE ("pedotransfer" OR "pedo transfer") AND KEY ("soil hydr*" OR "water retention" OR "field capacity" OR "available water" OR "permanent wilting point" OR "hydraulic conductivity")) AND (LIMIT-TO (DOCTYPE, "ar")) AND (LIMIT-TO (LANGUAGE, "English")).

The literature search was performed on June 8, 2021 and yielded a total of 444 articles, 58% (n = 257) published between 2012 and 2021, 32% (n = 144) published between 2002 and 2011 and 10% (n = 43) published between 1989 and 2001. With the 257 articles from 2012-2021 we performed a refined search assessing each article individually. The refined search aimed to select studies that exclusively focused on developing PTFs to predict SHP, excluding studies that, for instance, focused on the evaluation of existing PTFs.. This refined search selected 101 of the original 257 articles. Most of the discarded articles were studies using and comparing existing PTFs in Earth system models.

The selected 101 articles were submitted to data extraction. Data in tables were directly registered in the study database, whereas data from graphs and figures were extracted using the WebPlotDigitizer (https://apps.automeris.io/wpd/). In total, 1844 different PTFs were recorded and used to generate a database containing qualitative and quantitative information. Qualitative data were: PTF areal coverage/domain, predominant climate (tropical or temperate), specific location or database used to develop PTFs, country, target variable, type of PTF (continuous or class), method used to derive PTF, algorithms used (in case of machine learning techniques), variables used as predictors, model sensitivity assessment (yes or no), uncertainty quantification in the final prediction (yes or no). Countries considered as tropical were those that have most of their territory between the tropics (e.g., Brazil), or at least have considerable influence of the tropical climate in part of their territory (e.g., Australia). Countries located between the tropics in the Middle-East (i.e., Saudi Arabia) and North Africa (i.e., Algeria and Egypt) were not considered as tropical countries as they have a predominantly arid climate. This distinction was made to identify tropical countries that are developing PTFs, as much effort has been spent in recent years to develop representative PTFs for tropical soils, as mentioned in the introduction to this study.

Word clouds were made to investigate the most used predictors of each target variable. It is noteworthy that soil water retention was considered as a single variable, despite being composed of several variables (e.g., field capacity, permanent wilting point, available water, parameters of empirical models) and each publication may have had a distinct purpose of prediction. However, for the sake of simplicity, all variables related to soil water retention were considered in the same group (Fig. 3), while hydraulic conductivity was evaluated separately (Fig. 4). The predictors were subdivided into four groups: soil physical properties, soil chemical properties, soil morphological properties and environmental covariates. Word size is proportional to frequency of occurrence.

Quantitative information for each function were: number of samples for model training, number of samples for independent testing, as well as several error evaluation metrics and goodness-of-fit criteria, the most recorded being the coefficient of determination (R^2), the root mean square error (RMSE), the mean absolute error (MAE) and the mean bias error (MBE). The set of all recorded error and goodness-of-fit metrics is fully availably in the supplementary material of this work.

Qualitative data were analyzed according to the frequency of use. Error and goodness-of-fit metrics were evaluated in histograms to show the frequency density and bring to light the reliability of the predictions made during the studied decade (2012-2021). For the specific case of hydraulic conductivity estimation, different target variables were found, such as Ks, ln Ks and $\log_{10}$Ks. Therefore, given the impossibility of standardizing the error metric values without accessing the original dataset of each article, we selected only the metric values of PTFs that used $\log_{10}$Ks as its target variable, which was the most frequent in our review (60% of hydraulic conductivity articles). Given that, 17 articles were discarded and data were collected from 26 articles out of the 43 articles that addressed hydraulic conductivity.

The quantification of uncertainties associated with PTFs predictions was investigated in each article of this review, considering "yes" or "no" for the quantification of these uncertainties, without extracting values or magnitude from the data. The uncertainty quantification was considered in works that quantified metrics such as standard deviation (standard uncertainty), standard error, or those at least presented confidence intervals or prediction intervals in the results, e.g., through boxplots. Works that performed sensitivity analyses were considered as those that quantified model sensitivities, e.g., systematically changing one variable while keeping the others constant (i.e., local sensitivity analysis), excluding works that only mentioned the term “sensitivity” speculatively.

2.2.2 Case study: performance of existing PTFs to predict soil hydraulic properties of a soil database from Mato Grosso state (Brazil)

A dataset containing hydraulic properties of soils from the Mato Grosso state (Brazil) was used to perform a practical evaluation of existing PTFs, using both classical PTFs and recently developed Brazilian PTFs. The Mato Grosso dataset was assembled to perform this study and part of this database was used in previous works in the region (Bocuti, 2020; Bocuti et al., 2021). The dataset contains soil data from the Cerrado and Amazon region, the largest biomes in Brazil and extremely important for the country’s agriculture. The dataset (n = 228) contains SHP information such as soil water retention related properties (e.g., field capacity, permanent wilting point, available water capacity) and saturated hydraulic conductivity. The data refer to topsoil layer (0-20 cm) of Acrisols, Cambisols, Gleysols, Ferralsols, Arenosols, Nitisols, Planosols and Plinthosols (WRB, 2015), corresponding to *Argissolos*, *Cambissolos*, *Gleissolos*, *Latossolos*, *Neossolos*, *Nitossolos*, *Planossolos* e *Plintossolos* in the Brazilian Soil Classification System (Santos et al., 2018). Descriptive statistics of the database used for this section is provided in Table 1.

Table 1. Descriptive statistics of the soil database used for PTFs evaluation (n = 228). Data refer to several soils from Mato Grosso state (Brazil).

Soil properties	Min	Median	Max	Mean	SD
Total sand (%)	1.59	58.25	95.30	55.62	22.61
Coarse sand (%)	0.36	18.61	48.34	20.02	11.60
Fine sand (%)	1.22	34.40	85.86	35.60	19.88
Silt (%)	1.22	9.23	63.79	13.34	12.64
Clay (%)	1.34	31.09	67.21	31.03	16.46
Bulk density (Mg m^{-3})	1.02	1.47	1.84	1.46	0.18
Organic matter (%)	0.34	0.89	1.95	0.99	0.41
Total porosity (m^3 m^{-3})	0.25	0.44	0.75	0.45	0.08
Effective porosity (m^3 m^{-3})	0.07	0.26	0.60	0.26	0.07
θ_{10kPa} (m^3 m^{-3})	0.11	0.22	0.42	0.22	0.05
θ_{33kPa} (m^3 m^{-3})	0.07	0.19	0.38	0.19	0.05
$\theta_{1500kPa}$ (m^3 m^{-3})	0.04	0.13	0.35	0.13	0.05
Ks (cm day^{-1})	0.16	66.51	1893.25	146.46	224.04
log[Ks/(cm day^{-1})]	-0.80	1.82	3.28	1.77	0.65

θ_{10kPa} = soil water content at matric potential of -10 kPa; θ_{33kPa} = soil water content at matric potential of -33 kPa; $\theta_{1500kPa}$ = soil water content at matric potential of -1500 kPa; Ks = soil hydraulic conductivity; logKs = logarithm of soil hydraulic conductivity.

The predictions were made to estimate the field capacity (assumed as soil water content at matric potential of -10 kPa), the permanent wilting point (assumed as soil water content at matric potential of -1500 kPa) and the saturated hydraulic conductivity. The target variables were predicted using PTFs from the literature and the mean value of each target variable (i.e., null model). The null model is the simplest possible prediction and its use is motivated to verify the parsimony of using complex equations and data modeling techniques. For instance, if complex models do not show substantial advantages compared to the null model, their use is therefore not justified. For soil water retention of matric potentials at -10 kPa and -1500 kPa, four PTFs were selected, one of global coverage, one of continental coverage, one of national coverage and one of regional coverage. The PTF of global coverage was the H3 function from Rosetta software (Schaap, 2001) using its weighted recalibrated version (Zhang and Schaap, 2017), which has sand content, silt content, clay content and bulk density as input data. The PTF of continental coverage was the model 21 from Tóth et al. (2015), recently developed for European soils using silt, clay, organic carbon and bulk density as predictors. The national PTF was the one by Tomasella et al. (2000) which uses coarse sand, fine sand, silt, clay, organic carbon and bulk density as input data. Even though this latter model was not developed recently as the others, it is a classical reference for water retention prediction of Brazilian soils and remains widely used. Finally, as a representation of regional PTF with a geographic domain similar to our dataset we used the PTF of Medrado and Lima (2014), who proposed a water retention estimation PTF for soils from the Brazilian Cerrado biome, which covers 40% of the Mato Grosso state area. This latter model used sand, silt, clay, bulk density and organic matter as predictors.

To evaluate soil hydraulic conductivity predictions, two functions (WM330 and WMF330) from Ottoni et al. (2019) were selected. The authors proposed global PTFs using data from tropical and temperate soils for training the models. The WM330 and the WMF330 models consider effective porosity as a predictor of hydraulic conductivity, either directly measured (WM330) or estimated (WMF330) by linear regression based on basic soil properties (sand, silt, clay and bulk density). Two functions of the classical Rosetta software (Schaap, 2001) were also used in their weighted recalibrated version (Zhang and Schaap, 2017). The H3w version of Rosetta3 consider only basic soil variables (i.e., sand, silt, clay, and bulk density) as input data, while the H5w model considers, in addition to basic variables, the soil water content at -33 kPa and -1500 kPa as predictors of hydraulic conductivity.

The same existing PTFs were also tested for the same target variables after splitting the database by soil texture groups, evaluating the performance of each PTF for soils of coarse textures (< 15% of clay content), medium textures (> 15% and < 35% of clay content), and fine textures (> 35% of clay content). To evaluate the PTFs goodness-of-fit we used the R^2, RMSE, MAE, MBE and Lin's Concordance Correlation Coefficient (LCCC).

2.3 Results and Discussion

2.3.1 Overview

Altogether, 101 scientific articles were included in this review. However, to evaluate soil water retention and hydraulic conductivity separately, we considered articles that modeled both properties as different "works" in the same article. Thus, 76 PTF works were registered for water retention and 31 PTF works were registered for hydraulic conductivity, totaling 107 individual works. Gathering all these works, we recorded 1844 different PTFs, the vast majority for soil water retention (n = 1627) and 217 for hydraulic conductivity.

Regardless of the target variables, the expressive majority of PTFs were of continuous type, with few records of class-type PTFs. In total, 96% of the registered PTFs were continuous-PTFs (n = 1774). For soil water retention, of 1627 recorded PTFs, 1561 (96%) were continuous-PTFs and 66 (4%) were class-PTFs. For soil hydraulic conductivity, 213 (98%) of 217 PTFs were continuous-PTFs, whereas only 4 were class-PTFs. These results indicate that little has been done to develop class-PTFs, not only in the humid tropics as shown by Botula et al. (2014), but in the world in general. Despite having several facilities for the user, such as a little, simple and cheap information input (e.g., textural class) (Al Majou et al., 2018, 2008), class-PTFs may have poorer prediction than continuous-PTFs, which explains the low motivation to build them. Class-PTFs are more general applicable and have limitations in recognizing specific site conditions, as soils with the same textural class may show great differences among their features (Hodnett and Tomasella, 2002; Wösten et al., 1999).

3.2 PTFs location and their coverage proposal

PTFs for estimating SHPs were developed in all continents in the period between 2012 and 2021. Figure 1 highlights all countries where PTFs have been developed, with a special emphasis on tropical countries, colored green. Among the 18 works with an international purpose, i.e., which involved datasets from more than one country, 7 studies contained data from tropical countries. The tropical country that presented most studies on PTFs development was Brazil (n = 9), being the third country in the ranking of PTFs development in the world during 2012-2021 and according to the selection criteria, behind China (n = 12) and Iran (n = 11).

The results showed a considerable production of PTFs by tropical countries, mainly Brazil (n = 9), India (n = 7) and Vietnam (n = 4). This greater role of tropical countries in the development of PTFs indicates that the urgent need for accurate information on soil properties for the tropics, as highlighted by Minasny and Hartemink (2011), is being pursued. Minasny and Hartemink (2011) found that most PTFs developed in the world came from temperate climate countries, especially from Europe and North America. Our results showed an outlook change from 2012 to 2021, for example, more studies of PTF development were registered in Brazil than in the USA and any European country.

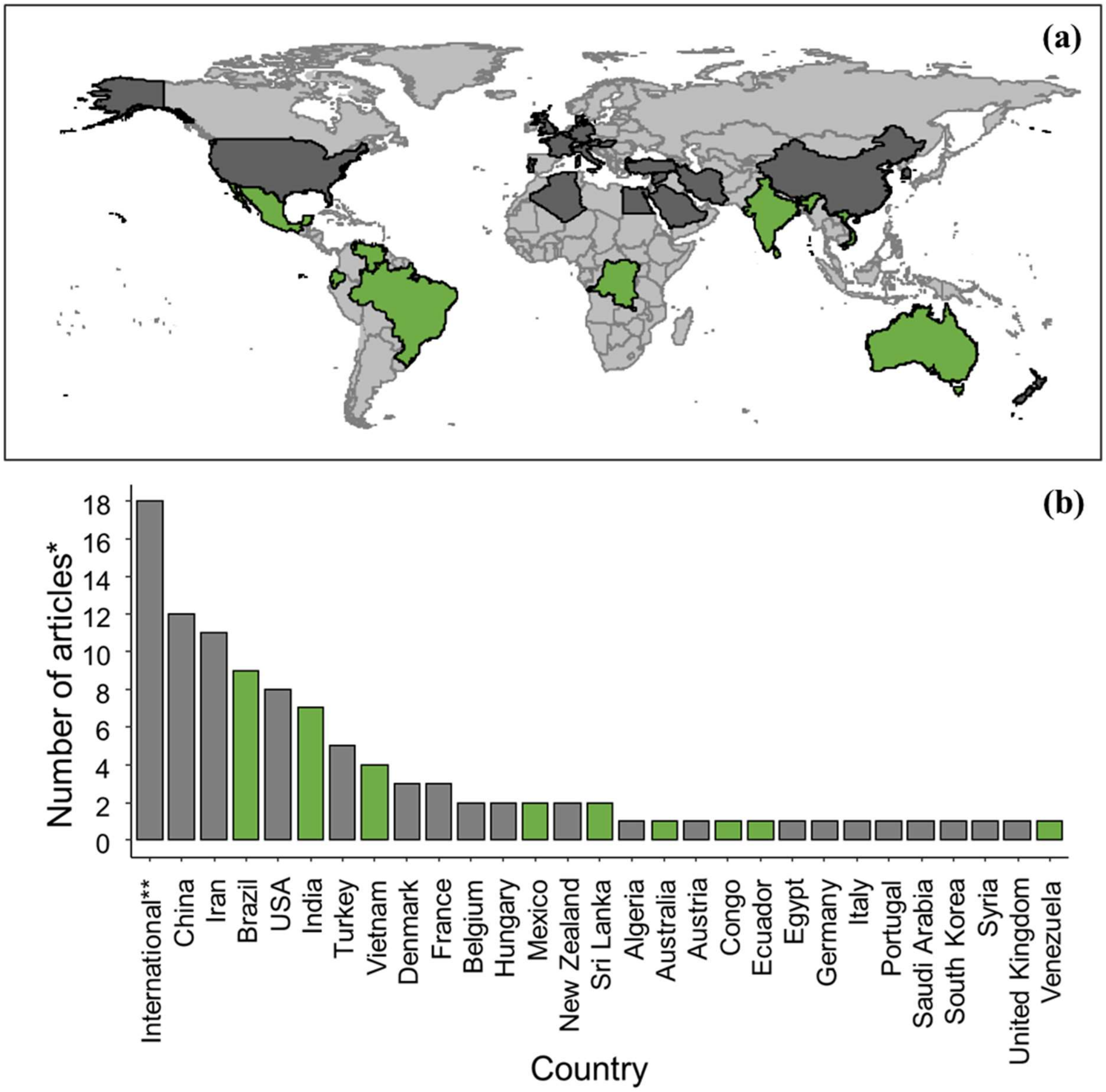

Figure 1. (a) Countries with studies evaluated in the review and (b) number of articles per country. Countries considered as predominantly tropical or under considerable influence of tropical climate are highlighted in green. *The sum of articles may differ from the total of publications (n = 101) as the same publication may have encompassed one or more countries. **The term "international" refers to publications that involve data from more than one country in their database.

The PTFs areal coverage proposal was divided by the following purposes: global, continental, national, regional and local (Figure 2). PTFs that used datasets from more than one country were considered to have global coverage, e.g., Ottoni et al. (2019) who used datasets from a region with a temperate climate (Denmark) and a tropical climate (Brazil). It is noteworthy that the subdivision of coverage purposes is subjective and was conditioned by the authors' point of view. It should also be noted that the coverage proposal does not concern the area of PTFs coverage in quantitative terms, since regional functions can cover large areas. For instance, the work by Achieng (2019) proposes PTFs for estimating soil available water in the Ogallala aquifer region (USA), covering an area of ~450 thousand km^2 and was considered as of regional purpose, whereas the work by Pollacco et al. (2017) proposes functions for New Zealand (~268 thousand km^2) and was considered as of national coverage purpose. The number shown in Fig. 2 may differ from the total number of publications (n = 101) as a same article may have produced PTFs with different coverage proposals.

Most of the studies had a regional coverage purpose (n = 59). Studies with national (n = 19) and global (n = 14) purpose came next, having greater representation than studies with a local (n = 9) and continental coverage purpose (n = 2). It is physically plausible that PTFs are developed on a regional basis, by means of recognizing environmental patterns and natural similarities, rather than arbitrary human-made divisions such as country borders. A typical example of a regional study where natural similarities overpass human-made boundaries is the study by Kalumba et al. (2020) in which the authors developed PTFs for SHP of a river basin whose area spans several countries in southern Africa.

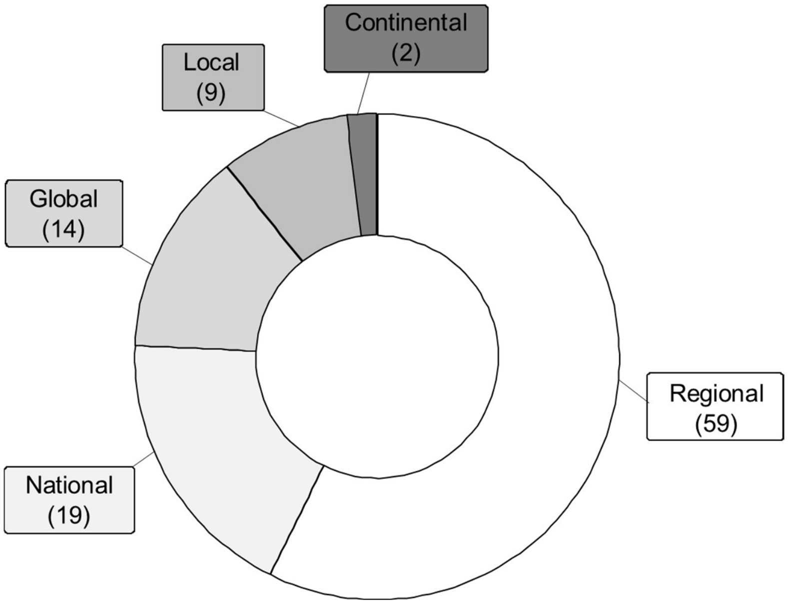

Figure 2. Number of articles by PTFs coverage proposal. The number in parentheses corresponds to the number of studies, which may differ from the total publications (n = 101) as a single study may have developed PTFs with different coverage proposals.

National PTFs can be interesting to support public policies in a generalized way. Future efforts may be towards combining regional datasets to form databases and PTFs that are nationally representative, especially in countries with large areas, such as China, USA, and Brazil, which were among the largest developers of PTFs in the last years. For instance, Brazil has five major biomes and establishing a database and PTFs representative of all these environmental units is an ongoing challenge (Ottoni et al., 2018). Meanwhile, regional PTFs have been developed for different environmental units, such as Brazilian semiarid region (Barros et al., 2013), Brazilian Coastal Plains (Silva et al., 2015), Cerrado biome (Medrado and Lima, 2014) and Amazon (Tomasella and Hodnett, 1998). These are typical examples of geographically specific PTFs, which tend to continue being produced until generic PTFs are developed and validated (Patil and Singh, 2016).

Studies to build generic global PTFs are still incipient. Even large and multinational datasets as UNSODA (Nemes et al., 2001) and HYPRES (Wösten et al., 1999) have gaps and limitations to develop PTFs able to reliably recognize and generalize global SHP patterns (Singh et al., 2020). Studies with a global proposal are emphatic about the underrepresentation of SHP data in many parts of the world, which leads to poor prediction performances (Gupta et al., 2021; Tian et al., 2021). For the specific case of soil hydraulic conductivity, Gupta et al. (2021) emphasizes the scarce data availability in tropical regions.

2.3.3 Predictors: soil properties and environmental covariates

The results in Fig. 3 and Fig. 4 indicate that PTFs are strongly based on “basic” soil properties, such as soil texture (i.e., sand, silt, and clay), bulk density and organic carbon, as also reported by Fatichi et al. (2020) and Patil and Singh (2016). This finding is in accordance with PTFs principles, which is to predict complex attributes by means of attributes with simpler acquisition. Nonetheless, the expressive presence of several alternative predictors was noticed, the most common being soil physical properties, followed by soil chemical properties, morphological attributes and environmental covariates. The selection of alternative predictors to improve PTF prediction has been discussed for a considerable time (Pachepsky et al., 1996; Rawls and Pachepsky, 2002a, 2002b). For instance, Pachepsky and Rawls (2003) suggest structural attributes as potential predictors and Lilly and Lin (2004) extensively discuss the use of morphological attributes (e.g., horizons, soil color and concretions). Some authors found promising results using alternative predictors in PTFs, such as the plastic limit (Khlosi et al.,2016; Nguyen et al., 2014) , which is related to other soil properties such as clay and mineralogy.

Morphological attributes can be important predictors as they carry correlations with properties measured in the laboratory and are commonly available in soil survey databases (Lilly and Lin, 2004). For instance, Nguyen et al. (2014) found that qualitative soil structure information improved the prediction of soil water retention and Moncada et al. (2014) concluded that visual soil quality description is a promising feature to be used in predictions as it merges several relations of different soil attributes into a single information. Nevertheless, reports about the use of morphological attributes in PTFs structure are scarce (Padarian et al., 2018; Patil and Singh, 2016) and could be further explored. However, advances in remote sensing techniques have allowed the use of other data that are also easily obtained and widely available as potential predictors, such as topographic attributes. Terrain data can be extracted from digital elevation models and may enable the generation of variables directly related to soil properties and geomorphological, hydrological and climatic processes (Patil and Singh, 2016). The use of terrain attributes such as slope, altitude and relief for PTF development is present in several SHP models (e.g., Gupta et al., 2021; Kalumba et al., 2020; Lim et al., 2020; Picciafuoco et al., 2019; Romano and Chirico, 2004; Romano and Palladino, 2002; Szabó et al., 2019; Wang et al., 2012; Yang et al., 2018).

Another important factor used as predictor was the parent material, which is a soil formation factor along with climate, organisms, topography and time. Bruand (2004) recommended the preliminary grouping of soils based on parent materials for the development of PTFs, indicating the importance of this pedogenetic factor in distinguishing soil properties. Bruand (2004b) also suggests the possibility of using chemical and mineralogical attributes as potential predictors of PTFs, as they are directly related to the pedogenesis process. Tóth et al. (2015) found that using pH, CEC and $CaCO_3$ improved the prediction of some SHP, Khlosi et al. (2013) found $CaCO_3$ among the most important predictors of water retention of calcareous soils and Hodnett and Tomasella (2002) showed that using pH in PTFs structure improved the prediction of the van Genuchten (1980) equation parameters. Moreover, CEC was an important predictor for soil water retention in some studies (Botula et al., 2013; Hodnett and Tomasella, 2002) as well as DCB-Fe which reduced uncertainty in soil water retention predictions of tropical soils (Botula et al., 2012). However, diverging from the classification of parent materials and terrain attributes, chemical and mineralogical variables impose greater difficulty in obtaining them, as they depend on sample collection and laboratory analyses.

Climate attributes also have potential as predictors, althoug they are considerably less observed compared to topographical attributes and parent material. The ease of obtaining these data by automatic stations and the availability of data distributed over time make climate variables worthy of attention regarding PTF development, as already observed in recent literature (Gupta et al., 2021; Hollis et al., 2015; Szabó et al., 2019). A promising example is the work by Malone et al. (2020), who concluded that climatic variables (e.g., temperature and rainfall) were important spatial predictors of SHP in agricultural soils of Australia. As a final and general example, the study by Szabó et al. (2019) is an important reference concerning the use of alternative variables as predictors of SHP. The authors used the following information as predictors: chemical properties, topographical attributes, vegetation attributes, climatic attributes, geographic coordinates and parent material, in addition to the commonly used basic attributes.

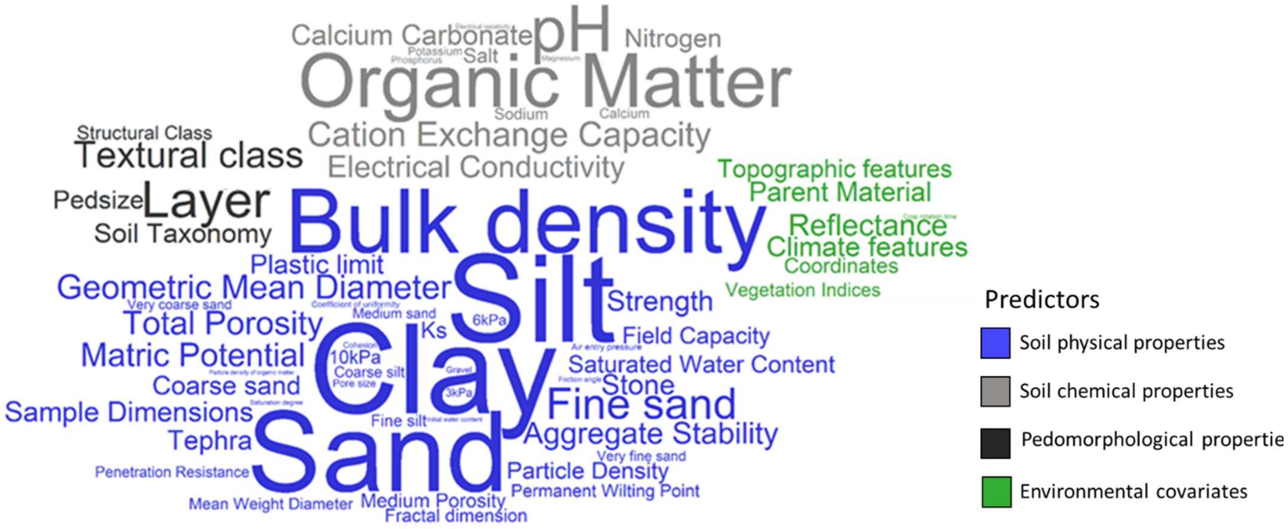

Figure 3. Word cloud of predictors used in PTFs to predict soil water retention. Word size is proportional to the occurrence of use as predictors.

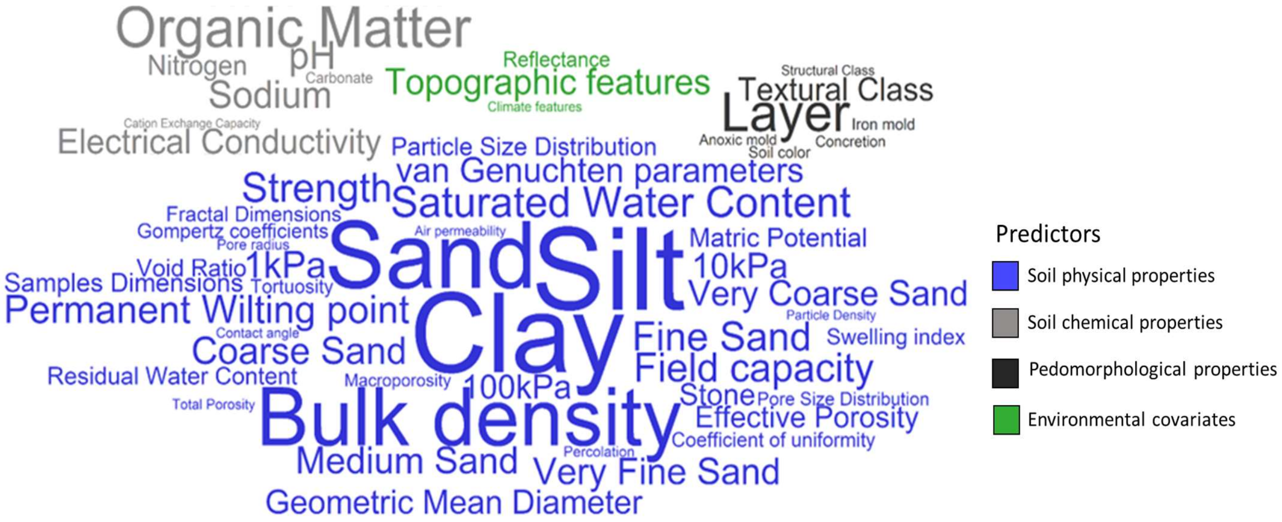

Figure 4. Word cloud of predictors used in PTFs to predict soil hydraulic conductivity. Word size is proportional to the occurrence of use as predictors.

2.3.4 Methods to derive PTFs: the machine learning age?

Our question about the machine learning age relies on the well-known increasing use of artificial intelligence techniques in the scientific process over the past decade, including hydrology (Nearing et al., 2020) and soil science (Padarian et al., 2020). We investigated the methods used to develop PTFs and the results were separated into three broad groups of methods: semi-physical, conventional regression and machine learning. Figure 5 presents the total number of articles subdivided by the method used to develop PTFs. The results show that the use of conventional regression is predominant (n = 61) over the use of machine learning techniques (n = 51) and semi-physical approach (n = 10). It is noteworthy that the number of what we consider as "studies" (n = 122) differs from the total number of publications evaluated in this review (n = 101), as the same publication may have tested different methods of obtaining PTFs.

The statement made by Schaap and Leij (1998) that most PTFs are empirically-based and use conventional regression techniques for their modeling remains valid. Our results indicate that 92% of studies developing PTFs for SHP prediction between 2012 and 2021 used empirical approaches and only 8% of studies developed semi-physical PTFs. This result is consistent with Botula et al. (2014) who reviewed PTFs for soil water retention and found only 9% of PTFs built on a semi-physical approach. The low development of physically-based PTFs is explained by the usual need to obtain very detailed information on particle size distribution (Schaap, 2005), which makes prediction almost as difficult as obtaining the target variables by direct measurements. Depending on the equipment, budget and technology available, this drawback may even violate the premise of efficiency for PTFs development proposed by Minasny and McBratney (2002b), who stated that PTFs should not be built if direct measurement is easier or cheaper. Nevertheless, the advance of laser diffraction methods has shown promising results towards the determination of detailed particle size distribution (Bittelli et al., 2022) and can make physically-based PTF development easier, although discrepancies compared to standard methods are still an issue (Svensson et al., 2022).

The predominance of empirical modeling in developing PTFs remains current though the expressive number of studies using machine learning - almost equivalent with the use of conventional regression - endorses recent statements that highlight the rapid growth of new data science techniques for PTFs development (Gebauer et al., 2020; Van Looy et al., 2017). For instance, in the review by Botula et al. (2014), 97% of empirically-based PTFs were developed using conventional regression techniques and only 3% used machine learning techniques. These results strongly differ from the results shown in Fig. 5, where machine learning techniques were used in 45% of studies that developed empirically-based PTFs, whereas conventional regression accounted for 55%. This observed growth in machine learning as a tool to build PTFs supports the statements that this increased use of new data science techniques is a characteristic of the last decade (Padarian et al., 2020), confirming that we are experiencing a new information age in soil science and hydrology (Nearing et al., 2020).

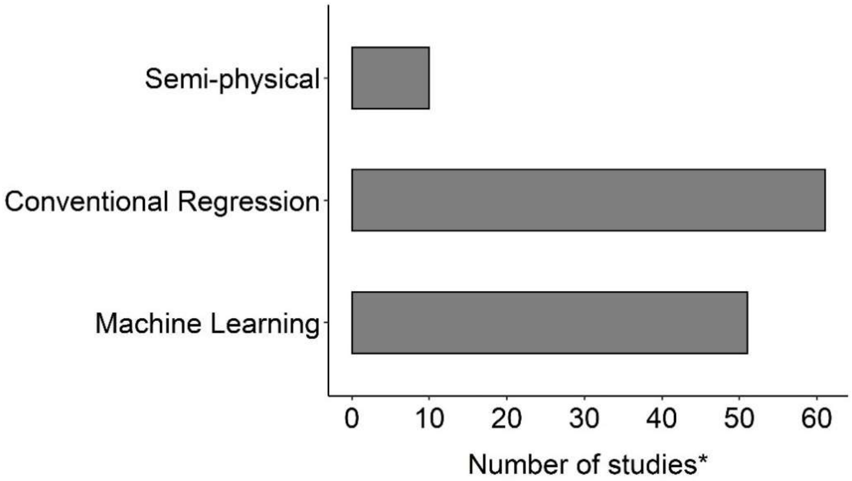

Figure 5. Methods used to develop PTFs. *The sum of studies may differ from the total number of publications (n = 101) as the same publication may have used one or more methods to derive PTFs.

Given the importance of machine learning to develop PTFs in the present time, we further investigated this matter. By investigating the algorithms used in the studies based on machine learning (Fig. 6a), 14 different algorithms were found. Artificial neural networks (ANN) were predominant (n = 30), followed by k-Nearest Neighbor (KNN, n = 10), Random Forest (RF, n = 8) and Support Vector Machine (SVM, n = 7) algorithms. As stated by Botula et al. (2013), ANN, KNN and SVM have been successful for predictive modeling in the field of soil hydrology. As ANN is a classical and powerful method for predictive modeling used to estimate SHP since the 1990s (Pachepsky et al., 1996; Schaap and Leij, 1998), the high adoption of this algorithm is plausible. KNN is one of the most widely used and well-known algorithms in the scientific community and SVM is considered one of the most robust for regressions (Wu et al., 2008), which also explains their wide use. The RF algorithm has been widely used for PTF development in recent years, given its high predictive performance (Gunarathna et al., 2019a; Rastgou et al., 2020). For detailed information on the use of KNN, SVM, and RF algorithms for PTF development the reader is referred to Van Looy et al. (2017).

Despite the great availability of machine learning models, choosing which one to use a priori is not a simple task. As stated by Wolpert (1996) in the “No Free Lunch” Theorem, the performance of a model is strongly conditioned by the modeling problem, so there is no model that will always perform better than the others. Thus, it is recommended to test several algorithms before deciding which one to use in the final prediction (Kuhn and Johnson, 2013). In this sense, we quantified the number of algorithms tested in studies that developed PTFs using the machine learning approach (Fig. 6b). Most works used only one algorithm (n = 23), followed by the use of two algorithms (n = 20) and few examples of works testing three (n = 4) and four different algorithms (n = 4). No papers were found comparing five or more machine learning algorithms.

The importance of evaluating different algorithms when applying machine learning techniques is confirmed by recent works, as divergent results were found even when the modeling process had the same purpose. As examples of irregular results, when modeling soil water retention, Gunarathna et al. (2019a) concluded that RF algorithm was better than ANN, KNN and multiple linear regression, Rastgou et al. (2020) found that polynomial regression performed better than RF, whereas Cueff et al. (2021) observed similar poor performances for RF, regression trees and multiple linear regression. Patil et al. (2013) found a better performance of KNN than ANN, Khlosi et al. (2016) concluded that SVM was better than ANN and multiple linear regression, while Kalumba et al. (2020) found similar results for ANN, RF, SVM and multiple linear regression. As a final example, Kotlar et al. (2019b) developed different PTFs for several matric potentials and all algorithms were listed at least once as the best.

The divergent results regarding models' performance for predicting soil water retention strongly agree with the "No Free Lunch" Theorem, indicating that the model's goodness-of-fit is also conditioned by characteristics of the training database. The evaluation of different algorithms seems not to have been considered in the evaluated period (2012-2021), since only 16% of the machine learning studies tested three or more algorithms. Considering the recent computational advances, which provide a reasonable ease of implementation of machine learning algorithms in programming environments, it is worth to consider the evaluation of different algorithms for PTF development, rather than arbitrarily choosing an algorithm a priori.

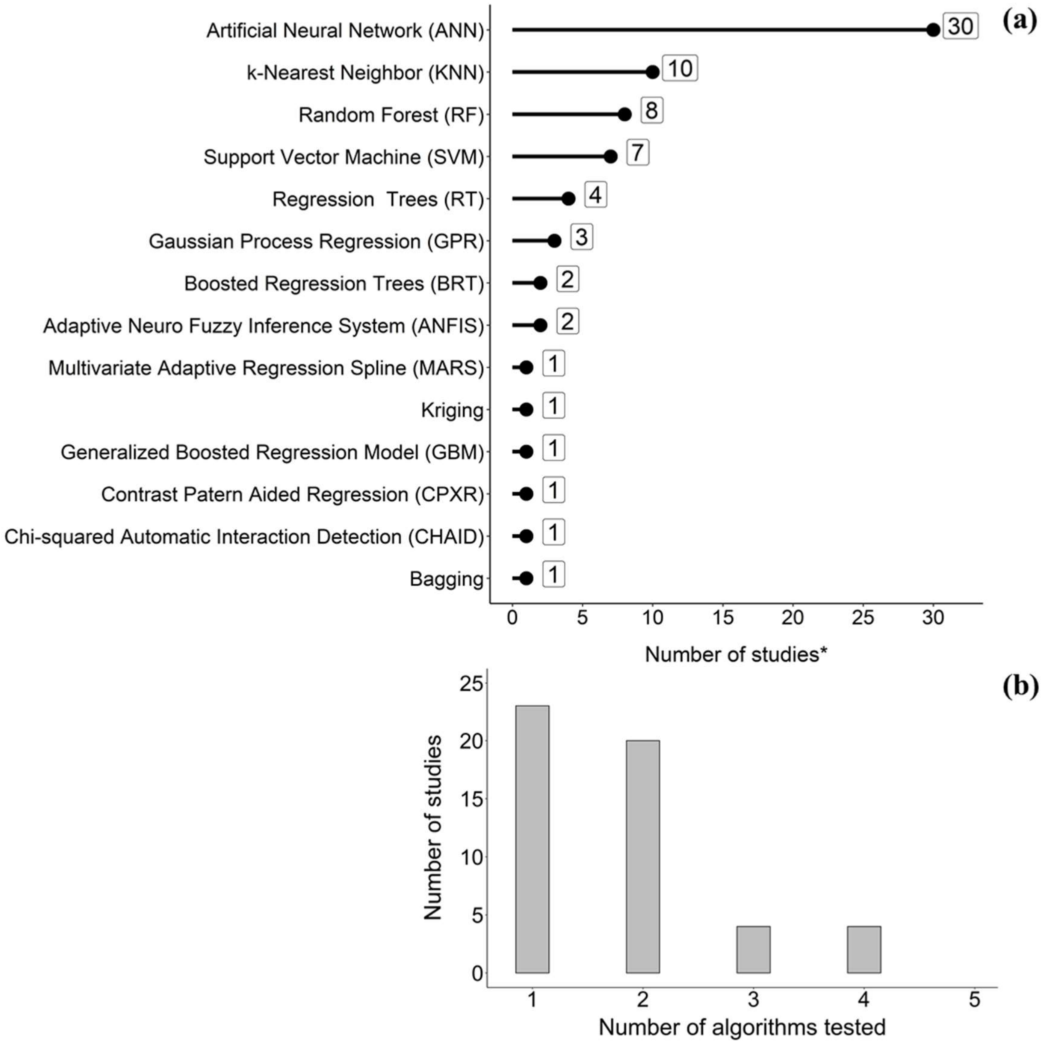

Figure 6. (a) Algorithms used for PTFs development based on machine learning and (b) number of algorithms tested per study. *The sum of studies in (a) may differ from the total number of machine learning studies (n = 51) as the same publication may have used one or more algorithms. Reader is referred to the supplementary material where every reference is listed along with the chosen algorithms at the column “algorithm”.

2.3.5 PTF development and evaluation criteria

The dataset sizes (i.e., the number of samples) used for training and independent testing of the PTFs for each target variable are shown in Figure 7. For this purpose, the 1844 PTFs encompassed in this review were stratified by datasets, since a single dataset can originate several functions. The histograms show the number of samples per number of datasets. Altogether, 214 distinct datasets were found, 171 used to predict water retention and 43 to predict hydraulic conductivity. Out of this total, 122 and 33 datasets were stratified for independent testing of water retention and hydraulic conductivity PTFs, respectively. It is noteworthy that we did not consider cross-validation as testing. Concerning datasets that were not stratified for independent testing of PTFs for water retention (n = 49), 31 used cross-validation and 18 only trained PTFs. For hydraulic conductivity PTFs, the datasets that were not stratified for independent testing (n = 8), 4 used cross-validation and 4 only trained the PTFs. For both target variables, the peak size of datasets was around 100 samples and studies rarely used more than 1000 samples (Fig. 7a and 7b), either for training or testing the PTFs.

To provide a zoom into the most frequent range of dataset sizes, additional histograms were made considering only datasets with less than 1000 samples (Fig. 7c and 7d). For both soil water retention and hydraulic conductivity most datasets contained less than 200 samples. Other authors also emphasize the majority development of PTFs from small databases, especially in tropical conditions (Ottoni et al., 2019; Silva et al., 2017), which may lead to limitations in PTFs application (Haghverdi et al., 2012; Nguyen et al., 2015a), such as high bias (Nguyen et al., 2015b) and contradictory results when applying the PTFs using different databases (Ghanbarian et al., 2017). It is well known that SHP commonly exhibit high spatial variability (Schaap, 2001), which may cause error propagation in final predictions when this variability is underrepresented (Minasny et al., 1999).

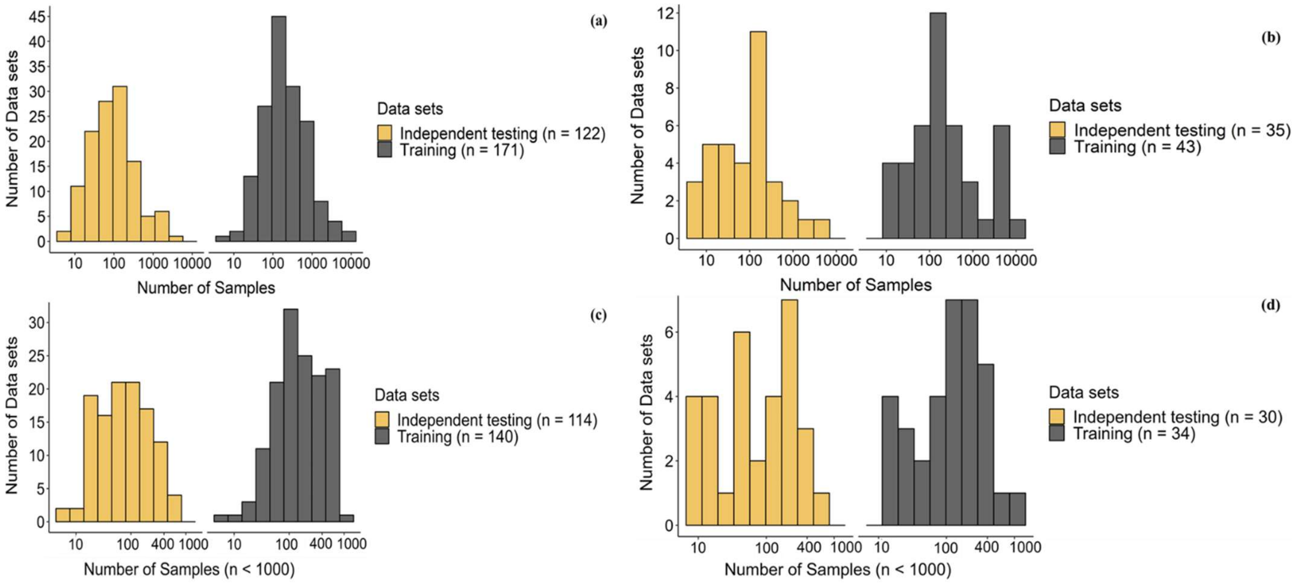

Figure 7. Frequency density histograms of datasets size used for training and testing PTFs for predicting soil water retention (a) and hydraulic conductivity (b). Density of datasets size considering only datasets with less than 1000 samples for predicting soil water retention (c) and hydraulic conductivity (d).

The most common statistical metrics used for PTFs performance evaluation were: R^2, RMSE, MAE and MBE. Figure 8 (water retention) and Figure 9 (hydraulic conductivity) show the frequency density of metric values, stratified by the type of method used to generate the PTFs. For soil water retention PTFs (n = 1627), 922 used R^2, 1335 used RMSE, 300 used MAE and 888 used MBE (Fig. 8). It is noteworthy that R^2 does not consider the bias in a model and may be optimistic in some cases. For hydraulic conductivity PTFs (n = 217), 137 evaluated the performance of PTFs using R^2, 187 used RMSE and 61 used MBE (Fig. 9). The Akaike Information Criteria (AIC) was also a metric widely used to compare models within the same work (n = 196 for water retention and n = 53 for hydraulic conductivity), but, as its raw value is not interpretable, they were not extracted. The metrics extracted from the literature concern the performance of independent test sets or cross-validation of functions.

Results showed higher accuracy of machine learning-based PTFs compared to conventional regression for soil water retention. PTFs based on machine learning showed a distribution of R^2, MAE and RMSE values with medians of 0.820, 0.033 m^3 m^{-3} and 0.049 m^3 m^{-3}, respectively. PTFs based on conventional regression showed a distribution of R^2 MAE and RMSE values with medians of 0.698,0.048 m^3 m^{-3} and 0.051 m^3 m^{-3}, respectively. It is important to note that these metrics include results of different matric potentials combined and must be seen as an overview on how PTFs for soil water retention prediction performed. The accuracy of soil water retention predictions is not uniform and varies according to specific targets, e.g., usually lower accuracy is found in the wet range compared to the dry range of the soil water retention curve (Kotlar et al., 2020; Wassar et al., 2016). Higher accuracy of machine learning-based PTFs was also observed when the prediction target was soil hydraulic conductivity. R^2 values for machine learning-based PTFs were substantially higher (median = 0.867) than values recorded for conventional regression-based PTFs (median = 0.427). RMSE values also demonstrated the best performance of machine learning algorithms in predicting soil hydraulic conductivity, showing a median of 0.528 compared to 0.785 for conventional regression-based PTFs. Semi-physical models also generally had high goodness-of-fit indices;

however, as already discussed, they are underrepresented in this study, as 92% of the works used in this review developed PTFs using conventional regression or machine learning techniques.

The fact that PTF performance improves using machine learning techniques for calibration was corroborated by our results. Padarian et al. (2020) highlight the ability of machine learning algorithms to capture non-linear relationships between soil properties, generally presenting better performance than the conventional approaches. The better ability of machine-leaning-based PTFs to capture complex relationships between soil properties is also highlighted by Araya and Ghezzehei (2019), giving as an example the relationship between bulk density and soil hydraulic conductivity, that is hardly described by conventional regression and physically-based methods. Compared to conventional regression and numerical models, machine learning models performed better in generating PTFs for soil water retention curve in the study by Amanabadi et al. (2019). Similarly, Gunarathna et al. (2019a) concluded that machine-learning-based PTFs improved the accuracy of PTFs for water retention prediction of a tropical soils database. Nonetheless, despite the general advantages found in the use of machine learning techniques, specific caveats require attention.

Evaluating the development of PTFs for soil water retention, Kotlar et al. (2019b) concluded that conventional regression techniques performed better than machine learning algorithms in some cases. Kalumba et al. (2020) found similar results for machine learning and conventional regression performances for SHP prediction. These results demonstrate that the use of machine learning should be encouraged, although it is reasonable to compare machine learning with conventional regression, since the latter has advantages such as requiring less computational effort and higher ease to use and interpret (Van Looy et al., 2017). Finally, there is still a quest to improve the performance of current machine learning algorithms in predicting SHP, as noted in the review by Zhang and Schaap (2019), specifically concerning hydraulic conductivity. According to the authors, despite the several recent advances of machine learning models, the reliability of hydraulic conductivity predictions is still modest.

