

RELATÓRIO DE PROJETO

DEPARTAMENTO DE HIDROLOGIA



Propriedades Hidráulicas em Solos Brasileiros

ESTUDO DE PEDOFUNÇÕES DE PROPRIEDADES HIDRÁULICAS EM SOLOS BRASILEIROS

Em Editoração

Serviço Geológico do Brasil - CPRM

Dezembro 2021

SERVIÇO GEOLÓGICO DO BRASIL – CPRM
DIRETORIA DE HIDROLOGIA E GESTÃO TERRITORIAL – DHT
DEPARTAMENTO DE HIDROLOGIA

Relatório de Projeto
Departamento de Hidrologia

ÁREA: RECURSOS HÍDRICOS SUBTERRÂNEOS
SUBÁREA: LEVANTAMENTO DE RECURSOS HÍDRICOS SUBTERRÂNEOS

**ESTUDO DE PEDOFUNÇÕES DE
PROPRIEDADES HIDRÁULICAS EM
SOLOS BRASILEIROS**

(Relatório de Projeto – Ano 2021)

REALIZAÇÃO

Divisão de Hidrologia Aplicada

AUTORES

Marta Vasconcelos Ottoni

Em Editoração

Rio de Janeiro/RJ, 2021

MINISTÉRIO DE MINAS E ENERGIA

Ministro de Estado

Bento Albuquerque

Secretário de Geologia, Mineração e Transformação Mineral

Pedro Paulo Dias Mesquita

SERVIÇO GEOLÓGICO DO BRASIL – CPRM

DIRETORIA EXECUTIVA

Diretor-Presidente

Esteves Pedro Colnago

Diretora de Hidrologia e Gestão Territorial

Alice Silva de Castilho

Diretor de Geologia e Recursos Minerais

Marcio José Remédio

Diretor de Infraestrutura Geocientíficas

Paulo Afonso Romano

Diretor de Administração e Finanças

Cassiano de Souza Alves

COORDENAÇÃO TÉCNICA

Chefe do Departamento de Hidrologia

Frederico Claudio Peixinho

Chefe da Divisão de Hidrologia Aplicada

Adriana Dantas Medeiros

Consultora do Departamento de Hidrologia

Andrea de Oliveira Germano

EQUIPE TÉCNICA

Marta Vasconcelos Ottoni - CPRM

Marcos Vinícius Ferreira - CPRM

Guilherme Ferreira - CPRM

Alexandre Evsukoff – COPPE/UFRJ

João Herbert Moreira Viana - EMBRAPA

Quirijn de Jong van Lier – CENA/USP

Michele Bruna de Souza do Nascimento - CPRM

Pedro Gomes de Campos - CPRM

Wenceslau Geraldes Teixeira - EMBRAPA

ESTUDO DE PEDOFUNÇÕES DE PROPRIEDADES HIDRÁULICAS EM SOLOS BRASILEIROS (Relatório de Projeto – Ano 2021)

REALIZAÇÃO

Divisão de Hidrologia Aplicada

AUTORES

Marta Vasconcelos Ottoni

FOTOS DA CAPA: produção Luis Candido Gomes de Campos

Dados Internacionais de Catalogação-na-Publicação (CIP)

O91e	Ottoni, Marta Vasconcelos. Estudo de pedofunções de propriedades hidráulicas em solos brasileiros / Marta Vasconcelos Ottoni ; realização Divisão de Hidrologia Aplicada. – Rio de Janeiro : CPRM, 2021. 1 recurso eletrônico : PDF. – (Relatório de projeto – ano 2021) Relatório de projeto Departamento de Hidrologia: área recursos hídricos subterrâneos, subárea: levantamento de recursos hídricos subterrâneos. ISBN 978-65-5664-203-1 1.Ciência do solo. 2.Hidráulica. 3.Hidrologia. I. Ottoni, Marta Vasconcelos. II. Título. III. Série. CDD 631.4
------	---

Ficha catalográfica elaborada pela bibliotecária Teresa Rosenhayme CRB-7/5662

Direitos desta edição: Serviço Geológico do Brasil – CPRM
Permitida a reprodução desta publicação desde que mencionada a fonte

Serviço Geológico do Brasil - CPRM
www.cprm.gov.br
seus@cprm.gov.br

APRESENTAÇÃO

A necessidade de se investigar os processos hidrológicos de maneira integrada tem imperado no campo disciplinar da hidrologia, com o principal viés de se garantir uma gestão integrada e mais eficiente dos recursos hídricos. As relações existentes entre os fluxos de água de superfície e subsuperfície ainda não são claras e a complexidade pode aumentar quando se amplia a escala espacial de observação. Aliado a esse fato, o conhecimento dos processos hidrológicos passa pelo levantamento de dados que nem sempre encontram-se disponíveis na escala de interesse de estudo. No caso dos solos, estudos de movimento de água na zona vadosa demandam informações de propriedades hidráulicas, que muitas vezes também não estão acessíveis. Esse fato tem motivado o avanço no desenvolvimento das chamadas funções de pedotransferência que a partir de medidas de mais fácil determinação do solo no campo e em laboratório permitem estimar variáveis do solo de determinação mais complexa, como são a maioria das propriedades hidráulicas.

O Projeto “Estudos de Pedofunções de Propriedades Hidráulicas em Solos Brasileiros” é um estudo institucional promovido pelo Departamento de Hidrologia do Serviço Geológico do Brasil e tem como objetivo disponibilizar informações de propriedades hidráulicas por meio da produção de diversos modelos de predição nacionais dessas variáveis de solo que levem em consideração preditores que sejam facilmente determinados e/ou disponibilizados em amplas bases de dados, contribuindo para estudos de modelagens, iniciativas de mapeamento de solos e projetos de engenharia.

RESUMO

A simulação de fluxo de água na matriz porosa dos solos requer informações de propriedades hidráulicas do solo que nem sempre se encontram disponíveis. As funções de pedotransferência, terminologia proposta por Bouma (1989), vem cobrir essa deficiência de dados dessas propriedades de solo, que permitem estimar essas informações de mais complexa determinação a partir de atributos do solo de mais fácil medição no campo e em laboratório. No Brasil, há diversos modelos para predição das propriedades hidráulicas, mas carece ainda de iniciativas mais atuais desenvolvidas na escala do território nacional e que levem em consideração uma estrutura hierarquizada das variáveis de entrada. Ainda também são poucos no país os estudos de validação funcional dessas pedofunções que avaliam o desempenho desses modelos na simulação de um determinado processo hidrodinâmico no solo. Diante deste cenário, o Departamento de Hidrologia do Serviço Geológico do Brasil propôs o Projeto “Estudo de Pedofunções de Propriedades Hidráulicas em Solos Brasileiros” com prazo de duração de 5 (cinco) anos a partir de 2022, que tem como objetivo gerar modelos de predição de propriedade hidráulicas do solo em escala nacional a partir de atributos do solo comumente disponíveis em bases de dados de solos e de fácil determinação no campo ou em laboratório. As variáveis de saída dos modelos são os dados de retenção de água para diferentes potenciais matriciais, os mais comumente utilizados, como o potencial representativo da capacidade de campo, do ponto de murcha e da saturação, além dos parâmetros da equação de van Genuchten que descrevem esses dados de retenção de água, bem como a condutividade hidráulica saturada. Estudos de validação funcional dos modelos ajustados também estão previsto no contexto deste trabalho. Este projeto pode contribuir para diferentes estudos de modelagens, iniciativas de mapeamento de solos e projetos de engenharia. Pretende-se estender o escopo deste estudo para incluir a condutividade hidráulica não saturada como variável de saída, quando um volume mais expressivo desses dados se encontrarem disponíveis, bem como incluir o desenvolvimento de outros modelos de predição que levem em consideração os dados morfológicos de perfis de solos e covariáveis ambientais extraídas dos modelos digitais de terreno, estes últimos com a vantagem de se poder estimar as propriedades de solo em larga escala territorial.

