

Universidade Federal do Espírito Santo  
Departamento de Engenharia Civil  
Programa de Pós-Graduação em Engenharia Civil

Dissertação de mestrado

**ANÁLISE COMPARATIVA DA ESTIMATIVA DO COEFICIENTE DE  
PERMEABILIDADE DE SOLOS POR REDES NEURAIS ARTIFICIAIS E  
MÉTODOS ESTATÍSTICOS**

**Aluna:** Morgana Moreschi

**Orientadora:** Prof<sup>a</sup>. Dra. Kátia Vanessa Bicalho

**Coorientador:** Prof<sup>o</sup> Dr. Wilian Hiroshi Hisatugu

Vitória  
2023

Dissertação de Mestrado

**ANÁLISE COMPARATIVA DA ESTIMATIVA DO COEFICIENTE DE  
PERMEABILIDADE DE SOLOS POR REDES NEURAIAS ARTIFICIAIS E  
MÉTODOS ESTATÍSTICOS**

Dissertação de mestrado a ser apresentada ao corpo docente do Programa de Pós-Graduação em Engenharia Civil da Universidade Federal do Espírito Santo para obtenção do título de Mestre em Engenharia Civil.

Orientadora: Prof<sup>a</sup>. Dra. Kátia Vanessa Bicalho

Coorientador: Prof<sup>o</sup> Dr. Wilian Hiroshi Hisatugu

Vitória

2023

Ficha catalográfica disponibilizada pelo Sistema Integrado de Bibliotecas - SIBI/UFES e elaborada pelo autor

---

M835a Moreschi, Morgana, 1989-  
Análise comparativa da estimativa do coeficiente de permeabilidade de solos por redes neurais artificiais e métodos estatísticos / Morgana Moreschi. - 2023.  
198 f. : il.

Orientadora: Kátia Vanessa Bicalho.  
Coorientador: Wilian Hiroshi Hisatugu.  
Dissertação (Mestrado em Engenharia Civil) - Universidade Federal do Espírito Santo, Centro Tecnológico.

1. Solos - Permeabilidade. 2. Mecânica do solo. 3. Redes neurais (Computação). 4. Correlação (Estatística). I. Bicalho, Kátia Vanessa. II. Hisatugu, Wilian Hiroshi. III. Universidade Federal do Espírito Santo. Centro Tecnológico. IV. Título.

CDU: 624

---

# UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO

## ANÁLISE COMPARATIVA DA ESTIMATIVA DO COEFICIENTE DE PERMEABILIDADE DE SOLOS POR REDES NEURAIAS ARTIFICIAIS E MÉTODOS ESTATÍSTICOS

**Morgana Moreschi**

Dissertação apresentada ao Curso de Mestrado em Engenharia Civil do Programa de Pós-Graduação em Engenharia Civil da Universidade Federal do Espírito, como requisito parcial para obtenção do título de Mestre em Engenharia Civil, área de Estruturas.

Aprovada no dia **17 de fevereiro de 2023** por:

---

Profa. Dra. Kátia Vanessa Bicalho  
Doutora em Engenharia Civil  
Orientadora – UFES

---

Prof. Dr. Wilian Hiroshi Hisatugu  
Doutor em Engenharia Eletrica  
Coorientador - UFES

---

Prof. Dr. Élcio Cassimiro Alves  
Doutor em Engenharia e Tecnologia Espaciais  
Examinador Interno – UFES

---

Prof. Dr. Silvio Romero de Melo Ferreira  
Doutor em Engenharia Civil  
Examinador Externo - UFPE

---

Prof. Dr. Celso Romanel  
Doutor em Engenharia Civil  
Examinador Externo - PUC-RIO

**Vitória – ES, fevereiro de 2023**



## **AGRADECIMENTOS**

A Deus, em primeiro lugar, por me acompanhar a cada passo dado e por me fortalecer, por tantas vezes, ao longo de toda essa trajetória.

À Universidade Federal do Espírito Santo - UFES e ao Programa de Pós-Graduação em Engenharia Civil - PPGEC por essa valiosa oportunidade de crescimento e por toda aprendizagem proporcionada.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pelo apoio à expansão e consolidação do PPGEC e à divulgação e fomento das produções científicas.

À professora Kátia Vanessa Bicalho, uma professora e pesquisadora com tanta dedicação nas orientações e instruções para a elaboração desta pesquisa. Ao professor Wilian, por todas as orientações e auxílios no estudo das redes neurais e por sua sempre disposição em orientar. Agradeço a vocês, meus orientadores, por cada tempo dedicado para que esse trabalho chegasse até aqui e por toda compreensão ao longo desse caminho.

A todos os demais professores, pelo dom de ensinar, inspirar e compartilhar sem ressalvas os seus conhecimentos. Aos funcionários da Universidade, por toda atenção e trabalho dedicados. Aos colegas de curso, pela parceria e apoio durante todo esse período.

Aos meus pais, que são meus alicerces, agradeço por todo amor e por cada esforço e suor que derramaram para que eu pudesse chegar até aqui. À minha irmã, meu exemplo de disciplina e dedicação.

Aos meus amigos, que sempre compreenderam as minhas ausências e me incentivaram a prosseguir até o fim.

A todos que, de alguma forma, estiveram presentes nessa jornada e contribuíram para a realização de mais essa etapa da minha vida e celebraram essa vitória ao meu lado, os meus sinceros agradecimentos.

## SUMÁRIO

LISTA DE FIGURAS.....	8
LISTA DE TABELAS .....	12
LISTA DE GRÁFICOS.....	15
LISTA DE SÍMBOLOS.....	17
RESUMO.....	19
1 INTRODUÇÃO .....	20
1.1 Objetivos .....	21
1.1.1 Objetivo geral .....	22
1.1.2 Objetivos específicos.....	22
1.2 Justificativa.....	22
1.3 Estrutura da dissertação .....	23
2 REVISÃO BIBLIOGRÁFICA.....	25
2.1 Lei de Darcy e o fluxo de água líquida no solo .....	25
2.2 Fatores que afetam o $k_{sat}$ .....	26
2.2.1 Influência da temperatura e do fluido percolante.....	27
2.2.2 Influência dos finos .....	28
2.2.2.1 Efeito dos finos no índice de vazios .....	30
2.2.2.2 Efeitos dos finos na distribuição granulométrica e no $d_{10}$ .....	33
2.2.2.3 Efeito da plasticidade do fino no $k_{sat}$ .....	35
2.2.3 Conclusões parciais do item 2.2.....	39
2.3 Equações empíricas para determinação do $k_{sat}$ .....	41
2.4 Desenvolvimento de métodos para previsão de $k_{sat}$ .....	49

2.4.1	Métodos estatísticos .....	49
2.4.1.1	Etapas da criação de uma correlação .....	49
2.4.1.2	Regressão linear simples e múltipla.....	52
2.4.1.3	Avaliação da regressão .....	54
2.4.2	Redes neurais artificiais.....	61
2.4.3	Aplicações de inteligência artificial na previsão de propriedades hidráulicas de solos.....	67
3	MATERIAIS E MÉTODOS .....	76
3.1	Etapa de revisão de literatura .....	78
3.2	Etapa do tratamento dos dados.....	78
3.3	Correlações empíricas.....	80
3.4	Análises estatísticas .....	80
3.5	Análises de redes neurais artificiais.....	82
3.6	Avaliação e comparação entre resultados.....	86
4	BANCOS DE DADOS.....	87
4.1	Bancos de dados obtidos da literatura .....	87
4.2	Dados disponibilizados em Wang e Huang (1984).....	91
5	RESULTADOS E DISCUSSÕES.....	102
5.1	Pré-tratamento dos dados experimentais .....	102
5.2	Resultados da Análise 1 .....	105
5.2.1	Resultados das análises estatísticas.....	106
5.2.2	Resultados das redes neurais .....	110
5.3	Análise 2.....	113

5.3.1	Resultados das análises estatísticas.....	113
5.3.2	Resultados das redes neurais .....	117
5.4	Análise 3.....	118
5.4.1	Resultados das análises estatísticas.....	118
5.4.2	Resultados das redes neurais .....	122
5.5	Análise 4.....	123
5.5.1	Resultados das análises estatísticas.....	124
5.5.2	Resultados das redes neurais .....	127
5.6	Análise 5.....	128
5.6.1	Resultados das análises estatísticas.....	128
5.6.2	Resultados das redes neurais .....	132
5.7	Conclusões parciais das equações de regressão e redes neurais.....	133
5.8	Resultados de correlações empíricas e semiempíricas.....	134
5.9	Análise comparativa dos resultados das regressões, RNA e correlações empíricas .....	138
6	CONSIDERAÇÕES FINAIS E SUGESTÕES .....	141
6.1	Sugestões para trabalhos futuros.....	144
	REFERÊNCIAS BIBLIOGRÁFICAS .....	145
	ANEXO 1 .....	151
	APÊNDICE 1.....	152
	APÊNDICE 2.....	175
	APÊNDICE 3.....	187
	APÊNDICE 4.....	189

APÊNDICE 5.....	190
APÊNDICE 6.....	194

## LISTA DE FIGURAS

Figura 1 – Relação entre velocidade de descarga e gradiente de carga hidráulica..	26
Figura 2 - Variação do (A) Índice de vazios, e correspondente (B) $k_{sat}$ em função do teor de finos em misturas de solos grossos e finos (Dados de: BELKHATIR <i>et al.</i> , 2014; TAIBA <i>et al.</i> , 2022).	32
Figura 3 - Distribuição granulométrica de misturas com diferentes teores de finos para misturas de areia e silte (Dados de: BELKHATIR <i>et al.</i> , 2014; ALDAOOD, 2020).	34
Figura 4 - Efeito da adição de finos com diferentes plasticidades no $k_{sat}$ resultante de misturas com solos granulares (Dados de: BARBER; SAWYER, 1952)	37
Figura 5 – Fluxograma para escolha da equação empírica para estimativa de $k_{sat}$ ..	48
Figura 6 – Exemplo de histograma	51
Figura 7 – Exemplo de diagrama de caixa	51
Figura 8 – Padrões do diagrama de resíduos	57
Figura 9 – Gráfico de probabilidade dos resíduos	59
Figura 10 - Modelo esquemático do <i>perceptron</i> de Rosenblatt (1958)	62
Figura 11 - Esquema de rede multicamadas	64
Figura 12 – Fluxograma das etapas da pesquisa	77
Figura 13 – Classificação relativa de permeabilidade	79
Figura 14 – Triângulo textural – Distribuição de 42 amostras, com teor de pedregulho igual a zero, dos dados de Wang e Huang (1984)	93
Figura 15 – Triângulo textural – 56 amostras de Benson <i>et al.</i> (1984)	153
Figura 16 – Triângulo textural – 419 amostras de Ottoni <i>et al.</i> (2018)	156
Figura 17 – Triângulo textural – amostras de 172 amostras de Nemes <i>et al.</i> (2001)	160

Figura 18 – Triângulo textural – 196 amostras de Rahmati <i>et al.</i> (2018) .....	163
Figura 19 – Triângulo textural – 7227 amostras de Gupta <i>et al.</i> (2020) .....	168
Figura 20 – Triângulo textural – 42 amostras de Pucket <i>et al.</i> (1985) .....	172
Figura 21 –Histograma do (A) %argila, (B) %silte, (C) %areia e (D) %pedregulho das amostras não plásticas (NP) – 30 amostras .....	175
Figura 22 –Histograma do (A) %finos, (B) $d_{10}$ , (C) $d_{50}$ e (D) $d_{60}$ das amostras não plásticas (NP) – 30 amostras .....	176
Figura 23 –Histograma do (A) $C_u$ , (B) $e$ , (C) $G_s$ e (D) $\rho_d$ das amostras não plásticas (NP) – 30 amostras .....	177
Figura 24 –Histograma do $\log(k_{sat})$ das amostras não plásticas (NP) – 30 amostras .....	178
Figura 25 –Histograma do (A) %argila, (B) %silte, (C) %areia e (D) %pedregulho das amostras plásticas – 81 amostras .....	179
Figura 26 – Histograma do (A) %finos, (B) $d_{10}$ , (C) $d_{50}$ e (D) $d_{60}$ das amostras plásticas – 81 amostras .....	180
Figura 27 – Histograma do (A) $C_u$ , (B) $e$ , (C) $G_s$ e (D) $\rho_d$ das amostras plásticas – 45 amostras .....	181
Figura 28 – Histograma do (A) LL, (B) LP, (C) IP e (D) $\log(k_{sat})$ das amostras plásticas – 81 amostras .....	182
Figura 29 – Diagramas de caixa do (A) %argila, (B) %silte, (C) %areia, (D) %pedregulho, (E) %finos, (F) $d_{10}$ , (G) $d_{50}$ , (H) $d_{60}$ e (I) $C_u$ das amostras não plásticas (NP) – 30 amostras .....	183
Figura 30 – Diagramas de caixa do (A) $e$ , (B) $G_s$ , (C) $\rho_d$ , (D) $\log(k_{sat})$ das amostras não plásticas (NP) – 30 amostras .....	184
Figura 31 – Diagramas de caixa do (A) %argila, (B) %silte, (C) %areia, (D) %pedregulho, (E) %finos, (F) $d_{10}$ , (G) $d_{50}$ , (H) $d_{60}$ e (I) $C_u$ das amostras plásticas – 81 amostras .....	185

Figura 32 – Diagramas de caixa do (A) $e$ , (B) $G_s$ , (C) $\rho_d$ , (D) LL, (E) LP, (F) IP e (G) $\log(k_{sat})$ das amostras plásticas – 45 amostras .....	186
Figura 33 – Análise 1 – Diagnóstico de resíduos – RL1 .....	194
Figura 34 - Análise 1 – Diagnóstico de resíduos – RNL1 .....	194
Figura 35 - Análise 2 – Diagnóstico de resíduos – RL2 .....	194
Figura 36 - Análise 2 – Diagnóstico de resíduos – RNL2 .....	194
Figura 37 - Análise 3 – Diagnóstico de resíduos – RL3 .....	195
Figura 38 - Análise 3 – Diagnóstico de resíduos – RNL3 .....	195
Figura 39 - Análise 4 – Diagnóstico de resíduos – RL4 .....	195
Figura 40 - Análise 4 – Diagnóstico de resíduos – RNL4 .....	195
Figura 41 - Análise 5 – Diagnóstico de resíduos – RL5 .....	196
Figura 42 - Análise 5 – Diagnóstico de resíduos – RNL5 .....	196

## LISTA DE QUADROS

Quadro 1 - Equações matemáticas para estimativa de $k_{sat}$ .....	43
Quadro 2 – Tabela de decisão em função de $d_u$ e $d_L$ .....	58
Quadro 3 - Resumo das funções de ativação .....	63
Quadro 4 – Uso de inteligência artificial e métodos estatísticos para estimativa de valores de $k_{sat}$ .....	73

## LISTA DE TABELAS

Tabela 1 – Dados experimentais - Resumo das arquiteturas de RNA treinadas.....	84
Tabela 2 – Número de neurônios por camada intermediária .....	85
Tabela 3 – Sumarização dos bancos de dados (N/I = não informado).....	88
Tabela 4 - Classificação relativa da permeabilidade das amostras de Wang e Huang (1984).....	92
Tabela 5 - Permeabilidade de solos grossos e finos puros (Dados de: WANG; HUANG, 1984) .....	95
Tabela 6 - Permeabilidade misturas de solos com 20% de finos (Dados de: WANG; HUANG, 1984) .....	97
Tabela 7 - Variação de $k_{sat}$ com a adição de finos em matrizes granulares de diferentes percentuais de areia e pedregulho para amostras compactadas com energia 90% do Proctor Normal (Dados de: WANG; HUANG, 1984) .....	97
Tabela 8 – Estatística descritiva das amostras NP dos dados selecionados de Wang e Huang (1984) – 30 amostras .....	103
Tabela 9 - Estatística descritiva das amostras plásticas dos dados selecionados de Wang e Huang (1984) – 81 amostras.....	104
Tabela 10 - Análise 1 – Resumo das equações de regressão propostas .....	106
Tabela 11 – Resumo dos resultados dos testes estatísticos para a equação RL1 .	107
Tabela 12 - Resumo dos resultados dos testes estatísticos para a equação RNL1	109
Tabela 13 – Análise 1 – Resumo das arquiteturas de RNA treinadas.....	110
Tabela 13 – Análise 1 – Resumo das arquiteturas de RNA treinadas – cont.....	111
Tabela 14 - Análise 2 – Resumo das equações de regressão propostas .....	113
Tabela 15 – Análise 2 - Resumo dos resultados dos testes estatísticos para a equação RL2.....	115

Tabela 16 - Análise 2 - Resumo dos resultados dos testes estatísticos para a equação RNL2.....	116
Tabela 17 - Análise 2 – Resumo das arquiteturas de RNA treinadas.....	117
Tabela 18 - Análise 3 – Resumo das equações de regressão propostas .....	119
Tabela 19 - Análise 3 - Resumo dos resultados dos testes estatísticos para a equação RL3.....	120
Tabela 20 - Análise 3 - Resumo dos resultados dos testes estatísticos para a equação RNL3.....	121
Tabela 21 - Análise 3 – Resumo das arquiteturas de RNA treinadas.....	122
Tabela 22 - Análise 4 – Resumo das equações de regressão propostos .....	124
Tabela 23 - Análise 4 - Resumo dos resultados dos testes estatísticos para a equação RL4.....	125
Tabela 24 - Análise 4 - Resumo dos resultados dos testes estatísticos para a equação RNL4.....	126
Tabela 25 - Análise 4 – Resumo das arquiteturas de RNA treinadas.....	127
Tabela 26 - Análise 5 – Resumo das equações de regressão propostas .....	129
Tabela 27 - Análise 5 - Resumo dos resultados dos testes estatísticos para a equação RL5.....	130
Tabela 28 - Análise 5 - Resumo dos resultados dos testes estatísticos para a equação RNL5.....	131
Tabela 29 - Análise 5 – Resumo das arquiteturas de RNA treinadas.....	132
Tabela 30 - Equações de predição de $k_{sat}$ selecionadas para o estudo comparativo .....	136
Tabela 31 – Resultados de correlações selecionadas da literatura aplicadas ao conjunto de 111 dados experimentais.....	136

Tabela 32 - Classificação relativa da permeabilidade das 56 amostras de Benson <i>et al.</i> (1984) .....	153
Tabela 33 – Classificação relativa da permeabilidade das amostras de Ottoni <i>et al.</i> (2018).....	157
Tabela 34 - Classificação relativa da permeabilidade das amostras de Nemes <i>et al.</i> (2001).....	160
Tabela 35 – Classificação relativa da permeabilidade das amostras de Rahmati <i>et al.</i> (2018).....	163
Tabela 36 - Valores de $k_{sat}$ para as amostras com maiores e menores teores de finos (Dados de: RAHMATI <i>et al.</i> , 2018) .....	165
Tabela 37 - Classificação relativa da permeabilidade das 32 amostras de Alyamani e Sen (1993).....	166
Tabela 38 - Classificação relativa da permeabilidade das amostras de Gupta <i>et al.</i> (2020).....	167
Tabela 39 - Permeabilidade de amostras menores $k_{sat}$ e de amostras com maiores teores de finos (Dados de: GUPTA <i>et al.</i> , 2020) .....	170
Tabela 40 - Classificação relativa da permeabilidade das amostras de Pucket <i>et al.</i> (1985).....	171

## LISTA DE GRÁFICOS

Gráfico 1 - Variação de $k_{sat}$ com o teor de finos para 111 diferentes misturas controladas (amostras compactadas) (Dados de: BARBER; SAWYER, 1952; WANG; HUANG, 1984) .....	94
Gráfico 2 - Variação de $k_{sat}$ com a adição de finos em matrizes granulares com diferentes percentuais de areia e pedregulho para amostras compactadas com energia 90% do Proctor Normal (Dados de: WANG; HUANG, 1984) .....	99
Gráfico 3 - Ábaco de Casagrande para as 81 amostras com plasticidade (Dados de: WANG; HUANG, 1984).....	102
Gráfico 4 – Análise 1 – Ajuste dos valores medidos e previstos de $\log(k_{sat})$ .....	106
Gráfico 5 - Análise 1 - Resultados estatísticos das RNA (fase de validação cruzada) e das equações de regressão propostas .....	112
Gráfico 6 - Análise 2 – Valores medidos e previstos de $\log(k_{sat})$ .....	114
Gráfico 7 - Análise 2 - Resultados estatísticos das RNA (fase de validação cruzada) e das equações de regressão propostas .....	118
Gráfico 8 - Análise 3 – Valores medidos e previstos de $\log(k_{sat})$ .....	119
Gráfico 9 - Análise 3 - Resultados estatísticos das RNA (fase de validação cruzada) e das equações de regressão propostasFonte: Autora (2022). .....	123
Gráfico 10 - Análise 4 – Valores medidos e previstos de $\log(k_{sat})$ .....	124
Gráfico 11 - Análise 4 - Resultados estatísticos das RNA (fase de validação cruzada) e das equações de regressão propostas .....	128
Gráfico 12 - Análise 5 – Valores medidos e previstos de $\log(k_{sat})$ .....	129
Gráfico 13 - Análise 5 - Resultados estatísticos das RNA (fase de validação cruzada) e das equações de regressão propostas .....	133
Gráfico 14 - Valores medidos e previstos de $\log(k_{sat})$ pelas correlações selecionadas da literatura .....	137

Gráfico 15 – Comparação de R <sup>2</sup> , RMSE e MAE entre as equações de melhor desempenho das regressões e redes neurais .....	139
Gráfico 16 - Variação de k <sub>sat</sub> com o teor de finos para diferentes misturas controladas (Dados de: BARBER; SAWYER, 1952; BENSON <i>et al.</i> , 1984) .....	154
Gráfico 17 - Variação de k <sub>sat</sub> com o teor de finos para as 419 amostras de diferentes solos de Ottoni <i>et al.</i> (2018) (Dados de: BARBER; SAWYER, 1952; OTTONI <i>et al.</i> , 2018) .....	158
Gráfico 18 - Variação de k <sub>sat</sub> com índice de vazios para diferentes solos (Dados de: OTTONI <i>et al.</i> , 2018) .....	159
Gráfico 19 - Variação de k <sub>sat</sub> com o teor de finos para 172 amostras de diferentes solos de Nemes <i>et al.</i> (2001) (Dados de: BARBER; SAWYER, 1952; NEMES <i>et al.</i> , 2001) .....	161
Gráfico 20 - Variação de k <sub>sat</sub> com o teor de finos para 196 amostras de diferentes solos de Rahmati <i>et al.</i> (2018) (Dados de: BARBER; SAWYER, 1952; RAHMATI <i>et al.</i> , 2018) .....	164
Gráfico 21 - Variação de k <sub>sat</sub> com o teor de finos para diferentes misturas controladas (Dados de: BARBER; SAWYER, 1952; GUPTA <i>et al.</i> , 2020) .....	169
Gráfico 22 - Variação de k <sub>sat</sub> com o teor de finos para diferentes solos naturais (Dados de: BARBER; SAWYER, 1952; PUCKET <i>et al.</i> , 1985) .....	173
Gráfico 23 - Variação de k <sub>sat</sub> com índice de vazios para diferentes solos naturais (Dados de: PUCKET <i>et al.</i> , 1985) .....	174

## LISTA DE SÍMBOLOS

$k_{sat}$  - coeficiente de permeabilidade do solo saturado

$k$  - condutividade hidráulica do solo não saturado

$R^2$  - coeficiente de determinação

$R^2_{ajust}$  – coeficiente de determinação ajustado

RMSE – raiz do erro quadrático médio

MAE – erro médio absoluto

$Q$  – vazão de fluxo que atravessa a amostra de solo

$\nabla h$  – gradiente hidráulico

$A$  – área da seção transversal da amostra de solo, atravessada pelo fluxo

$\Delta h$  – diferença de carga hidráulica total a montante e a jusante da amostra de solo

$L$  – comprimento da amostra na direção do fluxo

$K$  – permeabilidade intrínseca

$\mu$  – viscosidade do fluido percolante

$\rho$  – massa específica do fluido percolante

$g$  – aceleração da gravidade

%argila – percentual de argila

%silte – percentual de silte

%areia – percentual de areia

%pedregulho – percentual de pedregulho

%finos – percentual de finos

$G_s$  – densidade real dos grãos

LL – limite de liquidez

LP – limite de plasticidade

IP – índice de plasticidade

OC(%) – percentual de carbono orgânico

%MO – percentual de matéria orgânica

pH – potencial hidrogeniônico

$d_{10}$  – diâmetro tal que 10% do solo, em peso, tem diâmetros menores que ele (mm).  
(diâmetro efetivo)

$d_{50}$  – diâmetro tal que 50% do solo, em peso, tem diâmetros menores que ele (mm).  
(diâmetro médio das partículas)

$d_{60}$  – diâmetro tal que 60% do solo, em peso, tem diâmetros menores que ele (mm)

$C_u$  – coeficiente de não uniformidade do solo

$C_c$  – coeficiente de curvatura

$e$  – índice de vazios global

$D_r$  – compacidade relativa

$n$  - porosidade

$w$  – umidade

$\theta_s$  – umidade na saturação

$\rho_d$  – massa específica aparente seca

$\rho_s$  – massa específica real dos grãos

$\rho_t$  – massa específica total

$\gamma_d$  - peso específico aparente seco

## RESUMO

Esta pesquisa analisa e compara o desempenho de redes neurais artificiais (RNA), métodos estatísticos e correlações empíricas e semiempíricas na previsão de valores do coeficiente de permeabilidade de solos saturados ( $k_{sat}$ ), a partir de propriedades índices que caracterizam a distribuição granulométrica e a fração fina. Foram compilados e comparados 8258 dados experimentais de  $k_{sat}$  de solos compostos por grãos grossos e finos ( $2,50 \times 10^{-13} \leq k_{sat} \text{ (m/s)} \leq 4,50 \times 10^{-2}$ ), publicados em 08 bancos de dados na literatura, com o intuito de auxiliar na compreensão das propriedades hidráulicas de solos saturados e na descrição de problemas de predição de  $k_{sat}$ . Posteriormente, um conjunto de amostras foi selecionado e analisado com uma combinação de diferentes variáveis de entrada para previsão do  $\log(k_{sat})$ , utilizando regressão linear e polinomial múltipla e RNA. As variáveis de entrada consideradas foram percentual de finos (silte e argila) (%Finos), limite de liquidez (LL), diâmetro efetivo ( $d_{10}$ ), coeficiente de uniformidade ( $C_u$ ) e índice de vazios ( $e$ ). Os resultados foram avaliados a partir dos valores do coeficiente de determinação, da raiz do erro quadrático médio e do erro médio absoluto. O desempenho das RNA superou as regressões e as correlações da literatura. De todos os resultados das análises realizadas, a RNA302, que considerou como variáveis independentes o %Finos, LL,  $C_u$  e  $d_{10}$ , apresentou, numericamente, os melhores resultados. O acréscimo de uma terceira camada oculta reduziu a acurácia das redes. As regressões e as RNA foram melhores que as correlações empíricas para a previsão de  $k_{sat}$ , para a base de dados investigada, e mostraram que a escolha das variáveis que caracterizam a distribuição granulométrica e a fração fina foi satisfatória para a base de dados experimentais. Considerando que  $k_{sat}$  é uma propriedade altamente variável e função de diversas propriedades interdependentes, a técnica de RNA mostrou-se viável, principalmente, por não exigir o conhecimento prévio da relação matemática entre as variáveis e por sua capacidade de descrever problemas de maior complexidade. Destaca-se a importância de se incluir informações sobre a granulometria e a natureza dos finos em bancos de dados de  $k_{sat}$ , principalmente, para caracterização da permeabilidade de amostras com propriedades hidráulicas dominadas pela fração fina.

**Palavras-chaves:** Permeabilidade. Mecânica dos Solos. Redes neurais artificiais. Métodos estatísticos. Solo saturado.

## 1 INTRODUÇÃO

As propriedades hidráulicas de um solo afetam diretamente várias situações em campo que envolvem análises de fluxo, adensamento e estabilidade de obras de engenharia, e são importantes para os estudos de infiltração no solo, escoamento superficial, dispersão de poluentes subterrâneos, extração de águas de aquíferos, drenagem e controle de nível d'água freático, processo de adensamento, sistemas de irrigação, aplicação de fertilizantes, crescimento de vegetação etc. (SINGH *et al.*, 2020; KASHANI *et al.*, 2020; GOOTMAN; KELLNER; HUBBART, 2020).

Várias equações têm sido propostas que relacionam permeabilidade e propriedades do solo. O principal parâmetro que caracteriza a permeabilidade de um solo, quando saturado, é o chamado coeficiente de permeabilidade ( $k_{sat}$ ), geralmente expresso em m/s, que caracteriza a capacidade de um meio poroso de permitir a passagem de fluidos pelos poros interconectados (TAYLOR, 1948; LAMBE; WHITMAN, 1969; MITCHELL; SOGA, 2005; GOOTMAN; KELLNER; HUBBART, 2020; SHAKER *et al.*, 2022).

O valor de  $k_{sat}$  pode ser determinado por meio de ensaios em campo ou conduzidos em laboratório (MASROURI; BICALHO; KAWAI, 2008; CHAPUIS, 2012). Na ausência ou limitação de resultados de ensaios para determinação de  $k_{sat}$ , o comportamento hidráulico do solo saturado pode ser estimado por métodos indiretos, em função de propriedades índices das amostras desses solos (PARK, 2011; YILMAZ *et al.*, 2011; CHAPUIS, 2012; ROGIERS *et al.*, 2012; TIZPA *et al.*, 2014; SINGH *et al.*, 2020).