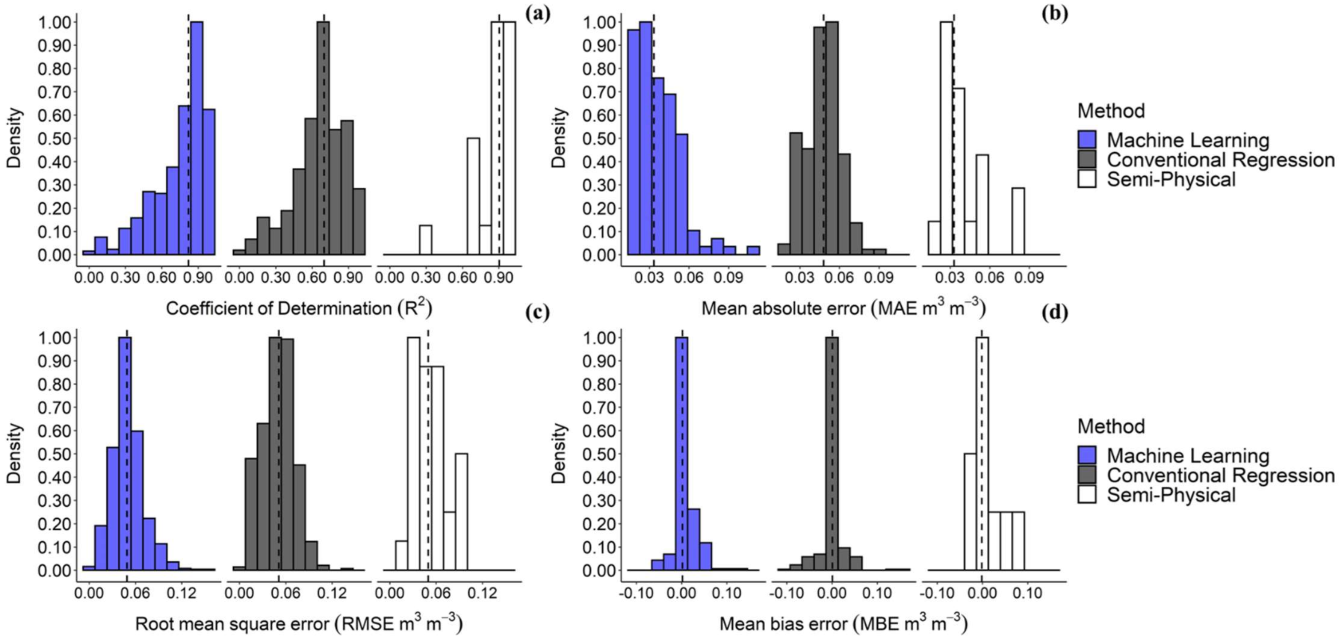

Figure 8. Frequency density histograms of the most used statistical metrics to evaluate the performance of soil water retention PTFs, stratified by different methods of PTF development. Dotted lines represent medians.

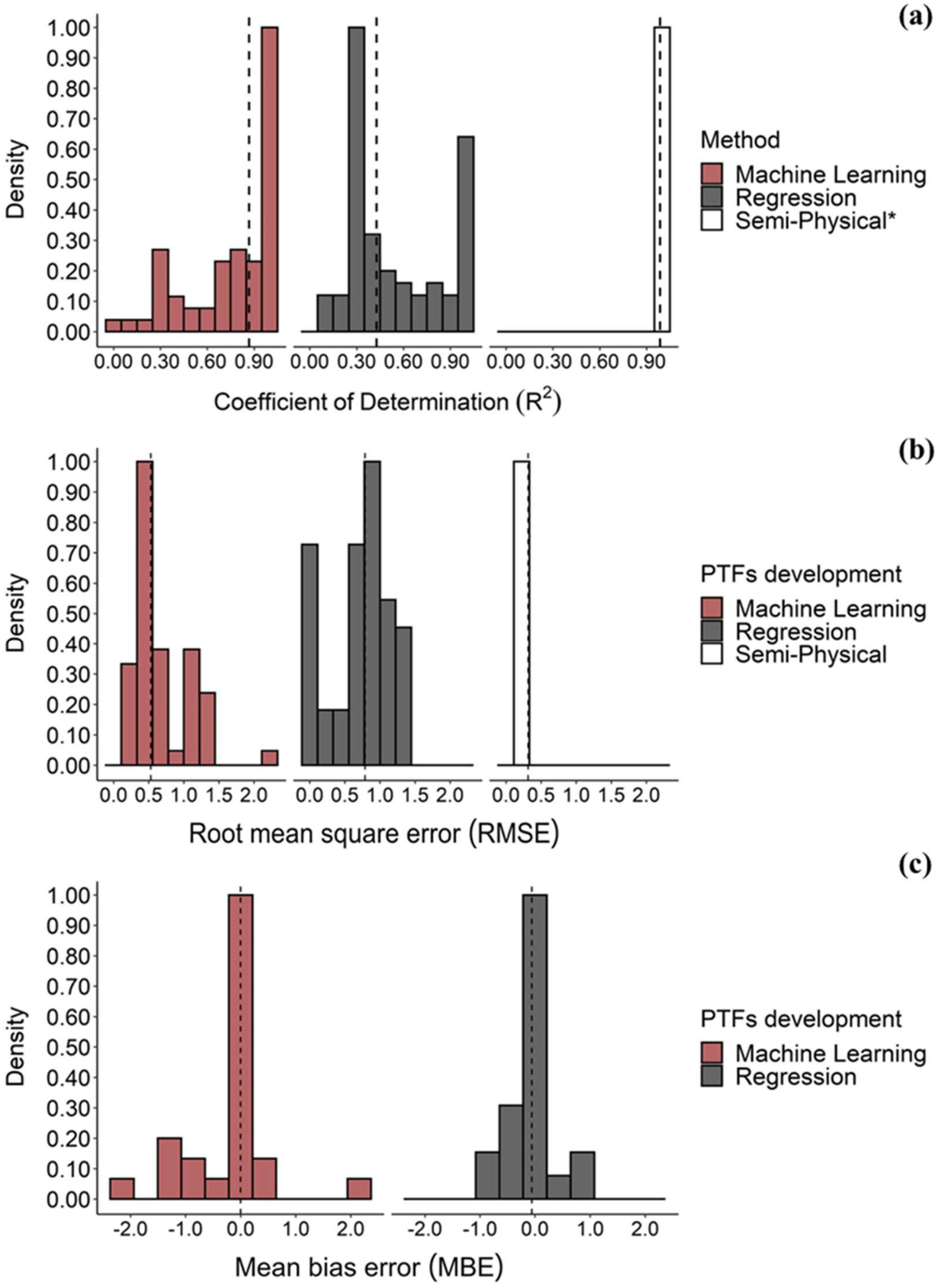

Figure 9. Frequency density histograms of the most used statistical metrics to evaluate the performance of hydraulic conductivity PTFs, stratified by different methods of PTF development. Dotted lines represent medians. *Only one value was recorded in (a) and three values were recorded in (b) for the semi-physical approach.

2.3.6 Uncertainty and model sensitivity assessment

Firstly, we observed that there is a lot of confusion between the terms “error” and “uncertainty”, as several works mention “uncertainty” when referring to error metrics, e.g., RMSE. Therefore, verifying the need for a careful use of these terms, we refer the reader to the conceptualization described in Minasny and Heuvelink (2008), who state that: error is “The difference between the quantity obtained by a model and the true value”, represented by metrics such as RMSE and MAE. Conversely, uncertainty is “Lack of assurance about the truth of a statement or about the exact magnitude of an unknown parameter”, represented by metrics such as standard deviation, standard error, confidence interval and prediction interval. The careful use of these terms may avoid misinterpretations.

The results in Table 2 show that both uncertainty quantification and sensitivity analysis are scarce in the PTF literature for SHP prediction. Out of the publications assessed, 30% (n = 30) of the studies quantified uncertainties and 16% (n = 16) performed sensitivity analysis of the models. Only 8% (n = 8) of the articles performed both analyses, whereas 62% (n = 63) performed neither. The results agree with the statement by Liao et al. (2014) that few studies perform uncertainty quantification and/or sensitivity analysis for PTF models.

The importance of quantifying PTFs uncertainties is highlighted by several authors (Donatelli et al., 2004; Heuvelink, 2018; Minasny et al., 1999; Minasny and McBratney, 2002; Schaap, 2004). Sensitivity analyses are considered as an extension of uncertainty quantification, identifying the importance and quantifying the contribution of each input variable to the inaccuracies of the outputs (i.e., results) of the models (Minasny and Heuvelink, 2008; Touil et al., 2016). Sensitivity analysis has a particular importance in predictions, especially in simulations that involve considerable amounts of variables and input parameters (Minasny and Heuvelink, 2008).

Uncertainties arising from input data rely on the variability of these variables (Pachepsky et al., 2015) and indicate possible impacts on model’s final uncertainty. As examples of studies evaluating uncertainties and sensitivity of SHP prediction models, Liao et al. (2014) observed a high coefficient of variation in predictors such as soil texture data and soil organic matter content. Touil et al. (2016) reported high uncertainty on PTFs outputs due to heterogeneities in their sand data inputs. These results, depending on the models' sensitivity to each variable, are indicative of the contribution of each variable to the final uncertainty.

The same authors also reported high uncertainty associated with saturated water content and with α parameter of the van Genuchten (1980). Other authors have also reported the high uncertainty associated with the α parameter of the van Genuchten model (Kotlar et al., 2019b; Ottoni et al., 2018; Wyatt et al., 2021).

Table 2. Number of studies that performed uncertainty quantification and/or sensitivity analysis for PTFs considering all publications (n = 101).

	Uncertainty	Sensitivity	Both
Yes	30	16	8
No	71	85	63

2.3.7 Case study: performance of existing PTFs to predict soil hydraulic properties of a soil database from Mato Grosso state (Brazil)

2.3.7.1 General comments

The Mato Grosso state (Brazil) has a notorious importance in the Brazilian agricultural production and economy, leading the national production of soybean, maize, cotton and beef, as well as having a key ecological function, as the state integrates three important biomes: Amazon, Cerrado and Pantanal. Mato Grosso has a great variation in SHP as a result of the great pedological and land use diversity in its vast territorial extension (~90 million hectares). Within this context of agricultural and environmental importance, we evaluated existing PTFs for the prediction of water retention (section 3.7.2) and hydraulic conductivity (section 3.7.3) for soils from Mato Grosso. In addition to the existing PTFs, we also made predictions using the null model (i.e., mean value), to assess whether the use of PTFs generates any gain in accuracy compared to the simplest model of all.

2.3.7.2 Predicting soil water retention

The results of predictions using the entire database for field capacity (θ_{FC}) and for the permanent wilting point (θ_{PWP}) are shown in Figure 10 and Figure 11, respectively. The prediction of θ_{FC} was substantially better when using PTFs developed with tropical soils database, i.e., those by Medrado and Lima (2014) and Tomasella et al. (2000), compared to PTFs developed using databases mainly composed of soils under temperate environments, i.e., those by Tóth et al. (2015) and the Rosetta3 model, by Zhang and Schaap (2017). The PTFs by Medrado and Lima (2014) and Tomasella et al. (2000) showed RMSE of 0.0637 m^3 m^{-3} and 0.0674 m^3 m^{-3}, respectively, lower than the RMSE values of 0.0937 m^3 m^{-3} and 0.1411 m^3 m^{-3} using the PTFs by Tóth et al. (2015) and Rosetta3 model, respectively. Similarly, MAE values were also lower for the tropical PTFs, being 0.0512 m^3 m^{-3} for Medrado and Lima (2014) PTF and 0.0564 m^3 m^{-3} for the PTF of Tomasella et al. (2000), compared with MAE of 0.0767 m^3 m^{-3} using the PTF of Tóth et al. (2015) and 0.1234 m^3 m^{-3} using the Rosetta3 model. The MBE calculated for predictions using the tropical PTFs were relatively low; conversely, the temperate PTFs showed considerable MBE values, both overestimating θ_{FC}, highlighting the high MBE calculated for the Rosetta3 model (-0.1157 m^3 m^{-3}). The high bias showed by Rosetta3 model led to the lower LCCC among all PTFs (0.2586), contrasting with the PTF by Medrado and Lima (2014) which had the lower MBE (0.0369) and showed the best model concordance (LCCC = 0.5477).

The overestimation of θ_{FC} of tropical soils by the temperate PTFs is possibly related to the clay fraction of these soils, as tropical soils have predominantly low activity clay minerals, resulting from an intense weathering process. Conversely, soils under the influence of temperate climates and with a lower weathering degree typically have higher concentration of 2:1 minerals, which leads to a greater specific surface area (van den Berg et al., 1997), contributing to a greater water retention capacity (Reichert et al., 2009). These results are also supported by Gaiser et al. (2000), who observed significant differences in water retention patterns comparing soils with contrasting mineralogy, concluding that soils mainly composed of low activity clay, as observed in most of tropical soils, retain less water compared to soils with high activity clay, which predominate in temperate climate environments.

It is well known that the mineralogy resulting from an intense weathering process plays a crucial role in the structural development of tropical soils, presenting remarkable differences from temperate soils (Ferreira et al., 1999; Martinez and Souza, 2020; Pessoa and Libardi, 2022; Silva et al., 2022). As soil structure directly influences soil hydrology, especially in the wet range of the soil water retention curve, as near to θ_{FC} (Botula et al., 2013; Pachepsky et al., 2006), it is plausible to assume that PTFs built for temperate soils may not be able to satisfactorily capture water retention patterns of tropical soils. Therefore, the results shown in Fig. 10 reinforce previous results in the literature that discuss the mineralogical and structural particularities of tropical soils and their influence on different water retention patterns compared to temperate soils (Hodnett and Tomasella, 2002; Ottoni et al., 2018; Tomasella and Hodnett, 2004). Furthermore, none of the tested PTFs have mineralogical features in their structure, which may hinder their reliability, and should deserve attention in future studies.

Looking beyond the differences between tropical and temperate PTFs accuracy, a result that deserves attention in Fig. 10 is that none of the existing PTFs had lower values of RMSE and MAE than the null model. This result supports the current need for further development of PTFs aiming at the recognition of tropical soil patterns. The PTF of Tomasella et al. (2000), considered a reference for predicting water retention in Brazilian soils, performed worse than the null model. The same occurred with the PTF by Medrado and Lima (2014), recently developed for soils in the Cerrado biome, which covers a large part of the state of Mato Grosso. Even with the reasonable geographic and environmental proximity, Brazilian PTFs did not show greater reliability than the null model when predicting θ_{FC} of soils from Mato Grosso, indicating that it would be better and parsimonious to predict it usinghe average value. Nevertheless, it is important to note that the null model was built using the same data it predicted, giving it a slight advantage as the other models did not know the data before they were applied. Given that, the null model performance must be seen as a training performance rather than an independent testing and it is likely that the null model applied in new datasets would perform worse.

The θ_{PWP} predictions were, in general, better than θ_{FC} predictions. Three of the four evaluated PTFs performed better in predicting θ_{PWP} than θ_{FC} and only the PTF by Tomasella et al. (2000) performed worse. The best performance was obtained by the model of Medrado and Lima (2014), with RMSE and MAE values of 0.0378 m^3 m^{-3} and 0.0291 m^3 m^{-3}, respectively. This PTF also presented the best model agreement (LCCC = 0.7646). The worst results were observed when using the PTF of Tomasella et al. (2000), which showed RMSE and MAE values of 0.1120 m^3 m^{-3} and 0.0990 m^3 m^{-3}, respectively. The PTF of Tomasella et al. (2000) also stood out for resulting a high MBE magnitude (-0.0966), showing a high bias towards overestimating θ_{PWP}. This high MBE for Tomasella et al. (2000) PTF resulted in the poorest model concordance among all PTFs (LCCC = 0.0737). The performance of the PTFs built with mainly temperate soils database (i.e., PTFs by Tóth et al., 2015 and Rosetta3 model) showed an intermediate performance between the two Brazilian PTFs.

A result that merits special notation was the better performance of PTFs for the prediction of θ_{PWP} compared to the null model, which showed RMSE and MAE of 0.0511 m^3 m^{-3} and 0.0394 m^3 m^{-3}, respectively. Only the PTF of Tomasella et al. (2000) did not perform better than the null model in predicting θ_{PWP}. It is a remarkable result especially considering that the null model had the advantage of knowing the data before predicting, playing a role as a training prediction whereas the other PTFs were an actual independent testing. This result means that applying these PTFs to predict θ_{PWP} for soils of Mato Grosso would provide gains in accuracy and worth the effort of using more complex models. Similar to our results, Wassar et al. (2016) also found lower errors when predicting the dry range of the soil water retention curve, which indicates that a single PTF may be able to model well some parts of the soil water retention curve whereas performing poorly in others. Kotlar et al. (2020) also exemplified the need to use different predictors when modeling each part of the curve to achieve satisfactory accuracies.

It is well known that modeling different ranges of the soil water retention curve may result in different accuracies, even when using the same function, as different ranges of matric potential are governed by different forces. The “dry” range of the soil water retention curve (i.e., near to permanent wilting point) is mainly controlled by adsorption forces, whereas the “wet” range (i.e., near to field capacity) is mainly governed by capillary forces (Jensen et al., 2015). Thus, the dry range of the curve is mainly influenced by particles size, known as “textural porosity”, whereas the wet range is strongly influenced by aggregates (Nguyen et al., 2014) and mainly controlled by pore size distribution, known as “structural porosity” (Carducci et al., 2013; Dexter, 2004). As a consequence, several studies highlight the best performances when PTFs are built for specific segments of the soil water retention curve (Vereecken et al., 2010), since they are able to capture different relationships between the most influential variables in each range of the curve (Cueff et al., 2021; Dashtaki et al., 2010; Pachepsky et al., 1996; Tomasella et al., 2003).

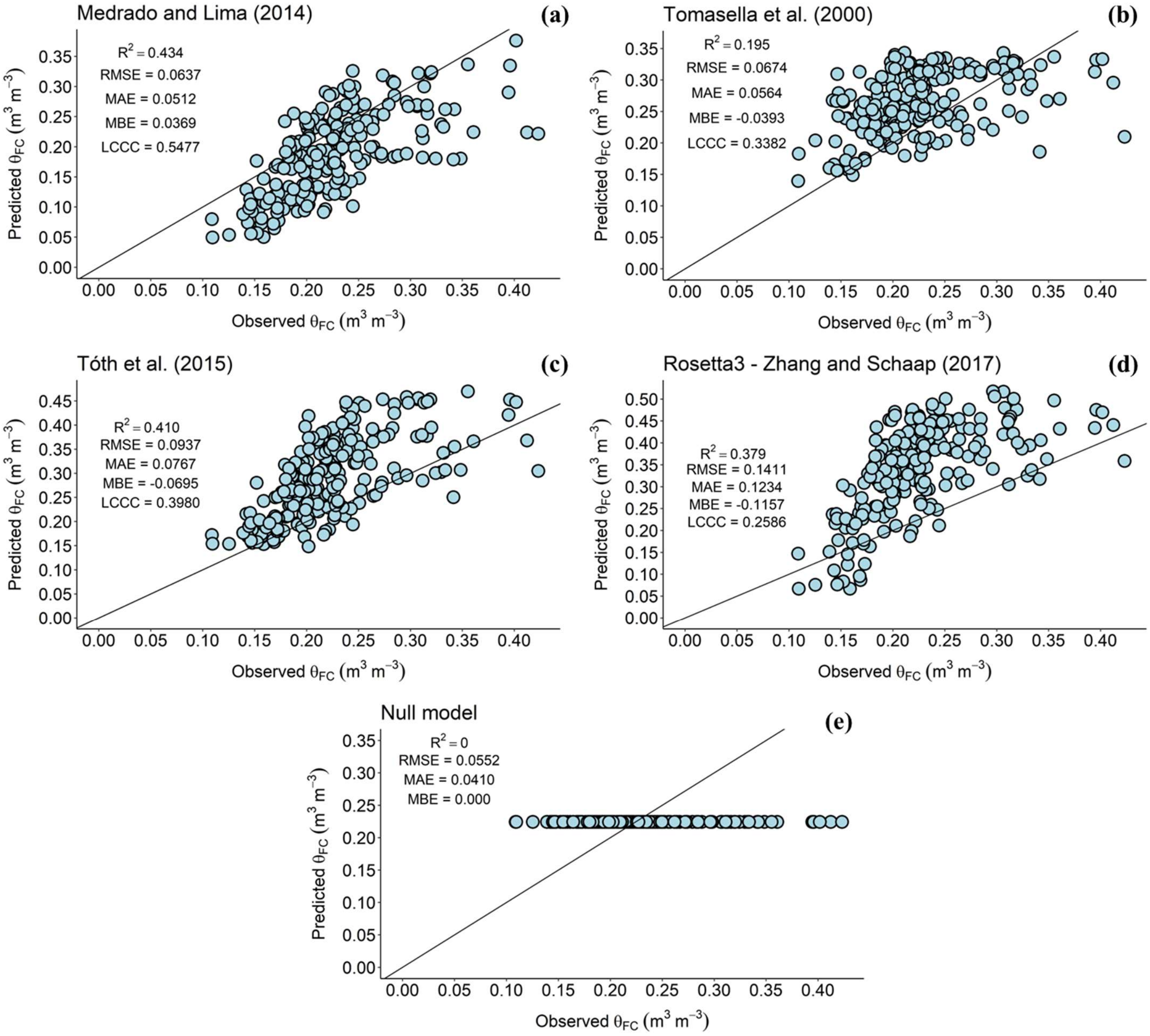

Figure 10. Performance of PTFs by (a) Medrado and Lima (2014), (b) Tomasella et al. (2000), (c) Tóth et al. (2015), (d) Zhang and Schaap (2017) and (e) Null model, for predicting field capacity (θ_{FC}) of soils from Mato Grosso state (Brazil). R^2 = coefficient of determination, RMSE = root mean square error, MAE = mean absolute error, MBE = mean bias error and LCCC = Lin's Concordance Correlation Coefficient.

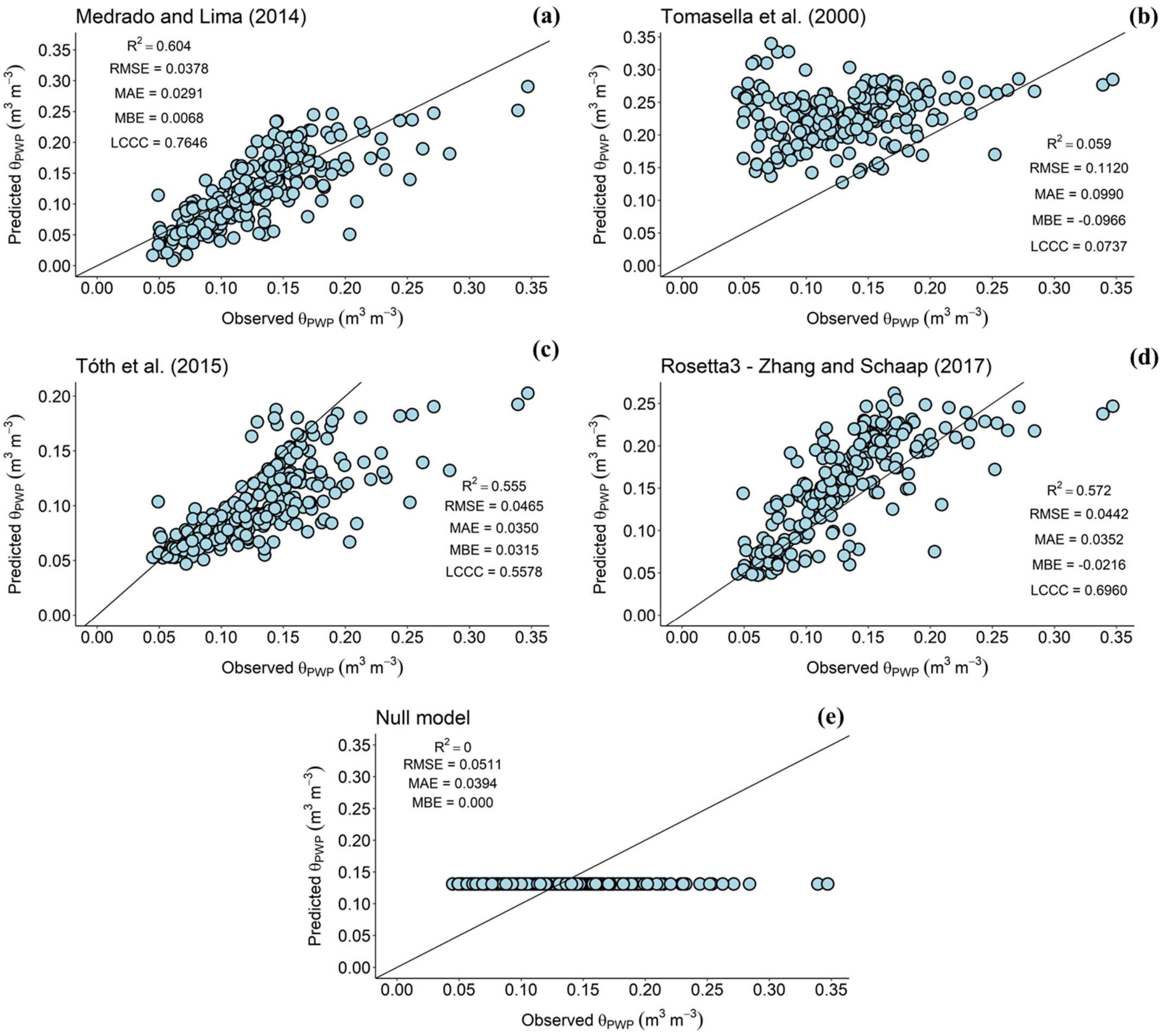

Figure 11. Performance of PTFs by (a) Medrado and Lima (2014), (b) Tomasella et al. (2000), (c) Tóth et al. (2015), (d) Zhang and Schaap (2017) and (e) Null model, for predicting permanent wilting point (θ_{PWP}) of soils from Mato Grosso state (Brazil). R^2 = coefficient of determination, RMSE = root mean square error, MAE = mean absolute error, MBE = mean bias error and LCCC = Lin's Concordance Correlation Coefficient.

The results of water retention predictions stratified by soil textural groups (coarse, medium and fine texture) are shown in Table 3. It is noteworthy that coarse-textured soils are less represented in our database (n = 32) than medium-textured (n = 101) and fine-textured soils (n = 95). The general results of evaluated PTFs when stratifying the database by soil textural groups did not differ from those observed in predictions using the entire database, as the PTFs by Medrado and Lima (2014), Tóth et al. (2015) and Rosetta3 model (Zhang and Schaap, 2017) consistently showed better predictions for θ_{PWP} than for θ_{FC}. Similar to what was observed in predictions using the entire database, the PTF of Tomasella et al. (2000) was the only one that performed better in predicting θ_{FC} than θ_{PWP}, regardless of soil textural groups. Nonetheless, stratifying the database into soil textural groups provided important insights regarding the performance of the existing PTFs in predicting SHP of tropical soils.

The two Brazilian PTFs performed worse than the PTFs by Tóth et al. (2015) and Rosetta3 model in predicting θ_{FC} for coarse-textured soils. It is well-known that, in general, a great portion of Brazilian soils tend to have a finer texture, resulting from the intense weathering process (Resende et al., 2014; Silva et al., 2020). This predominance of fine-textured soils may lead to a poor predictive capacity for coarse-textured soils by Brazilian PTFs, due to insufficient training. Concerning the proportionality between high data availability for training and the PTF's prediction ability, it is observed that only the PTF of Tóth et al. (2015) performed better – although slightly – than the null model in predicting θ_{FC} of coarse-textured soils. This result is possibly explained by the massive training database used in Tóth et al. (2015) (n = 4749), compared to 2134 samples used in Rosetta3 model training, 1092 samples used in Medrado and Lima (2014) and 517 samples used in Tomasella et al. (2000). In terms of accuracy gain, only the PTF of Tóth et al. (2015) could be applied to predict θ_{FC} in coarse-textured soils of Mato Grosso, whereas θ_{PWP} would be more satisfactorily and parsimoniously predicted by the null model.

Considering the medium-textured soils, the performance of existing PTFs were very similar, highlighting only the poor performance of Rosetta3 model when predicting θ_{FC}. For prediction of θ_{FC} in medium-textured soils, none of the PTFs showed accuracy gains compared to the null model. For the prediction of θ_{PWP}, only the PTF by Medrado and Lima (2014) performed better, decreasing RMSE and MAE compared to the null model. However, the decrease in error metrics were so slight that it is questionable whether its use would be parsimonious compared to the easy use of the null model. In general, the results allow us to infer that the most outstanding differences among the existing PTFs were observed for fine-textured soils.

As already discussed, the great dissimilarity between temperate and tropical soils is mainly remarkable in clay mineralogy and its effects, such as in soil structure formation. This dissimilarity was reflected in the divergent performance of Brazilian PTFs and temperate PTFs when predicting θ_{FC} for fine-textured soils of Mato Grosso. The PTFs based on temperate soils databases (Tóth et al., 2015 and Rosetta3 model) showed high RMSE (0.1263 $m^3\ m^{-3}$ and 0.1837 $m^3\ m^{-3}$, respectively) and MAE values (0.1179 $m^3\ m^{-3}$ and 0.1751 $m^3\ m^{-3}$, respectively). Furthermore, both temperate PTFs showed high magnitudes of MBE (-0.1159 $m^3\ m^{-3}$ and -0.1751 $m^3\ m^{-3}$), thus having a high bias towards overestimating θ_{FC} of fine-textured soils from Mato Grosso and resulting in the worst model concordance (LCCC = 0.1352 and 0.0326, respectively). The PTF by Medrado and Lima (2014) showed the best prediction of θ_{FC} for fine-textured soils, being the only PTF that performed better than the null model.

The performance of PTFs in predicting θ_{PWP} of fine-textured soils did not show strong differences as observed for θ_{FC}. Again, the PTF of Medrado and Lima (2014) had the best results and was the only one PTF that generated an accuracy gain – although slight – compared to the null model. The PTF by Tomasella et al. (2000) showed the highest errors for θ_{PWP} prediction, as was observed for all soil textural groups, which allows to conclude that the PTF by Tomasella et al. (2000) does not well capture θ_{PWP} patterns of soils from Mato Grosso, regardless of soil texture. The temperate PTFs showed an intermediate performance for θ_{PWP} prediction, demonstrating that the low reliability of these PTFs in predicting water retention of Brazilian soils is mainly associated with the wet range of the soil water retention curve (e.g., θ_{FC}). It is noteworthy that the PTF of Medrado and Lima (2014) is the only PTF that included in its training data samples collected within the state of Mato Grosso, which possibly led this PTF to recognize regional patterns of water retention in the state of Mato Grosso, contributing to its better accuracy. These results reinforce what was highlighted by Patil and Singh (2016), who stated that there is a demand for the development of geographically specific PTFs, at least until generic PTFs are developed and validated in large and representative databases.

Table 3. Performance of PTFs by Medrado and Lima (2014), Tomasella et al. (2000), Tóth et al. (2015), Zhang and Schaap (2017), and Null model to predict field capacity (θ_{FC}) and permanent wilting point (θ_{PWP}) of soils from Mato Grosso state (Brazil), stratified by three textural groups (coarse, medium and fine texture). R^2 = coefficient of determination, RMSE = root mean square error, MAE = mean absolute error, MBE = mean bias error and LCCC = Lin's Concordance Correlation Coefficient.

	Coarse-textured soils (n = 32)									
	θFC					θPWP				
Evaluated PTFs	R^2	RMSE ($m^3 m^{-3}$)	MAE ($m^3 m^{-3}$)	MBE ($m^3 m^{-3}$)	LCCC	R^2	RMSE ($m^3 m^{-3}$)	MAE ($m^3 m^{-3}$)	MBE ($m^3 m^{-3}$)	LCCC
Medrado and Lima (2014)	0.354	0.0860	0.0823	0.0823	0.0948	0.124	0.0470	0.0373	0.0363	0.1435
Tomasella et al. (2000)	0.279	0.0845	0.0699	-0.0684	0.2161	0.004	0.1711	0.1581	-0.1581	0.0079
Tóth et al. (2015)	0.130	0.0300	0.0244	-0.0064	0.2689	0.269	0.0350	0.0213	0.0197	0.1273
Rosetta3 - Zhang and Schaap (2017)	0.364	0.0478	0.0399	0.0086	0.4923	0.252	0.0325	0.0199	0.0165	0.2077
Null model	0.000	0.0312	0.0245	0.0000		0.000	0.0313	0.0214	0.0000	
	Medium-textured soils (n = 101)									
	θFC					θPWP				
Evaluated PTFs	R^2	RMSE ($m^3 m^{-3}$)	MAE ($m^3 m^{-3}$)	MBE ($m^3 m^{-3}$)	LCCC	R^2	RMSE ($m^3 m^{-3}$)	MAE ($m^3 m^{-3}$)	MBE ($m^3 m^{-3}$)	LCCC
Medrado and Lima (2014)	0.133	0.0673	0.0547	0.0480	0.2162	0.287	0.0333	0.0243	0.0125	0.4949
Tomasella et al. (2000)	0.003	0.0589	0.0481	-0.0237	0.0447	0.014	0.1088	0.0993	-0.0974	-0.0212
Tóth et al. (2015)	0.128	0.0675	0.0546	-0.0458	0.2313	0.163	0.0416	0.0323	0.0276	0.2281
Rosetta3 - Zhang and Schaap (2017)	0.178	0.1107	0.1013	-0.0993	0.1256	0.339	0.0337	0.0273	-0.0165	0.5126
Null model	0.000	0.0452	0.0325	0.0000		0.000	0.0336	0.0253	0.0000	
	Fine-textured soils (n = 95)									
	θFC					θPWP				
Evaluated PTFs	R^2	RMSE ($m^3 m^{-3}$)	MAE ($m^3 m^{-3}$)	MBE ($m^3 m^{-3}$)	LCCC	R^2	RMSE ($m^3 m^{-3}$)	MAE ($m^3 m^{-3}$)	MBE ($m^3 m^{-3}$)	LCCC
Medrado and Lima (2014)	0.268	0.0491	0.0370	0.0099	0.5011	0.328	0.0387	0.0315	-0.0093	0.5433
Tomasella et al. (2000)	0.087	0.0694	0.0608	-0.0462	0.1587	0.108	0.0875	0.0787	-0.0749	0.1079
Tóth et al. (2015)	0.248	0.1263	0.1179	-0.1159	0.1352	0.326	0.0541	0.0425	0.0397	0.3510
Rosetta3 - Zhang and Schaap (2017)	0.080	0.1837	0.1751	-0.1751	0.0326	0.220	0.0560	0.0490	-0.0400	0.2150
Null model	0.000	0.0523	0.0415	0.0000		0.000	0.0443	0.0309	0.0000	

2.3.7.3 Predicting soil hydraulic conductivity

Predictions of soil hydraulic conductivity (Ks) using the entire database of soils from Mato Grosso were expressed in terms of logarithm with base 10 ($\log_{10}$Ks) and are shown in

Figure 12. The WMF330 model by Ottoni et al. (2019) had the best performance, although modest, showing the lowest values of error metrics for $\log_{10}$Ks prediction (RMSE = 0.5872 and MAE = 0.4527), while the WM330 model, also by Ottoni et al. (2019), had the worst performance (RMSE = 0.6831 and MAE = 0.5137). However, both WMF330 and WM330 models showed the lowest concordance between observed and predicted data (LCCC = 0.2645 and 0.2289, respectively) among all models, indicating that they were not able to properly capture data variance. Conversely, both H3w and H5w models from Rosetta3 showed the highest concordances between predicted and observed data (LCCC = 0.4319 and 0.3464, respectively), indicating that although they had high errors due to dispersion, they were somewhat able to represent data variation. The WMF330 and H3w models (the latter from Rosetta3 model) were the only that systematically showed lower RMSE and MAE values compared to those found for the null model (RMSE = 0.6523 and MAE = 0.5108). The WMF330 model, by Ottoni et al. (2019), include Brazilian soil samples in its training database, which may explain the lower error metrics. Nevertheless, the worst performance of WM330 model demonstrates that the key difference between WMF330 and WM330 performances in predicting $\log_{10}$Ks was in the model's structure, since they were built based on the same training database.

Both WMF330 and WM330 models use effective porosity (i.e., total porosity minus soil water content at matric potential of -33kPa) as the unique predictor of Ks. However, WMF330 model uses soil textural properties (sand, silt, and clay) and bulk density to estimate effective porosity, whereas WM330 model uses the measured effective porosity. When analyzing the models of Rosetta3, something similar occured, as H3w uses only basic properties as Ks predictors (soil textural properties and bulk density), while H5w uses, in addition to basic properties, the soil water content at matric potentials of -33 kPa and -1500 kPa. Thus, the better performance of WMF330 and H3w models compared to the null model is a relatively surprising result, as they use only basic properties as Ks predictors and, therefore, are simpler than WM330 and H5w models.

WM330 and H5w models could be considered a priori as more robust models since they use more complex properties in their structure (i.e., effective porosity and soil water contents). However, the worst prediction performance of those models, without providing accuracy gains compared to the null model, is an intriguing result, since it is expected that the effort spent to collect and measure more complex properties is rewarded with gains in accuracy, which was not achieved. Therefore, using more complex properties might have "confused" the PTFs for $\log_{10}$Ks prediction, performing worse than the more basic models (WMF330 e H3w). This kind of result is referred to as the Hughes phenomenon (Hughes, 1968).

In short, the Hughes phenomenon states that a model may have the number of predictors increased until reaching an optimal number, and, from this number onwards, inserting more predictors degrades the model performance. This performance reduction is mainly due to the insertion of more noise in model structure, leading the model to have more opportunity to overfit to more noise, resulting in a poorer generalization. It is noteworthy that we are not assuming that WM330 and H5w models have concerns regarding the number of predictors, but the issue of adding noise to the model seems plausible, considering that the models may not be sufficiently trained to recognize all patterns, resulting in a lower predictive capacity.

This hypothesis of adding noise when effective porosity and soil water contents were added to PTFs structure may make sense since SHP present high variance (Minasny and McBratney, 2002a; Schaap, 2001). As a very sensitive property, values of Ks are strongly influenced by many soil properties (Araya and Ghezzehei, 2019; Zhang and Schaap, 2019) and can present large variations when small variations in other properties occur (Morgan et al., 1993). Ks values are very sensitive to variations in porous media, such as pores continuity, pores tortuosity and pore size distribution (Jačka et al., 2018), as well as being sensitive to variations of properties not very often used in PTFs, such as chemical properties (Candemir and Gülser, 2012). Thus, given this wide sensitivity of Ks, it is possible that including complex variables in PTFs explain the decreased predictive capacity, due to the already mentioned insertion of noise and patterns the models were not able to recognize, "confusing" the models and decreasing the generalization capacity. Overall, when predicting $\log_{10}$Ks for soils from Mato Grosso, PTFs based on basic soil properties (WMF330 and H3w models) provided greater performance – although modest – along with a higher ease-of-use.

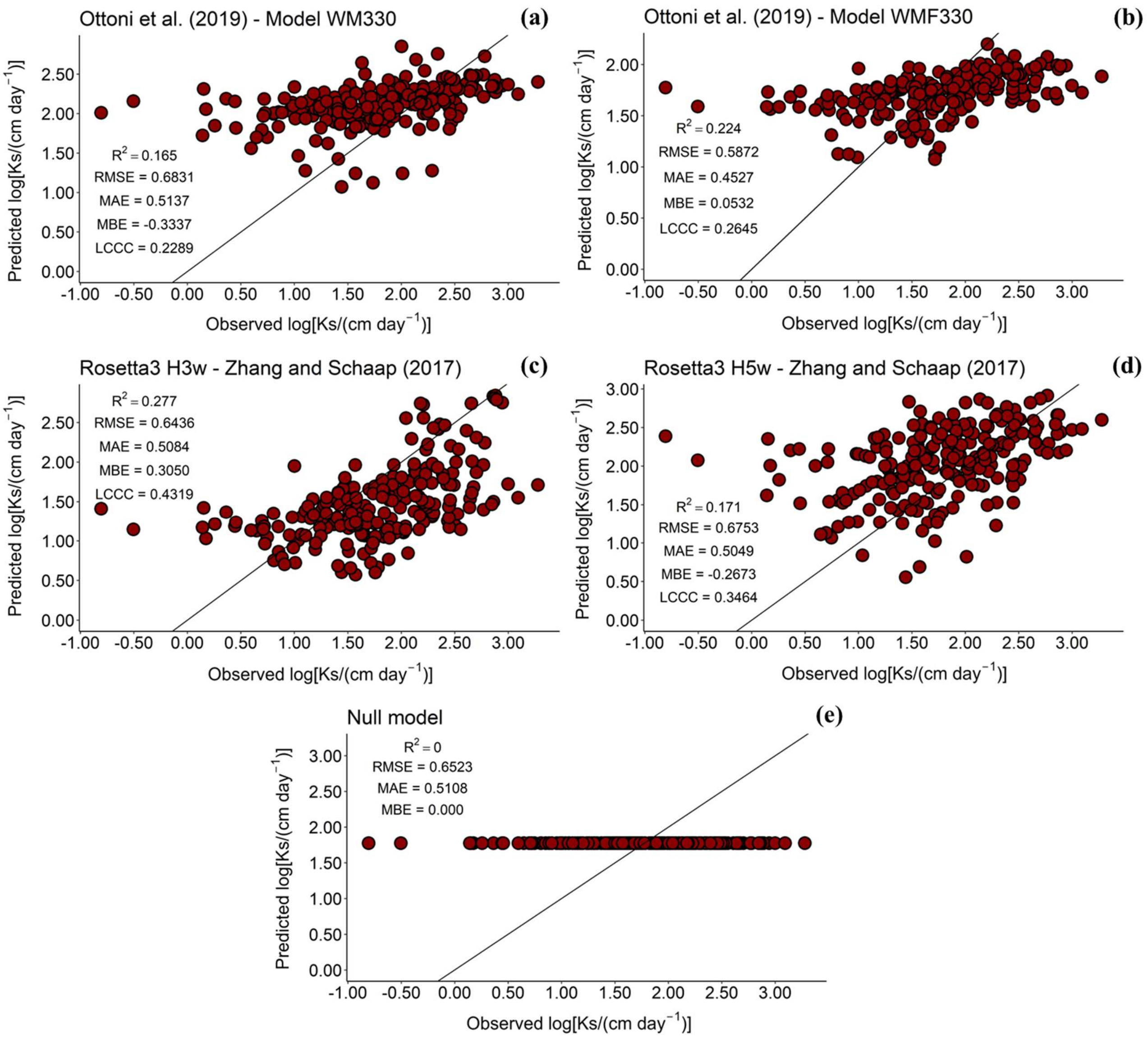

Figure 12. Performance of (a) WM330 and (b) WMF330 models (Ottoni et al., 2019), (c) H3w and (d) H5w models (Rosetta3 - Zhang and Schaap, 2017) and (e) and Null model for predicting the log of hydraulic conductivity of soils from Mato Grosso state (Brazil). R^2 = coefficient of determination, RMSE = root mean square error, MAE = mean absolute error, MBE = mean bias error and LCCC = Lin's Concordance Correlation Coefficient.

As was done for soil water retention, we stratified the database into soil textural groups to evaluate the performance of existing PTFs to predict $\log_{10}$Ks (Table 4). PTFs followed a general pattern, systematically performing better for coarse-textured soils, intermediate for medium-textured soils and the worst performance was observed for fine-textured soils. The only model that was an exception to this pattern was the WMF330 model, which performed better on medium-textured soils compared to coarse-textured soils; however, it followed the general pattern of worst predictions for fine-textured soils. As already mentioned in the former section, coarse-textured soils are less represented in our database (n = 32) than medium-textured (n = 101) and fine-textured soils (n = 95).

Evaluating coarse-textured soils, the best performance was found for WM330 model by Ottoni et al. (2019), with lower error metrics (RMSE = 0.2699 and MAE = 0.2018), whereas the worst prediction was performed by WMF330 model (RMSE = 0.5369 and MAE = 0.4484). The WMF330 model also showed a high bias towards underestimating predictions (MBE = 0.3859), which impaired the concordance between observed and predicted data and resulted in lowest LCCC (0.1762). The models from Rosetta3 (Zhang and Schaap, 2017) had intermediate performance, with RMSE of 0.2951 and 0.3150 and MAE of 0.2289 and 0.2478 for H3w and H5w models, respectively. Only the WMF330 model performed worse than the null model, which indicates that all other PTFs provided accuracy gains when predicting $\log_{10}$Ks for coarse-textured soils from Mato Grosso. For medium-textured soils, the overall performance was considerably worse, and the WM330 model, which had the best performance for coarse-textured soils, had the worst performance (RMSE = 0.6224, MAE = 0.5073 and LCCC = 0.1246). Conversely, the model that had the worst performance on coarse-textured soils (WMF330) had the best performance on medium-textured soils (RMSE = 0.4679 and MAE = 0.3897) and the lowest bias (MBE = -0.0442). Again, the models from Rosetta3 performed intermediately between the two models of Ottoni et al. (2019). For medium-textured soils, the only model that provided accuracy gains compared to the null model was WMF330 model.

The highest values of RMSE and MAE were observed for all PTFs when predicting $\log_{10}$Ks for fine-textured soils. The worst performances were found for more complex models, i.e., WM330 and H5w, with RMSE of 0.8267 and 0.8689, respectively, and MAE of 0.6255 and 0.6665, respectively. These models also showed the highest bias magnitude, both overestimating the predictions (MBE = -0.4617 and -0.4722, respectively). The worst performance of PTFs that used more complex predictors add an insight into what was discussed regarding the Hughes phenomenon observed in the predictions using the entire database, indicating that the worst performances were mainly influenced by the poor predictions of $\log_{10}$Ks for fine-textured soils. This implies that the models were possibly not sufficiently trained to recognize the patterns of effective porosity and soil water content in fine-textured soils, reinforcing the discussion about SHP particularities in clayey tropical soils, mainly conditioned by clay mineralogy and its influences in soil structure. Ottoni et al. (2019) also found remarkable differences in Ks prediction for fine-textured soils and discussed about the need to include other predictors aiming at better recognition of Ks patterns in these soils. For example, Tóth et al. (2015) found that pH, CEC and $CaCO_3$ are potential predictors to be explored in Ks prediction.

The particularity of SHP in fine-textured tropical soils is well studied, especially considering the literature concerning the main Brazilian soil class (Ferralsols) and its highly-stable granular structure. These soils present high macroporosity and high hydraulic conductivity even with high clay contents (Ferreira et al., 1999; Moura et al., 2021; Pessoa and Libardi, 2022; Silva et al., 2021; 2022), which may significantly differ from the pattern observed in clayey soils formed under temperate climate conditions. The mineralogy of Brazilian Ferralsols is essentially composed by kaolinite, gibbsite and hematite, which have a great influence on the formation of the granular structure (Silva et al., 2022) and consequently on the hydrological behavior, generally promoting high Ks values (Pessoa and Libardi, 2022). The granular structure is mainly formed by microaggregates of flocculated clay, also known as pseudo-sand (Martinez and Souza, 2020) which lead these soils to behave similarly to sandy soils in terms of water transmission, due to the high proportion of large pores (Tomasella et al., 2000). Finally, our results strongly support what was formerly highlighted by Botula et al. (2013) and Ottoni et al. (2018) regarding the inability of the most existing PTFs in predicting SHP of tropical soils. This generates a demand for further investigation and studies aiming at modeling porous media and well recognizing SHP patterns in clayey tropical soils.