ABSTRACT

The simulation of water flow in the porous matrix of soils requires information on soil hydraulic properties that are not always available. Pedotransfer functions, terminology proposed by Bouma (1989), allow the estimation of soil information that is more difficult to determine, for example, the soil hydraulic properties, from soil attributes that are easier to measure in the field and in the laboratory. In Brazil, there are several models for predicting soil hydraulic properties, but there is still a lack of more current initiatives developed at the national territory scale and that take into account a hierarchical structure of input variables. The functional validation studies of these pedofunctions that assess the performance of these models in the simulation of a certain hydrodynamic process in the soil are still scarce in the country. Given this scenario, the Department of Hydrology of the Geological Survey of Brazil proposed the Project "Study of Pedofunctions for Hydraulic Properties in Brazilian Soils" with a duration of 5 (five) years from 2022, which aims to generate soil hydraulic properties models on a national scale from soil attributes commonly available in soil databases and easy to determine in the field or in the laboratory. The output variables of the models are the water retention data for different matric potentials, the most commonly used, such as the representative potential of field capacity, wilting point and saturation, in addition to the parameters of the van Genuchten equation that describe these water retention data, as well as the saturated hydraulic conductivity. Functional validation studies of the generated models are also provided in the context of this work. This project can contribute to different modeling studies, soil mapping initiatives and engineering projects. It is intended to extend the scope of this study to include unsaturated hydraulic conductivity as an output variable, when a more expressive volume of this data is available, as well as to include the development of other prediction models that take into account the morphological data of soil profiles and environmental covariates extracted from digital terrain models, the latter with the advantage of being able to estimate soil properties on a large territorial scale.

SUMÁRIO

Sumário

1. INTRODUÇÃO	12
2. OBJETIVO GERAL.....	14
3. OBJETIVOS ESPECÍFICOS.....	14
4. EQUIPE	15
5. MATERIAIS E MÉTODOS	16
5.1. Base de dados	16
5.1.1 Base de dados complementares	18
5.1.2. Processamento dos dados	19
5.1.3. Análise dos dados e composição das bases de dados de solos para ajuste dos modelos	21
5.2 Variáveis de entrada e de saída das pedofunções	22
5.3. Desenvolvimento das pedofunções	23
5.3.1 Modelos de ajuste	24
5.3.2 Calibração x Validação e Teste dos modelos ajustados	25
5.3.3 Pedofunções da literatura	25
5.4. Análise de desempenho das pedofunções	26
5.5. Intervalo de Confiança	27
5.6. Domínio dos modelos.....	28
6. RESULTADOS ESPERADOS	28
7. PRODUTO ESPERADO	29
8. CRONOGRAMA DAS ATIVIDADES	29

9. CONCLUSÕES	30
10. Referências Bibliográficas	31

LISTA DE FIGURAS

FIGURA 01 – TRIÂNGULO ESTRUTURAL.	19
FIGURA 02 – DETECÇÃO DE OUTLIERS PELAS MÉDIAS DE DISTÂNCIA EUCLIDIANA EM ORDEM CRESCENTE. OUTLIERS EM VERMELHO.	23
FIGURA 03 – REPRESENTAÇÃO ESQUEMÁTICA DOS MODELOS DE PROPRIEDADES HIDRÁULICAS PARA AJUSTE.	26

LISTA DE TABELAS

TABELA 01 – EXEMPLOS DE FUNÇÕES DE PEDOTRANSFERÊNCIAS DE PROPRIEDADES HIDRÁULICAS AJUSTADAS PARA DOIS TRABALHOS DA LITERATURA.	28
---	----

1. INTRODUÇÃO

A simulação do fluxo de água no solo é realizada normalmente utilizando-se a equação de Richards, com exceção de situações em que os fluxos preferenciais predominam, quando modelos de porosidade dupla (dual-porosity), permeabilidade dupla (dual permeability), modelos multi-porosidade (multi-porosity) e/ou multi-permeabilidade (multi-permeability) (PRUESS; WANG, 1987; GERKE; VAN GENUCHTEN, 1993; GWO *et al.*, 1995; JARVIS, 1998) podem ser mais adequados. Na formulação de Richards demandam-se informações de propriedades hidráulicas do solo, a curva de retenção de água e condutividade hidráulica, cujas determinações no campo e em laboratório são laboriosas e que demandam elevados custos.

No Brasil, os dados dessas propriedades encontravam-se dispersos em teses, dissertações e publicações e o HYBRAS - *Hydrophysical Database for Brazilian Soils* (OTTONI *et al.*, 2018) foi uma iniciativa do Serviço Geológico do Brasil de compilar essas informações e disponibilizá-las para amplo uso. No entanto, a distribuição dos dados disponíveis no HYBRAS ainda não cobre a ampla variedade de solos e biomas existentes no Brasil, carecendo de esforços conjuntos de diferentes instituições do país produtoras de dados de solos, para a inclusão de mais informações nessa base de dados.

Nesse sentido, estimativas das propriedades hidráulicas a partir de informações mais facilmente medidas e amplamente disponíveis em bases de dados de solos no Brasil se fazem ainda necessárias. Funções de pedotransferência (FPTs), termo introduzido por Bouma (1989), traduzem a relação existente entre aquilo que é difícil medir e o que se quer estimar a partir de informações de fácil determinação. Essas pedofunções tem sido úteis como dados de entrada de solo em modelos climatológicos, hidrológicos, de erosão, em estudos de irrigação e drenagem e para trabalhos de mapeamento digital de solos (LIN, 2010; ZHAO *et al.*, 2018; RIBEIRO *et al.*, 2018).

Segundo recente levantamento pela autora de trabalhos que desenvolveram FPTs de propriedades hidráulicas no Brasil (OTTONI, 2019), foram catalogados 57 modelos no país no período de 1987 até 2019 (quando foi encerrada a pesquisa), a maioria concentrada no estado do Rio Grande do Sul, e quase todos destinados a estimar alguns pontos da curva de retenção, como a capacidade de campo e ponto de murcha.

Em escala nacional encontram-se disponíveis apenas as FPTs de Tomasella, Hodnett e Rossato (2000) e Tomasella *et al.* (2003), os dois para predição dos parâmetros da equação de Van Genuchten (1980), incluindo o último outros modelos para predição de alguns pontos da curva de retenção. Quanto à condutividade hidráulica saturada (Ksat), apenas a FTP de Ottoni *et al.* (2019)

foi calibrada para o território nacional. São inexistentes pedofunções de condutividade hidráulica não saturada no país. Além disso, os modelos brasileiros levam em consideração variáveis preditoras que nem sempre são medidas pelo usuário, havendo a demanda, dessa forma, por pedofunções com uma estrutura hierarquizada de variáveis de solos de entrada, como realizado em Rosetta (SCHAAP *et al.*, 2001).

Constatou-se também nesse estudo de revisão de pedofunções no Brasil a necessidade de se explorar diferentes técnicas de inteligência computacional para a geração dessas FPTs. Destacou-se que em torno de 82% das pedofunções brasileiras foram geradas a partir de modelos de regressão multilinear. Além disso, registraram-se poucos trabalhos com estudo de validação dos modelos e um número ainda menor com validação funcional, o que indicaria o desempenho da FPT gerada na simulação de um determinado processo hidrodinâmico no solo.

Outra carência observada nesse trabalho de revisão de FPTs no país diz respeito à aplicabilidade de uma determinada pedofunção para uma outra base de dados não incorporada na calibração. Não é usual a indicação nos trabalhos do domínio dos modelos, ou seja, a região do conjunto de dados de solos utilizado no ajuste das FPTs, limitando, portanto, sua aplicação em outras bases de dados. Por fim, esse estudo também permitiu concluir que ainda não há registros de pedofunções de propriedades hidráulicas no país que levem em consideração características morfológicas do perfil de solo, como cor, estrutura, consistência, profundidade. Essas pedofunções seriam úteis na transformação do legado de dados de solos existentes no país que hoje, segundo BDSOLOS ([Sistema de Informação de Solos Brasileiros \(embrapa.br\)](http://embrapa.br)), concentra em torno de 35,000 observações de solos de aproximadamente 5300 pontos amostrais com descrição morfológica, em informações de armazenamento e fluxo de água que raramente são descritos nessas bases de dados.

Demanda-se, portanto, no Brasil, avanços nos estudos de FTPs de propriedades hidráulicas, em especial, na geração de modelos preditores nacionais que levem em consideração a diversidade de solos e biomas existentes no país, que sejam facilmente disponíveis e aplicáveis a diferentes bases de dados e que produzam estimativas de qualidade e em formatos de saída que atendam às diferentes demandas dos usuários por essas informações.

Diante desse cenário, o Departamento de Hidrologia da CPRM instituiu o projeto Estudos de Pedofunções de Propriedades Hidráulicas em solos Brasileiros inserido no contexto do programa Estudos de Caracterização Hidrológica de Solos, este último com o objetivo de ampliar o conhecimento dos processos hidrológicos no solo e da sua relação com os fluxos de água superficiais e subterrâneos nas diferentes escalas de espaço-tempo.

O período de execução deste projeto é de 2022 a 2026, mas espera-se que seja de caráter continuado, visando atualizar recorrentemente os modelos de predição das propriedades hidráulicas a medida que um volume mais expressivo de dados estiver disponível e incluir outros temas de estudo que se fizerem necessário desenvolver no contexto desse projeto.