Todavia, a multiplicidade e variabilidade de correlações previamente publicadas para estimativa de  $k_{sat}$  (SHAHABI *et al.*, 1984; SIVAPPULLAIAH *et al.*, 2000; CHAPUIS, 2012; GOOTMAN; KELLNER; HUBBART, 2020) indicam a necessidade de analisar os critérios para a seleção da melhor correlação empírica e sua capacidade de previsão. Haja vista que muitas são desenvolvidas para condições e regiões específicas e foram obtidas a partir de um intervalo de valores bem definidos de  $k_{sat}$ , a extrapolação de seus resultados deve ser verificada a fim de reduzir sua utilização de forma indiscriminada.

Por vezes, faz-se necessário o estudo e desenvolvimento de novas correlações, mais adequadas ao problema particular, a partir de bancos de dados, utilizando análises estatísticas, como regressão linear múltipla (RLM), bem como o uso de técnicas de

inteligência artificial como redes neurais artificiais (RNA), máquinas de vetores de regressão, árvore de decisão, floresta aleatória, sistema de inferência neuro-*fuzzy* etc. (AKBULUT, 2005; PARK, 2011; YILMAZ *et al.*, 2011; TIZPA *et al.*, 2014; SEDAGHAT; BAYAT; SINEGANI, 2016; KASHANI *et al.*, 2020; SINGH *et al.*, 2020; TREJO-ALONSO *et al.*, 2021; WILLIAMS; OJURI, 2021).

No entanto, o uso dessas técnicas de predição exige a construção de bancos de dados que incluam a seleção de variáveis que caracterizem adequadamente o problema e que tenham direta influência sobre  $k_{sat}$ , para evitar-se o desenvolvimento de uma equação não representativa e/ou que leve a resultados tendenciosos.

Vários bancos de dados com valores de  $k_{sat}$  e outras propriedades índices estão disponíveis na literatura (BENSON *et al.*, 1984; WANG; HUANG, 1984; PUCKET *et al.*, 1985; NEMES *et al.*, 2001; OTTONI *et al.*, 2018; RAHMATI *et al.*, 2018; GUPTA *et al.*, 2020) e, para problemas de previsão de  $k_{sat}$ , a seleção de quais propriedades considerar é uma tarefa complexa, pois vários fatores afetam a permeabilidade de um solo saturado.

Dentre os fatores que exercem influência sobre  $k_{sat}$ , pode-se citar a granulometria, a densidade relativa do solo, a porosidade, a forma da partícula, a composição e a microestrutura do solo, as características do fluido percolante, a natureza dos minerais que compõem o solo etc. (TAYLOR, 1948; MITCHELL; SOGA, 2005; BOADU, 2002; TIZPA *et al.*, 2014; WILLIAMS; OJURI, 2021). Particularmente, o teor de finos de um solo afeta diversos parâmetros que exercem influência no  $k_{sat}$  (YANG; LACASSE; SANDVEN, 2006; ALDAOOD, 2020). Finos será o termo utilizado para se referir às partículas menores que 75  $\mu\text{m}$  (ASTM, 2022).

O objetivo desta dissertação é, portanto, compilar e comparar uma grande quantidade de dados experimentais disponibilizados na literatura para estudo da variação de  $k_{sat}$  de solos compostos por grãos grossos e finos, com o intuito de auxiliar na compreensão das propriedades hidráulicas de solos saturados e na descrição de problemas de predição de  $k_{sat}$ . A partir dos dados compilados, treinar arquiteturas de RNA e equações de regressão, para a previsão do valor do  $\log(k_{sat})$  a partir de banco de dados. Inicialmente, apresenta-se revisão bibliográfica sobre o tema, e a seguir apresentam-se e analisam-se os dados experimentais avaliados nesta dissertação.

## 1.1 Objetivos

### 1.1.1 Objetivo geral

Esta pesquisa tem como objetivo geral analisar e comparar o desempenho de RNAs e de equações de regressão, desenvolvidas a partir de dados experimentais disponibilizados na literatura, com os resultados obtidos de correlações empíricas selecionadas previamente na previsão de  $k_{sat}$  de solos compostos por grãos grossos e finos

### 1.1.2 Objetivos específicos

Os objetivos específicos desta pesquisa são:

- 1 Organizar e agrupar dados com as informações de propriedades índices (tais como percentuais de frações granulométricas, índice de vazios, limites de consistência, densidade real dos grãos etc.) utilizadas na previsão empírica do  $k_{sat}$ ;
- 2 Discutir possíveis limitações do uso de banco de dados compilados na literatura e a aplicabilidade de equações empíricas a essas amostras;
- 3 Treinar diversas RNA com diferentes arquiteturas para melhor aproximação da previsão de propriedades hidráulicas de solos saturados com base nas propriedades índices selecionadas;
- 4 Propor e avaliar equações de regressão para previsão de  $k_{sat}$ ;
- 5 Comparar o desempenho na previsão de  $k_{sat}$  entre as RNA e equações de regressão desenvolvidas com os resultados obtidos de correlações empíricas selecionadas previamente da literatura.

## 1.2 Justificativa

O conhecimento das propriedades hidráulicas de um solo saturado é importante para o estudo de várias situações em campo que envolvem fluxo de água por seus poros, como infiltração no solo, escoamento superficial, dispersão de poluentes subterrâneos, extração de águas de aquíferos, drenagem e controle de nível d'água freático, processo de adensamento, sistemas de irrigação, aplicação de fertilizantes, crescimento de vegetação etc.

Mesmo para o estudo da condutividade hidráulica de solos não saturados, é imprescindível o conhecimento de  $k_{sat}$  para a determinação da função de condutividade hidráulica e da permeabilidade relativa, esta última necessária para a aplicação de algumas correlações empíricas em estudos de solos não saturados.

As propriedades hidráulicas de solos saturados são normalmente determinadas a partir de resultados de ensaios em campo ou conduzidos em laboratório. Na ausência ou limitação de resultados de ensaios para determinação de tais parâmetros, o comportamento hidráulico do solo saturado pode ser estimado por correlações empíricas em função de propriedades índices das amostras desses solos.

Todavia, a multiplicidade e variabilidade de correlações previamente publicadas indicam a necessidade de analisar os critérios para a seleção da melhor correlação empírica e sua capacidade de previsão. Haja vista que muitas são desenvolvidas para condições e regiões específicas e foram obtidas a partir de um intervalo de valores bem definidos de  $k_{sat}$ , a extrapolação de seus resultados deve ser verificada a fim de reduzir sua utilização de forma indiscriminada.

Adicionalmente, as técnicas de inteligência artificial representam um recurso tecnológico computacional e matemático também capaz de prever resultados de  $k_{sat}$  e outras características hidráulicas em função de propriedades índices dos solos, podendo trazer ganhos substanciais à Engenharia Civil. Além disso, as técnicas de inteligência artificial são ferramentas que permitem o processamento de uma grande quantidade de dados e a realização de tarefas que seriam dificilmente executadas pelo ser humano e num intervalo de tempo tão abreviado.

Neste sentido, o esforço de avaliar o uso de RNA e comparar seu desempenho com o uso de correlações empíricas, disponíveis na literatura, e com equações de regressão desenvolvidos por métodos estatísticos para a previsão de propriedades hidráulicas dos solos se justifica pelo ganho significativo em qualidade e tempo em estudos preliminares de Geotecnia.

### **1.3 Estrutura da dissertação**

O trabalho está organizado em capítulos, conforme segue:

- **Capítulo 1 – Introdução** – apresenta as considerações iniciais, justificativa, objetivos e a forma como a dissertação é organizada.
- **Capítulo 2 – Revisão bibliográfica** – apresenta uma revisão bibliográfica acerca dos temas envolvidos nesta pesquisa, com o embasamento teórico necessário para entendimento dos fenômenos relacionados às propriedades hidráulicas de solos saturados, das RNA, bem como de correlações empíricas e métodos estatísticos de regressão.
- **Capítulo 3 – Materiais e métodos** – é feita a apresentação inicial dos bancos de dados compilados da literatura e apresentada a metodologia empregada para a aplicação das técnicas selecionadas. As condições para seleção das correlações empíricas da literatura são discutidas, bem como são apresentados os critérios para desenvolvimento das RNA e equações de regressão. Apresentam-se, por fim, as análises a serem realizadas sobre os resultados de cada técnica aplicada.
- **Capítulo 4 – Análise e discussão dos bancos de dados** – são apresentados e descritos os bancos de dados levantados na fase de revisão bibliográfica. Os dados apresentados pelos bancos são analisados sob a perspectiva da revisão de literatura, principalmente, para compreensão dos fatores que afetam a propriedade  $k_{sat}$ .
- **Capítulo 5 – Apresentação dos resultados e discussões** – são apresentados os resultados obtidos após o treinamento das arquiteturas de RNA, bem como a análise dos resultados. São apresentadas comparações de desempenho entre diferentes conjuntos de dados de entrada e arquiteturas de RNA para a estimativa de  $k_{sat}$ . Os resultados das RNA são comparados com os resultados obtidos em correlações disponíveis na literatura para estimativa do  $k_{sat}$ , bem como com os resultados das equações de regressão desenvolvidas.
- **Capítulo 6 – Conclusões e sugestões** – são apresentadas e consolidadas as conclusões obtidas pela pesquisa, as eventuais limitações encontradas e apresentadas sugestões de pesquisas que poderão ser realizadas no desenvolvimento de trabalhos futuros.

Por fim, são apresentadas as referências bibliográficas, anexos e apêndices.

## 2 REVISÃO BIBLIOGRÁFICA

As interações entre forças e energias associadas aos sistemas de múltiplas fases implicam na necessidade de separar a análise do comportamento de solos saturados e de solos não saturados (LU; LIKOS, 2004). Assim, a permeabilidade de um solo deve ser avaliada sob duas condições muito distintas: quando o solo está saturado, objeto desta pesquisa, e quando o solo está não saturado.

A depender do grau de saturação do solo, o seu comportamento quanto à permeabilidade é variável. Todavia, mesmo para análise de solos não saturados, o conhecimento do valor de  $k_{sat}$  é imprescindível para determinação da função de condutividade hidráulica do solo.

Uma revisão bibliográfica acerca dos temas envolvidos nesta pesquisa é, então, apresentada neste capítulo.

### 2.1 Lei de Darcy e o fluxo de água líquida no solo

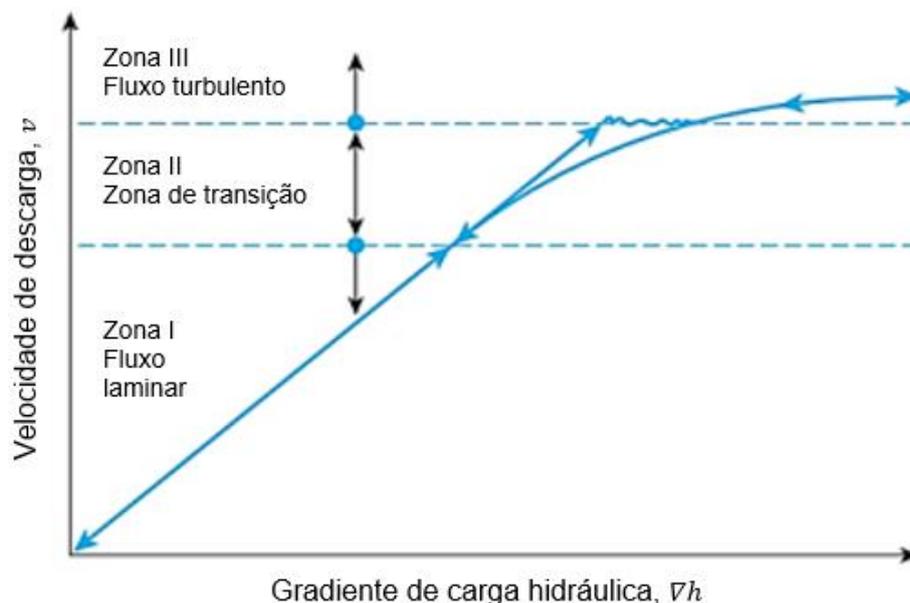
Henry Darcy, em 1856, estudou o fluxo de água por entre os vazios de uma massa de solo, utilizando um aparelho chamado permeâmetro. A Lei de Darcy foi originalmente desenvolvida para solos em condições muito específicas: solos arenosos uniformes, homogêneos, isotrópicos, saturados, não deformáveis e com temperatura constante (BICALHO; MACHADO; GERSCOVICH, 2015).

Trata-se de uma abordagem macroscópica que afirma que a taxa de fluxo hidráulico é linearmente proporcional ao gradiente de perda de carga (também chamado de gradiente hidráulico ou gradiente de carga hidráulica) (MASROURI; BICALHO; KAWAI, 2008). Em condições de regime de fluxo permanente, pode-se enunciar a Lei de Darcy como “a velocidade de descarga de um fluido em um meio poroso,  $v$ , é linearmente proporcional ao gradiente de carga hidráulica,  $\nabla h$ ” (LU; LIKOS, 2004, p. 328):

$$v = -k \times \nabla h \tag{1}$$

Importante reforçar que a Lei de Darcy, como enunciada, é válida para regime de fluxo laminar, no qual existe proporcionalidade linear entre  $\nabla h$  e a velocidade de fluxo  $v$ . Conforme apresentado na Figura 1, existe linearidade entre a velocidade de descarga e o gradiente hidráulico até um certo valor.

Figura 1 – Relação entre velocidade de descarga e gradiente de carga hidráulica



Fonte: Adaptado de Živković, Šafran e Zelić (2021).

À medida que o gradiente hidráulico aumenta, o regime de escoamento entra na zona de transição entre laminar e turbulento e passa a não existir mais linearidade entre  $v$  e  $\nabla h$ ; o fluxo apresenta um comportamento hidráulico intermediário entre laminar e turbulento. Por fim, caso o gradiente hidráulico assuma valores ainda maiores, o fluxo entra na zona III, que caracteriza o regime turbulento. Assim posto, a Lei de Darcy é válida para escoamentos que ocorram dentro da Zona I.

O termo de proporcionalidade na Equação (1),  $k$ , caracteriza a habilidade que um meio poroso possui de permitir a transmissão de um fluido pelo seu interior e é apresentado em unidade de comprimento por unidade de tempo. Nas condições consideradas por Darcy (1856), isto é, fluxo laminar em solos saturados, o termo de proporcionalidade é uma constante chamada *coeficiente de permeabilidade* e, por ter sido desenvolvida para solos em condição de saturação, é representado por  $k_{sat}$ .

## 2.2 Fatores que afetam o $k_{sat}$

Vários fatores contribuem para que um solo seja mais ou menos permeável. Dentre os quais, pode-se citar a densidade do solo, a umidade, o grau de saturação (em solos não saturados), o índice de vazios, a composição e a microestrutura do solo, as

características do fluido percolante, a natureza dos minerais que compõem o solo etc. (WILLIAMS; OJURI, 2021).

### 2.2.1 Influência da temperatura e do fluido percolante

O fluxo de fluidos pelos vazios do solo, tanto na condição total ou parcialmente saturada, depende da viscosidade do fluido, que é uma função da temperatura; por esta razão, convencionou-se como referência o valor de  $k_{sat}$  para fluidos a 20°C. Nos ensaios para determinação de  $k_{sat}$ , o valor encontrado a uma temperatura  $T$  qualquer é, então, corrigido para caso a água estivesse a 20°C.

As propriedades do fluido percolante influenciam na facilidade ou não para que fluxo ocorra por entre os vazios de um meio poroso. Para um mesmo solo, fluidos com propriedades diferentes fornecerão diferentes  $k_{sat}$ . Torna-se importante, então, a definição do conceito de *permeabilidade intrínseca*,  $K$ , que é dada por:

$$K = k \times \frac{\mu}{\rho \times g} \quad (2)$$

Onde  $k$  é a condutividade hidráulica do solo,  $g$  é a aceleração da gravidade ( $m/s^2$ ) e  $\mu$  e  $\rho$  são, respectivamente, a viscosidade dinâmica ( $N \cdot s/m^2$ ) e a massa específica ( $kg/m^3$ ) do fluido percolante.

A equação acima evidencia que a permeabilidade de um meio poroso depende tanto de propriedades do solo como também de propriedades do fluido percolante. Sendo assim, o  $k_{sat}$  do solo é um parâmetro resultante das propriedades intrínsecas do solo e do fluido que está percolando.

Bicalho, Machado e Gerscovich (2015) trazem à discussão que, a depender da natureza do mineral que compõem os grãos e partículas de solo, as propriedades do fluido percolante podem exercer maior ou menor influência sobre a permeabilidade saturada e não saturada. Em solos compostos por grãos considerados inertes, como as areias, a estrutura da matriz sólida é o principal fator que afeta a permeabilidade intrínseca.

Todavia, em solos compostos por minerais argila, cuja atividade superficial é intensa, devido ao desbalanceamento das cargas elétricas, a interação entre as partículas de sólido e do fluido percolante é evidente, principalmente quando o fluido é polar. Um

dos efeitos dessa interação elétrica e química é a formação da camada difusa, que reduz o canal disponível para fluxo.

Características químicas e elétricas do fluido percolante podem influenciar substancialmente a permeabilidade, se provocarem alterações na microestrutura dos solos (efeitos de inchamento, contração etc.) (LU; LIKOS, 2004). Como essa interação elétrica e química não é desprezível e não é contemplada na equação da definição de permeabilidade intrínseca, é importante ressaltar que esta equação deve ter uso priorizado para solos granulares (BICALHO; MACHADO; GERSCOVICH, 2015).

### 2.2.2 Influência dos finos

A influência do teor de finos pode ser identificada com o estudo de misturas controladas de solos grossos e finos e que foi realizado por diversos pesquisadores (BARBER; SAWYER, 1952; BOADU, 2000; YANG; LACASSE; SANDVEN, 2006; ALDAOOD, 2020). Por *misturas controladas* entende-se amostras com a mesma matriz grossa às quais são adicionados teores controlados de finos, normalmente em percentual em massa (%massa).

As pesquisas realizadas evidenciam que existe um percentual de finos que divide o comportamento do solo, isto é, se controlado pela matriz grossa ou pelos finos. Além disso, no geral, para misturas controladas de solo, o aumento do percentual de finos reduz o valor de  $k_{sat}$  até este teor de finos que é, então, definido como *teor de finos de transição* (TFT) (BARBER; SAWYER, 1952; BOADU, 2002; BANDINI; SATHISKUMAR, 2009; ALDAOOD, 2020). A partir deste TFT, o teor de finos passa, praticamente, a não afetar o valor de  $k_{sat}$  e este aproxima-se do valor do  $k_{sat}$  para o fino puro.

Na literatura, não existe, entretanto, um consenso a respeito do valor desse teor de finos de transição. Os resultados publicados mostram que, a partir de um teor de finos entre 20% e 30%, ocorre uma alteração na forma com a qual varia, não somente o  $k_{sat}$ , mas também as propriedades de compactação, de compressibilidade, de resistência ao cisalhamento e as propriedades índices, como coeficiente de uniformidade, massa específica e índice de vazios máximo e mínimo (BARBER; SAWYER, 1952; YANG; LACASSE; SANDVEN, 2006; BANDINI; SATHISKUMAR,

2009; BELKHATIR *et al.*, 2014; CHEMMAM *et al.*, 2015; ALAKAYLEH; CLEMENT; FANG, 2018; KHAN *et al.*, 2019; ALDAOOD, 2020; SHAKER *et al.*, 2022).

Na perspectiva da resistência ao cisalhamento, para baixo teor de finos, o fino é considerado um vazio e não contribui para a resistência ao cisalhamento, em outras palavras, para baixos teores de finos, a partícula fina não é capaz de perturbar o empacotamento dos grãos grossos (KOLTERMANN; GORELICK, 1995). À medida que o teor de finos aumenta, a fração granular é que passa a ser considerada um "vazio" e a não contribuir com a resistência ao cisalhamento, haja vista que o teor de finos é tão elevado que os grãos da matriz granular perdem contato entre si e passam a "flutuar" dispersos sobre a matriz fina do solo (YANG; LACASSE; SANDVEN, 2006).

Esse conceito foi introduzido por Carraro *et al.* (2003) como *matriz granular flutuante*, isto é, quando o índice de vazios da matriz granular é maior que o índice de vazios máximo da areia pura e os grãos não estão mais em contato. Na ocasião, em estudos de liquefação do solo, Carraro *et al.* (2003) concluiu que essa transição ocorria para um percentual de 15% de finos.

Embora não haja um consenso sobre o teor de finos no qual ocorre essa transição entre as matrizes, particularmente, sob a perspectiva da permeabilidade, quando observado o efeito dos finos em misturas controladas de solo, observa-se que pequenos teores de finos não plásticos (menores que 5%, por exemplo) já são capazes de reduzir sensivelmente o  $k_{sat}$  (BARBER; SAWYER, 1952; ALDAOOD, 2020), mas que esse efeito pode ser maior ou menor a depender, por exemplo, da natureza desse fino, como será discutido à frente. A adição de apenas 1% de montmorilonita em uma mistura com areia foi capaz de reduzir em mais da metade o  $k_{sat}$  da areia pura nas amostras de Denson, Shindala e Fenn (1968).

Além disso, ainda que o aumento do percentual de finos em misturas controladas de solo diminua o  $k_{sat}$ , considerando apenas o tamanho da partícula de fino, como a bentonita, seu efeito sobre o preenchimento dos poros pode não ser expressivo para percentuais abaixo de 5%, a ponto de afetar o valor de  $k_{sat}$  (SHAKER *et al.*, 2022).

Esse efeito do fino sobre o  $k_{sat}$  está relacionado às alterações que a presença da fração fina provoca em propriedades do solo, em particular na geometria dos poros, e que afetam a permeabilidade, como o índice de vazios e o diâmetro efetivo.

### 2.2.2.1 Efeito dos finos no índice de vazios

Alterações no índice de vazios do solo são observadas com o aumento da fração fina, independentemente da plasticidade do fino. Nos estudos realizados com misturas controladas, nos quais adiciona-se gradativamente diferentes teores de finos a uma matriz granular pura, observa-se que o aumento do teor de finos provoca a redução do índice de vazios.

Entretanto, à medida que o teor de finos na mistura aumenta, o índice de vazios passa novamente a aumentar, aproximando-se, inclusive, do índice de vazios da fração fina pura. Yang, Lacasse e Sandven (2006), Papadopoulou e Tika (2008) e Dash *et al.* (2010) reportaram que essa transição na variação do índice de vazios ocorre para um percentual de finos em torno de 20 a 30%.

Isso se dá, pois, à medida que se aumenta a fração fina, inicia-se, gradativamente, o preenchimento dos vazios da matriz granular e um índice de vazios mínimo é atingido quando todos os poros dos grãos grossos estão preenchidos por finos. Ao atingir esse valor mínimo do índice de vazios, o aumento do teor de finos induz a um afastamento dos grãos da matriz granular e, conseqüentemente, o aumento do índice de vazios. O índice de vazios final, correspondente a uma mistura composta 100% por finos, é maior que o índice de vazios da matriz granular pura (isto é, com 0% de finos). De fato, partículas finas têm a tendência de se agruparem em um empacotamento menor (maior índice de vazios) (LADE; LIGGIO; YAMAMURO, 1998).

Logicamente, essa variação no índice de vazios e, conseqüentemente, na porosidade, com o percentual de finos de um solo, exerce influência sobre a permeabilidade do solo saturado. À medida que o tamanho das partículas diminui, outras forças além das gravitacionais se tornam importantes na interação entre partículas e essas forças interpartículas controlam a formação do arranjo entre essas. As forças interpartículas impedem que as partículas de finos deslizem pelos vazios e se encaixem entre os grãos maiores, produzindo arranjos empacotados de partículas.

As partículas de finos tendem a se unir em aglomerados que, com os efeitos de atrito, formam arranjos soltos e estruturas com altas taxas de vazios entre si. As misturas de areia com finos tenderão, conseqüentemente, a apresentar taxas de vazios

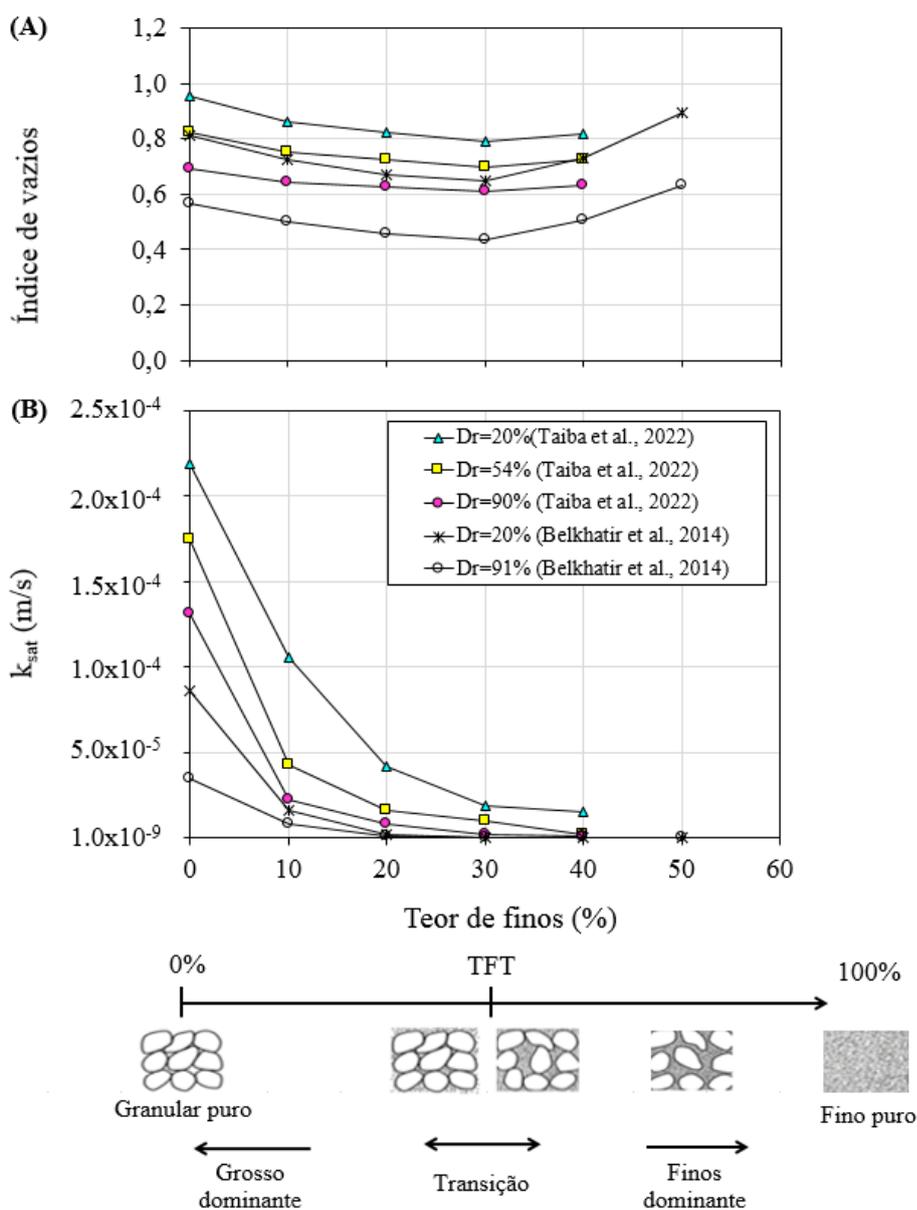
crescentes com o aumento do teor de finos, a partir de um teor elevado de finos na mistura (LADE; LIGGIO; YAMAMURO, 1998).

Na perspectiva da permeabilidade, com a redução do índice de vazios, até o TFT, reduz-se o espaço livre disponível para o fluxo e, conseqüentemente, há uma diminuição do valor de  $k_{sat}$  dos solos saturados. Porém, com o aumento do percentual de finos, mesmo que um aumento do índice de vazios seja observado, o valor de  $k_{sat}$  já não sofre variações acentuadas, tendendo ao valor do  $k_{sat}$  do fino puro.

Isso porque, tomando uma mistura com teor de finos acima do TFT, a quantidade de finos é suficiente para que a geometria e a distribuição dos poros, que dominam a permeabilidade, sejam controladas pela fração dos finos. É por essa razão, por exemplo, que misturas controladas que tenham teor de finos em torno de 30% já apresentam o valor de  $k_{sat}$  próximo do valor do  $k_{sat}$  do fino puro.

Para misturas com teores de finos abaixo do TFT, o percentual de grãos grossos é proporcionalmente maior e, assim, menos finos estão entre os poros da matriz granular. A condutividade da fração grossa passa a dominar o  $k_{sat}$  da mistura. No teor de 0% de finos, não há componente fino na mistura e o  $k_{sat}$  é o  $k_{sat}$  apenas dos grãos grossos (granular puro). A Figura 2 ilustra esse comparativo entre a variação do índice de vazios e a permeabilidade de misturas controladas de solos compostos por grãos grossos e finos, a partir dos resultados de Belkhatir *et al.* (2014) e Taiba *et al.* (2022).

Figura 2 - Variação do (A) Índice de vazios, e correspondente (B)  $k_{sat}$  em função do teor de finos em misturas de solos grossos e finos (Dados de: BELKHATIR *et al.*, 2014; TAIBA *et al.*, 2022).



Fonte: Autora (2022).