Table 4. Performance of WM330 and WMF330 models (Ottoni et al., 2019), H3w and H5w models (Rosetta3 - Zhang and Schaap, 2017), and Null model for predicting the log of hydraulic conductivity of soils from Mato Grosso state (Brazil), stratified by three textural groups (coarse, medium and fine texture). R^2 = coefficient of determination, RMSE = root mean square error, MAE = mean absolute error, MBE = mean bias error and LCCC = Lin’s Concordance Correlation Coefficient.

	Coarse-textured soils (n = 32)				
	log Ks				
Evaluated PTFs	R^2	RMSE	MAE	MBE	LCCC
Ottoni et al. (2019) - WM330	0.729	0.2699	0.2018	0.1061	0.7338
Ottoni et al. (2019) - WMF330	0.438	0.5369	0.4484	0.3859	0.1762
Rosetta3 H3w - Zhang and Schaap (2017)	0.574	0.2951	0.2289	0.0339	0.7534
Rosetta3 H5w - Zhang and Schaap (2017)	0.499	0.3150	0.2478	0.0581	0.6793
Null model	0.000	0.4347	0.3364	0.0000	

	Medium-textured soils (n = 101)				
	log Ks				
Evaluated PTFs	R^2	RMSE	MAE	MBE	LCCC
Ottoni et al. (2019) - WM330	0.046	0.6224	0.5073	-0.3525	0.1246
Ottoni et al. (2019) - WMF330	0.157	0.4670	0.3897	-0.0442	0.2179
Rosetta3 H3w - Zhang and Schaap (2017)	0.145	0.5984	0.4924	0.3546	0.2537
Rosetta3 H5w - Zhang and Schaap (2017)	0.136	0.5366	0.4344	-0.1776	0.3281
Null model	0.000	0.5038	0.4224	0.0000	

	Fine-textured soils (n = 95)				
	log Ks				
Evaluated PTFs	R^2	RMSE	MAE	MBE	LCCC
Ottoni et al. (2019) - WM330	0.199	0.8267	0.6255	-0.4617	0.1963
Ottoni et al. (2019) - WMF330	0.147	0.7060	0.5210	0.0447	0.2057
Rosetta3 H3w - Zhang and Schaap (2017)	0.193	0.7643	0.6194	0.3437	0.2555
Rosetta3 H5w - Zhang and Schaap (2017)	0.140	0.8689	0.6665	-0.4722	0.2644
Null model	0.000	0.7591	0.5642	0.0000	

2.4 Conclusions

In this review, we brought to light what was done in ten years (2012-2021) regarding the prediction of soil hydraulic properties by PTFs. The results showed a relative outlook change, as a considerable portion of the studies was performed in tropical regions, which was seldom in past decades. Most studies had a regional purpose and studies that developed PTFs with continental or global purposes are incipient. The development of PTFs is still strongly based on basic soil properties as predictors (mainly soil physical properties). However, the use of several alternative variables was observed, such as topographical attributes, climatic attributes, geological attributes, vegetation attributes, geographic coordinates, in addition to several soil morphological and chemical attributes. Promising results using alternative features in PTFs structure are found in the recent literature and it could be a matter to be further explored. The use of machine learning techniques to build PTFs was observed in 45%

of the studies and, although conventional regression is still prevailing (55%), this result reinforces the recent spread of advanced data science techniques. Nevertheless, results exposed a limitation in machine learning studies as the vast majority tested only one or two algorithms, which can restrict the PTFs reliability. Hence, testing several algorithms should be considered in future studies. Most databases for PTFs training and testing used up to 200 samples and there are few works that built PTFs using more than 1000 samples. Error and goodness-of-fit metrics (R^2, RMSE and MAE) indicated that, overall, machine learning techniques have provided better performances than conventional regression. Only 30% of the studies quantified uncertainties in the final predictions, 16% evaluated the models' sensitivity and only 8% performed both. This lack of assessment of uncertainties and sensitivities omits important information about the reliability of PTFs and deserves attention in future studies.

The results of the case study indicated that predicting soil water retention at permanent wilting point was easier than at field capacity, for which even Brazilian PTFs performed poorly. In the evaluation of field capacity, none of the tested PTFs performed better than the null model (mean value). Conversely, the prediction of permanent wilting point was better and most PTFs showed accuracy gains compared to the null model. The worst performances were found when PTFs based on temperate soils attempted to predict the field capacity of fine-textured soils from Mato Grosso, reinforcing the inability of PTFs built in temperate climates to recognize tropical soil patterns. In general, the best performance was observed for a PTF built with Brazilian soils which included samples from the Mato Grosso state in its training database, highlighting the importance of regional characteristics in improving the application of PTFs.

The prediction of hydraulic conductivity in the case study showed intriguing results indicating that basic models performed better than models considered as more complex and robust. The use of complex predictors by the robust models seems to have “confused” the PTFs and generated worse performances. The best predictions of hydraulic conductivity were observed for coarse-textured soils, especially for a PTF that included Brazilian soils in its training database. The most remarkable result was that all PTFs systematically presented worse predictions of hydraulic conductivity for fine-textured soils, reinforcing the particularities (e.g., clay mineralogy and soil structure) of clayey tropical soils. These results expose a demand for more studies aiming to modeling the porous media and well recognizing the hydraulic patterns of fine-textured tropical soils.

Acknowledgments

We acknowledge the financial contribution of the Minas Gerais State Research Foundation (FAPEMIG), the Brazil National Council of Technological and Scientific Development (CNPq - PROC. Nº 313154/2020-1 e PROC. Nº 408231/2021-1), the Coordination for the Improvement of Higher Education Personal (CAPES – Finance Code 001), and the Department of Agricultural Engineering at the Federal University of Viçosa (DEA – UFV).

References

Achieng, K.O., 2019. Modelling available water capacity of topsoil in a Bayesian paradigm. Environ. Model. Softw. 120, 104500. https://doi.org/10.1016/j.envsoft.2019.104500

Al Majou, H., Bruand, A., Duval, O., Le Bas, C., Vautier, A., 2008. Prediction of soil water retention properties after stratification by combining texture, bulk density and the type of horizon. Soil Use Manag. 24, 383–391. https://doi.org/10.1111/j.1475-2743.2008.00180.x

Al Majou, H., Hassani, B., Bruand, A., 2018. Transferability of continuous- and class-pedotransfer functions to predict water retention properties of semiarid Syrian soils. Soil Use Manag. 34, 354–369. https://doi.org/10.1111/sum.12424

Amanabadi, S., Vazirinia, M., Vereecken, H., Vakilian, K.A., Mohammadi, M.H., 2019. Comparative Study of Statistical, Numerical and Machine Learning-based Pedotransfer Functions of Water Retention Curve with Particle Size Distribution Data. Eurasian Soil Sci. 52, 1555–1571. https://doi.org/10.1134/S106422931930001X

Araya, S.N., Ghezzehei, T.A., 2019. Using Machine Learning for Prediction of Saturated Hydraulic Conductivity and Its Sensitivity to Soil Structural Perturbations. Water Resour. Res. 55, 5715–5737. https://doi.org/10.1029/2018WR024357

Barros, A.H.C., van Lier, Q.D.J., Nunes, A.D.H., Scarpare, F.V., 2013. Pedotransfer functions to estimate water retention parameters of soils in northeastern Brazil. Rev. Bras. Ciência do Solo 37, 379–391.

Benke, K.K., Norng, S., Robinson, N.J., Chia, K., Rees, D.B., Hopley, J., 2020. Development of pedotransfer functions by machine learning for prediction of soil electrical conductivity and organic carbon content. Geoderma 366, 114210.

https://doi.org/10.1016/j.geoderma.2020.114210

Bittelli, M., Pellegrini, S., Olmi, R., Andrenelli, M.C., Simonetti, G., Borrelli, E. and Morari, F., 2022. Experimental evidence of laser diffraction accuracy for particle size analysis. Geoderma, 409, p.115627. https://doi.org/10.1016/j.geoderma.2021.115627

Bocuti, E.D., 2020. Condições físico-hídricas de solos mato-grossenses e sua relação com a produtividade de soja: da pesquisa à transposição didática. Tese - Univ. Fed. Mato Grosso, Fac. Agron. e Med. Veterinária, Programa Pós- Grad. em Agric. Trop. Cuiabá, 2020. 200.

Bocuti, E.D., Amorim, R.S.S., Kavasaki, K.F.L., Prado, M.R. V, Santos, C.L.R., Di Raimo, L.A.L., 2021. Soil structure and its relationship with soybean yield. Rev. Bras. Eng. Agrícola e Ambient. 25, 168–173.

Botula, Y.D., Cornelis, W.M., Baert, G. and Van Ranst, E., 2012. Evaluation of pedotransfer functions for predicting water retention of soils in Lower Congo (DR Congo). Agricultural water management, 111, pp.1-10. https://doi.org/10.1016/j.agwat.2012.04.006

Botula, Y.-D., Nemes, A., Mafuka, P., Van Ranst, E., Cornelis, W.M., 2013. Prediction of Water Retention of Soils from the Humid Tropics by the Nonparametric k -Nearest Neighbor Approach . Vadose Zo. J. 12, vzj2012.0123. https://doi.org/10.2136/vzj2012.0123

Botula, Y.-D., Van Ranst, E., Cornelis, W.M., 2014. Pedotransfer functions to predict water retention for soils of the humid tropics: a review. Rev. Bras. Ciência do Solo 38, 679–698. https://doi.org/10.1590/s0100-06832014000300001

Bouma, J., 1989. Using Soil Survey Data for Quantitative Land Evaluation, in: Advances in Soil Science.

Bruand, A., 2004a. Preliminary grouping of soils. Pachepsky Y., Rawls W.J. Dev. Pedotrans. Funct. soil Hydrol. Dev. soil Sci. 30, 159–174. https://doi.org/10.1016/S0166-2481(04)30010-3

Bruand, A., 2004b. Utilizing mineralogical and chemical information in PTFs. Pachepsky Y., Rawls W.J. Dev. Pedotrans. Funct. soil Hydrol. Dev. soil Sci. 30, 153–158. https://doi.org/10.1016/S0166-2481(04)30009-7

Candemir, F., Gülser, C., 2012. Influencing Factors and Prediction of Hydraulic Conductivity in Fine-Textured Alkaline Soils. Arid L. Res. Manag. 26, 15–31. https://doi.org/10.1080/15324982.2011.631686

Carducci, C.E., Oliveira, G.C. De, Zeviani, W.M., Lima, V.M.P., Serafim, M.E., 2013.

Bimodal pore distribution on soils under conservation management system for coffee crop. Eng. Agrícola 33, 291–302. https://doi.org/http://dx.doi.org/10.1590/S0100-69162013000200008.

Chirico, G.B., Medina, H., Romano, N., 2010. Functional evaluation of PTF prediction uncertainty: An application at hillslope scale. Geoderma 155, 193–202. https://doi.org/10.1016/j.geoderma.2009.06.008

Cueff, S., Coquet, Y., Aubertot, J.N., Bel, L., Pot, V., Alletto, L., 2021. Estimation of soil water retention in conservation agriculture using published and new pedotransfer functions. Soil Tillage Res. 209. https://doi.org/10.1016/j.still.2021.104967

Dashtaki, S.G., Homaee, M., Khodaverdiloo, H., 2010. Derivation and validation of pedotransfer functions for estimating soil water retention curve using a variety of soil data. Soil Use Manag. 26, 68–74. https://doi.org/10.1111/j.1475-2743.2009.00254.x

Dexter, A.R., 2004. Soil physical quality: Part I. Theory, effects of soil texture, density, and organic matter, and effects on root growth. Geoderma 120, 201–214. https://doi.org/10.1016/j.geoderma.2003.09.004

Dick, D.P., 1986. Characterization of Iron Oxides and Phosphorus Adsorption in the Clay Fraction of Latosolic B Horizons. UFRGS, Porto Alegre.Dobarco, M.R., Cousin, I., Le Bas, C., Martin, M.P., 2019. Pedotransfer functions for predicting available water capacity in French soils, their applicability domain and associated uncertainty. Geoderma 336, 81–95. https://doi.org/10.1016/j.geoderma.2018.08.022

Donagemma, G.K., Ruiz, H.A., Fontes, M.P.F., Ker, J.C. and Schaffer, C.E.G.R., 2003. Dispersão de Latossolos em resposta à utilização de pré-tratamentos na análise textural. Revista Brasileira de Ciência do Solo, 27, pp.765-772. https://doi.org/10.1590/S0100-06832003000400021

Donatelli, M., Wösten, J.H.M., Belocchi, G., 2004. Methods to evaluate pedotransfer functions, in: Y. Pachepsky and W.J. Rawls. Development of Pedotransfer Functions in Soil Hydrology. pp. 357–411. https://doi.org/10.1016/S0166-2481(04)30020-6

Fatichi, S., Or, D., Walko, R., Vereecken, H., Young, M.H., Ghezzehei, T.A., Hengl, T., Kollet, S., Agam, N., Avissar, R., 2020. Soil structure is an important omission in Earth System Models. Nat. Commun. 11. https://doi.org/10.1038/s41467-020-14411-z

Ferreira, M.M., Fernandes, B., Curi, N., 1999. Influência da mineralogia da fração argila nas propriedades físicas de latossolos da região sudeste do Brasil. Rev. Bras. Ciência do Solo 23, 515–524. https://doi.org/10.1590/s0100-06831999000300004

Gaiser, T., Graef, F., Cordeiro, J.C., 2000. Water retention characteristics of soils with contrasting clay mineral composition in semi-arid tropical regions. Aust. J. Soil Res. 38, 523–536. https://doi.org/10.1071/SR99001

Galvão, T.C.B., Schulze, D.G., 1996. Mineralogical properties of a collapsible lateritic soil from Minas Gerais, Brazil. Soil Sci. Soc. Am. J. 60 (6), 1969–1978. https://doi.org/10.2136/sssaj1996.03615995006000060050x.Gebauer, A., Ellinger, M., M. Brito Gomez, V., Ließ, M., 2020. Development of pedotransfer functions for water retention in tropical mountain soil landscapes: Spotlight on parameter tuning in machine learning. Soil 6, 215–229. https://doi.org/10.5194/soil-6-215-2020

Ghanbarian, B., Taslimitehrani, V., Pachepsky, Y.A., 2017. Accuracy of sample dimension-dependent pedotransfer functions in estimation of soil saturated hydraulic conductivity. Catena 149, 374–380. https://doi.org/10.1016/j.catena.2016.10.015

Goedert, W.J., 1983. Management of the Cerrado soils of Brazil : a review. J. Soil Sci. 34, 405–428.

Guio Blanco, C.M., Brito Gomez, V.M., Crespo, P., Ließ, M., 2018. Spatial prediction of soil water retention in a Páramo landscape: Methodological insight into machine learning using random forest. Geoderma 316, 100–114. https://doi.org/10.1016/j.geoderma.2017.12.002

Gunarathna, M.H.J.P., Sakai, K., Nakandakari, T., Momii, K., Kumari, M.K.N., 2019a. Machine learning approaches to develop pedotransfer functions for tropical Sri Lankan soils. Water (Switzerland) 11. https://doi.org/10.3390/w11091940

Gunarathna, M.H.J.P., Sakai, K., Nakandakari, T., Momii, K., Kumari, M.K.N., Amarasekara, M.G.T.S., 2019b. Pedotransfer functions to estimate hydraulic properties of tropical Sri Lankan soils. Soil Tillage Res. 190, 109–119. https://doi.org/10.1016/j.still.2019.02.009

Gupta, S., Hengl, T., Lehmann, P., Bonetti, S., Papritz, A., 2020. Global prediction of soil saturated hydraulic conductivity using random forest in a Covariate-based Geo Transfer Functions (CoGTF) framework 1–22.

Gupta, S., Lehmann, P., Bonetti, S., Papritz, A., Or, D., 2021. Global Prediction of Soil Saturated Hydraulic Conductivity Using Random Forest in a Covariate-Based GeoTransfer Function (CoGTF) Framework. J. Adv. Model. Earth Syst. 13, 1–15. https://doi.org/10.1029/2020MS002242

Haghverdi, A., Cornelis, W.M., Ghahraman, B., 2012. A pseudo-continuous neural network approach for developing water retention pedotransfer functions with limited data. J. Hydrol. 442–443, 46–54. https://doi.org/10.1016/j.jhydrol.2012.03.036

Heuvelink, G.B.M., 2018. Chapter 14 Uncertainty and Uncertainty Propagation in Soil

Mapping and Modelling, in: McBratney, A. B., Minasny, B., and Stockmann, U. (Eds.). Pedometrics. Springer. pp. 335–336. https://doi.org/10.1111/j.1365-2389.2011.01370.x

Hodnett, M.G., Tomasella, J., 2002. Marked differences between van Genuchten soil water - retention parameters. Geoderma 108, 155–180.

Hollis, J.M., Lilly, A., Higgins, A., Jones, R.J.A., Keay, C.A., Bellamy, P., 2015. Predicting the water retention characteristics of UK mineral soils. Eur. J. Soil Sci. 66, 239–252. https://doi.org/10.1111/ejss.12186

Hughes, G.F., 1968. On the Mean Accuracy of Statistical Pattern Recognizers. IEEE Trans. Inf. Theory 14, 55–63. https://doi.org/https://doi.org/10.1109/TIT.1969.1054314

Jačka, L., Pavlásek, J., Kalibová, J., Bašta, P., Kovář, M., Kuráž, V., 2018. The layering of a mountain podzol can strongly affect the distribution of infiltrated water in the soil profile. J. Hydrol. Hydromechanics 66, 348–356. https://doi.org/10.2478/johh-2018-0014

Jensen, D.K., Tuller, M., de Jonge, L.W., Arthur, E., Moldrup, P., 2015. A New Two-Stage Approach to predicting the soil water characteristic from saturation to oven-dryness. J. Hydrol. 521, 498–507. https://doi.org/10.1016/j.jhydrol.2014.12.018

Kalumba, M., Bamps, B., Nyambe, I., Dondeyne, S., Van Orshoven, J., 2020. Development and functional evaluation of pedotransfer functions for soil hydraulic properties for the Zambezi River Basin. Eur. J. Soil Sci. 1–16. https://doi.org/10.1111/ejss.13077

Keesstra, S., Mol, G., de Leeuw, J., Okx, J., Molenaar, C., de Cleen, M., Visser, S., 2018. Soil-related sustainable development goals: Four concepts to make land degradation neutrality and restoration work. Land 7. https://doi.org/10.3390/land7040133

Keesstra, S.D., Bouma, J., Wallinga, J., Tittonell, P., Smith, P., Cerdà, A., Montanarella, L., Quinton, J.N., Pachepsky, Y., Van Der Putten, W.H., Bardgett, R.D., Moolenaar, S., Mol, G., Jansen, B., Fresco, L.O., 2016. The significance of soils and soil science towards realization of the United Nations sustainable development goals. Soil 2, 111–128. https://doi.org/10.5194/soil-2-111-2016

Khlosi, M., Alhamdoosh, M., Douaik, A., Gabriels, D., Cornelis, W.M., 2016. Enhanced pedotransfer functions with support vector machines to predict water retention of calcareous soil. Eur. J. Soil Sci. 67, 276–284. https://doi.org/10.1111/ejss.12345

Khlosi, M., Cornelis, W.M., Douaik, A., Hazzouri, A., Habib, H. and Gabriels, D., 2013. Exploration of the interaction between hydraulic and physicochemical properties of Syrian soils. Vadose Zone Journal, 12(4), pp.vzj2012-0209. https://doi.org/10.2136/vzj2012.0209

Kotlar, A.M., de Jong van Lier, Q., de Souza Brito, E., 2020. Pedotransfer functions for water contents at specific pressure heads of silty soils from Amazon rainforest. Geoderma 361, 114098. https://doi.org/10.1016/j.geoderma.2019.114098

Kotlar, A.M., Iversen, B. V., Jong van Lier, Q., 2019a. Evaluation of Parametric and Nonparametric Machine-Learning Techniques for Prediction of Saturated and Near-Saturated Hydraulic Conductivity. Vadose Zo. J. 18, 1–13. https://doi.org/10.2136/vzj2018.07.0141

Kotlar, A.M., Jong van Lier, Q., Barros, A.H.C., Iversen, B. V., Vereecken, H., 2019b. Development and Uncertainty Assessment of Pedotransfer Functions for Predicting Water Contents at Specific Pressure Heads. Vadose Zo. J. 18, 190063. https://doi.org/10.2136/vzj2019.06.0063

Kuhn, M., Johnson, K., 2013. Applied predictive modeling, Applied Predictive Modeling. https://doi.org/10.1007/978-1-4614-6849-3

Lal, R., 1979. Physical properties and moisture retention characteristics of some nigerian soils. Geoderma 21, 209–223. https://doi.org/10.1016/0016-7061(78)90028-9

Lehmann, P., Bickel, S., Wei, Z., Or, D., 2020. Physical Constraints for Improved Soil Hydraulic Parameter Estimation by Pedotransfer Functions. Water Resour. Res. 56, 1–13. https://doi.org/10.1029/2019WR025963

Liao, K., Xu, S., Wu, J., Zhu, Q., 2014. Uncertainty analysis for large-scale prediction of the van Genuchten soil-water retention parameters with pedotransfer functions. Soil Res. 52, 431–442. https://doi.org/10.1071/SR13230

Lilly, A., Lin, H., 2004. Using soil morphological attributes and soil structure in pedotransfer functions, in: Pachepsky Y., Rawls W.J. (Eds) Development of Pedotransfer Functions in Soil Hydrology, Developments in Soil Science. pp. 115–141. https://doi.org/10.1016/S0166-2481(04)30007-3

Lim, H., Yang, H., Chun, K.W., Choi, H.T., 2020. Development of Pedo-Transfer Functions for the Saturated Hydraulic Conductivity of Forest Soil in South Korea Considering Forest Stand and Site Characteristics. Water (Switzerland) 12, 1–18.

Malone, B.P., Luo, Z., He, D., Viscarra Rossel, R.A., Wang, E., 2020. Bioclimatic variables as important spatial predictors of soil hydraulic properties across Australia's agricultural region. Geoderma Reg. 23. https://doi.org/10.1016/j.geodrs.2020.e00344

Martinez, P., Souza, I.F., 2020. Genesis of pseudo-sand structure in Oxisols from Brazil – A review. Geoderma Reg. 22, e00292. https://doi.org/10.1016/j.geodrs.2020.e00292

Medrado, E., Lima, J.E.F.W., 2014. Development of pedotransfer functions for estimating water retention curve for tropical soils of the Brazilian savanna. Geoderma Reg. 1, 59–66. https://doi.org/10.1016/j.geodrs.2014.08.003

Méndez-Toribio, M., Ibarra-Manríquez, G., Paz, H., Lebrija-Trejos, E., 2020. Atmospheric and soil drought risks combined shape community assembly of trees in a tropical dry forest. J. Ecol. 108, 1347–1357. https://doi.org/10.1111/1365-2745.13355

Minasny, B., Hartemink, A.E., 2011. Predicting soil properties in the tropics. Earth-Science Rev. 106, 52–62. https://doi.org/10.1016/j.earscirev.2011.01.005

Minasny, B., Heuvelink, G., 2008. Analysing uncertainty, in: Guidelines for Surveying Soil and Land Resources. 2nd Ed. NJ McKenzie, MJ Grundy, R Webster, AJ Ringrose-Voase. p. 557.

Minasny, B., McBratney, A.B., 2002a. Uncertainty analysis for pedotransfer functions. Eur. J. Soil Sci. 53, 417–429. https://doi.org/10.1046/j.1365-2389.2002.00452.x

Minasny, B., McBratney, A.B., 2002b. The efficiency of various approaches to obtaining estimates of soil hydraulic properties. Geoderma 107, 55–70. https://doi.org/10.1016/S0016-7061(01)00138-0

Minasny, B., Mcbratney, A.B., Bristow, K.L., 1999. Comparison of different approaches to the development of pedotransfer functions for water-retention curves. Geoderma 93, 225–253.

Moncada, M.P., Gabriels, D. and Cornelis, W.M., 2014. Data-driven analysis of soil quality indicators using limited data. Geoderma, 235, pp.271-278. https://doi.org/10.1016/j.geoderma.2014.07.014

Morgan, M.T., Holmes, R.G., Wood, R.K., 1993. A system for measuring soil physical properties in the field. Soil Tillage Res. 26, 301–325. https://doi.org/10.1016/0167-1987(93)90003-8

Moura, M.S., Silva, B.M., Mota, P.K., Borghi, E., Resende, A.V. de, Acuña-Guzman, S.F., Araújo, G.S.S., da Silva, L. de C.M., de Oliveira, G.C., Curi, N., 2021. Soil management and diverse crop rotation can mitigate early-stage no-till compaction and improve least limiting water range in a Ferralsol. Agric. Water Manag. 243. https://doi.org/10.1016/j.agwat.2020.106523

Nearing, G.S., Kratzert, F., Sampson, A.K., Pelissier, C.S., Klotz, D., Frame, J.M., Prieto, C., Gupta, H. V., 2020. What Role Does Hydrological Science Play in the Age of Machine Learning? Water Resour. Res. https://doi.org/10.1029/2020wr028091

Nemes, A., Schaap, M.G., Leij, F.J., Wösten, J.H.M., 2001. Description of the unsaturated soil hydraulic database UNSODA version 2.0. J. Hydrol. 251, 151–162.

https://doi.org/10.1016/S0022-1694(01)00465-6

Nguyen, P.M., De Pue, J., Le, K. Van, Cornelis, W., 2015a. Impact of regression methods on improved effects of soil structure on soil water retention estimates. J. Hydrol. 525, 598–606. https://doi.org/10.1016/j.jhydrol.2015.04.014

Nguyen, P.M., Le, K. Van, Cornelis, W.M., 2014. Using categorical soil structure information to improve soil water retention estimates of tropical delta soils. Soil Res. 52, 443–452. https://doi.org/10.1071/SR13256

Nguyen, P.M., Van Le, K., Botula, Y.D., Cornelis, W.M., 2015b. Evaluation of soil water retention pedotransfer functions for Vietnamese Mekong Delta soils. Agric. Water Manag. 158, 126–138. https://doi.org/10.1016/j.agwat.2015.04.011

Ottoni, M.V., Ottoni Filho, T.B., Lopes-Assad, M.L.R.C., Rotunno Filho, O.C., 2019. Pedotransfer functions for saturated hydraulic conductivity using a database with temperate and tropical climate soils. J. Hydrol. 575, 1345–1358. https://doi.org/10.1016/j.jhydrol.2019.05.050

Ottoni, M.V., Ottoni Filho, T.B., Schaap, M.G., Lopes-Assad, M.L.R.C., Rotunno Filho, O.C., 2018. Hydrophysical Database for Brazilian Soils (HYBRAS) and Pedotransfer Functions for Water Retention. Vadose Zo. J. 17, 170095. https://doi.org/10.2136/vzj2017.05.0095

Pachepsky, Y.A., Rajkai, K., Tóth, B., 2015. Pedotransfer in soil physics: Trends and outlook - A review. Agrokem. es Talajt. 64, 339–360. https://doi.org/10.1556/0088.2015.64.2.3

Pachepsky, Y.A., Rawls, W.J., 2003. Soil structure and pedotransfer functions. Eur. J. Soil Sci. 54, 443–451.

Pachepsky, Y.A., Rawls, W.J., Lin, H.S., 2006. Hydropedology and pedotransfer functions. Geoderma 131, 308–316. https://doi.org/10.1016/j.geoderma.2005.03.012

Pachepsky, Y. A., Timlin, D., Varallyay, G., 1996. Artificial Neural Networks to Estimate Soil Water Retention from Easily Measurable Data. Soil Sci. Soc. Am. J. 60, 727–733. https://doi.org/10.2136/sssaj1996.03615995006000030007x

Pachepsky, Ya. A., Timlin, D., Varallyay, G., 1996. Artificial Neural Networks to Estimate Soil Water Retention from Easily Measurable Data. Soil Sci. Soc. Am. J. 60, 727–733. https://doi.org/10.2136/sssaj1996.03615995006000030007x

Padarian, J., Minasny, B., McBratney, A.B., 2020. Machine learning and soil sciences: A review aided by machine learning tools. Soil 6, 35–52. https://doi.org/10.5194/soil-6-35-2020

Padarian, J., Morris, J., Minasny, B., McBratney, A.B., 2018. Chapter 7 Pedotransfer

Functions and Soil Inference Systems, in: Pedometrics. pp. 195–220. https://doi.org/10.1007/978-3-319-63439-5_7

Patil, N.G., Singh, S.K., 2016. Pedotransfer Functions for Estimating Soil Hydraulic Properties: A Review. Pedosphere 26, 417–430. https://doi.org/10.1016/S1002-0160(15)60054-6

Patil, N.G., Tiwary, P., Pal, D.K., Bhattacharyya, T., Sarkar, D., Mandal, C., Mandal, D.K., Chandran, P., Ray, S.K., Prasad, J., Lokhande, M., Dongre, V., 2013. Soil Water Retention Characteristics of Black Soils of India and Pedotransfer Functions Using Different Approaches. J. Irrig. Drain. Eng. 139, 313–324. https://doi.org/10.1061/(asce)ir.1943-4774.0000527

Pessoa, T.N., Libardi, P.L., 2022. Physical-hydric properties of Oxisols as influenced by soil structure and clay mineralogy. Catena 211, 106009. https://doi.org/10.1016/j.catena.2021.106009

Pham, K., Kim, D., Yoon, Y., Choi, H., 2019. Analysis of neural network based pedotransfer function for predicting soil water characteristic curve. Geoderma 351, 92–102. https://doi.org/10.1016/j.geoderma.2019.05.013

Picciafuoco, T., Morbidelli, R., Flammini, A., Saltalippi, C., Corradini, C., Strauss, P., Blöschl, G., 2019. A Pedotransfer Function for Field-Scale Saturated Hydraulic Conductivity of a Small Watershed. Vadose Zo. J. 18, 1–15. https://doi.org/10.2136/vzj2019.02.0018

Poggio, L., Gimona, A., Brown, I., Castellazzi, M., 2010. Soil available water capacity interpolation and spatial uncertainty modelling at multiple geographical extents. Geoderma 160, 175–188. https://doi.org/10.1016/j.geoderma.2010.09.015

Pollacco, J.A.P., Webb, T., McNeill, S., Hu, W., Carrick, S., Hewitt, A., Lilburne, L., 2017. Saturated hydraulic conductivity model computed from bimodal water retention curves for a range of New Zealand soils. Hydrol. Earth Syst. Sci. 21, 2725–2737. https://doi.org/10.5194/hess-21-2725-2017

Quentin, S., Philippe, L., 2021. Uncertainty assessment of soil available water capacity using error propagation: A test in Languedoc-Roussillon. Geoderma 391, 114968. https://doi.org/10.1016/j.geoderma.2021.114968

Rab, M.A., Chandra, S., Fisher, P.D., Robinson, N.J., Kitching, M., Aumann, C.D., Imhof, M., 2011. Modelling and prediction of soil water contents at field capacity and permanent wilting point of dryland cropping soils. Soil Res. 49, 389–407. https://doi.org/10.1071/SR10160

Ramcharan, A., Hengl, T., Beaudette, D., Wills, S., 2017. A Soil Bulk Density Pedotransfer Function Based on Machine Learning: A Case Study with the NCSS Soil Characterization Database. Soil Sci. Soc. Am. J. 81, 1279–1287. https://doi.org/10.2136/sssaj2016.12.0421

Rastgou, M., Bayat, H., Mansoorizadeh, M., Gregory, A.S., 2020. Estimating the soil water retention curve: Comparison of multiple nonlinear regression approach and random forest data mining technique. Comput. Electron. Agric. 174, 105502. https://doi.org/10.1016/j.compag.2020.105502

Rawls, W.J., Pachepsky, Y.A., 2002a. Soil Consistence and Structure as Predictors of Water Retention. Soil Sci. Soc. Am. J. 66, 1115–1126. https://doi.org/10.2136/sssaj2002.1115

Rawls, W.J., Pachepsky, Y.A., 2002b. Using field topographic descriptors to estimate soil water retention. Soil Sci. 167, 423–435. https://doi.org/10.1097/01.ss.0000022697.98626.c6

Reichert, J.M., Albuquerque, J.A., Kaiser, D.R., Reinert, D.J., Urach, F.L., Carlesso, R., 2009. Estimation of water retention and availability in soils of Rio Grande do Sul. Rev. Bras. Cienc. do Solo 33, 1547–1560. https://doi.org/10.1590/s0100-06832009000600004

Resende, M., Curi, N., Rezende, S.B., Corrêa, G.F., Ker, J.C., 2014. Pedologia: base para distinção de ambientes. Lavras Ed. UFLA 6. ed. rev, 322 p.

Romano, N., Chirico, G.B., 2004. The role of terrain analysis in using and developing pedotransfer functions. Pachepsky Y., Rawls W.J. Dev. Pedotrans. Funct. soil Hydrol. Dev. soil Sci. 30, 273–294. https://doi.org/10.1016/S0166-2481(04)30016-4

Romano, N., Palladino, M., 2002. Prediction of soil water retention using soil physical data and terrain attributes. J. Hydrol. 265, 56–75. https://doi.org/10.1016/S0022-1694(02)00094-X

Santos, H.G., Jacomine, P.K.T., Anjos, L.H.C., Oliveira, V.A., Lumbreras, J.F., Coelho, M.R., Almeida, J.A., Araújo Filho, J.C., Oliveira, J.B., Cunha, T.J.F., 2018. Sistema brasileiro de classificação de solos, Embrapa Solos.

Santos, R.C.V., Vargas, M.M., Timm, L.C., Beskow, S., Siqueira, T.M., Mello, C.R., Soares, M.F., de Moura, M.M., Reichardt, K., 2021. Examining the implications of spatial variability of saturated soil hydraulic conductivity on direct surface runoff hydrographs. Catena 207. https://doi.org/10.1016/j.catena.2021.105693

Santra, P., Kumar, M., Kumawat, R.N., Painuli, D.K., Hati, K.M., Heuvelink, G.B.M., Batjes, N.H., 2018. Pedotransfer functions to estimate soil water content at field capacity and permanent wilting point in hot Arid Western India. J. Earth Syst. Sci. 127, 1–16.

https://doi.org/10.1007/s12040-018-0937-0

Schaap, M.G., 2005. Models for indirect estimation of soil hydraulic properties, in: Anderson, M., Ed. Encyclopedia of Hydrological Sciences. New York, John Wiley and Sons. pp. 1145–1150.

Schaap, M.G., 2004. Accuracy and uncertainty in PTF predictions, in: Y. Pachepsky and W.J. Rawls. Development of Pedotransfer Functions in Soil Hydrology. pp. 33–43. https://doi.org/10.1016/S0166-2481(04)30003-6

Schaap, M.G., 2001. Rosetta v1.2: A computer program for estimating soil hydraulic parameters with hierarchical pedotransfer functions 251, 163–176.

Schaap, M.G., Leij, F.J., 1998. Using neural networks to predict soil water retention and soil hydraulic conductivity. Soil Tillage Res. 47, 37–42. https://doi.org/10.1016/S0167-1987(98)00070-1

Schaefer, C.E.G.R., Fabris, J.D., Ker, J.C., 2008. Minerals in the clay fraction of Brazilian Latosols (Oxisols): a review. Clay Miner. 43, 137–154. https://doi.org/10.1180/claymin.2008.043.1.11

Scheffler, R., Neill, C., Krusche, A. V., Elsenbeer, H., 2011. Soil hydraulic response to land-use change associated with the recent soybean expansion at the Amazon agricultural frontier. Agric. Ecosyst. Environ. 144, 281–289. https://doi.org/10.1016/j.agee.2011.08.016

Silva, A.C., Armindo, R.A., dos Santos Brito, A., Schaap, M.G., 2017. SPLINTEX: A physically-based pedotransfer function for modeling soil hydraulic functions. Soil Tillage Res. 174, 261–272. https://doi.org/10.1016/j.still.2017.07.011

Silva, B.M., Silva, É.A. da, Oliveira, G.C. de, Ferreira, M.M., Serafim, M.E., 2014. Plant-available soil water capacity: estimation methods and implications. Rev. Bras. Ciência do Solo 38, 464–475. https://doi.org/10.1590/s0100-06832014000200011

Silva, L.C.M., Avanzi, J.C., Peixoto, D.S., Merlo, M.N., Borghi, E., Resende, Á.V. de, Acuña-Guzman, S.F., Silva, B.M., 2021. Ecological intensification of cropping systems enhances soil functions, mitigates soil erosion, and promotes crop resilience to dry spells in the Brazilian Cerrado. Int. Soil Water Conserv. Res. 9, 591–604. https:/

Silva, L.C.M., Peixoto, D.S., Gomes, J.B.V., Avanzi, J.C., Amorim, R.S.S., Borghi, E., de Resende, A.V., Silva, B.M., Mancini, M. and Curi, N., 2022. Mineralogy and pore size distribution of clayey Oxisols with granular structure and the effect of management systems. Soil and Tillage Research, 223, p.105479. https://doi.org/10.1016/j.iswcr.2021.06.006

Silva, E., Curi, N., Ferreira, M.M., Volpato, M.M.L., Santos, W.J.R., Silva, S.H.G., 2015. Pedotransfer functions for water retention in the main soils from the Brazilian Coastal Plains. Cienc. e Agrotecnologia 39, 331–338.

Silva, S.H.G., Weindorf, D.C., Pinto, L.C., Faria, W.M., Acerbi Junior, F.W., Gomide, L.R., de Mello, J.M., de Pádua Junior, A.L., de Souza, I.A., Teixeira, A.F. dos S., Guilherme, L.R.G., Curi, N., 2020. Soil texture prediction in tropical soils: A portable X-ray fluorescence spectrometry approach. Geoderma 362. https://doi.org/10.1016/j.geoderma.2019.114136

Singh, A., Haghverdi, A., Öztürk, H.S., Durner, W., 2020. Developing pseudo continuous pedotransfer functions for international soils measured with the evaporation method and the hyprop system: I. the soil water retention curve. Water (Switzerland) 13. https://doi.org/10.3390/w13060878

Svensson, D.N., Messing, I. and Barron, J., 2022. An investigation in laser diffraction soil particle size distribution analysis to obtain compatible results with sieve and pipette method. Soil and Tillage Research, 223, p.105450. https://doi.org/10.1016/j.still.2022.105450

Szabó, B., Szatmári, G., Takács, K., Laborczi, A., Makó, A., Rajkai, K., Pásztor, L., 2019. Mapping soil hydraulic properties using random-forest-based pedotransfer functions and geostatistics. Hydrol. Earth Syst. Sci. 23, 2615–2635. https://doi.org/10.5194/hess-23-2615-2019

Tian, Z., Chen, J., Cai, C., Gao, W., Ren, T., Heitman, J.L., Horton, R., 2021. New pedotransfer functions for soil water retention curves that better account for bulk density effects. Soil Tillage Res. 205, 104812. https://doi.org/10.1016/j.still.2020.104812

Tomasella, J., Hodnett, M., 2004. Pedotransfer functions for tropical soils. Dev. Soil Sci. 30, 415–429. https://doi.org/10.1016/S0166-2481(04)30021-8

Tomasella, J., Hodnett, M.G., 1998. Estimating soil water retention characteristics from limited data in Brazilian Amazonia. Soil Sci. 163, 190–202. https://doi.org/10.1097/00010694-199803000-00003

Tomasella, J., Hodnett, M.G., Rossato, L., 2000. Pedotransfer Functions for the Estimation of Soil Water Retention in Brazilian Soils. Soil Sci. Soc. Am. J. 64, 327–338. https://doi.org/10.2136/sssaj2000.641327x

Tomasella, J., Pachepsky, Y., Crestana, S., Rawls, W.J., 2003. Comparison of Two Techniques to Develop Pedotransfer Functions for Water Retention. Soil Sci. Soc. Am. J. 67, 1085–1092. https://doi.org/10.2136/sssaj2003.1085

Tóth, B., Weynants, M., Nemes, A., Makó, A., Bilas, G., Tóth, G., 2015. New generation of hydraulic pedotransfer functions for Europe. Eur. J. Soil Sci. 66, 226–238. https://doi.org/10.1111/ejss.12192

Touil, S., Degre, A., Chabaca, M.N., 2016. Sensitivity analysis of point and parametric pedotransfer functions for estimating water retention of soils in Algeria. Soil 2, 647–657. https://doi.org/10.5194/soil-2-647-2016

Van Den Berg, M., Klamt, E., Van Reeuwijk, L.P., Sombroek, W.G., 1997. Pedotransfer functions for the estimation of moisture retention characteristics of Ferralsols and related soils. Geoderma 78, 161–180. https://doi.org/10.1016/S0016-7061(97)00045-1

van Genuchten, M.T., 1980. A Closed-form Equation for Predicting the Hydraulic Conductivity of Unsaturated Soils. Soil Sci. Soc. Am. J. https://doi.org/10.2136/sssaj1980.03615995004400050002x

Van Looy, K., Bouma, J., Herbst, M., Koestel, J., Minasny, B., Mishra, U., Montzka, C., Nemes, A., Pachepsky, Y.A., Padarian, J., Schaap, M.G., Tóth, B., Verhoef, A., Vanderborght, J., van der Ploeg, M.J., Weihermüller, L., Zacharias, S., Zhang, Y., Vereecken, H., 2017. Pedotransfer Functions in Earth System Science: Challenges and Perspectives. Rev. Geophys. 55, 1199–1256. https://doi.org/10.1002/2017RG000581

Vereecken, H., Weynants, M., Javaux, M., Pachepsky, Y., Schaap, M.G., van Genuchten, M.T., 2010. Using Pedotransfer Functions to Estimate the van Genuchten-Mualem Soil Hydraulic Properties: A Review. Vadose Zo. J. 9, 795–820. https://doi.org/10.2136/vzj2010.0045

Visser, S., Keesstra, S., Maas, G., de Cleen, M., Molenaar, C., 2019. Soil as a basis to create enabling conditions for transitions towards sustainable land management as a key to achieve the SDGs by 2030. Sustain. 11. https://doi.org/10.3390/su11236792

Vitorino, A.C.T., Ferreira, M.M., Curi, N., Lima, J.M., Silva, M.L.N., Motta, P.E.F., 2003. Mineralogia, química e estabilidade de agregados do tamanho de silte de solos da Região Sudeste do Brasil. Pesq. Agropec. Bras. 38, 133–141. https://doi.org/10.1590/S0100-204X2003000100018Wang, S., Huang, G.H., He, L., 2012. Development of a clusterwise-linear-regression-based forecasting system for characterizing DNAPL dissolution behaviors in porous media. Sci. Total Environ. 433, 141–150. https://doi.org/10.1016/j.scitotenv.2012.06.045

Wassar, F., Gandolfi, C., Rienzner, M., Chiaradia, E.A., Bernardoni, E., 2016. Predicted and measured soil retention curve parameters in Lombardy region north of Italy. Int. Soil Water Conserv. Res. 4, 207–214. https://doi.org/10.1016/j.iswcr.2016.05.005

Witten, I.H., Frank, E., Mark, A., 2011. Data Mining Practical Machine Learning Tools and Techniques.

Wolpert, D.H., 1996. The Lack of a Priori Distinctions between Learning Algorithms. Neural Comput. 8, 1341–1390. https://doi.org/10.1162/neco.1996.8.7.1341

Wösten, J.H.M., Lilly, A., Nemes, A., Le Bas, C., 1999. Development and use of a database of hydraulic properties of European soils. Geoderma 90, 169–185. https://doi.org/10.1016/S0016-7061(98)00132-3

WRB, W.R.B. for S.R., 2015. International soil classification system for naming soils and creating legends for soil maps,. World Soil Resour. Reports 106, 1–191.

Wu, X., Kumar, V., Ross, Q.J., Ghosh, J., Yang, Q., Motoda, H., McLachlan, G.J., Ng, A., Liu, B., Yu, P.S., Zhou, Z.H., Steinbach, M., Hand, D.J., Steinberg, D., 2008. Top 10 algorithms in data mining. Knowl. Inf. Syst. 14, 1–37. https://doi.org/10.1007/s10115-007-0114-2

Wyatt, B.M., Ochsner, T.E., Brown, W.G., Diggins, D.C., Illston, B.G., Fiebrich, C.A., 2021. MesoSoil v2.0: An updated soil physical property database for the Oklahoma Mesonet. Vadose Zo. J. 1–9. https://doi.org/10.1002/vzj2.20134

Yamaç, S.S., Şeker, C., Negiş, H., 2020. Evaluation of machine learning methods to predict soil moisture constants with different combinations of soil input data for calcareous soils in a semi arid area. Agric. Water Manag. 234. https://doi.org/10.1016/j.agwat.2020.106121

Yang, Y., Jia, X., Wendroth, O., Liu, B., 2018. Estimating Saturated Hydraulic Conductivity along a South-North Transect in the Loess Plateau of China. Soil Sci. Soc. Am. J. 82, 1033–1045. https://doi.org/10.2136/sssaj2018.03.0126

Zhang, Y., Schaap, M.G., 2019. Estimation of saturated hydraulic conductivity with pedotransfer functions: A review. J. Hydrol. 575, 1011–1030. https://doi.org/10.1016/j.jhydrol.2019.05.058

Zhang, Y., Schaap, M.G., 2017. Weighted recalibration of the Rosetta pedotransfer model with improved estimates of hydraulic parameter distributions and summary statistics (Rosetta3). J. Hydrol. 547, 39–53. https://doi.org/10.1016/j.jhydrol.2017.01.004

Zhang, Y., Schaap, M.G., Wei, Z., 2020. Development of Hierarchical Ensemble Model and Estimates of Soil Water Retention With Global Coverage. Geophys. Res. Lett. 47, 0–1. https://doi.org/10.1029/2020GL088819

3. Paper II: Bridging the gap: creating regional-specific pedotransfer functions through machine learning algorithms

Abstract

Quantitative data for soil hydraulic properties (SHP) is a crucial input for several environmental and agricultural applications. As SHP can be laborious and expensive to determine, pedotransfer functions (PTFs) are a prospective alternative to estimate SHP. Although PTFs have been used to predict SHP, improvements are still pursued. The region-specific nature of PTFs generates a demand for accurate PTFs for specific locations and purposes. Therefore, this study aimed to develop new PTFs for a major agricultural center in Brazil, the Mato Grosso state. PTFs were developed using machine learning techniques and four algorithms were tested (RF, SVM, KNN and XGBoost) in addition to the conventional multiple linear regression. PTFs were developed for predicting soil hydraulic conductivity and soil water content at matric potentials of 0, -4, -6, -10, -33, -100, -500 and -1500 kPa. Three different predictor sets were compared to test whether the increase of features can overcome limited datasets for PTFs training. Predictors included sand, silt, clay, water dispersible clay and bulk density for the basic predictor set. The other two predictor sets included organic carbon and organic carbon plus chemical properties. The PTFs domain of application was determined comparing the similarity of calibration datasets to a comprehensive database of soils from Mato Grosso by means of Mahalanobis distances. Overall, results showed that RF was the best model for predicting soil water retention regardless of predictor set, and KNN was the best model for predicting soil hydraulic conductivity when additional features were added to basic predictors. Prediction of SHP was systematically improved by the addition of features to predictor sets, except for water retention at drier matric potentials. Water dispersible clay was an important feature to predict soil water retention and chemical properties improved predictions of hydraulic conductivity. PTFs are well calibrated for 91% of the Mato Grosso state area if using basic predictors, but this area of application decreases substantially when organic carbon and chemical properties are included as predictors.

Keywords: pedotransfer functions; tropical soils; soil hydraulic conductivity; available water capacity; random forest

1. Introduction

Water is the most limiting factor for agricultural crop growth (Naorem et al., 2023; Silva et al., 2019; Srayeddin and Doussan, 2009) and knowledge about soil hydraulic properties (SHP) is essential for assessing water availability for plants (Qiao et al., 2019; Silva et al., 2014; Torres et al., 2021) and productivity of cropping systems (Silva et al., 2023a, 2021). Moreover, SHP are required as input data for Earth system models (Gupta et al., 2020; Lehmann et al., 2020; Shao et al., 2023), playing a key role for several hydrological and agricultural applications. Nevertheless, SHP are not easily found and made available in public databases, mainly because they are laborious and expensive to determine via conventional methods (Botula et al., 2014; Romano and Palladino, 2002; Santra et al., 2018). This difficulty in obtaining data leads to a lack of information on SHP, which restricts their usage and application across the globe (Patil and Singh, 2016; Van Looy et al., 2017). While data scarcity is an issue, predictive models can be developed to indirectly estimate SHP (Botula et al., 2014), such as pedotransfer functions (PTFs).