2. OBJETIVO GERAL

O projeto tem como propósito disponibilizar informações de propriedades hidráulicas do solo por meio da produção de diversos modelos de predição nacionais que levem em consideração preditores que sejam facilmente determinados e/ou disponibilizados em amplas bases de dados, contribuindo para diferentes estudos de modelagem, iniciativas de mapeamento de solos e projetos de engenharia.

3. OBJETIVOS ESPECÍFICOS

- Desenvolver FPTs nacionais para as seguintes propriedades hidráulicas:
 - retenção de água em diferentes potenciais matriciais;
 - condutividade hidráulica saturada;
 - parâmetros de equações que descrevem os dados de retenção de água (ex: equação de van Genuchten, van Genuchten, 1980)
- Elaborar essas pedofunções de propriedades hidráulicas levando em conta uma estrutura hierarquizada de preditores. Os preditores escolhidos serão aqueles usualmente utilizados na literatura para se estimar as propriedades hidráulicas do solo;
- Avaliar os modelos de predição gerados neste trabalho, além daqueles registrados na literatura, para uma base de dados teste que não esteja incluída no conjunto de dados utilizado na calibração dos modelos;
- Realizar estudos de validação funcional desses modelos no sentido de se investigar o desempenho dessas FPTs para diferentes cenários de modelagem de água no solo.

Neste trabalho, utilizaremos como referência de base de dados de solo o HYBRAS, que também faz parte de um projeto institucional da CPRM e que é alimentado continuamente, seja por dados gerados internamente pela instituição, como também por informações compiladas da literatura. Pretende-se também fazer uso de outras bases de dados de solos existentes no país e internacionais para se ampliar o universo de dados para geração dos modelos.

4. EQUIPE

Nome	Atribuições	Instituição
Marta Vasconcelos Ottoni	Coordenação, proposição da metodologia e acompanhamento da execução das atividades do projeto, geração dos modelos, elaboração dos relatórios e trabalhos científicos	DIHAPI/DEHID/DHT/CPRM
Marcos Vinícius Ferreira	Ajuste dos modelos e programação em Python	DISEGE/DEGEO/DGM/CPRM
Guilherme Ferreira	Análise dos dados	DIGECO/DEREM/DGM/CPRM
Alexandre Evsukoff	Consultoria em aprendizado de máquina	COPPE/UFRJ
João Herbert Moreira Viana	Análise dos resultados	Embrapa Sete Lagoas-MG
Quirijn de Jong van Lier	Validação funcional dos modelos	CENA/USP
Michele Bruna de Souza do Nascimento	Digitação das equações de pedofunções da literatura e aplicação desses modelos para a base de dados do estudo	DIHEXP/DEHID/DHT/CPRM
Pedro Gomes de Campos	Digitação e processamento de dados de solos	DIHAPI/DEHID/DHT/CPRM
Wenceslau Geraldes Teixeira	Revisor geral do trabalho	Embrapa Solos-RJ

5. MATERIAIS E MÉTODOS

O desenvolvimento de pedofunções requer a disponibilidade de bases de dados de solos que cubram a diversidade dos ambientes que se quer reproduzir, contendo as variáveis de entrada, numéricas ou categóricas, e de saída a serem modeladas, além da escolha de uma ou mais técnicas de aprendizagem de máquina capazes de explorar e modelar as relações existentes entre os preditores e preditos. Uma análise exploratória dos dados visando identificar os erros grosseiros e outliers e eliminá-los, se necessários, também deve ser efetuada antes do estudo da geração das pedofunções.

5.1. Base de dados

A base de dados HYBRAS, versão 1, foi utilizada neste estudo para desenvolvimento das FPTs. Ela detém 1075 amostras com cobertura em diversos estados do Brasil e nas principais classes de solos do país, Latossolo e Argissolos, contendo informações de densidade do solo, densidade das partículas, porosidade total, teor de carbono orgânico, dados de retenção de água em diferentes potenciais matriciais, todos determinados em amostras indeformadas, os parâmetros ajustados da equação de van Genuchten (1980) para os dados de retenção de água do HYBRAS e condutividade hidráulica saturada, além de informações de uso e cobertura do solo e descrição geral do local de amostragem com coordenadas geográficas para muitos casos. Para maiores informes dessa base de dados acessar: [Serviço Geológico do Brasil \(cprm.gov.br\)](http://cprm.gov.br).

Desde 2018 quando o HYBRAS foi lançado no site da CPRM, deu-se continuidade ao levantamento de dados de propriedades hidráulicas no país e nesse período até os tempos atuais foi possível reunir essas informações de 300 amostras de solos adicionais, essas ainda não oficialmente incluídas no HYBRAS.

No total, 1375 amostras de solos foram reunidas para serem aplicadas no presente estudo de pedofunções e a esse conjunto de dados chamaremos aqui de HYBRAS completo, para se diferenciar do HYBRAS original com 1075 amostras.

A base de dados deste estudo é um extrato da base de dados do HYBRAS completo, o que significa que nem todas os campos do HYBRAS constam no conjunto de dados para geração das pedofunções. A esta base de dados de solos deste estudo chamaremos de BDPEDO. Na BDPEDO estão contidos os seguintes campos do HYBRAS (em negrito), conforme abaixo. Quanto aos dados de retenção de água, extraíram-se do HYBRAS apenas aqueles de interesse deste projeto (dados

de retenção de água para os potenciais 0kpa, 6kPa, 10kpa, 33kpa e 1500kpa), recebendo cada uma dessas variáveis de retenção uma denominação específica, conforme apresentada abaixo.

- Identificação da amostra: campo **code**
- Referência de onde os dados foram extraídos: campo **reference** (no HYBRAS chamado de publicn);
- Descrição geral do local de amostragem: campos: **general description**, com descrição sucinta do local onde as amostras foram extraídas; **Profi_id2** que indica o número do perfil a qual amostra de solo pertence; **LatitudeOR** e **LongitudeOR**, referem-se as coordenadas geográficas em graus decimais; **city**: município; **state**: estado; Coordinates Comments (campo não presente no HYBRAS - versão 1, a ser incluído na próxima versão): comentários a respeito da origem da coordenada que pode ter ser obtida diretamente do trabalho de origem ou estimada; **Soil Class_BR**: classe pedológica segundo SiBCS (Sistema Brasileiro de Classificação do Solos) mais recente (primeiro nível categórico); **Soil Class_FAO**: classe pedológica segundo WRB/FAO;
- Descrição do local da amostra: campos: **depth**: profundidade em cm onde a amostra foi extraída; **horizon**: descrição do horizonte; **horizon class**: classificação do horizonte em superficial e subsuperficial (Ottoni *et al.*, 2018); **texture**: classe textural segundo USDA;
- Atributos do solo: **clay (%)**: argila; **silt (%)**: silte; **sand (%)**: areia; **bulk density (g/cm3)**: densidade do solo; **particle density (g/cm3)**: densidade das partículas; **organic carbon (%)**: teor de carbono orgânico;
- Propriedades hidráulicas: **ksat (mm/h)**: condutividade hidráulica saturada; **TH0 (cm3/cm3)**: umidade volumétrica no potencial matricial de 0kPa; **TH6 (cm3/cm3)**: umidade volumétrica no potencial matricial de 6kPa; **TH10 (cm3/cm3)**: umidade volumétrica no potencial matricial de 10kPa; **TH33 (cm3/cm3)**: umidade volumétrica no potencial matricial de 33kPa; **TH1500 (cm3/cm3)**: umidade volumétrica no potencial matricial de 1500kPa;
- Parâmetros da equação de van Genuchten: **dvg_sat**, **dg_res**, **dg_alpha**, **dg_n**

Muitos registros do HYBRAS completo dispunham de informações de umidade volumétrica no potencial matricial de 30kPa. Essas foram presumidas como umidade no potencial de 33kPa (TH33) no BDPEDO, além dos próprios registros com umidade volumétrica no potencial matricial de 33kPa. Os dados de TH0 referem-se à umidade volumétrica de saturação, mas algumas vezes presumiu-se para essa variável a porosidade total, quando a primeira não se encontrava disponível.

5.1.1 Base de dados complementares

Em Ottoni (2017) foi apresentada uma nova metodologia de classificação da estrutura dos vazios do solo, na qual é sugerida um triângulo estrutural similar ao textural (Figura 01) para classificação de solos segundo três faixas de tamanhos dos poros relativo, chamados de microespaço (MICRO), mesoespaço (MESO) e macroespaço (MACRO).