As amostras de Belkhatir *et al.* (2014) foram obtidas a partir da mistura de silte (IP=5%) em teores de 0% a 50%, em um solo de matriz arenosa. Corpos de prova foram constituídos a partir de duas densidades relativas ( $Dr = 20\%$  e  $91\%$ ) e saturados para determinação de  $k_{sat}$ . Na

Figura 2, os resultados da variação de  $k_{sat}$  em função do índice de vazios global das misturas, para as duas densidades relativas, são mostrados.

Já Taiba *et al.* (2022), com o objetivo de prever  $k_{sat}$  em função do contato intergrãos, em termos de fração ativa de finos e da razão  $d_{50}/d_{10}$ , adicionaram diferentes teores de fino de baixa plasticidade ( $IP=5\%$ ), em percentuais de 0% a 40%, em uma matriz granular composta por areia. Foram determinadas diversas propriedades índices, incluindo o índice de vazios global das misturas, obtido a partir de três densidades relativas ( $D_r = 20\%$ , 54% e 90%). Amostras foram saturadas e ensaiadas à carga constante nas três densidades relativas.

Observa-se, nos resultados da Figura 2, considerando a mesma matriz granular (composição granulométrica dos grãos grossos fixada), a adição de finos (em percentual em massa) reduz o índice de vazios global de forma acentuada até um teor de finos de, aproximadamente, 30%, para as amostras de Belkhatir *et al.* (2014) e Taiba *et al.* (2022).

A partir desse teor, há um aumento do índice de vazios global das misturas, ainda que de forma mais suave nas amostras de Taiba *et al.* (2022). Observa-se, que a redução de  $k_{sat}$  acompanha a redução dos vazios até o teor de 30%. A partir desse percentual, não são observadas alterações significativas de  $k_{sat}$  até os teores de finos ensaiados nas amostras, pois o comportamento hidráulico da mistura passa a ser controlado pela fração fina.

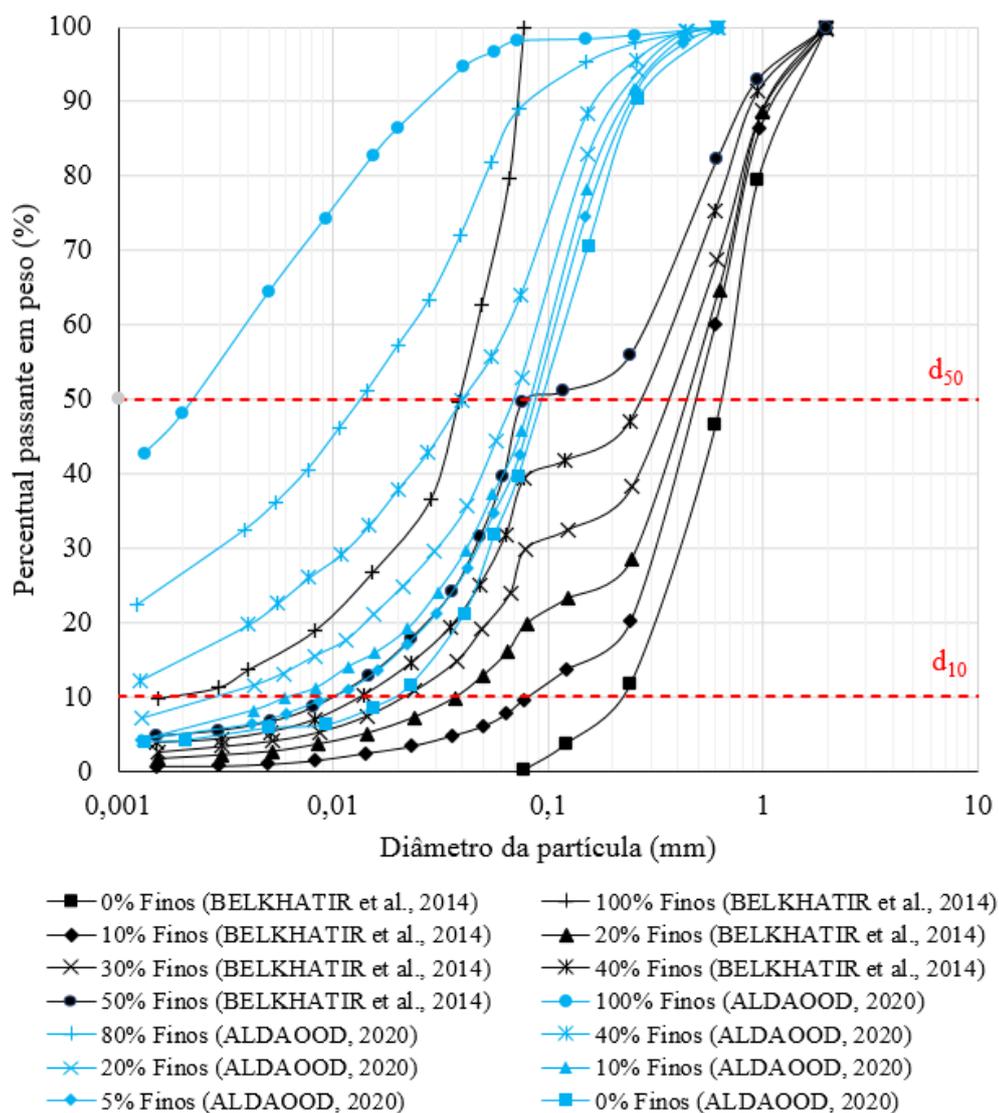
#### 2.2.2.2 Efeitos dos finos na distribuição granulométrica e no $d_{10}$

Efeitos na curva de distribuição granulométrica também são observados quando ocorre a adição de finos em misturas controladas. Observa-se que o aumento do teor de finos pouco altera o diâmetro médio das partículas até esse teor de finos de transição, a partir do qual é sensivelmente reduzido (YANG; LACASSE; SANDVEN, 2006; ALDAOOD, 2020).

Aldaoood (2020) estudou o efeito da variação do teor de finos em propriedades índices, propriedades da compactação e alterações na microestrutura, a partir de uma mistura de um solo tipicamente argiloso a uma areia siltosa, nas proporções de 0%, 5%, 10%,

20%, 40%, 60% e 80% em peso seco. O efeito da adição de finos observado na distribuição granulométrica, nas amostras de Aldaood (2020), é apresentado na Figura 3, que inclui também as alterações na distribuição granulométrica para os diferentes teores de adição de silte à areia nas amostras de Belkhatir *et al.* (2014).

Figura 3 - Distribuição granulométrica de misturas com diferentes teores de finos para misturas de areia e silte (Dados de: BELKHATIR *et al.*, 2014; ALDAOOD, 2020).



Fonte: Autora (2022).

Observa-se que o diâmetro médio das partículas sempre reduz com a adição de finos, mas ocorre uma redução mais expressiva a partir de 20% de adição de finos. O mesmo pode ser observado no valor de  $d_{10}$ .

O diâmetro efetivo,  $d_{10}$ , é conhecidamente relacionado com o  $k_{sat}$ . O valor de  $k_{sat}$  é tão menor quanto menor o valor de  $d_{10}$  (KHAN *et al.*, 2019), que apresenta melhor concordância com  $k_{sat}$  do que outros tamanhos de partículas, como  $d_{20}$  e  $d_{50}$  (CHAPUIS, 2012). Importante ressaltar que, em misturas controladas, o aumento do teor de finos reduz o valor de  $d_{10}$ ; conseqüentemente, um aumento no coeficiente de uniformidade é observado (CHEMMAM *et al.*, 2015).

Os menores grãos do solo são mais responsáveis por alterar o diâmetro médio dos poros que outros tamanhos de grãos que compõem a distribuição granulométrica. Portanto, o tamanho médio dos poros é afetado principalmente, por mudanças no  $d_{10}$  de um solo. O diâmetro efetivo controla mais a geometria dos poros do que o  $d_{50}$  e o  $d_{60}$ , por exemplo. Por esse motivo, a influência sobre os valores de  $k_{sat}$  (MAHMOODLU *et al.*, 2016).

A escolha por uma medida específica de tamanho de grão representativo depende do tipo de matriz que controla o comportamento hidráulico do solo (matriz granular ou fina) e depende dos teores de cada uma dessas frações na mistura. Recomenda-se, quando disponível toda a curva granulométrica do solo, o uso do diâmetro médio geométrico para misturas com teor de finos abaixo do valor do fino de transição e o uso do diâmetro médio harmônico para solos com teor de finos maiores ou igual ao valor de transição (KOLTERMANN; GORELICK, 1995).

Mesmo assim, outros parâmetros da distribuição granulométrica e da microestrutura, como o coeficiente de uniformidade e o índice de vazios, por exemplo, devem complementar a análise do coeficiente de permeabilidade do solo, uma vez que afetam a interconectividade dos poros (MUJTABA *et al.*, 2021). E para solos plásticos, especificamente, parâmetros de sua superfície específica ou limites de consistência devem ser considerados. Isso por que partículas de argila possuem interação com a água que não é contemplada, normalmente, nas equações empíricas (CHAPUIS; AUBERTIN, 2003).

### 2.2.2.3 Efeito da plasticidade do fino no $k_{sat}$

A influência da composição mineral de solos granulares e finos não plásticos exerce pouca influência nas propriedades hidráulicas, exceto quanto à existência de mica e

matéria orgânica. Mas quanto aos finos plásticos, a natureza do mineral argila, os íons trocáveis, a existência de impurezas (óxidos de ferro livres, por exemplo) e a capacidade de formação da dupla camada difusa, que afeta o espaço disponível para fluxo nos poros, torna a caracterização da natureza desse fino importante para análises de permeabilidade (LAMBE, 1954). Além disso, o tamanho dos grãos também é responsável por determinar a tortuosidade do caminho disponível para fluxo (DOLINAR, 2009).

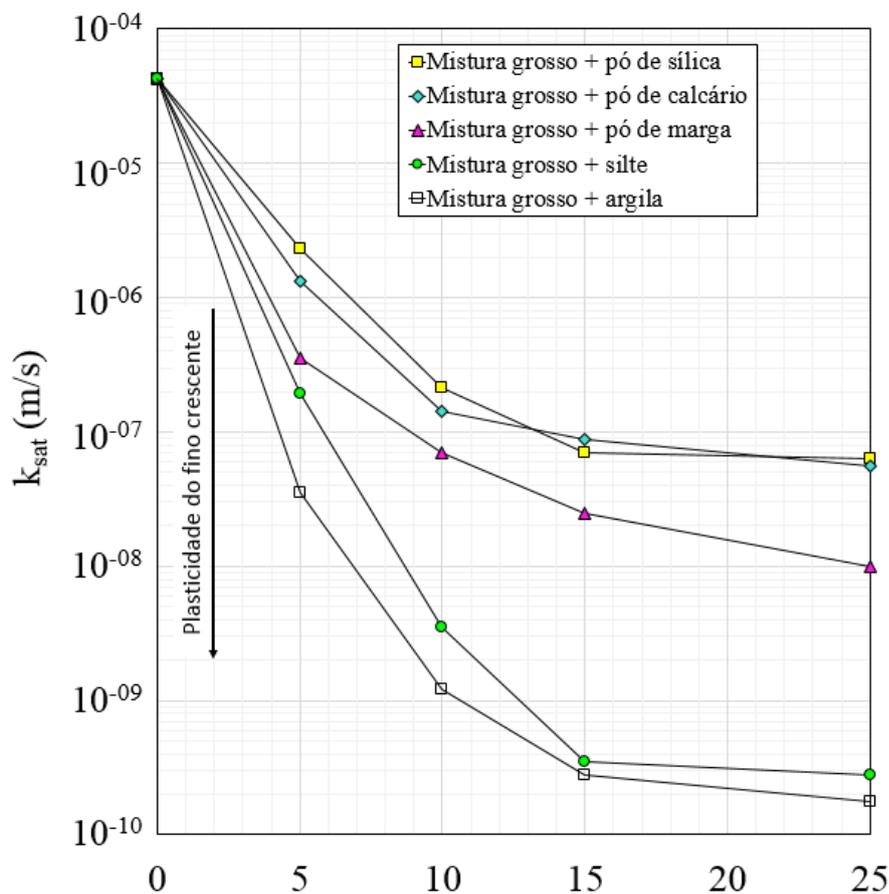
A depender do mineral argílico, a influência sobre os valores de  $k_{sat}$  pode ser bem distinta: a adição de 2,5% de montmorilonita reduziu em 97% o  $k_{sat}$  de uma areia pura, isto é, reduziu de  $1,08 \times 10^{-3}$  m/s para  $3 \times 10^{-5}$  m/s, em um estudo com misturas controladas de areia e argila conduzido por Denson, Shindala e Fenn (1968). No mesmo estudo, para alcançar uma redução de 80% no  $k_{sat}$  da mesma areia pura, foram necessários 10% de caulinita.

A plasticidade do argilomineral deve ser considerada, sempre que possível, na caracterização do solo. A adição de pequenos teores de finos de elevada plasticidade (montmorilonita, por exemplo) em uma areia pura pode levar a mistura a ter um comportamento de argila, pelo Ábaco de Casagrande, apesar de ser classificada como areia em sistemas de classificação, como o SUCS, por exemplo, quando considerada apenas a distribuição granulométrica (LEE; SUEDEKAMP, 1972).

A Figura 4 mostra o efeito da plasticidade do fino no  $k_{sat}$ . Diferentes teores de finos, plásticos e não plásticos, foram adicionados a uma mistura de cascalho e areia com composição granulométrica fixada e  $k_{sat}$  de  $4,23 \times 10^{-5}$  m/s. Os finos utilizados, que apresentavam diferentes plasticidades e permeabilidade saturada, foram: fino de sílica ( $k_{sat} = 9,17 \times 10^{-8}$  m/s), fino de calcário ( $k_{sat} = 8,82 \times 10^{-8}$  m/s), fino de marga ( $k_{sat} = 4,59 \times 10^{-9}$  m/s), silte ( $k_{sat} = 1,41 \times 10^{-9}$  m/s) e argila ( $k_{sat} = 7,06 \times 10^{-11}$  m/s).

As misturas com adição de sílica e as com adição de fino de origem calcária não apresentaram plasticidade para nenhum teor de fino na mistura. As misturas com fino com origem de marga apresentaram índice de plasticidade a partir da adição de 15% de finos. Já as misturas com silte e argila apresentaram plasticidade, independente do teor do fino na mistura. As misturas com argila apresentaram o dobro da plasticidade das misturas com silte.

Figura 4 - Efeito da adição de finos com diferentes plasticidades no  $k_{sat}$  resultante de misturas com solos granulares (Dados de: BARBER; SAWYER, 1952)



Fonte: Autora (2022).

Nota-se da Figura 4 que, à medida que o teor de finos aumenta, uma redução expressiva de  $k_{sat}$  é observada, até um teor de 25% de adição à mistura, e independentemente da plasticidade do fino. A partir desse percentual, o acréscimo de fino pouco altera o valor de  $k_{sat}$ , que se aproxima do valor para o fino puro.

Observa-se também que  $k_{sat}$  é tão menor quanto maior a plasticidade do fino e que a maior redução de  $k_{sat}$  ocorreu com a adição do fino de maior plasticidade. Isto é, a adição de 25% de fino de sílica reduziu o  $k_{sat}$  da matriz granular pura de  $4.23 \times 10^{-5}$  m/s para  $6.35 \times 10^{-8}$  m/s. Já a mesma adição de 25% de argila reduziu o  $k_{sat}$  do granular puro para  $1,76 \times 10^{-10}$  m/s; ambos os valores bem próximos ao  $k_{sat}$  desses finos puros.

Estudos sugerem que a natureza do fino influencia no teor de finos de transição e que, devido à plasticidade do fino, existe uma diferença do percentual que afeta o comportamento do solo quanto à resistência e quanto à permeabilidade, como mencionado anteriormente. O que vale ressaltar é que essa diferença entre o teor de

transição matriz granular x matriz de finos está relacionada à plasticidade do fino. Reportam-se que misturas de maior plasticidade apresentam teor de fino de transição menor que misturas de menor plasticidade, na perspectiva da permeabilidade (EKICI; HUVAJ; AKGUNER, 2022).

Existe, portanto, uma relação entre o  $k_{sat}$  e os limites de consistência da fração fina. Sendo assim, o valor de  $k_{sat}$  reduz, em geral, com o teor de finos, mas também com a plasticidade do solo; para solos com mesmo teor de finos, por exemplo, o  $k_{sat}$  tende a ser menor para o solo de maior plasticidade (ALAKAYLEH; CLEMENT; FANG, 2018; EKICI; HUVAJ; AKGUNER, 2022).

Argilominerais não são quimicamente inertes e, devido à sua elevada superfície específica, possuem forte interação com a água ao redor, como mencionado anteriormente, que é responsável pela plasticidade. As partículas de argilominerais possuem cargas elétricas desbalanceadas em sua superfície, o que faz ocorrer uma forte interação com o líquido circundante e seus íons dissolvidos (ALAKAYLEH; CLEMENT; FANG, 2018). Finos não plásticos possuem menor superfície específica e menor afinidade pela água (DOLINAR, 2009).

Moléculas de água são adsorvidas na superfície de argilominerais e formam uma camada de água de elevada resistência viscosa e praticamente imóvel, denominada dupla camada difusa. A água adsorvida, por apresentar resistência viscosa elevada, não está livre e bloqueia parte do canal disponível para fluxo; quando considerado apenas o espaço livre para percolação, a porosidade é chamada de *eficiente*, pois não inclui a água imóvel retida na superfície de argilominerais (DENSON; SHINDALA; FENN, 1968; ALAKAYLEH; CLEMENT; FANG, 2018).

Às cargas elétricas desbalanceadas, presentes na superfície da partícula, é associado um potencial elétrico, que vai determinar o “alcance” desse efeito. A espessura dessa camada difusa pode, então, ser estimada e vai depender de fatores como a força iônica, que é função dos cátions presentes no fluido dos poros, a constante dielétrica e a temperatura deste fluido. A espessura da camada difusa é tão maior quanto menor a força iônica do fluido, função da valência e da concentração das espécies iônicas dissolvidas, e, conseqüentemente, menor será o coeficiente de permeabilidade (GLEASON; DANIEL; EYKHOLT, 1997).

Como exemplo, a bentonita sódica, quando hidratada com água destilada, apresenta um limite de liquidez e de plasticidade muito maior do que a bentonita cálcica. Em gasolina, ambas não apresentam plasticidade. Essa particular interação com o fluido percolante de argilominerais reflete em suas propriedades e, conseqüentemente, no coeficiente de permeabilidade. Na análise do fluxo de água, o aumento do teor de bentonita sódica produz uma redução no coeficiente de permeabilidade mais acentuada que a adição de bentonita cálcica (GLEASON; DANIEL; EYKHOLT, 1997).

Quando comparados os efeitos da adição de caulinita e montmorilonita em misturas com areia, reduções muito expressivas no coeficiente de permeabilidade são obtidas para percentuais muito baixos de montmorilonita. Enquanto para caulinita seria necessário um percentual maior de adição na mistura, para produzir a mesma redução na permeabilidade (DENSON; SHINDALA; FENN, 1968).

Todavia, apesar das diferenças entre as naturezas da fração fina de um solo, as equações empíricas para previsão de  $k_{sat}$  não contemplam, em sua maioria, um termo que relaciona diretamente a composição mineralógica da fração argila, o que torna limitado o uso de muitas dessas equações, principalmente para solos plásticos. Porém, a superfície específica é um indicativo mais próximo da composição mineralógica dessa fração (LAMBE, 1954).

Salienta-se, no entanto, que determinar a superfície específica também não é uma tarefa trivial, por estar relacionada com a esfericidade, angulosidade e gradação (LIU; JENG, 2019). Quando se trata de um argilomineral expansivo, os limites de consistência ainda dependem das superfícies externas e internas e dos íons intercamadas (DOLINAR, 2009; EKICI; HUVAJ; AKGUNER, 2022) e a presença de matéria orgânica aumenta a superfície específica desses solos (CHAPUIS, 2012).

Para teores baixos de argilominerais expansivos em misturas com solos granulares, o inchamento da partícula está contido nos vazios da areia, sem alterar a estrutura da mistura. À medida que o teor de argilomineral expansivo aumenta, o vazio é todo preenchido e os grãos de areia são empurrados, resultando no inchamento da mistura de areia e argila (SHAKER *et al.*, 2022).

### 2.2.3 Conclusões parciais do item 2.2

A partir da revisão de literatura, conclui-se que  $k_{sat}$  é um parâmetro de complexa caracterização, pois é muito sensível ao estado e a outras propriedades do solo, como distribuição granulometria e natureza dos finos. Alguns outros apontamentos podem ser realizados:

1. Os resultados com misturas controladas mostram que geralmente o teor de finos exerce acentuada influência no  $k_{sat}$  até um percentual em torno de 20% a 30% de finos; a partir desse teor, o  $k_{sat}$  dos solos compostos por grãos grossos e finos se aproxima do valor do fino puro. Todavia, pequenos teores de finos de uma argila muito plástica, na ordem de 5%, já são responsáveis por uma significativa redução de  $k_{sat}$  quando adicionados a um solo granular puro;
2. A influência do teor de finos nos valores de  $k_{sat}$  se dá, pois, a alteração do teor de finos induz variações em outras propriedades que afetam o valor de  $k_{sat}$ , como o índice de vazios e diâmetro efetivo;
3. Misturas com finos plásticos tendem a apresentar menor valor de  $k_{sat}$  do que com finos não plásticos. A redução de  $k_{sat}$  com o teor de finos também varia a com plasticidade: quanto maior a plasticidade maior será a redução de  $k_{sat}$  com a adição do fino.

É importante, portanto, que bancos de dados para uso no desenvolvimento de correlações para previsão de  $k_{sat}$  contemplem as propriedades que exerçam influência sobre esse parâmetro e que auxiliem na interpretação dos resultados. O fornecimento de uma caracterização mais completa das amostras permite a execução de tarefas, como a análise de dados discrepantes; para a tomada de decisão sobre a exclusão de uma amostra, quanto mais informações estiverem disponíveis, mais confiável será análise sobre a eventual discrepância de um valor.

Para um parâmetro do solo como o  $k_{sat}$ , que sofre influência de tantas outras propriedades interdependentes do solo, é fundamental que uma caracterização completa das amostras seja fornecida, principalmente abrangendo o teor de finos, as características da distribuição granulométrica e propriedades que se relacionam a mineralogia do fino (limite de liquidez, por exemplo), permitindo, assim, serem avaliadas e consideradas em estudos para estimativa de  $k_{sat}$ .

Ainda, a partir da revisão de literatura, foi possível a compreensão das propriedades de afetam o  $k_{sat}$  e, com isso, um melhor direcionamento para a definição de

parâmetros de entrada a serem utilizados no desenvolvimento de RNA e equações de regressão.

O processo de desenvolvimento de uma equação de predição de  $k_{sat}$  é, portanto, complexo, pois envolve a uma investigação por variáveis que melhor descrevam sua caracterização e que abordam propriedades que podem influenciar na permeabilidade. Importante escolher variáveis para a equação que caracterizem a microestrutura do solo e característica do tamanho da partícula ou superfície específica.

### **2.3 Equações empíricas para determinação do $k_{sat}$**

A variabilidade espacial natural, por vezes, encontrada no  $k_{sat}$  exige que, para uma adequada caracterização da permeabilidade em campo, um número representativo de amostras sejam ensaiadas. Na impossibilidade de realização de ensaios em quantidade suficiente, dados os custos e o tempo necessário para sua execução, recorrer às correlações empíricas é uma alternativa. Muitas dessas correlações fazem uso de propriedades do solo de simples determinação para, a partir delas, estimar o valor de  $k_{sat}$ .

A interconectividade e a distribuição dos poros do solo são fatores, dentre outros, que afetam  $k_{sat}$ ; porém, dada a complexidade em descrever o arranjo da microestrutura dos solos, parâmetros como a distribuição granulométrica e o índice de vazios (ou a porosidade), por exemplo, têm sido utilizados para a definição de correlações empíricas para previsão de  $k_{sat}$ .

Živković, Šafran e Zelić (2021) realizaram uma pesquisa cujo objetivo foi comparar os resultados do  $k_{sat}$  do solo obtidos por ensaios de laboratório em permeâmetro de carga constante com os resultados de correlações empíricas, baseadas na distribuição granulométrica do solo, disponíveis na literatura. Os pesquisadores identificaram que, para certos tipos de solos, os resultados entre ensaio de laboratório e correlações empíricas podem apresentar-se muito divergentes.

Concluíram também que as correlações apresentam resultados mais satisfatórios quando utilizadas dentro dos critérios de validade em que foram determinadas. Reforçaram que pequenas diferenças na estimativa do  $k_{sat}$  podem causar efeitos

consideráveis, a depender do tipo de obra a que se destina, como contenção de poluentes.

Os pesquisadores defenderam que não há um método único a ser utilizado em todas as fases de um projeto e que deve ser priorizado sempre o uso de mais de um método, combinados. Para a condição saturada, algumas equações matemáticas foram reunidas da literatura e apresentadas no Quadro 1. Equações matemáticas para previsão de propriedades hidráulicas de solos não saturados foram reunidas e apresentadas em Bicalho, Machado e Gerscovich (2015).

Quadro 1 - Equações matemáticas para estimativa de  $k_{sat}$ 

Nº	Autor (es)	Equação	Definição das variáveis	Comentários
C1	Hazen (1911) <sup>(3)</sup>	$k_{sat} = C_H \times d_{10}^{-2}$	$d_{10}$ – diâmetro efetivo (mm) $C_H = 1,157(0,70+0,03T)$ – coeficiente empírico de Hazen, onde T é a temperatura do fluido em graus Celsius $k_{sat}$ em cm/s	Válido para areias uniformes, naturais e com e próximo ao $e_{máx}$ . Aplicável para intervalo entre 0,1 mm e 3 mm de $d_{10}$ e $C_u \leq 5$ .
C2	Kozeny-Carman (1927, 1937) <sup>(1)</sup>	$k_{sat} = \frac{\gamma}{\mu \times C_{KC} \times S_s^2} \frac{e^3}{1+e}$	$\gamma$ – peso específico do líquido permeante ( $10^{-3}$ N/cm <sup>3</sup> ) $\mu$ – viscosidade do permeante (g/cm.s) $C_{KC}$ – coeficiente empírico de Kozeny-Carman $S_s$ – superfície específica por unidade de volume (1/cm) $e$ – índice de vazios $k_{sat}$ em cm/s	Válida para areias.
C3	Kozeny-Carman (CHAPUIS; AUBERTIN, 2003) <sup>(2)</sup>	$\log(k_{sat}) = 0,5 + \log\left(\frac{e^3}{G_s^2 \times S_s^2 (1+e)}\right)$	$S_s$ – superfície específica por unidade de volume (m <sup>2</sup> /kg) $G_s$ – densidade real dos grãos $e$ – índice de vazios $k_{sat}$ em m/s	Válida para solos plásticos e não plásticos.
C4	NAVFAC DM7 (1974) <sup>(3)</sup>	$k_{sat} = 10^{1,291e-0,6435} (d_{10})^{10,5504-0,2937e}$	$d_{10}$ – diâmetro efetivo (mm) $e$ – índice de vazios $k_{sat}$ em cm/s	Válida para areia ou misturas de areia e pedregulho, com $0,1\text{mm} \leq d_{10} \leq 2\text{mm}$ , $2 \leq C_u \leq 12$ , $d_{10}/d_5 \leq 1,4$ e $0,3 < e < 0,7$
C5	Chapuis (2004) <sup>(4)</sup>	$k_{sat} = 2,4622 \left( d_{10}^2 \frac{e^3}{1+e} \right)^{0,7825}$	$d_{10}$ – diâmetro efetivo (mm) $e$ – índice de vazios $k_{sat}$ em cm/s	Válida para areias e pedregulhos, com $0,003\text{mm} < d_{10} < 3\text{mm}$ e $0,3 < e < 1,0$
C6	Sivappullaiah <i>et al.</i> (2000) <sup>(5)</sup>	$\log k_{sat} = \frac{e - 0,0535 w_L - 5,286}{0,0063 w_L + 0,2516}$	$w_L$ – limite de liquidez $e$ – índice de vazios $k_{sat}$ em m/s	Válida para solos com limite de liquidez maior que 50%.
C7	Shahabi <i>et al.</i> (1984) <sup>(3)</sup>	$k_{sat} = 1,2 \times C_u^{0,735} \times d_{10}^{0,89} \times \frac{e^3}{1+e}$	$d_{10}$ – diâmetro efetivo (mm) $e$ – índice de vazios $C_u$ – coeficiente de uniformidade $k_{sat}$ em cm/s	Válido para areias, com $0,15\text{mm} \leq d_{10} \leq 0,59\text{mm}$ , $1,2 \leq C_u \leq 8$ e $0,38 < e < 0,73$

Observações:  
(1) Carrier (2003);  
(2) Chapuis; Aubertin (2003);  
(3) Chapuis (2012);  
(4) Chapuis (2004);  
(5) Sivappullaiah, Sridharan e Stalin (2000);

Fonte: Autora (2022).

Quadro 1 - Equações matemáticas para estimativa de  $k_{sat}$  – cont.