PTFs were introduced by Bouma (1989) as an alternative method to predict complex soil attributes from easy-to-measure properties and have been widely used for estimating soil water retention (Botula et al., 2014; Cueff et al., 2021; Rab et al., 2011; Dobarco et al., 2019; Quentin and Philipe, 2021) and soil water movement (Araya and Ghezzehei, 2019; Gupta et al., 2020; Ottoni et al., 2019; Shao et al., 2023; Zhang and Schaap, 2019), especially at regional scales (Silva et al., 2023b). However, despite several advantages, PTFs carry sources of error and uncertainty (Chirico et al., 2010; Liao et al., 2014; Mcbratney et al., 2002; Minasny et al., 1999) and their improvement continues to be pursued (Van Looy et al., 2017). For this purpose, machine learning techniques have been used as an attempt to improve the calibration of PTFs (Araya and Ghezzehei, 2019; Benke et al., 2020; Guio Blanco et al., 2018; Gunarathna et al., 2019; Patil and Singh, 2016; Pham et al., 2019; Ramcharan et al., 2017; Van Looy et al., 2017; Yamaç et al., 2020). Machine learning-based PTFs have several advantages compared to classical regression techniques, e.g., not requiring strict assumptions about data distribution (Pham et al., 2019) and being convenient for fitting complex and non-linear relationships (Witten et al., 2011). Moreover, machine learning-based PTFs in general present better performance for predicting SHP than conventional regression (Silva et al., 2023b). Nevertheless, although consistent advances in data mining techniques have been achieved, the reliable application of PTFs is a current challenge (Benke et al., 2020; Kotlar et al., 2019; Van Looy et al., 2017) especially due to the region-specific nature of PTFs

(Choudhury et al., 2023), which brings a demand for more regional PTFs, at least until accurate generic PTFs are developed (Patil and Singh, 2016).

A well-known challenge concerning PTF usage is their application in different environmental conditions from which they were calibrated. Generalised PTFs may have limited reliability due to regional soil singularities (Minasny et al., 1999; Tomasella and Hodnett, 2004). For example, most of the available PTFs were developed in temperate climates (Botula et al., 2013; Gunarathna et al., 2019; Ottoni et al., 2019), which hinders their application to tropical soils (Ottoni et al., 2018) due to divergent patterns caused by specificities such as soil structure and clay mineralogy (Ferreira et al., 1999; Ker, 1997; Goedert, 1983; Schaefer et al., 2008; Silva et al., 2022). In this sense, several studies have been performed to develop PTFs (Botula et al., 2014; da Silva et al., 2017; Kotlar et al., 2020, 2019; Medrado and Lima, 2014; Ottoni et al., 2018; Tomasella et al., 2000; Tomasella and Hodnett, 2004, 1998; Veloso et al., 2022) and to assemble a databases regarding SHPs of tropical soils, e.g., the Hydrophysical Database for Brazilian soils (HYBRAS; Ottoni et al., 2018). Despite efforts in Brazil, few studies have addressed a major agricultural center in the country, the Mato Grosso state. For instance, the HYBRAS library does not contain data from Mato Grosso and the recent PTF by Medrado and Lima (2014) for soils of the Brazilian Cerrado included only 25 samples from Mato Grosso in their calibration dataset. Furthermore, local studies (e.g., Amorim et al., 2022; Rosseti et al., 2022) have addressed limited properties and geographical coverage. Therefore, specific, and comprehensive predictive models for depicting soil hydraulic patterns in the state of Mato Grosso are currently lacking.

The Mato Grosso state is a globally significant agricultural region (Cohn et al., 2016) that leads Brazil’s production of soybean, corn, cotton (Cassol et al., 2020; Di Raimo et al., 2022) and beef (Teixeira et al., 2020). Mato Grosso is the third largest state in Brazil with a large variation in soils and land use diversity across its ~90 million hectares. Besides agriculture, Mato Grosso is environmentally significant as it embraces three important biomes (i.e., Amazon, Cerrado and Pantanal) which have been threatened by deforestation (Scheffler et al., 2011; Simoes et al., 2020). Hence, specific and comprehensive predictions for hydraulic patterns of Mato Grosso soils would benefit both environmental and agricultural applications.

In this context, a comprehensive database was used to develop PTFs for water content at matric potentials of 0, -4, -6, -10, -33, -100, -500 and -1500 kPa as well as for hydraulic conductivity of soils from Mato Grosso state. Several machine learning techniques (i.e., Random Forest, Support Vector Machine, k-Nearest Neighbors, and eXtreme Gradient Boosting) were tested and compared against each other as well as against classical Multiple

Linear Regression and the null model. Three sets of predictors with different sizes were applied to test whether the increase of features improves the accuracy of PTFs when the number of samples for training is limited. Furthermore, the PTFs geographical domain of application was determined within the state area, evaluating the similarity between the PTFs calibration dataset and a comprehensive database of predictors from the Socioeconomic and Ecological Zoning of the Mato Grosso state (ZSEE/MT). The PTFs domain of application shows where the PTFs are well-calibrated within the state territory. Settling the domain for application can guide the decision making of potential users regarding what predictors and where they could apply the models with reasonable reliability. Furthermore, the domain of application provides directions towards where further data collection is needed, along with insights about what variables and range of values are desired, aiming to either develop more representative models or widen their geographical applicability.

2. Material and Methods

2.1 Study area

The study was performed with data from the Mato Grosso State, in the Central-west region of Brazil. The characteristic climate is hot and semi-humid tropical, with an average temperature ranging between 18ºC and 24ºC, an annual average maximum of 34ºC, comprising humid summers and dry winters (May to September), and annual precipitation ranging from 1,200 to 2,000 mm. The predominant soil classes in the state of Mato Grosso are Ferralsols, Acrisols and Arenosols (IUSS, 2015) corresponding to *Latossolos*, *Argissolos* e *Neossolos Quartzarênicos* in the Brazilian Soil Classification System (Santos et al., 2018). These soil classes cover, respectively, 41%, 25% and 13% of the state’s area (Di Raimo et al., 2019).

2.2 Development of pedotransfer functions

2.2.1 Database pre-processing and predictor sets

A comprehensive database containing hydraulic properties of soils from Mato Grosso was assembled to generate PTFs for the prediction of soil water content at matric potentials of 0, -4, -6, -10, -33, -100, -500, and -1500 kPa as well as for the prediction of soil hydraulic conductivity. The database encompasses soil data from the three biomes of the state: Cerrado (neotropical savanna), Amazon and Pantanal. The database contains information about several soil physical and chemical properties which were determined according to Teixeira et al. (2017). Part of the data was used in previous works addressing soil hydraulic properties from Mato Grosso, and determination methods are described in Bocuti (2020) and Bocuti et al. (2021). All data in this database refer to soils under agricultural use, mainly grain (soybean and corn) and pastures. Table 1 shows descriptive statistics of the database. Data refer to the topsoil layer (0-20 cm) of Acrisols, Cambisols, Gleysols, Ferralsols, Arenosols, Nitisols, Planosols and Plinthosols (IUSS, 2015), corresponding to *Argissolos*, *Cambissolos*, *Gleissolos*, *Latossolos*, *Neossolos*, *Nitossolos*, *Planossolos* and *Plintossolos* in the Brazilian Soil Classification System (Santos et al., 2018).

In order to test diverse predictors in the modeling process, the database was split into three datasets with different sizes (Table 1). The first dataset is referred to as "Basic" (n = 669) and contains records of the following predictors: sand (2.0 > ⌀ > 0.05 mm), silt (0.05 > ⌀ > 0.002 mm), clay (⌀ < 0.002 mm), bulk density (Bd), and water dispersible clay (Wdc). For predicting soil hydraulic conductivity, the Basic dataset was reduced (n = 566) due to missing data from the dependent variable. The second dataset is "Basic + OC" (n = 228), for which data of organic carbon was added to the Basic dataset variables. The Basic + OC dataset size was conditioned by the available number of OC data, as shown in Table 1. The third dataset is "Basic + OC + chemical" (n = 114), for which soil chemical properties (i.e., pH in water, cation exchange capacity, and base saturation) were added to the second dataset. It is important to note that the chemical properties were determined using disturbed samples collected in the 0-20 cm soil layer, whereas soil physical and hydraulic properties were determined separately for 0-10 and 10-20 cm at each sampling point. Therefore, chemical properties are associated with the mean value of physical and hydraulic properties in the layer 0-20 cm, which led to equal average values of soil physical and hydraulic properties between Basic + OC and Basic + OC + chemical datasets. The size of the third dataset was conditioned by the number of records of soil chemical properties.

Table 1. Descriptive statistics of datasets used for developing pedotransfer functions.

Basic dataset (n = 669)[a]	Min	Median	Mean	Max	SD
Sand (%)	1.43	59.46	56.28	97.77	22.75
Silt (%)	0.53	9.27	13.08	63.81	12.36
Clay (%)	1.00	30.49	30.58	69.79	16.49
Bd (Mg m^{-3})	0.46	1.50	1.47	1.91	0.21
Wdc (g kg^{-1})	0.00	5.52	7.33	47.06	6.87

0 kPa (m^{-3} m^{-3})	0.25	0.44	0.45	0.79	0.08
-4 kPa (m^{-3} m^{-3})	0.20	0.37	0.37	0.74	0.06
-6 kPa (m^{-3} m^{-3})	0.15	0.33	0.34	0.71	0.07
-10 kPa (m^{-3} m^{-3})	0.11	0.31	0.31	0.66	0.08
-33 kPa (m^{-3} m^{-3})	0.08	0.28	0.28	0.65	0.09
-100 kPa (m^{-3} m^{-3})	0.06	0.25	0.25	0.50	0.08
-500 kPa (m^{-3} m^{-3})	0.06	0.19	0.19	0.37	0.07
-1500 kPa (m^{-3} m^{-3})	0.02	0.14	0.14	0.31	0.07
Ks (mm h^{-1})	0.06	23.69	59.26	849.40	100.38
$\log_{10}$ (Ks (mm h^{-1}))	-1.25	1.37	1.27	2.93	0.79

Basic + OC (n = 228)	Min	Median	Mean	Max	SD
Sand (%)	1.59	58.25	55.62	95.30	22.61
Silt (%)	1.22	9.23	13.34	63.79	12.64
Clay (%)	1.34	31.09	31.03	67.21	16.46
Bd (Mg m^{-3})	1.02	1.47	1.46	1.84	0.18
Wdc (g kg^{-1})	0.01	5.92	7.31	27.05	6.36
OC (g kg^{-1})	1.99	5.15	5.73	11.35	2.39
0 kPa (m^{-3} m^{-3})	0.32	0.43	0.44	0.66	0.07
-4 kPa (m^{-3} m^{-3})	0.07	0.25	0.25	0.45	0.08
-6 kPa (m^{-3} m^{-3})	0.06	0.23	0.23	0.41	0.08
-10 kPa (m^{-3} m^{-3})	0.06	0.21	0.21	0.38	0.08
-33 kPa (m^{-3} m^{-3})	0.05	0.18	0.18	0.35	0.07
-100 kPa (m^{-3} m^{-3})	0.04	0.15	0.15	0.34	0.06
-500 kPa (m^{-3} m^{-3})	0.04	0.12	0.13	0.34	0.06
-1500 kPa (m^{-3} m^{-3})	0.03	0.11	0.11	0.33	0.06
Ks (mm h^{-1})	0.07	27.71	61.02	788.86	93.35
$\log_{10}$ (Ks (mm h^{-1}))	-1.19	1.44	1.40	2.90	0.65

Basic + OC + chemical (n = 114)	Min	Median	Mean	Max	SD
Sand (%)	1.87	58.01	55.62	94.72	22.56
Silt (%)	1.47	8.79	13.34	62.01	12.51
Clay (%)	1.88	31.14	31.03	62.79	16.35
Bd (Mg m^{-3})	1.06	1.48	1.46	1.78	0.17
Wdc (g kg^{-1})	0.05	5.90	7.31	23.54	6.23
OC (g kg^{-1})	1.99	5.19	5.73	11.35	2.39
pH (H_2O)	4.60	5.80	5.82	8.00	0.61
CEC (cmolc dm^{-3})	3.31	7.15	7.36	15.68	2.42
BS (%)	5.11	44.74	46.27	94.80	16.51
0 kPa (m^{-3} m^{-3})	0.33	0.43	0.44	0.62	0.06
-4 kPa (m^{-3} m^{-3})	0.07	0.25	0.25	0.44	0.08

-6 kPa (m^{-3} m^{-3})	0.07	0.23	0.23	0.41	0.08
-10 kPa (m^{-3} m^{-3})	0.07	0.21	0.21	0.37	0.07
-33 kPa (m^{-3} m^{-3})	0.06	0.17	0.18	0.33	0.07
-100 kPa (m^{-3} m^{-3})	0.05	0.15	0.15	0.31	0.06
-500 kPa (m^{-3} m^{-3})	0.04	0.12	0.13	0.30	0.05
-1500 kPa (m^{-3} m^{-3})	0.04	0.11	0.11	0.29	0.05
Ks (mm h^{-1})	0.60	26.15	61.02	461.53	79.02
$\log_{10}$ (Ks (mm h^{-1}))	-0.22	1.37	1.40	2.51	0.58

[a]Ks and $\log_{10}$Ks have n = 566 in the Basic dataset. Min = minimum value, Max = maximum value, SD = standard deviation, Bd = bulk density, Wdc = water-dispersible clay, OC = organic carbon, CEC = cation exchange capacity, BS = base saturation, "n" kPa = water content at "n" matric potential.

Each dataset was pre-processed to identify possible outliers, redundant variables and missing data that could potentially disrupt or hinder the modeling process. In the Basic dataset two outliers were found and arbitrarily removed from the database. These outliers were: Bulk density = 2.441 Mg m^{-3}, which is highly unlikely for non-gravel soils under agricultural use; 33 kPa = 1.392 m^{-3} m^{-3}, which is physically impossible. Multicollinearity was evaluated through Variable Inflation Factor (VIF), adopting VIF = 10 as the upper limit to detect multicollinearity. VIF values exceeded the limit only for particle size data (sand, silt and clay). To handle this, the silt/clay ratio was used instead of the raw values of silt and clay in all models of all datasets. This ratio was chosen not only to avoid multicollinearity but also because the silt/clay ratio holds important information about the weathering degree of tropical soils (Van Wambeke, 1962), which influences soil hydraulic behavior (Silva et al., 2022). Missing values of hydraulic conductivity (Ks) were found in the Basic dataset and were removed. Given that, the PTFs for Ks prediction using the Basic dataset had 566 samples in total, instead of 669 as used for modeling water retention. To fit the PTFs, the Random Forest (RF), Support Vector Machine (SVM), k-Nearest Neighbors (KNN), eXtreme Gradient Boosting (XGBoost), Multiple Linear Regression (MLR) and Null model (mean value) were tested.

2.2.2 Machine learning models overview

Random Forest (RF) an update of the Decision Trees method, introduced by (Breiman, 2001) in order to improve the accuracy of predictions. The method consists of ensemble techniques by building several decision trees with the random selection of a combination of input variables (Van Looy et al., 2017). Then each tree generates a prediction, and these predictions are averaged to obtain the forest's prediction (Araya and Ghezzehei, 2019; Kuhn and Johnson, 2013). Models based on the Random Forest method are commonly robust to errors and outliers (Gunarathna et al., 2019; Van Looy et al., 2017).

SVM method for regression (Drucker et al., 1997) is regarded as one of the most robust, flexible, and efficient algorithms (Kuhn and Johnson, 2013). Moreover, it is one of the most frequently used for building PTFs (Van Looy et al., 2017). The method is based on mapping input variables in a high-dimensional space using mapping functions (kernel functions) (Araya and Ghezzehei, 2019). The algorithm then finds a hyperplane that best represents the relationships between input variables and their corresponding output values, generating continuous estimations.

KNN method is one of the most used and well-known algorithms in the scientific community (Wu et al., 2008). The method is a nonparametric classifier, introduced by Cover and Hart, 1967) and can be used to solve classification and regression problems. The method is based on instances, and the prediction of a new instance is made by averaging the values of the nearest neighbors "k" from the training data (Cover and Hart, 1967). The quantification of the distance between the objects and the determination of the nearest neighbors of the labeled data is performed by calculating the Euclidean distance between the objects in the space of the predictor parameters (Araya and Ghezzehei, 2019; Yamaç et al., 2020).

XGBoost, is a machine learning algorithm based on the theorem of "strength of week learnability" by Schapire (1990). XGBoost is a scalable tree boosting that supports both classification and regression predictive modeling (Chen and Guestrin, 2016). XGBoost is an ensemble technique which combines several models and is able to solve real-world problems with limited resources (Chen and Guestrin, 2016) and noisy data (Friedman, 2001). XGBoost starts with a single model and generates new models based on previous trials, aiming to reduce prediction errors. This process is repeated several times and once no further improvements are observed, a final model is provided (Choudhury et al., 2023).

A null model is defined as a model which has only one parameter (i.e., the intercept) and no predictors (Crawley, 2013). In summary, the null model simply predicts using the mean value (James et al., 2013). Its application is mainly motivated by the need to assess whether using a complex model is parsimonious and worthwhile (Silva et al., 2023b). For instance, if intricate models fail to demonstrate reasonable better predictions in comparison to the null model, their utilization is not justified. Silva et al. (2023b) showed that popular PTFs performed worse than the null model in some scenarios.

2.2.3 Training, testing and evaluating performance

The repeat holdout method was applied to each dataset for partitioning them randomly into training and testing subsets. This method was applied 100 times aiming to avoid the potential split of biased training and testing subsets. Training data corresponded to 75% of each dataset and test data to 25%. For hyperparameter fitting a k-fold cross validation was performed ($k = 10$, $n = 3$).

The PTFs performance analysis was carried out using the independent test subset (25% of each dataset), which was separated before the beginning of training process. Root mean square error (RMSE), Mean absolute error (MAE), Coefficient of determination (R^2), Mean bias error (MBE) and Lin's concordance correlation coefficient (LCCC) were used to compare the models. RMSE was chosen as the metric for selecting the best models for each prediction.

Ranks of variables importance in predictions made by MLR and RF were determined using the "varImp" function of *caret* package in R software. For predictions by SVM and KNN, Recursive Feature Elimination with k-fold cross validation ($n = 10$) was performed and the score of each variable was considered as the importance. For XGBoost predictions, "xgb.importance" function was used, which calculates feature importance based on the gain, i.e., the improvement in accuracy brought by each predictor.

2.3 Domain of application

The domain of application of a PTF is defined based on the similarities in the distribution of data between that used to fit the models and a from broader database. The procedures followed the methodology proposed by Tranter et al. (2009) and insights raised by Dobarco et al. (2019). In general terms, the methodology proposes using metric distances such as Euclidean distance and Mahalanobis distance to quantify the similarity between the dataset used for PTFs calibration and an independent dataset with potential predictors. In this sense, a quantitative criterion of metric distance was adopted, investigating whether the data distribution of the independent set is represented by the PTFs calibration set.

Table 2 shows descriptive statistics of the independent dataset used to determine the PTFs for the domain of application. Our independent dataset for this purpose was the Socioeconomic and Ecological Zoning of the Mato Grosso state (ZSEE/MT), which contains information about 566 soil profiles across the state of Mato Grosso. In this study, the Mahalanobis distances between the centroids of predictors of the calibration dataset and each point of the independent dataset were used to determine the overall similarity. This process was carried out three times, i.e., comparing each calibration dataset (Basic, Basic + OC and Basic + OC + chemical) with the independent test (ZSEE/MT). Thus, three different domains of application were obtained based on the predictors.

Table 2. Descriptive statistics of the dataset extracted from the Socioeconomic and Ecological Zoning of Mato Grosso state.

ZSEE-MT dataset (n = 566)	Min	Median	Mean	Max	SD
Sand (%)	3.00	62.00	60.82	97.00	21.29
Silt (%)	1.00	16.00	17.77	58.00	12.02
Clay (%)	1.00	20.00	21.33	62.00	12.68
Bd (Mg m^{-3})	0.65	1.27	1.25	1.61	0.14
Wdc (g kg^{-1})	0.00	0.00	1.88	34.00	3.61
OC (g kg^{-1})	0.70	10.60	13.10	73.80	9.79
pH (H_2O)	3.70	4.80	4.90	7.30	0.60
CEC (cmolc dm^{-3})	0.05	6.27	7.13	37.04	3.81
BS (%)	2.00	17.00	24.00	100.00	18.94

Min = minimum value, Max = maximum value, SD = standard deviation, Bd = bulk density, Wdc = water-dispersible clay, OC = organic carbon, CEC = cation exchange capacity, BS = base saturation.

Before calculating the Mahalanobis distances, soil particle size data was transformed in all datasets applying the additive log-ratio transformation (Aitchison, 1982, Dobarco et al., 2019). This process is recommended (Dobarco et al., 2019; Lark and Bishop, 2007) as soil particle size data are compositional, i.e., the amount of sand, silt and clay are positive and sum to a constant (Aitchison, 1982). Therefore, clay and silt contents were transformed as following:

$clay_t = \ln(clay/sand)$

$silt_t = \ln(silt/sand)$

After transforming the variables, the covariance matrix for each calibration dataset was calculated along with the means for each variable. The means of each variable correspond to the centroids of each calibration dataset. Next, Mahalanobis distance was calculated between each point of the ZSEE/MT dataset and the centroid of the calibration dataset, for all variables. To determine whether each datapoint is within or outside the domain of application, the cutoff limit was set as the 97.5% percentile of the cumulative chi-squared distribution of the calibration dataset, with degrees of freedom corresponding to the number of independent variables (Tranter et al., 2009; Dobarco et al., 2019). When the data point exceeded the cutoff limit, it was considered out of the domain of application, or dissimilar to calibration dataset.

3. Results and Discussion

3.1 Prediction of soil water retention

Models with best performance for predicting water retention of soils from Mato Grosso are shown in Figure 1. Overall, the tested models showed gains in accuracy and are worth further consideration, as they performed better than the null model (mean value). To check the complete results comparing all models for each prediction target, the reader is referred to the supplementary material. Considering the best models shown in Fig. 1, all models showed low bias (MBE < |0.003| m^3 m^{-3}), indicating that overestimation or underestimation of the observed values were not a problem in our study. Among the most accurate models, only XGBoost did not produce the best performance in any case. All others (i.e., MLR, RF, SVM and KNN) had the best prediction performance at least twice, while RF stood out with the highest performance in 16 out of the 24 best models for soil water retention prediction. KNN was the best model 4 times followed by SVM and MLR twice each.

The best results provided by RF rely on its robustness (van Looy et al., 2017) and high capacity for generalization (Rastgou et al., 2020), which has been reported by several studies addressing PTFs for SHP (e.g., Beniaich et al., 2023; Gunarathna et al., 2019; Rezaei et al., 2023; Veloso et al., 2022). Nevertheless, RF was outperformed by other algorithms in specific cases. SVM performed the best in generalizing data of 0 kPa (RMSE = 0.048 m^3 m^{-3}, MAE = 0.033 m^3 m^{-3}, LCCC = 0.743) and 4 kPa (RMSE = 0.047 m^3 m^{-3}, MAE = 0.033 m^3 m^{-3}, LCCC = 0.622) using the Basic predictor set, although it was outperformed by RF when Basic + OC and Basic + OC + chemical predictor sets were used. KNN presented the best results for predicting drier potentials of soil water retention (i.e., 33 kPa, 100 kPa, 500 kPa and 1500 kPa) using the Basic + OC predictor set, possibly having more ability to handle organic carbon as a feature than the other algorithms.

MLR was benefited by the addition of features, showing the best performance for predicting 10 kPa with data from Basic + OC (RMSE = 0.039 m^3 m^{-3}, MAE = 0.030 m^3 m^{-3}, LCCC = 0. 849) and Basic + OC + chemical (RMSE = 0.032 m^3 m^{-3}, MAE = 0.026 m^3 m^{-3}, LCCC = 0.889) predictor sets. The best performance of MLR for water content predictions at

10 kPa is particularly important as this value is usually adopted as field capacity for Brazilian soils (Brito et al., 2011; Reichardt, 1988; Reichert et al., 2020; Silva et al., 2014) and MLR is a quite easy model to apply. MLR provides an equation as output which makes the application of a PTF more accessible. These results agree with the "No Free Lunch" Theorem (Wolpert, 1966) which states that a model's performance is highly conditioned by the modeling problem, and choosing a specific algorithm *a priori* is not recommended (Kuhn and Johnson, 2013).

Regardless of algorithms, prediction of soil water retention at wetter to mid ranges of matric potentials (i.e., 0 kPa, -4 kPa, -6 kPa, -10 kPa, -33 kPa and -100 kPa) was systematically improved using additional features to the Basic predictor set. Water retention results at these matric potentials followed the exact same pattern of accuracy improvement and error metrics decrease (RMSE and MAE) according to the predictor set. Models had their worst performance using the Basic predictor set, had an intermediate performance using Basic + OC and achieved their best performance using Basic + OC + chemical. Conversely, this pattern changed for the drier water retention potentials (i.e., -500 kPa and -1500 kPa). Prediction of water retention at -500 kPa had the lowest error metrics using Basic + OC + chemical (RMSE = 0.026 m^3 m^{-3}, MAE = 0.021 m^3 m^{-3}), however, Basic predictor set had intermediate error (RMSE = 0.031 m^3 m^{-3}, MAE = 0.023 m^3 m^{-3}) and the best agreement between predicted and observed values (LCCC = 0.901). In contrast, Basic + OC presented the highest error and poorest agreement between predicted and observed values (RMSE = 0.035 m^3 m^{-3}, MAE = 0.026 m^3 m^{-3}, LCCC = 0.783). For -1500 kPa, prediction using Basic predictor set was the most accurate (RMSE = 0.027 m^3 m^{-3}, MAE = 0.020 m^3 m^{-3}), but inferior performance was found using Basic + OC (RMSE = 0.037 m^3 m^{-3}, MAE = 0.028 m^3 m^{-3}).

Differences in predicting the wet and dry range of soil water retention curve are well reported (Dashtaki et al., 2010; Haghverdi et al., 2012; Pachepsky et al., 1996; Tomasella et al., 2003). Usually, the wetter segment is more complicated to predict than the drier (Kotlar et al., 2020; Silva et al., 2023b; Wassar et al., 2016) which explains why the prediction of wetter points of soil water retention were benefited by the addition of different predictors (i.e., using Basic + OC and Basic + OC + chemical predictor sets) whereas drier points were not. Vereecken et al. (2010) highlighted the importance of building PTFs for different ranges of water retention, as different models can capture the most important features and generalize the patterns separately for each range of water retention. Moreover, these differences could be somewhat related to the Hughes phenomenon (Hughes, 1968), which states that adding features increases accuracy until an optimal number is achieved but can degrade the performance by introducing more noise from the optimal number onwards. Therefore, using organic carbon and chemical properties to predict drier ranges of water retention potentially introduced noise to the process and diminished the model generalization capacity. It may be related to the fact that the water retention pattern at drier points is mainly governed by soil texture (Dexter, 2004; Manrique et al., 1991; Nguyen et al., 2014) so that the optimal prediction was achieved using only basic variables, mainly due to particle size information.

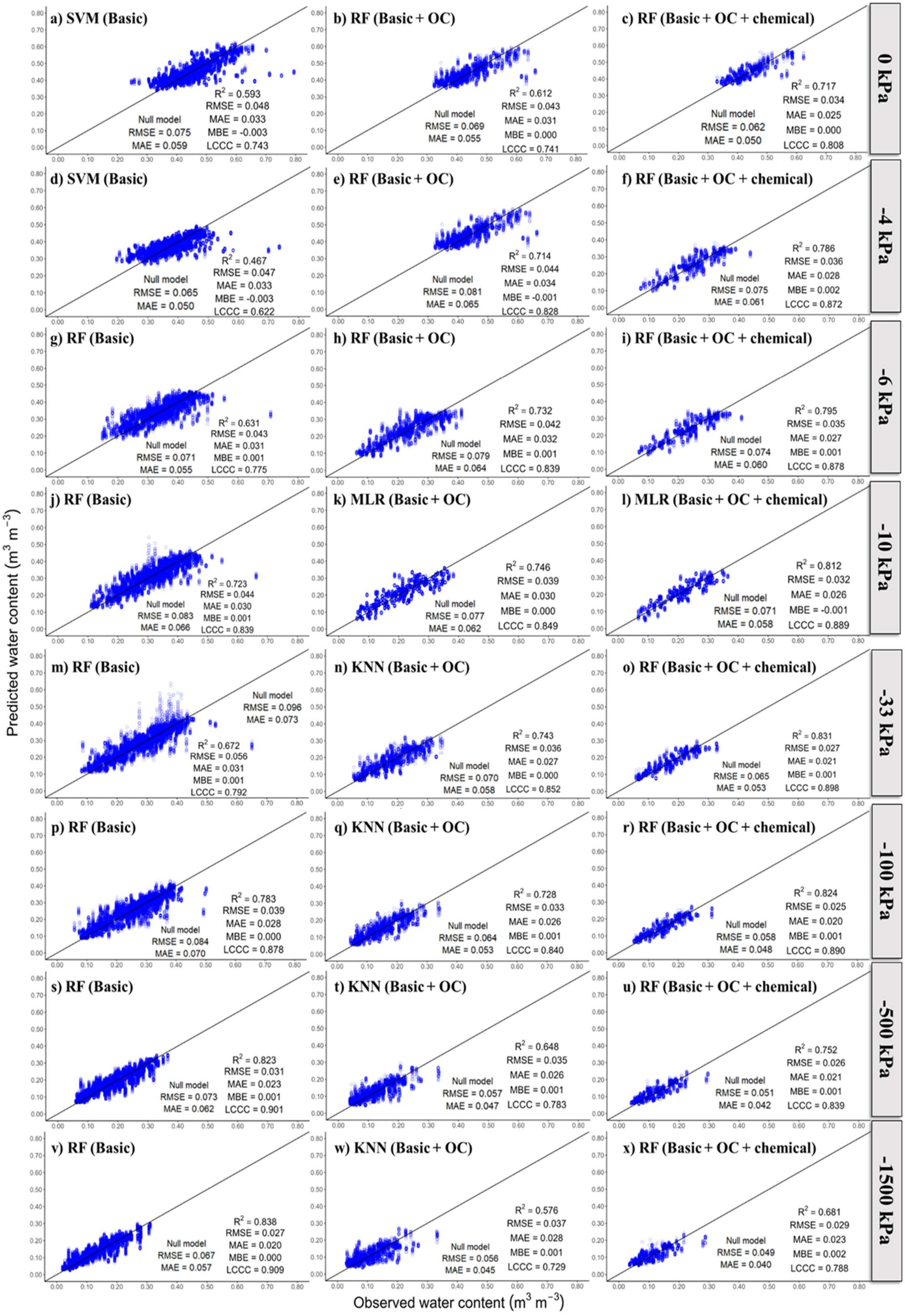
a) SVM (Basic)
R² = 0.593
RMSE = 0.048
MAE = 0.033
MBE = -0.003
LCCC = 0.743
Null model
RMSE = 0.075
MAE = 0.059
b) RF (Basic + OC)
R² = 0.612
RMSE = 0.043
MAE = 0.031
MBE = 0.000
LCCC = 0.741
Null model
RMSE = 0.069
MAE = 0.055
c) RF (Basic + OC + chemical)
R² = 0.717
RMSE = 0.034
MAE = 0.025
MBE = 0.000
LCCC = 0.808
Null model
RMSE = 0.062
MAE = 0.050
0 kPa
d) SVM (Basic)
R² = 0.467
RMSE = 0.047
MAE = 0.033
MBE = -0.003
LCCC = 0.622
Null model
RMSE = 0.065
MAE = 0.050
e) RF (Basic + OC)
R² = 0.714
RMSE = 0.044
MAE = 0.034
MBE = -0.001
LCCC = 0.828
Null model
RMSE = 0.081
MAE = 0.065
f) RF (Basic + OC + chemical)
R² = 0.786
RMSE = 0.036
MAE = 0.028
MBE = 0.002
LCCC = 0.872
Null model
RMSE = 0.075
MAE = 0.061
-4 kPa
g) RF (Basic)
R² = 0.631
RMSE = 0.043
MAE = 0.031
MBE = 0.001
LCCC = 0.775
Null model
RMSE = 0.071
MAE = 0.055
h) RF (Basic + OC)
R² = 0.732
RMSE = 0.042
MAE = 0.032
MBE = 0.001
LCCC = 0.839
Null model
RMSE = 0.079
MAE = 0.064
i) RF (Basic + OC + chemical)
R² = 0.795
RMSE = 0.035
MAE = 0.027
MBE = 0.001
LCCC = 0.878
Null model
RMSE = 0.074
MAE = 0.060
-6 kPa
j) RF (Basic)
R² = 0.723
RMSE = 0.044
MAE = 0.030
MBE = 0.001
LCCC = 0.839
Null model
RMSE = 0.083
MAE = 0.066
k) MLR (Basic + OC)
R² = 0.746
RMSE = 0.039
MAE = 0.030
MBE = 0.000
LCCC = 0.849
Null model
RMSE = 0.077
MAE = 0.062
l) MLR (Basic + OC + chemical)
R² = 0.812
RMSE = 0.032
MAE = 0.026
MBE = -0.001
LCCC = 0.889
Null model
RMSE = 0.071
MAE = 0.058
-10 kPa
m) RF (Basic)
Null model
RMSE = 0.096
MAE = 0.073
R² = 0.672
RMSE = 0.056
MAE = 0.031
MBE = 0.001
LCCC = 0.792
n) KNN (Basic + OC)
R² = 0.743
RMSE = 0.036
MAE = 0.027
MBE = 0.000
LCCC = 0.852
Null model
RMSE = 0.070
MAE = 0.058
o) RF (Basic + OC + chemical)
R² = 0.831
RMSE = 0.027
MAE = 0.021
MBE = 0.001
LCCC = 0.898
Null model
RMSE = 0.065
MAE = 0.053
-33 kPa
p) RF (Basic)
R² = 0.783
RMSE = 0.039
MAE = 0.028
MBE = 0.000
LCCC = 0.878
Null model
RMSE = 0.084
MAE = 0.070
q) KNN (Basic + OC)
R² = 0.728
RMSE = 0.033
MAE = 0.026
MBE = 0.001
LCCC = 0.840
Null model
RMSE = 0.064
MAE = 0.053
r) RF (Basic + OC + chemical)
R² = 0.824
RMSE = 0.025
MAE = 0.020
MBE = 0.001
LCCC = 0.890
Null model
RMSE = 0.058
MAE = 0.048
-100 kPa
s) RF (Basic)
R² = 0.823
RMSE = 0.031
MAE = 0.023
MBE = 0.001
LCCC = 0.901
Null model
RMSE = 0.073
MAE = 0.062
t) KNN (Basic + OC)
R² = 0.648
RMSE = 0.035
MAE = 0.026
MBE = 0.001
LCCC = 0.783
Null model
RMSE = 0.057
MAE = 0.047
u) RF (Basic + OC + chemical)
R² = 0.752
RMSE = 0.026
MAE = 0.021
MBE = 0.001
LCCC = 0.839
Null model
RMSE = 0.051
MAE = 0.042
-500 kPa
v) RF (Basic)
R² = 0.838
RMSE = 0.027
MAE = 0.020
MBE = 0.000
LCCC = 0.909
Null model
RMSE = 0.067
MAE = 0.057
w) KNN (Basic + OC)
R² = 0.576
RMSE = 0.037
MAE = 0.028
MBE = 0.001
LCCC = 0.729
Null model
RMSE = 0.056
MAE = 0.045
x) RF (Basic + OC + chemical)
R² = 0.681
RMSE = 0.029
MAE = 0.023
MBE = 0.002
LCCC = 0.788
Null model
RMSE = 0.049
MAE = 0.040
-1500 kPa
Predicted water content (m³ m⁻³)
Observed water content (m³ m⁻³)
0.00
0.10
0.20
0.30
0.40
0.50
0.60
0.70
0.80

Figure 1. Models with best performance for predicting water retention of soils from Mato Grosso state, Brazil, for each predictor set. SVM - a) and d); RF - b), c), e), f), g), h), i), j), m), o), p), r), s), u), v), x); MLR – k), l); KNN – n), q), t), w); "n"kPa = soil water content at "n" matric potential. Basic = basic predictor set with sand, silt, clay, bulk density, and water-dispersible clay. Basic + OC = basic predictor set + organic carbon. Basic + OC + chemical = basic predictor set + organic carbon + pH in water + cation exchange capacity + base saturation. Results represent the average for 100 rounds of dataset splitting, training, and testing.

Ranking the importance of variables for soil water retention prediction is presented in Table 3 where "1º" refers to the most important variable. The raw scores of importance were not shown as each model had a different method to determine the variables importance (section 2.2.3), therefore, the raw scores cannot be fairly compared. However, the reader is referred to the supplementary material to check raw scores for each variable in each model developed. Sand and Wdc were the two most frequent variables among the 1º and 2º positions considering all predictor sets regardless the water retention range. Bd had its highest importance for higher matric potentials (i.e., higher water contents) and Silt/clay did not exhibit relevant importance. The Bd importance in predicting soil water retention wetter points is explained by the forces controlling this process, which are mainly capillary forces (Jensen et al., 2015), strongly associated with soil structure and pore size distribution (Carducci et al., 2013; Dexter, 2004). Conversely, the soil water retention drier range is mainly governed by adsorption forces, which are related to soil texture (Manrique et al., 1991; Nguyen et al., 2014), explaining why sand and Wdc were prevalent in the prediction of lower matric potentials (i.e., lower water contents).

The importance of water-dispersible clay (Wdc) as a predictor of soil water retention warrants further discussion as this variable is not usually used as a feature in PTFs for such a purpose. For example, in the review carried out by Silva et al. (2023b), which assessed more than 100 articles on PTFs development, Wdc did not appear among the variables used as predictors for water retention, neither the flocculation degree, which is calculated from the Wdc value. However, in few studies, Wdc has been found to be a significant predictor of water retention (Amorim et al., 2022; Reichert et al., 2020). Wdc is strongly related to soil aggregation and soil structural quality (Brubaker et al., 1992; Cañasveras et al., 2010) reflecting on soil water dynamics (Rabot et al., 2018). Wdc may be particularly important concerning our database, which has data from agricultural soils where no-tillage is prevalent and practices such as surface liming are often applied. Surface liming can increase Wdc and promote the migration of dispersed clay, causing pore obstruction and increased bulk density (Nunes et al., 2017), consequently influencing water retention. Furthermore, Wdc is as easily obtained in laboratory analysis as soil texture, therefore it is suitable and well aligned with the PTF principle of using easy-to-measure data.

For Basic + OC and Basic + OC + chemical datasets, organic carbon had an intermediate importance for the prediction of water retention. Although organic carbon influence in soil water retention is well reported (Panagea et al., 2021; Rawls et al., 2003; Zhao et al., 2015), tropical soils may not show increases in organic carbon content even under conservation cropping systems (Silva et al., 2021). Therefore, organic carbon, despite providing improvements in accuracy, was not a significant feature in depicting water retention patterns for the developed models, mainly because this property had a maximum content of 2 g kg^{-1} and low variability (SD ~ 2 g kg^{-1}), as shown in Table 1. Chemical properties have been used in several studies as alternative predictors of water retention (e.g., Botula et al., 2012; Bruand 2004; Hodnett and Tomasella 2002; Tóth et al., 2015). Results in Table 3 show that chemical properties were somewhat important (Botula et al., 2012; Bruand, 2004; Hodnett and Tomasella, 2002; Tóth et al., 2015) in recognizing water retention patterns, although not major. CEC had, in general, the highest ranking compared to pH and BS. CEC has already been proven as an important predictor of soil water retention by other authors (Botula et al., 2013; Hodnett and Tomasella, 2002). These results reinforce the possible gains in accuracy using alternative variables – e.g., chemical properties - to predict water retention via PTFs, which has been a barely applied approach (Silva et al., 2023b).

Table 3. Ranking of importance of variables used as predictors for water retention of soils from Mato Grosso.

Variable importance ranking	**Basic predictor set**							
	0 kPa	-4 kPa	-6 kPa	-10 kPa	-33 kPa	-100 kPa	-500 kPa	-1500 kPa
1º	Bd	Sand	Sand	Sand	Sand	Sand	Sand	Sand
2º	Sand	Bd	Wdc	Wdc	Bd	Bd	Wdc	Wdc
3º	Wdc	Silt/clay	Bd	Bd	Silt/clay	Silt/clay	Silt/clay	Silt/clay
4º	Silt/clay	Wdc	Silt/clay	Silt/clay	Wdc	Wdc	Bd	Bd

Variable importance ranking	**Basic + OC**							
	0 kPa	-4 kPa	-6 kPa	-10 kPa	-33 kPa	-100 kPa	-500 kPa	-1500 kPa
1º	Bd	Sand	Sand	Sand	Wdc	Wdc	Wdc	Wdc
2º	Sand	Wdc	Wdc	Bd	Sand	Sand	Sand	Sand
3º	OC	OC	OC	Silt/clay	OC	OC	OC	Silt/clay
4º	Wdc	Bd	Bd	OC	Silt/clay	Silt/clay	Silt/clay	OC
5º	Silt/clay	Silt/clay	Silt/clay	Wdc	Bd	Bd	Bd	Bd

Variable importance ranking	**Basic + OC + chemical**							
	0 kPa	-4 kPa	-6 kPa	-10 kPa	-33 kPa	-100 kPa	-500 kPa	-1500 kPa
1º	Bd	Wdc	Sand	Bd	Wdc	Wdc	Wdc	Wdc
2º	Wdc	Sand	Wdc	Sand	Sand	Sand	Sand	Sand
3º	OC	CEC	CEC	Wdc	OC	OC	OC	OC
4º	Sand	OC	OC	BS	CEC	CEC	CEC	Silt/clay
5º	BS	BS	Bd	pH	Silt/clay	Bd	Bd	Bd
6º	pH	Bd	BS	Silt/clay	Bd	Silt/clay	Silt/clay	pH
7º	Silt/clay	Silt/clay	Silt/clay	OC	BS	BS	pH	CEC
8º	CEC	pH	pH	CEC	pH	pH	BS	BS

Basic = basic predictor set with sand, silt, clay, bulk density and water-dispersible clay. Basic + OC = basic predictor set + organic carbon. Basic + OC + chemical = basic predictor set + organic carbon + pH in water + cation exchange capacity + base saturation. “n”kPa = soil water content at “n” matric potential, Bd = bulk density, Wdc = water-dispersible clay, OC = organic carbon, CEC = cation exchange capacity, BS = base saturation, pH = pH (H_2O). Results consider 100 rounds of dataset splitting, training and testing.

1.1 Prediction of soil hydraulic conductivity

Models with the best performances to predict the logarithm of soil hydraulic conductivity (log$_{10}$Ks) using the three different datasets are shown in Figure 2. RF had the best performance for predicting log$_{10}$Ks using the Basic predictor set whereas KNN performed best with Basic + OC and Basic + OC + chemical predictor sets. Similarly to what occurred in wetter ranges of soil water retention, the best models for hydraulic conductivity also showed a progressive improvement in accuracy according to the addition of features to the predictor set. Considering the best models, RF using Basic predictor set showed the lowest performance (RMSE = 0.633, MAE = 0.488 and LCCC = 0.494), KNN using Basic + OC had intermediate performance (RMSE = 0.497, MAE = 0.377, LCCC = 0.639) and KNN using Basic + OC + chemical had the best performance (RMSE = 0.403, MAE = 0.314 and LCCC = 0.711). KNN with Basic + OC + chemical had the highest bias towards underestimating $\log_{10}$Ks values, though the effects were slight (MBE = -0.082).

Our results agree with the statement by Zhang and Schaap (2019) that despite consistent advances in machine learning, predictions of soil hydraulic conductivity remain modest. Nevertheless, all models presented higher accuracy than the null model (please see supplementary material), justifying their use in potential future applications. As well, an improvement in accuracy was achieved by using additional features in the predictor sets. Silva et al. (2023b) highlighted the importance of using alternative variables to predict SHP as most PTFs have been built based on soil physical properties. Likewise, Ottoni et al. (2019) pointed out that using predictors beyond basic features (i.e., particle size and bulk density) may help PTFs towards a better prediction of soil hydraulic conductivity.

In accordance with our results, Veloso et al. (2022) observed a progressive increase in accuracy according to the addition of features to their predictor sets. In our case, however, each predictor set has a different size and the number of records in a dataset can make a difference in the final prediction. For instance, larger datasets likely have more noise and less clean data, making the assemble of homogeneous and consistent databases a crucial concern for improving PTFs. Moreover, as hydraulic conductivity usually has high variability (Santos et al., 2021), it is likely that larger datasets of hydraulic conductivity will present more considerable variability, which may jeopardize the training process leading to a less capacity for pattern recognition. The relationship between variability and the dataset size was confirmed in our study. The largest dataset (Basic) showed a standard deviation (SD) of 0.793 for $\log_{10}$Ks values, whereas Basic + OC showed SD = 0.654 and Basic + OC + chemical had SD = 0.579. Nonetheless, while the most extensive dataset may have more noise, it also has more data to train and test, increasing its chance to perform well, relative to limited datasets. Therefore, it is reasonable to assume that the addition of variables to models' training improved their accuracy and it should be considered in further studies of PTF development. Overall, PTFs for hydraulic conductivity showed gains in accuracy compared to the null model and all models were able to capture data variation beyond the inherent data variability, as all models presented RMSE < SD for all predictor sets.

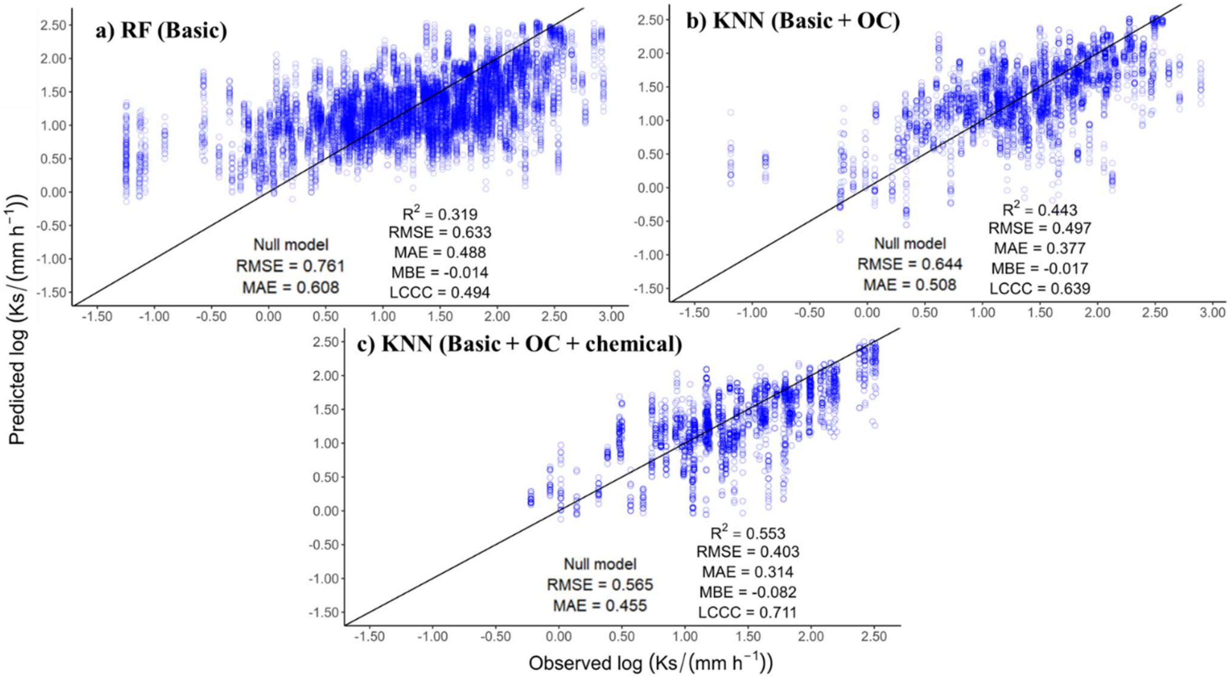

Figure 2. Models with best performance for predicting hydraulic conductivity of soils from Mato Grosso state, Brazil, for each predictor set. RF - a); KNN – b) and c). Basic = basic predictor set with sand, silt, clay, bulk density and water-dispersible clay. Basic + OC = basic predictor set + organic carbon. Basic + OC + chemical = basic predictor set + organic carbon + pH in water + cation exchange capacity + base saturation. Results represent the average for 100 rounds of dataset splitting, training and testing.

