



**LARA MOTA CORINTO**

**GROUNDWATER RECHARGE POTENTIAL AND ITS  
INTERACTION WITH SOIL PHYSICAL QUALITY IN  
BRAZIL**

**LAVRAS – MG  
2025**

**LARA MOTA CORINTO**

**GROUNDWATER RECHARGE POTENTIAL AND ITS INTERACTION WITH SOIL  
PHYSICAL QUALITY IN BRAZIL**

Dissertation presented to the Federal University of Lavras as part of the requirements of the Postgraduate Program in Soil Science, with an area of concentration in Environmental Resources and Land Use, for the Master's degree.

Prof. Dr. Bruno Montoani Silva  
Orientador

Prof. Dr. Junior Cesar Avanzi  
Co-orientador

Dra. Marta Vasconcelos Ottoni  
Co-orientadora

**LAVRAS-MG  
2025**

**Ficha Catalográfica elaborada pelo Sistema de Geração  
de Ficha Catalográfica da Biblioteca Universitária da UFLA, com  
dados informados pelo(a) próprio(a) autor(a).**

Corinto, Lara Mota.

Groundwater recharge potential and its interaction with soil physical quality in  
Brazil / Lara Mota Corinto. - 2025.

84 p. : il.

Orientador: Bruno Montoani Silva

Coorientador: Junior Cesar Avanzi

Coorientadora: Marta Vasconcelos Ottoni

Dissertação (Mestrado Acadêmico) - Universidade Federal de Lavras, 2025.

Bibliografia.

1. Water recharge areas. 2. Ecosystem services. 3. Soil quality index. I. Silva,  
Bruno Montoani. II. Avanzi, Junior Cesar. III. Ottoni, Marta Vasconcelos. IV.  
Universidade Federal de Lavras. V. Título.

**LARA MOTA CORINTO**

**GROUNDWATER RECHARGE POTENTIAL AND ITS INTERACTION WITH SOIL  
PHYSICAL QUALITY IN BRAZIL**

**POTENCIAL DE RECARGA DE ÁGUA SUBTERRÂNEA E SUA INTERAÇÃO COM  
A QUALIDADE FÍSICA DO SOLO NO BRASIL**

Dissertation presented to the Federal University of Lavras as part of the requirements of the Postgraduate Program in Soil Science, with an area of concentration in Environmental Resources and Land Use, for the Master's degree.

APROVADA em 24 de fevereiro de 2025.

Dr. Bruno Montoani Silva UFLA

Dr. Carlos Rogerio De Mello UFLA

Dr. Eduardo Da Costa Severiano IF GOIANO

Prof. Dr. Bruno Montoani Silva  
Orientador

Prof. Dr. Junior Cesar Avanzi  
Co-orientador

Dra. Marta Vasconcelos Ottoni  
Co-orientadora

**LAVRAS-MG  
2025**

*À minha mãe Maria por todo cuidado e amor durante meu caminho até aqui.  
Ao meu pai João por todo apoio e carinho em todas as etapas da minha vida.  
Ao meu irmão Matheus que pela distância não vi crescer, mas é o meu maior tesouro.  
Aos meus avós Maria Rodrigues e Osvaldo pelo exemplo de determinação e força.  
À minha avó Maria Mamides (in memorian) pelo meu maior exemplo de vida.  
Ao meu avô Polinário (in memorian), que nos deixou no início do meu mestrado e que mesmo  
longe me ensinou a beleza da vida, que é a família.  
Ao meu noivo Victor por nunca ter deixado eu desistir.  
DEDICO.*

## ACKNOWLEDGMENTS

I thank God...

To my parents for always supporting and encouraging me at every stage of my life, for their courage in seeing their daughter leave home at an early age, and for never letting them lack anything, regardless of their circumstances.

To my fiancé Victor, who has always been by my side, helping in any way he could, especially at the beginning of my master's degree.

To the people who always guided me and motivated me to build this work, who showed me the beauty of research, Prof. Bruno Montoani, Prof. Junior Cesar, Prof. Nilton Curi, and Marta Ottoni.

To my colleagues at DCS, especially Michel, Victor, Dayane, Olivia, and Samara, for the time spent together and for all the exchanges of knowledge.

To the *Coordenação de Aperfeiçoamento de Pessoal de Nível Superior* (CAPES) for the study grant.

To the Department of Soil Science (DCS/UFLA) for the opportunity to do my master's degree. And to everyone who contributed in some way to this stage of my life...

Thank you so much!

*If you don't know where you're going, any road will take you there*

- Lewis Carroll (Alice in Wonderland)

## RESUMO

Diante da crescente escassez hídrica global e intrincado processo de mudanças climáticas, as águas subterrâneas emergem como recurso fundamental para garantir a qualidade e disponibilidade hídrica. Dessa forma, o presente estudo investiga o papel dos atributos do solo no potencial de recarga de águas subterrâneas, por meio do desenvolvimento e avaliação de Índice de Qualidade do Solo (SQI – *Soil Quality Index*) utilizando dados do *Hydrophysical Database for Brazilian Soils* (HYBRAS). Inicialmente foi desenvolvido o SQI para Potencial de Recarga de Água Subterrânea (SQIra), com base na avaliação de onze atributos físicos do solo, utilizando-se da ferramenta de Análise de Componentes Principais Supervisionada (SPCA) para seleção dos indicadores principais: Condutividade Hidráulica Saturada (Kslab\_log), Porosidade Drenável (DP - *Drainable Porosity*) e Silte. O SQIra demonstrou eficácia ao destacar diferenças na qualidade do solo entre usos da terra e classes de solo. O índice foi validado com dados de Velocidade de Infiltração Básica (BIR) e com dados de deflúvio subterrâneo, destacando sua utilidade na gestão sustentável de bacias hidrográficas. Ademais, o padrão do SQI criado juntamente com outros disponíveis na literatura para o potencial de recarga foi avaliado quanto a influência da classe de solo, classe textural, estrutura funcional, uso da terra e biomas. Amazônia e Mata Atlântica apresentaram maior potencial de recarga, enquanto o Pampa teve valores menores. Áreas de vegetação nativa mostraram maior potencial de recarga. Embora a textura do solo seja um fator relevante na condução de água no solo e, portanto, na recarga hídrica, a combinação entre textura e classificação do solo se mostrou mais satisfatório para mapear o potencial de recarga no Brasil.

**Palavras-chave:** Análise de Componentes Principais (PCA); Random Forest; HYBRAS; XGBoost; áreas de recarga hídrica

## ABSTRACT

In the face of increasing global water scarcity and the complex process of climate change, groundwater emerges as a fundamental resource to ensure water quality and availability. Thus, the present study investigates the role of soil attributes in groundwater recharge potential by developing and evaluating Soil Quality Index (SQI) using data from the Hydrophysical Database for Brazilian Soils (HYBRAS). Initially, a Soil Quality Index for Groundwater Recharge Potential (SQIra) was developed based on the evaluation of eleven physical soil attributes, employing the Supervised Principal Component Analysis (SPCA) tool to select the main indicators: Saturated Hydraulic Conductivity (Kslab\_log), Drainable Porosity (DP), and Silt content. The SQIra proved effective in highlighting differences in soil quality among land uses and soil classes. The index was validated with Basic Infiltration Rate (BIR) data and baseflow data, emphasizing its usefulness for sustainable watershed management. Furthermore, the pattern of the SQI created, along with other indices available in the literature for recharge potential, was evaluated for the influence of soil class, textural class, functional structure, land use, and biomes. The Amazon and Atlantic Forest biomes showed higher recharge potential, while the Pampa biome exhibited lower values. Native vegetation areas showed greater recharge potential. Although soil texture is a relevant factor in water conduction through the soil and, therefore, in groundwater recharge, the combination of soil texture and soil classification proved more satisfactory for mapping recharge potential in Brazil.

**Keywords:** Principal Component Analysis (PCA); Random Forest; HYBRAS; XGBoost; water recharge áreas

## **INDICADORES DE IMPACTO**

A pesquisa possibilitou a criação de um Índice de Qualidade do Solo podendo ser utilizado como ferramenta para subsidiar políticas públicas e estratégias de manejo do solo e da água em territórios brasileiros sujeitos à escassez hídrica, degradação ambiental ou expansão agrícola. Os dados utilizados cobrem diversos biomas e regiões brasileiras, refletindo um território de abrangência nacional focado principalmente na área temática de Meio Ambiente. O caráter extensionista da pesquisa está evidenciado pela integração com agentes externos à Universidade Federal de Lavras (UFLA), por meio da utilização e validação de dados do HYBRAS. O trabalho está alinhado com pelo menos cinco Objetivos de Desenvolvimento Sustentável (ODS) da ONU, sendo eles: ODS 2 (Fome zero), ODS 6 (Água potável e saneamento), ODS 12 (Consumo e produção responsáveis) e ODS 13 (Ação contra a mudança global do clima). Ao oferecer suporte técnico-científico para o monitoramento da recarga hídrica e da capacidade de retenção de água nos solos brasileiros, o estudo contribui para a gestão integrada dos recursos hídricos com potencial de beneficiar diretamente agricultores, gestores públicos e comunidades que dependem do solo e da água para sua subsistência e qualidade de vida.

## **IMPACT INDICATORS**

The research enabled the development of a Soil Quality Index, which can be used as a tool to support public policies and strategies for soil and water management in Brazilian territories affected by water scarcity, environmental degradation, or agricultural expansion. The data used encompass various biomes and regions of Brazil, reflecting a territory of national scope, primarily focused on the thematic area of Environment. The extensionist nature of the research is evidenced by the integration with external agents to the Federal University of Lavras (UFLA), through the use and validation of data from the HYBRAS database. The work is aligned with at least five of the United Nations Sustainable Development Goals (SDGs), namely: SDG 2 (Zero hunger), SDG 6 (Clean water and sanitation), SDG 12 (Responsible consumption and production), and SDG 13 (Climate action). By providing technical-scientific support for monitoring groundwater recharge and the water retention capacity of Brazilian soils, the study contributes to the integrated management of water resources, with the potential to directly benefit farmers, public managers, and communities that depend on soil and water for their livelihood and quality of life.

## SUMMARY

<b>PART ONE</b> .....	<b>13</b>
<b>1. INTRODUCTION</b> .....	<b>13</b>
<b>1.1. General Introduction</b> .....	<b>13</b>
<b>1.2. Organization of the dissertation</b> .....	<b>14</b>
<b>2. LITERATURE REVIEW</b> .....	<b>14</b>
<b>2.1. Groundwater Recharge Potential in the Soil</b> .....	<b>14</b>
<b>2.2. Conceptual Aspects of Soil Quality</b> .....	<b>15</b>
<b>2.2.1. Soil texture</b> .....	<b>16</b>
<b>2.2.2. Soil structure</b> .....	<b>17</b>
<b>2.2.2.1. Soil particle density</b> .....	<b>18</b>
<b>2.2.2.2. Bulk density</b> .....	<b>18</b>
<b>2.2.2.3. Soil porosity</b> .....	<b>19</b>
<b>2.2.3. Basic Infiltration Rate</b> .....	<b>21</b>
<b>2.2.4. Hydraulic conductivity of soil</b> .....	<b>23</b>
<b>2.3. Soil Quality Assessment</b> .....	<b>24</b>
<b>2.3.2. Statistical methods</b> .....	<b>25</b>
<b>2.4. Soil Physical Quality and Groundwater Recharge</b> .....	<b>26</b>
<b>2.5. Hydrophysical Database for Brazilian Soils (HYBRAS)</b> .....	<b>26</b>
<b>REFERENCES</b> .....	<b>27</b>
<b>PART TWO - PAPERS</b> .....	<b>31</b>
<b>PAPER I: Development and Validation of a Statistical Index for Estimating Groundwater Recharge Potential</b> .....	<b>32</b>
<b>1. INTRODUCTION</b> .....	<b>35</b>
<b>2. MATERIAL AND METHODS</b> .....	<b>37</b>
<b>2.1. Data collection</b> .....	<b>37</b>
<b>2.2. Index determination</b> .....	<b>37</b>
<b>2.2.1. Selecting the Minimum Data Set (MDS)</b> .....	<b>37</b>
<b>2.2.2. Scoring and integration of indicators</b> .....	<b>38</b>
<b>2.2.1. Validation of the SQI for groundwater recharge (SQIgr)</b> .....	<b>39</b>
<b>2.2.1.1. Validation using BIRn</b> .....	<b>39</b>
<b>2.2.1.2. Validation through baseflow pattern</b> .....	<b>40</b>
<b>3. RESULTS AND DISCUSSION</b> .....	<b>41</b>

3.1.	Selecting the Minimum Data Set (MDS) .....	41
3.2.	Scoring and integration of indicators .....	44
3.3.	Validation of the SQIgr .....	44
3.3.1.	Simplification of soil classes.....	44
3.3.1.1.	Validation using BIRn.....	47
3.3.2.	Validation through baseflow pattern.....	49
4.	CONCLUSION.....	51
	REFERENCES .....	51
	<b>PAPER II: Determinants of Groundwater Recharge in Brazil: A Comprehensive Analysis</b>	
	56	
1.	INTRODUCTION.....	59
2.	METHODS .....	61
2.1.	Data collection .....	61
2.2.	Calculation of Soil Quality Index .....	61
2.3.	SQIs comparison .....	63
2.4.	Evaluation of the importance of factors .....	63
3.	RESULTS.....	65
3.1.	SQIs comparison .....	65
3.2.	Evaluation of the importance of factors and prediction indexes.....	70
4.	DISCUSSION .....	71
4.1.	Influence of factors on SQI.....	71
4.2.	Evaluation of the Importance of Factors.....	75
5.	CONCLUSION.....	76
	REFERENCES .....	77
	<b>PART 3 - FINAL CONSIDERATIONS .....</b>	<b>82</b>
1.	LIMITATIONS AND FINAL CONSIDERATIONS.....	83

**PART ONE**

**GROUNDWATER RECHARGE POTENTIAL AND ITS INTERACTION WITH SOIL  
PHYSICAL QUALITY IN BRAZIL**

## 1. INTRODUCTION

### 1.1. General Introduction

The sustainable management of water resources, especially regarding to groundwater recharge, is a topic of growing relevance in the current context, given the increasing demand for drinking water and the impacts of climate change. Groundwater represents around 99% of the liquid freshwater available on the planet (UNESCO, 2022) and can be used in various sectors, such as public supply, agriculture, livestock, industry and mining. However, excessive and inappropriate use of this source can lead to the degradation and depletion of aquifers (CHAMINÉ *et al.*, 2015; JAAFARZADEH *et al.*, 2021), especially when recharge is less than exploitation.

The soil has a direct impact on aquifer recharge, as factors such as soil texture, soil structure, porosity, bulk density and hydraulic conductivity affect the process of water infiltration into the soil (RESENDE *et al.*, 1998; MELLO, 2013). Vegetation also plays a crucial role in this process, helping to maintain water storage capacity in the soil, especially in areas with native vegetation (ROA-GARCÍA *et al.*, 2011; GERMER *et al.*, 2010).

Several studies have used the Soil Quality Index (SQI) to estimate groundwater recharge potential, considering indicators such as soil saturated hydraulic conductivity, drainable porosity and basic infiltration rate (SANTANA *et al.*, 2023; ALVARENGA *et al.*, 2012). However, they were developed restricted to sub-basins and were based solely on expert opinion, making the analysis subjective. Other studies have used remote sensing techniques combined with geographic information systems (GIS) to map areas with groundwater recharge potential based on the integration of various environmental factors. Nonetheless, they are limited to assessing the influence of the soil, adopting isolated variables such as soil textural class or soil class, which by far are able to predict alone the potential to recharge aquifers (ACHU *et al.*, 2020; JAAFARZADEH *et al.*, 2021). Both approaches can result in an underestimation or overestimation of the recharge potential, especially in the context of Brazil, which has a great diversity of soils with different physical and hydric characteristics and properties.

The publication of the Hydrophysical Database for Brazilian Soils (HYBRAS) in 2018 represented a significant advance, bringing together information on soil properties such as soil texture, porosity, bulk density and saturated hydraulic conductivity (OTTONI *et al.*, 2018). This database allows for a more accurate assessment of the physical quality of soils, which is essential for developing management strategies aimed at conserving groundwater resources.

The aim of this study is to statistically integrate soil properties to create a Soil Quality Index for recharge potential applicable to the entire national territory. In addition, the study

investigates how factors related to textural class, soil class, soil structure, land use and different Brazilian biomes influence soil quality for groundwater recharge potential. The purpose of this analysis is to identify the most representative soil quality factors for recharge, based on data from the HYBRAS database, to support the mapping of groundwater recharge zones in Brazil and contribute to the sustainable management of water resources, in line with the Sustainable Development Goals (SDGs), especially SDG 6, which aims to guarantee the availability and sustainable management of water.

### **1.2. Organization of the dissertation**

The work is divided into two parts, the first consisting of a general introduction and theoretical reference and the second consisting of two articles.

The first paper seeks to integrate various physical and hydric soil properties to create a robust Soil Quality Index (SQI) for groundwater recharge potential. Using data from the Hydrophysical Database for Brazilian Soils (HYBRAS), statistical analyses were applied to select the most representative indicators, ensuring the independence of subjective opinions in the choice of the most influential soil parameters in water recharge.

In the second paper, the SQI developed was compared with other indexes proposed in the literature, such as those by Santana *et al.* (2023) and Alvarenga *et al.* (2012), assessing its effectiveness in estimating recharge potential across different soil classes and land uses. In addition, this study serves as a basis for mapping areas with groundwater recharge potential in Brazil, considering the diversity of soils and environmental conditions in the country.

## **2. LITERATURE REVIEW**

### **2.1. Groundwater Recharge Potential in the Soil**

Groundwater consists of all the water resources found below the Earth's surface. It accounts for approximately 99% of all liquid freshwater on Earth (UNESCO, 2022). In the context of growing global water scarcity, coupled with climate change, groundwater has the potential to provide quality water for domestic use, irrigation, livestock watering and industry. However, there is a need for better management of these resources, as well as an understanding of the potential for groundwater recharge to control the volume of exploitation.

Data presented in the United Nations World Water Development Report 2022 show that total global groundwater abstraction in 2017 was estimated at 959 km<sup>3</sup>. In addition, the same report highlighted the main services for which groundwater resources are needed, including: i) supply for human consumption and other purposes, such as irrigation, industry, mining, among

others; ii) regulation, helping with water security, such as recharging aquifers, maintaining rivers and lakes, regulating the climate, protecting against saline intrusion; and iii) supporting dependent ecosystems (UNESCO, 2022).

Although there are many services that depend on groundwater resources, groundwater must be exploited sustainably through effective management. Evaluating and controlling groundwater potential is essential to prevent the depletion of this resource (CHAMINÉ *et al.*, 2015) and the natural recharge capacity of aquifers must be taken into account. The recharge potential is influenced by the infiltration of water into the soil from a region with favorable conditions for receiving rainfall, as well as the geomorphological characteristics of the watersheds (MELLO, 2013).

Thus, considering the infiltration process, it is necessary to understand how and which soil quality conditions are fundamental to establishing areas with greater potential for groundwater recharge. JAAFARZADEH *et al.* (2021) analyzed fifteen parameters that affect the potential for groundwater recharge, highlighting the importance of understanding the soil, as the permeability of the soil affects the amount of water that can infiltrate and recharge the aquifer.

## **2.2. Conceptual Aspects of Soil Quality**

The process of water infiltration into the soil and its redistribution of water is basically governed by two processes: the capacity of the soil to allow the movement of water through it (hydraulic conductivity) and the energy required to carry out this work through the difference in water potential in the soil (PINHEIRO; VAN LIER; BEZERRA, 2017).

These processes are influenced by indicators mainly related to the physical quality of the soil (BÜNEMANN *et al.*, 2018):

*“Soil Physical Quality is the ability of a given soil to meet the requirements of plants and ecosystems in relation to water demand, aeration and mechanical resistance over time, as well as to resist and recover from processes that may diminish this capacity”* (MCKENZIE; TISDALL; VANCE, 2011).

Therefore, the textural classification of the soil, the mutual and organizational arrangement of solid particles and soil aggregates, the composition of the pore space, the influence of land use, among others, affect the magnitude of the water infiltration and

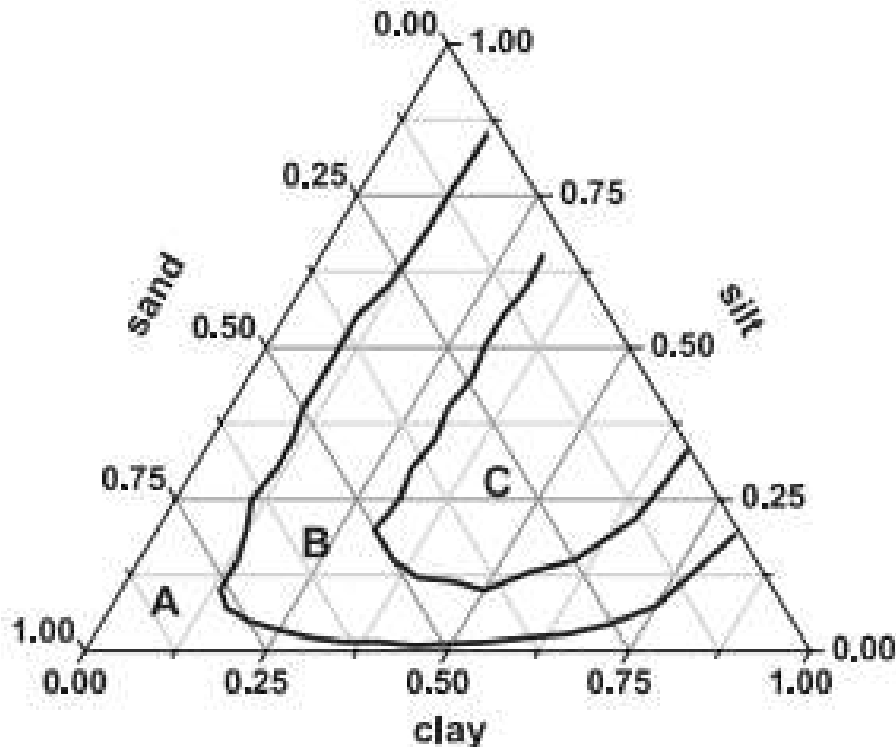
redistribution process in the soil profile for groundwater recharge, which will be presented in the following subsections.

### 2.2.1. Soil texture

Soil texture is a physical indicator related to the size distribution of the primary mineral particles in the soil, smaller than 2 mm, conveniently quantified in terms of sand, silt, and clay (TEIXEIRA *et al.*, 2017). Texture is commonly used in the mapping evaluation of areas with groundwater recharge potential (GUERRÓN-OREJUELA *et al.*, 2023; LENTSWE; MOLWALEFHE, 2020; THANH *et al.*, 2022), as well as a predictor for estimating the hydraulic properties of soils (MANTOVANELLI, 2021).

The different soil textures have varying impacts on the physical quality of the soil, as shown in the ternary diagram of S index values (Figure 1.1) proposed by DEXTER (2004). The S index is well known as an important indicator of soil physical quality. It represents the slope of the soil water release curve (SWRC) on a mass base at its inflection point on a logarithmic matric potential scale. In area A, there are values of  $S > 0.04$ , in area B, there are values of  $0.03 < S < 0.04$ , and in area C, there are values of  $S < 0.03$ . Thus, higher S values indicate better physical quality and, consequently, structural heterogeneity.

**Figure 1.1** - Ternary diagram of S Index values\* for different continuous texture functions



\* Area A:  $S > 0.04$ ; Area B:  $0.03 < S < 0.04$ ; Area C:  $S < 0.03$

Source: DEXTER (2004)

The influence of soil texture on the physical quality of the soil significantly affects the properties of hydrological flow (MURPHY; WATERHOUSE; DAHLKE, 2021), showing, above all, a linear correlation between groundwater recharge rates and the average clay content in the subsurface. Significantly lower clay content tends to have higher saturated hydraulic conductivity and infiltration capacity (TAO *et al.*, 2021), so sandy soils are classified as having greater groundwater recharge potential because they have low water retention and high saturated hydraulic conductivity.

However, soil texture is more descriptive when applied to soils in temperate regions (UEHARA; GILLMAN, 1985), and is not representative of soils in tropical regions due to the presence of highly weathered clays (MARTINEZ; SOUZA, 2020). FERREIRA; FERNANDES; CURI (1999) proposed an analysis of the influence of the mineralogy of the clay fraction on the physical properties of Latossolos in Brazil. They identified a positive correlation between the increase in clay content and the increase in saturated hydraulic conductivity, showing that the granular structure of these soils with a high gibbsite content may influence the process of water movement in the soil as a function of mineralogy.