Nº	Autor (es)	Equação	Definição das variáveis	Comentários
C8	Mbonimpa <i>et al.</i> (2002) <sup>(3)</sup>	$k_{sat} = C_G \times \frac{\gamma_w}{\mu_w} \times C_u^{1/3} \times d_{10}^2 \times \frac{e^{3+x}}{1+e}$	$d_{10}$ – diâmetro efetivo (cm) $e$ – índice de vazios $C_u$ – coeficiente de uniformidade $x$ – usualmente, 2 $C_G$ – coeficiente empírico para solos granulares, usualmente, 0,1 $\gamma_w$ - peso específico da água (kN/m <sup>3</sup> ) $\mu_w$ – viscosidade dinâmica da água (Pa.s) $k_{sat}$ em cm/s	Válida para solos não plásticos.
C9		$k_{sat} = C_p \times \frac{\gamma_w}{\mu_w} \times \frac{e^{3+x}}{1+e} \times \frac{1}{\rho_s^2 \times LL^{2x}}$	$e$ – índice de vazios $C_p$ – coeficiente empírico para solos plásticos, usualmente, 5,6 g <sup>2</sup> /m <sup>4</sup> $x$ – usualmente, 1,5 $x = 7,7 LL^{-0,15} - 3$ $\gamma_w$ - peso específico da água (kN/m <sup>3</sup> ) $\mu_w$ – viscosidade dinâmica da água (Pa.s) $k_{sat}$ em cm/s	Válida para solos plásticos.
C10	Pucket <i>et al.</i> (1985) <sup>(6)</sup>	$k_{sat} = 4,36 \times 10^{-5} \times e^{-0,1975 \times Arg}$	Arg – percentual de argila (%) $e$ – número de Euler $k_{sat}$ em m/s	Não são apresentadas restrições explícitas. Condições de desenvolvimento da PTF*: $8,33 \times 10^{-10} \leq k_{sat} \leq 3,94 \times 10^{-05}$ m/s. Solos plásticos e não plásticos.
C11	Jabro (1992) <sup>(7)</sup>	$\log k_{sat} = 9,56 - 0,81 \log Si - 1,09 \log Arg - 4,64 \rho_d$	Si – percentual de silte (%) Arg – percentual de argila (%) $\rho_d$ – massa específica total (g/cm <sup>3</sup> ) $k_{sat}$ em cm/h	Não são apresentadas restrições explícitas. Condições de desenvolvimento da PTF*: $6,39 \times 10^{-09} \leq k_{sat} \leq 4,83 \times 10^{-04}$ m/s. Solos plásticos e não plásticos.
C12	Salarashayeri e Siosemarde (2012) <sup>(8)</sup>	$k_{sat} = 10,06 + 118,54 d_{10} - 12,50 d_{50} - 7,32 d_{60}$	$d_{10}$ – diâmetro efetivo (mm) $d_{50}$ – diâmetro médio dos grãos (mm) $d_{60}$ – diâmetro tal que 60% do solo, em peso, tem diâmetros menores que ele (mm) $k_{sat}$ em m/d	Válida para areias. $1,00 \times 10^{-02} \leq k_{sat} \leq 4,19 \times 10^{-05}$ m/s
C13	Carrier (2003) <sup>(9)</sup>	$k_{sat} = 1,99 \times 10^4 \left( \frac{100\%}{\left\{ \sum \left[ \frac{f_i}{D_i^{0,404} \times D_{si}^{0,595}} \right] \right\}} \right)^2 \left( \frac{1}{SF^2} \right) \times \left[ \frac{e^3}{(1+e)} \right]$	$f_i$ - fração de partículas entre dois tamanhos de peneiras: maior (l) e menor (s) (%) $D_l$ e $D_s$ – tamanho da partícula maior e menor, respectivamente (cm) SF – fator de forma da partícula $k_{sat}$ em cm/s	Válida para água a 20°C; Reformulação da equação de Kozeny-Carman. Valores típicos de SF em Fair e Hatch (1933).
<p>Observações: *PTF – função de pedotransferência (7) Jabro (1992); (3) Chapuis (2012); (8) Salarashayeri e Siosemarde (2012); (6) Pucket <i>et al.</i> (1985); (9) Carrier (2003).</p>				

Fonte: Autora (2022).

Observa-se, nas equações empíricas e semiempíricas selecionadas da literatura, a necessidade de determinação de propriedades índices que caracterizem, pelo menos, a distribuição granulométrica, os vazios e/ou a plasticidade do solo.

Um cuidado deve ser tomado na escolha da equação para estimativa do  $k_{sat}$ ; atenção deve ser dada para os critérios de validade da equação (algumas apresentam mais de um critério). Além disso, as equações são, na maioria, aplicáveis ou para solos plásticos ou para solos não plásticos.

A equação de Hazen (1911) foi uma das pioneiras para a previsão de  $k_{sat}$ . Hazen (1911) desenvolveu uma equação empírica que correlaciona o diâmetro efetivo e o coeficiente de permeabilidade. Considerando o coeficiente empírico de Hazen ( $C_H$ ) igual a 1,5 (para uma temperatura de 20°C), a equação é válida para intervalo de  $k_{sat}$  entre  $1,5 \times 10^{-4}$  m/s e  $1,35 \times 10^{-1}$  m/s, desde que o coeficiente de uniformidade seja menor que 5.

Embora valores de  $C_H$  já foram reportados na literatura entre de 1 a 1000 (CARRIER, 2003), por sua simplicidade, é muito utilizada até os dias atuais. Todavia, por apresentar um intervalo de validade muito restrito e pela falta de um consenso sobre o valor da constante, a aplicabilidade da equação é prejudicada (CARRIER, 2003).

Uma equação semiempírica que também apresenta boa concordância para solos com menor teor de finos (KHAN *et al.*, 2019) e muito discutida na literatura é Kozeny-Carman (1927, 1937). Desenvolvida para solos granulares, é uma equação que envolve tanto um macroparâmetro (índice de vazios) quanto um microparâmetro (superfície específica) do solo. Todavia, seu uso não é tão popularizado, justamente por necessário o conhecimento da superfície específica dos sólidos.

Os métodos para determinação experimental da superfície específica são divididos em duas categorias: adsorção de gás ou adsorção de moléculas, que envolve líquidos polares, normalmente. São métodos que envolvem tempo e custo; uma comparação entre métodos experimentais é apresentada e discutida em Yuksele e Kaya (2006).

A determinação da superfície específica não é corriqueira nos laboratórios e isso dificulta, inclusive, o uso de algumas correlações empíricas, como Kozeny-Carman (1927, 1937) (CARRIER, 2003). A superfície específica pode ser medida ou estimada,

porém os valores medidos da superfície específica podem ser maiores que os valores a serem utilizados na predição do coeficiente de permeabilidade.

Isso se dá devido ao fato de que algumas partículas muito irregulares apresentam reentrâncias em sua superfície e porosidade intragranular, isto é, irregularidades na superfície que formam canais que não contribuem com o fluxo de líquido no interior do solo (CARRIER, 2003). Métodos de determinação da superfície específica apresentam maior acurácia para solos granulares (CHAPUIS; AUBERTIN, 2003).

Como alternativa, a superfície específica pode ser estimada por métodos e correlações (CHAPUIS, 2012). Quando disponível a caracterização da distribuição granulométrica completa, isto é, incluindo o peneiramento e a sedimentação, é possível estimar a superfície específica de solos não plásticos pelo método proposto por Chapuis e Légaré (1992).

Já para solos plásticos, uma alternativa é o uso do limite de liquidez, parâmetro de mais fácil obtenção, por estar relacionado à superfície específica e a mineralogia da partícula de argila (CHAPUIS, 2012; BILARDI; IELO; MORACI, 2020). Uma correlação entre a superfície específica ( $S_s$ ) e o limite de liquidez (LL) foi apresentada por Muhunthan (1991). Para solos com limite de liquidez menor que 110%, pode ser escrita uma relação linear entre  $1/S_s$  e  $1/LL$ , que é apresentada na Equação 3.

$$\frac{1}{S_s} = \frac{1,3513}{LL} - 0,0089 \quad (3)$$

Contudo, tanto os métodos experimentais quanto o uso de correlações para determinação da superfície específica trazem incertezas às predições de  $k_{sat}$  através das equações empíricas e semiempíricas e pode ser responsável pela baixa acurácia de alguns resultados.

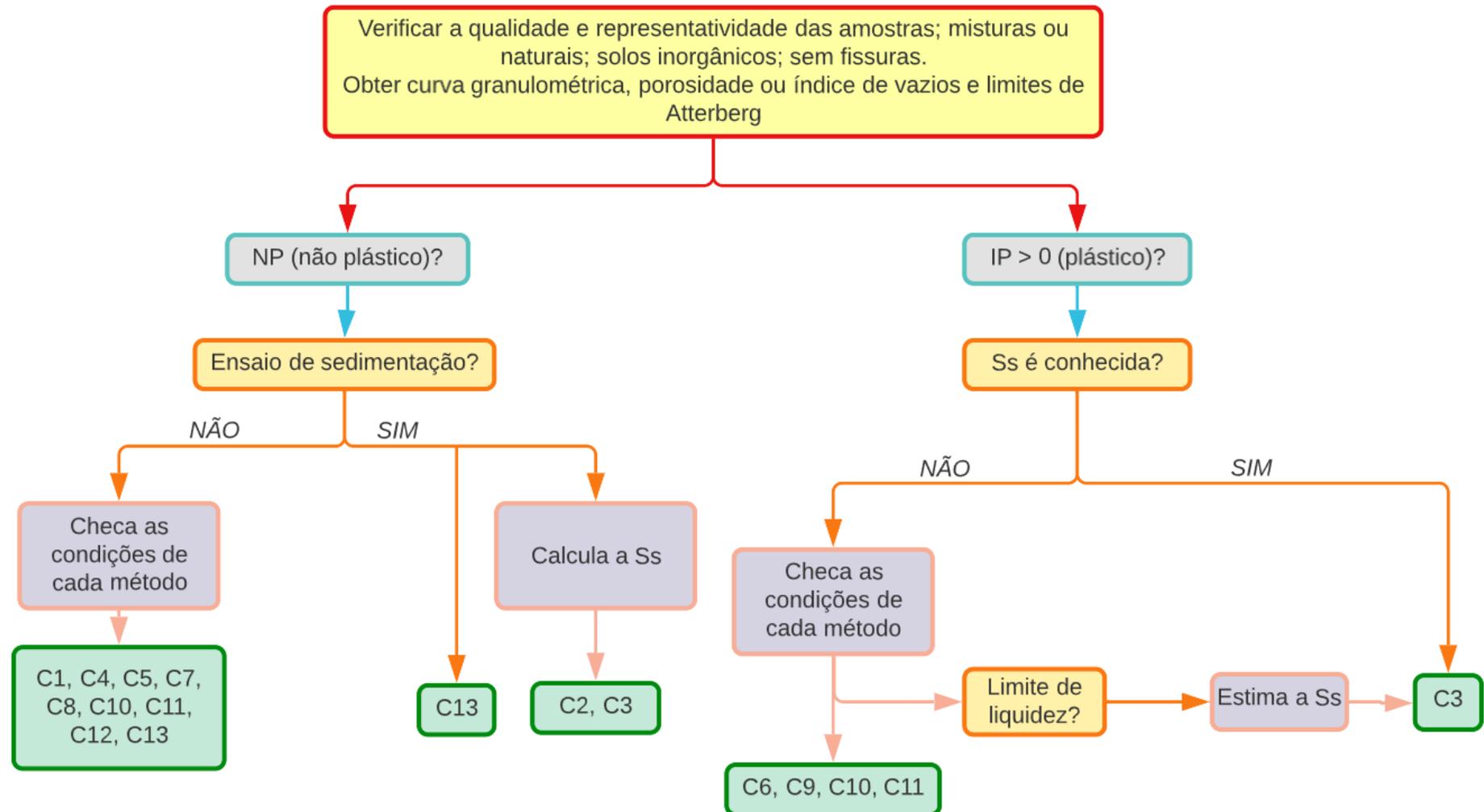
Chapuis e Aubertin (2003), após analisarem em detalhes a equação original de Kozeny-Carman (1927, 1937), concluíram que ela pode ser utilizada para solos plásticos, desde que apresentada na forma trazida no Quadro 1 (correlação C3). Os pesquisadores simplificaram parte da equação, mas ainda é necessária a superfície específica.

Outro ponto a se destacar nas equações empíricas e semiempíricas é que, em geral, poucas delas apresentam características do fluido percolante. Apenas Kozeny-

Carman (1927, 1937) e Mbonimpa *et al.* (2002) apresentam parâmetros do fluido percolante.

A variedade de equações e das condições em que foram desenvolvidas, além de suas diferentes restrições, justifica a necessidade de um estudo comparativo entre os métodos apresentados. Corroborando com Živković, Šafran e Zelić (2021), uma análise de previsão do  $k_{sat}$  deve ser complementada com outras técnicas de previsão deste parâmetro a fim de incorrer em menor erro no resultado. Para facilitar a escolha dentre as equações empíricas apresentadas no Quadro 1, o seguinte fluxograma foi elaborado. Nesta pesquisa, a seleção das equações empíricas e semiempíricas ainda considerou um critério adicional: a equação deveria ser aplicável a toda base de dados, como será detalhado no Capítulo 3.

Figura 5 – Fluxograma para escolha da equação empírica para estimativa de  $k_{sat}$



Fonte: Autora (2022).

## 2.4 Desenvolvimento de métodos para previsão de $k_{sat}$

Como mencionado anteriormente, haja vista a multiplicidade, variabilidade e restrições de correlações previamente publicadas na literatura, por vezes, é necessário o estudo e desenvolvimento de novas correlações que permitam uma previsão mais ajustada do valor de  $k_{sat}$ . Além disso, é importante desenvolver correlações que melhor incluam as variáveis que afetam esse parâmetro.

Neste intuito, existem métodos estatísticos e técnicas de inteligência artificial que permitem realizar tarefas de predição de dados e que podem ser aplicadas para prever resultados de  $k_{sat}$  em função de propriedades índices dos solos, através do processamento de uma grande quantidade de dados.

### 2.4.1 Métodos estatísticos

Tomar decisões e fazer conclusões diante da variabilidade de alguns fenômenos é possível com o auxílio da Estatística. A Estatística é uma ciência de dados que inclui a coleta, apresentação, análise e uso dos dados na solução de problemas. A partir dos métodos estatísticos, é possível entender a variabilidade e quais os potenciais fatores que mais influenciam nos resultados de um fenômeno observado (MONTGOMERY; RUNGER, 2012).

Dentre os objetivos específicos desta pesquisa, um deles é encontrar e avaliar correlações entre o coeficiente de permeabilidade do solo e propriedades índices de mais fácil determinação experimental. Equações de correlação objetivam explicar matematicamente a relação que existe entre variáveis aleatórias, considerando a influência de fontes de variabilidade sobre essas variáveis, observadas em experimentos (CREVELIN, 2018).

#### 2.4.1.1 Etapas da criação de uma correlação

O desenvolvimento de uma equação de correlação envolve diversas etapas sequenciadas, como descrito por Santos (2017). A partir de um banco de dados, com um número de amostras ( $n$ ) que permita a caracterização representativa do fenômeno que está sendo estudado, informações detalhadas das características, considerações

e limitações dessas amostras devem ser levantadas. Assim, evita-se o desenvolvimento de uma equação não representativa e/ou que leve a resultados tendenciosos.

Após o levantamento do banco de dados e uma investigação sobre a origem e representatividade dessas amostras, passa-se ao tratamento dos dados. O objetivo desta fase é remover informações discrepantes e verificar dados inconsistentes ou ausentes, para melhorar a qualidade do conjunto e evitar que esses erros possam prejudicar o alcance dos objetivos.

Aqui, incluem-se operações de verificação da consistência dos dados, correção de eventuais erros e preenchimento de lacunas, ou seja, atributos faltantes, quando assim for possível. Uso de ferramentas estatísticas podem ser empregadas para auxiliar nessa fase, como o uso de tabelas e gráficos (GOLDSCHIMIDT; PASSOS, 2005).

Medidas estatísticas de dispersão das variáveis do problema devem ser obtidas com o objetivo de avaliar a dispersão dos resultados em torno de seu valor médio ( $\bar{x}$ ); o desvio-padrão ( $s$ ), a variância ( $s^2$ ), os valores máximo, médio e mínimo e o coeficiente de variação (CV) são algumas das principais medidas utilizadas para este fim (SANTOS, 2017). As Equações 4, 5 e 6 definem a variância, o desvio-padrão e o coeficiente de variação, respectivamente.

$$s^2 = \frac{1}{n-1} \sum (i - \bar{x})^2 \quad (4)$$

$$s = \sqrt{s^2} \quad (5)$$

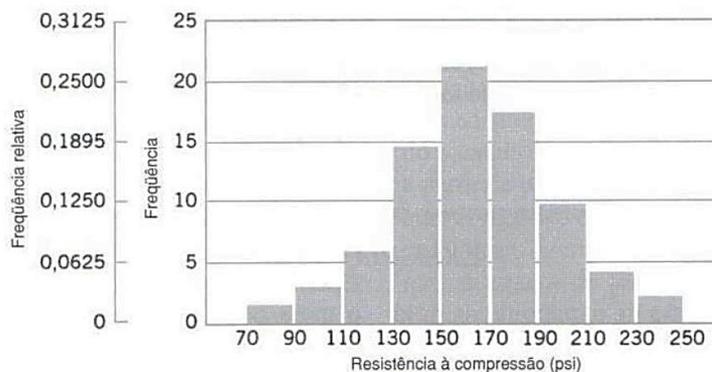
$$CV = \frac{s}{\bar{x}} 100 \quad (6)$$

Onde  $i$  é a observação,  $\bar{x}$  é a média das observações e  $n$  o tamanho da amostra.

Um tipo de gráfico que auxilia na interpretação dos dados são os histogramas. A construção de histogramas dos dados fornece uma informação visual de como os dados estão distribuídos, além da tendência central e a dispersão em torno desse valor; trata-se de uma disposição visual da distribuição de frequências e auxilia na escolha, por exemplo, de modelos de distribuição de probabilidades (MONTGOMERY;

RUNGER, 2012). A Figura 6 ilustra um exemplo histograma, cuja dispersão em torno da tendência central assemelha-se a uma distribuição normal.

Figura 6 – Exemplo de histograma

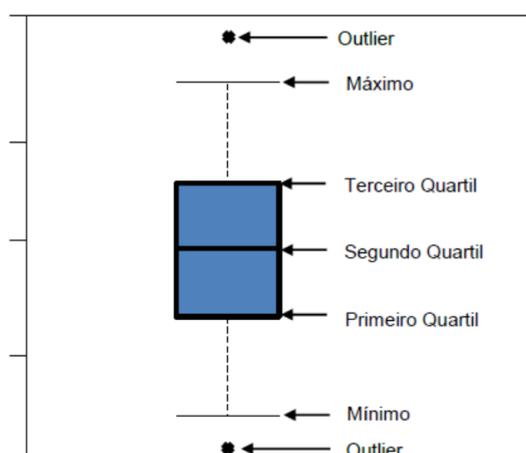


Fonte: Montgomery e Runger (2012).

Salientam Montgomery e Runger (2012), no entanto, que histogramas são sensíveis à quantidade de dados e ao número de intervalos (ou número de barras) definido; destacam que histogramas são mais estáveis para um número de dados acima de 75 e que, na prática, uma ordem de grandeza para o número de intervalos pode ser tomada como a raiz quadrada do número de observações.

Outro gráfico que auxilia nessa fase de reconhecimento dos dados são os diagramas de caixa, ou também chamados de *box-plots*. Trata-se de um diagrama que apresenta de forma gráfica informações como a mediana, a dispersão dos dados em torno desse valor, o desvio da simetria e a existência de eventuais valores discrepantes (ou *outliers*). A Figura 7 ilustra um exemplo de diagrama de caixa.

Figura 7 – Exemplo de diagrama de caixa



Fonte: Oliveira Filho (2019).

No diagrama de caixa, destacam-se os três quartis, os limites máximo e mínimo e, se houver, os *outliers* (observações discrepantes). A caixa delimitada pelo primeiro e terceiro quartis é chamada de faixa interquartil. A mediana dos dados é desenhada no diagrama na forma da linha que marca o segundo quartil. As linhas tracejadas na Figura 7, que se estendem da faixa interquartil até os limites máximo e mínimo, representam o maior e o menor valor observado no conjunto de dados, dentro da faixa 1,5 interquartil, respectivamente.

Ocasionalmente, símbolos diferenciados são utilizados para identificar pontos individuais mais afastados da caixa, fora do intervalo definido pelo limite superior e inferior; um ponto assim é o chamado *outlier*. Um dado discrepante no conjunto pode indicar um erro de amostragem, por exemplo, e sua eliminação auxilia a remover observações infieis (CREVELIN, 2018).

#### 2.4.1.2 Regressão linear simples e múltipla

A regressão é uma técnica estatística para investigar e modelar relações entre variáveis e largamente utilizada para problemas em diversas áreas, incluindo engenharia, física, economia etc. (MONTGOMERY; PECK; VINING, 2012). Uma equação de regressão objetiva prever a relação existente entre a variável resposta e variáveis preditoras (ou regressoras). A equação que envolve apenas uma variável preditora, numa relação linear entre as variáveis, é chamada de regressão linear simples e é escrita na forma da Equação 7.

$$y = \beta_0 + \beta_1 x + \varepsilon \quad (7)$$

Onde  $y$  é a variável resposta,  $x$  a variável preditora,  $\beta_0$  é o intercepto,  $\beta_1$  é o coeficiente angular da reta de regressão e  $\varepsilon$  é o erro estatístico ou resíduo.

Nem sempre a relação entre as variáveis será previamente conhecida a partir de relações teóricas. Quando não há conhecimento teórico prévio entre as variáveis, um diagrama de dispersão pode auxiliar na escolha da equação de regressão que melhor descreva a relação matemática entre as variáveis, se linear, exponencial, logarítmica etc.

Além de conhecer previamente a forma com as variáveis se relacionam, é importante escolher regressores que estejam realmente relacionadas com o problema, em um sentido de causa. Por isso, o conhecimento sobre o assunto do problema é fundamental para evitar desenvolver correlações que não sejam estatisticamente significantes.

As equações construídas são válidas apenas para os intervalos dos dados originais. Assim sendo, utilizar uma equação de regressão para valores das variáveis fora do intervalo considerado, pode levar a uma extrapolação indevida da equação e comprometer a validade do resultado (MONTGOMEY; RUNGER, 2012).

Feitas essas ressalvas, para a construção de uma regressão é necessário estimar os valores dos coeficientes  $\beta$ . Utiliza-se o Método dos Mínimos Quadrados para este fim. Considerando uma regressão linear simples, o equacionamento dos coeficientes  $\beta$  fica:

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \quad (8)$$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n x_i y_i - \frac{(\sum_{i=1}^n y_i)(\sum_{i=1}^n x_i)}{n}}{\sum_{i=1}^n x_i^2 - \frac{(\sum_{i=1}^n x_i)^2}{n}} \quad (9)$$

Onde  $y$  é a variável resposta,  $x$  a variável preditora,  $\bar{x}$  e  $\bar{y}$  são as médias das amostras para a variável  $x$  e  $y$ , respectivamente,  $n$  é o número de observações.

Equações que envolvem mais de uma variável regressora são chamadas de regressão múltipla. Quando a função entre as variáveis for linear, usa-se o termo regressão linear múltipla e segue a equação para  $k$  variáveis regressoras:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \varepsilon \quad (10)$$

Equações mais complexas, como funções polinomiais de grau  $j > 1$ , podem ser analisadas por técnicas de regressão linear múltipla, utilizando uma mudança de variável  $x_i = x^j$ . A mesma estratégia pode ser aplicada em equações que incluem efeitos de interação entre variáveis, como o produto de regressores. Nestes casos, as superfícies gráficas resultantes não serão hiperplanos.

### 2.4.1.3 Avaliação da regressão

Após o desenvolvimento de uma equação de regressão é necessário avaliar sua validade e, assim, definir se é adequada matematicamente para expressar a relação entre a variável resposta e as variáveis preditoras (SANTOS, 2017).

Diversos parâmetros estatísticos são utilizados nessa análise com o objetivo de avaliar os resultados fornecidos pela regressão. Como exemplo, pode-se citar o coeficiente de determinação ( $R^2$ ), a raiz do erro quadrático médio (RMSE) e o erro médio absoluto (MAE). Todavia, a análise de regressão deve ser complementada por inferência estatística, com o uso dos testes de hipóteses, incluindo uma análise dos resíduos.

Muito utilizado em análise de regressão, o coeficiente de determinação é referido como a quantidade de variabilidade nos dados que é explicada ou considerada pela equação de regressão (MONTGOMERY; RUNGER, 2012). Sendo assim, o valor de  $R^2$  pode variar de 0 a 1. Valores próximos a 1 implicam que a maior parte da variabilidade da variável resposta é explicada pela regressão. A Equação 11 é empregada para calcular  $R^2$ :

$$R^2 = 1 - \frac{\sum_{i=1}^n (k_{sat,estimado\ i} - k_{sat,medido\ i})^2}{\sum_{i=1}^n (k_{sat,medido\ i} - \bar{k}_{sat,medido\ m})^2} \quad (11)$$

Onde  $\bar{k}_{sat,medido\ m}$  é o valor médio das entradas de  $k_{sat}$  medidos em laboratório,  $k_{sat,medido\ i}$  representa o valor medido em laboratório de  $k_{sat}$  e  $k_{sat,estimado\ i}$  é o valor de  $k_{sat}$  estimado.

Contudo,  $R^2$  deve ser utilizado e interpretado com cautela, pois seu valor tende a aumentar à medida que se adicionam observações suficientes à análise. Outros parâmetros estatísticos também devem ser associados ao coeficiente de determinação ajustado e auxiliar na avaliação da equação.

A raiz do erro quadrático médio (*root mean squared error* – RMSE) é um outro parâmetro estatístico que pode auxiliar na interpretação da validade da regressão. É calculado como a raiz quadrada do somatório das diferenças quadradas entre os valores estimados e medidos, dividido pelo número total de observações. Valores mais próximos de zero indicam melhor desempenho da regressão. A vantagem da

RMSE é que um maior peso é atribuído aos maiores erros da equação de regressão. RMSE é calculada de acordo com a Equação 12.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (k_{sat, estimado\ i} - k_{sat, medido\ i})^2} \quad (13)$$

O erro médio absoluto (*mean absolute error* – MAE) é definido como a média aritmética de todos os erros da equação de regressão, isto é, é o módulo da diferença entre os valores estimados e medidos dividido pelo número total de observações. Por representar um erro médio dos resultados da equação, quanto mais próximo de zero, melhor o desempenho da regressão. É calculado pela Equação 13:

$$MAE = \frac{1}{n} \sum_{i=1}^n |k_{sat, estimado\ i} - k_{sat, medido\ i}| \quad (14)$$

Uma vez desenvolvido a equação de regressão, também é necessário verificar o nível de significância estatística da relação. Para analisar se a equação proposta é adequada para expressar a relação matemática entre as variáveis resposta e preditoras, utilizam-se de testes estatísticos de hipóteses, que pertencem a um ramo da Estatística chamado de Inferência Estatística.

Segundo Crevelin (2018), Inferência Estatística é a conclusão acerca de hipóteses sobre uma população, a partir da análise de dados amostrais. No teste de hipótese, duas condições são inicialmente estabelecidas: aceitar ou rejeitar a hipótese. Matematicamente, essa condição de aceitação/rejeição é representada por uma hipótese nula,  $H_0$ , e por uma hipótese alternativa,  $H_1$ , excludentes entre si, isto é: aceitando-se  $H_0$ , rejeita-se  $H_1$  e vice-versa.

A aceitação ou rejeição é realizada pela comparação entre o resultado do teste e um valor crítico pré-estabelecido, que normalmente fornece uma medida da credibilidade da hipótese nula,  $H_0$  (MONTGOMERY; RUNGER, 2012). Esse valor crítico é definido em termos do chamado valor-p, que está associado ao nível de significância que conduz à aceitação da hipótese nula, e geralmente é considerado igual à 0,05.

Dessa forma, nos testes de hipóteses, calcula-se o valor-p e o compara com o nível de significância adotado. Tal comparação conduzirá à rejeição de  $H_0$ , caso o valor-p seja superior ao nível de significância estabelecido.

Para a regressão proposta ser considerada *estatisticamente significativa*, alguns critérios devem ser atendidos. Como são requisitos que devem ser verdadeiros, os testes de hipóteses ajudam a confirmar e a conduzir as análises sobre a validação e atendimento destes critérios:

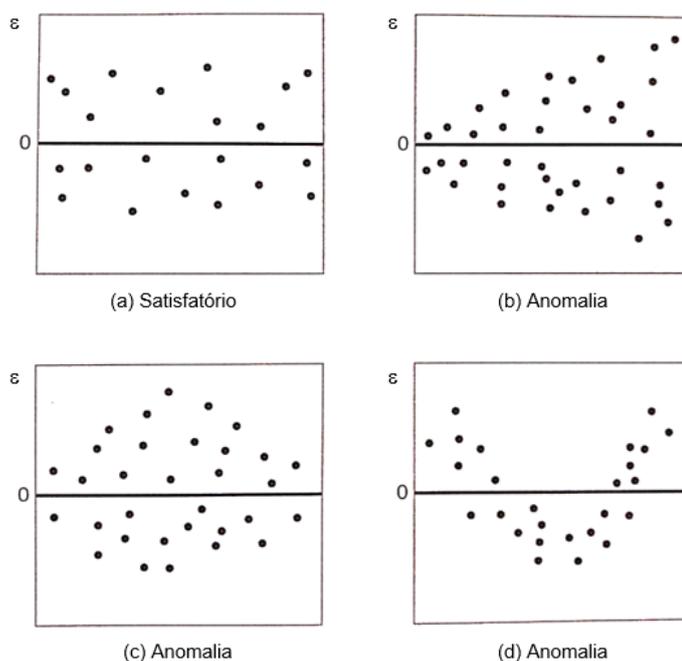
- (1) os termos da equação de regressão devem, de fato, estar contribuindo para o resultado da regressão;
- (2) a relação matemática proposta entre a variável resposta e as variáveis preditoras deve ser verdadeira, isto é, o fenômeno deve realmente se comportar de maneira linear, por exemplo;
- (3) os resíduos devem ser variáveis aleatórias independentes (não correlacionadas), com média zero e variância constante (homoscedasticidade);
- (4) os resíduos devem ser normalmente distribuídos.