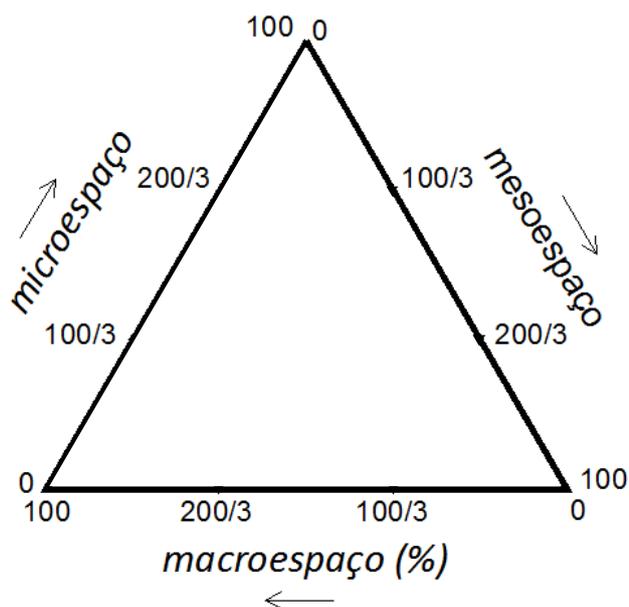


Figura 01 – Triângulo Estrutural

Essas frações são definidas a partir de três dados de retenção de água (TH6, TH33, TH1500) que representam as variações de formato da curva de retenção de água, além dos parâmetros da equação de van genuchten (equação 04), ajustados exclusivamente para esses dados de retenção, incluindo a retenção de água na saturação. Nas equações abaixo são apresentadas as formulações dessas três frações porosas. Maiores detalhes sobre essa metodologia de classificação e aplicações são apresentados em Ottoni (2017).

$$\text{MICRO} = \frac{\text{TH1500} - \theta_r}{\theta_s - \theta_r} \quad (01)$$

$$\text{MESO} = \frac{\text{TH6} - \text{TH1500}}{\theta_s - \theta_r} \quad (02)$$

$$\text{MACRO} = \frac{\theta_s - \text{TH}_6}{\theta_s - \theta_r} \quad (03)$$

em que θ_s e θ_r são a umidade volumétrica de saturação e residual, respectivamente, da equação de van Genuchten (1980) (equação 4).

Neste trabalho utilizaremos essa proposição para classificar os solos do BDPEDO, além de outras amostras de solos internacionais disponíveis na literatura provenientes de ambientes de clima temperado e que contém os dados de retenção de água para classificação estrutural segundo Ottoni (2017). Faremos isso para verificar se há alguma sobreposição no triângulo entre os dados nacionais e internacionais e em havendo, é um forte indicador que esses solos sobrepostos são semelhantes entre si do ponto de vista da estrutura de seus vazios. Buscaremos tirar proveito dessa semelhança para aumentar a abrangência da base de dados de solos a ser utilizado no ajuste das pedofunções, o que significa que incluiremos mais dados na BDPEDO a partir daquelas amostras de solos internacionais que apresentaram sobreposição no triângulo estrutural com os dados nacionais. Não utilizaremos os dados de solos internacionais, sem antes realizar essa triagem, pois se reconhece na literatura que informações de solos de clima temperado são incapazes de estimar com acurácia propriedades hidráulicas em solos tropicais brasileiros, em especial, os de textura fina (OTTONI *et al.*, 2018, 2019; TOMASELLA; HODNETT; ROSSATO, 2000)

Outras bases de dados de solos brasileiros que estiverem disponíveis na literatura mais recente também serão enfoque de investigação neste estudo, mas não precisarão passar obrigatoriamente por esse procedimento de classificação, como realizado para os solos internacionais.

5.1.2. Processamento dos dados

Os dados publicados na versão original do HYBRAS foram consistidos segundo metodologia apresentada em Ottoni (2018), sendo esses procedimentos de consistência replicados para os dados complementares deste estudo, o que incluiu as 300 amostras de solos nacionais adicionais, conforme comentado na seção 5.1.

De modo geral o trabalho de consistência dos dados que serão utilizados neste estudo compreende as seguintes atividades:

- a) Padronização das classes pedológicas para o SiBCS (Sistema Brasileiros de Classificação de Solos) mais recente;

- b) Padronização das classes de uso e cobertura dos solos para o sistema europeu LUCAS-Land Use/Cover Area Frame Survey (EUROPEAN COMMISSION, 2009), adotado como padrão no HYBRAS, mas que existe previsão de alteração para uso das classes do MapBiomas;
- c) Otimização dos quatro parâmetros da equação de van Genuchten (θ_s , θ_r , α e n) (equação 04) aos dados experimentais de retenção de água, considerando $m=1-1/n$. Será escolhido um algoritmo de ajuste que leve em consideração o encontro de uma solução de parâmetros que tenha o menor valor da função objetivo. Inicialmente, o ajuste do modelo se dará considerando valores fixos para os parâmetros iniciais. Os resultados gerados para os parâmetros de van Genuchten (equação 4) serão então presumidos como o novo valor inicial dos parâmetros e o modelo de van Genuchten (equação 4) será ajustado uma vez mais;
- d) Avaliação daqueles registros que contém valores extremos nas variáveis numéricas de solos, segundo faixas de valores usualmente aceitáveis e registradas na literatura. Os registros que tiveram valores enquadrados como extremos em qualquer uma das variáveis numéricas da base de dados serão excluídos do BDPEDO para evitar maiores perturbações no processamento dos dados.
- e) Checagem do somatório das frações granulométricas. Todos os registros com somatório das frações no intervalo entre 97% a 103%, excluindo o de 100%, terão os valores do teor de silte alterados para que a soma das suas frações resulte em 100%. Amostras com resultados fora dessa faixa serão excluídos da base de dados;
- f) Investigação do comportamento dos valores dos dados de retenção de água pelo potencial matricial. Os dados de retenção de água decrescem com o aumento do valor do potencial matricial e quando essa tendência não for observada para pelo menos dois pontos de medição do conjunto dos dados de retenção de água de uma amostra, optar-se-á por excluir a amostra da base de dados do estudo. Em havendo essa inconsistência para apenas um ponto desse conjunto de dados, a exclusão desse ponto se dará apenas quando o mesmo estiver muito fora da tendência das demais medições. Se a medição em questão apresentar apenas uma suave tendência anômala em relação as outras determinações, ela será mantida na base de dados.

Caso sejam utilizadas aquelas amostras de solos internacionais, conforme previsto na seção 5.1.1, adotar-se-á apenas as atividades de consistência c), d) e e), f), pois as atividades a) e

b) dependem de informações que não se encontram usualmente disponíveis nessas bases de dados, segundo os padrões adotados neste estudo.

5.1.3. Análise dos dados e composição das bases de dados de solos para ajuste dos modelos

Nesta etapa do projeto prevê-se a execução das seguintes atividades nessa sequência:

- a) Elaboração de estatísticas da base dos dados com valores médios, mínimos, máximos, desvio padrão e quartis da BDPEDO;
- b) Definição do percentual de amostras de solo da BDPEDO nos estados brasileiros e por grupos de solos, como as classes texturais, pedológicas e de uso e cobertura;
- c) Elaboração de gráficos de distribuição de frequência dos valores das variáveis de entrada e saída (seção 5.2) e de correlação entre as variáveis;
- d) Padronização das variáveis de solo da BDPEDO pelo método z-score (padronização é realizada pela a média e desvio padrão do conjunto de dados) ou por qualquer outro método de padronização mais adequado ao conjunto de dados do estudo, se for necessário;
- e) Elaboração de gráficos box-plot para as variáveis de solo de interesse;
- f) Avaliação da presença de valores ausentes na BDPEDO. Neste trabalho opta-se por não eliminar todos os registros ausentes das variáveis de solo de interesse simultaneamente, pois isso implicará numa redução significativa de amostras de solo da BDPEDO para ajuste dos modelos. Opta-se, portanto, por compor uma base de dados de solos específica a partir da base de dados completa (BDPEDO), para cada modelo que se pretende ajustar neste trabalho (seção 5.3), contendo em cada uma dessas bases de dados registros com informações completas (sem falhas) daquelas variáveis de solo envolvidas no ajuste do modelo de interesse.
- g) Avaliação de outliers para as bases de dados do estudo. Essa avaliação será realizada pelo método baseado em distâncias que leva em consideração todas as variáveis da base de dados de interesse simultaneamente. Segundo esse método, calcula-se a distância de um registro para todos os outros, a partir de uma matriz de distâncias que neste trabalho será a euclidiana, ordenando-se, em seguida, os valores médios dessas distâncias, obtidas para cada um dos registros, em ordem crescente. Os outliers são classificados como aqueles registros com maiores valores de distância média. Eles podem ser definidos visualmente a partir do gráfico das distâncias médias ordenadas, escolhendo-se a distância referente à inflexão da curva de tendência dos pontos de distância média ordenada (vide Figura 02, com

exemplo). Todos os registros com distância superiores à daquela identificada na inflexão poderão ser eliminados da base de dados para a geração dos modelos.