Therefore, considering that Latossolos represent approximately 30% of the Brazilian territory, it is necessary to evaluate other physical indicators related to soil structure, instead of considering only soil texture, in order to analyze areas with groundwater recharge potential in Brazilian soils.

### **2.2.2. Soil structure**

Soil structure refers to the arrangement of soil particles and the pore space between them, including the size, shape and arrangement of aggregates formed by the grouping of primary particles (MARSHALL, 1962). This indicator governs a significant part of the hydrological processes in a watershed (MELLO, 2013). A soil structure with a predominance of the clay fraction composed of smaller aggregate units facilitates the drainage of water in the soil due to macroporosity, such as soils with a granular structure. On the other hand, a soil structure with coarse-sized units can inhibit the flow of water due to the face-to-face contact (FERREIRA; FERNANDES; CURI, 1999).

The soil structure is subject to change depending on the use and management of the soil, and is, therefore, dynamic in nature, with great spatial variability and various factors influencing it (VAN LIER, 2010). Thus, soil structure can be assessed using indicators that relate the solid, liquid and gaseous phases of the soil.

Understanding soil structure in the context of Brazilian soils with highly weathered clays is more relevant than soil texture (MELLO; CURI, 2012). *Latossolos* found in Brazil with a clayey or very clayey texture and granular structure exhibit high infiltration capacity due to the formation of macropores between the aggregates (FERREIRA; FERNANDES; CURI, 1999).

#### **2.2.2.1. Soil particle density**

Soil particle density is considered an intrinsic characteristic of the soil and depends on the relative proportions of mineral and organic particles. It is only related to the volume occupied by the solid phase of the soil and therefore does not take porosity into account, nor does it take the soil's structuring conditions (TEIXEIRA *et al.*, 2017).

Like soil texture, this indicator allows for an indirect understanding of the dynamics of hydrological flow in different types of soil at depth, since it is used to calculate soil porosity (BÜNEMANN *et al.*, 2018), as well as at the soil surface to evaluate the particle cohesion and the corresponding effect on surface runoff of a given soil (AKPAN *et al.*, 2016). In addition, the organic matter content in the soil contributes to lower particle density values, which in turn directly influences on soil permeability and aeration, affected by the volume of micropores and mesopores (NEHLS *et al.*, 2006).

#### **2.2.2.2. Bulk density**

Bulk density is the ratio of dry soil mass to its natural volume. This property is directly linked to soil structure and mineralogy as well as to other soil processes such as soil compaction and water infiltration, due to its sensitivity to land use and management, and is included in the calculation of other indicators such as porosity and soil carbon stock (SILVA *et al.*, 2020).

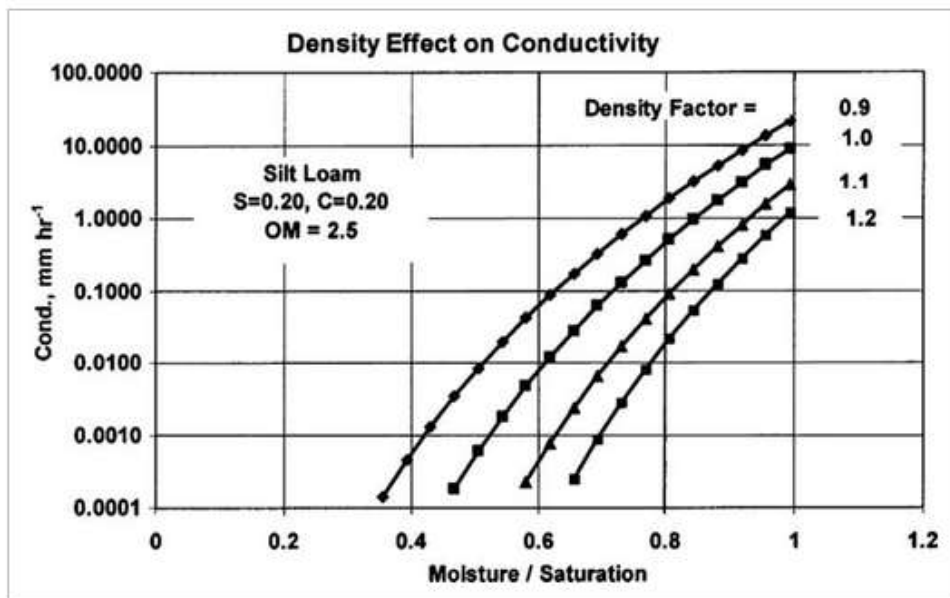
Measuring bulk density is relatively quick, simple, low cost and does not require sophisticated laboratory equipment. It is therefore the most widely used parameter for assessing the physical quality of Brazilian soils (SIMON *et al.*, 2022). However, although bulk density is an indicator of soil infiltration conditions, interpreting this property requires comparability with data from other sampling campaigns due to its variability (BÜNEMANN *et al.*, 2018).

Less weathered, shallow soils with primary minerals, such as *Neossolos Litólicos*, tend to have high bulk density values ( $1.87 \text{ Mg m}^{-3}$ ), while more weathered soils have low bulk density values ( $1.15 \text{ Mg m}^{-3}$ ) (DE SOUZA *et al.*, 2019). In relation to soil structure, lower bulk density values correspond to soils with a granular structure, while higher bulk density values

are associated with structures with coarse-sized aggregate units (blocky or prismatic) (VAN LIER, 2010).

The depth of the layer evaluated can also influence bulk density values, with greater variation in surface layers reflecting land use and management (SANTANA *et al.*, 2023). Inadequate soil management practices can influence the arrangement of soil particles and lead to an increase in bulk density, due to a reduction in macro-pores (DEXTER, 2004). When this occurs, it affects the soil's water retention capacity by reducing hydraulic conductivity (Figure 1.2) (SAXTON; RAWLS, 2006). However, the sensitivity of this soil property is related to the soil texture and should therefore be analyzed alongside it.

**Figure 1.2 - Effect of Bulk Density on Hydraulic Conductivity**



Source: SAXTON; RAWLS (2006)

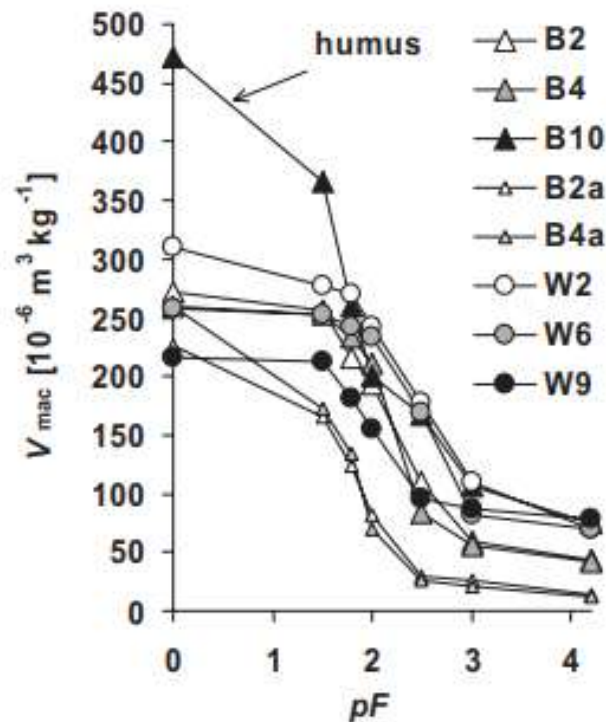
### 2.2.2.3. Soil porosity

Soil porosity corresponds to the fraction of the soil occupied by water or air. The analysis of properties such as pore size distribution is useful for understanding the processes that occur within porous system and the influence on water flow in the soil (FERREIRA *et al.*, 2019).

Considering that the pore space of the soil can be represented by the capillarity model, the soil pores are segregated into size classes, with a diameter of 0.05 mm serving as the threshold between the classes (TEIXEIRA *et al.*, 2017). Micropores are then considered to be capillary pores and macropores are considered to be non-capillary pores. Soils with high macroporosity values tend to have a high infiltration rate, while soils with high microporosity values support greater water retention capacity (HUNTLEY, 2023).

Thus, the process of water infiltration in the soil is strongly governed by the soil permeability, with Macroporosity being the most important indicator when analyzed in the surface layer of the soil (ALVARENGA *et al.*, 2012). In this way, the greater the volume of soil occupied by macropores, the lower the water retention, the higher the permeability, and consequently the greater the contribution to groundwater recharge process (NEHLS *et al.*, 2006) (Figure 1.3).

**Figure 1.3** - Water content curves used to calculate the volume of macropores



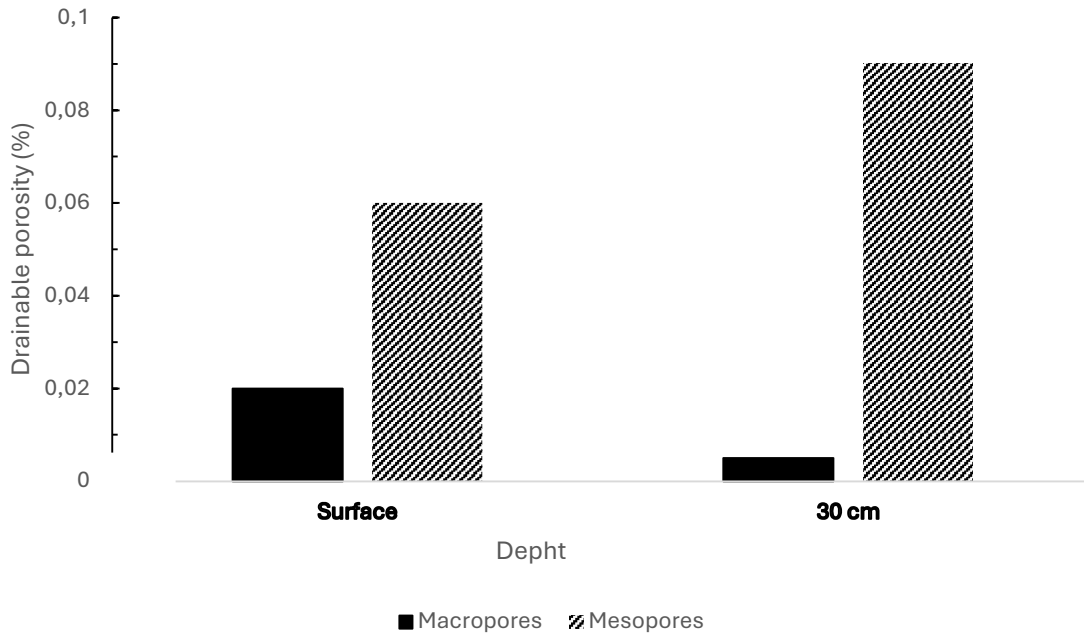
Source: NEHLS *et al.* (2006)

Other authors, still focusing on analyzing the influence of porosity on surface and subsurface water flow, as well as soil quality index for groundwater recharge potential, use drainable porosity or effective porosity as a basis (CAMEIRA, FERNANDO, PEREIRA, 2003; SANTANA *et al.*, 2023). Effective porosity is equivalent to the fraction of total porosity in which water moves freely under the action of gravity, determined by the difference between total porosity, or saturation water content, and the water content corresponding to field capacity (QUEIROZ, 1995).

For this purpose, the pore classes are subdivided into macropores (larger than 0.05 mm), mesopores (0.015 to 0.050 mm) and micropores (smaller than 0.015 mm), since the main function of the mesopores is to conduct the water after the macropores have been emptied (VAN

LIER, 2010) and thus influence the drainable porosity, especially in the subsurface of conventional plantations (Figure 1.4), (CAMEIRA; FERNANDO; PEREIRA, 2003).

**Figure 1.4** - Effect of pore distribution on the calculation of drainable porosity in conventional planting



Source: CAMEIRA; FERNANDO; PEREIRA (2003)

However, it should be noted that both drainable porosities based on macroporosity and mesoporosity do not take into account tortuosity or interruptions (BATISTA *et al.*, 2020) and have a different influence on unsaturated and saturated zones. In the former, pores are empty and permeability is hindered due to the adsorption process (soil-water), while in the latter the pores are filled with water, causing a greater flow of water due to the cohesion process (BRAGA, VELÁSQUEZ, FLEMING, 2020).

### 2.2.3. Basic Infiltration Rate

The infiltration process refers to the water that penetrates and moistens the soil. As rainfall enters the drier soil and the degree of saturation increases over time, the infiltration rate becomes constant, called the Basic Infiltration Rate (BIR), tending towards the value of the saturated hydraulic conductivity ( $K_{sat}$ ) (ANGULO-JARAMILLO; BAGARELLO; IOVINO, 2016).

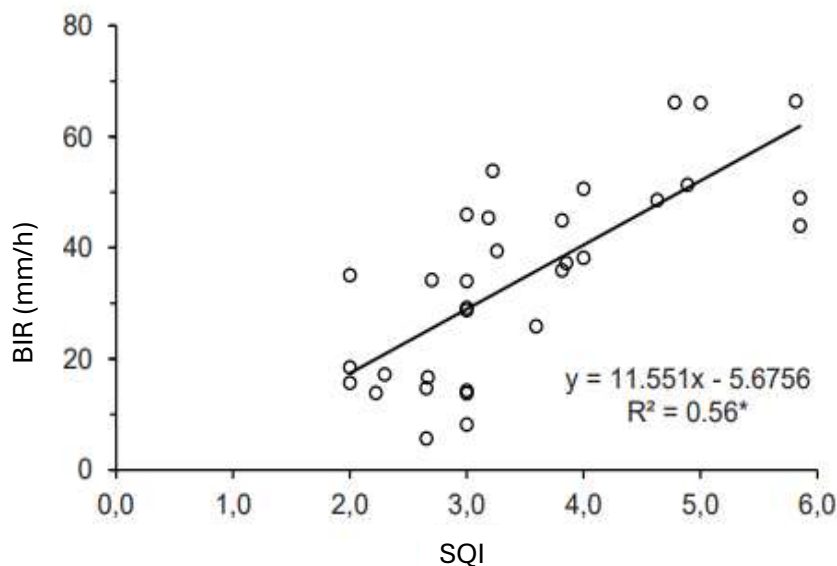
Water infiltration tests are usually carried out in the field using disk or tension infiltrometers in order to quantify the cumulative infiltration,  $I(t)$  (L), in the surface layer of

the soil. This is done by determining the infiltration rate,  $i(t)$  (L/T), the amount of water that infiltrates over time, which are related to each other by derivation (RAHMATI *et al.*, 2018).

BIR is related to soil texture. ILHA (2018) found by analyzing infiltration data with a rainfall simulator, a 24-hour test, that the spread of water distribution in soils varies according to soil texture. For sandy loam soils, the horizontal propagation distance reached 1.20 meters, while for clay loam and loam soils this distance was 0.90 and 0.50 meters. In addition, it can be influenced by vegetation cover, being greater in areas with vegetation cover than in areas without cover (ALVES; SUZUKI; SUZUKI, 2007). Furthermore, the process is enhanced in areas with natural vegetation than in cultivated areas (CAVENAGE *et al.*, 1999).

Higher BIR values are related to higher values of soil physical quality indicators, due to the favorable pore size distribution for root growth and water infiltration capacity in the soil (ALVES; SUZUKI; SUZUKI, 2007), showing a positive correlation between them (Figure 1.5) (DEBIASI *et al.*, 2023). On the other hand, degraded soils are associated with lower BIR values, due to the reduction in cross-sectional area for water flow, along with more tortuous paths for fluid movement, negatively affecting the infiltration process. (ALVES; SUZUKI; SUZUKI, 2007).

**Figura 1.5 – BIR and soil quality index relationship**



Source: DEBIASI *et al.* (2023)

In this way, BIR is one of the indicators that best reflects the physical conditions of the soil for water infiltration, and is commonly used in studies and design for irrigation projects, drainage systems, soil conservation processes and analysis of recharge potential. In a

comparison between double-ring infiltration tests and pumping tests (considered the most reliable field technique for estimating aquifer parameters, albeit at high cost), MASOUD; BASAHI; ZAIDI (2019) showed that infiltration tests (0.58 to 37.15 m/day) showed values close to those obtained from pumping tests (2.6 to 57.4 m/day), corroborating the evidence that BIR is a powerful soil physical indicator for groundwater recharge.

#### 2.2.4. Hydraulic conductivity of soil

Hydraulic conductivity is a parameter that describes the ease with which water flows through a porous medium (SEILER; GAT, 2007). When analyzed under saturated conditions, it is called soil saturated hydraulic conductivity ( $K_{sat}$ ) and can be expressed by the Darcy's equation.

The determination of  $K_{sat}$  is routinely carried out in the laboratory with constant head permeameters using undisturbed samples. However, other methods are reported in the literature using field infiltration measurements to estimate the saturated hydraulic conductivity of the soil. SALES *et al.* (1999) showed a correlation between the value of BIR and  $K_{sat}$  for the two soils studied (*Latosolo Vermelho* and *Argissolo Vermelho-Amarelo*) and in the surface and subsurface layers. The study shows that the influence of BIR can vary depending on the soil layer. In the 0-20 cm layer of the *Argissolo Vermelho-Amarelo*, the correlation between BIR and  $K_{sat}$  was weaker than in the 60-80 cm layer. For the *Latosolo Vermelho*, on the other hand, the BIR value was similar to the average value for both layers. Therefore, the use of BIR to estimate  $K_{sat}$  may depend on the soil layer and the specific characteristics of the soil analyzed.

$K_{sat}$  is influenced by the size of the pores and their interconnectivity, in which the macropores are mainly responsible for conducting water through the soil profile. Thus, a small reduction in macroporosity is accompanied by a large reduction in hydraulic conductivity values (SILVA; KATO, 1997). It can therefore be affected by the different land use systems, which as pointed out by SANTANA *et al.* (2023), higher  $K_{sat}$  values were associated with native forest areas.

In addition, soil hydraulic conductivity can also be affected by soil texture, where higher values of sand content usually present a positive correlation with hydraulic conductivity (BOCUTI *et al.*, 2020). Soil structure also strongly affects hydraulic conductivity depending on the type of soil (FERREIRA; FERNANDES; CURI, 1999).

Therefore, in addition to the sand content, it is important to analyze how the attributes silt, clay and organic matter influence the soil's structural stability. Silt and clay are smaller particles than sand and can therefore clog the soil's pores, reducing its ability to conduct water.

Organic matter, in turn, can affect soil structure and the formation of aggregates, which influence the pore size distribution of the soil. The relationship between structural stability and basic soil properties was proposed by REYNOLDS *et al.* (2009) (Equation 1). The structural stability index (SSI, %) is a metric for the risk of soil structure degradation.

$$SSI (\%) = \frac{1,724 \times OC}{Silt + Clay} \times 100 \quad \text{Equation 1}$$

Where OC is the organic carbon content of the soil (g/kg), the van Bemmelen factor (1.724) is used to convert OC into OM (CAMBARDELLA *et al.*, 2001). The structural quality index ranges from 0 to 1, with higher values indicating better physical quality of the soil in relation to its structure, thus presenting greater hydraulic conductivity in the soil.

## 2.3. Soil Quality Assessment

### 2.3.1. Additive-weighted function

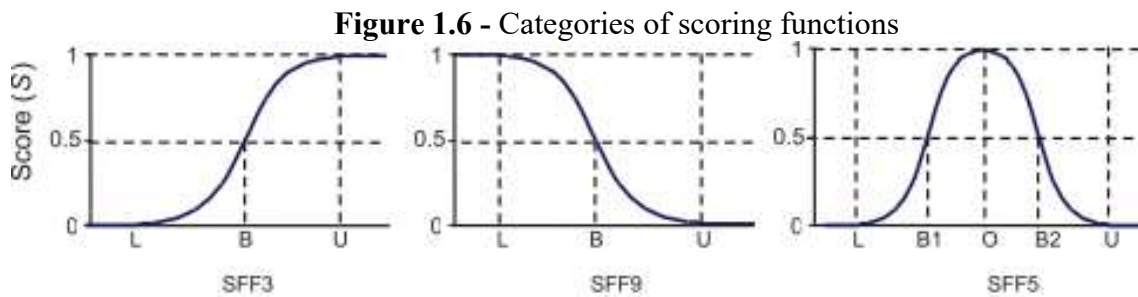
The assessment of soil quality is complex due to the dynamics of physical, chemical and biological processes, in addition to the heterogeneity and diversity of land uses (KELTING *et al.*, 1999). Despite these complexities, the creation of Soil Quality Index (SQIs) has been proposed to provide a numerical representation of soil quality (DORAN; PARKIN, 1994; KARLEN; STOTT, 1994; WANG; GONG, 1998). These indexes involve the selection of indicators, to which weights are assigned and the combination of these weighted values to produce an overall index, providing a quantitative metric for comparative analysis (DIACK; STOTT, 2001).

The SQI have emerged from approaches focused on evaluating the productive capacity of the soil, incorporating soil properties that represent soil functions such as root support, nutrient retention, water conductivity and biodiversity support. Thus, KARLEN; STOTT (1994) proposed the additive-weighted function model based on the definition of weights for each soil function, based on a specific set of representative indicators and scores for each indicator, highlights the relative importance through on expert Opinion (Equation 2).

$$SQI = \frac{\sum_{i=1}^n v_i \times p_i}{\sum_{i=1}^n p_i} \quad \text{Equation 2}$$

Where,  $n$  is the total number of soil quality indicators,  $v_i$  is the score value assigned to indicator  $i$ ,  $p_i$  is the weight of the indicator based on the degree of importance assigned.

These scores are transformed using non-linear scoring functions (SSF), which standardize the values on a scale from zero to one as proposed by WYMORE (1993). WYMORE (1993) defined twelve categories of SSFs; however, KARLEN; STOTT (1994) suggested the use of only three main scoring functions soil quality indicators (Figure 1.6) across different soil types and land uses.



Source: CHAER (2004)

In the first type, “more is better”, the positive slope parameter ( $S$ ) starts at zero and reaches 1 between  $L$  (lower limit) and  $U$  (upper limit). The second type, “less is better”, with negative  $S$  starting at 1 and decreasing to 0 between  $L$  and  $U$ . The third type is a combination of the previous two, with the optimum point “ $O$ ” corresponding to 1.

However, this methodology relies on weights assigned according to the expert's opinion, making the evaluation subjective (RINOT *et al.*, 2019). Therefore, statistical tools such as multivariate ordinations and machine learning models are used as alternatives for selecting indicators to establish Soil Quality Index (SQI) indexes (CHAER, 2004).

### 2.3.2. Statistical methods

Among the multivariate ordering methods, Principal Component Analysis (PCA) stands out (PEARSON, 1901). The result of the multivariate ordering is a two-dimensional graph, where the indicators are analyzed based on their spatial distance from a reference dataset used. ABDEL-FATTAH *et al.* (2021) use PCA to reduce the data set into new variables, avoid multicollinearity and determine relative weights ( $p_i$ ) and soil indicators ( $S_i$ ), which were used to obtain the SQI. The results showed that the first three components the principal components of PCA explained 83.6% of the total variance of the soil data, confirming that PCA produces satisfactory results, being representative of the variability in the final measures of soil quality.

## 2.4. Soil Physical Quality and Groundwater Recharge

Sensitive soil quality indicators, which reflect the use and management of the soil, are commonly used as a tool for assessing soil quality by means of a Soil Quality Index (SQI). With regard to the function of water storage, infiltration and/or recharge, soil physical indicators for formulating the SQI have been used mainly in the last 5 years, focusing on the dynamic properties of the soil and on the inherent characteristics of the soil (BÜNEMANN *et al.*, 2018).

This is due to the relevant effect of soil physical and hydraulic properties on soil functioning (root growth, storing and supplying water and nutrients, gas exchange and biological activity) (ARSHAD; LOWERY; GROSSMAN, 1996). The main soil physical indicators include soil texture, bulk density, particle density, porosity, basic infiltration rate, hydraulic conductivity and aggregate stability (ARAÚJO *et al.*, 2012).

Considering the soil physical indicators for assessing the physical quality of soil for groundwater recharge potential, SANTANA *et al.* (2023) and ALVARENGA *et al.* (2012) proposed two different sets of indicators, both using the expert opinion methodology for determining the soil quality index.