O atendimento ao critério (1) pode ser verificado através da aplicação do teste  $t$  (em regressões lineares simples) ou pelo teste  $F$  parcial (em regressões lineares múltiplas), que, quando aplicado para uma única variável, é equivalente ao teste  $t$  (MONTGOMERY; RUNGER, 2012). Testar individualmente os termos da regressão é útil para produzir uma equação mais efetiva, seja a partir da retirada de variáveis que não sejam importantes para o resultado fornecido pela regressão ou pela inclusão de preditores que melhor contribuam para a previsão da variável resposta.

O segundo critério é verificado por meio da aplicação do teste  $F$ . Para regressões lineares, o objetivo do teste  $F$  é verificar se existe uma relação linear entre as variáveis. O procedimento é geralmente sumarizado em uma tabela de análise de variância (ANOVA). Além disso, tanto o teste  $t$  e o teste  $F$ , normalmente, são fornecidos na saída dos programas computacionais de regressão múltipla.

Para o atendimento aos critérios (3) e (4), deve ser realizada uma análise dos resíduos da regressão, que contempla uma parte gráfica e testes de hipóteses. Quando plotados os resíduos, em sequência temporal (se conhecida), *versus* os valores da variável resposta preditos e variáveis regressoras, padrões podem ser revelados na dispersão e indicar a existência de *outliers* ou que os resíduos seguem uma tendência (não são independentes). A Figura 8 ilustra alguns padrões que podem ser observados no Diagrama de Resíduos.

Figura 8 – Padrões do diagrama de resíduos



Fonte: Adaptado de Montgomery e Runger (2012).

Grandes resíduos indicam prováveis *outliers* nos dados, mas regras devem ser observadas para descartar esses valores, pois podem fornecer também informações de interesse sobre circunstâncias experimentais não usuais. A presença de valores discrepantes podem interferir na validade da regressão. Se os resíduos têm distribuição normal, então aproximadamente 95% dos resíduos padronizados devem estar no intervalo de  $(-2, +2)$ . Por resíduo padronizado, entende-se pelo resíduo da observação  $i$  dividido pela raiz quadrada da variância dos resíduos.

Quando ocorrem nos valores extremos do diagrama, podem ser um indicativo de que a variância não seja constante ou que a relação matemática entre as variáveis não seja linear. Estas possibilidades devem ser investigadas antes que os pontos sejam considerados *outliers* (MONTGOMERY; PECK; VINING, 2012; MONTGOMERY; RUNGER, 2012).

Como a análise gráfica do diagrama de resíduos pode ser subjetiva, o diagnóstico da independência dos resíduos pode ser realizado pelo teste de Durbin-Watson (DW). Esse teste objetiva testar se existe ou não correlação serial entre resíduos consecutivos, em outras palavras, verificar se o erro  $\varepsilon_i$  influencia no erro  $\varepsilon_{i+1}$ . As hipóteses são:

$H_0$ : Não existe autocorrelação dos resíduos

$H_1$ : Existe autocorrelação positiva dos resíduos

A estatística dada pela Equação 14 é comparada com valores críticos  $d_L$  e  $d_U$ , tabelados, e, então, a decisão sobre aceitação ou rejeição sobre a hipótese nula pode ser tomada.

$$d_w = \frac{\sum_{i=2}^n (\varepsilon_i - \varepsilon_{i-1})^2}{\sum_{i=1}^n (\varepsilon_i)^2} \quad (15)$$

A estatística de DW pode assumir valores entre  $0 \leq d_w \leq 4$ . O Quadro 2 orienta quanto à decisão sobre a hipótese nula, em função dos valores de  $d_w$ ,  $d_L$  e  $d_U$ .

Quadro 2 – Tabela de decisão em função de  $d_U$  e  $d_L$

Zona de Rejeição e de não-rejeição de $H_0$					
$d_w$	$[0; d_L[$	$[d_L; d_U[$	$[d_U; 4 - d_U[$	$[4 - d_U; 4 - d_L[$	$[4 - d_L; 4]$
Decisão	Rejeitar $H_0$	Nada se pode concluir	Não rejeitar $H_0$	Nada se pode concluir	Rejeitar $H_0$
	Auto-correlação positiva		Os resíduos são independentes		Auto-correlação negativa

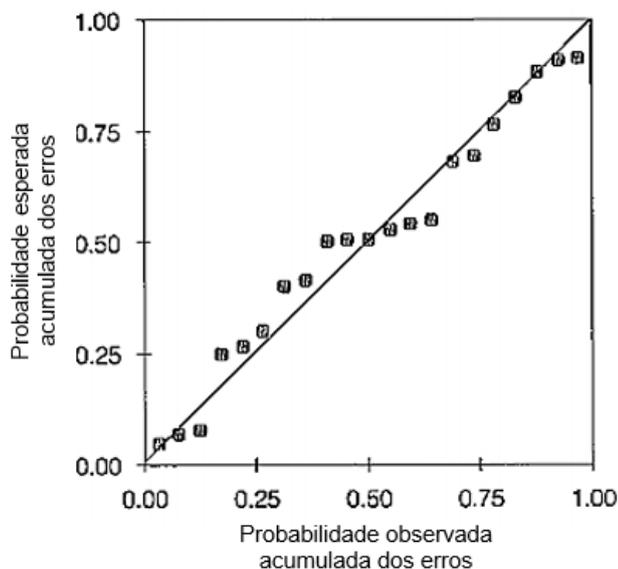
Fonte: Maroco (2007).

De acordo com o número de termos da equação (incluindo a constante) e o número de observações, são definidos os valores críticos de  $d_L$  e  $d_U$ , com os quais  $d_w$  deverá ser comparado. O Anexo 1 apresenta um quadro com valores de referência.

Para confirmar a distribuição normal dos resíduos (critério nº 4), tanto uma análise gráfica quanto testes de normalidade podem ser aplicados. A plotagem de um histograma dos resíduos deve apresentar um gráfico em formato de um sino. Contudo, para uma pequena quantidade de observações, a plotagem do histograma pode ficar comprometida. Para esses casos, é recomendado a plotagem da probabilidade normal.

Neste gráfico de probabilidade normal, a probabilidade observada acumulada dos erros é plotada nas abcissas e a probabilidade acumulada que se observaria, caso os resíduos apresentassem distribuição normal, é plotada nos eixos das ordenadas (MAROCO, 2007). A Figura 9 mostra um exemplo do gráfico de probabilidade.

Figura 9 – Gráfico de probabilidade dos resíduos



Fonte: Adaptado de Maroco (2007).

Se os resíduos apresentam distribuição normal, as observações registradas se aproximam da reta diagonal do gráfico, com nenhum afastamento significativo. Todavia, como toda análise gráfica, pode ser de interpretação subjetiva e, assim, os testes de hipóteses complementam a verificação da normalidade dos resíduos.

Há diversos testes de hipóteses que verificam a distribuição normal dos resíduos. Como exemplo, pode-se citar: Shapiro-Wilk (SW), Lilliefors (LF), Kolmogorov-Sminov (KS), Anderson-Darling (AD) e Shapiro-Francia (SF). Em geral, esses testes apresentam estatísticas de teste e critérios de decisão diferentes, mas as hipóteses testadas são as mesmas:

$H_0$  – a variável apresenta distribuição normal

$H_1$  – a variável não apresenta distribuição normal.

A decisão pela aceitação ou rejeição da hipótese nula se dá pela observação do valor-p do teste: se o valor-p retornado do teste for menor que o nível de significância adotado, rejeita-se a hipótese de normalidade (TORMAN; COSTER; RIBOLDI, 2012).

O critério para a seleção do teste de hipótese se dá pela eficiência do teste, que é influenciada, principalmente, pelo tamanho da amostra. Para amostras entre 4 e 30 observações, recomenda-se o uso dos testes de Shapiro-Wilk e Shapiro Francia. Amostras maiores que 50 observações, o teste de Anderson-Darling apresenta um

bom desempenho, assim como o teste de Lilliefors, que é uma correção aplicada ao teste de Kolmogorov-Smirnov. Este último, segundo Maroco (2007), não é recomendado para aplicações nas quais não se conhece a média nem o desvio padrão populacionais, sob o risco de incorrer no erro Tipo I (rejeitar a  $H_0$  quando ela for verdadeira).

Em um estudo comparativo entre testes de hipóteses para diferentes tamanhos de amostras, Torman, Coster e Riboldi (2012) perceberam elevadas taxas de erro Tipo I para o teste de Kolmogorov-Smirnov. Os pesquisadores verificaram melhor percentual de acerto para os testes Shapiro-Wilk e Shapiro-Francia. No estudo comparativo, foram considerados os testes Qui-Quadrado, Kolmogorov-Smirnov, Lilliefors, Shapiro-Wilk, Shapiro-Francia, Cramer-von Mises, Anderson-Darling e Jarque-Bera. Todos apresentaram taxa de acerto média de 60% para amostras com tamanho superior a 50 observações, exceto Kolmogorov-Smirnov, como já mencionado.

Por fim, em casos de regressão linear múltipla, ainda existem outros parâmetros que devem ser avaliados para verificar a qualidade da regressão: a existência de multicolinearidade. Basicamente, a multicolinearidade está presente na equação quando existe uma relação linear entre uma das variáveis independentes e as restantes (RODRIGUES, 2012).

A existência da multicolinearidade pode não afetar substancialmente o ajuste da equação (o valor de  $R^2$ , por exemplo), mas a sobrecarregará, sem ganhos significativos na qualidade (MAROCO, 2007). Para determinar se existe multicolinearidade pode ser feita a determinação do Fator de Inflação da Variância (do inglês, *variance inflation factor* – VIF) (RODRIGUES, 2012).

A determinação do valor de VIF se dá a partir do cálculo da variância, pela Equação 15, de cada um dos coeficientes da regressão linear múltipla,  $\hat{\beta}_i$ , associados às variáveis independentes:

$$var(\hat{\beta}_i) = \sigma^2 \left( \frac{1}{1-R_i^2} \right) \times \frac{1}{\sum_{j=1}^n (x_{ij} - \bar{x}_i)^2} \quad (16)$$

Onde  $R_i^2$  é o coeficiente de determinação da regressão entre a variável  $x_i$  (como variável dependente) e sobre as demais variáveis  $x_b$  ( $b \neq i = 1, \dots, p$ ). Essa variância é tanto maior quanto maior a correlação entre as variáveis independentes, pois, se a

variável  $x_i$  for fortemente correlacionada com as demais variáveis, o valor de  $R_i^2$  tende a estar próximo da unidade e o termo  $\left(\frac{1}{1-R_i^2}\right)$  “infla” o valor de  $var(\hat{\beta}_i)$ .

O termo  $\left(\frac{1}{1-R_i^2}\right)$  é o chamado VIF para o coeficiente de regressão  $\hat{\beta}_i$  associado à variável  $x_i$ . Segundo Montgomery, Peck e Vining (2012), valores de VIF maiores que 10 podem indicar problemas com multicolinearidade.

Por fim, cabe uma ressalva quanto à interpretação do valor do VIF. A depender da equação de regressão analisada, um alto valor do VIF para algum coeficiente não necessariamente significa um problema de multicolinearidade que prejudique a predição. Por exemplo, em funções polinomiais, que envolvam o produto entre variáveis, é normal que uma alta correlação exista entre os termos, mas cuja multicolinearidade não necessariamente impacte no desempenho da equação (JOHNSTON; JONES; MANLEY, 2018).

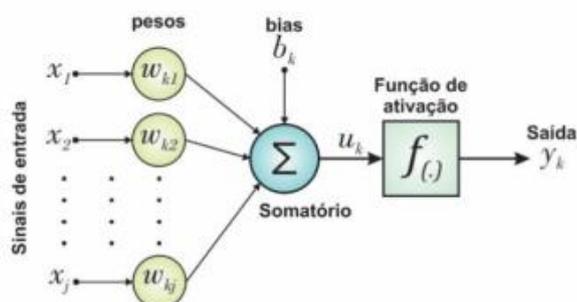
#### 2.4.2 Redes neurais artificiais

Dentre as técnicas alternativas para tarefas de predição de uma dada propriedade, uma rede neural artificial (RNA) é uma técnica computacional que simula o funcionamento do cérebro humano, com sua rede de neurônios e como eles estão interconectados, podendo ser escrita a partir de um algoritmo (HAYKIN, 2001; SHAHIN; JASKA; MAIER, 2001; AKBULUT, 2005). O objetivo de uma RNA é imitar como o cérebro humano realiza uma tarefa específica.

A RNA baseia-se numa unidade elementar conhecida por *neurônio artificial*, chamados de nós ou elementos de processamento, que são dispostos em *camadas*. Os precursores das redes neurais e dos neurônios artificiais foram McCulloch e Pitts, em 1943, que propuseram o neurônio MCP (MCCULLOCH; PITTS, 1990). Mas foi o trabalho de Rosenblatt (1958), que apresentou o modelo de aprendizagem supervisionada a partir de um neurônio chamado *perceptron*, que permitiu a construção das redes *perceptron* de múltiplas camadas, utilizadas nesta pesquisa.

A principal diferença entre o perceptron de Rosenblatt (1958) e o neurônio MCP é a possibilidade de associar mais de um neurônio, conectados entre si. O esquema do neurônio *perceptron* é mostrado na Figura 10.

Figura 10 - Modelo esquemático do *perceptron* de Rosenblatt (1958)



Fonte: Adaptado de Haykin (2001, *apud* OLIVEIRA FILHO, 2019)

Da Figura 10,  $x_i$  a  $x_j$  representam as variáveis de entrada, que são os atributos de pertencentes a um banco de dados. Cada entrada é multiplicada por um respectivo peso sináptico,  $w_{ki}$ . O índice  $k$  representa a identificação do neurônio; isso porque, como será visto a seguir, o *perceptron* possibilita a associação de diversos neurônios para a construção de uma rede (a rede neural). No núcleo do *perceptron*, é realizada a soma desses produtos  $x_i \times w_{ki}$  e incorporado um *bias* (leia-se *viés*).

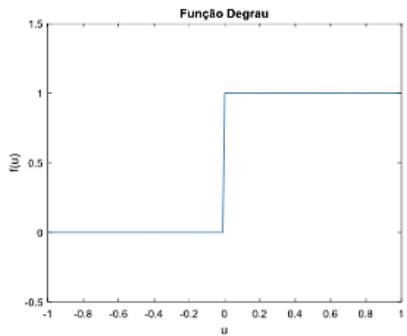
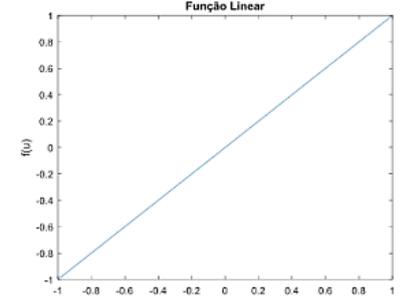
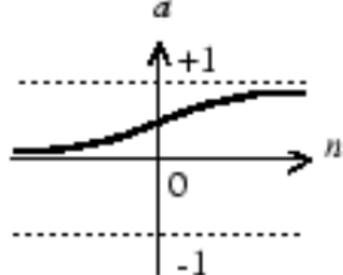
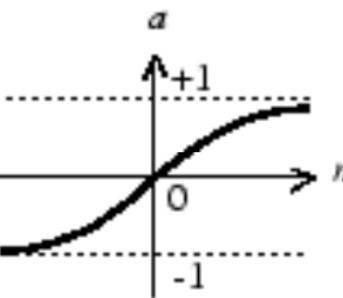
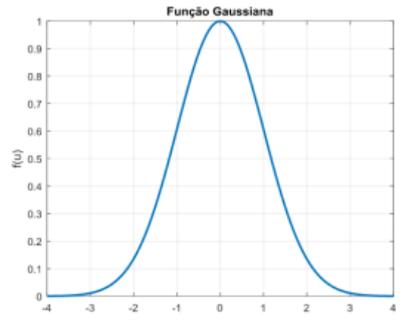
A soma resultante,  $u_k$ , é então aplicada à função de ativação e a saída  $y_k$  é produzida. Matematicamente, as operações realizadas pelo *perceptron* podem ser representadas pelas equações seguintes.

$$u_k = \sum_{i=1}^j (x_i \times w_{ki}) + b \quad (17)$$

$$y_k = f(u_k) \quad (18)$$

A função de ativação é a responsável por produzir uma resposta em termos do somatório  $u_k$ . Oliveira Filho (2019) reuniu e apresentou cinco tipos principais de função de ativação: linear, gaussiana, degrau, logística e tangente sigmóide. O Quadro 3 resume a função e o gráfico de cada tipo de função.

Quadro 3 - Resumo das funções de ativação

Tipo	Função	Gráfico
Degrau	$f(u_k) = \begin{cases} 1, & \text{se } u_k \geq 0 \\ 0, & \text{se } u_k < 0 \end{cases}$	
Linear <sup>(1)</sup>	$f(u_k) = \alpha \times u_k$	
Logística <sup>(1)</sup>	$f(u_k) = \frac{1}{1 + e^{-\alpha \times u_k}}$	
Tangente hiperbólica <sup>(1)</sup>	$f(u_k) = \frac{e^{\alpha \times u_k} - e^{-\alpha \times u_k}}{e^{\alpha \times u_k} + e^{-\alpha \times u_k}}$	
Gaussiana <sup>(2)</sup>	$f(u_k) = e \times \frac{-(u_k - \mu)^2}{r^2}$	

(1)  $\alpha$  é o coeficiente angular limitante dos valores de  $f$  entre  $[0;1]$  ou  $[-1;1]$ ;

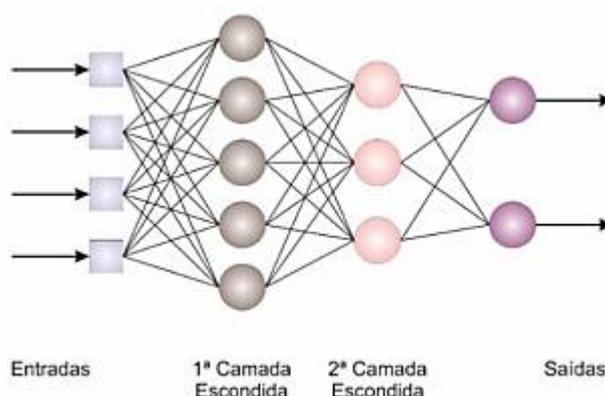
(2) onde  $\mu$  é o centro (ponto médio) e  $r$  é o raio de abertura da função.

Fonte: Adaptado de Oliveira Filho (2019).

Ao associar diversos *perceptrons*, constrói-se a *rede* de neurônios artificiais. Uma rede é formada, então, por uma quantidade de neurônios que são conectados entre si, segundo um arranjo definido pelo analista e conhecido por *arquitetura*. A arquitetura define a quantidade de neurônios envolvidos, a quantidade de camadas de neurônios e a quantidade de atributos de entrada na rede.

A RNA é formada por 03 tipos de camadas: a de entrada (que contém variáveis de entrada do problema), a camada intermediária (que contém os neurônios) e a camada de saída (que contém o dado de saída da rede). Porém, uma rede pode também apresentar mais de uma camada intermediária, também chamadas de camadas ocultas. Neste caso, a rede é chamada de multicamadas, pois apresenta mais de uma camada de neurônios conectadas entre si por pesos sinápticos, recebendo o nome de *rede perceptron multicamadas*. Os neurônios de cada camada se ligam aos neurônios da camada seguinte, segundo pesos sinápticos diferentes e que se atualizam a cada ciclo de iteração da rede, como será explicado adiante (AKBULUT, 2005). A Figura 11 ilustra um desenho esquemático deste tipo de rede.

Figura 11 - Esquema de rede multicamadas



Fonte: Adaptado de Haykin (2001, *apud*. OLIVEIRA FILHO, 2019).

O fluxo de treinamento de uma RNA se inicia com o fornecimento de um conjunto de dados com os parâmetros de entrada e as respectivas saídas esperadas à primeira camada de neurônios da rede. A informação é passada de camada para camada até que se produza um sinal de saída, que será, então, comparado ao valor esperado e o erro na predição é, então, calculado. A rede ajusta os pesos sinápticos, de acordo com o algoritmo de aprendizagem escolhido, até que se atinja uma condição pré-

definida para fim do treinamento, por exemplo, número de ciclos de treinamento e/ou erro aceitável (SHAHIN; JAKSA; MAIER, 2001).

Após treinada, a rede terá seu desempenho avaliado e poderá, então, ser utilizada na predição daquele parâmetro de saída específico. A tarefa de treinamento de uma rede neural objetiva encontrar uma função que relaciona os dados de entrada com os dados de saída, similar a uma equação de regressão. Contudo, ao contrário das tarefas estatísticas de regressão, não é necessário conhecimento prévio sobre a relação de linearidade ou não entre os dados. Além disso, o grau de não linearidade é facilmente ajustado ao longo do treinamento, o que confere às RNA grande aplicabilidade à problemas complexos e com alto grau de não linearidade.

A forma como a rede é organizada está relacionada ao tipo de algoritmo de aprendizagem que será utilizado para treinar a rede. Definir a arquitetura é um dos grandes desafios para os analistas, pois essa não é uma tarefa trivial. Além disso, a quantidade de camadas intermediárias e a quantidade de neurônios em cada camada afeta o desempenho final da rede e não há nenhum método analítico para obter o número ideal de camadas e neurônios em cada uma delas.

Em geral, redes com uma ou duas camadas intermediárias, desde que com um número adequado de neurônios, têm se mostrado suficientes para muitos problemas (TREJO-ALONSO *et al.*, 2021). Entretanto, o processo de modelagem da arquitetura ideal, isto é, que minimize o erro entre resultado fornecido pela rede e o resultado esperado, é realizado por calibração.

Outro aspecto importante sobre as redes neurais são os algoritmos de aprendizagem. Esses algoritmos são como as redes aprendem. Segundo Goldschmidt e Passos (2005) e Akbulut (2005), a aprendizagem da rede se dá por treinamento e, durante esse processo de aprendizagem, os pesos sinápticos e o *bias* (leia-se *viés*) vão sendo atualizados até que o resultado esperado e o fornecido pela rede difiram pouco um do outro.

O procedimento de como o treinamento da rede ocorrerá é feito pelo algoritmo de aprendizagem. Simplificando, os algoritmos são os códigos que dirão o que fazer com os dados de entrada, como processá-los, como aplicá-los na função de ativação e o que fazer com as saídas geradas.

Existem diversos algoritmos de aprendizagem e a diferença básica entre eles é a forma como serão atualizados os pesos sinápticos (HAYKIN, 2001). Os mais comuns são os algoritmos de retropropagação de erro (ou algoritmo retropropagação de erro) e o algoritmo Levenberg-Marquardt.

O algoritmo retropropagação de erro calcula o erro entre a saída da rede (valor estimado) e o valor esperado (registrado no banco de dados). Quanto menor essa diferença, mais treinada está a rede. Voltando ao erro, pode-se escrever a seguinte equação:

$$e_j(n) = d_j(n) - y_j(n) \quad (19)$$

Onde  $e_j(n)$  é o erro na saída do neurônio  $j$ , na  $n$ -ésima iteração. Cada iteração é a apresentação de um conjunto de dados à rede. O valor esperado na saída do neurônio  $j$ , na  $n$ -ésima iteração é  $d_j(n)$  e o valor fornecido pela rede, por este mesmo neurônio e nessa mesma iteração é  $y_j(n)$  (HAYKIN, 2001). Com base neste erro, o algoritmo recalcula o valor dos pesos sinápticos a partir da última para a primeira camada, num processo de trás para frente, daí o nome retropropagação de erro (WILLIAMS; OJURI, 2021).

O algoritmo verifica o quanto cada camada intermediária contribuiu para o erro na camada de saída. Essa retropropagação de erro é realizada com o auxílio do gradiente descendente da função erro. Como se trata de um fluxo que envolve derivadas parciais, é importante que as funções de ativação sejam diferenciáveis (OLIVEIRA FILHO, 2019).

Com novos pesos sinápticos, o fluxo é reiniciado (uma nova iteração é iniciada) e novas saídas são fornecidas, uma vez que os pesos sinápticos foram atualizados. Novamente, são comparados os valores fornecidos pela rede e os valores esperados. O erro é, então, calculado e retropropagado a fim de corrigir os pesos sinápticos.

Esse processo iterativo continua até que um critério de parada seja alcançado; esse critério é definido pelo analista e usualmente é um erro mínimo, considerado aceitável, ou um número máximo de iterações. As formulações matemáticas da sequência do algoritmo de retropropagação de erro podem ser consultadas em Haykin (2001) e Goldschmidt e Passos (2005).

O algoritmo Levenberg-Marquardt é uma melhoria do algoritmo de retropropagação de erro e baseia-se na minimização do erro pelo método de Newton, a fim de tornar a convergência mais ágil.

Segundo Diminski (2000), apesar do algoritmo Levenberg-Marquardt ser mais eficiente que o de retropropagação de erro, ele possui um inconveniente de necessitar de maior capacidade de armazenamento de memória computacional, principalmente quando se trabalha com um grande volume de dados. Trata-se, no entanto, de uma alternativa interessante haja vista o ganho de eficiência e tempo de processamento.

As redes neurais possuem capacidades que as tornam uma importante ferramenta na análise de grandes bancos de dados. Goldschmidt e Passos (2005) pontuam: (1) a habilidade de realizar trabalho de forma paralela; (2) a capacidade de aprender por exemplos e apenas a partir de dados à ela apresentados; (3) possibilidade de lidar com ruídos e dados distorcidos e, assim, ser capaz de generalizar os resultados; (4) habilidade de realizar associações e abstrações, isto é, de forma autônoma, a RNA é capaz de estabelecer relações entre padrões completamente distintos e de identificar as características de maior relevância dos dados; (5) por fim, apresentam robustez e degradação gradual, o que permite que se ajustem em caso de perda de um neurônio, por exemplo.

#### 2.4.3 Aplicações de inteligência artificial na previsão de propriedades hidráulicas de solos

A seguir, algumas pesquisas que envolveram a aplicação e avaliação do desempenho de técnicas de inteligência artificial para determinação da permeabilidade dos solos, em especial as técnicas abordadas nesta pesquisa, são apresentadas. O objetivo da apresentação dessas pesquisas é evidenciar as técnicas utilizadas, os parâmetros considerados, as limitações e considerações feitas pelos pesquisadores a fim de direcionar esta pesquisa.

Singh *et al.* (2020) propuseram utilizar algoritmos de processamento de dados para desenvolver técnicas para predição do coeficiente de permeabilidade de solos saturados baseados na análise textural (argila, silte ou areia), densidade real dos grãos, peso específico total e teor de carbono orgânico. Para tanto, após uma revisão de literatura, os pesquisadores optaram por aplicar cinco diferentes algoritmos na

previsão da permeabilidade de solos: *perceptron* de multicamadas (MLP) com algoritmo Levenberg-Marquardt, Máquinas de Vetores de Suporte (SVM), sistema de inferência neuro-*fuzzy* coativado (CANFIS), floresta aleatória e árvore de decisão. Também foram investigados os resultados entre algoritmos após a aplicação da transformada *Wavelet*, uma técnica de pré-processamento para identificação de distúrbios nos dados (BORGES, 2013).

Os resultados alcançados por Singh *et al.* (2020) mostraram que, das técnicas de aprendizado de máquinas empregadas, a floresta aleatória apresentou os melhores resultados e melhor desempenho para previsão do coeficiente de permeabilidade, seja fase de treinamento seja na fase de validação, seguida da árvore de decisão, Máquinas de Vetores de Suporte, sistema de inferência neuro-*fuzzy* coativado e *perceptron* de multicamadas com algoritmo Levenberg-Marquardt. Além disso, todas as técnicas apresentaram melhor desempenho quando submetidas à otimização com transformada *Wavelet*.

Os pesquisadores não detalharam as arquiteturas das redes que apresentaram melhores desempenhos, mas salientaram que todas as técnicas apresentaram resultados satisfatórios para previsão do coeficiente de permeabilidade das amostras ensaiadas.

Outra pesquisa que aplicou RNAs para a previsão do coeficiente de permeabilidade saturada foi desenvolvida por Williams e Ojuri (2021). Os pesquisadores desenvolveram arquiteturas de RNA, do tipo *perceptron* de multicamadas, alimentada a diante, com algoritmo de retropropagação de erro, para prever a permeabilidade e os compararam com análise de regressão linear múltipla.

Um banco de dados foi levantado a partir de pesquisas publicadas na literatura e foi composto por 144 conjuntos de dados de entrada e respectiva saída, incluindo amostras de diferentes regiões do globo. Como parâmetros de entrada para os dois métodos considerou-se o percentual de areia, de finos (material passante na peneira #200) e de argila, o índice de plasticidade e características de compactação, como umidade ótima, peso específico aparente seco máximo.