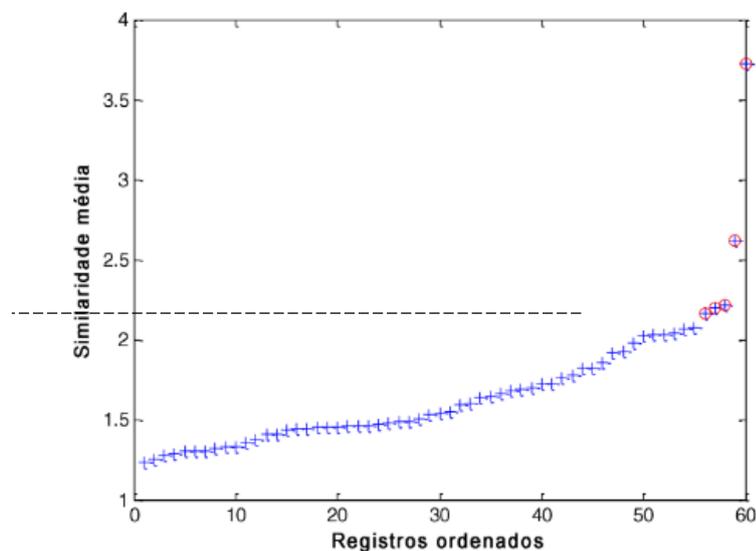


Figura 02 – Detecção de outliers pelas médias de distância euclidiana em ordem crescente. Outliers em vermelho. (Figura extraída do livro Inteligência Computacional: Fundamentos e aplicações, EVSUKOFF, 2020, p. 52)

- h) Representação gráfica do conjunto de dados transformados nas coordenadas transformadas pelos dois primeiros componentes principais, com a identificação dos correspondentes outliers detectados, que na sua maioria devem se encontrar na região mais extrema do domínio de dados transformados.

5.2 Variáveis de entrada e de saída das pedofunções

As propriedades hidráulicas a serem estimadas por modelos de pedofunções são:

- A retenção de água de potenciais matriciais representativos da capacidade de campo, para os quais, no Brasil, adotam-se usualmente os de 6 kPa (TH6), 10 kPa (TH10) e 33 kPa (TH33);
- Ponto de Murcha Permanente, tido como a retenção de água no potencial matricial de 1500 kPa (TH1500);
- Umidade de saturação representada pela retenção de água no potencial de 0 cm (TH0) ou pela porosidade total, na inexistência de informação do primeiro. Entende-se a porosidade total como a relação entre a densidade do solo e das partículas pela expressão: $1 - \text{densidade do solo} / \text{densidade das partículas}$;

- Parâmetros da equação de van Genuchten (1980) (θ_s , θ_r , α e n), conforme expressas na equação 04:

$$\theta(h) = \theta_r + \frac{\theta_s - \theta_r}{[1 + |\alpha h|^n]^{1-1/n}} \quad (04)$$

em que θ corresponde ao conteúdo de água ($m^3 m^{-3}$), θ_s e θ_r , respectivamente, aos teores de água residual e de saturação ($m^3 m^{-3}$), h ao potencial matricial da água no solo (kPa), α ao inverso da pressão de entrada de ar no maior poro do solo, e n parâmetro relacionado ao formato da curva.

- Condutividade hidráulica saturada;

Como variáveis preditoras pretende-se incluir aquelas presentes na base de dados do HYBRAS e que são usualmente utilizadas na predição de dados de retenção de água e K_{sat} , que são:

- Densidade do solo;
- Teor de carbono orgânico;
- Frações granulométricas (teor de areia, silte e argila);
- Retenção de água no potencial de 6kPa;
- Porosidade efetiva: Porosidade Total – TH33;

A retenção de água no potencial de 6kPa foi incluída como variável de entrada, além de ser uma variável a ser predita, pois é uma propriedade de mais fácil e rápida medição em laboratório, comumente determinada em estudo de manejo de solo, representando a microporosidade em solos brasileiros. Além disso, reconhece-se na literatura que o uso de dados de retenção de água tem se mostrado adequado na predição de K_{sat} e de retenção de água a outros potenciais (RAWLS; BRAKENSIEK, 1982; OTTONI *et al.*, 2019).

5.3. Desenvolvimento das pedofunções

São previstos sete modelos de predição, o primeiro mais simples, representado por três variáveis preditoras, as frações granulométricas e o último, mais complexo, caracterizado por todas as variáveis preditoras numéricas da base de dados, que no caso são seis, conforme Figura 03.

Os modelos intermediários têm como referência de preditores as frações granulométricas, presentes em todos os modelos, e que se encontram intercaladas com as outras variáveis de entradas em grupos de quatro a cinco variáveis, conforme ilustrado na Figura 03.

Os modelos ajustados para Ksat são os únicos a utilizarem o preditor porosidade efetiva, já que se reporta na literatura o seu bom desempenho de predição de Ksat (OTTONI *et al.*, 2019). Assim, para essa variável de saída são previstos oito modelos de ajuste, além dos sete antes comentado.

Assim, considerando o número de variáveis de saída de retenção de água (5), dos parâmetros de van Genuchten (4) e de condutividade hidráulica saturada (1) e os tipos de modelos a serem ajustados (Modelo 1 a 7 + Modelo 8 para Ksat) serão desenvolvidos 78 modelos de ajuste, isso, sendo considerar as técnicas de aprendizado de máquina a serem testadas, que segundo a seção 5.3.1, serão em quatro (regressão linear, random forest, catboost e rede neural), resultando em 312 modelos.

Em cada um dos sete modelos apresentados na Figura 03 pretende-se também investigar a importância de outras representações das variáveis de entrada de forma isolada (ex: logaritmo da relação entre duas frações granulométrica) ou em conjunto (ex: areia x silte ou areia x silte²).

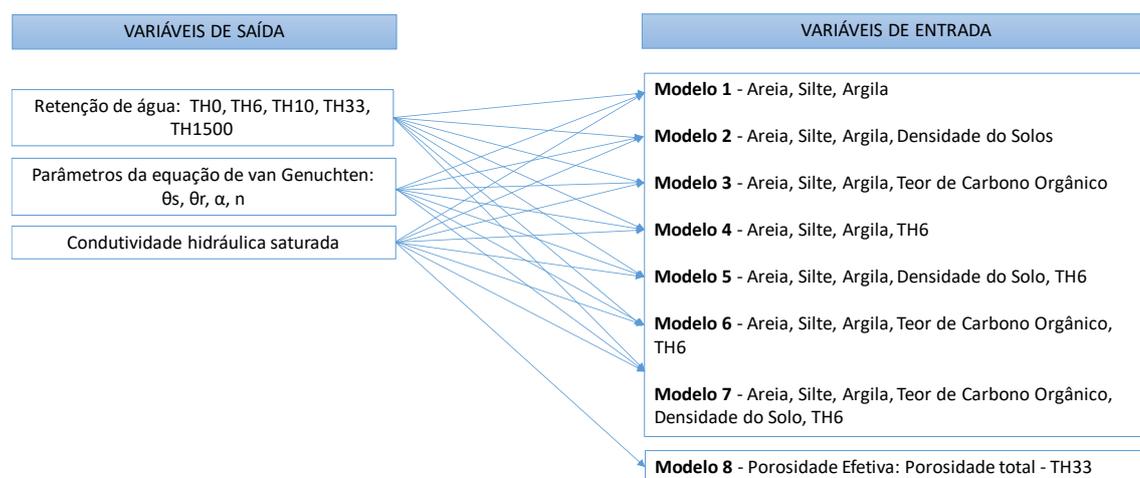


Figura 03 – Representação esquemática dos modelos para predição das propriedades hidráulicas do solo.

5.3.1 Modelos de ajuste

Neste trabalho pretende-se testar os modelos de regressão multilinear simples, os de árvore de decisão (RANDOM FOREST: HO, 1995; CATBOOST: PROKHORENKOVA *et al.*, 2018) e Redes Neurais (MAREN, 1990). Os modelos lineares já têm sido amplamente utilizados no Brasil para predição de propriedades hidráulicas e utilizaremos eles como referência e comparação das estimativas dos dois últimos modelos, também reportados na literatura como adequados na predição dessas propriedades de solo (SZABÓ; WEYNANTS; WEBER, 2021; SCHAAP *et al.*, 2001).