Similar to what was done for soil water retention prediction, the importance of variables for hydraulic conductivity PTF development was ranked (Table 4). Sand was the most important variable in all cases and Bd was the 2º most important variable for the Basic and Basic + OC + chemical datasets. Silt/clay was the least important variable for the Basic and Basic + OC + chemical, while it was the least but one for the Basic + OC dataset. The irrelevance of silt/clay in depicting hydraulic conductivity patterns may be related to the distinct hydraulic behavior of clayey Brazilian soils (Pessoa and Libardi, 2022).

A considerable portion of Brazilian soils (i.e., the main soil class in the country, Ferralsols) have high macroporosity and high hydraulic conductivity despite high clayey and very clayey textures (Moura et al., 2021; Ferreira et al., 1999; Silva et al., 2022, 2021). This is attributed to the mineralogy of Brazilian Ferralsols, which leads to the formation of a very stable granular structure (Silva et al., 2022), also known as pseudo-sand (Martinez and Souza, 2020). This granular structure has a great influence on the hydrological behavior of Ferralsols leading them to behave similarly to sandy soils in terms of water transmission: they have high hydraulic conductivity values (Pessoa and Libardi, 2022) resulting from the high proportion of large pores (Tomasella et al., 2000). Therefore, as Ferralsols are a considerable part of our datasets, silt/clay variability might not have sufficient variation to aid models in recognizing hydraulic conductivity patterns, even though this ratio is an important indicator of soil weathering degree. This result reinforces what was highlighted by Ottoni et al. (2019) and Silva et al. (2023) concerning the inability of current models to handle the relationship between hydraulic conductivity and clay in tropical soils.

Similar to what was observed in soil water retention predictions, organic carbon did not appear among the most important variables for hydraulic conductivity, but it may be associated with the uniformly low rates of carbon in our datasets (Max ~ 11 g kg^{-1}), which may not have provided enough variation to train the models. Nevertheless, the interplay between carbon and other variables seems to have had an effect, as hydraulic conductivity prediction was consistently improved when this property was used (RMSE = 0.633 to RMSE = 0.497, Fig. 3). Soil chemical properties had a considerable importance in predicting soil hydraulic conductivity as the three chemical features used (i.e., BS, CEC and pH) were the most important variables after sand and Bd.

Soil hydraulic conductivity is an extremely sensitive property that is highly influenced by several other soil attributes (Araya and Ghezzehei, 2019; Zhang and Schaap, 2019). Hence, hydraulic conductivity is susceptible to large variations triggered by small variations in other properties (Morgan et al., 1993). Hydraulic conductivity values can be highly sensitive to variations of chemical properties (Candemir and Gülser, 2012), which explains the importance of chemical features for the predictions. Under these results, Tóth et al. (2015) found that pH and CEC improved the reliability of hydraulic conductivity predictions. Additionally, Hodnett and Tomasella (2002) found that pH and CEC were important for soil water retention, which might reflect in soil hydraulic conductivity if further investigated. Bruand (2004) highlights the potential of chemical properties as predictors of PTFs for SHP, as chemical features are directly related to pedogenesis.

Table 4. Ranking of importance of variables used as predictors for hydraulic conductivity of soils from Mato Grosso.

Variable importance ranking	**Basic**	**Basic + OC**	**Basic + OC + chemical**
	$\log_{10}$ (Ks (mm h^{-1}))	$\log_{10}$ (Ks (mm h^{-1}))	$\log_{10}$ (Ks (mm h^{-1}))
1º	Sand	Sand	Sand
2º	Bd	Wdc	Bd
3º	Wdc	Bd	BS
4º	Silt/clay	Silt/clay	CEC
5º	-	OC	pH
6º	-	-	Wdc
7º	-	-	OC
8º	-	-	Silt/clay

Basic = basic predictor set with sand, silt, clay, bulk density and water-dispersible clay. Basic + OC = basic predictor set + organic carbon. Basic + OC + chemical = basic predictor set + organic carbon + pH in water + cation exchange capacity + base saturation. Bd = bulk density, Wdc = water-dispersible clay, OC = organic carbon, CEC = cation exchange capacity, BS = base saturation, pH = pH (H_2O).

3.2 Domain of application

The domain for application of models developed in this study is shown in Figure 3. Results show that PTFs are well calibrated to be reliably applied using predictors from the ZSEE/MT database in 91% of the state using variables of the Basic predictor set (Fig. 3a), in 40% of the state using variables of Basic + OC and in 18% of the state using variables of Basic + OC + chemical dataset. For the Basic predictor set, 52 out of 566 data points from the ZSEE/MT laid outside the PTFs' domain of application, whereas 342 data points for Basic + OC and 461 data points using the Basic + OC + chemical dataset. Overall, results indicate that PTFs are well calibrated to be used with data from ZSEE/MT for basic predictors (i.e., sand, silt, clay, Wdc and Bd), whereas intermediately calibrated to be applied using organic carbon data from ZSEE/MT and poorly calibrated to be applied using chemical properties (i.e., pH, CEC and BS) from the ZSEE/MT database.

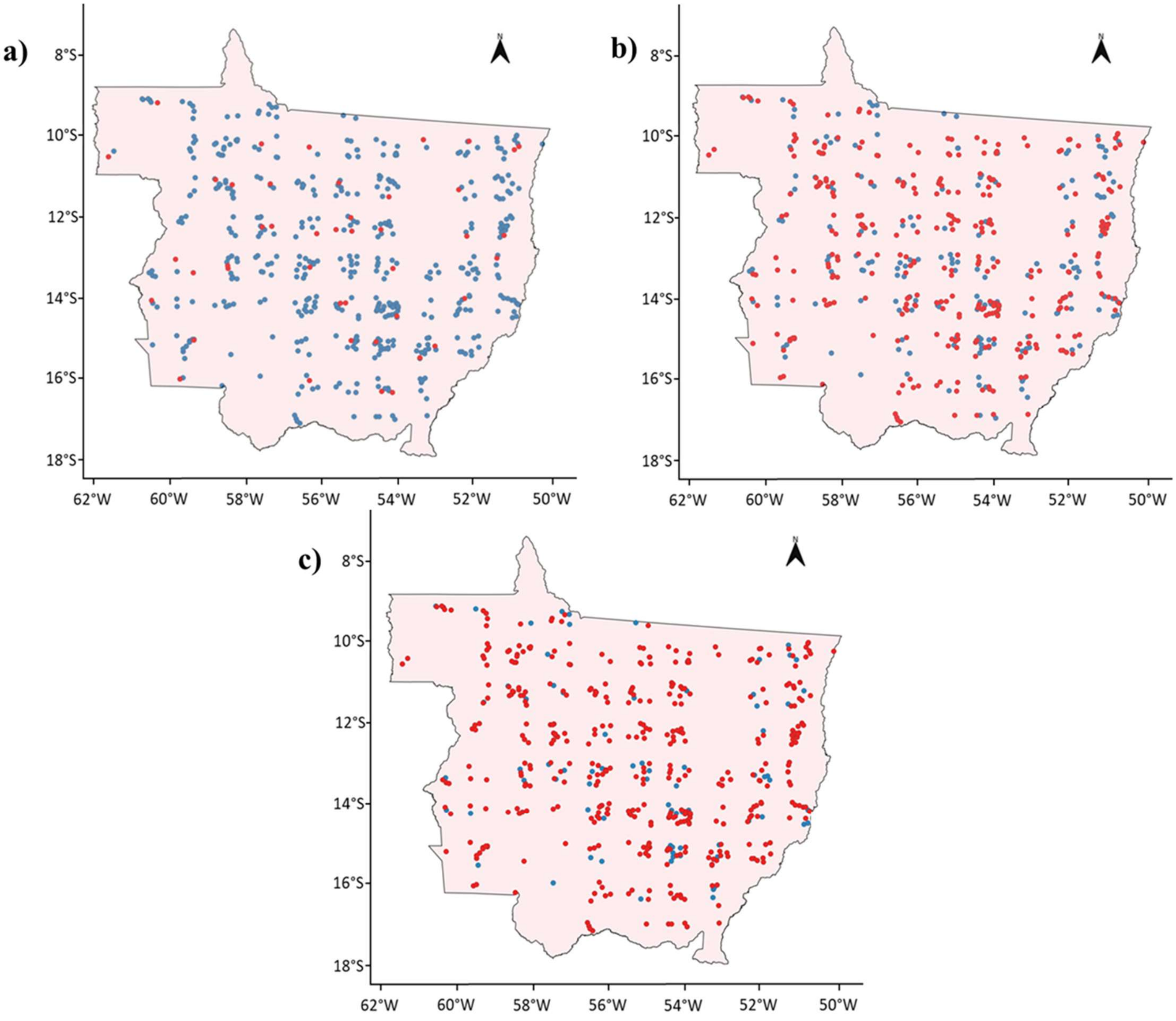

Figure 3. Domain of application for PTFs using the database of Socioeconomic and Ecological Zoning of the Mato Grosso state (ZSEE/MT). Blue dots are within the domain of application and red dots are outside the domain of application. a) refer to Basic = basic predictor set with sand, silt, clay, bulk density and water-dispersible clay. b) refer to Basic + OC = basic predictor set + organic carbon. c) refer to Basic + OC + chemical = basic predictor set + organic carbon + pH in water + cation exchange capacity + base saturation.

The reduced applicability of PTFs' within the state using organic carbon and chemical properties can be attributed to discrepancies between data from the calibration datasets and the data from the ZSEE/MT database. For organic carbon, our datasets for PTF calibration have a minimum value of ~2 g kg^{-1} and a maximum value of ~11 g kg^{-1}, with a standard deviation of ~2 g kg^{-1} (Table 1). Conversely, the ZSEE/MT database has a minimum value of 0.7 g kg^{-1} and goes to a maximum of 73.8 g kg^{-1}, with a standard deviation of 9.8 g kg^{-1} (Table 2). Hence, the variation of organic carbon in the ZSEE/MT database was much larger than the variation for which the PTFs were calibrated. While our datasets for PTFs calibration were strictly based on soils used for agricultural production, soil data for the ZSEE/MT database were collected from diverse types of land use.

The depletion of soil organic carbon when converting native lands to agriculture is well-reported (Aryal et al., 2018; Lal, 2007). Furthermore, increasing carbon in tropical agroecosystems is a slow and challenging process even when conservation cropping systems are adopted (Silva et al., 2021). Therefore, results indicate that PTFs developed in this study are not well calibrated to detect large organic carbon variations or high values, as may occur in soils under native vegetation. This result suggests that using organic carbon data from ZSEE/MT to apply our PTFs should be done with caution and previous analysis.

A similarly high level of dissimilarity occurred between calibration datasets and ZSEE/MT database (Fig. 3c) in chemical properties values, especially CEC and BS (Table 1 and Table 2). For example, the dataset Basic + OC + chemical has CEC values ranging from a minimum of 3.31 cmolc dm^{-3} to a maximum of 15.7 cmolc dm^{-3}, with mean = 3.76 cmolc dm^{-3} and median = 7.15 cmolc dm^{-3}, along with a standard variation of 2.42 cmolc dm^{-3}. By comparison, the ZSEE/MT database has CEC values between 0.05 cmolc dm^{-3} and 37.0 cmolc dm^{-3}, with mean = 7.13 cmolc dm^{-3} and median = 6.27 cmolc dm^{-3}, with a standard variation of 3.82 cmolc dm^{-3}. Hence, the ZSEE/MT database has a maximum value way larger than the calibration dataset. It may be related to the collection of samples in native areas with high carbon contents, which was not present in calibration datasets. It is well known that tropical soils have low values of CEC (Fontoura et al., 2022), even with high contents of clay (Yerima et al., 2020), due to the main presence of clays of low activity (Schaefer et al., 2008). Therefore, organic carbon plays a significant role in increasing CEC in these soils (Fageria, 2012; Shaheen et al., 2009), and the lack of organic carbon variability in our calibration datasets hinders the PTFs application using predictors from ZSEE/MT, which addressed larger variations of organic carbon and CEC.

The calibration dataset has values of BS with a mean two times higher than the mean for BS in ZSEE/MT. A significant portion of Brazilian soils are naturally poor in bases (Goedert, 1983; Lopes and Cox, 1977; Lopes, 1984), and so in agricultural systems landholders continuously apply fertilizers to overcome this issue (Castro and Crusciol, 2013).

Therefore, it is expected that soil chemical properties from agricultural lands will show different patterns compared to soils under native vegetation, explaining the high dissimilarity found for BS between the calibration dataset and ZSEE/MT database. Finally, as mentioned concerning organic carbon, PTFs are not well calibrated to run using chemical predictors from the ZSEE/MT database. Given that, we encourage potential users to calculate Mahalanobis distances between their datapoints and the centroids of calibration datasets (Table 5) before applying the PTFs within the state of Mato Grosso using organic carbon and chemical properties as predictors. Conversely, if using the Basic predictor set (sand, silt, clay, Wdc and Bd), PTFs are well calibrated to be applied using predictors from the ZSEE/MT database in 91 % of Mato Grosso state area. The results suggest a need to strengthen research on the relationships between hydraulic and other soil properties, aiming to establish more reliable and widely applicable PTFs. Currently, comprehensive databases that incorporate complete information on soil properties (e.g., hydraulic, chemical, physical, mineralogical) across a broad range of values are lacking.

Table 5. Centroids of variables for each dataset used in calibration of PTFs.

Predictor variable	Dataset		
	Basic	Basic + OC	Basic + OC + chemical
Sand	56.351	55.624	55.624
$Silt_t$	-1.739	-1.679	-1.658
$Clay_t$	-0.701	-0.653	-0.651
Wdc	7.329	7.305	7.305
Bd	1.465	1.458	1.458
OC	-	5.731	5.731
pH	-	-	5.820
CEC	-	-	7.355
BS	-	-	46.270
Cutoff	12.832	14.449	19.023

Basic = basic predictor set with sand, silt, clay, bulk density and water-dispersible clay. Basic + OC = basic predictor set + organic carbon. Basic + OC + chemical = basic predictor set + organic carbon + pH in water + cation exchange capacity + base saturation. $Clay_t$ = ln(clay(%)/sand(%)), $Silt_t$ = ln(silt(%)/sand(%)), Bd = bulk density (Mg m^{-3}), Wdc = water-dispersible clay (g kg^{-1}), OC = organic carbon (g kg^{-1}), CEC = cation exchange capacity (cmolc dm^{-3}), BS = base saturation (%), pH = pH (H_2O).

4. Conclusions

Pedotransfer functions were developed using machine learning algorithms to predict soil hydraulic properties of soils from the Mato Grosso state, Brazil. Overall, all models showed gains in accuracy compared to the null model (mean value), regardless of predictor set or prediction target. For soil water retention prediction, RF stood out showing best performances for 16 out of the 24 PTFs developed. However, MLR, a user-friendly model, showed the best performances for prediction of soil water content at -10 kPa, the field capacity of soils, using predictor sets with organic carbon and chemical properties added to basic properties (sand, silt, clay, water dispersible clay and bulk density). In general, models for soil water retention systematically improved their accuracies with the addition of features and best performances were achieved using organic carbon and chemical properties added to basic properties, except for the drier matric potentials (i.e., -500 kPa and -1500 kPa). Water-dispersible clay showed high importance in predicting soil water retention and deserves attention is future studies, as it is a simple property to obtain and rarely found as a predictor in pedotransfer functions.

Soil hydraulic conductivity was better predicted by RF using basic predictor set and KNN was benefited by the addition of features, showing the best performances using organic carbon and chemical properties along with basic properties. Although showing modest predictive accuracy, all models for hydraulic conductivity showed gains in accuracy and were able to capture patterns beyond the natural variability of the data. Models for hydraulic conductivity were also benefited by the addition of chemical properties. Silt/clay ratio was irrelevant for predicting soil hydraulic conductivity, highlighting the difficulty in capturing the relationship between clay and hydraulic conductivity in tropical soils, which is also highlighted by other authors.

The PTFs domain of application within the state of Mato Grosso showed distinct results according to predictor datasets. The PTFs developed in this study are well calibrated and can be applied to 91% of the state's area using predictors from the ZSEE/MT database if basic properties are chosen (i.e., sand, silt, clay, water dispersible clay and bulk density). However, adding organic carbon with or without chemical properties to the predictor set decreases the applicability domain to 40 or less % of the state's area. The PTFs calibration datasets need to be sampled to expand the magnitudes of organic carbon and chemical properties represented. We strongly recommend potential users identify the centroids of our calibration datasets to calculate whether PTFs are well calibrated to recognize patterns of their predictor sets. Models will be made available upon request.

Acknowledgements

We acknowledge the financial contribution of the Minas Gerais State Research Foundation (FAPEMIG), the Brazil National Council of Technological and Scientific Development (CNPq - PROC. Nº 313154/2020-1 e PROC. Nº 408231/2021-1), the Coordination for the Improvement of Higher Education Personal (CAPES – Finance Code 001, CAPES/PRINT Nº 88887.695902/2022-00), the Department of Agricultural Engineering at the Federal University of Viçosa (DEA – UFV) and the Department of Agricultural Sciences at Murdoch University.

References

Amorim, R.S.S., Albuquerque, J.A., Couto, E.G., Kunz, M., Rodrigues, M.F., Silva, L. de C.M. da, Reichert, J.M., 2022. Water retention and availability in Brazilian Cerrado (neotropical savanna) soils under agricultural use: Pedotransfer functions and decision trees. Soil Tillage Res 224. https://doi.org/10.1016/j.still.2022.105485

Araya, S.N., Ghezzehei, T.A., 2019. Using Machine Learning for Prediction of Saturated Hydraulic Conductivity and Its Sensitivity to Soil Structural Perturbations. Water Resour Res 55, 5715–5737. https://doi.org/10.1029/2018WR024357

Aryal, D.R., Ruiz, D.E.M., Marroquín, C.N.T., Ruiz, R.P., Hernández, F.G., Venegas, J.A.V., Mendoza, A.P., López, G.V., Lugo, F.C., Larramendi, L.A.R., de Coss, A.L., López, A.H., Jonapá, F.J.M., Sanabria, C.A.V., Aguilar, A.A., Chi, I.E., 2018. Soil organic carbon depletion from forests to grasslands conversion in Mexico: A review. Agriculture (Switzerland). https://doi.org/10.3390/agriculture8110181

Beniaich, A., Otten, W., Shin, H.C., Cooper, H. V., Rickson, J., Soulaimani, A., El Gharous, M., 2023. Evaluation of pedotransfer functions to estimate some of soil hydraulic characteristics in North Africa: A case study from Morocco. Front Environ Sci 11. https://doi.org/10.3389/fenvs.2023.1090688

Benke, K.K., Norng, S., Robinson, N.J., Chia, K., Rees, D.B., Hopley, J., 2020. Development of pedotransfer functions by machine learning for prediction of soil electrical conductivity and organic carbon content. Geoderma 366. https://doi.org/10.1016/j.geoderma.2020.114210

Bocuti, E.D., 2020. Condições físico-hídricas de solos mato-grossenses e sua relação com a produtividade de soja: da pesquisa à transposição didática. Tese - Univ. Fed. Mato

Grosso, Fac. Agron. e Med. Veterinária, Programa Pós- Grad. em Agric. Trop. Cuiabá, 2020. 200p.

Bocuti, E.D., Amorim, R.S.S., Kavasaki, K.F.L., Prado, M.R.V., Santos, C.L.R., Di Raimo, L.A.D.L., 2021. Soil structure and its relationship with soybean yield. Revista Brasileira de Engenharia Agricola e Ambiental 25, 168–173. https://doi.org/10.1590/1807-1929/agriambi.v25n3p168-173

Botula, Y.D., Cornelis, W.M., Baert, G., Van Ranst, E., 2012. Evaluation of pedotransfer functions for predicting water retention of soils in Lower Congo (D.R. Congo). Agric Water Manag 111, 1–10. https://doi.org/10.1016/j.agwat.2012.04.006

Botula, Y.-D., Nemes, A., Mafuka, P., Van Ranst, E., Cornelis, W.M., 2013. Prediction of Water Retention of Soils from the Humid Tropics by the Nonparametric k -Nearest Neighbor Approach. Vadose Zone Journal 12, 1–17. https://doi.org/10.2136/vzj2012.0123

Botula, Y.-D., Ranst, E. Van, Cornelis, W.M., 2014. Pedotransfer functions to predict water retention for soils of the humid tropics: a review. Rev Bras Cienc Solo 38, 679–698.

Breiman, L., 2001. Random forests. Machine learning, 45, pp.5-32.

Brito, A.D.S., Libardi, P.L., Mota, J.C.A. and Moraes, S.O., 2011. Estimativa da capacidade de campo pela curva de retenção e pela densidade de fluxo da água. Revista Brasileira de Ciência do Solo, 35, pp.1939-1948.

Bruand, A., 2004. Utilizing mineralogical and chemical information in PTFs. Developments in Soil Science. https://doi.org/10.1016/S0166-2481(04)30009-7

Brubaker, S. C., Holzhey, C. S., Brasher, B. R., 1992. Estimating the water-dispersible clay content of soils. Soil Science Society of America Journal, 56. 1226-1232.

Cañasveras, J.C., Barrón, V., del Campillo, M.C., Torrent, J., Gómez, J.A., 2010. Estimation of aggregate stability indices in Mediterranean soils by diffuse reflectance spectroscopy. Geoderma 158, 78–84. https://doi.org/10.1016/j.geoderma.2009.09.004

Candemir, F., Gülser, C., 2012. Influencing Factors and Prediction of Hydraulic Conductivity in Fine-Textured Alkaline Soils. Arid Land Research and Management 26, 15–31. https://doi.org/10.1080/15324982.2011.631686

Carducci, C.E., Oliveira, G.C.D., Zeviani, W.M., Lima, V.M. and Serafim, M.E., 2013. Bimodal pore distribution on soils under conservation management system for coffee crop. Engenharia Agrícola, 33, pp.291-302.

Cassol, H.L.G., Arai, E., Sano, E.E., Dutra, A.C., Hoffmann, T.B., Shimabukuro, Y.E., 2020. Maximum fraction images derived from year-based project for on-board autonomyvegetation (PROBA-V) data for the rapid assessment of land use and land cover areas in Mato Grosso state, Brazil. Land (Basel) 9. https://doi.org/10.3390/LAND9050139

Castro, G.S.A., Crusciol, C.A.C., 2013. Effects of superficial liming and silicate application on soil fertility and crop yield under rotation. Geoderma 195–196, 234–242. https://doi.org/10.1016/j.geoderma.2012.12.006

Chen, T. and Guestrin, C., 2016, August. Xgboost: A scalable tree boosting system. In Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining (pp. 785-794). https://doi.org/10.1145/2939672.2939785

Chirico, G.B., Medina, H., Romano, N., 2010. Functional evaluation of PTF prediction uncertainty: An application at hillslope scale. Geoderma 155, 193–202. https://doi.org/10.1016/j.geoderma.2009.06.008

Choudhury, B.U., Santra, P., Singh, N., Chakraborty, P., 2023. Development of land-use-specific pedotransfer functions for predicting bulk density of acidic topsoil in eastern Himalayas (India). Geoderma Regional 34. https://doi.org/10.1016/j.geodrs.2023.e00671

Cohn, A.S., Gil, J., Berger, T., Pellegrina, H., Toledo, C., 2016. Patterns and processes of pasture to crop conversion in Brazil: Evidence from Mato Grosso State. Land use policy 55, 108–120. https://doi.org/10.1016/j.landusepol.2016.03.005

Cover, T. and Hart, P., 1967. Nearest neighbor pattern classification. IEEE transactions on information theory, 13(1), pp.21-27.

Cueff, S., Coquet, Y., Aubertot, J.N., Bel, L., Pot, V., Alletto, L., 2021. Estimation of soil water retention in conservation agriculture using published and new pedotransfer functions. Soil Tillage Res 209. https://doi.org/10.1016/j.still.2021.104967

Crawley, M. J., 2013. The R Book. John Wiley & Sons. 1077p.

Dashtaki, S.G., Homaee, M., Khodaverdiloo, H., 2010. Derivation and validation of pedotransfer functions for estimating soil water retention curve using a variety of soil data. Soil Use Manag 26, 68–74. https://doi.org/10.1111/j.1475-2743.2009.00254.x

Dexter, A.R., 2004. Soil physical quality Part I. Theory, effects of soil texture, density, and organic matter, and effects on root growth. Geoderma 120, 201-214. https://doi.org/10.1016/j.geodermaa.2003.09.005

Di Raimo, L.A.D.L., Amorim, R.S.S., Couto, E.G., Hunter, M.O., Torres, G.N., Bocuti, E.D., da Silva, L. de C.M., 2022. Natural erosion potential in the Mato Grosso state, Brazil. Revista Brasileira de Engenharia Agricola e Ambiental 26, 655–661. https://doi.org/10.1590/1807-1929/agriambi.v26n9p655-661

Drucker, h., Burges, C.J.C., Kaufman, L. Smola, A. Vapnik, V., 1997. Support Vector Regression Machines. Advances in Neural Information Processing Systems, 155-161.

Fageria, N.K., 2012. Role of Soil Organic Matter in Maintaining Sustainability of Cropping Systems. Commun Soil Sci Plant Anal 43, 2063–2113. https://doi.org/10.1080/00103624.2012.697234

Ferreira, M.M., Fernandes, B., Curi, N., 1999. Mineralogia da fração argila e estrutura de Latossolos da região sudeste do Brasil. Revista brasileira de ciência do solo, 23, pp.507-514.

Fontoura, T.B., Coutinho, R.Q., da Silva, F.O.T., 2023. Geochemical and Mineralogical Contributions in the Study of Sedimentary Rock (Barreiras Formation) Soils, Recife/Brazil: Implications for Landslides. Geotechnical and Geological Engineering 41, 205–224. https://doi.org/10.1007/s10706-022-02274-8

IUSS Working Group, 2015. World reference base for soil resources 2014. International soil classification system for naming soils and creating legends for soil maps.

Friedman, J.H., 2001. Greedy function approximation: a gradient boosting machine. Annals of statistics, pp.1189-1232.

Guio Blanco, C.M., Brito Gomez, V.M., Crespo, P., Ließ, M., 2018. Spatial prediction of soil water retention in a Páramo landscape: Methodological insight into machine learning using random forest. Geoderma 316, 100–114. https://doi.org/10.1016/j.geoderma.2017.12.002

Gunarathna, M.H.J.P., Sakai, K., Nakandakari, T., Momii, K., Kumari, M.K.N., 2019. Machine learning approaches to develop pedotransfer functions for tropical Sri Lankan soils. Water (Switzerland) 11. https://doi.org/10.3390/w11091940

Gupta, S., Hengl, T., Lehmann, P., Bonetti, S., Or, D., n.d. SoilKsatDB: global soil saturated hydraulic conductivity measurements for geoscience applications. https://doi.org/10.5281/zenodo.3752721

Haghverdi, A., Cornelis, W.M., Ghahraman, B., 2012. A pseudo-continuous neural network approach for developing water retention pedotransfer functions with limited data. J Hydrol (Amst) 442–443, 46–54. https://doi.org/10.1016/j.jhydrol.2012.03.036

Hodnett, M.G., Tomasella, J., 2002. Marked differences between van Genuchten soil water-retention parameters for temperate and tropical soils: a new water-retention pedo-transfer functions developed for tropical soils. Geoderma, 108. 155-180.

Hughes, G.F., 1968. On the Mean Accuracy of Statistical Pattern Recognizers. IEEE Trans Inf Theory 14, 55–63. https://doi.org/10.1109/TIT.1968.1054102

James, G., Witten, D., Hastie, T., Tibshirani, R., 2013. An introduction to statistical learning. 112, New York: Springer. 441p.

Jensen, D.K., Tuller, M., de Jonge, L.W., Arthur, E., Moldrup, P., 2015. A New Two-Stage Approach to predicting the soil water characteristic from saturation to oven-dryness. J Hydrol (Amst) 521, 498–507. https://doi.org/10.1016/j.jhydrol.2014.12.018

Ker, J.C., 1997. Latossolos do Brasil: uma revisão. Geonomos.

Kotlar, A.M., de Jong van Lier, Q., Barros, A.H.C., Iversen, B. V., Vereecken, H., 2019. Development and Uncertainty Assessment of Pedotransfer Functions for Predicting Water Contents at Specific Pressure Heads. Vadose Zone Journal 18. https://doi.org/10.2136/vzj2019.06.0063

Kotlar, A.M., de Jong van Lier, Q., de Souza Brito, E., 2020. Pedotransfer functions for water contents at specific pressure heads of silty soils from Amazon rainforest. Geoderma 361. https://doi.org/10.1016/j.geoderma.2019.114098

Kuhn, M., Johnson, K., 2013. Applied predictive modeling. 26, New York: Springer. 595p. https://doi.org/10.1007/978-1-4614-6849-3

Lal, R., 2007. Carbon management in agricultural soils. Mitig Adapt Strateg Glob Chang. https://doi.org/10.1007/s11027-006-9036-7

Lehmann, P., Bickel, S., Wei, Z., Or, D., 2020. Physical Constraints for Improved Soil Hydraulic Parameter Estimation by Pedotransfer Functions. Water Resour Res 56. https://doi.org/10.1029/2019WR025963

Liao, K., Xu, S., Wu, J., Zhu, Q., 2014. Uncertainty analysis for large-scale prediction of the van Genuchten soil-water retention parameters with pedotransfer functions. Soil Research 52, 431–442. https://doi.org/10.1071/SR13230

Lopes, A.S., Cox, F.R. 1977. A Survey of the Fertility Status of Surface Soils Under "Cerrado" Vegetation in Brazil. Soil Science Society of America Journal, 41, 742-747.

Lopes, A. S. 1984. Solos sob "Cerrado": Características, propriedades e manejo (162p). Piracicaba: Associação Brasileira Para Pesquisa Da Potassa e Do Fosfato. 2a edição.

Goedert, W. J. Management of the Cerrado soils of Brazil: a review, 1983. Journal of Soil Science, 34, 405-428.

Manrique, L.A., Jones, C.A., Dyke, P.T., 1991. Predicting Soil Water Retention Characteristics From Soil Physical And Chemical Properties. Commun Soil Sci Plant Anal 22, 1847–1860. https://doi.org/10.1080/00103629109368540

Martinez, P., Souza, I.F., 2020. Genesis of pseudo-sand structure in Oxisols from Brazil – A review. Geoderma Regional. https://doi.org/10.1016/j.geodrs.2020.e00292

Mcbratney, A.B., Minasny, B., Cattle, S.R., Vervoort, R.W., 2002. From pedotransfer functions to soil inference systems, geoderma.

Medrado, E., Lima, J.E.F.W., 2014. Development of pedotransfer functions for estimating water retention curve for tropical soils of the Brazilian savanna. Geoderma Regional 1, 59–66. https://doi.org/10.1016/j.geodrs.2014.08.003

Minasny, B., Mcbratney, A.B., Bristow, K.L., 1999. Comparison of different approaches to the development of pedotransfer functions for water-retention curves, Geoderma.

Morgan, M.T., Holmes, R.G., Wood, R.K., 1993. A system for measuring soil physical properties in the fieldl, Soil & Tillage Research.

Moura, M.S., Silva, B.M., Mota, P.K., Borghi, E., Resende, A.V. de, Acuña-Guzman, S.F., Araújo, G.S.S., da Silva, L. de C.M., de Oliveira, G.C., Curi, N., 2021. Soil management and diverse crop rotation can mitigate early-stage no-till compaction and improve least limiting water range in a Ferralsol. Agric Water Manag 243. https://doi.org/10.1016/j.agwat.2020.106523

Naorem, A., Jayaraman, S., Dang, Y.P., Dalal, R.C., Sinha, N.K., Rao, C.S., Patra, A.K., 2023. Soil Constraints in an Arid Environment—Challenges, Prospects, and Implications. Agronomy. https://doi.org/10.3390/agronomy13010220

Nguyen, P.M., Le, K. Van, Cornelis, W.M., 2014. Using categorical soil structure information to improve soil water retention estimates of tropical delta soils. Soil Research 52, 443–452. https://doi.org/10.1071/SR13256

Nowlan, S., Wolpert, D.H., n.d. Communicated by The Lack of A Priori Distinctions Between Learning Algorithms.

Nunes, M.R., Vaz, C.M.P., Denardin, J.E., van Es, H.M., Libardi, P.L., da Silva, A.P., 2017. Physicochemical and Structural Properties of an Oxisol under the Addition of Straw and Lime. Soil Science Society of America Journal 81, 1328–1339. https://doi.org/10.2136/sssaj2017.07.0218

Ottoni, M.V., Ottoni Filho, T.B., Lopes-Assad, M.L.R.C., Rotunno Filho, O.C., 2019. Pedotransfer functions for saturated hydraulic conductivity using a database with temperate and tropical climate soils. J Hydrol (Amst) 575, 1345–1358. https://doi.org/10.1016/j.jhydrol.2019.05.050

Ottoni, M.V., Ottoni Filho, T.B., Schaap, M.G., Lopes-Assad, M.L.R.C., Rotunno Filho, O.C., 2018. Hydrophysical Database for Brazilian Soils (HYBRAS) and Pedotransfer Functions for Water Retention. Vadose Zone Journal 17, 1–17. https://doi.org/10.2136/vzj2017.05.0095

Pachepsky, Y.A., Timlin, D., Varallyay, G., n.d. Artificial Neural Networks to Estimate Soil Water Retention from Easily Measurable Data.

Panagea, I.S., Berti, A., Čermak, P., Diels, J., Elsen, A., Kusá, H., Piccoli, I., Poesen, J., Stoate, C., Tits, M., Toth, Z., Wyseure, G., 2021. Soil water retention as affected by management induced changes of soil organic carbon: Analysis of long-term experiments in europe. Land (Basel) 10. https://doi.org/10.3390/land10121362

Patil, N.G., Singh, S.K., 2016. Pedotransfer Functions for Estimating Soil Hydraulic Properties: A Review. Pedosphere 26, 417–430. https://doi.org/10.1016/S1002-0160(15)60054-6

Pessoa, T.N., Libardi, P.L., 2022. Physical-hydric properties of Oxisols as influenced by soil structure and clay mineralogy. Catena (Amst) 211. https://doi.org/10.1016/j.catena.2021.106009

Pham, K., Kim, D., Yoon, Y., Choi, H., 2019. Analysis of neural network based pedotransfer function for predicting soil water characteristic curve. Geoderma 351, 92–102. https://doi.org/10.1016/j.geoderma.2019.05.013

Qiao, J., Zhu, Y., Jia, X., Huang, L., Shao, M., 2019. Pedotransfer functions for estimating the field capacity and permanent wilting point in the critical zone of the Loess Plateau, China. J Soils Sediments 19, 140–147. https://doi.org/10.1007/s11368-018-2036-x

Quentin, S. Philipie, L. 2021. Uncertainty assessment of soil available water capacity using error propagation: A test in Languedoc-Roussillon. Geoderma 391. https://doi.org/10.1016/j.geoderma.2021.114968

Rab, M.A., Chandra, S., Fisher, P.D., Robinson, N.J., Kitching, M., Aumann, C.D., Imhof, M., 2011. Modelling and prediction of soil water contents at field capacity and permanent wilting point of dryland cropping soils. Soil Research 49, 389–407. https://doi.org/10.1071/SR10160

Rabot, E., Wiesmeier, M., Schlüter, S., Vogel, H.J., 2018. Soil structure as an indicator of soil functions: A review. Geoderma. https://doi.org/10.1016/j.geoderma.2017.11.009

Ramcharan, A., Hengl, T., Beaudette, D., Wills, S., 2017. A Soil Bulk Density Pedotransfer Function Based on Machine Learning: A Case Study with the NCSS Soil Characterization Database. Soil Science Society of America Journal 81, 1279–1287. https://doi.org/10.2136/sssaj2016.12.0421

Rastgou, M., Bayat, H., Mansoorizadeh, M., Gregory, A.S., 2020. Estimating the soil water retention curve: Comparison of multiple nonlinear regression approach and random forest data mining technique. Comput Electron Agric 174. https://doi.org/10.1016/j.compag.2020.105502

Rawls, W.J., Pachepsky, Y.A., Ritchie, J.C., Sobecki, T.M., Bloodworth, H., 2003. Effect of soil organic carbon on soil water retention. Geoderma 116, 61–76. https://doi.org/10.1016/S0016-7061(03)00094-6

Reichardt, K. Capacidade de Campo. Revista Brasileira de Ciência do Solo, 12, 211-216.

Reichert, J.M., Albuquerque, J.A., Solano Peraza, J.E., da Costa, A., 2020. Estimating water retention and availability in cultivated soils of southern Brazil. Geoderma Regional 21. https://doi.org/10.1016/j.geodrs.2020.e00277

Rezaei, M., Mousavi, S.R., Rahmani, A., Zeraatpisheh, M., Rahmati, M., Pakparvar, M., Jahandideh Mahjenabadi, V.A., Seuntjens, P., Cornelis, W., 2023. Incorporating machine learning models and remote sensing to assess the spatial distribution of saturated hydraulic conductivity in a light-textured soil. Comput Electron Agric 209. https://doi.org/10.1016/j.compag.2023.107821

Román Dobarco, M., Cousin, I., Le Bas, C., Martin, M.P., 2019. Pedotransfer functions for predicting available water capacity in French soils, their applicability domain and associated uncertainty. Geoderma 336, 81–95. https://doi.org/10.1016/j.geoderma.2018.08.022

Romano, N., Palladino, M., n.d. Prediction of soil water retention using soil physical data and terrain attributes.

Rosseti, R., A., C., Amorim, R.S.S., Di Loreto, L.A., Torres, G.N., da Silva, L. de C.M., Alves, I.M., 2022. Pedotransfer functions for predicting soil-water retention under Brazilian Cerrado. Pesqui Agropecu Bras 57. https://doi.org/10.1590/S1678-3921.PAB2022.V57.02474

Santos, H.G., Jacomine, P.K.T., Anjos, L.H.C., Oliveira, V.A., Lumbreras, J.F., Coelho, M.R., Almeida, J.A., Araújo Filho, J.C., Oliveira, J.B., Cunha, T.J.F., 2018. Sistema brasileiro de classificação de solos, Embrapa Solos.

Santos, R.C.V., Vargas, M.M., Timm, L.C., Beskow, S., Siqueira, T.M., Mello, C.R., Soares, M.F., de Moura, M.M. and Reichardt, K., 2021. Examining the implications of spatial variability of saturated soil hydraulic conductivity on direct surface runoff hydrographs. Catena, 207, p.105693.

Santra, P., Kumar, M., Kumawat, R.N., Painuli, D.K., Hati, K.M., Heuvelink, G.B.M., Batjes, N.H., 2018. Pedotransfer functions to estimate soil water content at field capacity and permanent wilting point in hot Arid Western India. Journal of Earth System Science 127. https://doi.org/10.1007/s12040-018-0937-0

Schaefer, C.E.G.R., Fabris, J.D., Ker, J.C., 2008. Minerals in the clay fraction of Brazilian Latosols (Oxisols): a review. Clay Miner 43, 137–154. https://doi.org/10.1180/claymin.2008.043.1.11

Schapire, R.E., 1990. The Strength of Weak Learnability.

Scheffler, R., Neill, C., Krusche, A. V., Elsenbeer, H., 2011. Soil hydraulic response to land-use change associated with the recent soybean expansion at the Amazon agricultural frontier. Agric Ecosyst Environ 144, 281–289. https://doi.org/10.1016/j.agee.2011.08.016

Shaheen, S.M., Tsadilas, C.D., Mitsibonas, T., Tzouvalekas, M., 2009. Distribution coefficient of Copper in different soils from Egypt and Greece. Commun Soil Sci Plant Anal 40, 214–226. https://doi.org/10.1080/00103620802625625

Shao, W., Chen, S., Su, Y., Dong, J., Ni, J., Yang, Z., Zhang, Y., 2023. Reduce uncertainty in soil hydrological modeling: A comparison of soil hydraulic parameters generated by random sampling and pedotransfer function. J Hydrol (Amst) 623. https://doi.org/10.1016/j.jhydrol.2023.129740

Silva, A.C., Armindo, R.A., dos Santos Brito, A., Schaap, M.G., 2017. SPLINTEX: A physically-based pedotransfer function for modeling soil hydraulic functions. Soil Tillage Res 174, 261–272. https://doi.org/10.1016/j.still.2017.07.011

Silva, B.M., Oliveira, G.C., Serafim, M.E., Silva, É.A., Guimarães, P.T.G., Melo, L.B.B., Norton, L.D., Curi, N., 2019. Soil moisture associated with least limiting water range, leaf water potential, initial growth and yield of coffee as affected by soil management system. Soil Tillage Res 189, 36–43. https://doi.org/10.1016/j.still.2018.12.016

Silva, B.M., Silva, É.A. da, Oliveira, G.C. de, Ferreira, M.M., Serafim, M.E., 2014. Plant-available soil water capacity: estimation methods and implications. Rev. Bras. Ci. Solo 38, 464–475.

Silva, L.C.M. Amorim, R.S.S. Fernandes Filho, E.I., Bocuti, E.D., Silva, D. D. 2023b. Pedotransfer functions and machine learning: Advancements and challenges in tropical soils. Geoderma Regional 35, e00720. https://doi.org/10.1016/j.geodrs.2023.e00720

Silva, L.C.M., Avanzi, J.C., Peixoto, D.S., Merlo, M.N., Borghi, E., Resende, Á.V. de, Acuña-Guzman, S.F., Silva, B.M., 2021. Ecological intensification of cropping systems enhances soil functions, mitigates soil erosion, and promotes crop resilience to dry spells in the Brazilian Cerrado. International Soil and Water Conservation Research 9, 591–604. https://doi.org/10.1016/j.iswcr.2021.06.006

Silva, L.C.M., Peixoto, D.S., Azevedo, R.P., Avanzi, J.C., Dias Junior, M. de S., Vanella, D., Consoli, S., Acuña-Guzman, S.F., Borghi, E., Resende, Á.V. de, Silva, B.M., 2023a. Assessment of soil water content variability using electrical resistivity imaging in an Oxisol under conservation cropping systems. Geoderma Regional 33. https://doi.org/10.1016/j.geodrs.2023.e00624

Silva, L.C.M., Peixoto, D.S., Gomes, J.B.V., Avanzi, J.C., Amorim, R.S.S., Borghi, E., Resende, Á.V. de, Silva, B.M., Mancini, M., Curi, N., 2022. Mineralogy and pore size distribution of clayey Oxisols with granular structure and the effect of management systems. Soil Tillage Res 223. https://doi.org/10.1016/j.still.2022.105479

Simoes, R., Picoli, M.C.A., Camara, G., Maciel, A., Santos, L., Andrade, P.R., Sánchez, A., Ferreira, K., Carvalho, A., 2020. Land use and cover maps for Mato Grosso State in Brazil from 2001 to 2017. Sci Data 7. https://doi.org/10.1038/s41597-020-0371-4

Srayeddin, I., Doussan, C., 2009. Estimation of the spatial variability of root water uptake of maize and sorghum at the field scale by electrical resistivity tomography. Plant Soil 319, 185–207. https://doi.org/10.1007/s11104-008-9860-5

Teixeira, L.A.C., Carvalho, F.T., Vallim, D.C., Pereira, R.C.L., Neto, A.C., Vieira, B.S., Carvalho, R.C.T., Figueiredo, E.E.S., 2020. Listeria monocytogenes in export-approved beef from mato grosso, brazil: Prevalence, molecular characterization and resistance to antibiotics and disinfectants. Microorganisms 8. https://doi.org/10.3390/microorganisms8010018

Teixeira, P. C., Donagemma, G. K., Fontana, A., Teixeira, W. G., 2017. Manual de Métodos de Análise de Solo.. ed. rev. e ampl (3). Brasilia, DF: Embrapa.

Tomasella, J., Hodnett, M., 2004. Pedotransfer functions for tropical soils. Developments in Soil Science. https://doi.org/10.1016/S0166-2481(04)30021-8

Tomasella, J., Hodnett, M.G., 1998. Estimating soil water retention characteristics from limited data in Brazilian Amazonia. Soil Sci 163, 190–202. https://doi.org/10.1097/00010694-199803000-00003

Tomasella, J., Hodnett, M.G., Rossato, L., 2000. Pedotransfer Functions for the Estimation of Soil Water Retention in Brazilian Soils. Soil Science Society of America Journal 64, 327–338. https://doi.org/10.2136/sssaj2000.641327x

Tomasella, J., Pachepsky, Y., Crestana, S., Rawls, W.J., n.d. Comparison of Two Techniques to Develop Pedotransfer Functions for Water Retention.

Torres, L.C., Keller, T., Lima, R.P. de, Tormena, C.A., Lima, H.V. de, Giarola, N.F.B., 2021. Impacts of soil type and crop species on permanent wilting of plants. Geoderma 384. https://doi.org/10.1016/j.geoderma.2020.114798

Tóth, B., Weynants, M., Nemes, A., Makó, A., Bilas, G., Tóth, G., 2015. New generation of hydraulic pedotransfer functions for Europe. Eur J Soil Sci 66, 226–238. https://doi.org/10.1111/ejss.12192

Tranter, G., McBratney, A.B., Minasny, B., 2009. Using distance metrics to determine the appropriate domain of pedotransfer function predictions. Geoderma 149, 421–425. https://doi.org/10.1016/j.geoderma.2009.01.006

Van Looy, K., Bouma, J., Herbst, M., Koestel, J., Minasny, B., Mishra, U., Montzka, C., Nemes, A., Pachepsky, Y.A., Padarian, J., Schaap, M.G., Tóth, B., Verhoef, A., Vanderborght, J., van der Ploeg, M.J., Weihermüller, L., Zacharias, S., Zhang, Y., Vereecken, H., 2017. Pedotransfer Functions in Earth System Science: Challenges and Perspectives. Reviews of Geophysics. https://doi.org/10.1002/2017RG000581

Van Wambeke, A.R., 1962. Criteria for classifying tropical soils by age. Journal of Soil Science 13, 125-132.

Veloso, M.F., Rodrigues, L.N., Filho, E.I.F., 2022. Evaluation of machine learning algorithms in the prediction of hydraulic conductivity and soil moisture at the Brazilian Savannah. Geoderma Regional. https://doi.org/10.1016/j.geodrs.2022.e00569

Vereecken, H., Weynants, M., Javaux, M., Pachepsky, Y., Schaap, M.G., van Genuchten, M.Th., 2010. Using Pedotransfer Functions to Estimate the van Genuchten–Mualem Soil Hydraulic Properties: A Review. Vadose Zone Journal 9, 795–820. https://doi.org/10.2136/vzj2010.0045

Wassar, F., Gandolfi, C., Rienzner, M., Chiaradia, E.A., Bernardoni, E., 2016. Predicted and measured soil retention curve parameters in Lombardy region north of Italy. International Soil and Water Conservation Research 4, 207–214. https://doi.org/10.1016/j.iswcr.2016.05.005

Witten, I.H., Frank, E., Mark, A., 2011. Data Mining Practical Machine Learning Tools and Techniques.