Os dados disponíveis foram divididos de maneira que 75% (108 dados) do banco de dados foi selecionado para a fase de treinamento e 25% (36 dados) utilizado para a fase de teste das arquiteturas. Os pesquisadores garantiram que cada uma dessas

frações contivessem dados que representassem de forma apropriada a ampla gama de valores de cada atributo.

Williams e Ojuri (2021) optaram por iniciar a modelagem da RNA com apenas uma camada oculta e, caso o desempenho não fosse satisfatório, aumentar-se-ia o número de neurônios ocultos antes de considerar o acréscimo de uma nova camada intermediária. No estudo realizado, o número de neurônios na camada oculta foi inicialmente adotado como metade do número de parâmetros de entrada, isto é, foram utilizados três neurônios na camada intermediária.

Após vários treinamentos, a arquitetura que apresentou o melhor desempenho foi composta da camada de entrada, uma camada intermediária com 10 neurônios e a camada de saída.

O desenvolvimento da análise de regressão se deu também, inicialmente, com os seis atributos de entrada mencionados anteriormente. Um estudo de nível de significância revelou que algumas variáveis de entrada não eram estatisticamente fortes para contribuir com o resultado da equação. Desse modo, uma nova equação foi desenvolvida adotando apenas três variáveis de entrada: índice de plasticidade, percentual de areia e peso específico aparente seco.

Uma análise de sensibilidade das variáveis na RNA também foi conduzida, utilizando o algoritmo de Garson. Ao contrário do observado na análise de regressão, o percentual de finos teve o maior impacto no valor de saída da rede enquanto o percentual de argila teve a menor influência.

As arquiteturas treinadas das RNA e as regressões foram analisadas pelos índices estatísticos: erro quadrático médio, raiz do erro quadrático médio, coeficiente de determinação e erro médio absoluto.

Ao analisar os parâmetros estatísticos que avaliam o desempenho das duas técnicas, é possível verificar que a RNA foi superior à regressão linear múltipla. Os resultados mostraram que, na fase de testes, a RNA apresentou coeficiente de determinação igual a 0,955 e a equação de regressão (com as três variáveis apenas) apresentou o valor de 0,404 para o mesmo coeficiente. Os demais parâmetros estatísticos apresentaram valores similares, porém com melhor desempenho para a RNA.

Os resultados obtidos por William e Ojuri (2021) corroboram que a RNA é uma ferramenta capaz de descrever relações complexas e não lineares entre atributos e

que, tendo um banco de dados cuidadosamente selecionado, é possível a condução da fase de treinamento e teste que permita resultados precisos da rede.

Trejo-Alonso *et al.* (2021) usaram RNA para prever o coeficiente de permeabilidade saturada no Distrito de Irrigação 023, localizado em Queretaro, no México. Para tanto, 900 amostras foram coletadas e enviadas para laboratório a fim de obter dados como: a textura do solo, pelo método do hidrômetro de Bouyucos; o peso específico total; o teor de umidade volumétrica na saturação; a capacidade de campo, isto é, a quantidade máxima de água retida nos microporos enquanto que os macroporos ficam vazios; o ponto de murcha permanente, isto é, o teor de umidade no qual uma planta não consegue mais retirar água do solo; e o coeficiente de permeabilidade do solo saturado, determinado, neste caso, pelo permeâmetro de carga variável.

Os pesquisadores treinaram várias RNA, todas com duas camadas intermediárias, e adicionalmente, uma arquitetura com três camadas ocultas. Das 900 amostras, 75% foram selecionadas para a etapa de treinamento e 25% para a etapa de teste. O processo de investigação da melhor arquitetura foi executado da seguinte forma: inicialmente, os sete atributos de entrada foram utilizados, a saber: percentual de areia, argila e silte, massa específica total, teor de umidade volumétrica na saturação, capacidade do solo e ponto de murcha permanente.

O processo de modelagem iniciou com dois neurônios na primeira camada intermediária e dois na segunda camada. Os testes de arquitetura prosseguiram, variando o número de neurônios da primeira camada intermediária de 2 a 10 e mantendo constante o número de neurônios na segunda camada.

Numa próxima etapa, aumentou-se o número de neurônios para 3 na segunda camada intermediária e, novamente, variou-se o número de neurônios na primeira camada oculta de 2 até 10. Esse procedimento foi realizado até o número de neurônios na segunda camada alcançar 10 unidades.

Finalizadas essas etapas, os pesquisadores alteraram o número de atributos na camada de entrada, diminuindo-os 1 a cada iteração, e repetindo todo o processo descrito nos parágrafos anteriores. A escolha da ordem de remoção de cada parâmetro foi construída baseada numa análise de importância de cada variável no valor do coeficiente de permeabilidade, realizada por uma função nativa do próprio *software* escolhido pelos pesquisadores.

O desempenho das inúmeras redes resultantes da pesquisa de Trejo-Alonso *et al.* (2021), na fase de modelagem, foi verificado pelos índices coeficiente de correlação, erro médio absoluto e raiz do erro quadrático médio, sendo que o valor da raiz do erro quadrático médio foi utilizado pelos pesquisadores para selecionar as melhores arquiteturas.

Em geral, os pesquisadores observaram que arquiteturas com menor quantidade de neurônios na primeira camada intermediária apresentaram resultados da raiz do erro quadrático médio melhores. Além disso, verificaram que pequenas alterações na arquitetura da rede pouco variavam nos índices de desempenho estatístico, sugerindo que os resultados da rede não são altamente sensíveis a sua configuração.

Para o estudo realizado, os valores da raiz do erro quadrático médio apresentaram diferença menor que 10% entre as diversas arquiteturas treinadas. Já os índices coeficiente de correlação e erro médio absoluto apresentaram diferença em torno de 5% e 11%, respectivamente. Os pesquisadores também notaram a tendência do uso de menos dados de entrada conduzir a resultados mais acurados do coeficiente de permeabilidade.

Trejo-Alonso *et al.* (2021), com o objetivo de explorar outros resultados, acrescentaram uma terceira camada oculta à rede e realizaram o mesmo procedimento descrito nos parágrafos anteriores. Verificaram uma melhora de 9% aproximadamente no desempenho da rede; contudo, utilizou-se 15 vezes mais tempo computacional para o treinamento, o que levou ao descarte dessas arquiteturas.

De todos os parâmetros utilizados como dados de entrada da rede, o teor de argila foi o que mais apresentou influência no valor do coeficiente de permeabilidade, seguido pelo peso específico, teor de umidade volumétrica na saturação, teor de silte, teor de areia e, por último, o ponto de murcha permanente e a capacidade do solo.

Resumiu-se, até aqui, alguns resultados do uso das redes neurais e das regressões estatísticas para a previsão de  $k_{sat}$  a partir de propriedades do solo. Outros exemplos são mostrados de forma resumida no Quadro 4 que traz as principais características de cada estudo listado: as técnicas de predição empregadas; o número de dados utilizados; quais os percentuais destinados à treinamento, teste e validação cruzada; as variáveis de entrada da rede; as variáveis de saída, caso tenham sido previstos

outras propriedades além de  $k_{\text{sat}}$ ; e os métodos estatísticos utilizados para mensurar o desempenho das arquiteturas treinadas.

Quadro 4 – Uso de inteligência artificial e métodos estatísticos para estimativa de valores de  $k_{sat}$ 

Autor(es)	Técnica(s) empregada(s)	Nº de dados utilizados	Amostras naturais?	Indefor-madas?	% Treino	% Teste	% Valid. cruzada	Variáveis de entrada	Tipo de solo	Medidas estatísticas de desempenho
Singh <i>et al.</i> (2020)	RNA-LM, Árvore de Decisão, MVS, Sistema de Inferência Neuro-fuzzy Coativado e Árvore Aleatória.	180	Sim	Não	N/I	N/I	N/I	%areia, %silte, %argila, $G_s$ , $\rho_t$ , OC(%)	Granulares e finos (amostras naturais)	coeficiente de eficiência de Nash-Sutcliffe, RMSE, coeficiente de correlação (r), índice de concordância de Willmott e índice de performance
Williams e Ojuri (2021)	RNA-BP e regressão linear múltipla	144	Sim	N/I	75%	25%	-	%areia, %finos, %argila, IP, teor de umidade ótima, %MO, $\rho_{d,máx}$	Granulares e finos (amostras naturais e misturas controladas)	MSE, RMSE, $R^2$ e MAE
Trejo-Alonso <i>et al.</i> (2021)	RNA-RBP	900	Sim	N/I	75%	25%	-	%areia, %silte, %argila, $\rho_t$ , ponto de murcha permanente, $\theta_s$ e capacidade de campo	Granulares e finos (amostras naturais)	MSE, RMSE e $R^2$
Kashani <i>et al.</i> (2020)	MVS, Aprendizado de Máquinas Extremo, <i>Splines</i> de Regressão Adaptativa Multivariada e Árvore de Regressão e RNA-BP híbrida	245	Sim	N/I	N/I	N/I	N/I	%areia, %silte, %argila, %MO, massa específica total, pH e resistividade elétrica	Granulares e finos (amostras naturais)	MSE, RMSE, coeficiente de eficiência de Nash-Sutcliffe e $R^2$
<b>Legenda auxiliar:</b> N/I – não informado										

Fonte: Autora (2022)

Quadro 4 – Uso de inteligência artificial para estimativa de valores de  $k_{sat}$  – cont.

Autor(es)	Técnica(s) empregada(s)	Nº de dados utilizados	Amostras naturais?	Indefor- madas?	% Treino	% Teste	% Valid. cruzada	Variáveis de entrada	Tipo de solo	Medidas estatísticas de desempenho
Sedaghat, Bayat e Sinagani (2016)	RNA (não menciona o algoritmo) e rede com função de base radial	260	Sim	Ambas	50%	50%	-	Razão de similaridade da distribuição do tamanho de partícula, dimensão fractal, tamanho do inicializador da distribuição de partículas, %argila, %areia e $p_t$	Granulares e finos (amostras naturais)	RMSE, critério de informação de Akaike e melhora relativa.
Akbulut (2005)	RNA (não menciona o algoritmo), regressão linear múltipla e equações empíricas (Hazen e Slichter)	95	N/I	Não	86,32%	13,68%	-	%finos, %areia, %pedregulho, $d_{10}$ , $d_{50}$ e o índice de vazios	Solos granulares (N/I)	$R^2$
Tizpa <i>et al.</i> (2014)	RNA (não menciona o algoritmo)	155	Sim	Não	90%	10%	-	módulo de finura, LL, %pedregulho, %areia, %finos e grau de compactação	Granulares e finos (amostras naturais e misturas controladas)	coeficiente de correlação (r), MSE e coeficiente de massa residual

**Legenda auxiliar:**

N/I – não informado pelo pesquisador

Fonte: Autora (2022)

Os exemplos apresentados até aqui e as outras publicações na literatura, resumidas no Quadro 4, têm, em comum, o uso de diferentes técnicas de inteligência artificial, como redes neurais, máquinas de vetores de suporte, árvore de decisão etc., para análises comparativas da previsão do coeficiente de permeabilidade, utilizando banco de dados. Além disso, por vezes, os resultados dessas técnicas são comparados com análises estatísticas dos dados, como as equações de regressão.

Métodos estatísticos como a raiz do erro quadrático médio, o erro absoluto e o coeficiente de determinação são alguns exemplos de parâmetros utilizados pelos pesquisadores para avaliarem o desempenho das redes neurais e das demais técnicas utilizadas.

Como dados de entrada para as equações de previsão do coeficiente de permeabilidade observou-se que são muito utilizados parâmetros da distribuição granulométrica, como teor de argila, silte, areia, finos,  $d_{10}$ ,  $d_{50}$  etc., e outras propriedades índices, como massa específica aparente seca, massa específica total, índice de vazios e densidade real dos grãos. Importante salientar aqui que, normalmente, não foi observada uma caracterização da natureza desse material fino, como, por exemplo, o mineral argílico presente.

Do banco de dados, normalmente, os pesquisadores separam entre 70% e 80% dos dados para a fase de treinamento da rede e o restante dos dados separados para a fase de teste. Alguns trabalhos também realizam a etapa de validação cruzada. Muitos pesquisadores, no entanto, não detalham critérios para a separação dos dados, se aleatória ou se seguem algum tipo de classificação para garantir a representatividade da seleção.

Outra característica dos estudos levantados da literatura é que a maioria dos trabalhos não apresentam o banco de dados utilizado nas modelagens, apenas um resumo estatístico contendo informações como valores máximo, médio e mínimo, e desvio padrão de cada parâmetro.

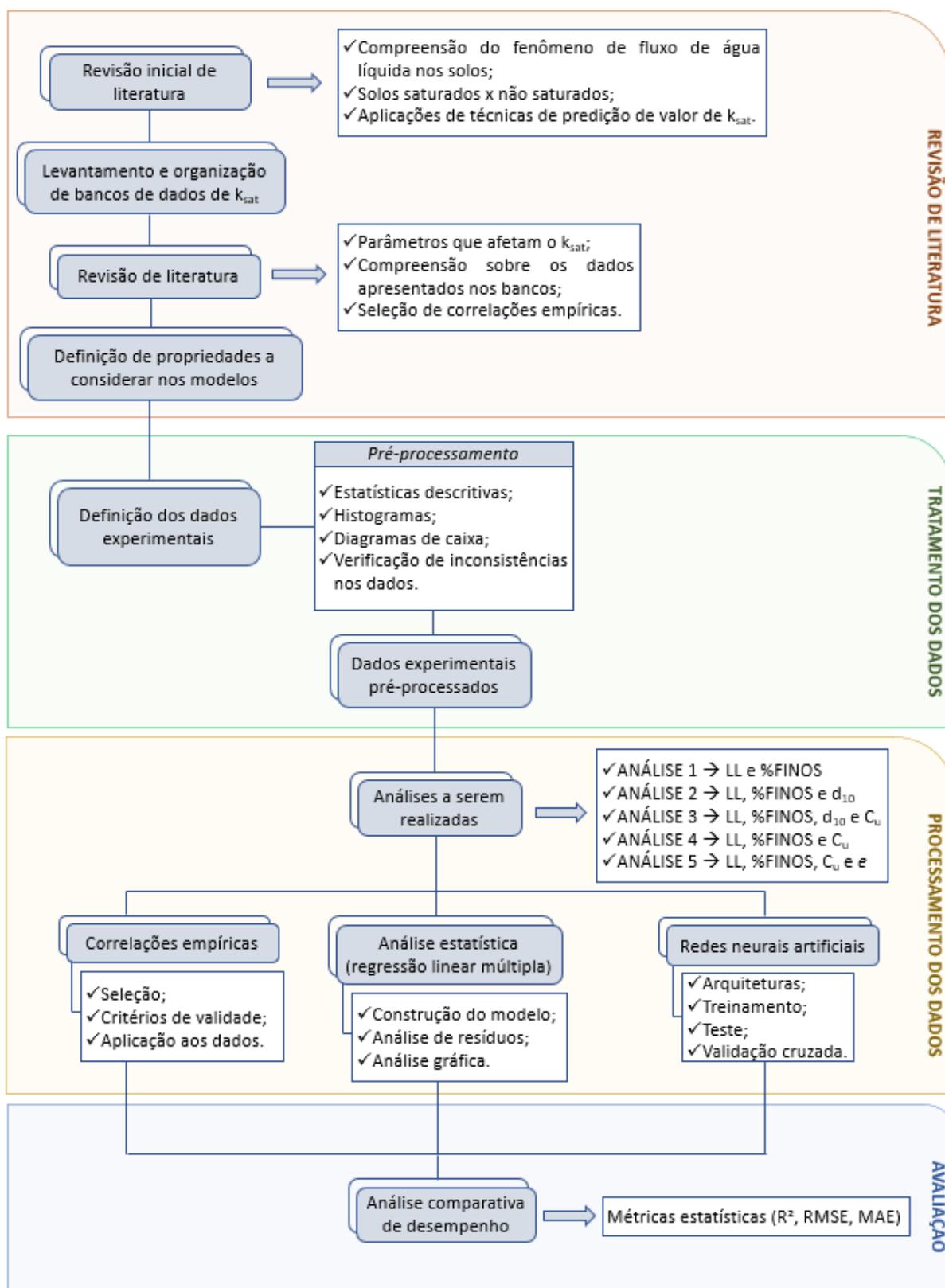
### 3 MATERIAIS E MÉTODOS

Neste capítulo, são apresentadas as bases de dados de  $k_{sat}$  compiladas da literatura (o tratamento e a discussão dos dados serão realizados no Capítulo 4). Além disso, são apresentados os critérios de seleção dos dados experimentais utilizado para o treinamento das RNA e desenvolvimento das análises estatísticas nesta pesquisa. Essa base de dados experimentais, selecionada a partir dos bancos de dados disponibilizados na literatura, é utilizada para o desenvolvimento das regressões lineares múltiplas, para o treinamento das redes neurais e aplicadas às equações empíricas. Cada uma dessas etapas de desenvolvimento é detalhada neste capítulo.

É descrito o método para estabelecer as correlações para previsão do valor de  $k_{sat}$ , considerando as ferramentas computacionais utilizadas, as análises gráficas e testes de hipóteses para verificação das equações de regressão propostas. Todas as análises realizadas, a partir de diferentes conjuntos de variáveis de entrada, seja para o treinamento das redes ou para o desenvolvimento das regressões são apresentadas, bem como os critérios para avaliação e análise comparativa entre as técnicas utilizadas.

A Figura 12 mostra o fluxograma das etapas realizadas no desenvolvimento desta pesquisa e que são detalhadas neste capítulo.

Figura 12 – Fluxograma das etapas da pesquisa



Fonte: Autora (2022).

### 3.1 Etapa de revisão de literatura

A etapa de revisão de literatura teve por objetivo o aprofundamento do tema para melhor compreensão da permeabilidade dos solos e das principais variáveis envolvidas neste fenômeno. A descrição correta do problema possibilita tanto a seleção das variáveis importantes para o desenvolvimento das etapas seguintes quanto uma análise crítica e discussão dos resultados.

A pesquisa por bancos de dados considerou apenas os disponibilizados em teses ou dissertações, publicados por agências e departamentos governamentais ou por pesquisadores em artigos científicos. Dados de amostras para os quais não fosse possível identificar a origem das informações não foram considerados. Ao todo, 8258 dados publicados foram compilados e são descritos e analisados de forma detalhada no Capítulo 4.

Embora a revisão de literatura tenha sido considerada a primeira etapa no fluxograma de atividades, ela foi atualizada ao longo de todo o desenvolvimento da pesquisa.

### 3.2 Etapa do tratamento dos dados

O levantamento dos dados de  $k_{sat}$  publicados nas fontes consideradas nesta pesquisa resultou em 8258 conjuntos de amostras. Por *conjunto de amostras* entenda-se a informação do  $k_{sat}$  e de alguma outra propriedade da amostra de solo, como propriedades índices e distribuição granulométrica, que possa ser utilizada na previsão de  $k_{sat}$ . Além disso, já foram previamente descartadas amostras cujos percentuais das frações granulométricas não somavam 100% e linhas de dados duplicadas.

A partir desses bancos de dados previamente reunidos, os dados experimentais para essa pesquisa seriam, então, selecionados, pré-processados e utilizados para o treinamento das redes neurais e análises estatísticas. Os 8258 conjuntos de dados estão distribuídos em 8 diferentes referências. Desse total:

- 114 dados foram localizados em Wang e Huang (1984);
- 56 amostras em Benson *et al.* (1984);
- 419 amostras disponibilizadas no Banco de Dados Hidrofísico de Solos no Brasil (HYBRAS) (Ottoni *et al.*, 2018);

- 172 dados presentes no *Unsaturated Soil Hydraulic Database (UNSODA)* (Nemes *et al.*, 2001);
- 196 dados coletados do *Soil Water Infiltration Global Database (SWIG)* (Rahmati *et al.*, 2018);
- 32 amostras em Alyamani e Sen (1993);
- 7227 amostras foram reunidas em Gupta *et al.* (2020);
- 42 amostras foram selecionadas em Puckett *et al.* (1985).

Cada referência citada organizou dados de  $k_{sat}$  com outros parâmetros do solo, como informações granulométricas, índice de vazios, porosidade, massa específica total e/ou aparente seca etc., e as estatísticas dos dados são apresentadas, de forma sucinta, no decorrer do Capítulo 4.

A Nota Técnica Nº 717 do *Soil Conservation Service* dos Estados Unidos da América (USDA, 1991) foi desenvolvida com foco na permeabilidade de solos para usos ambientais, como barreiras impermeáveis. Fornece diretrizes para algumas estimativas preliminares de permeabilidade, amostragem, ensaios de laboratório e de campo para este fim.

No documento, é apresentada uma classificação qualitativa da permeabilidade do solo com base no valor de  $k_{sat}$ . A classificação é apenas orientativa. De acordo com o valor de  $k_{sat}$ , a permeabilidade é classificada como *extremamente alta*, *muito alta*, *alta*, *média*, *baixa*, *muito baixa* e *praticamente impermeável*, como mostra a Figura 13. Importante reforçar que não se trata de um sistema de classificação de solos e, sim, apenas orienta a respeito da magnitude de algumas faixas de  $k_{sat}$ .

Figura 13 – Classificação relativa de permeabilidade

$k_{sat}$ (m/s)	$10^{-1}$	$10^{-2}$	$10^{-3}$	$10^{-4}$	$10^{-5}$	$10^{-6}$	$10^{-7}$	$10^{-8}$	$10^{-9}$	$10^{-10}$	$10^{-11}$
Perm. relativa	Extremamente alta	Muito alta	Alta		Média	Baixa	Muito baixa	Praticamente impermeável			
Tipos de solos	Pedregulho limpo (GP)	Areias limpas, misturas de areias e pedregulhos (GW, GP, SW, SP, SM)			Areias muito finas, siltes orgânicos e inorgânicos, misturas de areia, silte e argila, solos de origem glacial, depósitos estratificados de argila, etc. (GM, SM, ML, GC, CL)			Solos "impermeáveis". Argilas homogêneas (CL, CH)			
			Solos "impermeáveis" modificados por efeito de vegetação e clima (MH, ML, CL)								

Fonte: USDA (1991).

Além da classificação relativa da permeabilidade, também foi realizada a classificação da textura pelo sistema de classificação textural (USDA, 1993), que utiliza os percentuais de areia, silte e argila. São apresentadas as amostras no triângulo textural para visualização da distribuição dos tipos de solos, exceto amostras que apresentam pedregulho. Apenas as amostras de Wang e Huang (1984) e Benson *et al.* (1984) foram classificadas no SUCS, pois sistemas de classificação como HRB e SUCS, por exemplo, necessitam de informações, como LL, IP,  $C_c$  e  $C_u$ , que a maioria dos bancos não disponibilizam sobre as amostras.

Pela quantidade de bancos de dados e detalhes apresentados e discutidos, optou-se por apresentá-los em um capítulo à parte (Capítulo 4) desta dissertação. Neste referido capítulo, também é apresentada a base de dados utilizada nas análises e os critérios para seleção das amostras.

### **3.3 Correlações empíricas**

No Capítulo 2 desta dissertação, foram apresentadas diversas correlações publicadas na literatura para estimativa de  $k_{sat}$  em função de propriedades índices. Observando os requisitos de validade das equações e as variáveis necessárias para aplicação, algumas correlações apresentadas foram selecionadas para um estudo comparativo com os resultados das equações que são propostas nesta pesquisa. No processo de seleção de correlações, considerou-se o atendimento aos critérios de validade das equações e a possibilidade de aplicá-las a todo o conjunto de dados experimentais.

### **3.4 Análises estatísticas**

Com o objetivo de propor uma equação de regressão múltipla para a previsão de  $k_{sat}$ , a partir dos dados experimentais desta pesquisa, uma análise estatística foi realizada. Para tanto, foram utilizadas nesta etapa as ferramentas computacionais Microsoft Excel, Minitab (versão 19) e RStudio (versão 2022.07.2).

O Microsoft Excel foi utilizado para tabulação e organização dos dados, além da criação de gráficos e cálculos adicionais, no decorrer da pesquisa. O Minitab é programa estatístico que apresenta recursos para análise de dados. Nesta pesquisa,

foi utilizado para as estatísticas descritivas, criação dos diagramas de caixa e análises de regressão.

O programa computacional RStudio é um recurso desenvolvido para computação estatística e gráficos, utilizando a linguagem de programação R. O programa foi utilizado para a realização dos testes de hipóteses, cujo código está disponível no Apêndice 2.

A sequência das análises estatísticas seguiu o breve passo a passo:

- 1) Após o pré-processamento dos dados, equações de regressão linear múltipla e regressão polinomial (que inclui o produto entre variáveis e variáveis elevadas a expoentes maiores que a unidade) foram desenvolvidas no Minitab;
- 2) Com o Microsoft Excel, medidas estatísticas de desempenho foram calculadas, bem como gráficos foram construídos para melhor visualização dos resultados;
- 3) No RStudio, foram realizados os testes de hipóteses com o objetivo de verificar a significância estatística das regressões múltiplas propostas e a análise dos resíduos.

As equações de regressão propostas incluíram relações lineares e polinomiais. Para as equações polinomiais, foram investigadas relações matemáticas que incluíam termos resultantes do produto entre duas variáveis (na regressão polinomial, esses termos são chamados de *interações*) e termos que incluíam variáveis elevadas à potência 2 (ordem do polinômio). As equações de regressão investigadas nesta pesquisa são nos seguintes formatos, de acordo com a quantidade de variáveis de entrada:

- Regressão linear múltipla com  $n$  variáveis ( $n+1$  termos):

$$\log(k_{sat}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n$$

- Regressão polinomial de grau 2 e duas variáveis preditoras ( $x_1$  e  $x_2$ ) (6 termos):

$$\log(k_{sat}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \beta_{12} x_1 x_2$$

- Regressão polinomial de grau 2 e três variáveis preditoras ( $x_1$ ,  $x_2$  e  $x_3$ ) (10 termos):

$$\log(k_{sat}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \beta_{33} x_3^2 + \beta_{12} x_1 x_2 + \beta_{13} x_1 x_3 + \beta_{23} x_2 x_3$$

- Regressão polinomial de grau 2 e quatro variáveis preditoras ( $x_1, x_2, x_3$  e  $x_4$ ) (15 termos):

$$\begin{aligned} \log(k_{sat}) = & \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_{11} x_1^2 + \beta_{22} x_2^2 \\ & + \beta_{33} x_3^2 + \beta_{44} x_4^2 + \beta_{12} x_1 x_2 + \beta_{13} x_1 x_3 + \beta_{14} x_1 x_4 + \beta_{23} x_2 x_3 \\ & + \beta_{24} x_2 x_4 + \beta_{34} x_3 x_4 \end{aligned}$$

Para verificar se os termos da regressão estão contribuindo para o resultado da predição, foi aplicado o teste F parcial; já para testar se a relação matemática proposta é verdadeira (significância da regressão como um todo), foi aplicado o teste F. Para os testes de normalidade dos resíduos, além da análise gráfica (realizada com o Minitab), foram aplicados os testes de hipóteses de Shapiro-Francia, Anderson-Darling e Lilliefors.

A estatística de Durbin-Watson foi utilizada para verificar a independência dos resíduos. E, por se tratar de equações de regressão múltipla, o valor de VIF também foi calculado e analisado para fins de verificação da existência de multicolinearidade. A não existência de colinearidade/multicolinearidade será considerada para valores de  $VIF \leq 10$ .

Os resultados dos testes foram avaliados pela comparação do valor-p com o nível de significância da equação. Assim, para o teste de Durbin-Watson, valor-p < 0,05, indica que os resíduos não são independentes. E, para os testes de Shapiro-Francia, Anderson-Darling e Lilliefors, o valor-p > 0,05 confirma a normalidade da distribuição dos resíduos.

### 3.5 Análises de redes neurais artificiais

Para o desenvolvimento e treinamento de arquiteturas de redes neurais artificiais, foi utilizado o programa RapidMiner Studio, que é uma ferramenta para cientistas de dados e que apresenta diversos modelos de aprendizagem de máquinas, além de uma interface gráfica intuitiva.

No RapidMiner Studio, diversas arquiteturas de RNA multicamadas foram treinadas a partir de um conjunto de variáveis de entrada e saída previamente definido, utilizando o algoritmo de aprendizagem *backpropagation*. O programa fornece as medidas

estatísticas de RMSE, MAE e  $R^2$ , além de outras métricas de desempenho, caso desejado pelo usuário.

O desenvolvimento das RNA foi realizado em três fases: treinamento, teste e validação cruzada. Para tanto, 70% dos dados experimentais foram utilizados na fase de treinamento da rede, 15% na fase de teste e os outros 15% para a validação cruzada. A seleção das amostras para esses dois conjuntos foi realizada pelo *software* de maneira aleatória.

Inicialmente, para o conjunto de dados experimentais, 05 análises foram realizadas e são resumidas na Tabela 1:

Tabela 1 – Dados experimentais - Resumo das arquiteturas de RNA treinadas

Análise	RNA's	CAMADA DE ENTRADA	CAMADAS INTERMEDIÁRIAS						CAMADA DE SAÍDA		
		Dados de entrada	CAMADA 1		CAMADA 2		CAMADA 3		Dado de saída	Função de ativação	
			Nº neurônios	Função de ativação	Nº neurônios	Função de ativação	Nº neurônios	Função de ativação			
1	RNA101 - RNA125	Finos - LL						1 - 5	Tangente sigmóide		
2	RNA201 - RNA210	Finos - LL - d <sub>10</sub>						-	-		
3	RNA301 - RNA310	Finos - LL - d <sub>10</sub> - C <sub>u</sub>	1 - 10	Tangente sigmóide	0 - 10	Tangente sigmóide		-	-	log (k <sub>sat</sub> )	Linear
4	RNA401 - RNA410	Finos - LL - C <sub>u</sub>						-	-		
5	RNA501 - RNA510	Finos - LL - C <sub>u</sub> - e						-	-		

Unidades de medida dos dados de entrada:

Finos - [%]

LL - [%]

d<sub>10</sub> - [10<sup>-5</sup> mm]

C<sub>u</sub> - adimensional

e - adimensional

log (k<sub>sat</sub>) - [m/s]

Fonte: Autora (2022).