In the Cantareira System, Brazil, the soil quality index, when evaluated from the perspective of land use and management in three types of soil, considering drainable porosity, basic infiltration rate and saturated hydraulic conductivity in the laboratory as soil physical indicators, showed that the effects of land use on soil functions depend on the type of soil. The soil quality index related to potential groundwater recharge in the surface layer of the soil was higher in areas of native forest for soils of the *Cambissolos* (SANTANA *et al.*, 2023).

The set of indicators proposed by ALVARENGA *et al.* (2012), such as, bulk density, saturated hydraulic conductivity in the field and macroporosity - also proved to be sensitive to land use and management, proved sensitive to an environmental quality indicator for managing and analyzing water recharge in the sub-basins studied in the Serra da Mantiqueira region, Brazil.

Therefore, the use of soil physical quality index provides the necessary tools to study and evaluate not only the management of soil resources, but also the impact on the potential groundwater recharge system in the soil. This contributes to a more holistic view of the benefits that soil can provide for sustainable development.

## 2.5. Hydrophysical Database for Brazilian Soils (HYBRAS)

HYBRAS is a hydrophysical database of Brazilian soils, composed of water retention, and saturated hydraulic conductivity (Ks) data, associated with basic soil properties and the

methods used to determine these indicators. The database has two versions: the first has already been published and is publicly available, while version 2 is currently under development.

Version 1 comprises data from 163 scientific documents, whereas version 2 includes 212 studies, encompassing samples collected at various locations and depths, totaling approximately 8,000 samples. These sources originate from a wide range of national publications, including peer-reviewed articles, master's dissertations, doctoral theses, conference and symposium proceedings on soil science, and research bulletins from the *Empresa Brasileira de Pesquisa Agropecuária* (Embrapa).

Both versions of the database include general sample information, such as geographic location, soil classification, profile descriptions, and sampling depth. The database also contains volumetric water content ( $\theta$ ) data for various suction levels, along with the corresponding parameter values of the van Genuchten (VG) equation (van Genuchten, 1980). Furthermore, HYBRAS includes land cover classification data, hierarchically structured in two levels and standardized according to Collection 8.1 of the MapBiomas Project – Annual Mapping of Land Use and Land Cover in Brazil. Notably, version 2 incorporates additional variables, such as sulfuric attack and the weathering index (OTTONI *et al.*, 2018).

Considering both versions, the HYBRAS database provides a comprehensive and representative set of hydrophysical information for tropical weathered soils in Brazil, serving as a fundamental tool for soil modeling and related environmental studies.

## REFERENCES

- ABDEL-FATTAH, M. K. *et al.* Quantitative evaluation of soil quality using principal component analysis: The case study of El-Fayoum depression Egypt. **Sustainability (Switzerland)**, v. 13, n. 4, p. 1–19, 2 fev. 2021.
- AKPAN, G. E. *et al.* Soil Quality Studies for Some Physicochemical Parameters and Concentrations of Toxic Heavy Metals in Coastal Areas in Akwa Ibom State. *International Journal of Environment and Bioenergy Int. J. Environ. Bioener*, v. 11, n. 2, p. 122–139, 2016.
- ALVARENGA, C. C. *et al.* Índice de Qualidade do Solo Associado à Recarga de Água Subterrânea (IQS RA) na Bacia Hidrográfica do Alto Rio Grande, MG. 2012.
- ALVES, M. C.; SUZUKI, L. G. A. S.; SUZUKI, L. E. A. S. Densidade do solo e infiltração de água como indicadores da qualidade física de um latossolo vermelho distrófico em recuperação. **R. Bras. Ci. Solo**, v. 31, p. 617–625, 2007.
- ANDREWS, S. S.; KARLEN, D. L.; MITCHELL, J. P. A comparison of soil quality indexing methods for vegetable production systems in Northern California. *Agriculture, Ecosystems & Environment*, v. 90, n. 1, p. 25–45, 1 jun. 2002.

ANGULO-JARAMILLO, R.; BAGARELLO, V.; IOVINO, M. **Infiltration Measurements for Soil Hydraulic Characterization**. 2016

ARAÚJO, E. A. DE *et al.* Qualidade do solo: conceitos, indicadores e avaliação. **Revista Brasileira de Tecnologia Aplicada nas Ciências Agrárias**, v. 5, n. 1, p. 187–206, 30 abr. 2012.

ARSHAD, M. A.; LOWERY, B.; GROSSMAN, B. **Physical Tests for Monitoring Soil Quality. Em: Methods for Assessing Soil Quality**. v. 49p. 123–142.

BATISTA, L. DOS S. *et al.* Water infiltration rate in the soil under different uses and covers in the Poxim River basin, Sergipe, Brazil. **International Journal for Innovation Education and Research**, n. 11, 2020.

BOCUTI, E. D. *et al.* Effective hydraulic conductivity and its relationship with the other attributes of cerrado soils. **Revista Brasileira de Engenharia Agrícola e Ambiental**, v. 24, n. 6, p. 357–363, 2020.

BRAGA, L. T. P.; VELÁSQUEZ, L. N. M.; FLEMING, P. M. Groundwater recharge through the dolines in the semi-arid climate in Minas Gerais State, Brazil. **Environmental Earth Sciences**, v. 79, n. 1, 1 jan. 2020.

BÜNEMANN, E. K. *et al.* Soil quality – A critical review. **Soil Biology and Biochemistry**, v. 120, p. 105–125, 1 maio 2018.

CAMEIRA, M. R.; FERNANDO, R. M.; PEREIRA, L. S. Soil macropore dynamics affected by tillage and irrigation for a silty loam alluvial soil in southern Portugal. **Soil and Tillage Research**, v. 70, n. 2, p. 131–140, 2003.

CAVENAGE, A. *et al.* **Alterações nas propriedades físicas de um latossolo vermelho-escuro sob diferentes culturas**. v. 23, p. 997–1003, 1999.

CHAMINÉ, H. I. *et al.* **Role of hydrogeological mapping in groundwater practice: back to basics**. Disponível em: <<https://recipp.ipp.pt/handle/10400.22/18051>>. Acesso em: 13 set. 2024.

DE SOUZA, E. *et al.* Spatial and temporal potential groundwater recharge: The case of the doce river basin, Brazil. **Revista Brasileira de Ciencia do Solo**, v. 43, 2019.

DEBIASI, H. *et al.* Qualidade estrutural do solo e taxa de infiltração estável influenciadas por culturas de entressafra da soja. **Embrapa Solos**, 2023.

DEXTER, A. R. Soil physical quality: Part I. Theory, effects of soil texture, density, and organic matter, and effects on root growth. **Geoderma**, v. 120, n. 3–4, p. 201–214, 1 jun. 2004.

DIACK, M.; STOTT, D. E. **The Global Farm. Selected papers from the 10th International Soil Conservation Organization Meeting held May 24-29**. 2019

- DORAN, J.; PARKIN, T. Defining and Assessing Soil Quality. In: **Defining Soil Quality for a Sustainable Environment** p. 3–21. 1994
- FERREIRA, M. M.; FERNANDES, B.; CURI, N. **Influência da mineralogia da fração argila nas propriedades físicas de latossolos da região sudeste do Brasil**. 1999
- FERREIRA, T. R. *et al.* Water retention curve to analyze soil structure changes due to liming. **Anais da Academia Brasileira de Ciências**, v. 91, n. 3, 2019.
- GUERRÓN-OREJUELA, E. J. *et al.* Mapping Groundwater Recharge Potential in High Latitude Landscapes Using Public Data, Remote Sensing, and Analytic Hierarchy Process. **Remote Sensing**, v. 15, n. 10, 1 maio 2023.
- ILHA, R. **Escoamento superficial e infiltração de água no solo em sistema plantio direto com “vertical mulching”**. Tese de Doutorado—Santa Maria: Universidade Federal de Santa Maria, 2018.
- JAAFARZADEH, M. S. *et al.* Groundwater recharge potential zonation using an ensemble of machine learning and bivariate statistical models. **Scientific Reports**, v. 11, n. 1, 1 dez. 2021.
- KARLEN, D. L.; STOTT, D. E. **A Framework for Evaluating Physical and Chemical Indicators of Soil Quality**. 1994.
- KELTING, D. L. *et al.* Soil quality assessment in domesticated forests: a southern pine example. **Forest Ecology and Management**, v. 122, p. 167–185, 1999.
- LENTSWE, G. B.; MOLWALEFHE, L. Delineation of potential groundwater recharge zones using analytic hierarchy process-guided GIS in the semi-arid Motloutse watershed, eastern Botswana. **Journal of Hydrology: Regional Studies**, v. 28, p. 100674, 1 abr. 2020.
- MARTINEZ, P.; SOUZA, I. F. Genesis of pseudo-sand structure in *Latossolos* from Brazil – A review. **Geoderma Regional**, v. 22, p. e00292, 1 set. 2020.
- MASOUD, M. H. Z.; BASAHI, J. M.; ZAIDI, F. K. Assessment of artificial groundwater recharge potential through estimation of permeability values from infiltration and aquifer tests in unconsolidated alluvial formations in coastal areas. **Environmental Monitoring and Assessment**, v. 191, n. 1, 1 jan. 2019.
- MELLO, C. R. DE; CURI, N. Hydopedology. **Ciência e Agrotecnologia**, v. 36, n. 2, p. 137–146, abr. 2012.
- MURPHY, N. P.; WATERHOUSE, H.; DAHLKE, H. E. Influence of agricultural managed aquifer recharge on nitrate transport: The role of soil texture and flooding frequency. **Vadose Zone Journal**, v. 20, n. 5, 1 set. 2021.
- NEHLS, T. *et al.* Pore-system characteristics of pavement seam materials of urban sites. **Journal of Plant Nutrition and Soil Science**, v. 169, n. 1, p. 16–24, fev. 2006.
- PEARSON, K. On lines and planes of closest fit to systems of points in space. **Philosophy**, 1901.

- PINHEIRO, E.; VAN LIER, Q.; BEZERRA, A. Hydrology of a Water-Limited Forest under Climate Change Scenarios: The Case of the Caatinga Biome, Brazil. **Forests**, v. 8, n. 3, p. 62, 27 fev. 2017.
- QUEIROZ, J. E. Parâmetros hidrodinâmicos de um solo de várzea para fins de drenagem subterrânea. Tese de Doutorado, São Paulo – 1995.
- RAHMATI, M. *et al.* Development and analysis of the Soil Water Infiltration Global database. **Earth System Science Data**, v. 10, n. 3, p. 1237–1263, 10 jul. 2018.
- REYNOLDS, W. D. *et al.* Use of indicators and pore volume-function characteristics to quantify soil physical quality. **Geoderma**, v. 152, n. 3–4, p. 252–263, 15 set. 2009.
- RINOT, O. *et al.* Soil health assessment: A critical review of current methodologies and a proposed new approach. **Science of The Total Environment**, v. 648, p. 1484–1491, 15 jan. 2019.
- SANTANA, M. L. T. *et al.* Interactions between land use and soil type drive soil functions, highlighting water recharge potential, in the Cantareira System, Southeast of Brazil. **Science of The Total Environment**, v. 903, 2023.
- SAXTON, K. E.; RAWLS, W. J. Soil Water Characteristic Estimates by Texture and Organic Matter for Hydrologic Solutions. **Soil Science Society of America Journal**, v. 70, n. 5, p. 1569–1578, set. 2006.
- SILVA, P. A. DA *et al.* Spatial variation of soil carbon stability in sugarcane crops, central-south of Brazil. **Soil and Tillage Research**, v. 202, p. 104667, 1 ago. 2020.
- SIMON, C. DA P. *et al.* Soil quality literature in Brazil: A systematic review. **Revista Brasileira de Ciencia do Solo**, 2022.
- TAO, Z. *et al.* Groundwater recharge in hillslopes on the Chinese Loess Plateau. **Journal of Hydrology: Regional Studies**, v. 36, p. 100840, 1 ago. 2021.
- TEIXEIRA, P. C. *et al.* **Manual de Métodos de Análise de Solo**. Embrapa Solos: 2017
- THANH, N. N. *et al.* Global review of groundwater potential models in the last decade: Parameters, model techniques, and validation. **Journal of Hydrology**, v. 614, p. 128501, 1 nov. 2022.
- UNESCO. The United Nations World Water Development Report 2022: groundwater: making the invisible visible. 2022
- WANG, X.; GONG, Z. Assessment and analysis of soil quality changes after eleven years of reclamation in subtropical China. **Geoderma**. 1997.
- WYMORE, A. W. Model-Based Systems Engineering. **CRC Press**, 1993.

**PART TWO - PAPERS**

**PAPER I: Development and Validation of a Statistical Index for Estimating  
Groundwater Recharge Potential**

Lara Mota Corinto <sup>(1)</sup>; Samara Martins Barbosa <sup>(1)</sup>; Zilmar Divino de Carvalho Júnior <sup>(1)</sup>;  
Marta Vasconcelos Ottoni <sup>(2)</sup>; Nilton Curi <sup>(1)</sup>; Junior Cesar Avanzi <sup>(1)</sup>; Bruno Montoani Silva  
<sup>(1)</sup>

Standards of the Journal of Environmental Management (JEM).

---

<sup>(1)</sup> Federal University of Lavras, Department of Soil Science, 37200-000, Lavras, MG, Brazil.  
Email: laramcorinto@hotmail.com, leandcampos@yahoo.com.br, dayysant71@gmail.com,  
ntcuri@gmail.com, junior.avanzi@ufla.br; brunom.silva@ufla.br

<sup>(2)</sup> Geological Survey of Brazil, Department of Hydrology, 22290-255, Rio de Janeiro, RJ,  
Brazil. Email: marta.ottoni@sgb.gov.br

## ABSTRACT

In the context of increasing global water scarcity and the process of challenges of climate change, groundwater presents a viable source of high-quality water. With the influence of water infiltration, soil physical quality becomes an important factor in managing of these resources. The aim of this study was to develop a soil quality index associated with groundwater recharge. The study used the Hydrophysical Database for Brazilian Soils to evaluate the integration of soil physical attributes in the development of the index. Eleven physical attributes were assessed for the Supervised Principal Component Analysis (SPCA). Three attributes - Saturated Soil Hydraulic Conductivity (Kslab\_log), Drainable Porosity (DP), and Silt content - were selected to construct the Soil Quality Index for Groundwater Recharge Potential (SQIgr). For two land use classes (Agriculture and Forestry) according to MapBiomas, and four soil classes according to SiBCS (G2: *Argissolo Vermelho* and *Argissolo Vermelho-Amarelo*; G4: *Latossolo Vermelho* and *Latossolo Amarelo*) in the groups Agr\_G2 and For\_G4 there wasn't significant statistical difference between SQIgr and BIRn values, showing consistence results for the proposed index. The feasibility of the index's spatial coverage was validated by comparing the baseflow contribution in the runoff. SQIgr values for the For\_G4, Agr\_G4, and Agr\_G2 groups were 0.76, 0.58, and 0.50, respectively, indicating that agricultural use tends to decrease soil quality. However, the predominance of *Latossolos* was associated with higher soil quality scores. Thus, the proposed SQIgr proved to be a fundamental tool for the sustainable management of Brazilian watersheds, promising for distinguishing environments for preservation and conservation.

**Keywords:** Soil structure; Soil water infiltration; Ecosystem services

## RESUMO

No contexto do aumento da escassez global de água e dos desafios impostos pelas mudanças climáticas, as águas subterrâneas representam uma fonte viável de água de alta qualidade. Com a influência da infiltração de água, a qualidade física do solo torna-se um fator importante na gestão desses recursos. O objetivo deste estudo foi desenvolver um índice de qualidade do solo associado à recarga de aquíferos. Para isso, utilizou-se o Banco de Dados Hidrofísicos de Solos Brasileiros, a fim de avaliar a integração dos atributos físicos do solo no desenvolvimento do índice. Onze atributos físicos foram avaliados por meio da Análise de Componentes Principais Supervisionada (SPCA). Três atributos — Condutividade Hidráulica Saturada do Solo ( $K_{slab\_log}$ ), Porosidade Drenável (DP) e teor de Silte — foram selecionados para a construção do Índice de Qualidade do Solo para o Potencial de Recarga de Aquíferos (SQIgr). Para duas classes de uso da terra (Agricultura e Floresta), segundo o MapBiomas, e quatro classes de solo de acordo com a SiBCS (G2: Argissolo Vermelho e Argissolo Vermelho-Amarelo; G4: Latossolo Vermelho e Latossolo Amarelo), nos grupos Agr\_G2 e For\_G4 não foi observada diferença estatística significativa entre os valores de SQIgr e BIRn, indicando resultados consistentes para o índice proposto. A viabilidade da aplicação espacial do índice foi validada por meio da comparação da contribuição da vazão de base na vazão total. Os valores de SQIgr para os grupos For\_G4, Agr\_G4 e Agr\_G2 foram, respectivamente, 0,76; 0,58 e 0,50, indicando que o uso agrícola tende a reduzir a qualidade do solo. No entanto, a predominância de Latossolos esteve associada a escores mais elevados de qualidade do solo. Assim, o SQIgr proposto demonstrou ser uma ferramenta fundamental para a gestão sustentável das bacias hidrográficas brasileiras, sendo promissor na distinção de ambientes prioritários para preservação e conservação.

**Palavras -chave:** Estrutura do Solo, Infiltração de água no solo; Serviços Ecossistêmicos

## 1. INTRODUCTION

The sustainable management of groundwater is a topic of great importance for the society. Groundwater accounts approximately 99% of all liquid freshwater on Earth (UNESCO, 2022). It can be used for various purposes, such as public supply, irrigation, livestock farming, industry and mining. However, excessive and inadequate exploitation can lead to its degradation and depletion (Chaminé *et al.*, 2015).

Considering factors that interfere with groundwater recharge potential, Jaafarzadeh *et al.* (2021) evaluated fifteen parameters, among which highlighted the importance of understanding soils, as their permeability affects the amount of water that can infiltrate and recharge the aquifer.

Soil quality is understood as its capacity to promote plant growth, protect watersheds and prevent water and air pollution (Sims *et al.*, 1997). The concept of soil quality goes beyond crop productivity, serving as a tool for evaluating soil functions in the environment, which support the delivery of ecosystem services and environmental sustainability (Bünemann *et al.*, 2018).

Soil quality can be evaluated in relation to one or several soil functions (Baveye *et al.*, 2016; Bouma, 2014), which are influenced by several factors, such as soil texture, soil structure, porosity, bulk density and hydraulic conductivity (McKenzie *et al.*, 2011). The dynamics of these intervening factors, conditioned by physical, chemical and biological processes in the soil, added to the heterogeneity and diversity of land uses and classes (Daniel L. Kelting *et al.*, 1999), adds complexity to the assessment of a Soil Quality Index.

In the context of groundwater recharge, soil quality is often assessed through hydraulic properties, such as the Basic Infiltration Rate, in combination with landscape feature analysis. The Basic Infiltration Rate refers to steady-state rate of water infiltration into the soil, when the infiltration rate remains stable over time. Measuring it requires robust instrumentation and is not straightforward (Jaafarzadeh *et al.*, 2021).

The potential groundwater recharge can also be estimated using the integration of geographic information systems (GIS) techniques and Hierarchical Process Analysis with different types of environmental variability data to construct potential zones for recharge (Guerrón-Orejuela *et al.*, 2023b; Lentswe and Molwalefhe, 2020; Thanh *et al.*, 2022); otherwise, using simplified numerical representation through a Soil Quality Index (SQI), selecting soil attributes that are linked to the recharge process as employed by Santana *et al.* (2023) and Alvarenga *et al.* (2012).

To quantify the SQI, Karlen and Stott (1994) proposed an additive-weighted function model based on the definition of weights for representative indicators of the analyzed function. The application to estimating the potential groundwater recharge based on soil quality assessment (Santana *et al.*, 2023; Alvarenga *et al.*, 2012) presents two sets of different indicators, Soil Hydraulic Conductivity in the laboratory (Ksat\_lab), Basic Infiltration Rate (BIR) and Drainable Porosity in Santana *et al.* (2023); and Soil Hydraulic Conductivity in the field (Ksat\_field), Macroporosity and Bulk Density (BD) in Alvarenga *et al.* (2012). The resulting indices proved reliable for comparative analysis of soil quality, considering different uses and soil classes at watershed and management scales.

However, the methodology adopted by these authors takes into account the opinion of experts when selecting indicators, which can be subjective or biased. Therefore, alternatives such as Principal Component Analysis (PCA), a multivariate technique, have been used to improve the non-linear relationships between the soil quality indicators and the target variable (groundwater recharge) (Abdel-Fattah *et al.*, 2021b; Andrews *et al.*, 2002; Bandyopadhyay and Maiti, 2021; Bünemann *et al.*, 2018; Rinot *et al.*, 2019).

Currently, no studies on the use of these methodologies when associated with groundwater recharge for assessing tropical soils in Brazil, most likely due to the difficulty in accessing data on hydrophysical properties of these soils. To overcome this limitation, in 2018, the Hydrophysical Database for Brazilian Soils (HYBRAS) was published, a database that provides information on the physical and hydraulic properties of Brazilian soils such as soil texture, porosity, bulk density and saturated hydraulic conductivity, among others (Ottoni *et al.*, 2018). This opens promising opportunities for evaluating the physical quality of these soils, as well as for developing management strategies aimed at preserving groundwater resources (Bünemann *et al.*, 2018).

This study tests the hypothesis that soil hydrophysical properties can be used to compose a statistically robust Soil Quality Index for Groundwater Recharge Potential (SQIgr), independent of expert opinion and applicable across the entire national territory using HYBRAS data. To achieve this, the aim is to develop a soil quality index related to groundwater recharge and to validate it using measured values of Basic Infiltration Rate and baseflow patterns across Brazil's river basins. This proposal aims to support the sustainable management of groundwater resources in Brazil, with the aim of supporting the achievement of Sustainable Development Goal (SDG) 6.

## 2. MATERIAL AND METHODS

### 2.1. Data collection

The data were extracted from the Hydrophysical Database for Brazilian Soils (HYBRAS) version 1 (Ottoni *et al.*, 2018) and version 2, currently under development. The focus was on soil properties most relevant to the infiltration process and groundwater recharge, including: Basic Infiltration Rate (BIR\_log) (logarithm of BIR; cm/h); Saturated Hydraulic Conductivity (Ksat\_log) (logarithm of Ksat; cm/h), for methods determined in the laboratory; Macroporosity in cm<sup>3</sup>/cm<sup>3</sup> based on the difference between the total porosity and the volumetric moisture value at 6 kPa; Total Porosity (TP) in cm<sup>3</sup>/cm<sup>3</sup>; Drainable Porosity (DP) in cm<sup>3</sup>/cm<sup>3</sup>, calculated according to Otto (1988), considering the difference between the volumetric moisture at TP and the volumetric moisture at 10 KPa; Structural Stability Index (SSI), in percent (%), calculated according to Reynolds *et al.* (2009); Particle Density (PD) in g/cm<sup>3</sup>; Bulk Density (SD), in g/cm<sup>3</sup>; and, Particle Composition (texture) based on the percentage of Sand, Silt and Clay. In the absence of volumetric moisture data at 6 KPa and 10 KPa for a soil sample in HYBRAS, we used the corresponding result of fitting the sample's water retention data to the van Genuchten (1980) model with Mualem (1976) restriction ( $m=1-1/n$ ), when available in the database; The VG parameters were fitted for the HYBRAS's water retention data that contained at least five data points covering a wide range of water tensions (from 0 to -15000 cm).

### 2.2. Index determination

The Soil Quality Index was determined using the methodology proposed by Karlen and Stott (1994), which comprises three stages: selection of the Minimum Data Set (MDS); scoring of the data using the normalization process according to Wymore (1993); and integration of the indicators to compose the index.

#### 2.2.1. Selecting the Minimum Data Set (MDS)

To select the MDS, the Supervised Principal Component Analysis (SPCA) methodology was applied (Bair *et al.*, 2006), which utilizes correlation analyses with a reference variable to assign weights to the input variables before applying the Principal Component Analysis. Thus, the log of the Basic Infiltration Rate (BIR\_log) in cm/h, was used as a reference for the correlation analyses with the previously ones in section 2.1. It should be noted that, at this stage, the database from HYBRAS accounts for 138 data points sampled with all the input variables (BASE 1).

The Principal Components (PCs) explaining  $\geq 5\%$  of the variability in the data (Wander and Bollero, 1999) were selected. In each PC, only the variables with the highest factor loadings were retained. Highly weighted factor loadings were defined as those with absolute values  $\geq 0.40$  (Wander and Bollero, 1999). When more than one variable was retained in a single PC, correlation coefficients were used to determine whether the variables could be considered redundant (correlation coefficient  $> 0.60$ ) and, if classified as redundant, the variable with the lowest factor loading (absolute value) was eliminated from the MDS (Andrews *et al.*, 2002).