Wu, X., Kumar, V., Ross, Q.J., Ghosh, J., Yang, Q., Motoda, H., McLachlan, G.J., Ng, A., Liu, B., Yu, P.S., Zhou, Z.H., Steinbach, M., Hand, D.J., Steinberg, D., 2008. Top 10 algorithms in data mining. Knowl Inf Syst 14, 1–37. https://doi.org/10.1007/s10115-007-0114-2

Yamaç, S.S., Şeker, C., Negiş, H., 2020. Evaluation of machine learning methods to predict soil moisture constants with different combinations of soil input data for calcareous soils in a semi arid area. Agric Water Manag 234. https://doi.org/10.1016/j.agwat.2020.106121

Yerima, B.P.K., Enang, R.K., Kome, G.K., Van Ranst, E., 2020. Exchangeable aluminium and acidity in Acrisols and Ferralsols of the north-west highlands of Cameroon. Geoderma Regional 23. https://doi.org/10.1016/j.geodrs.2020.e00343

Zhang, Y., Schaap, M.G., 2019. Estimation of saturated hydraulic conductivity with pedotransfer functions: A review. J Hydrol (Amst) 575, 1011–1030. https://doi.org/10.1016/j.jhydrol.2019.05.058

Zhao, X., Othmanli, H., Schiller, T., Zhao, C., Sheng, Y., Zia, S., Müller, J., Stahr, K., 2015. Water use efficiency in saline soils under cotton cultivation in the Tarim River Basin. Water (Switzerland) 7, 3103–3122. https://doi.org/10.3390/w7063103

4. GENERAL CONCLUSIONS

The literature review performed in Paper I demonstrated an outlook change in PTFs development across the world, as a considerable portion of the studies was performed in tropical regions. Machine learning techniques were applied in 45% of the studies and conventional regression in 55%. Results exposed a limitation in machine learning studies as most works tested only one or two algorithms, which can restrict PTFs reliability. Error and goodness-of-fit metrics (R^2, RMSE and MAE) indicated that, overall, machine learning techniques have provided better performances than conventional regression. The results of the case study indicated that field capacity predictions had poor performance, as none of the tested PTFs performed better than the null model (mean value). The prediction of hydraulic conductivity exhibited intriguing results as basic models performed better than more complex and elaborated models. The most remarkable result was that all PTFs systematically presented worse predictions of hydraulic conductivity for fine-textured soils, reinforcing the particularities (e.g., clay mineralogy and soil structure) of clayey tropical soils. Results from Paper I paved the way for developing new PTFs specifically addressing soils from Mato Grosso state.

In Paper II, all models developed provided gains in accuracy compared to the null model, regardless of predictor set or prediction target. For soil water retention prediction, Random Forest stood out showing best performances for 16 out of the 24 PTFs developed. In general, models for soil water retention systematically improved their accuracies with the addition of features, except those for the drier matric potentials (i.e., -500 kPa and -1500 kPa). Soil hydraulic conductivity was better predicted by Random Forest using basic predictors and k-Nearest Neighbor was benefited by the addition of features, showing better performances using organic carbon and chemical properties along with basic predictors. Although modest, all models for hydraulic conductivity provided gains in accuracy. Establishing the PTFs domain of application within the state of Mato Grosso exposed distinct results according to predictor datasets. The PTFs developed in Paper II are well calibrated for 91% of the state's area using basic predictors (i.e., sand, silt, clay, water dispersible clay and bulk density). However, including organic carbon and chemical properties as predictors decreases the applicability domain to 40 or less % of the state's area. Authors recommend potential users to check the calibration dataset centroids to verify whether PTFs are well calibrated to recognize patterns of their predictor sets. Regional-specific PTFs for Mato Grosso were developed in R language and calibrated models will be made available upon request.

SECOND PART: Australian Soils

5. GENERAL INTRODUCTION

Sandy soils pose several challenges to rainfed agriculture such as low water holding capacity, subsoil constraints, topsoil water repellence, and susceptibility to wind erosion (Betti et al., 2016; Hall et al., 2020; Lowe et al., 2021; Roper et al., 2015). These limitations result in patchy seed germination, poor weed control, low soil coverage rates (Harper et al., 2000) and yield well below the water-limited yield potential (Hall et al. 2020). Across western and southern Australia, many hectares of sandy soils are used for rainfed agricultural production. These sandy soils cover an area highly susceptible to drought due to low annual rainfall, thereby requiring strict soil management practices to improve water availability and maintain agricultural productivity (Rahmati et al., 2020; Silva et al., 2021). In this sense, several soil managements practices have been adopted to improve soil water retention in Australian southwest agricultural fields, especially in regions that already face rainfall scarcity and/or irregularity. Management practices have been used over the years in an attempt to increase root growth and access to stored water in these soils, which mainly include clay delving and spreading (Betti et al., 2016; Cann, 2000; McKissock et al., 2002), mineral and organic amendments (Schapel et al., 2023) and strategic deep tillage (Scanlan and Davies, 2019).

Many of the 16 million hectares of sandy soils in western and southern Australia overlie clay-rich subsoils, which allows management strategies such as delving, deep tillage and subsequent turning and mixing of soil layers. Deep tillage has also been applied combining this operation with the incorporation of organic and inorganic amendments (Schapel et al., 2023). The incorporation of subsoil clay and amendments in the topsoil sandy layers can lead to several physical, chemical, and biological benefits (Bell and Oliveira, 2022), such as increased soil aggregates stability (Djajadi et al., 2012), decreased water repellence and increased water holding capacity (Hallett, 2008; Ruthrof et al., 2019), increased organic carbon, increased cation exchange capacity, increased nutrient retention and increased microbial biomass (Hallett, 2008; Müller and Deurer, 2011; Roper et al., 2015). Nonetheless, these soil alleviation practices are not feasible in all regions, due to high operation costs (Hallett, 2008), or due to unavailability or insufficiency of material (e.g., subsoil clay) to provide the desired benefits (Unkovich et al., 2020). Furthermore, the consequences of modifying soil profiles through amelioration practices are still not well understood (Roper et al., 2015), which results in a growing and current demand for research on the effectiveness and suitability of these management practices.

Management practices aiming at clay delving and spreading and deep tillage can lead to a heterogeneous soil profile (Hall et al., 2020; Scanlan and Davies, 2019) and adding mineral amendments modify the soil texture (Mohawesh and Durner, 2019). Hence, these modified soils may exhibit different characteristics from typical soils formed under natural conditions, and their hydrophysical pattern remains unclear. For instance, studies have used the Groenevelt and Grant equation (Grant et al., 2010; Groenevelt and Grant, 2004) to fit soil water retention curves of Western Australian sandy soils (Betti, 2019; Betti et al., 2016). However, few efforts – if any – have been made towards addressing whether Groenevelt and Grant equation better represents the water retention pattern of these human-modified sandy soils compared to other models, such as the classical equations proposed by van Genuchten (1980) or Brooks and Corey (1964). It is noteworthy that Groenevelt and Grant (2004) compared their equation with van Genuchten (1980) model but using European soil samples in a wide texture range, therefore probably resulting in quite different patterns from Western Australian sandy soils and their modified versions.

Furthermore, several further issues concerning the hydrophysical pattern of ameliorated sandy soils have been raised in recent studies and remain uncertain. As stated by Unkovich et al. (2020), the amount of clay required to avoid topsoil water repellence may be insufficient to improve soil water holding capacity. As also found by Betti et al. (2016), the hydrological pattern of sandy soils after clay spreading could not be predicted based only on texture, as similar-textured soils had different available water capacities and water retention curve shapes. Schapel et al. (2018) demonstrated that water availability is distinct in clay-amended sandy soils depending on the concentration of subsoil clay added and concluded that large amounts may be harmful. Mohawesh and Durner (2019) emphasized the need for carrying out experiments on a field-scale for the proper assessment of amendments effects. These findings are an important knowledge gap regarding water dynamics in sandy soils subjected to amelioration practices.

In this context, this second part aimed to: (i) evaluate the performance of different models to represent the water retention curves in Western Australian sandy soils under agricultural use, testing four widely-used equations: Brooks and Corey (1964), van Genuchten (1980), Kosugi (1996) and Groenevelt and Grant (2004). (ii) assess the hydrophysical properties of sandy soils modified by amelioration practices, such as subsoil clay incorporation and organic and mineral amendments. Soil water retention curves and pore size distribution were determined throughout the soil managements to evaluate their effects on plant-available water capacity. The outcomes of these studies aimed to fill a gap in understanding the water dynamics in sandy soils modified by subsoil clay, mineral and organic amendments associated with strategic deep tillage, in places where rainwater supply is limited.

References

Betti, G., 2018. Changing the physical properties of texture-contrast soils by clay delving. Ph.D. Thesis. The University of Adelaide, School of Agriculture, Food and Wine. 160p.

Betti, G., Grant, C.D., Murray, R.S., Churchman, G.J., 2016. Size of subsoil clods affects soil-water availability in sand-clay mixtures. Soil Research 54, 276–290. https://doi.org/10.1071/SR15115

Brooks, R.H., Corey, A.T., 1964. Hydraulic Properties of Porous Media, Hydrology Papers, Colorado State University. Hydrology Papers, Colorado State University.

Cann, M.A., 2000. Clay spreading on water repellent sands in the south east of South Australia promoting sustainable agriculture. J Hydrol (Amst) 231–232, 333–341. https://doi.org/10.1016/S0022-1694(00)00205-5

Djajadi, Abbott, L.K., Hinz, C., 2012. Synergistic impacts of clay and organic matter on structural and biological properties of a sandy soil. Geoderma 183–184, 19–24. https://doi.org/10.1016/j.geoderma.2012.03.012

Grant, C.D., Groenevelt, P.H., Robinson, N.I., 2010. Application of the Groenevelt-Grant soil water retention model to predict the hydraulic conductivity. Australian Journal of Soil Research 48, 447–458. https://doi.org/10.1071/SR09198

Groenevelt, P.H., Grant, C.D., 2004. A new model for the soil-water retention curve that solves the problem of residual water contents. Eur J Soil Sci. https://doi.org/10.1111/j.1365-2389.2004.00617.x

Hall, D.J.M., Davies, S.L., Bell, R.W., Edwards, T.J., 2020. Soil management systems to overcome multiple constraints for dryland crops on deep sands in a water limited environment on the south coast of western australia. Agronomy 10, 1–21. https://doi.org/10.3390/agronomy10121881

Hallett, P.D., 2008. A brief overview of the causes, impacts and amelioration of soil water repellency - A review. Soil and Water Research 3, S21–S29. https://doi.org/10.17221/1198-swr

Harper, R.J., McKissock, I., Gilkes, R.J., Carter, D.J., Blackwell, P.S., 2000. A multivariate framework for interpreting the effects of soil properties, soil management and landuse on water repellency. J Hydrol (Amst) 231–232, 371–383. https://doi.org/10.1016/S0022-1694(00)00209-2

Kosugi, K., 1996. Lognormal distribution model for unsaturated soil hydraulic properties. Water Resour Res 32, 2697–2703. https://doi.org/10.1029/96WR01776

Lowe, M.A., McGrath, G., Leopold, M., 2021. The Impact of Soil Water Repellency and Slope upon Runoff and Erosion. Soil Tillage Res 205, 104756. https://doi.org/10.1016/j.still.2020.104756

McKissock, I., Gilkes, R.J., Walker, E.L., 2002. The reduction of water repellency by added clay is influenced by clay and soil properties. Appl Clay Sci 20, 225–241. https://doi.org/10.1016/S0169-1317(01)00074-6

Müller, K., Deurer, M., 2011. Review of the remediation strategies for soil water repellency. Agric Ecosyst Environ 144, 208–221. https://doi.org/10.1016/j.agee.2011.08.008

Rahmati, O., Falah, F., Dayal, K.S., Deo, R.C., Mohammadi, F., Biggs, T., Moghaddam, D.D., Naghibi, S.A., Bui, D.T., 2020. Machine learning approaches for spatial modeling of agricultural droughts in the south-east region of Queensland Australia. Science of the Total Environment 699, 134230. https://doi.org/10.1016/j.scitotenv.2019.134230

Roper, M.M., Davies, S.L., Blackwell, P.S., Hall, D.J.M., Bakker, D.M., Jongepier, R., Ward, P.R., 2015. Management options for water-repellent soils in Australian dryland agriculture. Soil Research 53, 786–806. https://doi.org/10.1071/SR14330

Ruthrof, K.X., Hopkins, A.J.M., Danks, M., O'Hara, G., Bell, R., Henry, D., Standish, R., Tibbett, M., Howieson, J., Burgess, T., Harper, R., 2019. Rethinking soil water repellency and its management. Plant Ecol 220, 977–984. https://doi.org/10.1007/s11258-019-00967-4

Scanlan, C.A., Davies, S.L., 2019. Soil mixing and redistribution by strategic deep tillage in a sandy soil. Soil Tillage Res 185, 139–145. https://doi.org/10.1016/j.still.2018.09.008

Schapel, A., Bell, R., Yeap, S., Hall, D., 2023. Sandy soil constraints: Organic and clay amendments to improve the productivity of sandy soils, in: Soil Constraints and Productivity. pp. 1–46.

Schapel, A., Marschner, P., Churchman, J., 2018. Clay amount and distribution influence organic carbon content in sand with subsoil clay addition. Soil Tillage Res 184, 253–260. https://doi.org/10.1016/j.still.2018.08.001

Silva, L. de C.M. da, Avanzi, J.C., Peixoto, D.S., Merlo, M.N., Borghi, E., Resende, Á.V. de, Acuña-Guzman, S.F., Silva, B.M., 2021. Ecological intensification of cropping systems enhances soil functions, mitigates soil erosion, and promotes crop resilience to dry spells in the Brazilian Cerrado. International Soil and Water Conservation Research 9, 591–604. https://doi.org/10.1016/j.iswcr.2021.06.006

Unkovich, M., Therese, M., Llewellyn, R., Hall, J., Gupta, V.V., Macdonald, L.M., 2020. Challenges and opportunities for grain farming on sandy soils of semi-arid south and south-eastern Australia. Soil Research 58, 323–334.

van Genuchten, M.T., 1980. A Closed-form Equation for Predicting the Hydraulic Conductivity of Unsaturated Soils. Soil Science Society of America Journal 44, 892–898.

6. Paper III: Performance of soil water retention models in Western Australian sands under agricultural use

Lucas de Castro Moreira da Silva, Richard Bell, Luca De Prato, Wendy Vance, Ricardo Santos Silva Amorim, Elpídio Inácio Fernandes-Filho

Abstract

Soil water retention is the major determinant of crop productivity especially in dryland agriculture. This property is represented by the soil water retention curve (SWRC). While there are many models for the mathematical description of SWRC, specific soils and regions have unique characteristics, and hence need the selection of the most reliable model. Sandy soils under agricultural use within the Western Australian Wheatbelt are of particular interest concerning SWRC modeling, not only due to their water-limited environment or inherent constraints, but because their profiles have been re-engineered to enhance agricultural production. Re-engineering generates a soil profile distinct from typical soils formed under natural conditions, but their hydrophysical behavior is unclear. Therefore, the objective of this study was to evaluate four popular SWRC models in Western Australian sandy soils: Brooks and Corey (1964), van Genuchten (1980), Kosugi (1996) and Groenevelt and Grant (2004). The influence of different soil re-engineering practices on the models' performance was also assessed. Overall, Brooks and Corey model showed the highest goodness-of-fit, the lowest errors, and a consistent negligible bias across the entire range of soil water content considering the whole dataset. Similarly, Brooks and Corey model also had the best results when the dataset was stratified per soil management. van Genuchten and Groenevelt and Grant models had equivalent overall performance, although the van Genuchten model tended to underestimate soil water contents near saturation, while the Groenevelt and Grant model generally overestimated soil water content across the entire range of matric potentials. The Kosugi model had the poorest performance in all scenarios. In conclusion, results of this study support that comparing different SWRC models may be critical in sandy soils of Western Australia. The potential for minor estimation errors to induce considerable changes in soil water contents should be taken into account, as it can directly impact the water use efficiency in a region where rainfall is scarce.

Keywords: Sandy soils, duplex soils, hydraulic properties, soil water retention curve, delving

1. Introduction

Understanding soil water retention is crucial for optimizing water use in agriculture. Soil water retention refers to the capacity of a soil to hold water in its pores and obtaining quantitative data in this regard is essential for land use planning (Silva et al., 2014), assessing drought risk (Quentin and Philippe, 2021; Rahmati et al., 2020) and modeling crop production (Dobarco et al., 2019). As soil water retention regulates the uptake of water and nutrients by plants (Dick et al., 2018; Fontes et al., 2024), it directly impacts crops productivity (Razzaghi et al., 2020; Silva et al., 2021), especially in water-limited areas ((Blanco-Canqui et al., 2023; Silva et al., 2023b).

Soil water retention is governed by capillary and adsorption forces (Jensen et al., 2015), graphically depicted by the soil water retention curve (SWRC). The SWRC describes the relationship between soil water content and soil matric potential under equilibrium conditions (Assouline and Or, 2013). Despite the absence of an analytical equation to express this property (van Lier, 2020), many empirical models have been proposed for determining its mathematical description through experimental data. Initiatives for modeling SWRC date back decades (e.g., Brooks and Corey, 1964; Brutsaert, 1966; Campbell, 1974; Groenevelt and Grant, 2004; Kosugi, 1996; van Genuchten, 1980) although recent efforts have been made to develop new models (e.g., (Chen et al., 2021; Li et al., 2023; Rudiyanto et al., 2020) and to validate existing ones (e.g., (Amorim et al., 2022; Armindo et al., 2019; Pan et al., 2019).

There is a consensus that the models by van Genuchten (1980), Brooks and Corey (1964) and Kosugi (1996) are among the most popular equations for SWRC description (Du, 2020; Lal and Shukla, 2007; Li et al., 2023; Shukla, 2014; Streck and Weber, 2020). However, they are usually adopted arbitrarily without consideration of the fact that SWRC models can perform differently according to soil attributes (Amorim et al., 2022) and/or soil management (Silva et al., 2022). Hence, the comparison and selection of SWRC models merit attention when addressing specific regions (Pan et al., 2019) and soils altered by agricultural activities (Hallema et al., 2015).

In western and southern Australia, sandy soils used for agricultural production are of particular interest due to their various soil constraints in a water-limited environment. Several management practices have been implemented in the region to overcome intrinsic challenges such as low water holding capacity (Hall et al., 2020; Tennant et al., 1992) and topsoil water repellency (Ruthrof et al., 2019). These management strategies are especially based on clay incorporation (Betti et al., 2016; Cann, n.d.; McKissock et al., 2002), addition of organic and mineral amendments (Schapel et al., 2023), and strategic deep tillage (Scanlan and Davies, 2019). Given that many sandy areas overlie clay-rich subsoils, amelioration methods usually involve delving, deep tillage, turning, and mixing soil layers (Davenport et al., 2011) to bring clay up to the sandy topsoil (Betti et al., 2015; Schapel et al., 2018). While these approaches can lead to great improvements in productivity potential (Bell and Oliveira, 2022; Djajadi et al., 2012), they result in a heterogeneous soil profile (Hall et al., 2020; Scanlan and Davies, 2019), the consequences of which are not yet fully understood (Roper et al., 2015). Soils modified by amelioration interventions may exhibit distinct differences from typical soil profiles formed under natural conditions, and their hydrophysical behavior remains unclear.

Regional-wide studies addressing water retention and/or SWRC models are rarely found. Smettem and Gregory (1996) used the van Genuchten (1980) model to fit SWRC for sandy soils from Western Australia under natural conditions. Betti et al. (2016) assessed water retention in modified soil samples from South Australia using the Groenevelt and Grant (2004) model. Schapel et al. (2018) measured water retention in soils with subsoil clay addition at a matric potential of -10 kPa. Nevertheless, few efforts – if any – have been devoted to evaluate and compare different SWRC models on these sandy soils and their ameliorated profile forms.

In this context, the objective of this study was to evaluate different models to represent the water retention curves in sandy soils under agricultural use within the Western Australian Wheatbelt. Popular soil amelioration practices adopted in the region were compared to assess their effects on SWRC models performance. Four widely used SWRC models were tested: Brooks and Corey (1964), van Genuchten (1980), Kosugi (1996) and Groenevelt and Grant (2004). In addition, this study provides a test for Groenevelt and Grant model in a distinct database, which has been sought (Armindo et al., 2019).

2. Material and Methods

2.1 Study area and experimental area

This work was conducted using a dataset of soil attributes collected at two experimental areas in the Wheatbelt region of Western Australia: Bullaring (S32°28'37.8", E117°37'00.3") and South Kweda (S32º25"40", E117º24"18"). The Bullaring experimental area was established in June 2022 on a deep sandy soil paddock. The South Kweda experimental field was established in May 2022 on a pale, deep non-gravelly sand overlying a kaolin-dominant horizon at a depth of 0.4-0.6m. Both experimental areas have 4 blocks for each of several soil management practices. The climate in the region is mediterranean with wet winters (June-August) and dry hot summers (December-February). Monthly average temperatures range from 10 to 22ºC and the average annual rainfall of 440 mm falls mostly between April and October.

A description of the soil management practices addressed in this study are shown in Table 1. In the Bullaring experimental area, three soil management practices were chosen to analyses: control (i.e., no-tillage planting and no amendment application - B1), deep tillage using Bednar machinery (B2), and deep tillage with Bednar machinery plus clay incorporation at a rate of 100 t of subsoil/ha (B3). Four blocks of each management from 12 experimental plots of 30 m width and 6.4 m length. In the South Kweda site, four soil management practices were assessed: control (no-tillage planting) (K1); deep tillage to 0.35 m using Bednar machinery (K2); incorporation of humiclay compost "Cwise" (10 t/ha) by delving and mixing to 0.35 m using Bednar machinery (K3), and; incorporation of aluminosilicate mineral "Zeolite" at a rate of 30 t/ha by delving and incorporation to 0.35 m using Bednar machinery (K4). Four blocks form 16 experimental plots, with 8 plots with dimensions of 18 m width and 200 m length for the first two treatments and 8 plots with dimensions of 18m wide and 2 m long for the two latter. Soils from both sites were under barley cultivation in 2022.

Table 1. Description of sites where soil samples were taken.

Sampling site	Description (2022 cropping season)
B1	Barley crop without tillage or amendment application
B2	Barley crop with deep tillage
B3	Barley crop with deep tillage + clay-rich subsoil incorporation (100 t/ha)
K1	Barley crop without tillage or amendment application
K2	Barley crop with deep tillage
K3	Barley crop with deep tillage + humiclay compost incorporation (10 t/ha)
K4	Barley crop with deep tillage + aluminosilicate mineral incorporation (30 t/ha)

2.2 Soil sampling and laboratory analyses

Soil sampling campaigns occurred between the 2022 harvest and the 2023 sowing season. Disturbed and undisturbed samples were taken in each block at three soil depths (0-10 cm, 20-30 cm and 30-40 cm). In total, 84 undisturbed and 84 disturbed samples were collected. Undisturbed samples were collected using metallic cores with a height of 5 cm and a diameter of 3.5 cm and disturbed samples were collected with an auger. Disturbed samples were used to carry out particle size distribution analysis according to the Pipette method (Dane and Topp, 2018). For samples from soils with clay incorporation, clay clods were ground until they could pass through a sieve with a 2 mm aperture.

Undisturbed samples were saturated and equilibrated to matric potentials of -4, -6, -10, -33 and -100 kPa in a Richards' pressure chamber (Klute, 1986). After determining water retention at each matric potential, samples were dried in an oven at 105 ºC for 48 hours to determine soil bulk density (Bd). Gravimetric water content at the potentials of -500 and -1500 kPa were determined using disturbed samples in a WP4 psychrometry (Gubiani et al., 2013). Gravimetric water content was then multiplied by the bulk density to obtain the volumetric water content.

After laboratory analyses, the dataset for this study was assembled. Basic soil properties of each sampling site are shown in Table 2 and descriptive statistics for the whole dataset is shown in Table 3. Soil water retention curves were fitted for each data point and the performance of models was then evaluated overall and stratified by sampling site (i.e., soil management).

Table 2. Average basic soil properties of the 0-40 cm soil layer in each sampling site.

Sampling site	Sand	Silt	Clay	Bd
	-----	%	-----	Mg m^{-3}
B1	97.2	1.3	1.5	1.64
B2	97.8	0.8	1.5	1.63
B3	94.0	1.5	4.5	1.63
K1	95.4	1.3	3.3	1.64
K2	93.6	1.3	5.1	1.62
K3	91.8	2.4	5.8	1.61
K4	96.7	1.2	2.1	1.59

For sampling sites description, reader is referred to Table 1. Bd = Bulk density.

Table 3. Descriptive statistics for the soil dataset (n = 84).

	Min	Median	Mean	Max	SD
Sand (%)	86.6	96.5	95.2	98.6	2.95
Silt (%)	0.4	1.3	1.4	4.1	0.83
Clay (%)	0.3	2.4	3.4	11.8	2.91
Bd (Mg m^{-3})	1.30	1.63	1.62	1.86	0.12
0 kPa (m^{-3} m^{-3})	0.35	0.44	0.45	0.62	0.06
-4 kPa (m^{-3} m^{-3})	0.26	0.34	0.35	0.52	0.06
-6 kPa (m^{-3} m^{-3})	0.10	0.18	0.19	0.29	0.05
-10 kPa (m^{-3} m^{-3})	0.05	0.10	0.11	0.22	0.03
-33 kPa (m^{-3} m^{-3})	0.03	0.07	0.07	0.17	0.03
-100 kPa (m^{-3} m^{-3})	0.02	0.05	0.06	0.16	0.03
-500 kPa (m^{-3} m^{-3})	0.01	0.03	0.04	0.13	0.03
-1500 kPa (m^{-3} m^{-3})	0.00	0.02	0.03	0.11	0.02

Bd = Bulk density; “n” kPa = soil water content at matric potential of “n” kPa.

2.3 Soil water retention curve (SWRC) models

2.3.1 Brooks and Corey water retention model

Brooks and Corey (1964) model is one of the oldest and most widely used equations for fitting soil water retention across a range of soil water contents. It is sometimes preferred over other models due to its mathematical simplicity (Stankovich· and Lockington, 1995; van Lier, 2020). Brooks and Corey model is expressed as:

$$\theta(h) = \theta r + (\theta s - \theta r)(\alpha h)^{-\lambda}$$

where θ is the volumetric water content (m^{-3} m^{-3}); h is the matric potential (hPa); θs is the saturated water content (m^{-3} m^{-3}); θr is the residual water content (m^{-3} m^{-3}); α and λ are empirical shape parameters.

2.3.2 van Genuchten water retention model

van Genuchten (1980) is probably the most popular soil water retention model due to its high prediction capacity (Dexter, 2004; Ross et al., 1991), which has been validated in a wide range of soils (Cornelis et al., 2005; Madi et al., 2018; van Genuchten and Nielsen, 1985). In this study, we used van Genuchten model with Mualem restriction (m = 1 – 1/n) (Mualem, 1976). The van Genuchten model with Mualem restriction is expressed as:

$$\theta(h) = \theta r + (\theta s - \theta r)[1 + (\alpha h)^n]^{-m}$$

where θ is the volumetric water content (m^{-3} m^{-3}); h is the matric potential (hPa); θs is the saturated water content (m^{-3} m^{-3}); θr is the residual water content (m^{-3} m^{-3}); α, n and m are empirical shape parameters.

2.3.3 Kosugi water retention model

The Kosugi (1996) water retention model is also a popular model although less used compared to the two models described above. Kosugi model is a lognormal distribution model with four parameters:

$$\theta(h) = \theta r + \frac{1}{2}(\theta s - \theta r)\, erfc\, [\frac{ln\,(\frac{h}{hm})}{\sigma\sqrt{2}}]$$

where θ is the volumetric water content (m^{-3} m^{-3}); θs is the saturated water content (m^{-3} m^{-3}); θr is the residual water content (m^{-3} m^{-3}); $erfc$ is the complementary error function; h is the matric potential (hPa); hm is the matric potential corresponding to median pore radius (hPa); σ is a dimensionless parameter related to pore size distribution.

2.3.4 Groenevelt and Grant water retention model

Groenevelt and Grant (2004) proposed a mathematically versatile equation which can be anchored at any water content point. For works addressing the “plant-available” range of soil water retention (i.e., no higher matric potential than the wilting point), the authors recommend anchoring the curve at the measured wilting point (i.e., water content at matric potential of – 15000 hPa). The Groenevelt and Grant equation is expressed as:

$$\theta(h) = \theta wp + k1\{\exp\left(\frac{-k0}{15000^n}\right) - \exp\left(\frac{-k0}{hm^n}\right)\}$$

where θ is the volumetric water content (m^{-3} m^{-3}); θwp is the water content at the wilting point (m^{-3} m^{-3}); h is the matric potential (hPa); hm is the matric potential corresponding to median pore radius (hPa); $k0$, $k1$, and n are dimensionless fitting parameters.

2.4 Data analysis

SWRC were fitted using the “SWRC fit” software (Seki et al., 2023) and the “fitsoilwater app” (Silva and Lima, 2022). Both softwares determine the best SWRC fit based on least squares method. Model performances were evaluated based on the mean absolute error (MAE), root mean squared error (RMSE), mean bias error (MBE) and coefficient of determination (R^2). After checking assumptions, ANOVA was carried out followed by Tukey’s test ($p<0.05$) to compare averages of RMSE and MBE among the tested models. Comparisons for MBE were performed using the absolute values. Pearson’s correlation analysis was carried out to test whether soil attributes influenced a model’s performance.

3. Results and Discussion

3.1 Performance of soil water retention curve (SWRC) models

Performances of SWRC models for all measurements of soil water content are shown in Figure 1. Overall, all models could explain more than 95% of the measured data variability. The Brooks and Corey (Fig. 1a) was the most accurate and consistent model, exhibiting the highest goodness-of-fit, the lowest errors, and a negligible bias (R^2 = 0.988, RMSE = 0.016 m^3 m^{-3}, MAE = 0.012 m^3 m^{-3} and MBE = 0.000 m^3 m^{-3}). van Genuchten and Groenevelt and Grant models (Fig. 1b and Fig. 1d, respectively) had similar performance with intermediate errors (RMSE = 0.027 m^3 m^{-3} and MAE = 0.020 m^3 m^{-3}, RMSE = 0.029 m^3 m^{-3} and MAE = 0.021 m^3 m^{-3}, respectively). They also had comparable absolute bias, although van Genuchten showed an overall inclination for underestimating (MBE = -0.009 m^3 m^{-3}) and Groenevelt and Grant generally overestimated the soil water content (MBE = 0.006 m^3 m^{-3}). Kosugi model (Fig. 1c) had the poorest performance in terms of error (RMSE = 0.035 m^3 m^{-3} and MAE = 0.026 m^3 m^{-3}) and an intermediate absolute bias towards overestimation (MBE = 0.008 m^3 m^{-3}).

Bias-related errors showed a clear distinction among models. As previously noted, the Brooks and Corey model had consistent estimates across the entire range of soil water content, displaying a uniform dispersion of data points. Conversely, van Genuchten showed bias towards underestimation of soil water contents higher than 0.10 m^3 m^{-3}. A systematic bias error was observed between water contents of 0.20 and 0.35 m^3 m^{-3}, where van Genuchten model underestimated all data points. This water content range is related to matric potentials near saturation (> -10 kPa, see Table 3), and this result aligns with several studies which have reported poorer performances for modeling or predicting this segment of the SWRC (Kotlar et al., 2020; Silva et al., 2023a; Too et al., 2014; Wassar et al., 2016). As stated by Amorim et al. (2022), modeling SWRC at near saturation is even more complicated in sandy soils, where slight changes in matric potential may lead to substantial alterations in soil water content.

The Kosugi model had a general positive bias, yet its errors were contradictory depending on the soil water content range. Errors towards overestimating soil water content were observed from near dryness up to 0.25 $m^3 m^{-3}$, while underestimation occurred in data points with water content above 0.25 $m^3 m^{-3}$. This indicates that soil water content estimates by Kosugi model did not adhere to the same linear relationship as measured data, explaining its poorer performance. The Groenevelt and Grant model also showed an overall positive bias error especially within the range of 0.10 $m^3 m^{-3}$ to 0.20 $m^3 m^{-3}$. However, it had a more consistent and symmetric data dispersion compared to van Genuchten and Kosugi models. In general, the findings reinforce the importance of evaluating various models in specific soil datasets. Depending on the chosen model, estimates could be either overestimated or underestimated. These errors, even though subtle, may lead to significant errors for practical applications (Vogel et al., 2001).

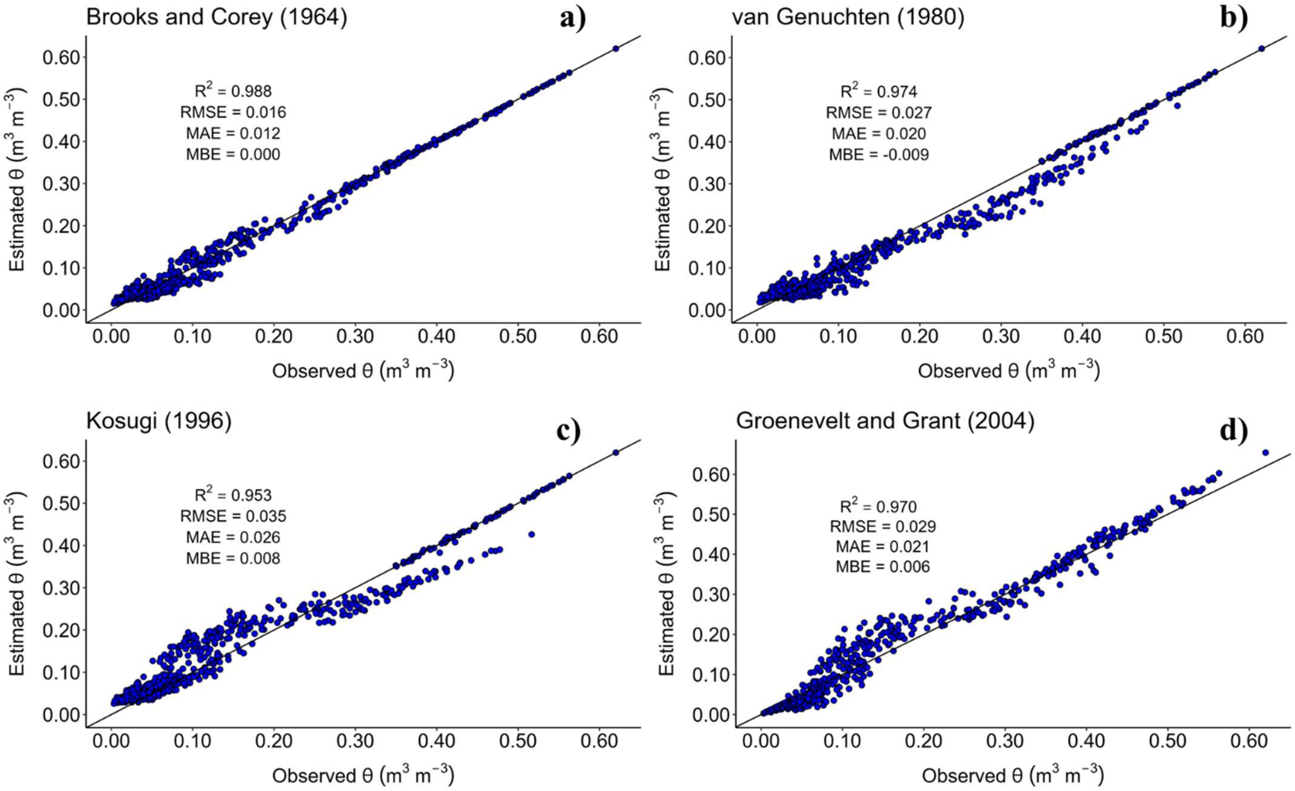

Figure 1. Performance of soil water retention models for Western Australian Wheatbelt sands considering the whole dataset. θ = soil water content, R^2 = coefficient of determination, RMSE = root mean square error, MAE = mean absolute error, MBE = mean bias error.

RMSE and MBE were selected as criteria for the statistical comparison of SWRC model performance (Figure 2). Results in Fig. 2 refer to averages of SWRC errors fitted for each soil sample. The Brooks and Corey model stood out, significantly outperforming all others by showing both the lowest RMSE (Fig. 2a) and the lowest absolute MBE (Fig. 2b). This result coincides with Amorim et al. (2022), who found that Brooks and Corey had lower errors compared to van Genuchten fitting SWRC in sandy soils. Despite that, van Genuchten model stands as the most popular choice, well-known for its high prediction capacity (Dexter, 2004; Madi et al., 2018) and flexibility (Lal and Shukla, 2007), validated in a diverse range of soils (Cornelis et al., 2005; van Genuchten & Nielsen, 1985). Thus, this flexibility often makes van Genuchten model a preferable option for generalization purposes. Nevertheless, findings exposed in Fig. 1 and Fig. 2 are sufficient to conclude that comparing different models is worthwhile when addressing specific soil datasets. This was also emphasized by Pan et al. (2019), who reported similar performances for Brooks and Corey and van Genuchten in specific soils from Tibet.

van Genuchten and Groenevelt and Grant models did not differ regarding RMSE and absolute values of MBE, although their overall biases were opposite. van Genuchten model systematically underestimated soil water content, whereas Groenevelt and Grant predominantly overestimated. Similar results comparing both models were found by Armindo et al. (2019) assessing Brazilian soils, as well as the original publication of Groenevelt and Grant (2004), that also found a comparable accuracy with van Genuchten model using a Dutch soil database. Thus, our results further contribute to testing the Groenevelt and Grant model in distinct databases beyond its original context. The Kosugi model presented the highest errors for the soil dataset used in this study (Fig. 2a), although equivalent bias to van Genuchten and Groenevelt and Grant models (Fig. 2b).

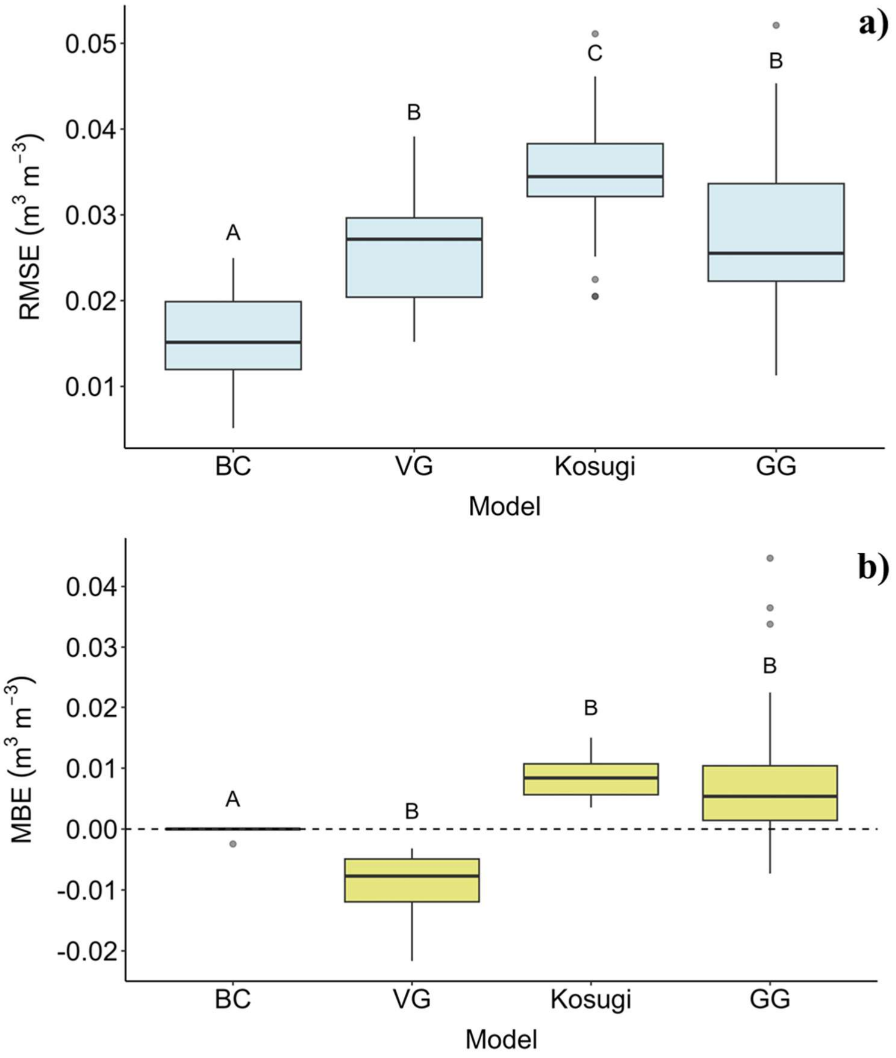

Figure 2. Boxplot of means of RMSE = root mean square error (a) and MBE = mean bias error (b). Means with same letters on top of boxes do not differ by Tukey's test ($p < 0.05$). BC = Brooks and Corey (1964) model; VG = Van Genuchten (1980) model, Kosugi = Kosugi (1996) model, GG = Groenevelt and Grant (2004) model.

3.2 Effects of soil management on SWRC models

Table 4 presents the results of SWRC model performance stratified by sampling site. The Brooks and Corey model ranked above all other models regardless of sampling site. The results using this model had the best goodness-of-fit with R^2 ranging from 0.984 to 0.991, along with the lowest error metrics (RMSE between 0.015 and 0.018 m^3 m^{-3}, and MAE between 0.010 and 0.014 m^3 m^{-3}). Furthermore, Brooks and Corey estimates were very consistent, exhibiting negligible bias error (MBE < 0.000 $m^3 m^{-3}$) across all sampling sites. In contrast, Kosugi model had consistently the poorest performances. Kosugi estimate errors ranged from 0.033 to 0.037 $m^3 m^{-3}$ for RMSE and from 0.024 to 0.028 $m^3 m^{-3}$ considering MAE. Positive bias related to Kosugi model had little variance across sampling sites, ranging between 0.008 and 0.009 $m^3 m^{-3}$. Overall, although Kosugi and Brooks and Corey models demonstrated divergent general performances, their results were regularly distributed across sampling sites.

Differences related to sampling sites were observed with respect to van Genuchten and Groenevelt and Grant models. van Genuchten had its best performance modeling samples from South Kweda experimental area (i.e., K1, K2, K3 and K4), where lower errors were detected, with RMSE values between 0.021 and 0.026 $m^3 m^{-3}$ and MBE values between -0.005 and -0.009 $m^3 m^{-3}$. Conversely, in sampling sites from the Bullaring experimental area (i.e., B1, B2 and B3), van Genuchten model presented poorer performances, with RMSE ranging from 0.028 to 0.030 $m^3 m^{-3}$ and MBE towards underestimation between -0.011 and -0.013 $m^3 m^{-3}$. Groenevelt and Grant model, in contrast, had its best performance modeling samples from Bullaring experimental area, with lower RMSE values (from 0.023 to 0.026 $m^3 m^{-3}$) than those found in samples from South Kweda area (from 0.027 to 0.033 $m^3 m^{-3}$). The same pattern was observed regarding MBE values, which ranged between 0.003 and 0.004 $m^3 m^{-3}$ in Bullaring area and from 0.006 to 0.011 $m^3 m^{-3}$ in sampling sites of South Kweda.

Nevertheless, no sufficient evidence was yielded to draw conclusions regarding soil management within the experimental areas. It indicates that the differences in model's performance were probably influenced by particularities of each experimental area, rather than by soil management practices. These particularities are not clearly distinguished through the assessed soil properties and may extend beyond the scope of this study. In summary, the Brooks and Corey model stood out with very accurate and consistent results regardless of sampling sites or experimental areas. As the dataset used in this study is quite specific, it suggests that the relative simplicity of the Brooks and Corey model (Lal and Shukla, 2007; van Lier, 2020) was beneficial, whereas more versatile models such as van Genuchten and Groenevelt and Grant might have been affected by subtle nuances, diminishing their performances.

Differences in performances of van Genuchten and Groenevelt and Grant models between the two experimental areas were systematic and worth further discussion. For instance, sampling sites from South Kweda had, in general, higher clay contents than the sampling sites of Bullaring (see Table 2), and this seems to have impacted both models performance. Given that, a Pearson correlation analysis was carried out to investigate interactions between basic soil properties and models performance in the following section.

Table 4. Performance of soil water retention models stratified by sampling site.

	Brooks and Corey (1964)				van Genuchten (1980)			
Sampling site	R^2	RMSE	MAE	MBE	R^2	RMSE	MAE	MBE
		----- m^3 m^{-3} -----				----- m^3 m^{-3} -----		
B1	0.984	0.017	0.014	0.000	0.965	0.029	0.023	-0.011
B2	0.986	0.018	0.014	0.000	0.971	0.030	0.023	-0.012
B3	0.988	0.017	0.013	0.000	0.970	0.028	0.023	-0.013
K1	0.988	0.017	0.012	0.000	0.970	0.026	0.020	-0.009
K2	0.989	0.016	0.012	0.000	0.981	0.022	0.017	-0.006
K3	0.989	0.016	0.012	0.000	0.977	0.023	0.018	-0.007
K4	0.991	0.015	0.010	0.000	0.984	0.021	0.016	-0.005

	Kosugi (1996)				Groenevelt and Grant (2004)			
Sampling site	R^2	RMSE	MAE	MBE	R^2	RMSE	MAE	MBE
		----- m^3 m^{-3} -----				----- m^3 m^{-3} -----		
B1	0.951	0.033	0.024	0.009	0.977	0.023	0.016	0.003
B2	0.954	0.034	0.025	0.008	0.981	0.024	0.015	0.004
B3	0.957	0.034	0.025	0.009	0.979	0.026	0.017	0.003
K1	0.950	0.037	0.028	0.008	0.966	0.029	0.023	0.010
K2	0.952	0.036	0.026	0.009	0.960	0.033	0.025	0.009
K3	0.951	0.036	0.026	0.008	0.961	0.031	0.024	0.011
K4	0.954	0.036	0.026	0.008	0.974	0.027	0.021	0.006

For description and basic soil properties of sampling sites reader is referred to Table 1 and Table 2, respectively. R^2 = coefficient of determination, RMSE = root mean square error, MAE = mean absolute error, MBE = mean bias error.

3.3 Correlation between soil properties and SWRC models performance

Correlations between basic soil properties and error metrics of SWRC models are shownin Figure 3. Only significant correlations ($p < 0.05$) are exhibited. Many studies have reported significant correlations between clay and parameters of SWRC models (e.g., (Babaeian et al., 2015; Liao et al., 2014; Zhang et al., 2015). Correlations were generally weak and significant interactions were only found for Kosugi model (sand and clay with RMSE) and van Genuchten model (sand and clay with MBE). Brooks and Corey errors did not show any significant correlation with basic soil properties, which aids to explain its consistency regardless of sampling sites.

The significant correlations between clay and error metrics of SWRC warrants further discussion concerning soils in the context of this study. Adding subsoil clay is a common practice in sandy soils of western and southern Australia (Davenport et al., 2011; Masters, 2014; Schapel et al., 2018). Furthermore, even subtle alterations in clay contents of these soils can lead to profound differences in other properties (Schapel et al., 2023), which may reflect on their hydrophysical behavior. Therefore, it is crucial to select models that can depict the patterns of these modified soils. Consequently, although weak, the positive correlations between clay and SWRC errors are worth of attention for further studies.

In order to detect changes in SWRC related to clay addition, more flexible models may be preferred a priori due to their best generalization capacity. For instance, van Genuchten model was benefited when clay contents were higher, decreasing its absolute bias error (Fig 3b). However, in specific scenarios, more complex and versatile models are susceptible to fit to slight nuances and noises, which may lead to worse performance compared to simpler ones (Silva et al., 2023a). It is possibly justified by the high variability that soil hydraulic properties usually present (Mcbratney et al., 2002; Schaap et al., 2001), which may lead to considerable alterations related to small variations in other properties (Morgan et al., 1993).

Alterations led by soil management practices may not show clear differences on SWRC models as pedogenetic differences do. Therefore, more versatile models, which are preferable when modeling diverse soils, might not perform as well as simpler models do when addressing specific datasets. The soil dataset evaluated in this study is highly homogeneous, with all sampling sites within the same region and with similar texture. Consequently, Brooks and Corey, a model mathematically simpler than the others, presented higher accuracy along with more consistent estimates regardless of slight changes in soil texture. In summary, the findings of this study support the approach that involves comparing different SWRC models when addressing specific soil datasets, as also stated by other authors (e.g., Hallema et al., 2015; Pan et al., 2019).