Os resultados de cada um desses modelos serão avaliados individualmente por meio dos índices estatísticos descritos na seção 5.4, mas pretende-se também investigar o potencial de predição do modelo combinado que seria a resultante da combinação linear das estimativas dos modelos individuais.

5.3.2 Calibração x Validação e Teste dos modelos ajustados

Da base de dados de solos selecionados neste estudo serão extraídos 20% dos dados para realização de teste dos modelos ajustados. Assim, somente 80% restantes serão utilizados na calibração e validação das pedofunções.

A calibração e validação dos modelos será efetuada por validação cruzada pelo método k-fold (STONE, 1974) que particiona a base de dados em k seções/dobras, sendo ajustados os modelos para k-1 seções e validado para a dobra restante. A rotina é executada novamente, considerando uma outra dobra para validação e os demais para calibração do modelo. O mesmo processo se repete até que todas as dobras da partição inicial sejam utilizadas no conjunto de validação (k vezes). Em cada uma das k repetições, são calculados índices estatísticos que descrevem o desempenho do ajuste do modelo, como o R², por exemplo, e ao final determina-se a média deles para representar o desempenho final do procedimento K-fold. Esse resultado é o que será utilizado para comparação de desempenho entre os modelos propostos no estudo.

Após a avaliação do modelo pelo método k-fold, ajusta-se o modelo novamente para a base de dados completa. Esse procedimento será adotado para todos os modelos previstos neste trabalho e as estimativas dos modelos individuais serão combinados linearmente para ajuste do modelo final, conforme antes comentado. Esse modelo final será então aplicado a base de dados selecionada para teste para avaliação de seu desempenho de estimativa (seção 5.4).

Todavia, os modelos individuais calibrados para a base de dados completa também serão avaliados separadamente na base de dados teste para investigar se o modelo combinado está efetivamente resultando em melhores estimativas. Caso o aprimoramento não seja estatisticamente significativo em relação ao desempenho dos modelos individuais, escolher-se-á apenas um modelo e será aquele que apresentar melhor resultado do procedimento k-fold.

5.3.3 Pedofunções da literatura

No estudo de levantamento de pedofunções de propriedades hidráulicas no Brasil realizado até o ano de 2019 pela autora deste trabalho, conforme já comentado, foram listadas todas as equações de ajuste das FPTs no Brasil, quando disponíveis, conforme exemplos registrados na Tabela 01 para dois trabalhos inventariados por esse estudo.

Tabela 01 – Exemplos de funções de pedotransferências de propriedades hidráulicas do solo de dois trabalhos da literatura.

referência	revista	ano	variável predita	variáveis predictoras	equação
Almeida et al. (2010)	V Congresso de Pesquisa e Inovação da Rede Norte Nordeste de Educação Tecnológica (CONNEPI)	2010	Capacidade de Campo, CC (theta-33kPa) (g/g)*100	areia (%), silte (%), argila (%)	CC=0,003(areia) ² -0,911(areia)+57,91
					CC=0,004(argila) ² -0,312(argila)+3,289
					CC=-0,030(silte) ² -1,462(silte)+1,987
			Ponto de murcha Permanente, PMP (theta-1500kPa) (g/g)*100		CC=-0,003(silte+argila) ² -0,180(silte+argila)+3,309
					PMP=129,0e ^{-0,04(areia)}
					PMP=0,379(argila) ^{0,905}
					PMP=1,384(silte) ^{0,654}
					PMP=0,213(silte+argila) ^{0,990}
Alvarenga et al. (2010)	XIX CONGRESSO DE PÓS-GRADUAÇÃO DA UFLA	2010	Capacidade de Campo, CC (theta-6kPa) (cm ³ /cm ³)	areia (%), silte (%), argila (%), densidade do dolo (g/cm ³), porosidade (cm ³ /cm ³)	CC = - 0,079A + 0,212ARG + 0,201S + 13,684Ds + 20,442P
				areia (%), silte (%), argila (%), densidade do dolo (g/cm ³)	CC = 0,119A + 0,410ARG + 0,396S + 6,242Ds
				areia (%), silte (%), argila (%)	CC = 0,208A + 0,474ARG + 0,456S
				areia (%), argila (%)	CC = 0,245A + 0,607ARG
			Ponto de Murcha Permanente, PMP (theta-1500kPa) (cm ³ /cm ³)	areia (%), silte (%), argila (%), densidade do dolo (g/cm ³), porosidade (cm ³ /cm ³)	PMP = - 0,124A + 0,156ARG + 0,060S + 10,811Ds + 5,517 P
				areia (%), silte (%), argila (%), densidade do dolo (g/cm ³)	PMP = - 0,071A + 0,210ARG + 0,113S + 8,803Ds
				areia (%), silte (%), argila (%)	PMP = 0,054A + 0,300ARG + 0,196S
				areia (%), silte (%), argila (%)	PMP = 0,070A + 0,358ARG
				areia (%), argila (%)	
				areia (%), argila (%)	

ARG - Argila, S - Silte, A - Areia, Ds - Densidade do Solo, P - porosidade

Algumas modificações nessas equações serão efetuadas para que as unidades das variáveis de entrada e saída estejam compatíveis com as das variáveis de solo da base de dados de solos deste estudo. Após isso, pretende-se aplicá-las à base de dados selecionados para teste e comparar seus resultados com os do modelo ajustado neste trabalho. Esse exercício tem como propósito investigar se o modelo atual proposto é de fato justificável, ou seja, se vale a pena o esforço em construí-la. Em não sendo, buscaremos investigar em quais circunstâncias (tipo de solo, classes texturais, etc) os modelos da literatura são melhores que os deste trabalho e se for o caso propor um algoritmo de predição que leve em consideração o universo mapeado de pedofunções no Brasil, incluindo as ajustadas neste estudo, e as restrições em que elas são mais adequadas nas predições.

5.4. Análise de desempenho das pedofunções

O desempenho dos modelos será acessado por meio dos seguintes índices:

$$RMSE \text{ (root mean squared error): } \sqrt{\frac{1}{n} \sum_{j=1}^n (y_j - \hat{y}_j)^2} \quad (05)$$

$$MAE \text{ (mean absolute error): } \frac{1}{n} \sum_{j=1}^n |y_j - \hat{y}_j| \quad (06)$$

$$ME \text{ (mean error): } \frac{1}{n} \sum_{j=1}^n y_j - \hat{y}_j \quad (07)$$

$$R2 \text{ (coeficiente de determinação): } 1 - \frac{\sum_{j=1}^n (y_j - \hat{y}_j)^2}{\sum_{j=1}^n (y_j - \bar{y}_j)^2} \quad (08)$$

Esses índices serão determinados para a base de dados teste completa e por grupos de solos definidos por classes texturais, pedológicas e de uso e cobertura desse conjunto de dados teste.

Associado ao resultado desses índices, pretende-se avaliar os gráficos dos erros de ajuste ($y_j - \hat{y}_j$) versus os valores observados das propriedades hidráulicas (variáveis de saída) e os dos preditores para averiguação se há alguma tendência dos erros de ajustes dos modelos em faixa de valores dos preditores e preditos. Gráficos dos valores observados e estimados pelas pedofunções também serão analisados.

5.5. Intervalo de Confiança

O texto abaixo foi extraído do livro Inteligência Computacional: Fundamentos e aplicações, pags. 101-102 (EVSUKOFF, 2020).

Se a estrutura selecionada pelo modelo representa adequadamente a relação entre o resultado do modelo e os valores observado da variável de saída, uma estimativa da variância dos resíduos é calculada como:

$$Var(e) = \hat{\sigma}^2 = \frac{1}{N-n} \sum_{t=1}^N e^2(t) = \frac{1}{N-n} \sum_{t=1}^N (y(t) - \hat{y}(t))^2 \quad (09)$$

onde n é o número de regressores do modelo.

Quando a distribuição do resíduo pode ser representada adequadamente por uma distribuição normal, é possível calcular um intervalo de confiança do resultado do modelo para um valor de entrada arbitrário $\hat{\mathbf{x}}_0$ do vetor de regressores. A estimativa da variância do modelo no ponto $\hat{\mathbf{x}}_0$ é calculada como:

$$Var(\hat{y}_0) = \hat{\sigma}^2 \hat{\mathbf{x}}_0^T (\hat{\mathbf{X}}^T \hat{\mathbf{X}})^{-1} \hat{\mathbf{x}}_0 \quad (10)$$

onde \hat{y}_0 é o valor do modelo no ponto $\hat{\mathbf{x}}_0$ e $\hat{\sigma}^2$ é a estimativa da variância do resíduo calculada pela equação 9.