Once all the MDS indicators were chosen, a final check for variable redundancy was carried out. For each variable selected, the factor loadings (in absolute value) of the previously selected PCs were summed. When the correlation coefficient of these variables was  $\geq 0.70$ , they were considered redundant and, thus, the one with the lowest absolute factor load (Andrews *et al.*, 2002; Martín-Sanz *et al.*, 2022).

To carry out the next steps, a new database of the selected indicators was established. It also considered the soil samples collected from the surface layer (0 - 20 cm), thus forming BASE 2, which consists of 336 samples.

### 2.2.2. Scoring and integration of indicators

The indicator scores were given by transforming their values using non-linear scoring functions (SSF), normalizing them on a scale from zero to one (Wymore, 1993), Eq. 1:

$$v = \frac{1}{1 + \left(\frac{B-L}{x-L}\right)^{2S(B+x-2L)}} \quad \text{Eq. 1}$$

Where  $v$  represents the normalized score;  $B$  is the value of the indicator when the normalized score ( $v$ ) is 0.5, specifically the average of the respective indicator;  $L$  is the lowest observed value of the indicator;  $S$  is the slope of the tangent line to the curve at  $B$ , meaning the second derivative of equation 1 equal to 0; and  $x$  is the value of the soil indicator associated with the normalized value ( $v = 0.5$ ), where  $x = B$ .

To determine the parameters of the curve, all available samples in HYBRAS were considered for each variable without applying any filters. In other words, the parameters were based on the environmental variability of the data, which includes a diversity of texture classes, soil classes, land uses, biomes, depths, among others. The base value ( $B$ ) was defined as the

mean of the analyzed variable; the value of L as the lowest observed value; and U as the highest observed value.

The calculation of the S value for each curve was performed using the Solver tool in Microsoft Excel, taking into account the final requirements of  $\vartheta = 0.5$  and  $x \approx B$ . To do this, the Mean (B) and Standard Deviation (SD) of each indicator were initially defined in order to determine the initial x value (Eq. 2). Then, the S value was calculated based on the initial x using Eq. 3. In cases where the indicator was considered “more is better,” the S value was positive; on the other hand, when the indicator was considered “less is better,” the S value was negative. The sign (positive or negative) was applied based on the correlation of the variable with the Basic Infiltration Rate (BIR). Thus, the Solver tool was used with the objective of setting  $\vartheta = 0.5$  by adjusting the variables S and  $x_{initial}$  so that the latter would approximate the B value.

$$x_{initial} = 0,5 * SD + B \quad \text{Eq. 2}$$

$$S = \frac{\log\left(\frac{1}{\vartheta}\right) - 1}{\log\log\left(\frac{B-L}{x-L}\right) * 2(B+x-2L)} \quad \text{Eq. 3}$$

After scoring the indicators, the SQI was calculated by multiplying the score of each indicator ( $v$ ) by its respective weight ( $p$ ) (Eq. 2). The weight of each indicator is defined by the ratio of the cumulative variance value to which the indicator belongs to the cumulative variance of the selected PCs. This proportion of the variances comes from evaluating the principal component analysis (MDS selection).

$$SQI_{gr} = v_1 x p_1 + v_2 x p_2 + v_3 x p_3 \quad \text{Eq. 4}$$

Where  $SQI_{gr}$  refers to the Soil Quality Index for assessing groundwater recharge potential.

## 2.2.1. Validation of the SQI for groundwater recharge (SQI<sub>gr</sub>)

### 2.2.1.1. Validation using BIR<sub>n</sub>

SQI<sub>gr</sub> was validated using the Basic Infiltration Rate values present in HYBRAS and collected in the surface layer. The BIR values were normalized according to Eq. 1 and the its

parametrization (L, B and S). The normalized BIR is referred as BIR<sub>n</sub>. It is understood that the higher the value of BIR<sub>n</sub>, the higher the value of SQI<sub>gr</sub> for groundwater recharge, since the groundwater recharge fundamentally depends on the permeability and infiltration capacity of the soil.

Given the limited number of HYBRAS samples representing the diversity of uses and soils in Brazil, it was often not feasible to obtain BIV values and the indicators that constitute the index for BASE 2 for the same sample. Thus, we opted for validation considering a comparison of the BIR<sub>n</sub> and SQI<sub>gr</sub> values, using box-plot graphs, under different data grouping conditions. The groupings used in this analysis were combinations of land use and soil classes according to SiBCS (second categorical level), which contained a minimum of ten soil samples from HYBRAS. To maximize the number of classes grouped in the validation, we grouped soil classes according to the similarity BIR values by the clustering method (K-modes), using the normalized BIR as a reference (Huang, 1997), described as: G1 corresponding to the *Neossolo Flúvico* class; G2 the *Argissolo Vermelho* and *Argissolo Vermelho-Amarelo* classes; G3 the *Gleissolo Háptico* and *Cambissolo Flúvico* classes; and, G4 the *Latossolo Amarelo* and *Latossolo Vermelho* classes.

The predominant land use classes in the HYBRAS database selected in the context of this study were Agriculture (Agr) and Forest (For) classes. In this way, the following combinations were created to validate the SQI<sub>gr</sub> data, considering the restriction of ten soil samples in each group: Agr\_G1, Agr\_G2, Agr\_G3, Agr\_G4, For\_G2 and For\_G4

The HYBRAS soil classes were described in accordance with the MapBiomias Project (Souza *et al.*, 2020) and the soil classes according to SiBCS - 5th edition (Santos *et al.*, 2018). The similarity between BIR<sub>n</sub> and ISQ<sub>ra</sub> in each soil grouping was analyzed using the Mann-Whitney statistical test (Mann and Whitney, 1947). The test was performed using RStudio software (version 4.4), and the comparisons between the indicators (SQI<sub>gr</sub> and BIR<sub>n</sub>) were adjusted using the Bonferroni method to control Type I error. The adjusted p-values were used to determine statistical significance, with results considered significant when the adjusted p-value was less than 0.05.

### **2.2.1.2. Validation through baseflow pattern**

A second validation process was carried out to evaluate the possibility of spatializing the SQI<sub>gr</sub> by integrating the diversity of soils and land uses in Brazil. Initially, 35 basins were randomly selected throughout Brazil. Next, a filter was applied to select those that met the representativeness requirement of > 50% in the predominant land uses in the study database

(Agriculture or Forestry). In this case, the MapBiomass map (Souza *et al.*, 2020) was used to extract the land use classes covering the river basins selected in the study. After this, a new filter was created to select the basins with more than 50% representation of the soil classes, G2 or G4. The soil classes determined in the basins were taken from the Soil Map of Brazil 1:5,000,000 (Santos *et al.*, 2011).

From these filters, three basins were selected for analysis, and their historical flow series were obtained from the HIDROWEB platform of the Brazilian National Water Agency (ANA). For each basin, the hydrological year was selected based on the sampling year of the SQIgr data points located within the basin. The same year was considered for extracting land use data from the MapBiomass platform.

The mean annual runoff (Rf) of the basins was calculated by averaging the monthly flow over the hydrological year. The corresponding baseflow (Bf) was then derived from the runoff (Rf), assuming an exponential pattern of the baseflow by identifying the inflection points in the hydrograph (Hümann *et al.*, 2011). Since baseflow is a groundwater recharge function, this methodological approach allows for validating the SQIra. The higher the index, the greater the baseflow contribution, reflecting the water stock accumulated during the rainy season.

### 3. RESULTS AND DISCUSSION

#### 3.1. Selecting the Minimum Data Set (MDS)

Figure 2.1 shows the correlation matrix used in the MDS selection process. The data in this figure indicates that BIR\_log is directly proportional and significant with Soil Hydraulic Conductivity (Kslab\_log), Total Porosity (TP), Macroporosity (Macro) and Drainable Porosity (DP). On the other hand, Silt content showed a negative correlation, which reduces soil infiltration capability, as well as the groundwater recharge.

Regarding TP, a positive correlation with BIR was also found by Basset *et al.* (2023). They highlight a relationship between BIR and soil texture, in which an increase in TP matches with an increase in the percentage of fine particles (silt and clay). This aspect is also clear in our study. The significant negative correlation between BIR\_log and silt content can be attributed to the high silt contents, which normally is associated with reduced aggregate stability and increased surface crusting, altering the initial values in the infiltration rate (van Es *et al.*, 1991). It should be noted that the correlation values between BIR and silt content were very low; however, we will adopt this trend between BIR and silt data, aiming to normalize scores for the silt indicator (see section 2.2.2).

**Fig. 2.1** Correlation matrix of the variables used for the selection process and BIR\_log as the reference variable. Values with an asterisk '\*' correspond to significant correlations between the variables.

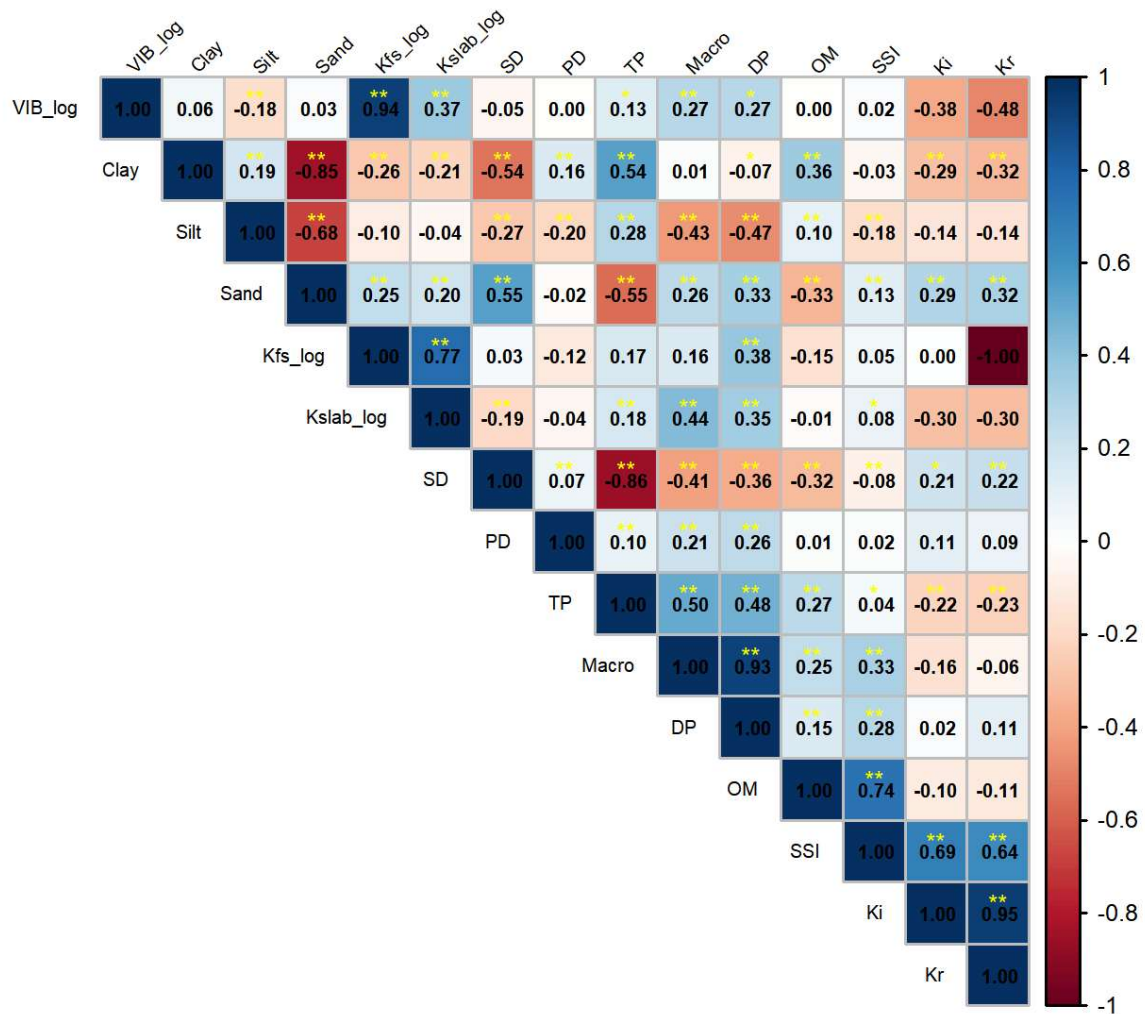


Table 2.1 shows the results of the supervised principal component analysis of the soil quality indicators, highlighting the first three PCs that explained  $> 0.05$  of the variance, and which were selected for the analysis.

The variables with the greatest weight in PC1 were Kslab\_log, Macro and DP, the two latter significantly correlated ( $r = 0.93^*$ , Fig 2.1). Macro showed the highest factor loading in relation to DP and was retained for the MDS along with Kslab\_log. In PC2, the same variables were repeated, but DP had a higher factor loading than Macro. Consequently, Kslab\_log and Macro were kept in the MDS, and DP was added to PC2. Only Silt was retained for the MDS with a higher factor load in PC3. Thus, the first selection of soil quality indicators comprised by Kslab\_log, Macro, DP and Silt.

**Table 2.1.** Results of the Supervised Principal Component Analysis of soil quality indicators.

<b>Principal Components</b>	<b>PC1</b>	<b>PC2</b>	<b>PC3</b>	<b>PC4</b>
Standard Deviation	0.476	0.248	0.184	0.102
Proportion of Variance <sup>a</sup>	<b>0.672</b>	<b>0.182</b>	<b>0.100</b>	0.031
Cumulative Variance	0.672	0.854	0.954	0.985
<b>Variables</b> <sup>b,c</sup>				
Clay	-0.019	-0.036	0.025	-0.422
Silt	-0.112	-0.042	<b><u>-0.904</u></b>	-0.377
Sand	-0.002	0.011	-0.096	0.147
Kslab_log	<b><u>-0.665</u></b>	<b>-0.730</b>	0.055	0.142
BD	-0.060	-0.007	0.119	-0.288
PD	0.000	0.000	0.000	0.000
TP	-0.166	0.031	0.393	-0.731
Macro	<b>-0.513</b>	<b>0.450</b>	-0.007	0.053
DP	<b>-0.500</b>	<b><u>0.510</u></b>	-0.010	0.134
OM	0.002	-0.002	-0.008	0.007
SSI	-0.014	0.020	0.034	-0.016

*a*: The eigenvalues in bold correspond to the PCs analyzed for the index.

*b*: Factor loadings in bold are considered highly weighted.

*c*: The factor loadings in bold and italics correspond to the indicators included in the MDS

Ksat\_log: Saturated Hydraulic Conductivity in log; BD: Bulk Density; PD: Particle Density; TP: Total Porosity; Macro: Macroporosity; DP: Drainable Porosity; OM: Organic Matter; SSI Structural Stability Index.

The variables with the greatest weight in PC1 were Kslab\_log, Macro and DP, the two latter significantly correlated ( $r = 0.93^*$ , Fig 2.1). Macro showed the highest factor loading in relation to DP and was retained for the MDS along with Kslab\_log. In PC2, the same variables were repeated, but DP had a higher factor loading than Macro. Consequently, Kslab\_log and Macro were kept in the MDS, and DP was added to PC2. Only Silt was retained for the MDS with a higher factor load in PC3. Thus, the first selection of soil quality indicators comprised by Kslab\_log, Macro, DP and Silt.

In the final correlation analysis between the selected indicators, DP and Macro were highly correlated ( $> 0.70$ ). The absolute sum of the factor loadings for the three PCs showed the highest value for DP, which was kept for the MDS. Thus, the final MDS was composed by Kslab\_log, DP and Silt.

In studies carried out in Brazilian soils, Alvarenga *et al.* (2012) selected several similar indicators by expert opinion, including Kslab, SD and Macro, and Santana *et al.* (2023) used BIR, Kslab and DP. It should be noted that macroporosity is physically close to drainable porosity, both indicate macro-spaces that guarantee water drainage in the soil profile and

subsequent deep percolation. Therefore, the indicators selected are appropriate for assessing the potential for groundwater recharge.

### 3.2. Scoring and integration of indicators

The parameters used to construct the normalization curves for each indicator are in Table 2.2, along with the basic statistics of the indicators, such as the mean, coefficient of variation and maximum and minimum values. Based on these parameters, the indicator data (BASE 2) was converted into normalized values ( $v$ ) and used to calculate the SQIra.

**Table 2.2.** Parameters for construct the normalization curves and basic statistics of the data.

Indicators	L	B	S	Mean	CV (%)	Max	Min
Kslab_log ( $cm\ h^{-1}$ )	-2.08	0.48	0.87	0.48	180.83	2.61	-2.08
DP ( $cm^3\ cm^{-3}$ )	0.01	0.21	6.98	0.21	52.88	0.86	0.01
Silt (%)	0.00	17.80	-0.10	17.80	68.80	86.00	0.00

After normalizing the indicator values, weights were assigned to each indicator, considering the proportion of variance explained by each PC (Table 2.2) in relation to the sum of the three PCs. PC1 refers to the Kslab\_log indicator, PC2 corresponds to DP and PC3 represents Silt. Finally, the SQIgr was composed and calculated using Eq. 3. It is important to highlight that the values of  $Kslab\_log_n$  and  $DP_n$  were derived from a “more is better” response curve. Conversely,  $Silt_n$  was derived from a “less is better” curve, indicating that higher actual silt contents result in lower normalized values, approaching zero.

$$SQI_{gr} = 0.70 Kslab\_log_n + 0.19 DP_n + 0.11 Silt_n \quad \text{Eq. 3}$$

The subscript  $n$  represents the value of the normalized indicator obtained using the aforementioned parameters. The SQI was calculated for 336 samples with data on the Kslab\_log, DP, and Silt

### 3.3. Validation of the SQIgr

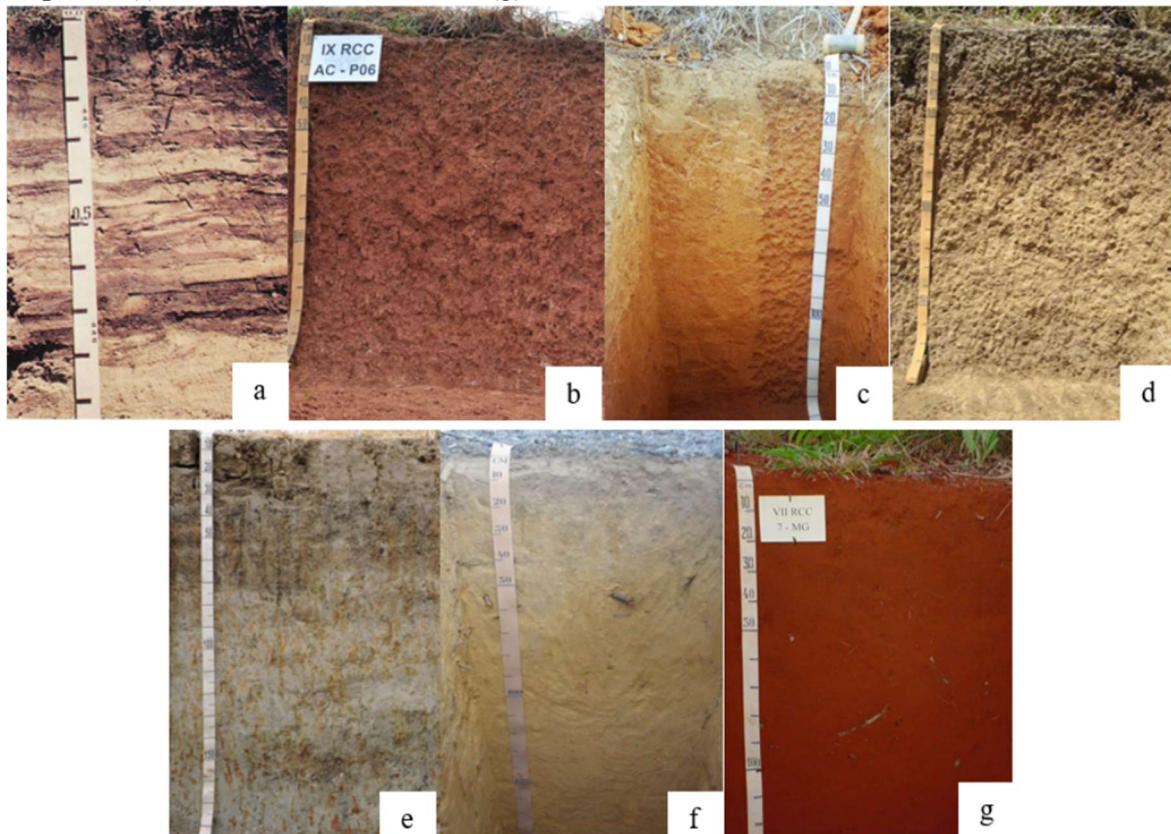
#### 3.3.1. Simplification of soil classes

Seven soil classes with BIRn values in HYBRAS have more than ten samples, e.g., *Neossolo Flúvico*; *Argissolo Vermelho*; *Argissolo Vermelho-Amarelo*; *Cambissolo Flúvico*; *Gleissolo Háptico*; *Latossolo Amarelo*; and, *Latossolo Vermelho*. These soils were used to simplify the classes so that they could be used to validate the SQIra. The clustering process

resulted in four groups of soils that were similar to each other for estimating the groundwater recharge potential. Fig 2 shows the representative profiles of each of these soils.

Some authors also simplified the soil classes to classify their influence on groundwater recharge (Costa *et al.*, 2022; Menezes *et al.*, 2025; Tenenwurcel, 2021). Table 2.3 shows the classification of groundwater recharge potential for different groups of soils. It is also included the results suggested in this study, which considered the average BIRn values for ranking in the recharge potential groups. The BIRn averages were 0.651 (G4), 0.616 (G2), 0.402 (G3), and 0.303 (G1).

**Figure 2.2.** Representative profiles of the soil classes analyzed (a) *Neossolo Flúvico*; (b) *Argissolo Vermelho*; (c) *Argissolo Vermelho-Amarelo*; (d) *Cambissolo Flúvico*; (e) *Gleissolo Háplico*; (f) *Latossolo Amarelo*; and, (g) *Latossolo Vermelho*



Source: Santos *et al.* (2018)

*Neossolo Flúvico* (RY) showed no similarity in terms of BIRn if compared to the other groups in the study database. The magnitudes of BIRn were low in RY, being classified as a soil with low groundwater recharge potential, a result consistent with Menezes *et al.* (2025) and Tenenwurcel (2021). This soil is formed by overlapping layers of sediment, and has an A horizon over a C horizon with a fluvial characteristic at a depth of 150 cm. These soils have a high texture and other physical properties variability, and a low BIRn conditioned by the soil

use on the surface and/or the intrinsic characteristics of the soil profile. Overall, studies characterizing the hydraulic properties of these soils are still not widely known in Brazil (?) (Araújo Pedron *et al.*, 2011).

**Table 2.3.** Comparative grouping of soil classes in terms of recharge potential for Brazilian soils

Potential for groundwater recharge	BIRn	MENEZES <i>et al.</i> , 2025	TENENWURCEL <i>et al.</i> , 2020	COSTA <i>et al.</i> , 2017
Very high	LA, LV (G4)	LA, LB, LV, LVA, NB, NV, NX	LVA	<i>Latossolo, Nitossolo</i>
High	PV, PVA (G2)	PA, PV, PVA	CX, RQ	<i>Argissolo, Espedossolo, Cambissolo</i>
Medium	CY, GX (G3)	RL, RR, CX, CH	RL	<i>Neossolo, Chernossolo</i>
Low	RY (G1)	CY, GM, GX, OX and RY	RY, GM	<i>Vertissolo, Planossolo, Gleissolo, Luvissolo, Organossolo, Plintossolo</i>

*CX - Cambissolos Háplicos, CH - Cambissolos Húmicos, CY - Cambissolos Flúvicos, GM - Gleissolos Melânicos, GX - Gleissolos Háplicos, LA - Latossolos Amarelos, LB - Latossolos Brunos, LVA - Latossolos Vermelho-Amarelos, LV - Latossolos Vermelhos, NB - Nitossolos Brunos, NV - Nitossolos Vermelhos, NX - Nitossolos Háplicos, PA - Argissolos Amarelos, OX - Organossolos Háplicos, PV - Argissolos Vermelhos, PVA - Argissolos Vermelho-Amarelos, RL - Neossolos Litólicos, RQ - Neossolo Quartzarênico, RR - Neossolos Regolíticos e RY - Neossolos Flúvicos.*

The *Argissolos* were grouped in the high groundwater recharge group (G2). *Argissolo Vermelho* (PV) and *Argissolo Vermelho-Amarelo* (PVA) have a deep red color due to their high iron oxide content. They have a thick B horizon due to the more intense accumulation of clay. They are found in regions with high rainfall, with favorable leaching conditions and represent areas suitable for groundwater recharge. Assessing non-linear regression models for groundwater recharge, Nolan *et al.* (2007) found a positive correlation between *Argissolos* and groundwater recharge potential. Despite great potential for groundwater recharge of these soils, the greater presence of kaolinite minerals that condition their blocky structure has constrained the water percolation. These minerals present face-to-face contact of the kaolinite sheets, reducing the permeability (Costa *et al.*, 2022; Menezes *et al.*, 2025).