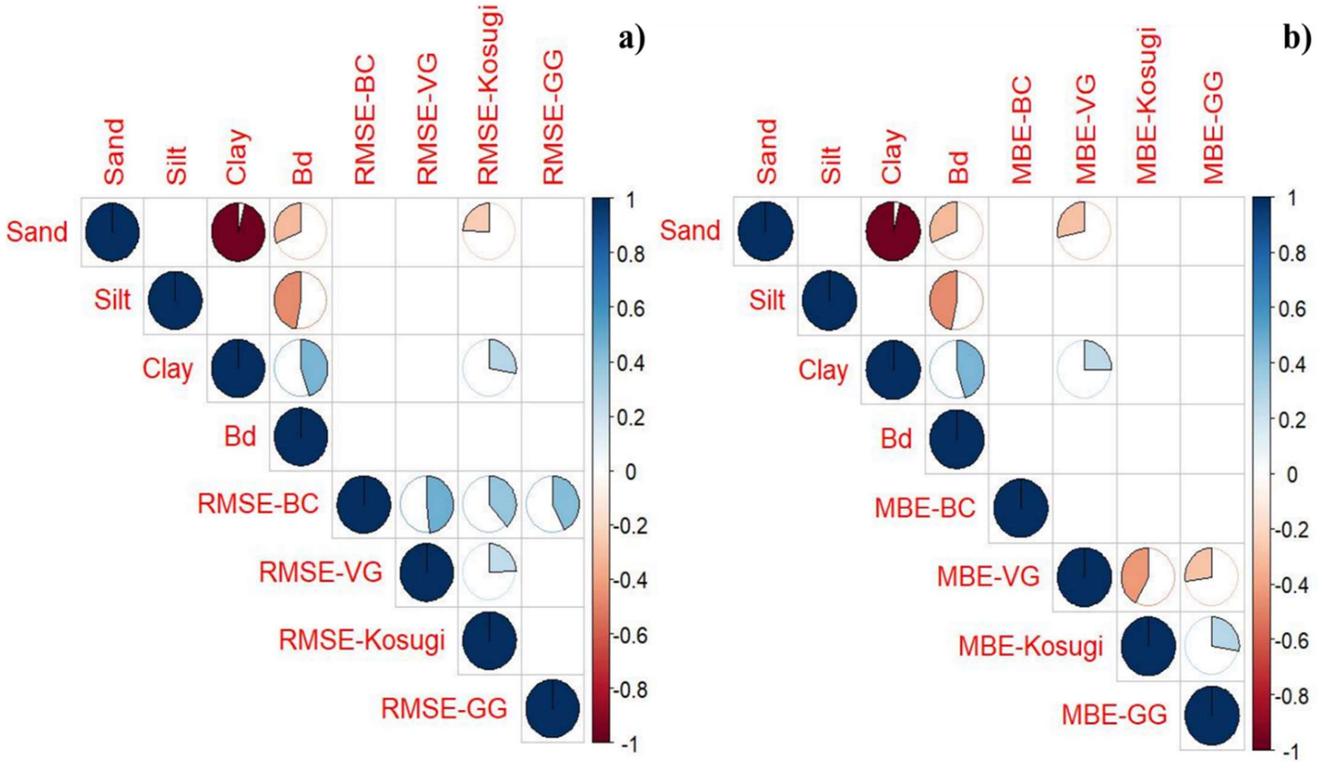

Figure 3. Pearson correlation matrices among soil basic properties and root mean square error (a) and mean bias error of four models (b). Only significant differences are shown ($p < 0.05$). RMSE = root mean square error, MBE = mean bias error, Bd = bulk density, BC = Brooks and Corey (1964) model, VG = Van Genuchten (1980) model, Kosugi = Kosugi (1996) model, GG = Groenevelt and Grant (2004) model.

4. Conclusions

This study evaluated four popular soil water retention models in a specific soil dataset from the Western Australian Wheatbelt. Performances of Brooks and Corey (1964), van Genuchten (1980), Kosugi (1996) and Groenevelt and Grant (2004) models were compared, as well as their interactions with popular soil amelioration practices adopted in the region.

The Brooks and Corey model ranked above all other models regardless of soil water content range, experimental area, or sampling site. For all scenarios, the Brooks and Corey model exhibited the highest goodness-of-fit, the lowest errors, and a consistent negligible bias. van Genuchten and Groenevelt and Grant models had similar overall performance, however, the van Genuchten model tended to underestimate soil water contents, especially near saturation, while the Groenevelt and Grant model generally overestimated the entire range of soil water contents. The Kosugi model had the poorest performance demonstrating higher errors in all scenarios, although its absolute bias was equivalent to van Genuchten and Groenevelt and Grant models. No sufficient evidence was achieved to draw conclusions regarding soil management effects on the accuracy of soil water retention curves.

Overall, Brooks and Corey model stood out as the most reliable and consistent modelfor predicting soil water retention on sandy profiles, including those that had been re-engineered by deep tillage. The findings of this study support the importance of comparing different soil water retention curve models when addressing specific soil datasets. This is particularly critical in sandy soils within water-limited regions, where minor estimation errors may result in significant changes in soil water content, directly impacting water use efficiency.

Acknowledgements

We acknowledge the financial contribution of the Minas Gerais State Research Foundation (FAPEMIG), the Brazil National Council of Technological and Scientific Development (CNPq), the Coordination for the Improvement of Higher Education Personal (CAPES – Finance Code 001, CAPES/PRINT Nº 88887.695902/2022-00), the Department of Agricultural Engineering at the Federal University of Viçosa (DEA – UFV) and the Department of Agricultural Sciences at Murdoch University.

References

Amorim, R.S.S., Albuquerque, J.A., Couto, E.G., Kunz, M., Rodrigues, M.F., Silva, L. de C.M. da, Reichert, J.M., 2022. Water retention and availability in Brazilian Cerrado (neotropical savanna) soils under agricultural use: Pedotransfer functions and decision trees. Soil Tillage Res 224. https://doi.org/10.1016/j.still.2022.105485

Armindo, R.A., Jong van Lier, Q. de, Turek, M.E., Abreu de Melo, M.L., Huf dos Reis, A.M., Palma Ramos, M.E., Ono, G.M., 2019. Performance of the groenevelt and grant model for fitting soil water retention data from brazilian soils. Rev Bras Cienc Solo 43. https://doi.org/10.1590/18069657rbcs20180217

Assouline, S., Or, D., 2013. Conceptual and Parametric Representation of Soil Hydraulic Properties: A Review. Vadose Zone Journal 12, 1–20. https://doi.org/10.2136/vzj2013.07.0121

Babaeian, E., Homaee, M., Vereecken, H., Montzka, C., Norouzi, A.A., van Genuchten, M.Th., 2015. A Comparative Study of Multiple Approaches for Predicting the Soil-Water Retention Curve: Hyperspectral Information vs. Basic Soil Properties. Soil Science Society of America Journal 79, 1043–1058. https://doi.org/10.2136/sssaj2014.09.0355

Bell, R.W., Oliveira, T.S., 2022. Subsoil Constraints for Crop Production: Recent Advances, New Technologies, and Priorities for Further Research, in: Subsoil Constraints for Crop Production. Springer International Publishing, pp. 397–450. https://doi.org/10.1007/978-3-031-00317-2

Betti, G., Grant, C., Churchman, G., Murray, R., 2015. Increased profile wettability in texture-contrast soils from clay delving: Case studies in South Australia. Soil Research 53, 125–136. https://doi.org/10.1071/SR14133

Betti, G., Grant, C.D., Murray, R.S., Churchman, G.J., 2016. Size of subsoil clods affects soil-water availability in sand-clay mixtures. Soil Research 54, 276–290. https://doi.org/10.1071/SR15115

Blanco-Canqui, H., Ruis, S.J., Francis, C.A., 2023. Do organic farming practices improve soil physical properties? Soil Use Manag. https://doi.org/10.1111/sum.12999

Brooks, R.H., Corey, A.T., 1964. Hydraulic Properties of Porous Media. Hydrol Pap 3. 37p.

Brutsaert, W., 1966. Probability laws for pore-size distributions. Soil Sci 101, 85–92.

Campbell, G.S., 1974. A simple method for determining unsaturated conductivity from moisture retention data. Soil Sci 117, 311–314. https://doi.org/10.1097/00010694-197406000-00001

Cann, M.A., n.d. Clay spreading on water repellent sands in the south east of South Australia-promoting sustainable agriculture.

Chen, L., Ghorbani, J., Zhang, C., Dutta, T.T., Kodikara, J., 2021. A novel unified model for volumetric hardening and water retention in unsaturated soils. Comput Geotech 140. https://doi.org/10.1016/j.compgeo.2021.104446

Cornelis, W.M., Khlosi, M., Hartmann, R., Van Meirvenne, M., De Vos, B., 2005. Comparison of Unimodal Analytical Expressions for the Soil-Water Retention Curve. Soil Science Society of America Journal 69, 1902–1911. https://doi.org/10.2136/sssaj2004.0238

Dane, J.H., Topp, C.G., 2018. Methods of Soil Analysis, Part 4: Physical Methods. John Wiley & Sons, Vol. 20.

Davenport, David., Hughes, B., Davies, S., Hall, D., 2011. Spread, delve, spade, invert : a best practice guide to the addition of clay to sandy soils. Grains Research and Development Corporation.

Dexter, A.R., 2004. Soil physical quality Part I. Theory, effects of soil texture, density, and organic matter, and effects on root growth. Geoderma 120, 201–214. https://doi.org/10.1016/j.geodermaa.2003.09.005

Dick, J., Tetzlaff, D., Bradford, J., Soulsby, C., 2018. Using repeat electrical resistivity surveys to assess heterogeneity in soil moisture dynamics under contrasting vegetation types. J Hydrol (Amst) 559, 684–697. https://doi.org/10.1016/j.jhydrol.2018.02.062

Djajadi, Abbott, L.K., Hinz, C., 2012. Synergistic impacts of clay and organic matter on structural and biological properties of a sandy soil. Geoderma 183–184, 19–24. https://doi.org/10.1016/j.geoderma.2012.03.012

Dobarco, M.R., Cousin, I., Le Bas, C., Martin, M.P., 2019. Pedotransfer functions for predicting available water capacity in French soils, their applicability domain and associated uncertainty. Geoderma 336, 81–95. https://doi.org/10.1016/j.geoderma.2018.08.022

Du, C., 2020. Comparison of the performance of 22 models describing soil water retention curves from saturation to oven dryness. Vadose Zone Journal 19. https://doi.org/10.1002/vzj2.20072

Fontes, G.P., Greer, K.D., Pittelkow, C.M., 2024. Does biochar improve nitrogen use efficiency in maize? GCB Bioenergy 16. https://doi.org/10.1111/gcbb.13122

Groenevelt, P.H., Grant, C.D., 2004. A new model for the soil-water retention curve that solves the problem of residual water contents. Eur J Soil Sci. https://doi.org/10.1111/j.1365-2389.2004.00617.x

Gubiani, P.I., Reichert, J.M., Campbell, C., Reinert, D.J., Gelain, N.S., 2013. Assessing Errors and Accuracy in Dew-Point Potentiometer and Pressure Plate Extractor Measurements. Soil Science Society of America Journal 77, 19–24. https://doi.org/10.2136/sssaj2012.0024

Hall, D.J.M., Davies, S.L., Bell, R.W., Edwards, T.J., 2020. Soil management systems to overcome multiple constraints for dryland crops on deep sands in a water limited environment on the south coast of western australia. Agronomy 10. https://doi.org/10.3390/agronomy10121881

Hallema, D.W., Périard, Y., Lafond, J.A., Gumiere, S.J., Caron, J., 2015. Characterization of Water Retention Curves for a Series of Cultivated Histosols. Vadose Zone Journal 14, 1–8. https://doi.org/10.2136/vzj2014.10.0148

Jensen, D.K., Tuller, M., de Jonge, L.W., Arthur, E., Moldrup, P., 2015. A New Two-Stage Approach to predicting the soil water characteristic from saturation to oven-dryness. J Hydrol (Amst) 521, 498–507. https://doi.org/10.1016/j.jhydrol.2014.12.018

Klute, A., 1986. Water retention: laboratory methods, Methods of soil analysis: part 1 physical and mineralogical methods.

Kosugi, K., 1996. Lognormal distribution model for unsaturated soil hydraulic properties. Water Resour Res 32, 2697–2703. https://doi.org/10.1029/96WR01776

Kotlar, A.M., de Jong van Lier, Q., de Souza Brito, E., 2020. Pedotransfer functions for water contents at specific pressure heads of silty soils from Amazon rainforest. Geoderma 361. https://doi.org/10.1016/j.geoderma.2019.114098

Lal, R., Shukla, M.K., 2007. Principles of Soil Physics. The Ohio State University Columbus, Ohio, USA. 699p.

Li, P., Zha, Y., Zuo, B., Zhang, Y., 2023. A Family of Soil Water Retention Models Based on Sigmoid Functions. Water Resour Res 59. https://doi.org/10.1029/2022WR033160

Liao, K., Xu, S., Wu, J., Zhu, Q., 2014. Uncertainty analysis for large-scale prediction of the van Genuchten soil-water retention parameters with pedotransfer functions. Soil Research 52, 431–442. https://doi.org/10.1071/SR13230

Madi, R., Huibert De Rooij, G., Mielenz, H., Mai, J., 2018. Parametric soil water retention models: A critical evaluation of expressions for the full moisture range. Hydrol Earth Syst Sci 22, 1193–1219. https://doi.org/10.5194/hess-22-1193-2018

Masters, B., 2014. A review and summary of information pertaining to the modification of sandy soils in the broadacre agriculture zones of Southern Australia. Government of South Australia, Department of Environment, Water and Natural Resources. (Rural Solutions SA: Adelaide)

McBratney, A.B., Minasny, B., Cattle, S.R., Vervoort, R.W., 2002. From pedotransfer functions to soil inference systems Geoderma 109, 41-73.

McKissock, I., Gilkes, R.J., Walker, E.L., 2002. The reduction of water repellency by added clay is influenced by clay and soil properties. Appl Clay Sci 20, 225–241.

Morgan, M.T., Holmes, R.G., Wood, R.K., 1993. A system for measuring soil physical properties in the field, Soil & Tillage Research 26, 301-325.

Mualem, Y., 1976. A New Model for Predicting the Hydraulic Conductivity of Unsaturated Porous Media. Water Resources Research 12, 513-522.

Pan, T., Hou, S., Liu, Y., Tan, Q., 2019. Comparison of three models fitting the soil water retention curves in a degraded alpine meadow region. Sci Rep 9. https://doi.org/10.1038/s41598-019-54449-8

Quentin, S., Philippe, L., 2021. Uncertainty assessment of soil available water capacity using error propagation: A test in Languedoc-Roussillon. Geoderma 391. https://doi.org/10.1016/j.geoderma.2021.114968

Rahmati, O., Falah, F., Dayal, K.S., Deo, R.C., Mohammadi, F., Biggs, T., Moghaddam, D.D., Naghibi, S.A., Bui, D.T., 2020. Machine learning approaches for spatial modeling of agricultural droughts in the south-east region of Queensland Australia. Science of the Total Environment 699. https://doi.org/10.1016/j.scitotenv.2019.134230

Razzaghi, F., Obour, P.B., Arthur, E., 2020. Does biochar improve soil water retention? A systematic review and meta-analysis. Geoderma. https://doi.org/10.1016/j.geoderma.2019.114055

Roper, M.M., Davies, S.L., Blackwell, P.S., Hall, D.J.M., Bakker, D.M., Jongepier, R., Ward, P.R., 2015. Management options for water-repellent soils in Australian dryland agriculture. Soil Research 53, 786–806. https://doi.org/10.1071/SR14330

Ross, P.J., Williams, J., Bristow, K.L., 1991. Equation for Extending Water-Retention Curves to Dryness. Soil Science Society of America Journal 55, 923–927.

Rudiyanto, Minasny, B., Shah, R.M., Setiawan, B.I., van Genuchten, M.T., 2020. Simple functions for describing soil water retention and the unsaturated hydraulic conductivity from saturation to complete dryness. J Hydrol (Amst) 588. https://doi.org/10.1016/j.jhydrol.2020.125041

Ruthrof, K.X., Hopkins, A.J.M., Danks, M., O’Hara, G., Bell, R., Henry, D., Standish, R., Tibbett, M., Howieson, J., Burgess, T., Harper, R., 2019. Rethinking soil water repellency and its management. Plant Ecol 220, 977–984. https://doi.org/10.1007/s11258-019-00967-4

Scanlan, C.A., Davies, S.L., 2019. Soil mixing and redistribution by strategic deep tillage in a sandy soil. Soil Tillage Res 185, 139–145. https://doi.org/10.1016/j.still.2018.09.008

Schaap, M.G., Leij, F.J., Van Genuchten, M.T., Brown, G.E., 2001. ROSETTA: a computer program for estimating soil hydraulic parameters with hierarchical pedotransfer functions. J Hydrol (Amst) 251, 163–176.

Schapel, A., Bell, R., Yeap, S., Hall, D., 2023. Sandy soil constraints: Organic and clay amendments to improve the productivity of sandy soils, in: Soil Constraints and Productivity. pp. 1–46. Bolan, N., and Kirkham, M.B. (Eds.). CRC Press. https://doi.org/10.1201/9781003093565

Schapel, A., Marschner, P., Churchman, J., 2018. Clay amount and distribution influence organic carbon content in sand with subsoil clay addition. Soil Tillage Res 184, 253–260. https://doi.org/10.1016/j.still.2018.08.001

Seki, K., Toride, N., Van Genuchten, M.T., 2023. Evaluation of a general model for multimodal unsaturated soil hydraulic properties. Journal of Hydrology and Hydromechanics 71, 22–34. https://doi.org/10.2478/johh-2022-0039

Shukla, M.K., 2014. Soil Physics An Introduction. CRC Press, Taylor & Francies Group. 466p.

Silva, A.R., Lima, R.P., 2022. Package “soilphysics” Type Package Title Soil Physical Analysis.

Silva, B.M., Silva, É.A. da, Oliveira, G.C. de, Ferreira, M.M., Serafim, M.E., 2014. Plant-available soil water capacity: estimation methods and implications. Rev. Bras. Ci. Solo 38, 464–475.

Silva, L. de C.M. da, Amorim, R.S.S., Fernandes Filho, E.I., Bocuti, E.D., Silva, D.D., 2023a. Pedotransfer functions and machine learning: Advancements and challenges in tropical soils. Geoderma Regional. https://doi.org/10.1016/j.geodrs.2023.e00720

Silva, L. de C.M. da, Peixoto, D.S., Azevedo, R.P., Avanzi, J.C., Dias Junior, M. de S., Vanella, D., Consoli, S., Acuña-Guzman, S.F., Borghi, E., Resende, Á.V. de, Silva, B.M., 2023b. Assessment of soil water content variability using electrical resistivity imaging in an Oxisol under conservation cropping systems. Geoderma Regional 33. https://doi.org/10.1016/j.geodrs.2023.e00624

Silva, L. de C.M. da, Avanzi, J.C., Peixoto, D.S., Merlo, M.N., Borghi, E., Resende, Á.V. de, Acuña-Guzman, S.F., Silva, B.M., 2021. Ecological intensification of cropping systems enhances soil functions, mitigates soil erosion, and promotes crop resilience to dry spells in the Brazilian Cerrado. International Soil and Water Conservation Research 9, 591–604. https://doi.org/10.1016/j.iswcr.2021.06.006

Silva, R.F., Moura, E.G., Liska, G.R., Oliveira, G.C., Ferreira, D.F., Dias Junior, M.S., Santos, G.G., 2022. Modelling soil water retention to identify management influence on soil pore characteristics. An Acad Bras Cienc 94. https://doi.org/10.1590/0001-3765202220190944

Smettem, K.R.J., Gregory, P.J., 1996. The relation between soil water retention and particle size distribution parameters for some predominantly sandy Western Australian soils. Aust. J. Soil Res 34, 695–708.

Stankovich·, J.M., Lockington, D.A., 1995. Brooks-Corey and van Genuchten Soil-Water-Retention Models. Journal of Irrigation and Drainage Engineering 121, 1–7.

Streck, T., Weber, T.K.D., 2020. Analytical expressions for noncapillary soil water retention based on popular capillary retention models. Vadose Zone Journal 19. https://doi.org/10.1002/vzj2.20042

Tennant, D., Scholz, G., Dixon, J., Purdie, B., 1992. Physical and chemical characteristics of duplex soils and their distribution in the south-west of Western Australia, Australian Journal of Experimental Agriculture. Australian Journal of Experimental Agriculture 32, 827-843.

Too, V.K., Omuto, C.T., Biamah, E.K., Obiero, J.P., 2014. Review of Soil Water Retention Characteristic (SWRC) Models between Saturation and Oven Dryness. Open Journal of Modern Hydrology 04, 173–182. https://doi.org/10.4236/ojmh.2014.44017

van Genuchten, M.T., 1980. A Closed-form Equation for Predicting the Hydraulic Conductivity of Unsaturated Soils. Soil Science Society of America Journal 44, 892–898.

van Genuchten, M.T., Nielsen, D.R., 1985. On Describing and Predicting the Hydraulic Properties of Unsaturated Soils. Ann Geophys 3, 615–628.

van Lier, Q. de J., 2020. Física do Solo baseada em processos. Universidade de São Paulo, Piracicaba. 413p.

Vogel, T., Van Genuchten, M.T., Cislerova, M., Brown, G.E., 2001. Effect of the shape of the soil hydraulic functions near saturation on variably-saturated flow predictions. Adv Water Resour 24, 133–144.

Wassar, F., Gandolfi, C., Rienzner, M., Chiaradia, E.A., Bernardoni, E., 2016. Predicted and measured soil retention curve parameters in Lombardy region north of Italy. International Soil and Water Conservation Research 4, 207–214. https://doi.org/10.1016/j.iswcr.2016.05.005

Zhang, L., Wang, J., Bai, Z., Lv, C., 2015. Effects of vegetation on runoff and soil erosion on reclaimed land in an opencast coal-mine dump in a loess area. Catena (Amst) 128, 44–53. https://doi.org/10.1016/j.catena.2015.01.016

7. Paper IV: Subsoil clay and amendment incorporation: impact on water dynamics in sandy soils

Lucas de Castro Moreira da Silva, Richard Bell, Luca De Prato, Wendy Vance, Ricardo Santos Silva Amorim, Elpídio Inácio Fernandes-Filho

Abstract

Sandy soils pose several challenges to rainfed agriculture owing to constraints such as low water holding capacity, water repellency, and susceptibility to wind erosion. In regions where rainfall is scarce, such as western and southern Australia, these challenges are particularly pronounced and result in crop yields well below the water-limited potential. In this context, many soil management practices have been employed to overcome such limitations, including deep tillage, subsoil clay incorporation, and the application of organic and mineral amendments. Although these practices may lead to yield benefits, they cause considerable physical modifications in soil profiles, creating heterogeneous soil profiles distinct from those formed under natural conditions. The consequences of these practices for enhancing soil water retention and availability are not well understood. Therefore, this study aimed to assess hydrophysical properties of sandy soils under agricultural use in the Western Australian Wheatbelt and evaluate the impacts of different soil management practices on soil pore size distribution and available water capacity. The results showed that subsoil clay incorporation significantly increased total porosity and microporosity but did not improve water availability. The application of organic and mineral amendments did not exhibit significant differences in soil pore size distribution or soil water retention. Conversely, the disturbance caused by deep tillage reduced the plant-water available capacity, possibly due to a loosened soil structure that harmed pore connectivity. In contrast with tillage, soil amendments may cause gradual modifications on soil porosity and water availability over time. Therefore, further investigations into the long-term effects of organic and inorganic amendments on the water availability of sandy soils are warranted.

Keywords: Duplex soils, plant-available water capacity, organic amendment, zeolite, clay incorporation

1. Introduction

Regions with characteristic dry periods require careful soil management to increase water availability and maintain agricultural productivity without degrading natural resources (Rahmati et al., 2020; Silva et al., 2021). Across western and southern Australia, many hectares of sandy soils are used for rainfed agricultural production. These soils, in addition to the region's inherent rainfall scarcity, pose several challenges to agriculture owing to their low water-holding capacity (Hall et al., 2020; Tennant et al., 1992), subsoil constraints (Bell and Oliveira, 2022), topsoil water repellency (Ruthrof et al., 2019), and susceptibility to wind erosion (Lowe et al., 2021; Roper et al., 2015). These limitations result in patchy seed germination, poor weed control, low soil coverage rates (Harper et al., 2000) and yield well below the water-limited yield potential (Hall et al., 2020). In this sense, several soil management practices have been tested to improve the water use in these agricultural fields in an attempt to increase root growth and access stored water in deeper soil layers.

Many of the 16 million hectares of sandy soils in western and southern Australia overlie clay-rich subsoils, allowing engineering interventions aiming to raise clay to the topsoil, such as delving, deep tillage and subsequent turning and mixing of soil layers (Betti et al., 2016; Cann, 2000; McKissock et al., 2002; Scanlan and Davies, 2019). In addition to subsoil clay, organic and mineral amendments have also been used (Schapel et al., 2023), and their association with deep tillage aims to mix the amendments into the sandy top layer (Betti et al., 2015; Schapel et al., 2018). The incorporation of amendments in sandy soils results in several physical, chemical, and biological benefits, such as increased soil aggregate stability (Djajadi et al., 2012b), decreased water repellency, increased water holding capacity (Hallett, 2008; Ruthrof et al., 2019), increased organic carbon, increased cation exchange capacity, increased nutrient retention and increased microbial biomass (Hallett, 2008; Hall et al. 2010; Müller and Deurer, 2011; Roper et al., 2015). Regarding water availability, Mohawesh and Durner (2019) concluded that mineral and organic amendments significantly improved water retention in a sandy soil, although emphasizing the need for further assessments on a field scale.

Despite the considerable potential of soil amelioration practices to improve productivity (Bell and Oliveira, 2022), they are not feasible in all regions, because of high operation costs (Hallett, 2008), unavailability, or insufficiency of material (e.g., subsoil clay) to provide the expected benefits (Unkovich et al., 2020). Furthermore, the consequences of modifying soil profiles through amelioration practices are not fully understood (Roper et al., 2015). The addition of subsoil clay does not homogeneously change the soil texture, as clay clods of diverse sizes are incorporated into the sandy matrix (Schapel et al., 2019). Hence, management practices based on deep tillage with clay delving and amendments incorporation can lead to a heterogeneous soil profile (Hall et al., 2020; Scanlan & Davies, 2019), which may exhibit different characteristics from typical soils formed under natural conditions.

The water dynamics in these modified soil profiles remains unclear and several doubts regarding their hydrophysical behavior have been raised in recent studies. Schapel et al. (2018) showed distinct results of soil water availability depending on the concentration of subsoil clay added to a sandy soil. Unkovich et al. (2020) concluded that the amount of clay required to reduce topsoil water repellency may be insufficient to improve soil water holding capacity. Betti et al. (2016) observed that the hydrological pattern of clay-amended sandy soils could not be predicted based only on texture, as similar-textured soils had different available water capacities and water retention curve shapes. The same authors also found that water retention capacity varied according to different sizes of clay clods. These findings are an important knowledge gap concerning water dynamics in sandy soils after the incorporation of subsoil clay, organic or mineral amendments.

In this context, this study assessed the hydrophysical properties of sandy soils used for agriculture in the Western Australian Wheatbelt. Common practices employed in the region, such as subsoil clay incorporation and organic and mineral amendments, were analyzed. Soil water retention curves and pore size distribution were determined in soil modified with deep tillage and amendment addition to evaluate their effects on plant-available water capacity. The outcomes of this study aimed to fill a gap in understanding the water dynamics in sandy soils modified by subsoil clay, mineral and organic amendments associated with strategic deep tillage, in places where rainwater supply is limited.

2. Material and Methods

2.1 Experimental sites

This study was conducted at two experimental fields within the Wheatbelt region of Western Australia: Bullaring (S32°28'37.8", E117°37'00.3") and South Kweda (S32º25’40”, E117º24’18”). The Bullaring experimental area was established in June 2022 on a deep sandy soil paddock. The South Kweda experimental field was established in May 2022 on a pale, deep non-gravelly sand overlying a kaolin-dominant horizon at a depth of 0.4-0.6m. Both experimental areas had four blocks of different soil management practices. The climate in the region is temperate with wet winters (June-August) and dry hot summers (December-March). Monthly average temperatures range from 10 to 22º C and the average annual rainfall of 440 mm falls mostly between April and October.

A description of the soil management practices is shown in Table 1. At Bullaring, three soil management treatments were assessed: control (i.e., without tillage or amendment application - B1), deep tillage using Bednar machinery (B2), and deep tillage with Bednar machinery with subsoil clay incorporation at a rate of 100 t of subsoil/ha (B3). Four replicates were set up, forming 12 experimental plots of 30 m length and 6.4 m width. At South Kweda, four soil management treatmens were assessed: control (K1), deep tillage using Bednar machinery (K2), incorporation of humiclay compost “Cwise” (10 t/ha) followed by delving and incorporation to 0.35 m using Bednar machinery (K3), and incorporation of aluminosilicate mineral zeolite at a rate of 30 t/ha followed by delving and incorporation to 0.35 m (K4). Treatments were repeated four times, forming 16 experimental plots in total: 8 plots had dimensions of 18 m width and 200 m length for the first two treatments and 8 plots had dimensions of 18m width and 2 m length for the two latter treatments. Soils from both sites were under barley cultivation in 2022.

Table 1. Description of soil management practices.

Soil managements	Description (2022 cropping season)
B1	Barley crop without tillage or amendment application
B2	Barley crop with deep tillage
B3	Barley crop with deep tillage + clay-rich subsoil incorporation (100 t/ha)
K1	Barley crop without tillage or amendment application
K2	Barley crop with deep tillage
K3	Barley crop with deep tillage + humiclay compost incorporation (10 t/ha)
K4	Barley crop with deep tillage + zeolite incorporation (30 t/ha)

2.2 Soil sampling and laboratory analyses

Fieldwork for soil sampling were done between the 2022 harvest and the 2023 sowing season. Disturbed and undisturbed samples were collected from each block at three soil depths (0-10 cm, 20-30 cm and 30-40 cm). In total, 84 undisturbed and 84 disturbed samples were taken. Metallic cores with a height of 5 cm and a diameter of 3.5 cm were using for collecting undisturbed soil samples. Disturbed samples were used to perform particle size distribution (soil texture) analysis following the pipette method (Dane and Topp, 2018). For samples from soils with clay incorporation, clay clods were ground until they could pass through a sieve with a 2 mm aperture. An image of clay clods within the sandy matrix in a clay-amended soil is shown in Figure 1a and clay clods sampled are presented in Figure 1b. The results of the soil particle size distribution are shown in Table 2.

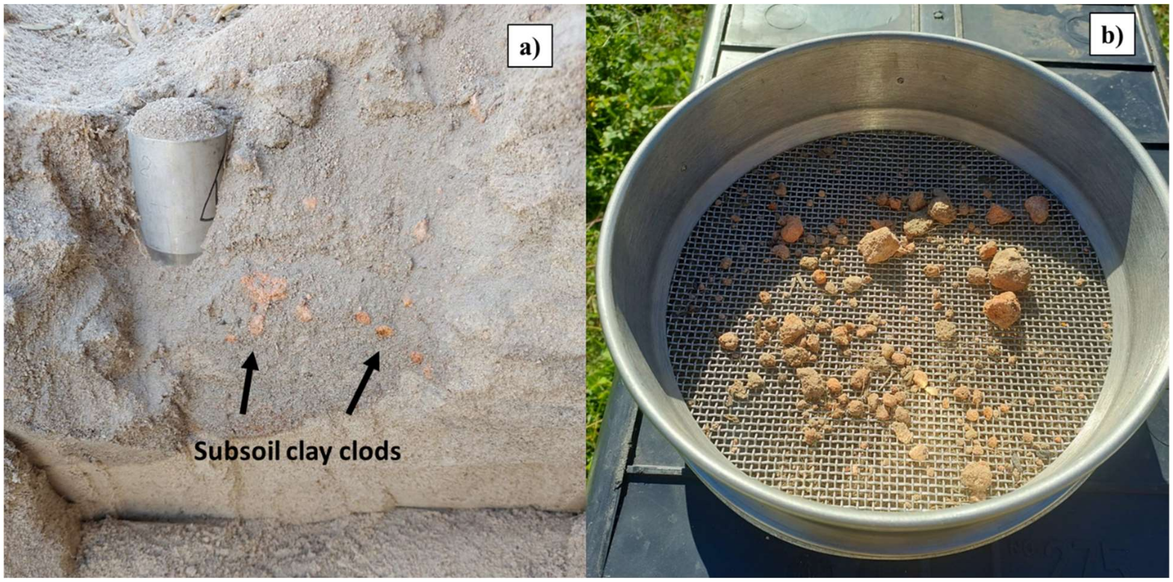

Figure 1. Subsoil clay clods distributed within the sandy top layer of the clay-amended soil (a). Different sizes of subsoil clay clods sampled from B3 soil management (b). The reader is referred to Table 1 for the soil management description.

Table 2. Particle size distribution of soil management treatments at three different soil depths.

Soil management	Depth 0-10 cm			
	Sand	Silt	Clay	Bd
	x-----	%	-----	Mg m^{-3}
B1	97.8	0.9	1.3	1.68
B2	97.0	0.4	2.6	1.63
B3	91.6	2.0	6.3	1.63
K1	96.4	1.3	2.4	1.49
K2	97.0	1.4	1.6	1.46
K3	94.5	4.1	1.4	1.43
K4	96.5	2.2	1.3	1.42

Soil management	Depth 20-30 cm			
	Sand	Silt	Clay	Bd
	x-----	%	-----	Mg m^{-3}
B1	97.0	2.6	0.3	1.63
B2	98.6	1.0	0.5	1.66
B3	97.6	1.0	1.3	1.66
K1	96.9	1.4	1.7	1.68
K2	93.6	2.2	4.2	1.66
K3	94.1	1.6	4.3	1.66
K4	96.1	0.9	3.0	1.63

Soil management	Depth 30-40 cm			
	Sand	Silt	Clay	Bd
	x-----	%	-----	Mg m^{-3}
B1	96.7	0.8	2.5	1.63
B2	97.7	0.9	1.4	1.62
B3	92.8	1.4	5.7	1.60
K1	93.0	1.3	5.8	1.76
K2	90.2	0.5	9.3	1.76
K3	86.6	1.6	11.8	1.75
K4	97.3	0.5	2.2	1.73

For soil management descriptions, the reader is referred to Table 1. Bd = Bulk density.

Undisturbed samples were saturated and equilibrated to matric potentials of -4, -6, -10, -33 and -100 kPa in a Richards' pressure chamber (Klute, 1986). Water content was determined for each matric potential and samples were then dried in an oven at 105 º C for 48 hours to determine soil bulk density (Bd). Gravimetric water content at potentials of -500 and -1500 kPa was determined using disturbed samples in a WP4 psychrometer (Gubiani et al., 2013). Finally, volumetric water content was obtained by multiplying the gravimetric water content by the bulk density.

The topsoil water repellency was determined by the Molarity of Ethanol Droplet (MED) test (King, 1981). The topsoil samples were dried in an oven (< 70º) for 48 h and cooled at room temperature. Then, samples were placed in Petri dishes to ensure a flat surface before adding water or aqueous ethanol solution drops. However, distilled water infiltrated in all samples in less than 5 seconds, indicating no issues concerning topsoil water repellency. Therefore, we assumed that water repellency did not merit further investigation in this study.

2.3 Data processing and statistical analysis

Soil water retention curves were fitted using the Brooks and Corey (1964) model. This model was compared with other popular soil water retention models in a previous study and was the most accurate for the soils under analysis (Silva et al., 2024). To provide a visualization of model uncertainty for each soil management, the bootstrap resampling method was used to generate a 95% confidence interval based on 100 iterations. Brooks and Corey model is expressed as:

$$\theta(h) = \theta r + (\theta s - \theta r)(\alpha h)^{-\lambda}$$

where θ is the volumetric water content (m^{-3} m^{-3}); h is the matric potential (hPa); θs is the saturated water content (m^{-3} m^{-3}); θr is the residual water content (m^{-3} m^{-3}); α and λ are empirical shape parameters.

Soil water retention curve data was used to calculate the following soil properties: total porosity, macroporosity, mesoporosity, microporosity, plant-available water capacity and relative field capacity. The plant-available water capacity was considered as the water content interval between matric potentials of -10 kPa (field capacity) and -1500 kPa (permanent wilting point). The relative field capacity was calculated according to Reynolds et al. (2008):

$$\mathrm{RFC} = \mathrm{FC}/\mathrm{Tp}$$

where RFC is the Relative field capacity (dimensionless), FC is the field capacity at a matric potential of -10 kPa (m^3 m^{-3}), and Tp is the total porosity (m^3 m^{-3}).

The pore diameter was determined according to Bouma (1991):

$$D = 4\ \sigma\ \mathrm{Cos}\ \theta/\Psi m$$

where D is the pore diameter (μm), σ is the water surface tension (73.43 kPa at 20 ◦C), θ is the contact angle between the meniscus and the capillary wall (0) and Ψm is the matric potential (kPa).

Pore size distribution was classified according to Kay (1990) and (Lal and Shukla, 2007), as follows: macropores refer to pores larger than 30 μm, mesopores are between 30 and 0.2 μm, and micropores are pores with sizes lower than 0.2 μm. Based on this classification, mesopores are associated with plant-available water capacity (White, 2006). Relative macroporosity, mesoporosity, and microporosity were determined using a proportional rule with anchoring total porosity at 100%.

The experiment was designed in blocks with four replicates. After verifying the assumptions, ANOVA was conducted followed by Tukey's test ($p < 0.05$) to compare soil properties among the different soil management systems.

3. Results and Discussion

3.1 Soil water retention curves

Soil water retention curves for soil management treatments of the Bullaring experimental site (B1, B2 and B3) are displayed in Figure 2. Considering all treatments and depths, water content at saturation ranged from 0.35 to 0.52 m^{-3} m^{-3}, water content at field capacity (-10 kPa) varied from 0.08 to 0.14 m^{-3} m^{-3}, and at permanent wilting point (-1500 kPa) between 0.00 and 0.03 m^{-3} m^{-3}. The results indicate that most of the water held in these soils is gravitational water, that is, the water content between saturation and field capacity (Liebig et al., 1995). This is a typical pattern observed in sandy soils with high hydraulic conductivity. These coarse-textured soils predominantly have larger pores, where water is primarily influenced by gravitational energy rather than capillary forces and does not remain available for plant uptake once the pores drain (Carvalho et al., 2020).

Overall, water retention curves were very similar across all treatments and depths, except for B3 at 0-10 cm (Fig. 2c). This treatment can be distinguished from other soil management systems within the same depth of analysis. B3 had the highest water content at saturation, highest field capacity, and highest permanent wilting point. This difference is explained by the addition of subsoil clay (see Table 1), which affects the pore size distribution of this treatment. B3 has 6% clay at 0-10 cm, B1 has 1%, and B2 has 2.6%. Even though clay is not homogeneously distributed throughout the soil profile (see Fig. 1), small clay contents among coarser particles may increase capillary forces (Pollacco, 2008), thereby raising soil water retention at field capacity. Schapel et al. (2018) also found increased water retention at -10 kPa for clay amended soils. Nevertheless, adsorptive forces are also increased by adding clay to the sandy matrix, which is confirmed by the increased water content at the permanent wilting point of the B3 treatment. Thus, increasing field capacity does not necessarily mean an improvement in available water, as a similar increase in permanent wilting point may occur. At lower potentials (more negative) than -1500 kPa, water is held in very small pores with a high adsorptive energy, thus not easily available for plant uptake (Silva et al., 2022).

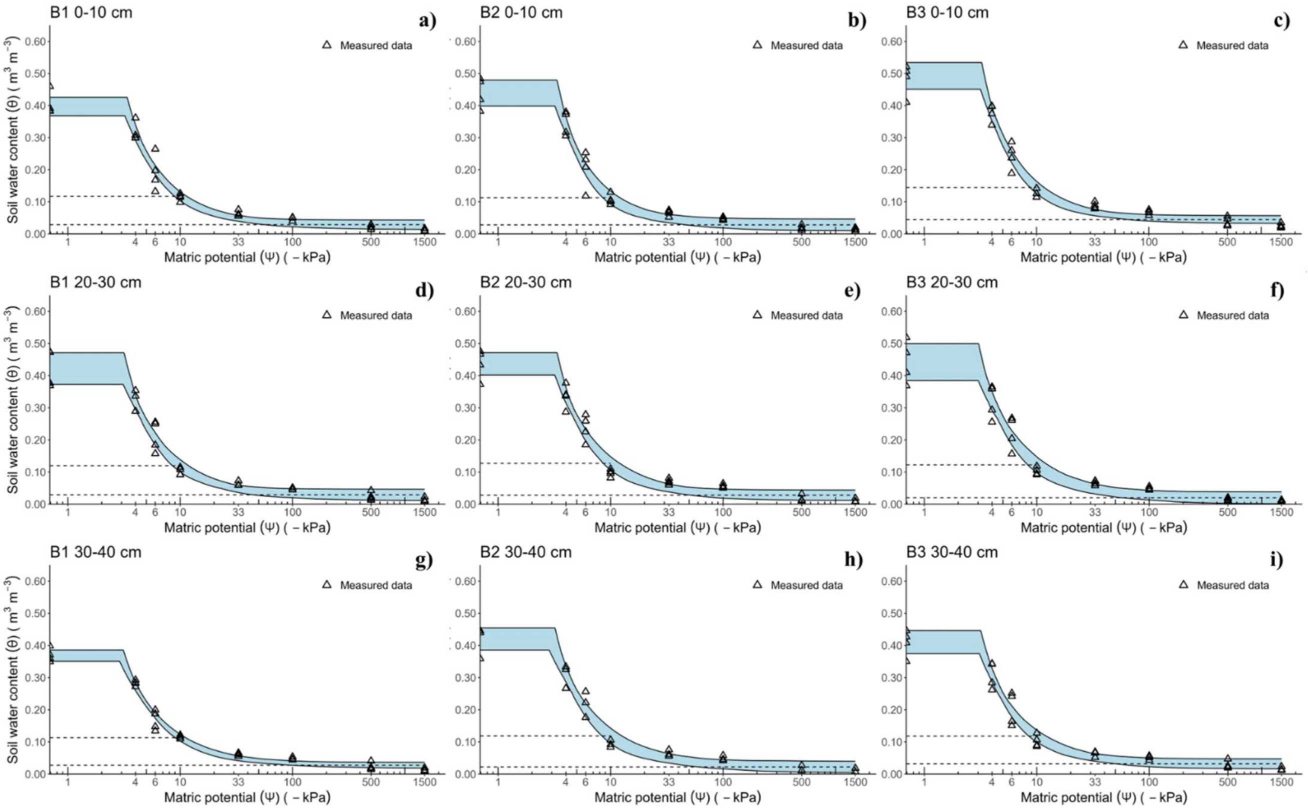

Figure 2. Soil water retention curves for each soil management of the Bullaring site at three different depths (0-10, 20-30, and 30-40 cm). B1, B2 and B3 refer to soil management treatments. For soil management descriptions, refer to Table 1. Dashed lines mark the water content at field capacity (-10 kPa) and permanent wilting point (-1500 kPa). The blue range represents the 95% confidence interval generated by the bootstrap resampling method.

In contrast with the results from the Bullaring site, soil water retention curves from the Kweda site (Figure 3) exhibited distinct patterns across depths. Regardless of soil management treatments, all soils had a clear difference in curve slopes between the topsoil (0-10 cm) and the deeper layer (30-40 cm). Considering the topsoil, all management treatments demonstrated notably steep curve slopes, indicating a rapid drainage rate and a substantial change in water content between saturation and field capacity. Conversely, at 30-40 cm depth, the curves had a gentler slope, suggesting a less abrupt transition of pore sizes compared to the top layer. This result is aligned with differences in clay content (Table 2), as all management treatments showed increased clay content in depth. The most pronounced difference in slope was observed for K3, which has ~1% of clay at the topsoil and ~12% of clay at 30-40 cm.

Water content at saturation varied from 0.48 to 0.62 m^{-3} m^{-3} in the top layer (0-10 cm), from 0.40 to 0.51 m^{-3} m^{-3} at the 20-30 cm depth, and from 0.36 to 0.43 m^{-3} m^{-3} in the deeper layer (30 – 40 cm). Field capacity ranged between 0.05 and 0.13 m^{-3} m^{-3} at 0-10 cm depth, from 0.06 to 0.17 m^{-3} m^{-3} at 20-30 cm, and from 0.07 to 0.22 m^{-3} m^{-3} at the 30-40 cm layer. Permanent wilting points exhibited similar patterns ranging from 0.01 to 0.03 m^{-3} m^{-3} at the top layer, from 0.01 to 0.07 m^{-3} m^{-3} at 20-30 cm, and between 0.02 and 0.11 m^{-3} m^{-3} at the deeper layer (30-40 cm depth). Considering all treatments, the results demonstrated a soil structure distinction between the top and the deeper layer as the topsoil exhibits lower bulk densities and greater overall volume of voids. This indicates that the deeper layer has a denser structural arrangement, which may be beneficial in terms of water retention compared with the very loose top layer. This was confirmed by the gradual decrease in total porosity (related to water content at saturation) and an increase in field capacity and permanent wilting point in depth.

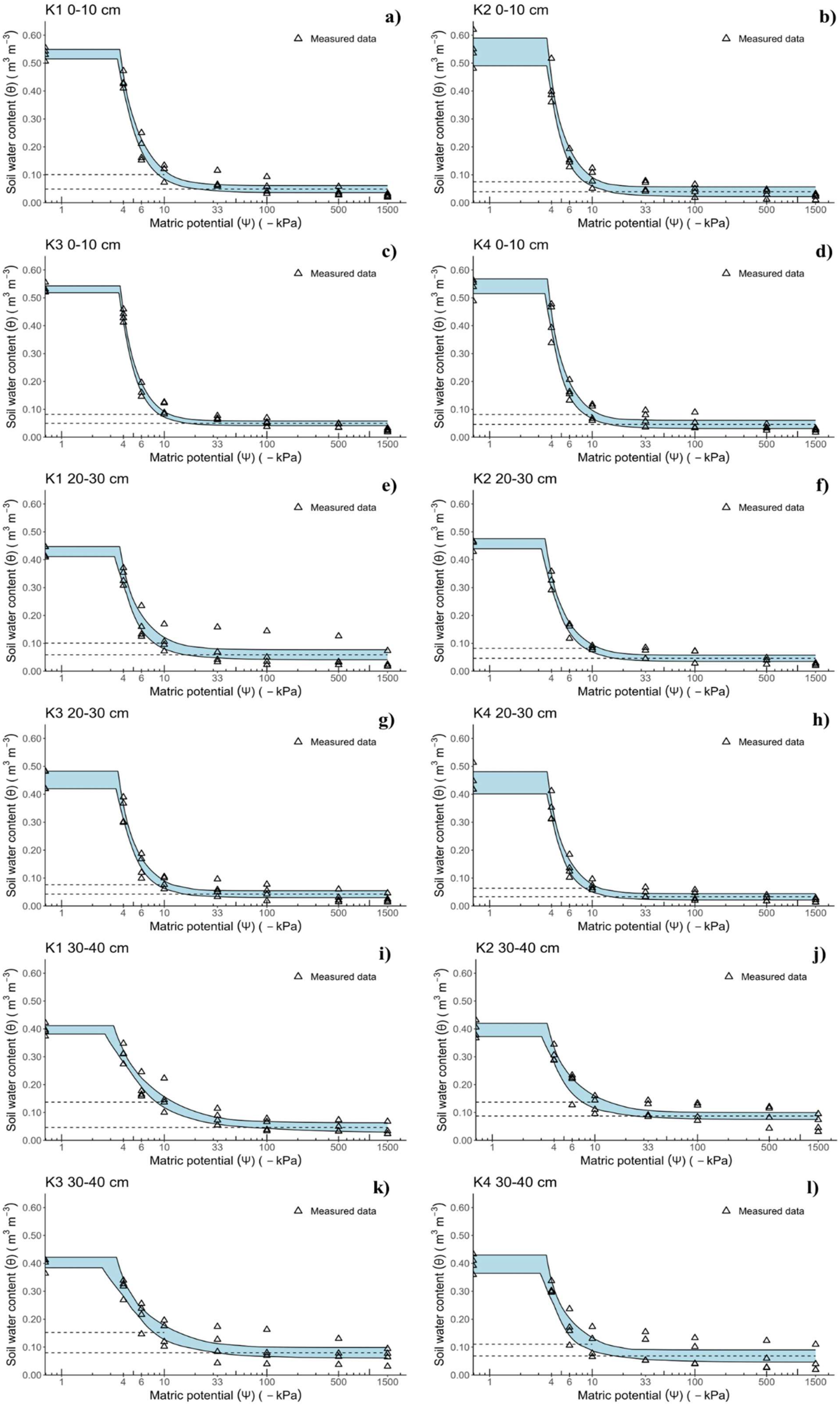
K1 0-10 cm
a)
K2 0-10 cm
b)
K3 0-10 cm
c)
K4 0-10 cm
d)
K1 20-30 cm
e)
K2 20-30 cm
f)
K3 20-30 cm
g)
K4 20-30 cm
h)
K1 30-40 cm
i)
K2 30-40 cm
j)
K3 30-40 cm
k)
K4 30-40 cm
l)
△ Measured data
Soil water content (θ) (m³ m⁻³)
0.00
0.10
0.20
0.30
0.40
0.50
0.60
Matric potential (Ψ) (− kPa)
1
4
6
10
33
100
500
1500

Figure 3. Soil water retention curves for each soil management treatment at the Kweda site in three different depths (0-10, 20-30, and 30-40 cm). K1, K2, K3 and K4 refer to soil management treatments. For soil management description, refer to Table 1. Dashed lines mark the water contents at field capacity (-10 kPa) and permanent wilting point (-1500 kPa). The blue range represents the 95% confidence interval generated by the bootstrap resampling method.