Ao todo, 05 combinações diferentes de variáveis de entrada foram testadas; como critério de escolha das variáveis, foram selecionadas propriedades da distribuição granulométrica e propriedades que permitam inferir acerca da natureza da fração fina e sua proporção na amostra. Foram testadas redes com 1, 2 ou até 3 camadas intermediárias com a quantidade de neurônios por camada resumida na Tabela 1 acima. O RapidMiner Studio só fornece como função de ativação para as camadas intermediárias a função tangente sigmoide. Na camada de saída, foi utilizada a função de ativação do tipo linear, única disponibilizada no programa.

Apenas na Análise 1 foram testadas RNA com três camadas ocultas. Como o desempenho foi inferior às redes com 1 ou 2 camadas intermediárias, não foi testada essa arquitetura nas demais análises. A quantidade de neurônio por camada oculta está resumida na Tabela 2.

Tabela 2 – Número de neurônios por camada intermediária

CAMADA INTERMEDIÁRIA 1	CAMADA INTERMEDIÁRIA 2	CAMADA INTERMEDIÁRIA 3
2		
4		
6		
8		
10		
2	1	
4	2	
6	3	
8	4	
10	5	
1	2	
2	4	
3	6	
4	8	
5	10	
2	1	1
4	2	2
6	3	3
8	4	4
10	5	5
2	2	1
4	4	2
6	6	3
8	8	4
10	10	5

Fonte: Autora (2022).

A terceira camada oculta, como dito anteriormente, foi testada apenas na Análise 1. Além disso, as arquiteturas com 2 camadas, nas quais o número de neurônios da segunda camada era o dobro do número de neurônios da primeira camada, também só foram testadas na Análise 1, pois não houve melhoria no desempenho; aliás, foi ligeiramente inferior às demais arquiteturas. Um esquema gráfico das RNA treinadas no RapidMiner Studio é mostrado no Apêndice 3.

Para apresentação dos resultados, foi adotada uma numeração utilizando 3 dígitos para as arquiteturas: o primeiro dígito se refere à indicação da Análise (de 1 a 5) e os dois últimos, à rede correspondente. Portanto, na Análise 1, foram treinadas 25 arquiteturas de RNA (RNA101 à RNA125) e, nas Análises 2, 3, 4 e 5, foram treinadas 10 arquiteturas cada uma: RNA201 a RNA210, RNA301 a RNA310, RNA401 a RNA410 e RNA501 a RNA510, respectivamente.

Para a paralisação do processo de retropropagação de erro, um critério de parada deve ser fornecido ao programa. Assim, dois parâmetros adicionais foram adotados para o treinamento das redes, a saber: o treinamento da rede é paralisado quando um número máximo de ciclos de treinamento igual a 1000 for alcançado ou um erro máximo de 0,0001 seja observado, o que acontecer primeiro.

### **3.6 Avaliação e comparação entre resultados**

Para cada conjunto de dados experimentais, correlações publicadas na literatura foram aplicadas, uma equação de regressão linear múltipla e uma equação de regressão polinomial foram desenvolvidas e redes neurais foram treinadas. Os resultados de cada método de predição de  $k_{sat}$  foram comparados segundo as medidas estatísticas de desempenho RMSE, MAE e  $R^2$  a fim de verificar quais dos métodos utilizados mais se adequaram ao banco de dados utilizado. Os resultados foram organizados em tabelas e gráficos para facilitar a visualização e análise comparativa.

## 4 BANCOS DE DADOS

Neste capítulo, o banco de dados utilizado para treinamento das RNA e análises estatísticas é descrito e analisado em detalhes. A base de dados experimentais foi obtida após a seleção e análise de bancos de dados disponibilizadas na literatura.

### 4.1 Bancos de dados obtidos da literatura

Como descrito no Capítulo 3, inicialmente, 8258 conjuntos de dados, distribuídos em 8 diferentes referências, foram reunidos, sendo estes:

- 114 dados foram localizados em Wang e Huang (1984);
- 56 amostras em Benson *et al.* (1984);
- 419 amostras disponibilizadas no Banco de Dados Hidrofísico de Solos no Brasil (HYBRAS) (Ottoni *et al.*, 2018);
- 172 dados presentes no *Unsaturated Soil Hydraulic Database (UNSODA)* (Nemes *et al.*, 2001);
- 196 dados coletados do *Soil Water Infiltration Global Database (SWIG)* (Rahmati *et al.*, 2018);
- 32 amostras em Alyamani e Sen (1993);
- 7227 amostras foram reunidas em Gupta *et al.* (2020);
- 42 amostras foram selecionadas em Puckett *et al.* (1985).

Para fins de seleção da base de dados a ser utilizada nesta pesquisa, uma detalhada investigação sobre as amostras apresentadas nestes bancos de dados foi realizada. Para não tornar a leitura exaustiva, a análise sobre os bancos de dados pode ser consultada no Apêndice 1. Este Capítulo 4 se concentrará, portanto, na apresentação das conclusões das análises realizadas e nos critérios para definir a base de dados.

Para auxiliar, a Tabela 3 mostra um resumo da quantidade e localidade dos sítios das amostras de solos investigadas. Além disso, inclui os dados disponibilizados por cada banco, o equipamento utilizado e o tipo de ensaio de permeabilidade, a natureza da amostra e a faixa de  $k_{sat}$ .

Tabela 3 – Sumarização dos bancos de dados (N/I = não informado)

Nº	Organização	Localidade	Dados disponíveis	Nº amostras	Equipamento de ensaio	Tipo de ensaio	Natureza da amostra	Faixa de $k_{sat}$ (m/s)	
								Mínimo	Máximo
1	Wang e Huang (1984)	Não disponibilizada	% argila, % silte, % areia, % pedregulho, $G_s$ , LL, LP, $d_{10}$ , $d_{50}$ , $C_u$ , $e$	114	Molde cilíndrico de compactação	Carga constante e carga variável	Mistura controlada	$2,50 \times 10^{-13}$	$1,70 \times 10^{-05}$
2	Benson <i>et al.</i> (1984)	América do Norte	w, % argila, % finos, % areia, % pedregulho, LL, IP, $\rho_d$	56	Permeâmetro de parede flexível (52 amostras), tubo Shelby (parede rígida) (07 amostras) ou em células de adensamento equipadas para medição direta de permeabilidade (08 amostras)	N/I	Indefornada, extraídas de aterro com solo argiloso natural	$2,70 \times 10^{-11}$	$6,70 \times 10^{-09}$
3	Otoni <i>et al.</i> (2018)	Brasil	% argila, % silte, % areia, $\rho_d$ , $\rho_s$ , n	419	N/I	Carga constante Permeâmetro de Guelph N/I	N/I	$1,35 \times 10^{-07}$	$4,50 \times 10^{-02}$
4	Nemes <i>et al.</i> (2001)	América do Norte, Ásia, Oceania e Europa	% argila, % silte, % areia, $\rho_t$ , $\rho_s$ , n	172	N/I	N/I	N/I	$7,52 \times 10^{-09}$	$3,20 \times 10^{-03}$
5	Rahmatl <i>et al.</i> (2018)	Irã e França	% argila, % silte, % areia, $\rho_t$ , $\rho_s$ , $d_{50}$	196	N/I	N/I	N/I	$7,18 \times 10^{-08}$	$4,15 \times 10^{-03}$
6	Alyamani e Sen (1993)	Arábia Saudita e Austrália	$d_{10}$ , $d_{50}$ , $d_{60}$ , $C_u$ , $e$	32	N/I	Carga constante N/I	Mistura controlada (22) N/I (10)	$1,32 \times 10^{-04}$	$1,10 \times 10^{-02}$
7	Gupta <i>et al.</i> (2020)	Várias localidades	% Argila, % Silte, % Areia, OC (%), $\rho_t$	7227	N/I	N/I	N/I	$2,15 \times 10^{-10}$	$2,34 \times 10^{-03}$
8	Pucket <i>et al.</i> (1985)	Estados Unidos da América	% Argila, % Silte, % Areia, n, $\rho_t$	42	N/I	Carga constante	Amostras naturais indefornadas	$8,33 \times 10^{-10}$	$3,94 \times 10^{-05}$

Fonte: Autora (2022).

Observando a disponibilidade de informações nos bancos de dados, é notória a carência de informações sobre algumas propriedades imprescindíveis do solo que apresentam influência direta ou indireta sobre o  $k_{sat}$ , conforme discutido no Capítulo 2.

Alguns bancos não apresentam informações sobre as frações granulométricas presentes, em termos de percentuais de areia, silte e argila, como Alyamani e Sen (1993). Outros, não apresentam propriedades que possam sugerir sobre a natureza do mineral argila, como o limite de liquidez ou plasticidade. A maioria dos bancos não mencionam informações sobre o mineral que compõem a fração argila.

De todos os dados levantados, nenhum banco contemplou informação sobre a superfície específica, exceto Pucket *et al.* (1985), corroborando para a dificuldade em se utilizar equações empíricas que necessitam desse parâmetro, como Kozeny-Carman (1927, 1937). Do total de conjunto de dados levantados, praticamente todos apresentam %argila, %silte e %areia, exceto Alyamani e Sen (1993), como já mencionado.

Porém, quando verificada a disponibilidade de informações complementares, há uma escassez de dados. Os limites de consistência, necessários para inferir sobre a fração fina das amostras, na ausência de informações sobre a superfície específica, estão presentes em apenas 167 amostras do total de 8258 dados compilados, que corresponde a 2.02% dos conjuntos de dados. O  $d_{10}$ , importante parâmetro relacionado à permeabilidade e necessário para aplicação de correlações empíricas, como Hazen (1911), só é apresentado para 143 amostras (1.73% dos dados). O  $C_u$  e o  $d_{60}$  são apresentados em 143 amostras. O índice de vazios apenas está presente para 754 conjuntos de dados (9,13% do total).

Ou seja, apesar dos bancos de dados apresentarem informações sobre o  $k_{sat}$ , muitas vezes, carecem de outras propriedades que são importantes para a caracterização e compreensão sobre a permeabilidade das amostras. Das conclusões parciais do Capítulo 2, verificou-se que a presença da fração fina no solo afeta os valores de  $k_{sat}$ , mas que essa influência é refletida em outros parâmetros, como na distribuição granulométrica, e dependente da natureza da fração fina. Sendo assim, é importante que a construção de um banco de dados experimentais contemplem essas características.

Em amostras provenientes de bancos de dados, a relação  $k_{sat}$  vs índice de vazios e  $k_{sat}$  vs teor de finos, reportada no Capítulo 2, nem sempre é possível de ser identificada, pois, normalmente, se trata de amostras de diferentes granulometrias na matriz grossa e fina; essas relações são mais visíveis em estudos com amostras que apresentam a mesma granulometria da matriz granular.

Nestes bancos de dados, notou-se que misturas com finos plásticos tendem a apresentar menor valor de  $k_{sat}$  do que com finos não plásticos. Além disso, diferentes composições da matriz granular (teor de areia e pedregulho) e as diferentes composições da matriz fina (teor de silte e argila) exercem influência nos valores de  $k_{sat}$  das amostras. A influência da composição da matriz fina (%silte e %argila) é mais expressiva do que a composição da matriz grossa, nos dados avaliados.

Em suma, a permeabilidade de solos com predominância de finos (ou com o comportamento hidráulico controlado pelos finos) não é caracterizada de forma adequada somente com a informação da fração granulométrica desses solos. São necessárias informações que caracterizem esse fino, como os limites de consistência.

Para um parâmetro do solo como o  $k_{sat}$ , que sofre influência de tantas outras propriedades interdependentes do solo, é fundamental que uma caracterização completa das amostras seja fornecida, principalmente abrangendo o teor de finos, as características da distribuição granulométrica e propriedades que se relacionam a mineralogia do fino (limite de liquidez, por exemplo), permitindo, assim, serem avaliadas e consideradas em estudos para estimativa de  $k_{sat}$ .

Tarefas que compõem uma análise de bancos de dados, como a avaliação de *outliers*, para a tomada de decisão sobre a exclusão de uma amostra, só é possível a partir do fornecimento de uma caracterização mais completa das amostras; quanto mais informações estiverem disponíveis, mais confiável será análise sobre a eventual discrepância de um valor.

Em estudos para previsão de  $k_{sat}$ , dada a natureza complexa e altamente variável dessa propriedade, é tão importante quanto uma quantidade representativa de dados a quantidade de propriedades informadas de cada uma das amostras.

A partir da revisão apresentada no Capítulo 2 e a análise dos bancos de dados realizada neste capítulo, foi possível definir critérios para a seleção dos dados amostras a serem utilizados no treinamento das RNA e nas análises estatísticas.

De todas as amostras, apenas os dados de Wang e Huang (1984) apresentam esses parâmetros de distribuição granulométrica, limites de consistência e a natureza do mineral argila. Sendo assim, essa pesquisa foi conduzida considerando, principalmente, apenas os conjuntos de dados com %finos, LL, Cu,  $d_{10}$  e  $e$ , logo, apenas os dados de Wang e Huang (1984). Nenhum outro banco de dados apresentou todas essas propriedades para que pudessem ser incorporados no conjunto de dados experimentais. Os dados de Wang e Huang (1984) são, portanto, apresentados a seguir.

#### **4.2 Dados disponibilizados em Wang e Huang (1984)**

Wang e Huang (1984) obtiveram amostras de solos a partir de misturas entre pedregulho, areia, pó de calcário e bentonita, cujo principal argilomineral constituinte era a montmorilonita sódica. Ao todo, 57 amostras foram construídas com diferentes dosagens entre esses componentes.

As amostras preparadas foram, então, compactadas em laboratório na energia *Proctor Normal*, para a obtenção dos dados de umidade ótima e massa específica aparente seca. Além do ensaio de compactação, ensaios de caracterização foram realizados e parâmetros como %areia, %silte e %argila, LL, LP, módulo de finura,  $d_{50}$ ,  $d_{10}$  e  $C_u$  foram determinados.

Após esse primeiro ensaio de compactação para as 57 diferentes dosagens, corpos de prova para ensaio de permeabilidade foram obtidos a partir de compactação com 90% e 95% da energia *Proctor Normal*. Ao final, 114 corpos de prova foram obtidos a partir das 57 dosagens: metade compactada com 90% do *Proctor Normal* e a outra metade com 95%, conseqüentemente, com diferentes índices de vazios resultantes.

De um exame visual dos dados, três conjuntos de amostras já precisaram ser previamente excluídos por apresentarem valores discrepantes (ou inconsistentes), como percentual de argila acima de “100%” e índice de vazios de 0,05.

As amostras de Wang e Huang (1984) foram as que apresentaram os menores valores de  $k_{sat}$  dentre todos os dados levantados por esta pesquisa. Das 111 amostras (já removidas as três com valores visivelmente discrepantes), 68 delas são classificadas como "praticamente impermeável" ( $k_{sat} < 3,53 \times 10^{-9}$  m/s), de acordo com os critérios apresentados na Nota Técnica Nº 717 do *Soil Conservation Service* dos Estados Unidos da América (USDA, 1991). A Tabela 4 resume a distribuição das amostras nos intervalos de  $k_{sat}$  e a respectiva classificação relativa de permeabilidade.

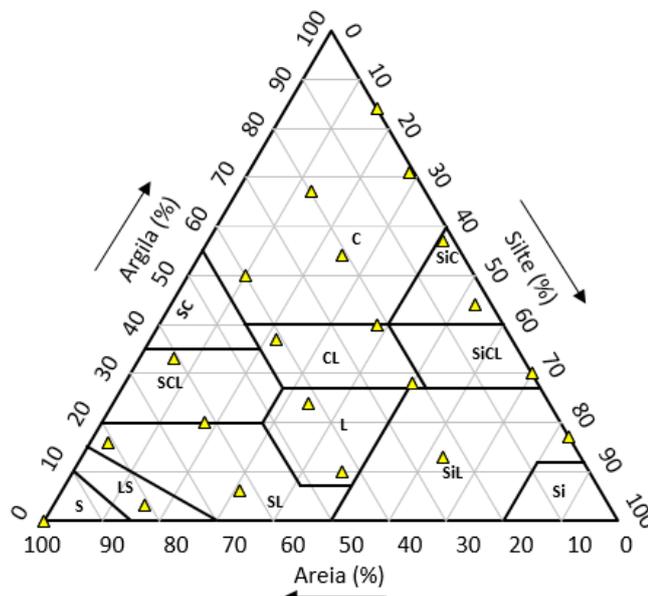
Tabela 4 - Classificação relativa da permeabilidade das amostras de Wang e Huang (1984)

Classificação relativa da permeabilidade	Intervalo de referência (m/s)	Amostras não plásticas (NP)	Amostras plásticas
		nº de amostras	nº de amostras
Extremamente alta	$k_{sat} \geq 10^{-2}$	-	-
Muito alta	$3,53 \times 10^{-4} \leq k_{sat} < 10^{-2}$	-	-
Alta	$3,53 \times 10^{-6} \leq k_{sat} < 3,53 \times 10^{-4}$	5	-
Média	$3,53 \times 10^{-7} \leq k_{sat} < 3,53 \times 10^{-6}$	7	-
Baixa	$3,53 \times 10^{-8} \leq k_{sat} < 3,53 \times 10^{-7}$	10	6
Muito baixa	$3,53 \times 10^{-9} \leq k_{sat} < 3,53 \times 10^{-8}$	7	8
Praticamente impermeável	$k_{sat} < 3,53 \times 10^{-9}$	1	67
TOTAL DE AMOSTRAS		30	81

Fonte: Autora (2022).

As demais amostras apresentam permeabilidade entre muito baixa e alta. A maior parte das amostras são compostas por solos plásticos praticamente impermeáveis. A distribuição de 42 amostras de Wang e Huang (1984) no triângulo textural, pois amostras com pedregulho (total de 69 amostras) não foram consideradas, é mostrada na Figura 14.

Figura 14 – Triângulo textural – Distribuição de 42 amostras, com teor de pedregulho igual a zero, dos dados de Wang e Huang (1984)



LEGENDA: C - Argila, SC - Argila Arenosa, CL - Franco Argilosa, SCL - Franco Argilo Arenosa, SL - Franco Arenoso, LS - Areia Franca, S - Areia, L - Franca, SiC - Argila Siltosa, SiCL - Franco Argilo Siltosa, SiL - Franco Siltosa, Si - Silte

Fonte: Autora (2022)

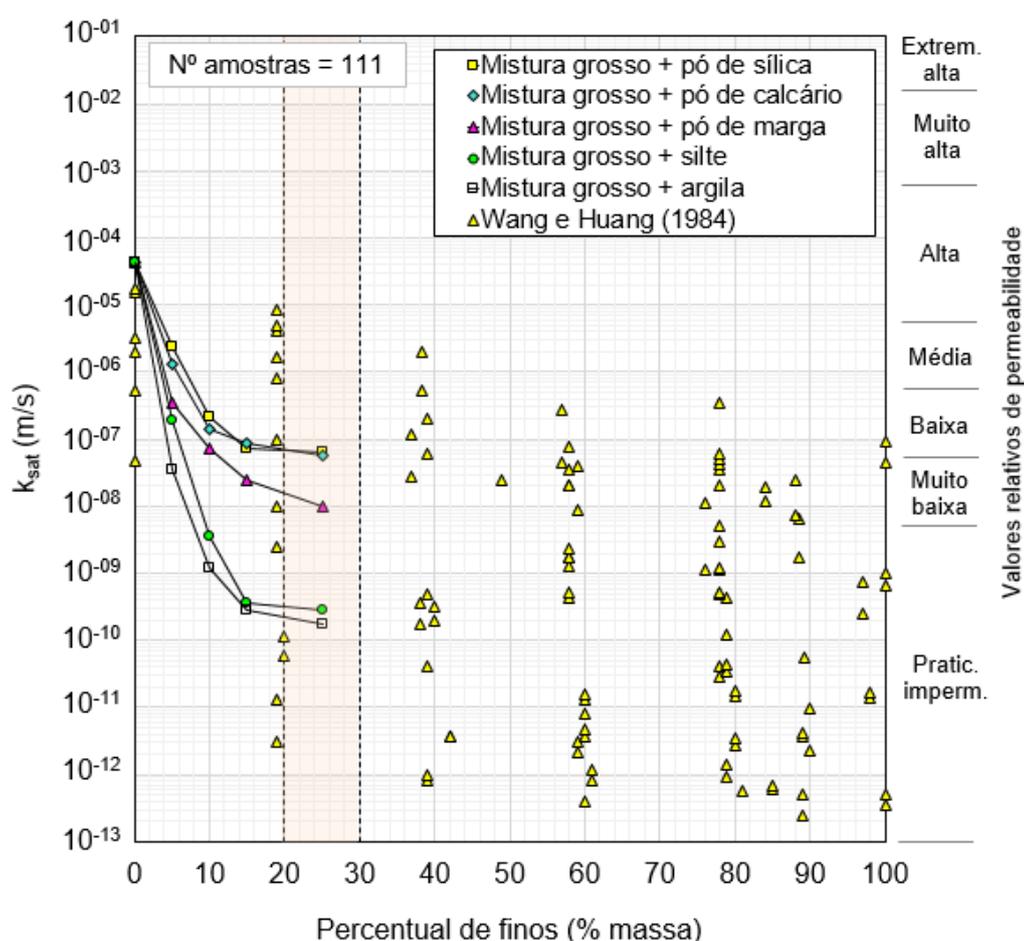
É observado que as amostras são bem distribuídas dentro do triângulo textural, fato reforçado pelo processo de dosagem ao qual foi submetido o solo para a construção dessas amostras. O triângulo textural reforça a variedade de tipos de solo presente no banco de dados de Wang e Huang (1984). A classificação no SUCS das 111 amostras compactadas é apresentada:

- Argila de alta plasticidade (CH): 58 amostras
- Argila de baixa plasticidade (CL): 2 amostras
- Silte de baixa plasticidade (ML): 16 amostras
- Areia argilosa (SC): 17 amostras
- Areia siltosa (SM): 12 amostras
- Areia mal graduada (SP): 6 amostras

Trata-se, portanto, de um banco de dados que contém desde solos granulares puros (areia pura ou areia com pedregulho) até solos finos puros, plástico e não plástico. Os dados de  $k_{sat}$  vs teor de finos para este primeiro banco de dados foram plotados no Gráfico 1.

No gráfico, os resultados obtidos por Barber e Sawyer (1952) também são apresentados. Os resultados de Barber e Sawyer (1952) aparecem como curvas, cada qual com os valores de  $k_{sat}$  em função do teor de finos de diferentes plasticidades, que foram obtidas no estudo de misturas controladas, apresentado no Capítulo 2 desta dissertação. Adicionalmente, a classificação relativa da permeabilidade também é apresentada, baseada no proposto na Nota Técnica n° 717 (USDA, 1991, p. 8).

Gráfico 1 - Variação de  $k_{sat}$  com o teor de finos para 111 diferentes misturas controladas (amostras compactadas) (Dados de: BARBER; SAWYER, 1952; WANG; HUANG, 1984)



Fonte: Autora (2022).

A apresentação dos resultados de Barber e Sawyer (1952) permitem uma visualização do efeito da plasticidade do fino nos valores de  $k_{sat}$  em misturas controladas de solos grossos e finos e a comparação deste mesmo efeito nos resultados de Wang e Huang (1984). O trabalho de Barber e Sawyer (1952) apenas difere de Wang e Huang (1984), pois utilizaram uma matriz granular fixa na construção de todas as misturas controladas.

Do gráfico, devido ao processo de construção das misturas controladas e à presença de finos de diferentes plasticidades, observa-se, primeiramente, que os resultados de  $k_{sat}$  das amostras de Wang e Huang (1984) são coerentes com as curvas de diferentes plasticidades de Barber e Sawyer (1952): amostras de maior plasticidade apresentam menor valor de  $k_{sat}$ .

Todas as amostras do banco de dados que apresentaram  $IP > 140\%$  tiveram  $k_{sat}$  menor que  $10^{-10}$  m/s. Conforme o  $IP$  das amostras diminuiu, os valores de  $k_{sat}$  situaram-se acima de  $10^{-10}$  m/s, com os maiores  $k_{sat}$  medidos nas amostras NP, mesmo com teor de finos acima de 80%, por exemplo. Para as amostras de solos granulares puros (teor de finos = 0%) e amostras de solos finos puros (teor de finos = 100%), um resumo é apresentado na Tabela 5.

Tabela 5 - Permeabilidade de solos grossos e finos puros (Dados de: WANG; HUANG, 1984)

Solos grossos puros – amostras compactadas								
%Argila	%Silte	%Areia	%Pedreg	%Finos	$k_{sat}$ (m/s)	IP (%)	e	$C_u$
0	0	100	0	0	$5,20 \times 10^{-07}$	-	0,653	2,95
0	0	90	10	0	$1,50 \times 10^{-05}$	-	0,490	4,17
0	0	81	19	0	$1,70 \times 10^{-05}$	-	0,390	5,14
0	0	100	0	0	$4,80 \times 10^{-08}$	-	0,566	2,95
0	0	90	10	0	$3,20 \times 10^{-06}$	-	0,412	4,17
0	0	81	19	0	$2,00 \times 10^{-06}$	-	0,317	5,14
Solos finos puros – amostras compactadas								
%Argila	%Silte	%Areia	%Pedreg	%Finos	$k_{sat}$ (m/s)	IP (%)	e	$C_u$
30	70	0	0	100	$1,00 \times 10^{-09}$	53	1,095	14,00
17	83	0	0	100	$9,00 \times 10^{-08}$	2	1,174	4,79
30	70	0	0	100	$6,40 \times 10^{-10}$	53	1,072	14,00
17	83	0	0	100	$4,50 \times 10^{-08}$	2	0,942	4,79
84	16	0	0	100	$5,20 \times 10^{-13}$	449	1,362	10,36
84	16	0	0	100	$3,60 \times 10^{-13}$	449	1,238	10,36

Fonte: Autora (2022).

Nas amostras de solos grossos sem a presença de finos, é possível perceber que a substituição de 10% da areia por pedregulho é responsável pelo aumento em 28 vezes do valor de  $k_{sat}$ , mas a substituição de 20% da areia por pedregulho, já não resultou em ganho significativo da permeabilidade saturada.

Também é possível notar a influência do índice de vazios em uma dada fração granulométrica. As amostras com mesmos percentuais de areia e pedregulho apresentam, conseqüentemente, as mesmas propriedades de distribuição

granulométrica ( $d_{10}$ ,  $d_{50}$ ,  $C_u$  etc.), nas amostras de Wang e Huang (1984). O que justifica a variação de  $k_{sat}$  entre essas amostras de mesma dosagem é o valor do índice de vazios, resultante dos diferentes graus de compactação.

O índice de vazios das amostras de Wang e Huang (1984) foi obtido a partir massa específica aparente seca ( $\rho_d$ ) e da massa específica dos grãos ( $\rho_s$ ). Portanto, para os solos granulares puros, os valores de  $k_{sat}$  vão de  $4,80 \times 10^{-8}$  m/s a  $1,70 \times 10^{-5}$  m/s e essa amplitude se deve às adições de pedregulho à areia e às diferentes compacidades. Ressalta-se que todas as amostras granulares puras são uniformes ( $C_u \leq 6$  para areias).

As amostras de solos finos puros são compostas todas por misturas de silte e argila, mas em diferentes proporções. Nestas amostras é possível visualizar a influência do tipo de fino e do índice de vazios. As 81 amostras plásticas de Wang e Huang (1984) apresentam um IP médio de  $209,07 \pm 135$  (%), com valores variando de 2% a 449%. Amostras com maiores percentuais de argila apresentam menores valores de  $k_{sat}$  quando comparadas com amostras com mais silte.

Por exemplo, amostras com 84% de argila (IP = 449%) apresentam  $k_{sat}$  da ordem de  $10^{-13}$  m/s, enquanto amostras com os mesmos 83%, porém, de silte (IP = 2%) possuem  $k_{sat}$  em torno de  $10^{-8}$  m/s. Portanto, além do teor de finos, o tipo de fino exerce efeito sobre o valor da permeabilidade saturada e recomenda-se que características desse fino sejam consideradas em equações de predição, como os limites de consistência.

Desta forma, destaca-se que diferentes composições da matriz granular (teor de areia e pedregulho) como também as diferentes composições da matriz fina (teor de silte e argila), refletem nos valores de  $k_{sat}$  das amostras. Entretanto, para as amostras estudadas, a composição da matriz fina exerceu maior influência nos valores de  $k_{sat}$  do que a composição da matriz granular.

Quando analisados outros conjuntos de amostras de Wang e Huang (1984), com diferentes teores de finos, percebe-se que a amplitude dos valores de  $k_{sat}$  é justificada por essas outras propriedades que influenciam na permeabilidade, além do teor de finos em si. Por exemplo, as amostras com percentual de finos de, aproximadamente,

20%, que são mostradas na Tabela 6. Nota-se que os valores de  $k_{sat}$  para amostras com teor de finos de 19% e 20% variam de  $8,40 \times 10^{-6}$  m/s a  $3,00 \times 10^{-12}$  m/s.