*Cambissolo Flúvico* (CY) and *Gleissolo Háplico* (GX) were grouped in the medium groundwater recharge potential class based on the results of the infiltration value cluster analysis. CY is a soil with a fluvial character, developed from alluvial sediments in the top 120 cm of the soil. GX has a gley horizon and is located in the relatively lower parts of the alluvial

plain. Both soils have limited drainage conditions, resulting in reduced infiltration capacity. Despite these limitations, the cluster analysis identified a moderate recharge pattern for both soils, indicating that, although the percolation of water is less intense compared to soils with higher permeability, it is still sufficient to support moderate groundwater recharge. This classification reflects the fact that, while these soils do not present a high recharge potential, they have an intermediate infiltration capacity that favors groundwater recharge. This contrasts with the discharge environment classification assigned by Menezes *et al.* (2025), but reinforces the idea that even in soils with drainage limitations, a moderate contribution to groundwater recharge can still be observed.

The *Latossolos* group was classified into the very high groundwater recharge potential class, with *Latossolo Amarelo* (LA) and *Latossolo Vermelho* (LV). These soils are deep and exhibit a relatively homogeneous clay distribution, which facilitates water infiltration (Cambráia Neto and Rodrigues, 2020; Krishnaswamy *et al.*, 2013; Mota *et al.*, 2012; Pessoa and Libardi, 2022). Additionally, *Latossolos* with a granular structure, characterized by a higher presence of gibbsite, have a high infiltration capacity, allowing for significant groundwater recharge (Ferreira *et al.*, 1999a, 1999b; Ottoni *et al.* 2024). In contrast, *Latossolos* with a blocky structure, more common in soils with a predominance of kaolinite, have greater resistance to water percolation due to their higher cohesion and lower porosity. Nonetheless, their great depth and homogeneous composition favor groundwater recharge, particularly in areas with high rainfall and favorable drainage conditions.

### 3.3.1.1. Validation using BIRn

After simplifying the soil classes, they were divided into Agriculture and Forestry. Therefore, the central tendency and variance measures for VIBn and SQIra were calculated for each grouping with a sample greater than 10 (Table 2.4). Only 3 groupings were selected for SQIgr. Considering the simplification of the soil classes within each use, the groups follow the trend presented in the literature for Brazilian soils (Costa *et al.*, 2022; Menezes *et al.*, 2025; Tenenwurcel, 2021) with increasing average values, i.e., G1 (NY) < G3 (CY and GX) < G2 (PV and PVA) < G4 (LA and LV).

The comparative analysis between the BIRn and SQIgr (Fig. 2.3) aimed to evaluate the applicability of using SQIgr as a substitute for BIRn in identifying areas with groundwater recharge potential. Comparisons were made within the Agr\_G2, Agr\_G4, and For\_G4 groupings.

**Table 2.4.** Measures of central tendency and variance for BIRn and SQIgr for each cluster with  $n > 10$

Groupings	n	Min	Q1	Median	Q3	Max	Mean	CV (%)
<i>BIRn</i>								
<b>Agr_G1</b>	23	0.00	0.00	0.08	0.62	0.98	0.30	116.82
<b>Agr_G2</b>	34	0.01	0.37	0.66	0.91	0.98	0.61	51.01
<b>Agr_G3</b>	45	0.00	0.02	0.36	0.79	0.97	0.40	92.06
<b>Agr_G4</b>	111	0.00	0.47	0.72	0.93	0.99	0.66	42.17
<b>For_G4</b>	12	0.00	0.24	0.76	0.82	0.98	0.56	69.44
<i>SQIgr</i>								
<b>Agr_G2</b>	23	0.02	0.37	0.55	0.67	0.90	0.50	47.88
<b>Agr_G4</b>	78	0.01	0.44	0.59	0.76	0.91	0.58	36.50
<b>For_G4</b>	19	0.28	0.71	0.85	0.89	0.97	0.76	29.06

Agr\_G1: Agriculture + *Neossolos Flúvicos* (NY); Agr\_G2: Agriculture + *Argissolo Vermelho* (PV) and *Argissolo Vermelho-Amarelo* (PVA); Agr\_G3: Agriculture + *Cambissolo Flúvico* (CY) and *Gleissolo Háplico* (GX); Agr\_G4: Agriculture + *Latossolo Amarelo* (LA) and *Latossolo Vermelho* (LV); For\_G4: Forest + *Latossolo Amarelo* (LA) and *Latossolo Vermelho* (LV)

n: number of samples; Min: minimum value; Q1: first quartile; Q3: third quartile; Max: maximum value; CV: coefficient of variation

The comparative analysis between the BIRn and SQIgr (Fig. 2.3) aimed to evaluate the applicability of using SQIgr as a substitute for BIRn in identifying areas with groundwater recharge potential. Comparisons were made within the Agr\_G2, Agr\_G4, and For\_G4 groupings.

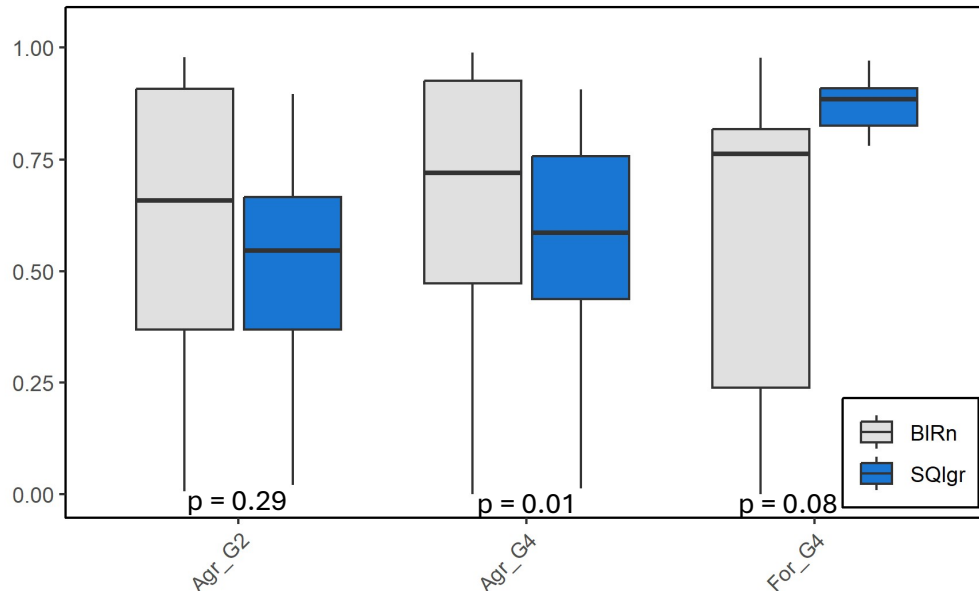
In the Agr\_G2 grouping, which represents agricultural areas with PVA and PV, the comparison resulted in an adjusted p-value of 0.29. This suggests that, in this scenario, the values of SQIgr and BIRn are relatively close. Therefore, using SQIgr can be considered for evaluate the groundwater recharge potencial. On the other hand, in the Agr\_G4 group, corresponding to agricultural areas with the presence of LA and LV, there was a significant difference between the BIRn and SQIgr (p-value = 0.01), which compromises the possibility of direct substitution, thus recommending further validation.

In the For\_G4 group, representing native vegetation areas in LA and LV, the test did not show a significant difference (p-value = 0.08), indicating similarity between the BIRn and SQIgr patterns in forested areas.

The SQIgr appears promising as an indirect tool for assessing recharge, but its application as a substitute for BIRn should take into account specific land use and soil groupings. It should be noted that although SQIgr was only validated for the Agr\_G2, Agr\_G4

and For\_G4 groupings, these three represent a large part of the environmental conditions in Brazil.

**Figure 2.3.** Evaluation of the medians between BIRn and SQIgr for use groupings (Agriculture - Agr and Forest - For) and simplified soil classes (G2 - PV and PVA and G4 - LV and LA).



Agr\_G2: Agriculture + *Argissolo Vermelho* (PV) and *Argissolo Vermelho-Amarelo* (PVA);  
 Agr\_G4: Agriculture + *Latossolo Amarelo* (LA) and *Latossolo Vermelho* (LV); For\_G4:  
 Forest + *Latossolo Amarelo* (LA) and *Latossolo Vermelho* (LV)

### 3.3.2. Validation through baseflow pattern

The validation process to verify the viability of the spatial coverage of the SQIgr also proved to be promising, since in the evaluation of the three groupings, the Bf/Rf ratio was close to the SQIgr values. The values of the Bf/Rf ratio and the average SQIgr for the corresponding grouping are shown in Fig 2.4.

In the Flo\_G4 basin, where forest cover predominates (52%) and agricultural use is reduced (38.3%), the highest value of Bf/Rf (0.76) was observed, along with an equally high SQIgr (0.76). This indicates more favorable environmental conditions and lower soil degradation. The presence of *Latossolos* in this basin, combined with preserved native vegetation, contributes to maintaining infiltration capacity. The Agro\_G4 basin, although characterized by the same soil group (G4), exhibits intensive agricultural use (82.7%) and lower forest cover (13.6%), resulting in intermediate values of Bf/Rf (0.61) and SQIgr (0.58). This highlights the influence of land use on the hydrological functioning of soils.

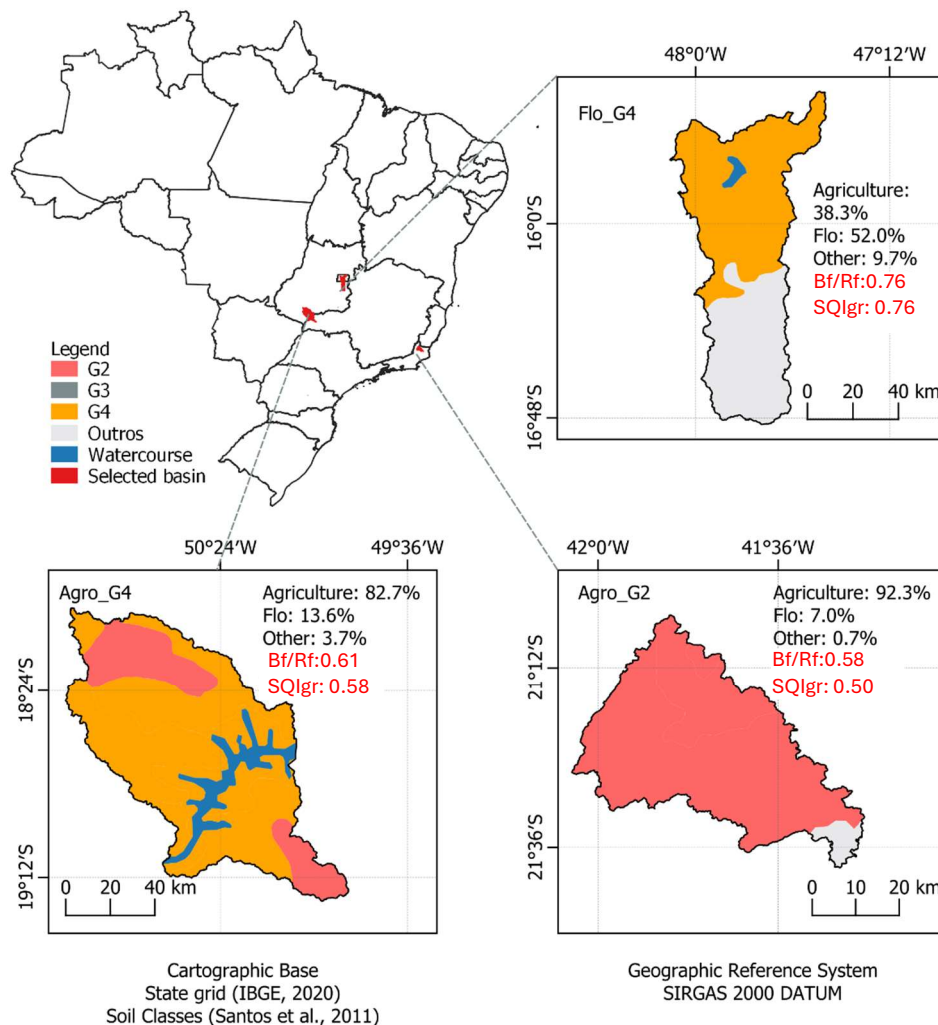
In contrast, the Agro\_G2 basin, with even more intensive agricultural use (92.3%) and soils from the G2 group (PVA and PA), presented the lowest values of both Bf/Rf (0.58) and

SQIgr (0.50). This reflects a condition of greater soil degradation, which may be associated with the lower infiltration capacity and loss of physical quality typical of these soils.

The results indicate a positive trend between Bf/Rf and SQIgr, reinforcing the potential use of the SQIgr as a substitute or complementary indicator to the Bf/Rf ratio in environmental assessments of watershed conditions. Furthermore, it is essential to consider the intrinsic characteristics of soils when interpreting these indicators. More structured and porous soils, such as *Latossolos* (G4), tend to perform better even under agricultural use. These findings are supported by Araújo (2006) and Alvarenga *et al.* (2012), who reported high recharge potential in areas dominated by *Latossolos*, as well as in regions with prevailing forest cover.

Thus, the integrated analysis of soil type, land use, and hydrological indicators is essential for assessing environmental quality and the hydrogeological functioning of landscapes.

**Figure 2.4.** Evaluation of Bf/Rf ratio values with average SQIgr values for use groupings (Agriculture - Agr and Forest - For) and simplified soil classes (G2 - PV and PVA and G4 - LV and LA).



#### 4. CONCLUSION

1. The indicators that make up the index, Ksat\_log, DP and Silt, were selected without subjective analysis and are representative of the entire national territory. Furthermore, they are soil properties that do not require robust equipment to measure;
2. The index captured the variability found in the sub-basins and, despite the limitations in the validation process, it is an unprecedented index, which considers the variability of infiltration conditions throughout the country based on statistical models;
3. Thus, assessing soil quality using the proposed index can be used as a fundamental tool for the sustainable management of sub-basins in line with SDG6. It can be used to identify areas with the greatest potential for groundwater recharge in the country, either to define areas for conservation and preservation, or to help in the process of valuing soil in relation to the ecosystem service.

#### ACKNOWLEDGEMENTS

*We gratefully acknowledge the financial support of the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Conselho Nacional de Pessoal de Nível Superior (CAPES), Conselho Nacional de Desenvolvimento Científico e Desenvolvimento Científico e Tecnológico (CNPq), Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG), Department of Soil Science of the Federal University of Lavras (DCS - UFLA), Geological Survey of Brazil (SGB) and Federal University of Lavras (UFLA). This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Process n88887.845541/2023-00. This study was carried out with support from the Conselho Nacional de Desenvolvimento Científico e Desenvolvimento Científico e Tecnológico (CNPq) and the Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG).*

#### REFERENCES

- Abdel-Fattah, M.K., Mohamed, E.S., Wagdi, E.M., Shahin, S.A., Aldosari, A.A., Lasaponara, R., Alnaimy, M.A., 2021. Quantitative Evaluation of Soil Quality Using Principal Component Analysis: The Case Study of El-Fayoum Depression Egypt. **Sustainability** 13, 1824. <https://doi.org/10.3390/su13041824>
- AGÊNCIA NACIONAL DE ÁGUAS – ANA, n.d. Hidroweb: Sistemas de informações hidrológicas [WWW Document]. URL <http://hidroweb.ana.gov.br/> (accessed 1.24.25).
- Alvarenga, C.C., Mello, C.R. de, Mello, J.M. de, Silva, A.M. da, Curi, N., 2012. Índice de qualidade do solo associado à recarga de água subterrânea (SQI RA) na Bacia Hidrográfica do Alto Rio

- Grande, M.G. **Rev Bras Cienc Solo** 36, 1608–1619. <https://doi.org/10.1590/S0100-06832012000500025>
- Andrews, S.S., Karlen, D.L., Mitchell, J.P., 2002. A comparison of soil quality indexing methods for vegetable production systems in Northern California. **Agric Ecosyst Environ** 90, 25–45. [https://doi.org/10.1016/S0167-8809\(01\)00174-8](https://doi.org/10.1016/S0167-8809(01)00174-8)
- Araújo, A.R. de, 2006. **Solos da bacia do Alto Rio Grande (MG): base para estudos hidrológicos e aptidão agrícola**. Universidade Federal de Lavras, Lavras.
- Araújo Pedron, F. de, Fink, J.R., Rodrigues, M.F., Azevedo, A.C. de, 2011. Hydraulic conductivity and water retention in leptosols-regosols and saptrolite derived from sandstone, Brazil. **Rev Bras Cienc Solo** 35, 1253–1262. <https://doi.org/10.1590/S0100-06832011000400018>
- Bair, E., Hastie, T., Paul, D., Tibshirani, R., 2006. Prediction by Supervised Principal Components. *J Am Stat Assoc* 101, 119–137. <https://doi.org/10.1198/016214505000000628>
- Bandyopadhyay, S., Maiti, S.K., 2021. Application of statistical and machine learning approach for prediction of soil quality index formulated to evaluate trajectory of ecosystem recovery in coal mine degraded land. **Ecol Eng** 170, 106351. <https://doi.org/10.1016/J.ECOLENG.2021.106351>
- Basset, C., Abou Najm, M., Ghezzehei, T., Hao, X., Daccache, A., 2023. How does soil structure affect water infiltration? A meta-data systematic review. **Soil Tillage Res** 226, 105577. <https://doi.org/10.1016/J.STILL.2022.105577>
- Baveye, P.C., Baveye, J., Gowdy, J., 2016. Soil “Ecosystem” Services and Natural Capital: Critical Appraisal of Research on Uncertain Ground. **Front Environ Sci** 4. <https://doi.org/10.3389/fenvs.2016.00041>
- Bouma, J., 2014. Soil science contributions towards Sustainable Development Goals and their implementation: linking soil functions with ecosystem services. **Journal of Plant Nutrition and Soil Science** 177, 111–120. <https://doi.org/10.1002/jpln.201300646>
- Bünemann, E.K., Bongiorno, G., Bai, Z., Creamer, R.E., De Deyn, G., de Goede, R., Flesskens, L., Geissen, V., Kuyper, T.W., Mäder, P., Pulleman, M., Sukkel, W., van Groenigen, J.W., Brussaard, L., 2018. Soil quality – A critical review. **Soil Biol Biochem** 120, 105–125. <https://doi.org/10.1016/j.soilbio.2018.01.030>
- Cambraia Neto, A.J., Rodrigues, L.N., 2020. Evaluation of groundwater recharge estimation methods in a watershed in the Brazilian Savannah. **Environ Earth Sci** 79, 140. <https://doi.org/10.1007/s12665-020-8884-x>
- Chaminé, H.I., Carvalho, J.M., Teixeira, J., Freitas, L., 2015. Role of hydrogeological mapping in groundwater practice: back to basics. **European Geologist** 40, 34–42.
- Costa, A.M. da, Viana, J.H.M., Evangelista, L.P., Carvalho, D.C. de, Pedras, K.C., Horta, I. de M., Salis, H.H. de C., Pereira, M.P.R., Sampaio, J.L.D., 2022. Ponderação de variáveis ambientais para a determinação do Potencial de Uso Conservacionista para o Estado de Minas Gerais. **Revista Geografias** 13, 118–133. <https://doi.org/10.35699/2237-549X..13439>
- Guerrón-Orejuela, E.J., Rains, K.C., Brigino, T.M., Kleindl, W.J., Landry, S.M., Spellman, P., Walker, C.M., Rains, M.C., 2023. Mapping Groundwater Recharge Potential in High Latitude

- Landscapes Using Public Data, Remote Sensing, and Analytic Hierarchy Process. **Remote Sens (Basel)** 15. <https://doi.org/10.3390/rs15102630>
- Huang, Z.X., 1997. Clustering Large Data Sets with Mixed Numeric and Categorical Values, in: **Proceedings of the 1st Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD)**. Trondheim, pp. 21–34.
- Hümann, M., Schüler, G., Müller, C., Schneider, R., Johst, M., Caspari, T., 2011. Identification of runoff processes – The impact of different forest types and soil properties on runoff formation and floods. **J Hydrol (Amst)** 409, 637–649. <https://doi.org/10.1016/J.JHYDROL.2011.08.067>
- Jaafarzadeh, M.S., Tahmasebipour, N., Haghizadeh, A., Pourghasemi, H.R., Rouhani, H., 2021. Groundwater recharge potential zonation using an ensemble of machine learning and bivariate statistical models. **Sci Rep** 11. <https://doi.org/10.1038/s41598-021-85205-6>
- Karlen, D.L., Stott, D.E., 1994. A Framework for Evaluating Physical and Chemical Indicators of **Soil Quality**. pp. 53–72. <https://doi.org/10.2136/sssaspecpub35.c4>
- Kelting, D.L., Burger, J.A., Patterson, S.C., Aust, W.M., Miwa, M., Trettin, C.C., 1999. Soil quality assessment in domesticated forests – a southern pine example. **For Ecol Manage** 122, 167–185. [https://doi.org/10.1016/S0378-1127\(99\)00040-7](https://doi.org/10.1016/S0378-1127(99)00040-7)
- Krishnaswamy, J., Bonell, M., Venkatesh, B., Purandara, B.K., Rakesh, K.N., Lele, S., Kiran, M.C., Reddy, V., Badiger, S., 2013. The groundwater recharge response and hydrologic services of tropical humid forest ecosystems to use and reforestation: Support for the “infiltration-evapotranspiration trade-off hypothesis.” **J Hydrol (Amst)** 498, 191–209. <https://doi.org/10.1016/J.JHYDROL.2013.06.034>
- Lentswe, G.B., Molwalefhe, L., 2020. Delineation of potential groundwater recharge zones using analytic hierarchy process-guided GIS in the semi-arid Motloutse watershed, eastern Botswana. **J Hydrol Reg Stud** 28. <https://doi.org/10.1016/j.ejrh.2020.100674>
- Mann, H.B., Whitney, D.R., 1947. On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other. **The Annals of Mathematical Statistics** 18, 50–60. <https://doi.org/10.1214/aoms/1177730491>
- Martín-Sanz, J.P., de Santiago-Martín, A., Valverde-Asenjo, I., Quintana-Nieto, J.R., González-Huecas, C., López-Lafuente, A.L., 2022. Comparison of soil quality indexes calculated by network and principal component analysis for carbonated soils under different uses. **Ecol Indic** 143, 109374. <https://doi.org/10.1016/J.ECOLIND.2022.109374>
- McKenzie, B.M., Tisdall, J.M., Vance, W.H., 2011. Soil Physical Quality. **Springer Nature**, pp. 770–777. [https://doi.org/10.1007/978-90-481-3585-1\\_153](https://doi.org/10.1007/978-90-481-3585-1_153)
- Mello, C.R. de M., da Silva, A.M., 2013. **Hidrologia: princípios e aplicações em sistemas agrícolas**. UFLA, Lavras.
- Menezes, M.D., Silva, S.H.G.S., Gelsleichter, Y.A., Monteiro, M.E.C., Ottoni, M. V., 2025. **Potentials and geospatial analyses of aquifers recharge in the South region of Minas Gerais state, Brazil**.