3.2 Pore size distribution

The total porosity and relative field capacity of soil management treatments at Bullaring and Kweda sites are presented in Figure 4 and Figure 5, respectively. At Bullaring site, in the topsoil (0-10 cm), B3 had higher total porosity than B1. An increase in total porosity is expected by the addition of subsoil clay in this treatment. Clay fills voids between coarser particles (Bradford and Blanchar, 1999) and its aggregates promote higher total porosity due to the presence of small pores within and among them. This result is consistent with what was observed in the soil water retention curves, where B3 (Fig. 2c) demonstrated higher amounts of water content at field capacity and permanent wilting point. By promoting an increase in total porosity through the formation of smaller pores, clay amendments enhance the capillary and adsorptive forces, which are responsible for water retention (Jensen et al., 2015). Soil management treatments at Kweda site did not show evidence of changes in total porosity.

Relative field capacity (RFC) did indicate any difference among treatments for both experimental sites. Reynolds et al. (2008) describe the RFC as the capacity of a soil to store water and air relative to its total porosity. The optimal range of RFC proposed by the authors is between 0.60 and 0.70 (dimensionless). All soil treatments at Bullaring (i.e., B1, B2 and B3) showed RFC values below half of the lower limit for the optimal range. Soil management treatments at Kweda site (i.e., K1, K2, K3 and K4) exhibited RFC values below 0.40. Therefore, soils from both sites were considered as water-limited based on this indicator, and none of the management practices significantly improved it. It is noteworthy that Reynolds et al. (2008) validated the RFC optimal range on a clay loam soil, and low values were expected in this study owing to the intrinsic characteristics of sandy soils. Nevertheless, this indicator serves as a guideline for agricultural lands and has been adopted to assess soil quality in coarser-textured soils (e.g., Nascimento et al., 2019).

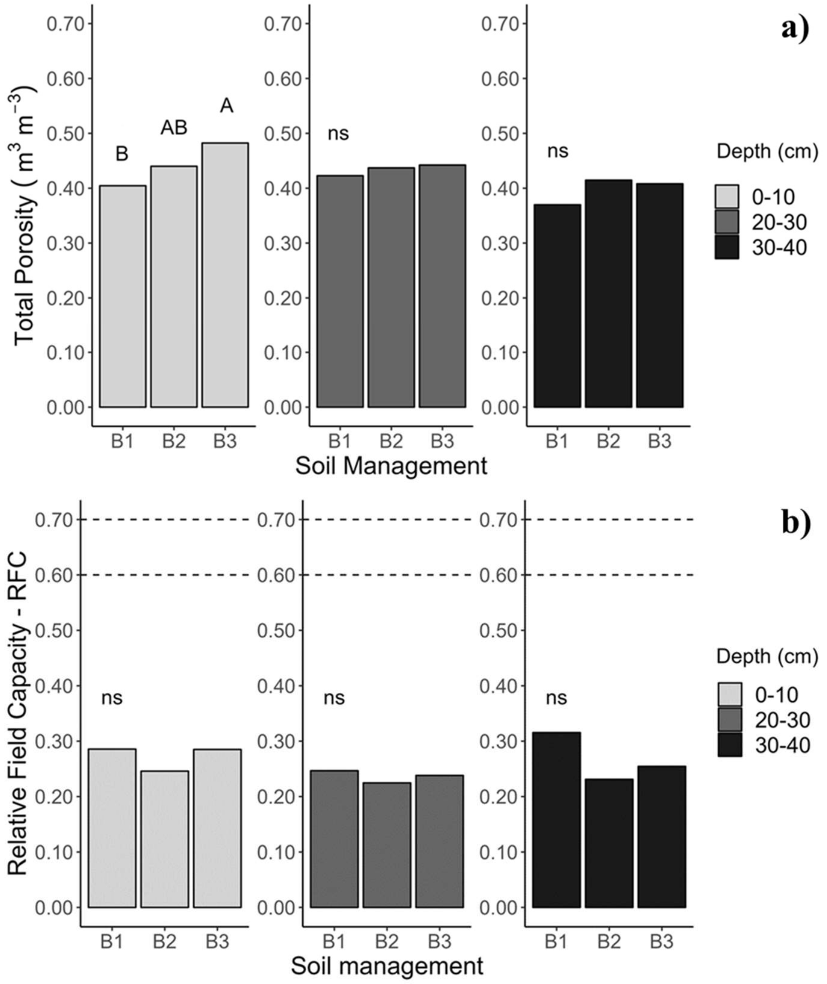

Figure 4. Total porosity (a) and Relative field capacity (b) of soil management treatments at the Bullaring experimental site in three depths (0-10, 20-30, and 30-40 cm). Bars with different letters on the top differ according to Tukey's test ($p<0.05$). ns = not significant. B1, B2 and B3 refer to soil management treatments. For soil management description, refer to Table 1. Dashed lines represent the optimal range of Relative field capacity defined by Reynolds et al. (2008).

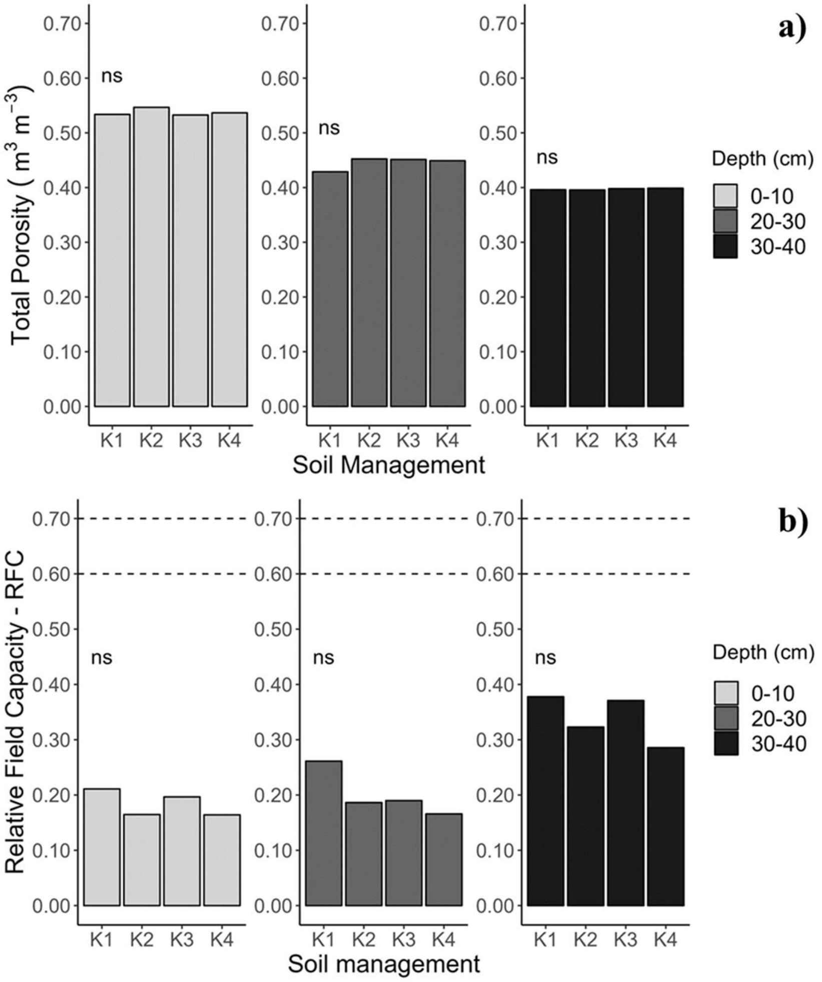

Figure 5. Total porosity (a) and Relative field capacity (b) of soil management treatments at the Kweda experimental site in three different depths (0-10, 20-30, and 30-40 cm). Bars with different letters on the top differ according to Tukey's test ($p<0.05$). ns = not significant. K1, K2, K3 and K4 refer to soil management treatments. For soil management description, refer to Table 1. Dashed lines represent the optimal range of Relative field capacity defined by Reynolds et al. (2008).

The pore size distribution for soil management treatments at Bullaring is displayed in Figure 6. Overall, more than 70% of their total porosity consisting of macropores and more than 20% was formed by mesopores, at all depths of analysis. Micropores accounted for a small portion of the total porosity, ranging from 1% to 5%. No significant differences were found for either macropores or mesopores, indicating that soil management treatments have similar drainage and aeration capacity, which are associated with macroporosity, and similar plant-available water capacity, which is related to mesopores as per the classification used.

Microporosity significantly differed for B3 compared to the other management treatments (i.e., B1 and B2) in the topsoil layer (0-10 cm). In B3, 5% of its total porosity was constituted by micropores, whereas B1 and B2 showed 3% of microporosity each. This difference was previously noted in soil water retention curves, where B3 exhibited higher field capacity and permanent wilting point than other treatments within the same depth of analysis. This increased microporosity is due to the addition of subsoil clay in the B3 treatment. However, microporosity is related to "textural porosity" (Carducci et al., 2013; Dexter, 2004), where water is retained with high adsorptive energy in very small pores, considered not readily available to plants ((Lal and Shukla, 2007; Silva et al., 2014; Silva et al., 2022). Therefore, although the B3 treatment provided a significant change in pore size distribution compared to unamended management systems, benefits of subsoil clay incorporation towards improving water availability were inconclusive.

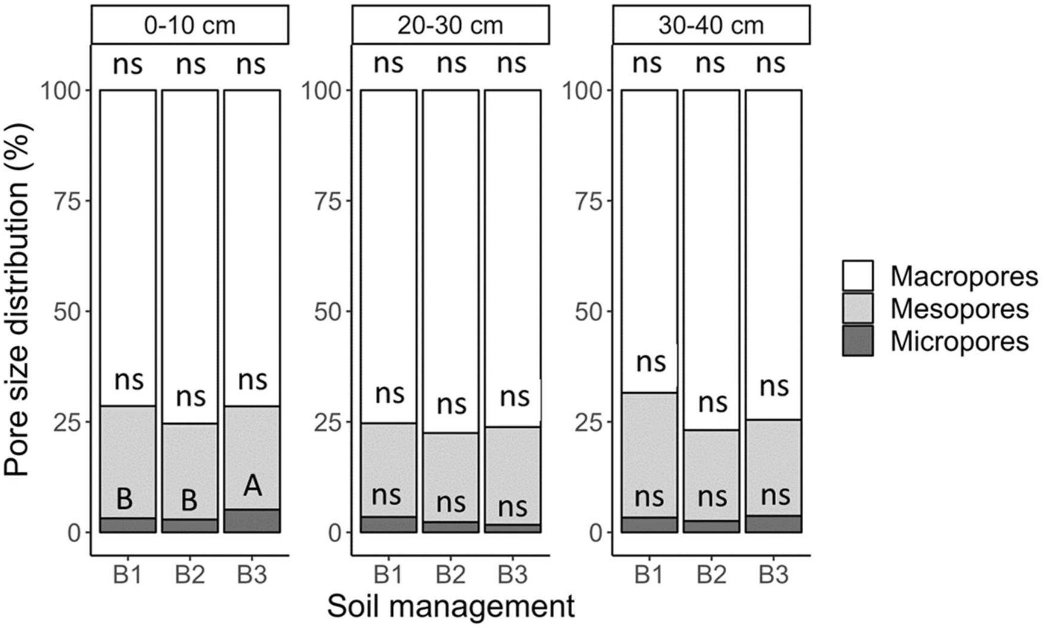

Figure 6. Pore size distribution of soil management treatments at the Bullaring experimental site in three depths (0-10, 20-30, and 30-40 cm). Macropores refer to pore diameters higher than 30 μm, mesopores have diameters between 30 and 0.2 μm, and micropores have diameters <0.2 μm. The results are relative to the total porosity, anchored at 100%. Bars with different letters on the top differ according to Tukey's test ($p<0.05$). ns = not significant. B1, B2 and B3 refer to soil management treatments. For soil management description, refer to Table 1.

Figure 7 displays the pore size distribution of soil management treatments at Kweda. In all soil management treatments, ~75% of porosity was formed by macropores. No differences were found at the first two depths of evaluation (0-10 and 20-30 cm) for any of the pore size classes. Nevertheless, differences in pore size distribution were observed in the deeper layer (30-40 cm). K1, which is the control, i.e., without any amendment application or tillage operation, had the highest relative mesoporosity. According to the pore size classification used in this study, mesoporosity is related to plant-available water. Therefore, this result indicates that K1 had a higher capacity to hold water available to plants than K2 and K4, which were subjected to deep tillage with Bednar machinery, and K4 was also amended with zeolite. K3 did not differ from K1, K2, or K4.

The higher mesoporosity in K1 at 30-40 cm depth suggests that the tillage led to a decline in the class of pores responsible for retaining readily available water. It is well known that a slight compaction may have positive effects in soils with high proportion of macropores (Arvidsson, 1999; Blanco-Canqui and Ruis, 2018; Głab, 2014). Light compaction induces a structural rearrangement and can improve pore connectivity in loose-structured soils (Blanco-Canqui, 2017), leading to heightened capillary forces and consequently increased mesoporosity (Silva et al., 2022). Thus, as the experimental area was recently established, it is plausible that a subsoil compaction either natural (Hall et al., 2020) or resulting from routine machinery traffic (Hamza and Anderson, 2005), had a positive effect on K1 mesoporosity. However, the disturbance caused by the Bednar machinery in K2 and K4 possibly loosened the soil structure, hindering pore connectivity and culminating in lower levels of relative mesoporosity.

The application of organic and mineral amendments (K3 and K4, respectively) did not provide conclusive findings with respect to pore size distribution and effects of tillage were more pronounced at this stage. However, soil reconsolidation is likely to occur after the first years of tillage (Hall et al., 2010), diminishing the differences between ripped and unripped treatments (Hall et al., 2020), making the potential impacts of amendments more noticeable. Furthermore, the advantages of applying organic and inorganic amendments may appear over time (Abiven et al., 2009; Jarosz et al., 2022; Zheng et al., 2018), and the long-term outcomes of these practices merit further investigation. It is also possible that higher rates of amendments are needed to achieve significant change in porosity. For example, Ippolito et al. (2011) found increased water retention between -100 and -300 kPa applying zeolite at a rate of 44.8 t ha^{-1} in a coarse-textured soil and no changes were found for a rate of 13.4 t ha^{-1}.

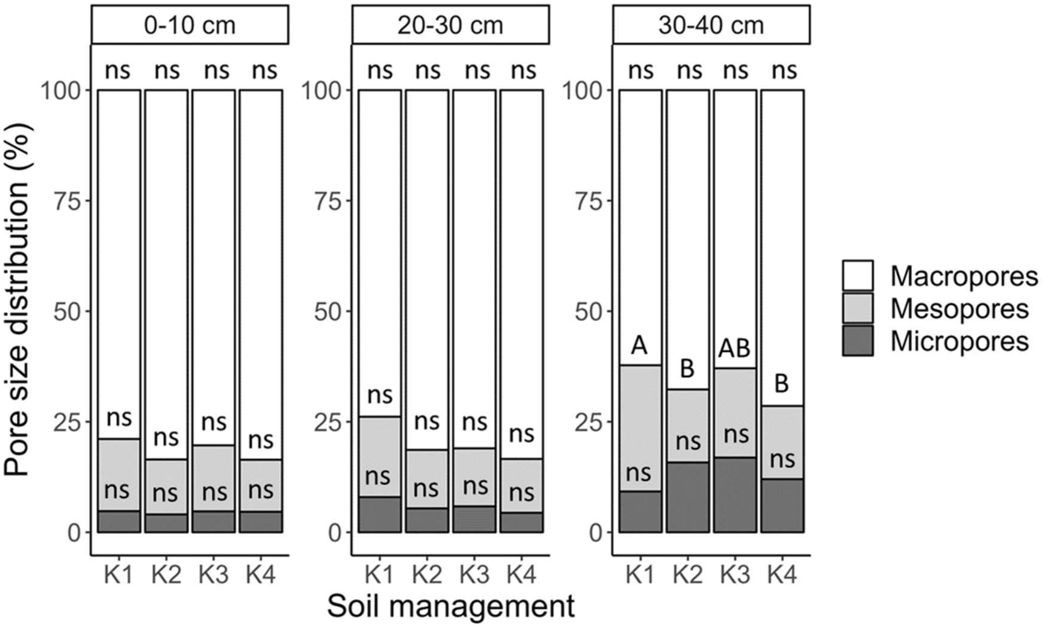

Figure 7. Pore size distribution of soil management treatments at Kweda at three depths (0-10, 20-30, and 30-40 cm). Macropores refer to pore diameters larger than 30 µm, mesopores have diameters between 30 and 0.2 µm, and micropores have diameters <0.2 µm. Results are relative to total porosity, anchored at 100%. Bars with different letters on top differ by Tukey's test ($p<0.05$). ns = not significant. K1, K2, K3 and K4 refer to soil management treatments. For soil management description, refer to Table 1.

3.3 Plant-available water capacity (PAWC)

PAWC values for all soil management treatments at Bullaring site fell within the "limited" and "poor" classes defined by White (2006), across all depths (Figure 8). PAWC values between 0.10 and 0.15 m^{-3} m^{-3} denotes limited available water, while PAWC < 0.10 m^{-3} m^{-3} indicates drought-prone soils. No significant difference was found among treatments, although the data distribution exposed by boxplots showed patterns that warrant further discussion. In the top layer (0-10 cm), B3 was fully within the limits of "limited" or "moderate" PAWC (White, 2006; Reynolds et al., 2008), whereas B1 and B2 had part of their quantiles lying below the poor PAWC limit. Conversely, in the deeper layer (30-40 cm) B1 showed ~75% of its distribution falling within the moderate PAWC class while at least 75% of B2 and B3 distribution fell into the poor PAWC category.

Although no significant differences were found for PAWC among treatments, results for B3 in the topsoil (0-10 cm) offered additional insights into its effects on soil water retention, which can be attributed to the higher clay content (Table 2) resulting from the incorporation of subsoil clay. In comparison to unamended soils, B3 exhibited increased water retention at field capacity and permanent wilting point (Fig. 2c), significantly higher total porosity (Fig. 4a), and significantly higher relative microporosity (Fig. 6). Despite the lack of statistical significance regarding the effects of subsoil clay incorporation on PAWC, these results merit further investigation. Future studies addressing this topic could benefit from increasing the sample size to enhance the statistical power. This could potentially provide more solid evidence either to confirm the advantages of soil management on PAWC or to refute them. Notably, Schapel et al. (2018) also reported higher water retention at field capacity in clay-amended soils. However, if clay amendments result in a proportional increase in water retained at the permanent wilting point, as shown in Fig. 2c, the range of available water remains narrow.

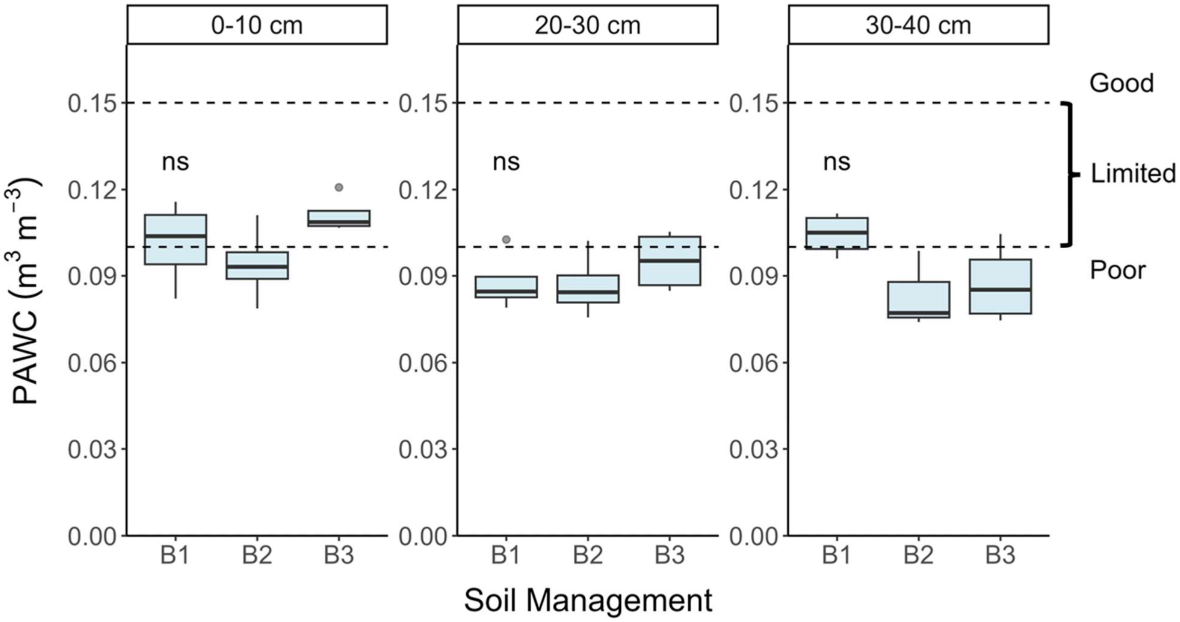

Figure 8. Plant-available water capacity (PAWC) of soil management treatments at Bullaring at three depths (0-10, 20-30, and 30-40 cm). Bars with different letters on top differ by Tukey's test ($p<0.05$). ns = not significant. Dashed lines represent PAWC limits: "Good" = $0.15 \leq \text{PAWC} < 0.20$ m^{-3} m^{-3}, "Limited" = $0.10 \leq \text{PAWC} < 0.15$ m^{-3} m^{-3}, "Poor" = $\text{PAWC} < 0.10$ m^{-3} m^{-3} (White, 2006). B1, B2 and B3 refer to soil management treatments. For soil management description, refer to Table 1.

PAWC results for soil management treatments at Kweda are provided in Figure 9. Overall, for the first two depths of analysis (0-10 and 20-30 cm), soil management treatments had values within the “poor” PAWC class. This indicates that all treatments fell within the category of drought-prone soils at these depths, regardless of soil management practices. In the topsoil (0-10 cm), K1 had more than 75% of its distribution below the boundary between limited and poor PAWC and K2, K3 and K4 had 100% of their PAWC values considered as poor. In the deeper soil layer (30-40 cm), K1 showed higher values of PAWC, with more than 75% of its distribution falling within the limited PAWC range, and a small portion lying above the limit between limited and good PAWC. In contrast, K2 and K4 had 100% of their PAWC distribution within the poor PAWC class, K3 had 75% considered as poor, and 25% categorized as limited PAWC.

Significant differences observed for PAWC at 30-40 cm were similar to those found for relative mesoporosity (Fig. 7). According to the pore size classification used in this study, they essentially convey the same information, except that the relative microporosity is conditioned by the total porosity, while the data presented in Fig. 9 express the measured PAWC values. As previously discussed, tillage with Bednar may have compromised pore connectivity and possibly have broken soil aggregates, which strongly influence capillary forces (Nguyen et al., 2014). Thus, by disrupting the pores responsible for holding available water (Silva et al., 2023), known as structural porosity (Carducci et al., 2013; Dexter, 2004), tillage practices can potentially reduce the PAWC. It is noteworthy that strategic tillage is performed primarily to alleviate compaction, and positive effects appear when penetration resistance is restrictive to plants growth (Hall et al., 2020). Nevertheless, despite this study did not assess soil penetration resistance, it is plausible to infer that soil strength was not an issue in the evaluated sites, as shown by high macroporosity rates (Fig. 6 and 7). Therefore, strategic tillage showed a negative effect in this study, reducing water availability.

Finally, deep tillage associated with organic amendment (K3) and zeolite incorporation (K4) did not improve the PAWC. K3 did not differ from K1 or from other soil management system, probably because of the high data variance. The experimental area was recently established, which could explain the absence of amendment effects and the pronounced influence of tillage on PAWC. Tillage has an immediate impact on soil structure whereas amendments may take longer to promote evident changes. For instance, clay-amended soils can increase organic carbon over time (Schapel et al., 2018), which has a great potential to improve soil structure by enhancing soil aggregation (Cui et al., 2014). In the context of organic amendments (K3 management), high-quality carbon material provides slow but long-term improvements in aggregate stability (Fang et al., 2021). Similarly, mineral amendments such as zeolites (K4 management) also provide long-term effects by modifying soil texture (Nakhli et al., 2017). Many chemical benefits achieved through the application of zeolites are mentioned in the reviews by Jarosz et al. (2022) and Nakhli et al. (2017). However, further studies are needed to assess the effectiveness of zeolite on water availability, as well as testing different rates of application. For instance, Ippolito et al. (2011) studied a sandy soil and did not observe improvements in water retention with 13.4 t ha^{-1} of zeolite, whereas Ravali et al. (2020) found increased water retention in a sandy loam soil with zeolite application in a dose of 7.5 t ha^{-1}.

Hence, the results of this study highlight the need for further investigation into the long-term effects of organic and inorganic amendments on water availability in sandy soils. Given that soil re-compaction is likely to occur eventually (Hall et al., 2020), new insights can be gained from these amended re-consolidated soils over time, especially regarding structural rearrangements induced by crops. Soil-root interactions lead to modifications in soil porosity and aggregation through several physical and biological processes, such as wetting and drying cycles (Materechera et al., 1992), formation of biopores (Silva et al., 2021; 2022), release of root exudates (Habib et al., 1990), and root-fungi associations (Tisdall and Oades, 1979), which directly influence soil structure and, thereby, soil water availability.

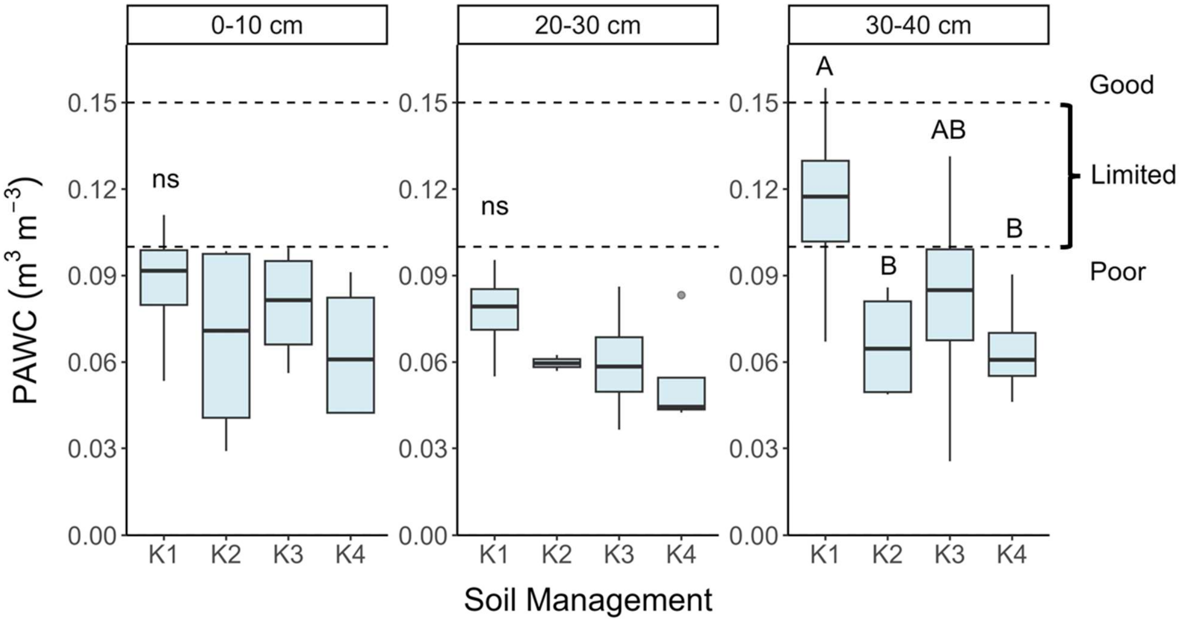

Figure 9. Plant-available water capacity (PAWC) of soil management treatments at Kweda at three depths (0-10, 20-30, and 30-40 cm). Bars with different letters on the top differ according to Tukey's test ($p<0.05$). ns = not significant. Dashed lines represent PAWC limits: "Good" = $0.15 \le PAWC < 0.20$ m^{-3} m^{-3}, "Limited" = $0.10 \le PAWC < 0.15$ m^{-3} m^{-3}, "Poor" =

PAWC < 0.10 m^{-3} m^{-3} (White, 2006). K1, K2, K3 and K4 refer to soil management treatments. For soil management description, refer to Table 1.

4. Conclusions

This study assessed common soil management practices employed in the Western Australian Wheatbelt to overcome the intrinsic constraints of sandy soils. Deep tillage in combination with subsoil clay incorporation and organic and inorganic amendments were compared for their effectiveness in enhancing soil water retention and availability at two experimental sites.

Soil water retention curves and pore size distribution showed that, regardless of management practices, the majority of their porosity consisted of macropores. Subsoil clay incorporation significantly increased the total porosity and microporosity in the topsoil (0-10 cm). However, no evidence emerged indicating improvement in water availability, as soil water retention rose proportionally at field capacity and permanent wilting point. Although no significant differences were found for available water, the modifications in soil porosity caused by subsoil clay incorporation warrant further investigation. Future studies addressing this topic may provide more solid evidence if larger sample size is used to enhance statistical power.

The incorporation of organic and mineral amendments did not exhibit modifications in soil pore size distribution and available water in the upper soil layers (0-10 and 20-30 cm), but significant differences were found among treatments in the subsurface (30-40 cm). The control (i.e., without tillage or amendment application) had significantly higher values of mesoporosity and plant-available water capacity compared to treatments subjected to deep tillage. This is explained by the immediate impact tillage has on loosening the soil structure, possibly breaking aggregates and disrupting pore connectivity. Given that soil reconsolidation eventually occurs, amended soils may demonstrate clearer effects over time, especially regarding structural rearrangements induced by crops. Soil-plant-microbiota interactions can be boosted by organic and mineral additives, guiding a gradual modification of soil porosity and aggregation, thereby influencing water availability. Therefore, further investigations into the long-term effects of organic and inorganic amendments on water availability of sandy soils are warranted.

Acknowledgements

The authors acknowledge the financial contribution of the Coordination for the Improvement of Higher Education Personal (CAPES – Finance Code 001, CAPES/PRINT Nº 88887.695902/2022-00), the Department of Agricultural Engineering at the Federal University of Viçosa (DEA – UFV) and the Department of Agricultural Sciences at Murdoch University.

References

Abiven, S., Menasseri, S., Chenu, C., 2009. The effects of organic inputs over time on soil aggregate stability - A literature analysis. Soil Biol Biochem 41, 1–12.

Arvidsson, J., 1999. Nutrient uptake and growth of barley as affected by soil compaction, Plant and Soil.

Bell, R.W., Oliveira, T.S., 2022. Subsoil Constraints for Crop Production: Recent Advances, New Technologies, and Priorities for Further Research, in: Subsoil Constraints for Crop Production. Springer International Publishing, pp. 397–450. https://doi.org/10.1007/978-3-031-00317-2

Betti, G., Grant, C., Churchman, G., Murray, R., 2015. Increased profile wettability in texture-contrast soils from clay delving: Case studies in South Australia. Soil Research 53, 125–136. https://doi.org/10.1071/SR14133

Betti, G., Grant, C.D., Murray, R.S., Churchman, G.J., 2016. Size of subsoil clods affects soil-water availability in sand-clay mixtures. Soil Research 54, 276–290. https://doi.org/10.1071/SR15115

Blanco-Canqui, H., 2017. Biochar and Soil Physical Properties. Soil Science Society of America Journal 81, 687–711. https://doi.org/10.2136/sssaj2017.01.0017

Blanco-Canqui, H., Ruis, S.J., 2018. No-tillage and soil physical environment. Geoderma 326, 164–200. https://doi.org/10.1016/j.geoderma.2018.03.011

Bouma, J., 1991. Influence of Soil Macroporosity on Environmental Quality. Advances in Agronomy 46, 1–37.

Bradford, J.M., Blanchar, R.W., 1999. Mineralogy and Water Quality Parameters in Rill Erosion of Clay-Sand Mixtures. Soil Science Society of America Journal 63, 1300–1307.

Brooks, R.H., Corey, A.T., 1964. Hydraulic Properties of Porous Media. Hydrology Papers 3.

Cann, M.A., 2000. Clay spreading on water repellent sands in the south east of South Australia promoting sustainable agriculture. J Hydrol (Amst) 231–232, 333–341. https://doi.org/10.1016/S0022-1694(00)00205-5

Carducci, C.E., Oliveira, G.C., Zeviani, W.M., Lima, V.M.P., Serafim, M.E., 2013. Bimodal pore distribution on soils under conservation management system for coffee crop. Engenharia Agrícola 33, 291–302.

Carvalho, M.L., de Moraes, M.T., Cerri, C.E.P., Cherubin, M.R., 2020. Biochar amendment enhances water retention in a tropical sandy soil. Agriculture (Switzerland) 10. https://doi.org/10.3390/agriculture10030062

Cui, J., Askari, M.S., Holden, N.M., 2014. Relation Soil Carbon and Soil Structure to Land Use Management, in: Soil Carbon. pp. 297–304.

Dane, J.H., Topp, C.G., 2018. Methods of Soil Analysis Part 4 Physical Methods.

Dexter, A.R., 2004. Soil physical quality Part I. Theory, effects of soil texture, density, and organic matter, and effects on root growth. Geoderma 120, 201–214. https://doi.org/10.1016/j.geodermaa.2003.09.005

Djajadi, Abbott, L.K., Hinz, C., 2012a. Synergistic impacts of clay and organic matter on structural and biological properties of a sandy soil. Geoderma 183–184, 19–24. https://doi.org/10.1016/j.geoderma.2012.03.012

Djajadi, Abbott, L.K., Hinz, C., 2012b. Synergistic impacts of clay and organic matter on structural and biological properties of a sandy soil. Geoderma 183–184, 19–24. https://doi.org/10.1016/j.geoderma.2012.03.012

Fang, Y., Singh, B.P., Zwieten, L. V, Collins, D., Pitt, W., Armstrong, R., Tavakkoli, E., 2021. Additive effects of organic and inorganic amendments can significantly improve structural stability of a sodic dispersive subsoil. Geoderma 404, 115281.

Głab, T., 2014. Effect of soil compaction and N fertilization on soil pore characteristics and physical quality of sandy loam soil under red clover/grass sward. Soil Tillage Res 144, 8–19. https://doi.org/10.1016/j.still.2014.05.010

Gubiani, P.I., Reichert, J.M., Campbell, C., Reinert, D.J., Gelain, N.S., 2013. Assessing Errors and Accuracy in Dew-Point Potentiometer and Pressure Plate Extractor Meaurements. Soil Science Society of America Journal 77, 19–24. https://doi.org/10.2136/sssaj2012.0024

Habib, L., Morel, J.L., Guckert, A., Plantureux, S., Chenu, C., 1990. Influence of root exudates on soil aggregation. Symbiosis 9, 87–91.

Hall, D.J.M., Davies, S.L., Bell, R.W., Edwards, T.J., 2020. Soil management systems to overcome multiple constraints for dryland crops on deep sands in a water limited environment on the south coast of western australia. Agronomy 10. https://doi.org/10.3390/agronomy10121881

Hall, D.J.M., Jones, H.R., Crabtree, W.L., Daniels, T.L., 2010. Claying and deep ripping can increase crop yields and profits on water repellent sands with marginal fertility in southern Western Australia. Australian Journal of Soil Research. https://doi.org/10.1071/SR09078

Hallett, P.D., 2008. A brief overview of the causes, impacts and amelioration of soil water repellency - A review. Soil and Water Research 3, S21–S29. https://doi.org/10.17221/1198-swr

Hamza, M.A., Anderson, W.K., 2005. Soil compaction in cropping systems: A review of the nature, causes and possible solutions. Soil Tillage Res. https://doi.org/10.1016/j.still.2004.08.009

Harper, R.J., McKissock, I., Gilkes, R.J., Carter, D.J., Blackwell, P.S., 2000. A multivariate framework for interpreting the effects of soil properties, soil management and landuse on water repellency. J Hydrol (Amst) 231–232, 371–383. https://doi.org/10.1016/S0022-1694(00)00209-2

Ippolito, J. A., Tarkalson, D. D., & Lehrsch, G. A. (2011). Zeolite soil application method affects inorganic nitrogen, moisture, and corn growth. Soil Science, 176, 136–142.

Jarosz, R., Szerement, J., Gondek, K., Mierzwa-Hersztek, M., 2022. The use of zeolites as an addition to fertilisers – A review. Catena (Amst). https://doi.org/10.1016/j.catena.2022.106125

Kay, B.D., 1990. Rates of Change of Soil Structure Under Different Cropping Systems, in: Advances in Soil Science. pp. 1–41.

King, P.M., 1981. Comparison of methods for measuring severity of water repellence of sandy soils and assessment of some factors that affect its measurement. Soil Research 19, 275–285.

Klute, A., 1986. Water retention: laboratory methods, Methods of soil analysis: part 1 physical and mineralogical methods.

Lal, R., Shukla, M.K., 2007. Principles of Soil Physics. 682p.

Liebig, M.A., Jones, A.J., Doran, J.W., Mielke, L.N., 1995. Potential Soil Respiration and Relationship to Soil Properties in Ridge Tillage. Soil Science Society of America Journal 59, 1430–1435.

Lowe, M.A., McGrath, G., Leopold, M., 2021. The Impact of Soil Water Repellency and Slope upon Runoff and Erosion. Soil Tillage Res 205, 104756. https://doi.org/10.1016/j.still.2020.104756

Materechera, S.A., Dexter, A.R., Alston, A.M., 1992. Formation of aggregates by plant roots in homogenised soils. Plant Soil 142, 69–79.

McKissock, I., Gilkes, R.J., Walker, E.L., 2002. The reduction of water repellency by added clay is influenced by clay and soil properties. Appl Clay Sci 20, 225–241. https://doi.org/10.1016/S0169-1317(01)00074-6

Mohawesh, O., Durner, W., 2019. Effects of Bentonite, Hydrogel and Biochar Amendments on Soil Hydraulic Properties from Saturation to Oven Dryness. Pedosphere 29, 598–607. https://doi.org/10.1016/S1002-0160(17)60426-0

Müller, K., Deurer, M., 2011. Review of the remediation strategies for soil water repellency. Agric Ecosyst Environ 144, 208–221. https://doi.org/10.1016/j.agee.2011.08.008

Nakhli, S.A.A., Delkash, M., Bakhshayesh, B.E., Kazemian, H., 2017. Application of Zeolites for Sustainable Agriculture: a Review on Water and Nutrient Retention. Water Air Soil Pollut 228. https://doi.org/10.1007/s11270-017-3649-1

Nascimento, D.M., Cavalieri-Polizeli, K.M. V, Silva, A.H., Favaretto, N., Parron, L.M., 2019. Soil physical quality under long-term integrated agricultural production systems. Soil Tillage Res 186, 292–299.

Nguyen, P.M., Le, K. Van, Cornelis, W.M., 2014. Using categorical soil structure information to improve soil water retention estimates of tropical delta soils. Soil Research 52, 443–452. https://doi.org/10.1071/SR13256

Pollacco, J.A.P., 2008. A generally applicable pedotransfer function that estimates field capacity and permanent wilting point from soil texture and bulk density. Can J Soil Sci 761–774.

Rahmati, O., Falah, F., Dayal, K.S., Deo, R.C., Mohammadi, F., Biggs, T., Moghaddam, D.D., Naghibi, S.A., Bui, D.T., 2020. Machine learning approaches for spatial modeling of agricultural droughts in the south-east region of Queensland Australia. Science of the Total Environment 699, 134230. https://doi.org/10.1016/j.scitotenv.2019.134230

Ravali, C., Rao, J.K., Anjaiah, T., Suresh, K., 2020. Effect of zeolite on soil physical and physico-chemical properties. J Sci Agric Eng. 776-781.

Reynolds, W.D., Drury, C.F., Yang, X.M., Tan, C.S., 2008. Optimal soil physical quality inferred through structural regression and parameter interactions. Geoderma 146, 466–474. https://doi.org/10.1016/j.geoderma.2008.06.017

Roper, M.M., Davies, S.L., Blackwell, P.S., Hall, D.J.M., Bakker, D.M., Jongepier, R., Ward, P.R., 2015. Management options for water-repellent soils in Australian dryland agriculture. Soil Research 53, 786–806. https://doi.org/10.1071/SR14330

Ruthrof, K.X., Hopkins, A.J.M., Danks, M., O’Hara, G., Bell, R., Henry, D., Standish, R., Tibbett, M., Howieson, J., Burgess, T., Harper, R., 2019a. Rethinking soil water repellency and its management. Plant Ecol 220, 977–984. https://doi.org/10.1007/s11258-019-00967-4

Ruthrof, K.X., Hopkins, A.J.M., Danks, M., O’Hara, G., Bell, R., Henry, D., Standish, R., Tibbett, M., Howieson, J., Burgess, T., Harper, R., 2019b. Rethinking soil water repellency and its management. Plant Ecol 220, 977–984. https://doi.org/10.1007/s11258-019-00967-4

Scanlan, C.A., Davies, S.L., 2019. Soil mixing and redistribution by strategic deep tillage in a sandy soil. Soil Tillage Res 185, 139–145. https://doi.org/10.1016/j.still.2018.09.008

Schapel, A., Bell, R., Yeap, S., Hall, D., 2023. Sandy soil constraints: Organic and clay amendments to improve the productivity of sandy soils, in: Soil Constraints and Productivity. pp. 1–46.

Schapel, A., Marschner, P., Churchman, J., 2019. Influence of clay clod size and number for organic carbon distribution in sandy soil with clay addition. Geoderma 335, 123–132. https://doi.org/10.1016/j.geoderma.2018.08.009

Schapel, A., Marschner, P., Churchman, J., 2018. Clay amount and distribution influence organic carbon content in sand with subsoil clay addition. Soil Tillage Res 184, 253–260. https://doi.org/10.1016/j.still.2018.08.001

Silva, B.M., Silva, É.A. da, Oliveira, G.C. de, Ferreira, M.M., Serafim, M.E., 2014. Plant-available soil water capacity: estimation methods and implications. Rev. Bras. Ci. Solo 38, 464–475.

Silva, L. de C.M. da, Amorim, R.S.S., Fernandes Filho, E.I., Bocuti, E.D., Silva, D.D., 2023. Pedotransfer functions and machine learning: Advancements and challenges in tropical soils. Geoderma Regional. https://doi.org/10.1016/j.geodrs.2023.e00720

Silva, L. de C.M. da, Avanzi, J.C., Peixoto, D.S., Merlo, M.N., Borghi, E., Resende, Á.V. de, Acuña-Guzman, S.F., Silva, B.M., 2021. Ecological intensification of cropping systems enhances soil functions, mitigates soil erosion, and promotes crop resilience to dry spells in the Brazilian Cerrado. International Soil and Water Conservation Research 9, 591–604. https://doi.org/10.1016/j.iswcr.2021.06.006

Silva, L. de C.M. da, Peixoto, D.S., Gomes, J.B.V., Avanzi, J.C., Amorim, R.S.S., Borghi, E., Resende, Á.V. de, Silva, B.M., Mancini, M., Curi, N., 2022. Mineralogy and pore size distribution of clayey Oxisols with granular structure and the effect of management systems. Soil Tillage Res 223. https://doi.org/10.1016/j.still.2022.105479

Tennant, D., Scholz, G., Dixon, J., Purdie, B., 1992. Physical and chemical characteristics of duplex soils and their distribution in the south-west of Western Australia, Australian Journal of Experimental Agriculture.

Tisdall, J.M., Oades, J.M., 1979. Stabilization of Soil Aggregates by the Root Systems of Ryegrass. Aust. J. Soil Res 17, 429–470.

Unkovich, M., McBeath, T., Llewellyn, R., Hall, J., Gupta, V.V.S.R., Macdonald, L.M., 2020. Challenges and opportunities for grain farming on sandy soils of semi-arid south and south-eastern Australia. Soil Research 58, 323–334.

White, R.E., 2006. Principles and Practice of Soil Science: The Soil as a Natural Resource. Blackwell Publishing. 387p.

Zheng, J., Chen, T., Wu, Q., Yu, J., Chen, W., Chen, Y., Siddique, K.H.M., Meng, W., Chi, D., Xia, G., 2018. Effect of zeolite application on phenology, grain yield and grain quality in rice under water stress. Agric Water Manag 206, 214–251.

8. GENERAL CONCLUSIONS

The Paper III evaluated four popular soil water retention models in a specific soil dataset from the Western Australian Wheatbelt. For all scenarios, the Brooks and Corey model exhibited the highest goodness-of-fit, the lowest errors, and a consistent negligible bias. van Genuchten and Groenevelt and Grant models had similar overall performance, however, the van Genuchten model tended to underestimate soil water contents, especially near saturation, while the Groenevelt and Grant model generally overestimated the entire range of soil water contents. No sufficient evidence was achieved to draw conclusions regarding soil management effects on the accuracy of soil water retention curves. Nevertheless, the findings of this study support the importance of comparing different soil water retention curve models when addressing specific soil datasets, which can be particularly critical in water-limited regions, where minor estimation errors may result in significant changes in soil water content, directly impacting water use efficiency.

In Paper IV, common soil management practices adopted in the Western Australian Wheatbelt were assessed. Subsoil clay incorporation significantly increased the total porosity and microporosity in the topsoil (0-10 cm). However, no evidence emerged indicating improvement in water availability, as soil water retention rose proportionally at field capacity and permanent wilting point. Considering the subsurface (30-40 cm), the control (i.e., without tillage or amendment application) had significantly higher values of mesoporosity and plant-available water capacity compared to treatments subjected to deep tillage. This is explained by the immediate impact tillage has on loosening the soil structure, possibly breaking aggregates and disrupting pore connectivity. Effects of organic and inorganic amends on soil water availability were inconclusive. However, as the experimental areas were recently established, their effects may appear over time. Therefore, further investigations into the long-term effects of organic and inorganic amendments on water availability of sandy soils are warranted.