O intervalo que contém o valor esperado da relação real $y_r = E(y_0)$, onde y_0 é o valor observado em \hat{x}_0 , com confiança estatística de $\xi\%$ é calculado como:

$$\hat{y}_0 - t_{\xi,d}\sqrt{\text{Var}(\hat{y}_0)} \leq y_r \leq \hat{y}_0 + t_{\xi,d}\sqrt{\text{Var}(\hat{y}_0)} \quad (11)$$

onde $d = N - n$ é o número de graus de liberdade e $t_{\xi,d}$ é o valor da distribuição t de Student. A distribuição t pode ser aproximada pela normal para $d \geq 30$, de forma que o intervalo de confiança com 95% de significância pode ser aproximado por:

$$\hat{y}_0 - 2.0\sqrt{\text{Var}(\hat{y}_0)} \leq y_r \leq \hat{y}_0 + 2.0\sqrt{\text{Var}(\hat{y}_0)} \quad (12)$$

5.6. Domínio dos modelos

A aplicação dos modelos ajustados é reprodutível quando a região do conjunto de dados utilizados para o ajuste, ou seja, o domínio do modelo, é capaz de cobrir a região do conjunto de dados que se quer obter as predições. Em caso dessa cobertura não ocorrer, a confiabilidade da predição deve ser baixa e caberá ao usuário utilizá-la ou não.

Uma forma simples de se checar a aplicabilidade desses modelos, segundo seus domínios, é rodar um algoritmo de clusterização de dados, como é o k-means (HARTIGAN, 1975), para a base de dados da calibração e checar se o novo ponto a se estimado está próximo a um dos centroides dos clusters propostos. Outros métodos para avaliação da cobertura dos pontos a serem estimados pelo o domínio dos modelos propostos neste estudo ainda se encontram em investigação.

6. RESULTADOS ESPERADOS

São esperados os seguintes resultados deste estudo:

- Bases de dados de solos para geração dos modelos de predição, após organização, harmonização, conferência e correção dos dados de solo;
- Modelos de predição nacionais de retenção de água para os seguintes potenciais matriciais: 0kPa, 6kPa, 10kPa, 33kPa e 1500kPa;
- Modelos de predição nacionais da condutividade hidráulica saturada;

- Modelos de predição dos parâmetros de equações que descrevem os dados de retenção de água;
- Resultados da validação dos modelos gerados neste trabalho e os compilados na literatura para uma base de dados teste;
- Resultados da validação funcional dos modelos gerados para diferentes estudos de casos geoambientais que envolvam simulações do movimento de água no solo;

7. PRODUTO ESPERADO

Espera-se produzir um aplicativo nos moldes do Qmin, desenvolvido pelo Serviço Geológico do Brasil ([Qmin \(cprm.gov.br\)](http://Qmin.cprm.gov.br)), onde o usuário insere sua base de dados de solos, e indica qual o modelo preditor quer utilizar, conforme relacionados na Figura 03, e como resultado final, será produzida uma outra base de dados com os resultados, incluindo os intervalos de confiança e se as estimativas encontram-se na área de domínio do modelo.

8. CRONOGRAMA DAS ATIVIDADES

Nesta seção descrevem-se as atividades a serem desenvolvidas no contexto deste estudo em seis tópicos temáticos e seus prazos de execução.

Tabela 02 – Relação das atividades do Projeto Estudo de Pedofunções de Propriedades Hidráulicas em Solos Brasileiros e prazos de execução no período de 2021 a 2026.

TOPICOS	Atividades Previstas	2021	2022	2023	2024	2025	2026
BASE DE DADOS	Compilação de dados de solo						
	Seleção de dados de solos internacionais, segundo metodologia da seção 5.1.1						
	Organização, harmonização e consistência dos dados, conforme seção 5.1.2						
	Análise dos dados, conforme seção 5.1.3						
	Composição das bases de dados do estudo, segundo as variáveis de saída e entrada a serem ajustados (item f da seção 5.1.3)						
AJUSTE DOS MODELOS PARA OS DADOS DE RETENÇÃO DE ÁGUA	Ajuste dos modelos lineares para os dados de retenção de água e teste de outras representações das variáveis de entrada, conforme seção 5.3, além daquelas previstas na seção 5.2. Uso da validação cruzada						
	Ajuste do modelo Random Forest para os dados de retenção e teste de outras representações das variáveis de entrada, conforme seção 5.3, além daquelas previstas na seção 5.2. Uso da validação cruzada						
	Ajuste de todos os modelos (linear, random forest, catboost e redes neurais) para as representações das variáveis de entrada testadas nos estudos anteriores						
	Análise dos resultados pelos scores do método de validação cruzada						
	Elaboração dos modelos combinados e						

Tabela 02 (continuação) – Relação das atividades do Projeto Estudo de Pedofunções de Propriedades Hidráulicas em Solos Brasileiros e prazos de execução no período de 2021 a 2026.

TOPICOS	Atividades Previstas	2021	2022	2023	2024	2025	2026
AJUSTE DOS MODELOS PARA OS DADOS DE RETENÇÃO DE ÁGUA	Seleção do modelo final de estimativa dos dados de retenção de água						
	Desenvolvimento do aplicativo para predições de dados de retenção de água em outras bases de dados de solos						
AJUSTE DOS MODELOS PARA OS PARÂMETROS DE VAN GENUCHTEN	Ajuste dos parâmetros da equação de van Geuncken para a base de dados de retenção de água						
	Ajustes dos modelos previstos no estudo						
	Análise dos resultados pelos scores do método de validação cruzada						
	Elaboração dos modelos combinados e comparação de desempenho com os dos modelos individuais para a base de dados teste						
	Definição do domínio do modelo						
	Seleção do modelo final de estimativa dos parâmetros						
	Aprimoramento do aplicativo para inclusão dos modelos de predição dos parâmetros da equação de van Genuchten						
AJUSTE DOS MODELOS PARA CONDUTIVIDADE HIDRÁULICA SATURADA	Ajustes dos modelos previstos no estudo						
	Análise dos resultados pelos scores do método de validação cruzada						
	Elaboração dos modelos combinados e comparação de desempenho com os dos modelos individuais para a base de dados teste						
	Definição do domínio do modelo						
	Seleção do modelo final de estimativa de Ksat						
	Aprimoramento do aplicativo para inclusão dos modelos de predição de Ksat						
VALIDAÇÃO FUNCIONAL	Escolha de cenários geoambientais que envolvam o armazenamento e fluxos de água no solo						
	Escolha do modelo para simulação dos processos hidrológicos no solo						
	Simulação dos cenários escolhidos, levando em consideração o uso das FPTs ajustadas no trabalho e aquelas usualmente utilizadas na literatura. Realizar comparação dos resultados obtidos pelos dois modelos de FPTs (ajustado neste trabalho e os da literatura)						
RELATÓRIO E TRABALHOS CIENTÍFICOS	Elaboração do relatório do projeto						
	Elaboração de relatórios parciais						
	Elaboração do relatório final						
	Elaboração de trabalhos científicos						

9. CONCLUSÕES

Este relatório apresenta a metodologia do Projeto institucional “Estudos de Pedofunções de Propriedades Hidráulicas em Solos Brasileiros”, bem como a relação das atividades a serem desenvolvidas e prazos de execução, resultados e produtos esperados. Este projeto é de médio

prazo com previsão para término em 2026, mas espera-se que esse trabalho seja continuamente revisto caso novos dados de solos encontrarem-se disponíveis. Este trabalho também pode ser estendido para predição de dados de condutividade hidráulica não saturada, quando essas informações forem acessíveis em um número mais significativo, como também para geração de modelos de predição de propriedades hidráulicas, as mesmas descritas neste trabalho, que levem em consideração variáveis morfológicas do solo e covariáveis ambientais produzidas a partir de modelos digitais de terrenos para permitir estimativas das propriedades hidráulicas em mais larga escala territorial.

10. REFERÊNCIAS BIBLIOGRÁFICAS

BOUMA, J. Using soil survey data for quantitative land evaluation. **Advances in Soil Science**, v. 9, p. 177-213, 1989. DOI: 10.1007/978-1-4612-3532-3_4.

EUROPEAN COMMISSION. **Lucas 2009 (Land Use / Cover Area Frame Survey)**. Technical reference, document C-3: Land use and Land Cover: Nomenclature (version of 20/03/2009), Eurostat.

EVSUKOFF, A. G. **Inteligência computacional: Fundamentos e aplicações**. 1. ed. Rio de Janeiro: E-papers, 2020. 526 p. Versão eletrônica. ISBN 978-65-8706-502-1.

GERKE, H. H.; VAN GENUCHTEN, M.T. H. A dual-porosity model for simulating the preferential movement of water and solutes in structured porous media. **Water Resources Research**, v. 29, n. 2, p. 305–319, 1993. DOI: <https://doi.org/10.1029/92WR02339>. Disponível em: https://www.ars.usda.gov/arsuserfiles/20360500/pdf_pubs/P1208.pdf. Acesso em: 10 jan. 2022.