- Mota, F.O.B., Ness, R.L.L., Mota, J.C.A., Clemente, C.A., Sousa, S.C., 2012. Physical quality of an oxisol under different uses. **Rev Bras Cienc Solo** 36, 1828–1835. <https://doi.org/10.1590/S0100-06832012000600017>
- Mualem, Y., 1976. A new model for predicting the hydraulic conductivity of unsaturated porous media. **Water Resour Res** 12, 513–522. <https://doi.org/10.1029/WR012i003p00513>
- Nolan, B.T., Healy, R.W., Taber, P.E., Perkins, K., Hitt, K.J., Wolock, D.M., 2007. Factors influencing ground-water recharge in the eastern United States. **J Hydrol (Amst)** 332, 187–205. <https://doi.org/10.1016/J.JHYDROL.2006.06.029>
- Otto, S.R.L., 1988. **Estimativa da porosidade drenável em função das propriedades de solo orgânico**. Universidade Federal de Viçosa, Viçosa.
- Otoni, M.V., Otoni 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>
- Pessoa, T.N., Libardi, P.L., 2022. Physical-hydric properties of *Latosolos* as influenced by soil structure and clay mineralogy. **Catena (Amst)** 211, 106009. <https://doi.org/10.1016/J.CATENA.2021.106009>
- Reynolds, W.D., Drury, C.F., Tan, C.S., Fox, C.A., Yang, X.M., 2009. Use of indicators and pore volume-function characteristics to quantify soil physical quality. **Geoderma** 152, 252–263. <https://doi.org/10.1016/J.GEODERMA.2009.06.009>
- Rinot, O., Levy, G.J., Steinberger, Y., Svoray, T., Eshel, G., 2019. Soil health assessment: A critical review of current methodologies and a proposed new approach. **Science of The Total Environment** 648, 1484–1491. <https://doi.org/10.1016/J.SCITOTENV.2018.08.259>
- RStudio Team, 2020. **RStudio: Integrated Development for R**.
- Santana, M.L.T., dos Santos, F.F., de Carvalho, K.M., Peixoto, D.S., Uezu, A., Avanzi, J.C., Serafim, M.E., Nunes, M.R., van Es, H.M., Curi, N., Silva, B.M., 2023. Interactions between land use and soil type drive soil functions, highlighting water recharge potential, in the Cantareira System, Southeast of Brazil. **Science of The Total Environment** 903, 166125. <https://doi.org/10.1016/J.SCITOTENV.2023.166125>
- Santos, H.G. dos, Carvalho Junior, W. de, Dart, R. de O., Aglio, M.L.D., Souza, J.S. de, Pares, J.G., Fontana, A., Martins, A.L. da S., Oliveira, A.P. de., 2011. **O novo mapa de solos do Brasil: legenda atualizada**. Embrapa Solos, Rio de Janeiro.
- Santos, H.G. dos, Jacome, P.K.T., Anjos, L.H.C. dos, Oliveira, V.A. de, Lumbreras, J.F., Coelho, M.R., Almeida, J.A. de, Araujo Filho, J.C. de, Oliveira, J.B. de, Cunha, T.J.F., 2018. **Sistema Brasileiro de Classificação de Solos**, 5. ed. rev. e ampl. ed. Embrapa, Brasília, DF.
- Sims, J.T., Cunningham, S.D., Sumner, M.E., 1997. Assessing Soil Quality for Environmental Purposes: Roles and Challenges for Soil Scientists. **J Environ Qual** 26, 20–25. <https://doi.org/10.2134/jeq1997.00472425002600010004x>
- Souza, C.M., Z. Shimbo, J., Rosa, M.R., Parente, L.L., A. Alencar, A., Rudorff, B.F.T., Hasenack, H., Matsumoto, M., G. Ferreira, L., Souza-Filho, P.W.M., de Oliveira, S.W., Rocha, W.F., Fonseca, A. V., Marques, C.B., Diniz, C.G., Costa, D., Monteiro, D., Rosa, E.R., Vélez-Martin,

- E., Weber, E.J., Lenti, F.E.B., Paternost, F.F., Pareyn, F.G.C., Siqueira, J. V., Viera, J.L., Neto, L.C.F., Saraiva, M.M., Sales, M.H., Salgado, M.P.G., Vasconcelos, R., Galano, S., Mesquita, V. V., Azevedo, T., 2020. Reconstructing Three Decades of Land Use and Land Cover Changes in Brazilian Biomes with Landsat Archive and Earth Engine. **Remote Sens (Basel)** 12, 2735. <https://doi.org/10.3390/rs12172735>
- Tenenwurcel, M.A., 2021. **Estimativa da recarga potencial de aquíferos na bacia hidrográfica do Rio Bicudo-MG** (master's thesis). Universidade Federal de Minas Gerais, Belo Horizonte.
- Thanh, N.N., Thunyawatcharakul, P., Ngu, N.H., Chotpantararat, S., 2022. Global review of groundwater potential models in the last decade: Parameters, model techniques, and validation. **J Hydrol (Amst)** 614, 128501. <https://doi.org/10.1016/j.jhydrol.2022.128501>
- UNESCO, 2022. **The United Nations World Water Development Report 2022: GROUNDWATER Making the invisible visible**. Paris.
- van Es, H.M., Cassel, D.K., Daniels, R.B., 1991. Infiltration Variability and Correlations with Surface Soil Properties for an Eroded Hapludult. **Soil Science Society of America Journal** 55, 486. <https://doi.org/10.2136/sssaj1991.03615995005500020032x>
- van Genuchten, M.Th., 1980. A Closed-form Equation for Predicting the Hydraulic Conductivity of Unsaturated Soils. **Soil Science Society of America Journal** 44, 892–898. <https://doi.org/10.2136/sssaj1980.03615995004400050002x>
- Wander, M.M., Bollero, G.A., 1999. Soil Quality Assessment of Tillage Impacts in Illinois. **Soil Science Society of America Journal** 63, 961–971. <https://doi.org/10.2136/sssaj1999.634961x>
- Wymore, A.W., 1993. **Model-Based Systems Engineering**. **CRC Press**. <https://doi.org/10.1201/9780203746936>

## **PAPER II: Determinants of Groundwater Recharge in Brazil: A Comprehensive Analysis**

Lara Mota Corinto <sup>(1)</sup>; Leandro Campos Pinto <sup>(1)</sup>; Dayane Saturnino de Santana <sup>(1)</sup>; Marta Vasconcelos Ottoni <sup>(2)</sup>; Nilton Curi <sup>(1)</sup>; Junior Cesar Avanzi <sup>(1)</sup>; Bruno Montoani Silva <sup>(1)</sup>

Standards of the Journal of Hydrology (J. Hydrol).

---

<sup>(1)</sup> Federal University of Lavras, Department of Soil Science, CEP 37200-000, Lavras, MG, Brazil. Email: laramcorinto@hotmail.com, leandcampos@yahoo.com.br, dayysant71@gmail.com, ntcuri@gmail.com, junior.avanzi@ufla.br; brunom.silva@ufla.br

<sup>(2)</sup> Geological Survey of Brazil, Department of Hydrology, Brazil, Avenida Pasteur, 404 - Urca CEP: 22290-255, Rio de Janeiro, RJ, Brazil. Email: marta.ottoni@sgb.gov.br

## ABSTRACT

Recharge zones play a fundamental role in the hydrological cycle, as they are regions where atmospheric water infiltrates the soil, initiating underground flow and replenishing aquifers. The infiltration process is influenced by factors such as soil class, vegetation cover and land use. In this sense, this study aims to understand how soil properties, land use and biome influence soil quality for groundwater recharge potential, considering the specificities and diversities of Brazilian soils. Three Soil Quality Index (SQIs) were calculated based on the literature, using data from the Hydrophysical Database for Brazilian Soils (HYBRAS). The indicators selected included: (i) Laboratory Soil Hydraulic Conductivity (Ksat\_lab), Drainable Porosity (DP) and Silt Content for (Corinto *et al.*, 2025a, preprint), here referred to as SQIco; (ii) Ksat\_lab, Basic Infiltration Rate (BIR) and DP (Santana *et al.*, 2023), SQIs; and (iii) Field Hydraulic Conductivity (Ksat\_field), Macroporosity and Bulk density (BD) (Alvarenga *et al.*, 2012), SQIal. The results showed that the Amazon and the Atlantic Forest had the highest values for the three SQIs calculated, reflecting their high vegetation cover and infiltration capacity. The Pampa showed the lowest values due to its flat topography, sub-humid climate and the presence of shallow water tables. The Cerrado showed variable results attributed to the diversity of vegetation physiognomies, while the Caatinga showed dispersion in SQIco values, suggesting edaphic and climatic heterogeneity. Areas of native vegetation had greater recharge potential, favoring infiltration and reducing surface runoff. Among the soil classes, the *Neossolos Quartzarênicos* had the highest SQI due to their high porosity, while the *Latossolos* had high recharge potential even with a clay texture. Soil texture was the most relevant factor in predicting the SQI, but in isolation it may not capture the pattern of soils with specific structural characteristics. Thus, the combination of soil texture and soil classification is essential to compose zoning maps of the water recharge potential of Brazilian soils.

**Keywords:** Soil Quality Index; HYBRAS; Amazon; soil infiltration

## RESUMO

As zonas de recarga desempenham um papel fundamental no ciclo hidrológico, pois são regiões onde a água atmosférica se infiltra no solo, iniciando o fluxo subterrâneo e reabastecendo os aquíferos. O processo de infiltração é influenciado por fatores como a classe de solo, a cobertura vegetal e o uso do solo. Nesse sentido, este estudo tem como objetivo compreender como os atributos do solo, o uso da terra e o bioma influenciam a qualidade do solo para o potencial de recarga das águas subterrâneas, considerando as especificidades e diversidades dos solos brasileiros. Foram calculados três Índices de Qualidade do Solo (SQI) com base na literatura, utilizando dados do *Hydrophysical Database for Brazilian Soils* (HYBRAS). Os indicadores selecionados incluíram: (i) Condutividade Hidráulica do Solo em Laboratório (Ksat\_lab), Porosidade Drenável (DP) e Teor de Silte para (Corinto *et al.*, 2025a, preprint), aqui denominado como SQIco; (ii) Ksat\_lab, Velocidade de Infiltração Básica (BIR) e PD (Santana *et al.*, 2023), SQIs; e (iii) Condutividade Hidráulica do Solo em Campo (Ksat\_field), Macroporosidade e Densidade do solo (BD) (Alvarenga *et al.*, 2012), IQSal. Os resultados mostraram que a Amazônia e a Mata Atlântica apresentaram os maiores valores para os três IQS calculados, refletindo sua alta cobertura vegetal e capacidade de infiltração. O Pampa apresentou os menores valores devido à sua topografia plana, clima subúmido e presença de lençóis freáticos rasos. O Cerrado apresentou resultados variáveis atribuídos à diversidade de fisionomias vegetais, enquanto a Caatinga apresentou dispersão nos valores do IQSc, sugerindo heterogeneidade edáfica e climática. As áreas de vegetação nativa apresentaram maior potencial de recarga, favorecendo a infiltração e reduzindo o escoamento superficial. Dentre as classes de solo, os Neossolos Quartzarênicos apresentaram o maior IQS devido à sua alta porosidade, enquanto os Latossolos apresentaram alto potencial de recarga mesmo com textura argilosa. A textura do solo foi o fator mais relevante na predição do IQS, mas isoladamente pode não captar o comportamento de solos com características estruturais específicas. Assim, a combinação da textura e da classificação do solo é essencial para compor mapas de zoneamento do potencial de recarga hídrica dos solos brasileiros.

**Palavras-chave:** Índice de Qualidade do Solo; HYBRAS; Amazônia; infiltração do solo

## 1. INTRODUCTION

Groundwater recharge is crucial for maintaining water resources, particularly in regions where water demand is increasing, and the effects of climate change are becoming more apparent. Groundwater recharge zones are vital, as they are areas where rain infiltrates into the soil, replenishing aquifers. Identifying and protecting these areas is essential to ensure water resources sustainability, especially in the face of challenges like overexploitation and groundwater contamination.

Several studies have employed remote sensing techniques to connect soil-related factors with climatic and topographic elements for mapping areas with groundwater recharge potential. Despite these advancements, most of these studies primarily rely on a single soil variable, typically the textural class, in conjunction with other environmental factors (Achu *et al.*, 2020; Jaafarzadeh *et al.*, 2021; Kaewdum and Chotpantararat, 2021; Selvam *et al.*, 2016). This limited approach can lead to a partial understanding of the soils' potential to contribute to the groundwater recharge process, especially in the Brazilian context, which, due to its vast territorial extension, exhibits a diversity of soils with varying characteristics and physical properties.

Various factors, including soil type, vegetation cover, and land use practices influence the process of water entering soils to recharge aquifers. Soils are active water reservoirs, with their physical, chemical, and biological properties determining their infiltration and storage capacity. Texture, structure, porosity, bulk density, and hydraulic conductivity affect the balance between infiltration, surface runoff, evapotranspiration, and underground storage (Resende *et al.*, 1998). Beyond evapotranspiration, vegetation plays a crucial role, directly impacting the physical properties of soil. For example, native vegetation enhances infiltration through improved soil aggregation and preferential flows within the soil profile, promoting percolation to groundwater recharge even during dry seasons (Roa-García *et al.*, 2011). Additionally, changes in land use, such as deforestation, lead to soil compaction, alterations in soil structure, and preferential flow, resulting in increased surface runoff due to decreased infiltration (Germer *et al.*, 2010).

Understanding how various soil characteristics influence groundwater recharge, along with the effects of different land uses, is essential for developing water management strategies that sustain groundwater resources. Corinto *et al.* (2025a, preprint), Santana *et al.* (2023), and Alvarenga *et al.* (2012) proposed soil quality indices that help identify areas more susceptible to groundwater recharge. These authors adopted various soil properties such as soil hydraulic conductivity in the laboratory (Ksat\_lab), drainable

porosity (DP), and silt in Corinto *et al.* (2025a, preprint); soil hydraulic conductivity in the laboratory (Ksat\_lab), basic infiltration velocity (VIB), and drainable porosity in Santana *et al.* (2023); and soil hydraulic conductivity in the field (Ksat\_field), macroporosity, and bulk density (BD) in Alvarenga *et al.* (2012). The resulting indices demonstrated effectiveness in the comparative analysis of soil quality for groundwater recharge across different uses and soil classes, making them a valuable tool for evaluating recharge zones.

Although studies address isolated aspects of soil quality and water recharge in watersheds, there remains a gap in integrating these factors within a broad spatial perspective that considers the specificities and diversities of the tropical soils in Brazil. One challenge in advancing research on groundwater recharge considering the Brazilian soils has been the lack of data on their physical and hydraulic properties. However, in 2018, the Hydrophysical Database for Brazilian Soils (HYBRAS) was published, providing detailed information on attributes such as texture, porosity, bulk density, and soil saturated hydraulic conductivity (Ottoni *et al.*, 2018). This database marks a significant advancement in assessing the physical quality of soils and in developing management practices aimed at conserving groundwater resources (Bünemann *et al.*, 2018).

In this context, this study aims to understand how soil, land use, and biome affect soil quality for groundwater recharge and identify the factors that best represent soil quality for mapping recharge zones, considering the specificities and diversities of Brazilian soils and biomes. To achieve this, three soil quality indices (SQIs), designed to identify potential areas for groundwater recharge in Brazil, were calculated using data from the HYBRAS database. More specifically, we sought to (i) evaluate the impact of soil texture identification of recharge zones; (ii) analyze the influence of land uses and soil classes on soil quality for groundwater recharge; and (iii) emphasize the most representative soil quality factors for mapping groundwater recharge potential areas to support further investigations of groundwater processes in Brazil.

This study is based on the hypotheses that (i) the gibbsite *Latossolos* are fundamental for the groundwater recharge zoning in Brazil, even if they possess a fine texture (very clayey/clayey); (ii) areas of native vegetation (e.g., forests) have a higher potential for groundwater recharge compared to areas used for agriculture and/or intensive grazing; (iii) the *Neossolos quartzarênicos* exhibit greater groundwater recharge

potential due to their textural and structural properties; (iv) soil structure and land use significantly influence the groundwater recharge process in Brazil.

## 2. METHODS

### 2.1. Data collection

The data was extracted from the Hydrophysical Database for Brazilian Soils (HYBRAS) version 1 (Otoni *et al.*, 2018) and version 2, which is under development. It focuses on the attributes used by Corinto *et al.* (2025a, preprint), Santana *et al.* (2023), and Alvarenga *et al.* (2012), including Basic Infiltration Rate (BIR\_log), in log cm/h; Saturated Hydraulic Conductivity (Ksat\_log), in log cm/h for laboratory-determined methods; Saturated Hydraulic Conductivity (Kfs\_log), in log m/d for field-determined methods; Macroporosity (mac), in m<sup>3</sup>/m<sup>3</sup>, calculated based on the difference between total porosity and the volumetric soil moisture at 6 kPa; Drainable Porosity (DP), in cm<sup>3</sup>/cm<sup>3</sup>, computed according considering the difference between volumetric soil moisture at saturation and at 10 kPa (field capacity) (Otto, 1988); Bulk Density (BD), in kg/dm<sup>3</sup>; and Silt content (%). If soil volumetric moisture data at 6 kPa and 10 kPa for a soil sample in HYBRAS are absent, the corresponding result from fitting the sample's water retention data to the van Genuchten (1980) model with the Mualem restriction ( $m=1-1/n$ ) (Mualem, 1976) will be used, when available. This fit was performed only for those soil samples that included at least five water retention data points covering a wide range of water tensions from 0 to -15,000 cmca.

### 2.2. Calculation of Soil Quality Index

The datasets were standardized on a scale from 0 to 1 based on non-linear scoring values (Wymore, 1993), Eq. 1:

$$v = \frac{1}{1 + \left(\frac{B-L}{x-L}\right)^{2S(B+x-2L)}} \quad \text{Eq. 1}$$

Where  $v$  represents the normalized score;  $B$  indicates the value of the indicator when the normalized score ( $v$ ) is 0.5, which corresponds to the average of the respective indicator;  $L$  denotes the lowest observed value of the indicator;  $S$  is the slope of the tangent line to the curve at  $B$ , meaning the second derivative of equation 1 equals 0; and

x signifies the value of the soil indicator associated with the normalized value ( $v = 0.5$ ), for  $x = B$ .

To determine the parameters of the curve, all available samples in HYBRAS were considered for each variable without applying any filters. In other words, the parameters were based on the environmental variability of the data, which includes a diversity of texture classes, soil classes, land uses, biomes, depths, among others. The base value (B) was defined as the mean of the analyzed variable; the value of L as the lowest observed value; and U as the highest observed value.

The calculation of the S value for each curve was performed using the Solver tool in Microsoft Excel, taking into account the final requirements of  $\vartheta = 0.5$  and  $x \approx B$ . To do this, the Mean (B) and Standard Deviation (SD) of each indicator were initially defined in order to determine the initial x value (Eq. 2). Then, the S value was calculated based on the initial x using Eq. 3. In cases where the indicator was considered “more is better,” the S value was positive; on the other hand, when the indicator was considered “less is better,” the S value was negative. The sign (positive or negative) was applied based on the correlation of the variable with the Basic Infiltration Rate (BIR). Thus, the Solver tool was used with the objective of setting  $\vartheta = 0.5$  by adjusting the variables S and  $x_{initial}$  so that the latter would approximate the B value.

$$x_{initial} = 0,5 * SD + B \quad \text{Eq. 2}$$

$$S = \frac{\log\left(\frac{1}{\vartheta}\right) - 1}{\log\log\left(\frac{B-L}{x-L}\right) * 2(B+x-2L)} \quad \text{Eq. 3}$$

After scoring the indicators, The SQI was calculated by multiplying the score of each indicator (v) by its respective weight (p), as indicated by each author: Eq. 4 Corinto *et al.* (2025a, preprint); Eq. 5 for Santana *et al.* (2023); e, Eq. 6 for Alvarenga *et al.* (2012).

$$SQIc = 0.70 Ksat_{lab_n} + 0.19 DP_n + 0.11 Silt_n \quad \text{Eq. 4}$$

$$SQI_{sa} = 0.33 BIR_n + 0.33 Ksat_{lab_n} + 0.33 DP_n \quad \text{Eq. 5}$$

$$SQI_{al} = 0.25 BD_n + 0.40 K_{sat_{field}_n} + 0.35 mac_n \quad \text{Eq. 6}$$

Where  $n$  represents the value of the indicator normalized between 0 and 1.

### 2.3. SQIs comparison

All analyses were conducted using R software version 4.4. The indices were evaluated based on the biomes, taking into account the average SQI for the soil profile, land use based on the average SQI for the surface layer (0-20 cm); soil class at the second categorical level of the Brazilian Soil Classification System, taking into account the average SQI for the soil profile; the functional structure of the soil regarding the diagnostic horizon; and texture for both the surface horizons (0-20 cm) and the texture of the B horizon. It is important to note that due to the limitations of the samples with BIR characterization, which was restricted to the surface layer, thus, the SQI<sub>sa</sub> is limited to this layer.

To analyze each factor in the SQIs, the data was filtered to exclude groups with fewer than five observations and samples where the analyzed factor was not specified. Outliers were detected and excluded using the interquartile range (IQR) method. A new check was then performed to ensure at least five observations in each group related to the analyzed factor. Because most residuals of the factor do not present normality, we used non-parametric methods. Thus, the Kruskal-Wallis test was applied to compare the SQI medians among the different groups within each analyzed parameter. This was followed by the Dunn test with Bonferroni correction for multiple comparisons, utilizing the FSA package.

For cases where the SQI of a specific group was calculated using Corinto *et al.* (2025a, preprint), Santana *et al.* (2023), or Alvarenga *et al.* (2012), the SQIs were evaluated within the group, allowing for the comparison of the three indices considered in this study.

### 2.4. Evaluation of the importance of factors

After comparing the groups in each factor, the SQI with the best distribution of data was selected to check the importance of the factors and select those that most influence the SQI. Then, the VSURF library was used based on the methodology proposed by Genuer *et al.* (2010). This methodology consists of two stages. The first, where the explanatory variables are classified according to their importance (VI), the

corresponding values of the standard deviation of VI are calculated to define the threshold value. Once this threshold value has been set, the factors selected will be those with a value greater than the threshold value. In the second stage, this subset is used, where the factors are tested gradually, being selected when  $errOOB$  decreases greater than the average variation obtained by adding the variable (Genuer *et al.*, 2015).

$$X_{sel} = \frac{1}{m - m'} \sum_{j=m'}^{m-1} |errOOB(j + 1) - errOOB(j)| \quad \text{Eq. 7}$$

Where  $m$  represents the initial model,  $m'$  is the model with the added variable, and  $errOOB$  denotes the average error from each decision tree constructed based on the  $j$  most important factors.

Considering the selected factors, machine learning techniques (e.g., Random Forest and XGBoost) were employed to develop a model that best estimates the SQI for identifying recharge zones in Brazil. The lack of availability of soil factors for the SQI across the entire Brazilian territory restricts the use of the SQI application on larger scales.

The models were constructed by randomly dividing the data into training (80% of the data) and test (20% of the data) sets. Two different scenarios were examined; the first included all factors such as biome class, land use, soil class (Soil\_2nc), structure pertaining to the diagnostic horizon, and textural class of the average granulometric composition of the profile (Texture\_prof); the second considered only the factors selected by the VSURF method. Overall, four models were developed: Random Forest with all variables (RFt), Random Forest with selected variables (RFs), XGBoost with all variables (XGBt), and XGBoost with selected variables (XGBs).

The parameters of the Random Forest-based models were standardized with the number of trees ( $n_{tree}$ ) set to 1500 and the number of variables used in each tree ( $m_{try}$ ) set to 4. To implement modeling with XGBoost, the factors were transformed using One-Hot Encoding via the `model.matrix` function. The hyperparameters of the model were defined as follows: a learning rate ( $\eta$ ) of 0.05, a maximum tree depth ( $max\_depth$ ) of 10, data sampling (`subsample`) at 0.8, variable sampling per tree (`colsample_bytree`) at 0.8, L2 regularization ( $\lambda$ ) of 2, L1 ( $\alpha$ ) of 1, and a minimum weight of 10 per leaf (`min_child_weight`). The training was conducted with 1500 iterations ( $n_{rounds}$ ), applying the early stopping technique after 500 rounds without improvement in performance.

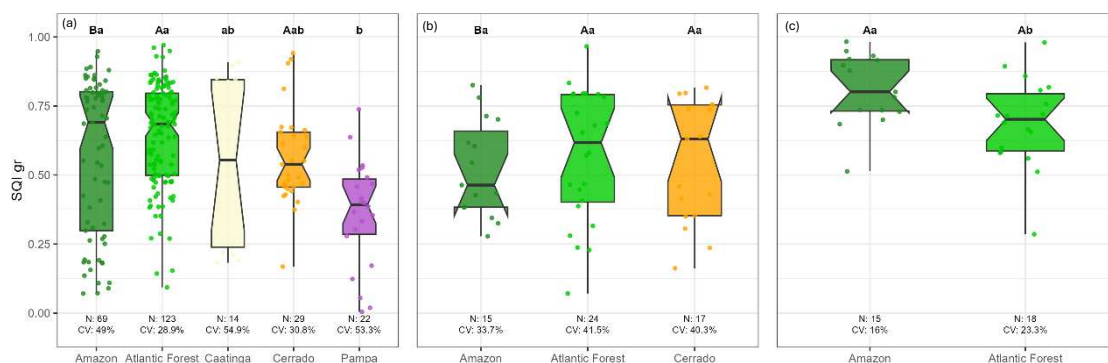
The performance of the models was evaluated using the Root Mean Square Error (RMSE), the Mean Absolute Error (MAE) and the coefficient of determination ( $R^2$ ), calculated from the predictions in the test set.

### 3. RESULTS

#### 3.1. SQIs comparison

The Amazon biome has the highest SQI for identifying potential groundwater recharge zones, distinguishing it among the analyzed biomes in Brazil. The Atlantic Forest also emerged prominently if SQI<sub>coo</sub> (Fig. 3.1a) and SQI<sub>al</sub> (Fig. 3.1b) were considered, showing no significant difference. This aspect also was observed for the Cerrado biome, i.e., these did not show statistical differences. The use of SQI<sub>sa</sub> in the Atlantic Forest revealed lower values than in the Amazon (Fig. 3.1c). Considering the SQI<sub>co</sub> index, the Caatinga displayed no significant differences compared to the Amazon, Cerrado, and Atlantic Forest; however, there was greater variation in the values for this biome. In contrast, the lowest SQI values were observed for the Pampa, indicating that this biome presented the least potential for groundwater recharge. The SQI<sub>coo</sub> and SQI<sub>al</sub> values for Amazon are considerably lower than those reported by the SQI<sub>sa</sub>. The values for the Atlantic Forest showed no significant difference, regardless of the index used to calculate the soil quality. In Cerrado, the comparison between SQI<sub>co</sub> and SQI<sub>al</sub> shows no significant difference.

**Fig. 3.1.** Comparison between Brazilian biomes considering the average SQI for potential groundwater recharge for each profile analyzed in three different methods (a) SQI<sub>co</sub>; (b) SQI<sub>al</sub>; and (c) SQI<sub>sa</sub>.



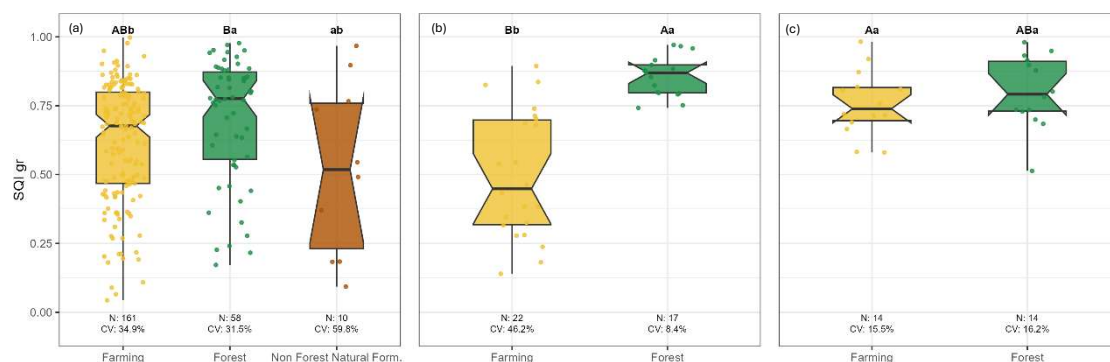
Lowercase letters differ between classes within each SQI (a,b); Uppercase letters differ between SQIs within each class (A, B).

SQI<sub>co</sub>: Soil Quality Index according to Corinto *et al.* (2025a); SQI<sub>al</sub>: Soil Quality Index according to Alavrenga *et al.* (2012); SQI<sub>sa</sub>: Soil Quality Index according to Santana *et al.* (2023) - SQI<sub>sa</sub>; N: number of soil profiles analyzed; CV: coefficient of variation of the SQI in the group.

Land use was analyzed in relation to the surface layer, which exerts the greatest impact. In all the SQIs, the forest exhibited the highest values, showing a significant difference compared to farming areas in the SQIco (Fig. 3.2a) and SQIal (Fig. 3.2b) indexes. This difference was not noted in the SQIsa index (Fig. 3.2c), possibly due to the small number of samples. Yet, considering the SQIco index, the planted forest formation did not display significant differences compared to farming and forest.

Using the SQIsa, the farming areas showed high values, while the SQIal indicated the lowest values for the same usage. SQIco showed no significant differences compared to the other models for farming areas. Regarding forests, the SQIco did not significantly differ from the SQIsa but exhibited lower values than the SQIal.

**Fig. 3.2.** Comparison between land uses considering the average SQI for groundwater recharge potential in the surface layer (0-20cm) analyzed using three different methods (a) SQIco; (b) SQIal; and, (c) SQIsa.



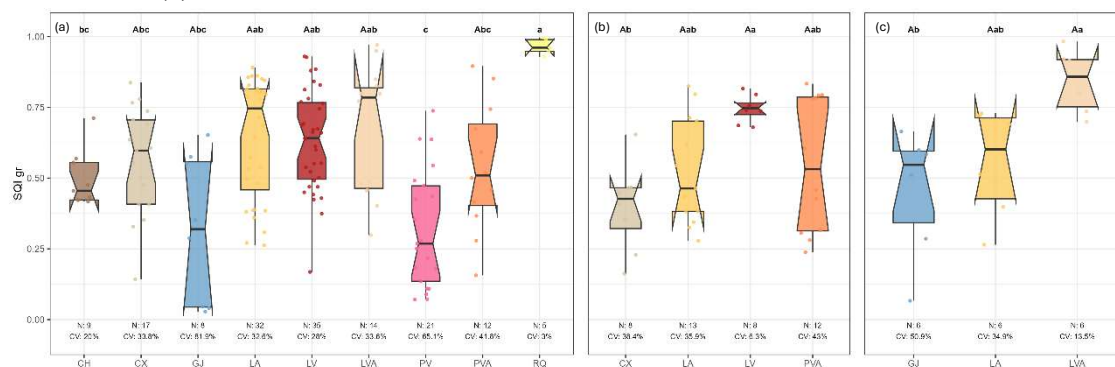
Lowercase letters differ between classes within each SQI (a,b); Uppercase letters differ between SQIs within each class (A, B).

SQIco: Soil Quality Index according to Corinto *et al.* (2025a); SQIal: Soil Quality Index according to Alavarenga *et al.* (2012); SQIsa: Soil Quality Index according to Santana *et al.* (2023) - SQIsa; N: number of soil profiles analyzed; CV: coefficient of variation of the SQI in the group.

The *Neossolos Quartzarênicos* (RQ) exhibited the highest SQI values, although they did not differ significantly from the *Latossolos* (LA, LV, and LVA), exhibiting high SQIco values (Fig. 3.3a). In the SQIco index, *Latossolos* did not differ significantly from *Cambissolos*, *Argissolo Vermelho-Amarelo* (PVA) and *Gleissolos Tiomórficos* (GJ). This similarity among *Latossolos*, *Cambissolos* and *Argissolos Vermelho-Amarelo* was also noted for the SQIal index (Fig. 3.3b), although the similarity with Inceptisols was limited to CX and LA. The similarity between GJ and *Latossolos* can be explained by the limited number of samples in the SQIco and SQIsa indexes (Fig. 3.3c), resulting in a high variation coefficient. The *Argissolos Vermelho* (PV) displayed the lowest SQI values. The

pattern remained consistent for the soil classes that appeared in more than one SQI, with no significant differences.

**Fig. 3.3.** Comparison between soil classes considering the average SQI for groundwater recharge potential in the soil profile analyzed in three different methods (a) SQIco; (b) SQIal; and, (c) SQIsa.



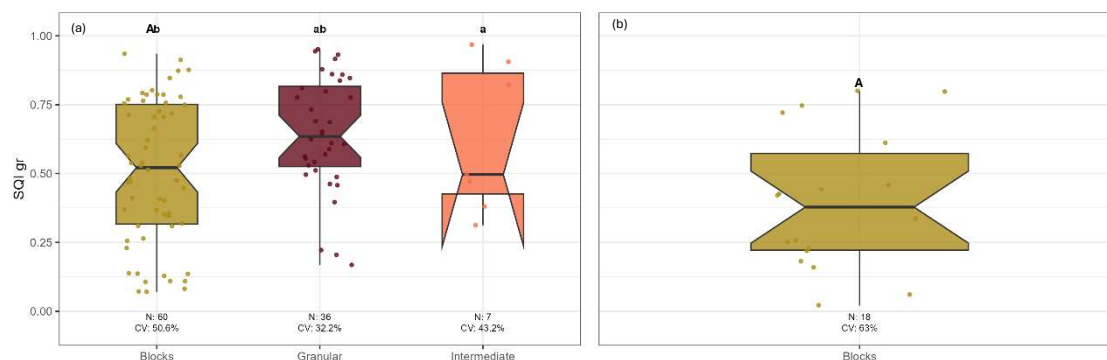
Lowercase letters differ between classes within each SQI (a,b); Uppercase letters differ between SQIs within each class (A, B).

*Cambissolo Húmico* (CH); *Cambissolo Háptico* (CX); *Gleissolo Tiomórfico* (GJ); *Latossolo Amarelo* (LA); *Latossolo Vermelho* (LV); *Latossolo Vermelho-Amarelo* (LVA); *Argissolo Vermelho* (PV); *Argissolo Vermelho-Amarelo* (PVA); *Neossolo Quartzarênico* (RQ); SQIco: Soil Quality Index according to Corinto *et al.* (2025a); SQIal: Soil Quality Index according to Alavarenga *et al.* (2012); SQIsa: Soil Quality Index according to Santana *et al.* (2023) - SQIsa; N: number of soil profiles analyzed; CV: coefficient of variation of the SQI in the group.

The assessment of soil structure was limited to the diagnostic horizon. In the SQIco index, the granular class is found primarily in gibbsite soils, exhibiting a significantly higher value than the block structure. Only SQIco (Fig. 3.4a) and SQIal (Fig. 3.4b) were analyzed due to the restriction of VIB values to the surface samples. Consequently, it was not possible to compare them with the SQIsa. Additionally, only the block structure was identified in SQIco and SQIal, which did not show significant differences.

The soil textural classes were analyzed on Horizon A (surface layer 0-20 cm) and Horizon B (diagnostic horizon). In Horizon A, the coarse fraction displayed higher values of SQIco (Fig. 3.5a) and SQIal (Fig. 3.5b). Using the SQIal index, the coarse texture did not significantly differ from the medium texture. The fine texture is responsible for the lowest values in the SQIal and SQIco indexes, with no significant difference from the medium texture. However, in the SQIsa index (Fig. 3.5c), the fine texture exhibited higher values than the medium texture.

**Fig. 3.4.** Comparison between soil classes considering the average SQI for potential groundwater recharge in the soil profile analyzed in three different methods (a) SQIco; and (b) SQIal.



Lowercase letters differ between classes within each SQI (a,b); Uppercase letters differ between SQIs within each class (A, B).

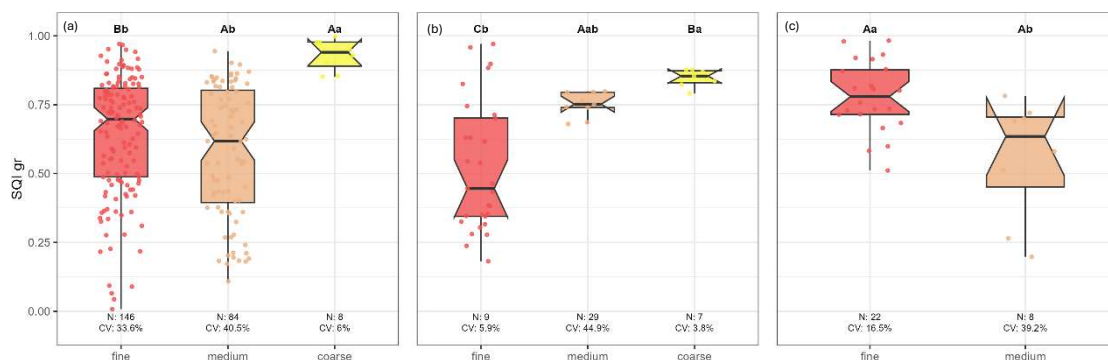
SQIco: Soil Quality Index according to Corinto *et al.* (2025a); SQIal: Soil Quality Index according to Alavarenga *et al.* (2012); N: number of soil profiles analyzed; CV: coefficient of variation of the SQI in the group.

The soil textural classes were analyzed on Horizon A (surface layer 0-20 cm) and Horizon B (diagnostic horizon). In Horizon A, the coarse fraction displayed higher values of SQIco (Fig. 3.5a) and SQIal (Fig. 3.5b). Using the SQIal index, the coarse texture did not significantly differ from the medium texture. The fine texture is responsible for the lowest values in the SQIal and SQIco indexes, with no significant difference from the medium texture. However, in the SQIal index (Fig. 3.5c), the fine texture exhibited higher values than the medium texture.

In the comparative analysis of indexes, the fine fraction displayed significant differences across all SQIs, following ascending order: SQIal < SQIco < SQIal. For the coarse fraction, SQIco was significantly higher than SQIal. No significant differences were obtained for the medium texture among the indexes.

As mentioned above, the limited number of samples restricted the analysis of Horizon B for the SQIco (Fig. 3.6a) and SQIal (Fig. 3.6b) indexes. The pattern was similar to that observed in Horizon A, except that in the SQIal, the fine texture showed a significant difference compared to the medium texture.

**Fig. 3.5.** Comparison between the medium texture classes considering the SQI for potential groundwater recharge in the surface layer (0-20cm) analyzed using three different methods (a) SQIco; (b) SQIal; and (c) SQIsa.

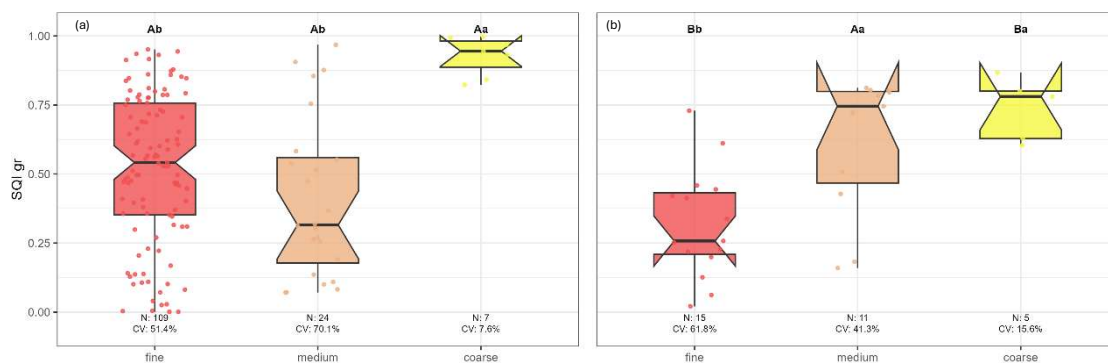


Lowercase letters differ between classes within each SQI (a,b); Uppercase letters differ between SQIs within each class (A, B).

SQIco: Soil Quality Index according to Corinto *et al.* (2025a); SQIal: Soil Quality Index according to Alavarenga *et al.* (2012); SQIsa: Soil Quality Index according to Santana *et al.* (2023) - SQIsa; N: number of soil profiles analyzed; CV: coefficient of variation of the SQI in the group.

As mentioned above, the limited number of samples restricted the analysis of Horizon B for the SQIco (Fig. 3.6a) and SQIal (Fig. 3.6b) indexes. The pattern was similar to that observed in Horizon A, except that in the SQIal, the fine texture showed a significant difference compared to the medium texture.

**Fig. 3.6.** Comparison between textural classes considering the mean SQI for potential groundwater recharge in the Horizon B analyzed using three different methods (a) SQIco; (b) SQIal; and (c) SQIsa.



Lowercase letters differ between classes within each SQI (a,b); Uppercase letters differ between SQIs within each class (A, B).

SQIco: Soil Quality Index according to Corinto *et al.* (2025a); SQIal: Soil Quality Index according to Alavarenga *et al.* (2012); SQIsa: Soil Quality Index according to Santana *et al.* (2023) - SQIsa; N: number of soil profiles analyzed; CV: coefficient of variation of the SQI in the group.

### 3.2. Evaluation of the importance of factors and prediction indexes

When analyzing the importance of the factors, the mean values for each SQIco using soil profile were considered due to the higher number of samples and their similarities with other soil quality indices for comparison of each factor. Our analysis revealed that Texture\_prof had the highest predictive weight in the model, with an average importance value of 0.016, followed by Soil\_2nc (0.012), Biome (0.008), Landuse (0.004), and Str\_prof (0.003) (Fig. 3.7a). The standard deviation of the importance of the variables also indicated that Texture\_prof was the most variable in terms of its impact on predictions (0.0009), and the others displayed less variation, suggesting greater consistency in their effects on the predictions (Fig. 3.7b).

The OOB (Out-Of-Bag) error in the nested models exhibited a notable reduction until the inclusion of the Soil\_2nc factor, with a declining error from 0.037 to 0.034. Following this inclusion, the error stabilized, indicating that adding other factors, such as Biome, Landuse, and Str\_prof would not significantly improve the model's performance (Fig. 3.7c). Furthermore, the OOB error in the predictive models demonstrated a steady decrease from 0.037 to 0.034, emphasizing the importance of the Texture\_prof and Soil\_2nc factors in modeling the Soil Quality Index (Fig. 3.7d).

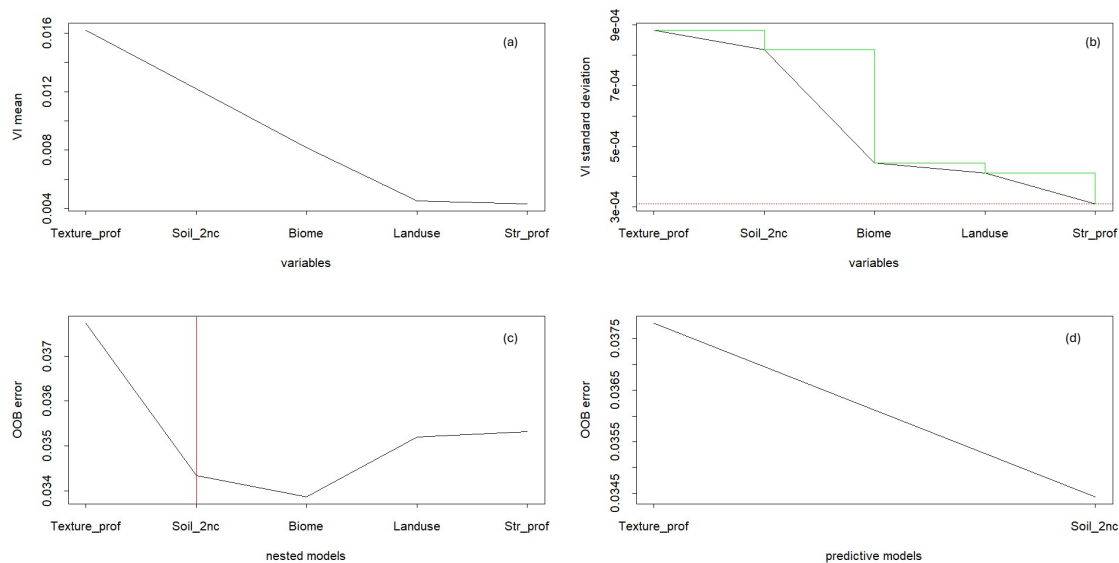
The results obtained demonstrate variations in the performance of the models to estimate the SQI concerning groundwater recharge in Brazil (Table 3.1). The XGB and RF models run with Texture\_prof and Soil\_2nc and perform better than those that utilize all factors for prediction. The XGB model exhibited the best performance, with  $R^2 = 0.59$  along with the lowest RMSE (0.16) and MAE (0.14), suggesting greater precision and lower error magnitude. Conversely, the XGBt model displayed the poorest fit ( $R^2 = 0.28$ ), despite lower RMSE and MAE values. The Random Forest-based models (RFt and RFs) showed intermediate performance, with  $R^2$  values ranging from 0.39 to 0.42.

**Table 3.1.** Performance of machine learning models in predicting the SQI for potential groundwater recharge in Brazil, with  $R^2$ , RMSE and MAE values.

Models	R2	RMSE	MAE
RFt	0,39	0,29	0,26
RFs	0,42	0,29	0,25
XGBt	0,28	0,17	0,15
XGBs	0,59	0,16	0,14

Random Forest with all variables (RFt), Random Forest with selected variables (RFs); XGBoost with all variables (XGBt) and XGBoost with selected variables (XGBs).

**Fig. 3.7.** Importance of the parameters in predicting SQI. (a) Average importance of the factors in the model; (b) Standard deviation of the importance of the factor; (c) Out-Of-Bag (OOB) error in nested models; (d) OOB error in predictive models.



## 4. DISCUSSION

### 4.1. Influence of factors on SQI

The results of this study indicated that Amazon had the highest SQI for potential groundwater recharge, and Pampa had the lowest values. These findings reflect the natural characteristics of each biome, including the greater vegetation cover and greater soil infiltration capability in the Amazon compared to the Pampa biome due to its flat topography and sub-humid climate. Groundwater in this biome is typically very shallow, leading to frequent flooding events.

These results corroborate previous studies highlighting the Amazon's significance for groundwater recharge, supporting infiltration during the rainy and dry seasons (Celentano *et al.*, 2017). They also note that the delayed and muted response to rainfall allows surface water to be absorbed during droughts, facilitating infiltration even in these dry periods.

In relation to the Atlantic Forest, Pinto *et al.* (2017) showed the importance of this biome in infiltration and groundwater recharge processes, emphasizing that the drainable porosity is high to moderate when associated with native forest fragments. This supports the results presented by the SQIco and SQIal, which indicate high values for the Atlantic Forest biome.

Regarding the Cerrado biome, Oliveira *et al.* (2020) contradicts expectations found in other biomes. In this biome, an increase in vegetation density tends to reduce the potential for groundwater recharge. The authors demonstrate that areas with denser vegetation types, such as “*cerrado sensu stricto*” and “*cerradão*”, typically generate low recharge rates, while open cerrado, like “*campo limpo*” and “*campo sujo*”, promote recharge rates greater than 350 mm per year. This mainly supports the results of our study, where the Cerrado does not significantly differ from any other biome, given that the diversity of vegetation types present in the biome can influence recharge potential.

The Caatinga biome displayed the greatest variation in SQI values, indicating heterogeneity in the region's edaphic and climatic characteristics, which affected recharge capacity. However, the limited water available in this biome, with an average annual rainfall of 800 mm (Assad, 2024), suggests a low recharge rate, as the Caatinga biome has high potential evapotranspiration and very irregular precipitation (Pineiro *et al.*, 2017, 2016).

We observed that native forest areas are more prone to groundwater recharge than areas with agriculture and livestock. This aligns with several studies that have shown that forest cover enhances infiltration and reduces surface runoff. This underscores the significance of native vegetation in maintaining soil quality and the hydrological cycle. Furthermore, the intensity of the use of planted forests and monoculture plantations will be a critical factor affecting soil water dynamics in current and future scenarios. The removal of native vegetation cover for agriculture adversely impacts soil-water dynamics, leading to a substantial increase in surface runoff and soil loss. Future extreme rainfall events will elevate surface flow in degraded areas, hindering the water recharge process. However, the SQI<sub>sa</sub> index did not reveal significant differences between forest and agricultural areas, which may be attributed to the smaller number of samples, the index's varying sensitivity to detect land use variations or the management practices employed in the samples analyzed. Previous studies, such as de Almeida *et al.* (2018), indicated that soil water infiltration is more influenced by vegetation cover - depending on land use more than the soil preparation system. Nevertheless, Sone *et al.* (2019) emphasized that agricultural management can impact the recharge process. The authors note that integrated agricultural systems enhance water infiltration into the soil and decrease soil erosion, particularly when an appropriate rotation period is implemented. Our study is constrained to the analysis of land use since the HYBRAS database lacks detailed data on soil management practices. For example, conservation practices that enhance soil

water infiltration have been reported for coffee (Silva *et al.*, 2021), citrus (Melo *et al.*, 2023), and crops under a no-tillage system (Bertollo *et al.*, 2021).