GWO, J. P.; JARDINE, P. M.; WILSON, G. V.; YEH, G. T. A multiple-pore-region concept to modeling mass transfer in subsurface media. **Journal of Hydrology**, v. 164, n. 1-4, p. 217–237, 1995. DOI: [https://doi.org/10.1016/0022-1694\(94\)02555-P](https://doi.org/10.1016/0022-1694(94)02555-P).

HARTIGAN, J. A. **Clustering algorithms**. New York: John Wiley & Sons, 1975. 369 p. ISBN 978-0-471-35645-5. Disponível em: [https://people.inf.elte.hu/fekete/algorithmok_msc/klaszterezes/John%20A.%20Hartigan-Clustering%20Algorithms-John%20Wiley%20&%20Sons%20\(1975\).pdf](https://people.inf.elte.hu/fekete/algorithmok_msc/klaszterezes/John%20A.%20Hartigan-Clustering%20Algorithms-John%20Wiley%20&%20Sons%20(1975).pdf). Acesso em: 10 jan. 2022.

HO, T. K. Random Decision Forests. *In*: INTERNATIONAL CONFERENCE ON DOCUMENT ANALYSIS AND RECOGNITION, 3., 1995, Montreal. **Anais [...]**. Montreal: IEEE Computer Society, 1995. p. 278–282. ISBN: 978-08-1867-128-9.

JARVIS, N. J. Modeling the impact of preferential flow on nonpoint source pollution. *In*: SELIM, H. M.; MA, L. (eds.). **Physical Nonequilibrium in Soils: Modeling and Application**. Boca Raton: CRC Press, 1998. p. 195–221. ISBN 978-10-0307-609-4.

LIN, H. Earth's Critical Zone and hydrogeology: concepts, characteristics, and advances. **Hydrology and Earth System Sciences**, v. 14, n. 1, p. 25–45, 2010. DOI: <https://doi.org/10.5194/hess-14-25-2010>.

MAREN, A. J. Neural Network Structures: Form Follows Function. *In*: HARSTON, C. T.; PAP, R. M. (eds.). **Handbook of Neural Computing Applications**. San Diego (California): Academic Press, 1990. p. 45-57. *eBook* ISBN: 978-14-8326-484-4.

OTTONI, M. V. (org.). **HYBRAS, Hydrophysical Database for Brazilian Soils: Banco de dados hidrofísicos em solos no Brasil para o desenvolvimento de funções de pedotransferências de propriedades hidráulicas, versão 1.0**. Rio de Janeiro: CPRM, 2018. 27 f. Relatório parcial do projeto Estudos de Caracterização Hidrológica dos Solos. Disponível em: <https://rigeo.cprm.gov.br/jspui/handle/doc/19021>. Acesso em: 07 jan. 2022.

OTTONI, M. V. Pedotransfer functions of hydraulic properties for brazilian soils: a review: preliminary results. *In*: BRAZILIAN SOIL PHYSICS MEETING, 5., 2019, Lavras, Minas Gerais. **Apresentação oral** [...]. Lavras: BSPM, 2019. Disponível em: <https://rigeo.cprm.gov.br/handle/doc/22568> . Acesso em: 22 jan. 2022.

OTTONI, M. V. **Sistema de Classificação dos Solos baseado na Estrutura do Espaço Poroso**. 2017. 206 f. Tese (Doutorado em Engenharia Civil) – Instituto Alberto Luiz Coimbra, Universidade Federal do Rio de Janeiro – UFRJ, Rio de Janeiro, 2017. Disponível em: <https://rigeo.cprm.gov.br/jspui/bitstream/doc/19000/1/Tese%20de%20Marta%20V.%20Ottoni.pdf> . Acesso em: 07 jan. 2021.

OTTONI, M. V.; OTTONI FILHO, T. B.; SCHAAP, M. G.; LOPES-ASSAD, M. L. R. C.; ROTUNNO FILHO, O. C. Hydrophysical database for Brazilian soils (HYBRAS) and Pedotransfer Functions for water retention. **Vadose Zone Journal**, v. 17, n. 1, 2018. DOI: <https://doi.org/10.2136/vzj2017.05.0095>.

OTTONI, M. V.; OTTONI FILHO, T. B.; SCHAAP, M. G.; LOPES-ASSAD, M. L. R. C.; ROTUNNO FILHO, O. C. Pedotransfer functions for saturated hydraulic conductivity using a database with temperate and tropical climate soils. **Journal of Hydrology**, v. 575, p. 1345-1358, 2019. DOI: <https://doi.org/10.1016/j.jhydrol.2019.05.050> .

PROKHORENKOVA, L.; GUSEV, G.; VOROBIEV, A.; DOROGUSH, A. V.; GULIN, A. **CatBoost: unbiased boosting with categorical features**. *In*: CONFERENCE ON NEURAL INFORMATION PROCESSING SYSTEMS (NEURIPS 2018), 32., Montréal, Canadá, 2018. **Anais** [...]. Montreal: NeurIPS, 2018. p. 6639 – 6649. Disponível em: https://www.researchgate.net/publication/328576065_CatBoost_gradient_boosting_with_categorical_features_support. Acesso em: 10 jan. 2022.

PRUESS, K.; WANG, J. S. Y. Numerical modeling of isothermal and non-isothermal flow in unsaturated fractured rock- a review. *In*: EVANS, D. D.; NICHOLSON, T. J. (eds.). **Flow and Transport through Unsaturated Fractured Rock**. Washington, DC: American Geophysical Union, 1987. p. 11–22. (Geophysics Monograph, 42). DOI: <https://doi.org/10.1029/GM042p0019>.

RAWLS, W. J.; BRAKENSIEK, D. L. Estimating soil water retention from soil properties. **Journal of the Irrigation and Drainage Division**, v. 108, p. 166–171, 1982.

RIBEIRO, B. T.; COSTA, A. M. DA; SILVA, B. M.; FRANCO, F. O.; BORGES, C. S. Assessing pedotransfer functions to estimate the soil water retention. **Bioscience Journal**, Uberlândia, v. 34, supplement 1, p. 177-188, 2018. DOI: <https://doi.org/10.14393/BJ-v34n6a2018-42371>.

SCHAAP, M. G.; LEIJ, F. J.; VAN GENUCHTEN, M. T. H. ROSSETA: computer program for estimating soil hydraulic parameters with hierarchical pedotransfer functions. **Journal of Hydrology**, v. 251, n. 3/4, p. 163-176, 2001. DOI: [https://doi.org/10.1016/S0022-1694\(01\)00466-8](https://doi.org/10.1016/S0022-1694(01)00466-8).

STONE, M. Cross-Validatory Choice and Assessment of Statistical Predictions. **Journal of the Royal Statistical Society: Series B (Methodological)**, v. 36, n. 2, p. 111–147, 1974. DOI: <https://doi.org/10.1111/j.2517-6161.1974.tb00994.x>.

SZABÓ, B.; WEYNANTS, M.; E WEBER, T. K. D. Updated European hydraulic pedotransfer functions with communicated uncertainties in the predicted variables (euptfv2). **Geoscientific Model Development**, v. 14, n. 1, p. 151–175, 2021. DOI: <https://doi.org/10.5194/gmd-14-151-2021>.

TOMASELLA, J.; HODNETT, M. G.; ROSSATO, L. Pedotransfer functions for the estimation of soil water retention in Brazilian soils. **Soil Science Society of America Journal**, v. 64, n. 1, p. 327-338, 2000. DOI: <https://doi.org/10.2136/sssaj2000.641327x>.

TOMASELLA, J.; PACHEPSKY, Y. A.; CRESTANA, S.; RAWLS, W. J. Comparison of two techniques to develop pedotransfer functions for water retention. **Soil Science Society of America Journal**, v. 67, n. 4, p. 1085-1092, 2003. DOI: <https://doi.org/10.2136/sssaj2003.1085>.

VAN GENUCHTEN, M. Th. A closed form equation for predicting the hydraulic conductivity of unsaturated soils. **Soil Science Society of America Journal**, v. 44, n. 5, p. 892-989, 1980. DOI: <https://doi.org/10.2136/sssaj1980.03615995004400050002x>.

ZHAO, H.; ZENG, Y.; LV, S.; SU, Z. Analysis of soil hydraulic and thermal properties for land surface modeling over the Tibetan Plateau. **Earth System Science Data**, v. 10, p. 1031–1061, 2018. DOI: <https://doi.org/10.5194/essd-10-1031-2018> .