Future research needs to further investigate these claims since water infiltration in the soil depends on other intrinsic factors, such as texture, porosity, bulk density, and compaction, among others. Some of these factors were explored in this study, particularly regarding land use and soil profile interactions. When considering the physical properties of both surface and subsurface horizons, the SQI did not differ between farming and forestry (Figure 4.1, Supplementary Files). It's important to note that Santana *et al.* (2023) conducted this analysis by profile and emphasized the importance of considering soil structure, which can vary between horizons. However, we understand that integrating all the factors that affect infiltration to define potential recharge zones can be a costly and complex evaluation.

The *Neossolo Quartzarênico* showed the highest SQI due to their high drainage porosity and low water retention capacity. When considering the relationship between texture and water retention concerning high clay content, sandy soils are assumed to possess low water retention and high permeability (Costa *et al.*, 2013; Fidalski *et al.*, 2013). Nevertheless, this substantial water recharge potential is also linked to the risks of leaching and contamination of water resources. Therefore, it is vital to implement soil and water conservation techniques to protect areas with recharge potential where this soil type exists (Donagemma *et al.*, 2016).

The *Latossolos* also exhibit high recharge potential, but these soils are characterized by their clay texture. Consequently, the high recharge potential values are associated with the presence of gibbsite, which, despite a high clay content, offers improved permeability, macroporosity, and aggregate stability (Ferreira *et al.*, 1999a, 1999b; Ottoni *et al.*, 2024), like *Latossolo Vermelho*, having a granular structure. The similarity with other *Latossolos* may relate to the proportion of pore volume present in these soils compared to others (Reatto *et al.*, 2007). This finding suggests that soil structure plays a more significant role than texture in groundwater recharge.

The absence of significant differences between *Latossolos*, *Cambissolos* and *Argissolo Vermelho-Amarelo* contrasts with the findings of Souza *et al.* (2019), who reported lower potential recharge in *Cambissolos* and *Argissolo Vermelho-Amarelo* due to their reduced porosity. However, in an assessment of transmissivity in the *Serra da Mantiqueira* region, Pinto *et al.* (2016) highlight considerable potential groundwater recharge, particularly in areas with native forests, even in steeper slopes. This emphasizes

the need for more comprehensive studies of the soil's physical and hydraulic properties to determine the potential recharge in relation to soil depth. Our study demonstrates that the hydraulic characteristics of the *Cambissolos* can be influenced by land use and the depth of the profile

The unexpected similarity between *Gleissolos Tiomórficos* and *Latosolos* may be attributed to the limited number of samples or the influence of specific local conditions. The potential high groundwater recharge of the *Gleissolos* is also noted by Menezes *et al.* (2009) where the favorable recharge potential is linked to their occurrence in flatter landscapes that enhance infiltration and receive increased water flow from steeper areas.

The granular structure, a defining characteristic of gibbsite *Latosolos*, exhibited the highest SQI in the SQIco index. This finding indicates that, even in highly weathered soils, the soil's physical structure can alleviate limitations associated with fine texture and prevailing mineralogy, facilitating water infiltration. This aligns with Ferreira *et al.* (1999a, 1999b) and Ottoni *et al.* (2024), who underscored the role of granular structure in promoting porosity and enhancing water movement in tropical soils. Thus, the data partially supports the hypothesis that the structure of gibbsite *Latosolos* alters the expected pattern based only on texture, underscoring the need to consider structural characteristics in recharge indexes.

Soil texture emerged as the most critical factor in predicting SQI, as indicated by the analysis of variable importance presented in section 3.2. Soils with a higher proportion of coarse texture had the highest SQI values, a finding consistent with literature that associates sandy soils with greater permeability and, consequently, greater recharge potential. However, in the SQI<sub>sa</sub> index, fine texture displayed higher values than medium texture, which contradicts the expectation that clay soils show lower infiltration. This discrepancy may reflect limitations in the SQI<sub>sa</sub> index or the influence of specific sample characteristics, such as variations in soil structure.

The variation observed among the indexes (SQIco, SQI<sub>al</sub>, and SQI<sub>sa</sub>) indicates that the selection of indicators significantly influences the results, particularly concerning the interpretation of soil texture. Each index adopts different methods for weighing physical characteristics, leading to discrepancies in the estimation of the potential recharge. Additionally, factors that are restricted to samples from a micro-watershed (SQI<sub>al</sub> and SQI<sub>sa</sub>) and the number of samples available in HYBRAS that correspond to the indicators utilized in each index also contribute to these differences.

#### 4.2. Evaluation of the Importance of Factors

The Texture\_prof variable emerged as the most influential factor in modeling SQL, demonstrating the highest predictive power and significant variability among soil profiles. These findings reinforce the central role of texture in groundwater recharge, as emphasized by Jaafarzadeh *et al.* (2021), who underscored the impact of textural heterogeneity on soil water infiltration. Therefore, the textural diversity of Brazilian soils is a critical consideration in recharge mapping studies.

Out-Of-Bag (OOB) error analysis revealed that including variables such as Soil\_2nc significantly enhanced the model's performance, while external factors like Biome and Landuse had a more limited effect. This result indicates that intrinsic soil characteristics, such as mineralogical composition and structure, have a more direct influence on potential recharge than external factors, such as, topographic and hydrological, being consistent with Jaafarzadeh *et al.* (2021). It reinforces the notion that the interactions between soil and water are closely intertwined and cannot be studied separately (Mello and Curi, 2012). Nonetheless, the interaction between land use and structure may be more relevant at local scales, highlighting the necessity for more detailed studies in various regional contexts. Several studies in the literature exclusively utilize soil texture as the basis for developing recharge zone maps (Achu *et al.*, 2020; Jaafarzadeh *et al.*, 2021; Kaewdum and Chotpantararat, 2021; Selvam *et al.*, 2016), whereas others prioritize soil classification as the primary indicator (Costa *et al.*, 2022; Lentswe and Molwalefhe, 2020). However, our findings indicate the need to integrate both information for a more accurate assessment. Texture alone provides valuable information on water infiltration capacity but may not adequately capture the patterns of soils with specific structural characteristics, such as gibbsite *Latosolos*, whose structure affects the pattern expected based on its texture (Ferreira *et al.*, 1999a). Conversely, soil classification encompasses information on pedogenetic processes, mineralogy, and structure but might not reflect the local texture variability, particularly in heterogeneous soils (Mello and Curi, 2012). The combination of these two approaches enables the capturing of both physical heterogeneity and the pedological processes that influence the potential recharge, offering a more comprehensive perspective tailored to the complexity of Brazilian soils.

The limitations observed in the model fitting results are associated with the significant edaphoclimatic and land use diversity present throughout the Brazilian territory, which poses challenges in predictive modeling. Specific regional factors, such as variations in soil texture, soil classes, and types of vegetation cover may not have been

fully captured by the global models, affecting the accuracy of predictions across different biomes and regions. This underscores the aim of this work, highlighting the primary soil-relative factors that influence potential groundwater recharge for future mapping efforts.

## 5. CONCLUSION

This study demonstrated that the Soil Quality Index (SQI) is effective for assessing groundwater recharge potential. The variation among the SQI indexes (SQIco, SQIal, and SQIsa) highlights the importance of selecting appropriate indicators, which significantly influence the results. This reinforces the need to consider different spatial scales. The Amazon and Atlantic Forests had the highest SQI values, reflecting their dense vegetation cover and greater infiltration capacity. In contrast, Pampa had the lowest values due to its flat topography, sub-humid climate, and shallow groundwater. The Cerrado exhibited variable results, attributed to the diversity of its vegetation physiognomies, while the Caatinga displayed considerable variation in SQI values, suggesting edaphic and climatic heterogeneity. Areas of native vegetation had greater recharge potential, promoting infiltration and reducing surface runoff. However, the SQIsa index showed no significant differences between forests and farming areas, possibly due to sampling limitations or specific management practices. Among the soil classes, *Neossolos Quartzarênicos* exhibited the highest SQI due to their high porosity, while *Latossolos* demonstrated high recharge potential despite their clay texture, owing to the presence of gibbsite, which enhances permeability and aggregate stability. Soil texture was the most relevant factor in predicting SQI, but in isolation, it may not adequately capture the pattern of soils with specific structural characteristics. Therefore, combining texture with soil classification is essential for a more accurate assessment when mapping areas with groundwater recharge potential.

## ACKNOWLEDGEMENTS

We gratefully acknowledge the financial support of the *Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES)*, *Conselho Nacional de Pessoal de Nível Superior (CAPES)*, *Conselho Nacional de Desenvolvimento Científico e Desenvolvimento Científico e Tecnológico (CNPq)*, *Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG)*, *Department of Soil Science of the Federal University of Lavras (DCS - UFLA)*, *Geological Survey of Brazil (SGB)* and *Federal University of Lavras (UFLA)*. This study was financed in part by the *Coordenação de Aperfeiçoamento de*

*Pessoal de Nível Superior – Brasil (CAPES) – Process n88887.845541/2023-00. This study was carried out with support from the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and the Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG).*

## REFERENCES

- Achu, A.L., Reghunath, R., Thomas, J., 2020. Mapping of Groundwater Recharge Potential Zones and Identification of Suitable Site-Specific Recharge Mechanisms in a Tropical River Basin. **Earth Systems and Environment** 4, 131–145. <https://doi.org/10.1007/s41748-019-00138-5>
- Alvarenga, C.C., Mello, C.R. de, Mello, J.M. de, Silva, A.M. da, Curi, N., 2012. Índice de qualidade do solo associado à recarga de água subterrânea (IQS RA) na Bacia Hidrográfica do Alto Rio Grande, MG. **Rev Bras Cienc Solo** 36, 1608–1619. <https://doi.org/10.1590/S0100-06832012000500025>
- Alves, M.A.B., de Souza, A.P., de Almeida, F.T., Hoshide, A.K., Araújo, H.B., da Silva, A.F., de Carvalho, D.F., 2023. Effects of Land Use and Cropping on Soil Erosion in Agricultural Frontier Areas in the Cerrado-Amazon Ecotone, Brazil, Using a Rainfall Simulator Experiment. **Sustainability** 15, 4954. <https://doi.org/10.3390/su15064954>
- Assad, M.L.L., 2024. **Caatinga: o bioma exclusivamente Brasileiro**. Fundação Getúlio Vargas.
- Bonell, M., Purandara, B.K., Venkatesh, B., Krishnaswamy, J., Acharya, H.A.K., Singh, U.V., Jayakumar, R., Chappell, N., 2010. The impact of forest use and reforestation on soil hydraulic conductivity in the Western Ghats of India: Implications for surface and sub-surface hydrology. **J Hydrol (Amst)** 391, 47–62. <https://doi.org/10.1016/j.jhydrol.2010.07.004>
- Breiman, L., 2001. Random Forests. **Mach Learn** 45, 5–32. <https://doi.org/10.1023/A:1010933404324>
- Bünemann, E.K., Bongiorno, G., Bai, Z., Creamer, R.E., De Deyn, G., de Goede, R., Flesskens, L., Geissen, V., Kuyper, T.W., Mäder, P., Pulleman, M., Sukkel, W., van Groenigen, J.W., Brussaard, L., 2018. Soil quality – A critical review. **Soil Biol Biochem** 120, 105–125. <https://doi.org/10.1016/j.soilbio.2018.01.030>
- Celentano, D., Rousseau, G.X., Engel, V.L., Zelarayán, M., Oliveira, E.C., Araujo, A.C.M., de Moura, E.G., 2017. Degradation of Riparian Forest Affects Soil Properties and Ecosystem Services Provision in Eastern Amazon of Brazil. **Land Degrad Dev** 28, 482–493. <https://doi.org/10.1002/ldr.2547>
- Chen, T., Guestrin, C., 2016. XGBoost, in: **Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining**. ACM, New York, NY, USA, pp. 785–794. <https://doi.org/10.1145/2939672.2939785>

- Corinto, L. M.; Barbosa, S. M.; Carvalho Júnior, Z. D.; Faria, V. L.; Ottoni, M. V.; Curi, Nilton; Avanzi, J. C.; Silva, B. M. 2025a. **Development and Validation of a Statistical Index for Estimating Groundwater Recharge Potential** (preprint).
- Costa, A.M. da, Salis, H.H.C. de, Araújo, B.J.R.S., Moura, M.S. de, Silva, V.C. da, Oliveira, A.R. de, Pereira, M.P.R., Viana, J.H.M., 2022. Potencial de uso conservacionista em bacias hidrográficas. **Revista Geografias** 15, 127–147. <https://doi.org/10.35699/2237-549X.2019.19891>
- Costa, André da, Albuquerque, J.A., Costa, Adriano da, Pértile, P., Silva, F.R. da, 2013. Water retention and availability in soils of the State of Santa Catarina-Brazil: effect of textural classes, soil classes and lithology. **Rev Bras Cienc Solo** 37, 1535–1548. <https://doi.org/10.1590/S0100-06832013000600010>
- de Almeida, W.S., Panachuki, E., de Oliveira, P.T.S., da Silva Menezes, R., Sobrinho, T.A., de Carvalho, D.F., 2018. Effect of soil tillage and vegetal cover on soil water infiltration. **Soil Tillage Res** 175, 130–138. <https://doi.org/10.1016/j.still.2017.07.009>
- Donagemma, G.K., Freitas, P.L. de, Balieiro, F. de C., Fontana, A., Spera, S.T., Lumbreras, J.F., Viana, J.H.M., Araújo Filho, J.C. de, Santos, F.C. dos, Albuquerque, M.R. de, Macedo, M.C.M., Teixeira, P.C., Amaral, A.J., Bortolon, E., Bortolon, L., 2016. Characterization, agricultural potential, and perspectives for the management of light soils in Brazil. **Pesqui Agropecu Bras** 51, 1003–1020. <https://doi.org/10.1590/s0100-204x2016000900001>
- Ferreira, M.M., Fernandes, B., Curi, N., 1999a. Mineralogia da fração argila e estrutura de latossolos da região sudeste do Brasil. **Rev Bras Cienc Solo** 23, 507–514. <https://doi.org/10.1590/S0100-06831999000300003>
- Ferreira, M.M., Fernandes, B., Curi, N., 1999b. Influência da mineralogia da fração argila nas propriedades físicas de latossolos da região sudeste do Brasil. **Rev Bras Cienc Solo** 23, 515–524. <https://doi.org/10.1590/S0100-06831999000300004>
- Fidalski, J., Tormena, C.A., Alves, S.J., Auler, P.A.M., 2013. Influência das frações de areia na retenção e disponibilidade de água em solos das formações Caiuá e Paranaíba. **Rev Bras Cienc Solo** 37, 613–621. <https://doi.org/10.1590/S0100-06832013000300007>
- Genuer, R., Poggi, J.-M., Tuleau-Malot, C., 2015. **VSURF: An R Package for Variable Selection Using Random Forests**. *R J* 7, 19–33.
- Genuer, R., Poggi, J.M., Tuleau-Malot, C., 2010. **Variable selection using random forests**. **Pattern Recognit Lett** 31, 2225–2236. <https://doi.org/10.1016/J.PATREC.2010.03.014>
- Germer, S., Neill, C., Krusche, A. V., Elsenbeer, H., 2010. Influence of land-use change on near-surface hydrological processes: Undisturbed forest to pasture. **J Hydrol (Amst)** 380, 473–480. <https://doi.org/10.1016/j.jhydrol.2009.11.022>
- Jaafarzadeh, M.S., Tahmasebipour, N., Haghizadeh, A., Pourghasemi, H.R., Rouhani, H., 2021. Groundwater recharge potential zonation using an ensemble of machine

- learning and bivariate statistical models. **Sci Rep** 11. <https://doi.org/10.1038/s41598-021-85205-6>
- Kaewdum, N., Chotpantararat, S., 2021. Mapping Potential Zones for Groundwater Recharge Using a GIS Technique in the Lower Khwae Hanuman Sub-Basin Area, Prachin Buri Province, Thailand. **Front Earth Sci (Lausanne)** 9. <https://doi.org/10.3389/feart.2021.717313>
- Kruskal, W.H., Wallis, W.A., 1952. Use of Ranks in One-Criterion Variance Analysis. **J Am Stat Assoc** 47, 583–621. <https://doi.org/10.1080/01621459.1952.10483441>
- Lentswe, G.B., Molwalefhe, L., 2020. Delineation of potential groundwater recharge zones using analytic hierarchy process-guided GIS in the semi-arid Motloutse watershed, eastern Botswana. **J Hydrol Reg Stud** 28, 100674. <https://doi.org/10.1016/j.ejrh.2020.100674>
- Margat, J., Gun, J. van der, 2013. Groundwater around the World. **CRC Press**. <https://doi.org/10.1201/b13977>
- Mello, C.R. de M., da Silva, A.M., 2013. **Hidrologia: princípios e aplicações em sistemas agrícolas**. UFLA, Lavras.
- Mello, C.R. de, Curi, N., 2012. Hydropedology. **Ciência e Agrotecnologia** 36, 137–146. <https://doi.org/10.1590/S1413-70542012000200001>
- Menezes, M.D. de, Junqueira Júnior, J.A., Mello, C.R. de, Silva, A.M. da, Curi, N., Marques, J.J., 2009. Dinâmica hidrológica de duas nascentes, associada ao uso do solo, características pedológicas e atributos físico-hídricos na sub-bacia hidrográfica do Ribeirão Lavrinha – Serra da Mantiqueira (MG). **Scientia Foretalis** 37, 175–184.
- Mualem, Y., 1976. A new model for predicting the hydraulic conductivity of unsaturated porous media. **Water Resour Res** 12, 513–522. <https://doi.org/10.1029/WR012i003p00513>
- Nosetto, M.D., Acosta, A.M., Jayawickreme, D.H., Ballesteros, S.I., Jackson, R.B., Jobbágy, E.G., 2013. Land-use and topography shape soil and groundwater salinity in central Argentina. **Agric Water Manag** 129, 120–129. <https://doi.org/10.1016/j.agwat.2013.07.017>
- Ogle, D.H., Doll, J.C., Wheeler, A.P., Dinno, A., 2015. FSA: Simple Fisheries Stock Assessment Methods. **CRAN: Contributed Packages**. <https://doi.org/10.32614/CRAN.package.FSA>
- Oliveira, P.T.S., Leite, M.B., Mattos, T., Nearing, M.A., Scott, R.L., de Oliveira Xavier, R., da Silva Matos, D.M., Wendland, E., 2017. Groundwater recharge decrease with increased vegetation density in the Brazilian cerrado. **Ecohydrology** 10. <https://doi.org/10.1002/eco.1759>
- Otto, S.R.L., 1988. **Estimativa da porosidade drenável em função das propriedades de solo orgânico**. Universidade Federal de Viçosa, Viçosa.
- Otoni, M.V., Otoni 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>
- Otoni, M.V., Souza, L.R., Barbosa, S.M., Andrade, R., Teixeira, W.G., Inda, A.V., Curi, N., 2024. **Mineralogia, estrutura e comportamento físico-hídrico de Latossolos brasileiros e comparação com solos de regiões temperadas** (No. 1), 1. Rio de Janeiro.
- Pinheiro, E., Van Lier, Q., Bezerra, A., 2017. Hydrology of a Water-Limited Forest under Climate Change Scenarios: The Case of the Caatinga Biome, Brazil. **Forests** 8, 62. <https://doi.org/10.3390/f8030062>
- Pinheiro, E.A.R., Metselaar, K., van Lier, Q. de J., de Araújo, J.C., 2016. Importance of soil-water to the Caatinga biome, Brazil. **Ecohydrology** 9, 1313–1327. <https://doi.org/10.1002/eco.1728>
- Pinto, L.C., de Mello, C.R., Norton, L.D., Owens, P.R., Curi, N., 2016. Spatial prediction of soil–water transmissivity based on fuzzy logic in a Brazilian headwater watershed. **Catena (Amst)** 143, 26–34. <https://doi.org/10.1016/j.catena.2016.03.033>
- Pinto, L.C., Mello, C.R. de, Norton, L.D., Silva, S.H.G., Taveira, L.R.S., Curi, N., 2017. Land-use effect on hydropedology in a mountainous region of Southeastern Brazil. **Ciência e Agrotecnologia** 41, 413–427. <https://doi.org/10.1590/1413-70542017414002017>
- Reatto, A., Bruand, A., Silva, E.M., Martins, E.S., Brossard, M., 2007. Hydraulic properties of the diagnostic horizon of Latosols of a regional toposequence across the Brazilian Central Plateau. **Geoderma** 139, 51–59. <https://doi.org/10.1016/j.geoderma.2007.01.003>
- Resende, S.B.; Lani, J.L.; Resende, M. Solo como reservatório dinâmico na produção de água. **Ação Ambiental**, Viçosa, v.3, p.23-26, 1998.
- Roa-García, M.C., Brown, S., Schreier, H., Lavkulich, L.M., 2011. The role of land use and soils in regulating water flow in small headwater catchments of the Andes. **Water Resour Res** 47. <https://doi.org/10.1029/2010WR009582>
- RStudio Team, 2020. RStudio: Integrated Development for R.
- Santana, M.L.T., dos Santos, F.F., de Carvalho, K.M., Peixoto, D.S., Uezu, A., Avanzi, J.C., Serafim, M.E., Nunes, M.R., van Es, H.M., Curi, N., Silva, B.M., 2023. Interactions between land use and soil type drive soil functions, highlighting water recharge potential, in the Cantareira System, Southeast of Brazil. **Science of The Total Environment** 903, 166125. <https://doi.org/10.1016/J.SCITOTENV.2023.166125>
- Santos, H.G. dos, Jacome, P.K.T., Anjos, L.H.C. dos, Oliveira, V.A. de, Lumbrellas, J.F., Coelho, M.R., Almeida, J.A. de, Araujo Filho, J.C. de, Oliveira, J.B. de, Cunha, T.J.F., 2018. **Sistema Brasileiro de Classificação de Solos**, 5. ed. rev. e ampl. ed. Embrapa, Brasília, DF.

- Selvam, S., Dar, F.A., Magesh, N.S., Singaraja, C., Venkatramanan, S., Chung, S.Y., 2016. Application of remote sensing and GIS for delineating groundwater recharge potential zones of Kovilpatti Municipality, Tamil Nadu using IF technique. **Earth Sci Inform** 9, 137–150. <https://doi.org/10.1007/s12145-015-0242-2>
- Sone, J.S., Sanches de Oliveira, P.T., Pereira Zamboni, P.A., Motta Vieira, N.O., Altrão Carvalho, G., Motta Macedo, M.C., Romeiro de Araujo, A., Baptaglin Montagner, D., Alves Sobrinho, T., 2019. Effects of Long-Term Crop-Livestock-Forestry Systems on Soil Erosion and Water Infiltration in a Brazilian Cerrado Site. **Sustainability** 11, 5339. <https://doi.org/10.3390/su11195339>
- Souza, E. de, Pontes, L.M., Fernandes Filho, E.I., Schaefer, C.E.G.R., Santos, E.E. dos, 2019. Spatial and Temporal Potential Groundwater Recharge: the Case of the Doce River Basin, Brazil. **Rev Bras Cienc Solo** 43. <https://doi.org/10.1590/18069657rbc20180010>
- UNESCO, 2022. **The United Nations World Water Development Report 2022: GROUNDWATER Making the invisible visible**. Paris.
- van Genuchten, M.Th., 1980. A Closed-form Equation for Predicting the Hydraulic Conductivity of Unsaturated Soils. **Soil Science Society of America Journal** 44, 892–898. <https://doi.org/10.2136/sssaj1980.03615995004400050002x>
- Wymore, A.W., 1993. **Model-Based Systems Engineering**. **CRC Press**. <https://doi.org/10.1201/9780203746936>

**PART THREE - FINAL CONSIDERATIONS**

## 1. LIMITATIONS AND FINAL CONSIDERATIONS

This study presents a Soil Quality Index (SQI) derived from a comprehensive dataset covering multiple regions of Brazil. However, its validation was limited to specific environmental conditions—primarily in areas where Latosols and Argisols predominate under agricultural and forest land uses, respectively—representing the most prevalent soil classes in the country. Despite the robustness of the HYBRAS database, limitations in the spatial and categorical coverage of key indicators restrict the quantitative spatial application of the SQI across the national territory. Furthermore, while the predictive performance of the model is considered reasonable given current data availability, the primary contribution of this work lies in proposing methodological alternatives for future efforts aimed at mapping groundwater recharge potential. These alternatives emphasize the integration of soil class and texture data with relevant environmental variables, thus offering a strategic framework to inform sustainable land and water management in Brazil.