Tabela 6 - Permeabilidade misturas de solos com 20% de finos (Dados de: WANG; HUANG, 1984)

Solos com 20% de finos – amostras compactadas							
%Argila	%Silte	%Areia	%Pedreg	%Finos	$k_{sat}$ (m/s)	IP	e
3	16	81	0	19	$5,00 \times 10^{-06}$	0	0,568
16	3	81	0	19	$1,30 \times 10^{-11}$	111	0,594
3	16	71	10	19	$8,00 \times 10^{-07}$	0	0,468
10	9	71	10	19	$1,00 \times 10^{-08}$	60	0,561
4	15	62	19	19	$4,00 \times 10^{-06}$	0	0,408
3	16	81	0	19	$8,40 \times 10^{-06}$	0	0,497
16	3	81	0	19	$3,00 \times 10^{-12}$	111	0,510
3	16	71	10	19	$1,00 \times 10^{-07}$	0	0,390
10	9	71	10	19	$2,50 \times 10^{-09}$	60	0,479
4	15	62	19	19	$1,70 \times 10^{-06}$	0	0,334
17	3	61	19	20	$1,15 \times 10^{-10}$	140	0,560
17	3	61	19	20	$6,00 \times 10^{-11}$	140	0,478

Fonte: Autora (2022).

Para discutir melhor a influência das matrizes e a evolução dos valores de  $k_{sat}$  com a variação do teor de finos, nas amostras de Wang e Huang (1984), três conjuntos de amostras, compactadas com 90% da energia Proctor Normal, foram destacados do banco de dados. Os valores de  $k_{sat}$  vs %finos são apresentados no Gráfico 2 e o resumo das amostras destacadas na Tabela 7.

Tabela 7 - Variação de  $k_{sat}$  com a adição de finos em matrizes granulares de diferentes percentuais de areia e pedregulho para amostras compactadas com energia 90% do Proctor Normal (Dados de: WANG; HUANG, 1984)

Amostras com 0% de pedregulho – amostras compactadas								
n.	%Argila	%Silte	%Areia	%Pedreg	%Finos	$C_u$	e	$k_{sat}$ (m/s)
6	17	83	0	0	100	4,79	1,174	$9,00 \times 10^{-08}$
7	13	63	24	0	76	0,47	0,983	$1,10 \times 10^{-08}$
8	10	47	43	0	57	80	0,762	$2,70 \times 10^{-07}$
9	6	31	63	0	37	225	0,717	$1,20 \times 10^{-07}$
10	3	16	81	0	19	71,43	0,568	$5,00 \times 10^{-06}$
11	0	0	100	0	0	2,95	0,653	$5,20 \times 10^{-07}$

Fonte: Autora (2022).

Tabela 7 - Variação de  $k_{sat}$  com a adição de finos em matrizes granulares de diferentes percentuais de areia e pedregulho para amostras compactadas com energia 90% do Proctor Normal (Dados de: WANG; HUANG, 1984) – cont.

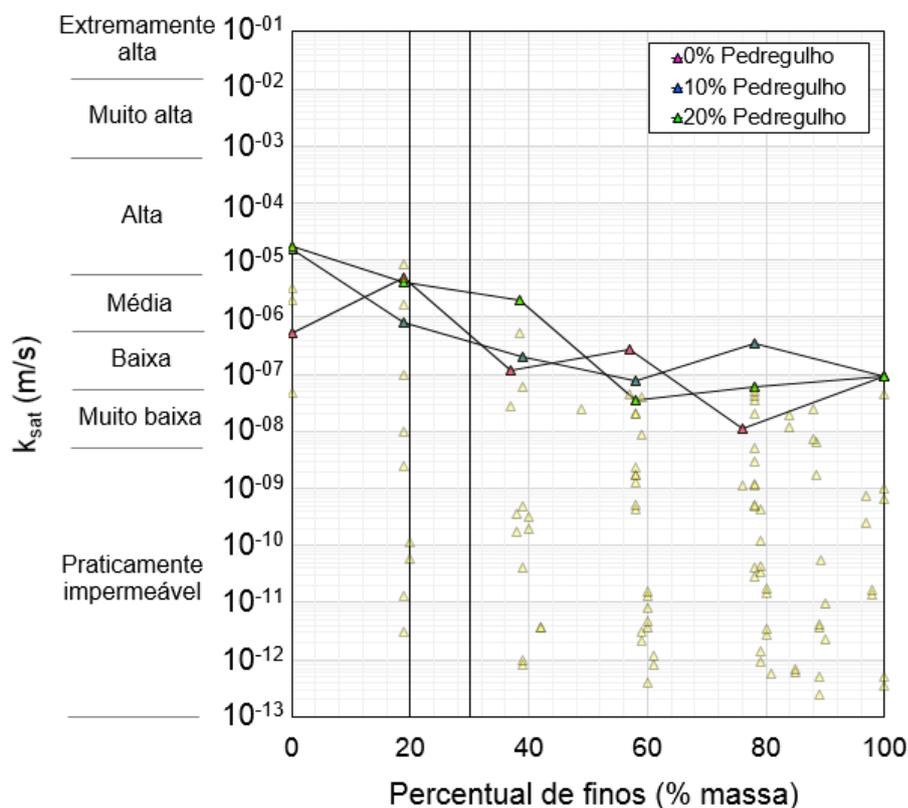
Amostras com 10% de pedregulho – amostras compactadas								
n.	%Argila	%Silte	%Areia	%Pedreg	%Finos	Cu	e	$k_{sat}$ (m/s)
6	17	83	0	0	100	4,79	1,174	$9,00 \times 10^{-08}$
26	14	64	12	10	78	5,94	0,862	$3,50 \times 10^{-07}$
25	10	48	32	10	58	63,16	0,747	$7,50 \times 10^{-08}$
24	7	32	51	10	39	140,74	0,594	$2,00 \times 10^{-07}$
23	3	16	71	10	19	96,67	0,468	$8,00 \times 10^{-07}$
22	0	0	90	10	0	4,17	0,490	$1,50 \times 10^{-05}$
Amostras com 20% de pedregulho – amostras compactadas								
n.	%Argila	%Silte	%Areia	%Pedreg	%Finos	Cu	e	$k_{sat}$ (m/s)
6	17	83	0	0	100	4,79	1,174	$9,00 \times 10^{-08}$
43	14	64	2	20	78	6,33	0,888	$6,00 \times 10^{-08}$
44	11	47	23	19	58	78,95	0,676	$3,50 \times 10^{-08}$
45	7,2	31,2	42,4	19	38,4	160,71	0,492	$2,00 \times 10^{-06}$
46	4	15	62	19	19	102,86	0,408	$4,00 \times 10^{-06}$
47	0	0	81	19	0	5,14	0,390	$1,70 \times 10^{-05}$

Fonte: Autora (2022).

O primeiro conjunto de amostras trata-se de uma matriz inicialmente granular, composta por 100% de areia e  $k_{sat}$  de  $5,20 \times 10^{-7}$  m/s. Parte dessa areia foi, então, substituída por porcentagens de 19%, 37%, 57% e 76% de finos (plástico e não plástico, nas proporções indicadas na tabela). O teor de 100% de finos (17% de argila e 83% de silte) apresenta  $k_{sat}$  de  $9,00 \times 10^{-8}$  m/s.

O segundo e o terceiro conjuntos de amostras diferem do primeiro pela quantidade de pedregulho que compõem a fração granular (10% e 19%, respectivamente). Os  $k_{sat}$  da matriz granular pura do segundo e terceiro conjunto são também apresentados na Tabela 7. Os percentuais de finos de substituição são aproximados entre os três conjuntos de amostras.

Gráfico 2 - Variação de  $k_{sat}$  com a adição de finos em matrizes granulares com diferentes percentuais de areia e pedregulho para amostras compactadas com energia 90% do Proctor Normal (Dados de: WANG; HUANG, 1984)



Fonte: Autora (2022).

Considerando, inicialmente, as amostras derivadas da matriz granular com 0% de pedregulho, observa-se que o valor de  $k_{sat}$  não apresenta redução constante com a substituição de parte da areia por finos; ao contrário, o valor oscila, sem uma tendência definida, à medida que maiores são os percentuais de substituição da areia por finos.

O índice de vazios das amostras reduz até o teor de 19%, a partir do qual, o aumento do teor de finos ocasiona um aumento no índice de vazios das amostras. Mudanças no coeficiente de uniformidade também são observadas, induzidas pelas mudanças da distribuição granulométricas com a substituição de parte da matriz granular por finos:  $C_u$  diminui até um teor de 37% de finos, a partir do qual reduz, exceto para a amostra com 76% de finos.

Quando analisada a amostra com matriz granular inicial composta por 90% de areia e 10% de pedregulho, verifica-se um comportamento similar das amostras. À medida que parte da areia é substituída por finos, ocorre, em geral, a redução do valor de  $k_{sat}$ , exceto para a amostra com 78% de finos. Mas também não foi possível visualizar a

relação entre  $k_{sat}$  e o índice de vazios. O valor do índice de vazios reduziu com a substituição da areia por 19% de finos e depois aumentou à medida que mais finos foram substituindo a fração de areia. Todavia, o valor de  $k_{sat}$  não apresentou a tendência observada por Taiba *et al.* (2022) e Belkhatir *et al.* (2014).

Quando a matriz granular foi composta, inicialmente, por 80% de areia e 20% de pedregulho, percebe-se que a substituição de parte da areia por finos provocou uma redução, em geral, nos valores de  $k_{sat}$  até o percentual de 40%. O índice de vazios do grupo de amostras com 20% de pedregulho também reduziu até o teor de 19% de finos, a partir do qual, aumentou.

Algumas observações podem ser destacadas destes três conjuntos de amostras extraídos dos dados de Wang e Huang (1984). De acordo com os resultados experimentais reportados na literatura, espera-se que o aumento do teor de finos reduza o  $k_{sat}$  até um teor de finos de, aproximadamente, 20%, a partir do qual o  $k_{sat}$  pouco se altera e se aproxima do valor do  $k_{sat}$  do fino puro. Entretanto, esse comportamento é possível ser visualizado quando a matriz granular é a mesma e teores de finos são adicionados (em percentual em massa).

Além disso, diferentemente dos estudos de Taiba *et al.* (2022) e Belkhatir *et al.* (2014), Wang e Huang (1984), quando substituíam parte da matriz granular por finos, alteravam apenas a proporção de areia das misturas e não reduziam proporcionalmente o teor de pedregulhos, ou seja, a granulometria dos grãos grossos nas misturas não poderia ser considerada a mesma e com apenas o teor de finos aumentando, em massa.

Wang e Huang (1984), inclusive, adicionaram dois tipos diferentes de finos, um plástico e outro não plástico, diferentemente dos estudos de Taiba *et al.* (2022) e Belkhatir *et al.* (2014). A cada percentual de fino substituído na matriz granular, as proporções de argila e silte na mistura se alteravam: os percentuais de argila, por exemplo, variaram de 0% a 17% em cada amostra. Como as características do tipo do fino influenciam no  $k_{sat}$ , principalmente finos não inertes, como a montmorilonita usada nas amostras Wang e Huang (1984), essas proporções entre diferentes finos interferiram na permeabilidade.

Destas amostras, conclui-se que:

- (1) amostras com mesmo teor de finos podem apresentar grande variação de  $k_{sat}$ ;
- (2) para mesmo teor de finos, em geral, as amostras com maior IP apresentam menor  $k_{sat}$ ;
- (3) nas amostras com mesmas frações granulométricas, as com menor índice de vazios apresentam menor  $k_{sat}$ ;
- (4) não somente o teor de grãos grossos e finos, mas a composição dessas matrizes granular e fina interferem no valor de  $k_{sat}$ ;
- (5) efeitos do índice de vazios nos valores de  $k_{sat}$  podem ser mais bem visualizados quando se utilizam misturas com mesma granulometria na matriz granular.

Portanto, equações de predição que não considerem, principalmente, as propriedades da fração fina, como os limites de consistência, podem incorrer em estimativas tendenciosas e de baixa acurácia de  $k_{sat}$ .

## 5 RESULTADOS E DISCUSSÕES

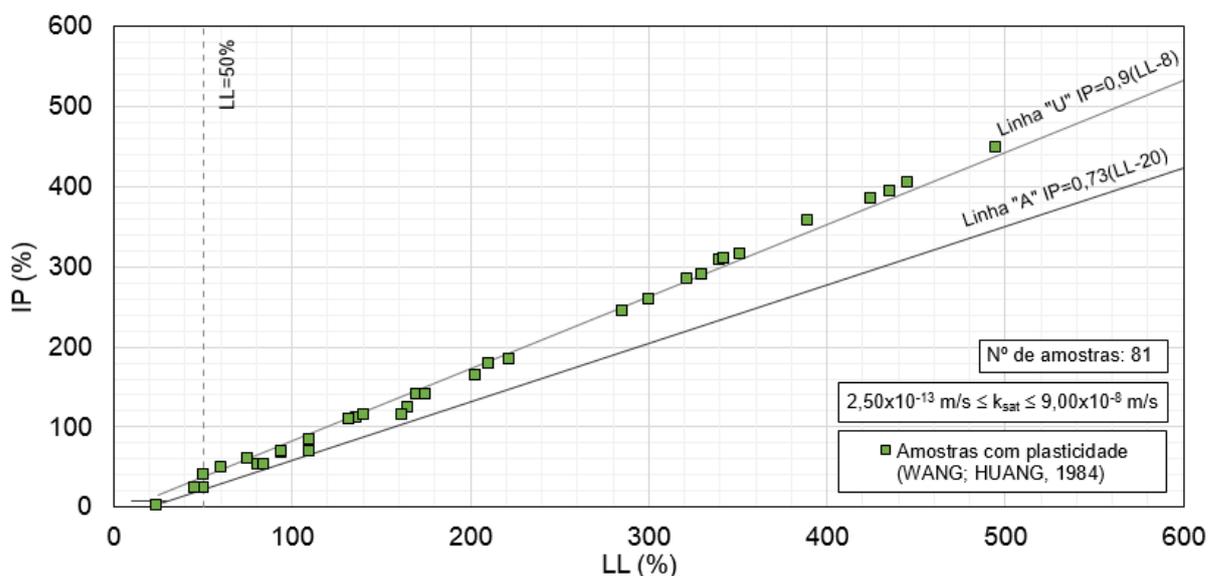
Neste capítulo são apresentados inicialmente os resultados das análises estatísticas e das redes neurais obtidos para dos principais dados experimentais investigados, selecionados a partir de Wang e Huang (1984). A análise comparativa com as correlações da literatura também é apresentada, bem como as correspondentes discussões sobre esses resultados.

### 5.1 Pré-tratamento dos dados experimentais

Os dados experimentais, que contemplam as amostras de Wang e Huang (1984), foram submetidos a uma etapa de pré-processamento dos dados. Um exame visual dos dados excluiu 3 amostras, resultando, portanto, os 111 conjuntos de dados apresentados no Capítulo 4. Essa base de 111 dados passou por uma etapa de pré-processamento mais detalhada para verificar a presença de valores discrepantes ou dados inconsistentes.

Primeiramente, 81 amostras, que apresentaram limites de consistência, foram verificadas no Ábaco de Casagrande e observou-se que, destas, 36 estavam acima da Linha U. O ábaco de Casagrande para as 81 amostras que apresentam plasticidade é apresentado no Gráfico 3. Observa-se a predominância de amostras de argilas de alta plasticidade (acima da linha A) e alta compressibilidade ( $LL > 50\%$ ).

Gráfico 3 - Ábaco de Casagrande para as 81 amostras com plasticidade (Dados de: WANG; HUANG, 1984)



Fonte: Autora (2022).

Algumas amostras localizaram-se ligeiramente acima da Linha U: 36 amostras. A diferença média entre o LP e 0,9(LL-8) (Limite da Linha U) foi de apenas 3,16%. Apenas 01 ponto amostral excedeu em mais de 10% o valor limite da Linha U. Wang e Huang (1984) determinaram os limites de consistência métodos da ASTM D 423-66 e ASTM D 424-59. Como os percentuais excedidos além da Linha U são relativamente pequenos e, dada a natureza operacional destes ensaios, optou-se por manter as amostras no conjunto de dados.

Sendo assim, das 111 amostras, 30 são amostras não plásticas e 81 são amostras plásticas. Os 111 dados originais podem ser consultados no Apêndice 4. A descrição estatística das 111 amostras selecionadas foi realizada no programa computacional Minitab e está organizada por plasticidade da amostra, mostrando os intervalos de valores máximo e mínimo de cada variável, além da média aritmética ( $\mu$ ), desvio padrão (SD) e o coeficiente de variação (CV), dado em porcentagem (%). São apresentadas na Tabela 8 e Tabela 9.

Tabela 8 – Estatística descritiva das amostras NP dos dados selecionados de Wang e Huang (1984) – 30 amostras

Variável	Média arit.	Desv. Padr.	CV (%)	Mínimo	Máximo
%argila	6,95	5,18	74,56	0,00	16,00
%silte	32,15	23,73	73,82	0,00	72,00
%areia	51,83	29,14	56,22	2,00	100,00
%pedregulho	9,07	7,48	82,48	0,00	19,00
%finos	39,09	28,89	73,91	0,00	88,00
d <sub>10</sub> (10 <sup>-5</sup> mm)	3955,00	7401,00	187,12	150,00	19000,00
d <sub>50</sub> (10 <sup>-4</sup> mm)	2623,00	2126,00	81,07	60,00	6200,00
d <sub>60</sub> (10 <sup>-4</sup> mm)	37671,00	29356,00	77,93	141,00	95090,00
C <sub>u</sub>	69,60	67,20	96,60	0,50	225,00
e	0,61	0,18	30,34	0,32	0,98
G <sub>s</sub>	2,73	0,07	2,41	2,67	2,88
$\rho_d$ (g/cm <sup>3</sup> )	1,72	0,17	9,84	1,38	2,03
k <sub>sat</sub> (m/s)	2,05x10 <sup>-6</sup>	4,24x10 <sup>-6</sup>	207	1,15x10 <sup>-9</sup>	1,70x10 <sup>-5</sup>
log(k <sub>sat</sub> ) (m/s)	-6,66	1,07	-16,08	-8,94	-4,77

Fonte: Autora (2022).

Tabela 9 - Estatística descritiva das amostras plásticas dos dados selecionados de Wang e Huang (1984) – 81 amostras

Variável	Média arit.	Desv. Padr.	CV (%)	Mínimo	Máximo
%argila	38,39	18,62	48,50	10,00	84,00
%silte	30,48	19,81	64,98	3,00	83,00
%areia	22,61	22,60	99,94	0,00	81,00
%pedregulho	8,39	7,69	91,62	0,00	20,00
%finos	68,87	22,96	33,34	19,00	100,00
LL (%)	2,75	0,04	1,61	2,69	2,87
LP (%)	241,40	151,00	62,54	24,00	495,00
IP (%)	32,32	9,63	29,79	3,00	47,00
d <sub>10</sub> (10 <sup>-5</sup> mm)	209,10	144,50	69,11	2,00	449,00
d <sub>50</sub> (10 <sup>-4</sup> mm)	35,56	46,61	131,10	3,80	220,00
d <sub>60</sub> (10 <sup>-4</sup> mm)	681,00	1307,00	191,79	4,00	4800,00
C <sub>u</sub>	11661,00	19250,00	165,08	58,00	73000,00
e	542,00	1078,00	198,86	2,00	6176,00
G <sub>s</sub>	0,88	0,21	24,41	0,45	1,38
ρ <sub>d</sub> (g/cm <sup>3</sup> )	1,48	0,16	10,97	1,16	1,87
k <sub>sat</sub> (m/s)	5,57x10 <sup>-09</sup>	1,52x10 <sup>-08</sup>	272,37	2,50x10 <sup>-13</sup>	9,00x10 <sup>-08</sup>
log(k <sub>sat</sub> ) (m/s)	-10,08	1,58	-15,71	-12,6	-7,05

Fonte: Autora (2022).

No Apêndice 5, são mostrados os histogramas das variáveis e os diagramas de caixa, separados por plasticidade das amostras. Tais gráficos foram construídos com o auxílio do programa computacional Minitab. Dos histogramas, infere-se que algumas variáveis seguem uma distribuição, aparentemente, do tipo normal. Quanto à análise dos diagramas de caixa, as variáveis G<sub>s</sub> e d<sub>10</sub> apresentam possíveis valores considerados estatisticamente como discrepantes.

Analisando os valores de d<sub>10</sub> das amostras não plásticas, existem 06 dados considerados discrepantes, estatisticamente. Após a verificação desses dados para verificar a origem dessa discrepância e tomada de decisão quanto à exclusão da base de dados ou não, constatou-se que são amostras com mais de 81% de fração grossa, correspondendo a valores considerados elevados de d<sub>10</sub> quando comparados aos demais dados. Sendo assim, optou-se por não excluir essas amostras, por ser considerada uma questão de variabilidade natural do solo e não resultante de um erro experimental.

O mesmo ocorreu no diagrama de caixa de d<sub>10</sub> das amostras com plasticidade; 06 dados foram identificados como discrepantes, mas correspondem a amostras que apresentaram percentual de pedregulho e areia, elevando o valor de d<sub>10</sub>. Pelas

mesmas razões, não foram removidos da base de dados. Sendo assim, os dados experimentais foram compostos por 111 conjuntos de amostras que contemplam os dados: %areia, %pedregulho, %silte, %argila, %finos,  $\gamma_d$ ,  $G_s$ , LL, LP, IP,  $d_{10}$ ,  $d_{50}$ ,  $d_{60}$ ,  $C_u$ ,  $e$ ,  $k_{sat}$  e  $\log(k_{sat})$ .

Diante da descrição dos dados experimentais e as considerações do Capítulo 2 e 4, foram selecionadas as variáveis de entrada e as cinco análises, descritas no Capítulo 3, foram realizadas sobre esse banco de dados.

Como o banco de dados amostral é composto por 27% de amostras NP, considerou-se os dados da granulometria, como o percentual de finos, o coeficiente de uniformidade e o índice de vazios. Uma parcela expressiva dos dados é composta por amostras plásticas e, como viu-se no Capítulo 2, o fator que influencia na permeabilidade deste tipo de solo é a característica da fração fina: o tipo de argilomineral, que terá influência sobre a dupla camada difusa e, conseqüentemente, na superfície específica destas partículas.

Por essa razão, optou-se por investigar o limite de liquidez na previsão dos valores de  $k_{sat}$ , como a característica a ser considerada para a fração fina. Na Análise 1, parte-se de um conjunto de variáveis de entrada que representam as características da fração fina: teor de finos e LL. Nas Análises seguintes, são acrescentadas diferentes características da granulometria: diâmetro efetivo, coeficiente de uniformidade e índice de vazios, que são propriedades principalmente importantes para caracterizar a permeabilidade de solos não plásticos.

Os resultados são apresentados a seguir divididos por Análise. Cada tópico seguinte apresentará os resultados das regressões lineares e polinomiais e das redes neurais considerando o conjunto de variáveis de entrada de cada Análise.

## **5.2 Resultados da Análise 1**

Na Análise 1, objetivou-se prever o  $\log(k_{sat})$  em função de apenas duas variáveis de entrada: LL e %Finos. Os 111 dados selecionados foram submetidos à análise estatística para a proposição de equações de regressão múltipla e ao treinamento de redes neurais. Posteriormente, estes resultados das equações propostas são comparados com os resultados fornecidos por correlações da literatura.

### 5.2.1 Resultados das análises estatísticas

Com o auxílio do programa Minitab, duas regressões foram propostas, considerando como variáveis independentes o LL e o %Finos: uma linear e uma polinomial, no caso, segundo o formato de polinômio de grau 2 com duas variáveis preditoras apresentado no Capítulo 3. O Tabela 10 resume as equações e os respectivos valores das medidas estatísticas de desempenho:  $R^2$ , RMSE e MAE.

Tabela 10 - Análise 1 – Resumo das equações de regressão propostas

Análise	Variáveis de entrada	Equação de regressão	$R^2$	RMSE	MAE	
1	Finos - LL	RL1	$\log(k_{sat}) = -6,961 - 0,010930 LL - 0,00437 \text{ Finos}$	0,805	0,926	0,720
		RNL1	$\log(k_{sat}) = -6,046 - 0,02361 LL - 0,0178 \text{ Finos} + 0,000010 LL^2 + 0,000024 \text{ Finos}^2 + 0,000113 LL * \text{Finos}$	0,881	0,723	0,587

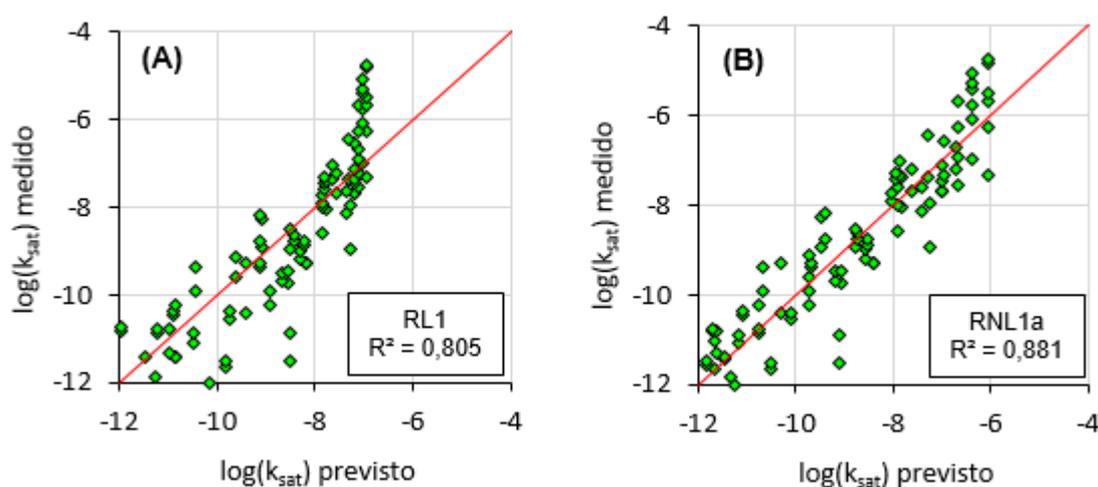
Onde:  
 LL – Limite de liquidez (%)  
 Finos – Percentual de finos (%)

Fonte: Autora (2022)

Analisando unicamente os valores de  $R^2$ , RMSE e MAE, observa-se que um ganho na qualidade do ajuste foi obtido quando proposta a equação polinomial. O

Gráfico 4, (a) e (b), mostra a dispersão entre valores previstos e medidos de  $\log(k_{sat})$  para RL1, e RNL1, respectivamente.

Gráfico 4 – Análise 1 – Ajuste dos valores medidos e previstos de  $\log(k_{sat})$



Fonte: Autora (2022).

Como descrito no Capítulo 2, testes de hipóteses foram realizados para verificar a significância estatística das equações acima. As duas equações foram submetidas à

avaliação dos resíduos e as verificações de normalidade da distribuição e da independência dos valores residuais. Os testes, conforme metodologia descrita no item 3.4, foram utilizados como critério para a validação da equação proposta. Assim, a homoscedasticidade pode ser verificada através dos gráficos de resíduos apresentados na Figura 33 e na Figura 34, no Apêndice 6, para as equações RL1 e RNL1, respectivamente.

Na Figura 33, é visualizada uma tendência nos resíduos, embora, por ser uma análise subjetiva, é necessário comprovar com a estatística de Durbin-Watson. Na distribuição dos resíduos da RNL1, não é possível identificar uma tendência dos valores de resíduos. Em todos os dois gráficos, são observados valores elevados de resíduos; pelo menos 04 observações apresentaram resíduo próximo ou acima de  $\pm 2$ . A Tabela 11 apresenta o resumo dos resultados dos testes de hipóteses realizados sobre a equação RL1.

Tabela 11 – Resumo dos resultados dos testes estatísticos para a equação RL1

Objetivo	Teste	valor-p	Conclusão
Teste dos termos da regressão	Teste t (intercepto)	$2,00 \times 10^{-16}$	Parâmetro significativo
	Teste t (LL)	$2,00 \times 10^{-16}$	Parâmetro significativo
	Teste t (Finos)	$4,25 \times 10^{-02}$	Parâmetro significativo
Teste da regressão	Teste F	$2,20 \times 10^{-16}$	Parâmetro significativo
Normalidade dos resíduos	Teste AD	$3,35 \times 10^{-01}$	Distribuição Normal dos resíduos
	Teste SF	$1,18 \times 10^{-01}$	Distribuição Normal dos resíduos
	Teste de LF	$3,39 \times 10^{-01}$	Distribuição Normal dos resíduos
Independência dos resíduos	Teste DW	$1,59 \times 10^{-03}$	Resíduos não são independentes
Multicolinearidade	VIF_LL	1,07	Não há multicolinearidade
	VIF_Finos	1,07	Não há multicolinearidade

Fonte: Autora (2022)

Os resultados dos testes t e F indicam que os coeficientes individuais da regressão são significantes em relação aos dados observados, uma vez que o valor-p obtido é menor do que a significância de 5%. Isto é, as variáveis selecionadas contribuem para a equação e a relação linear entre as variáveis preditoras e  $\log(k_{sat})$  é estatisticamente significativa.

O resultado do teste de aderência de Anderson-Darling (AD), Shapiro-Francia (SF) e Lilliefors (LF) apontam para a distribuição normal dos resíduos, com valor-p superior a significância de 5%. Todavia, o resultado da estatística de Durbin-Watson (DW), que

avalia a independência entre os resíduos, indica que os resíduos da equação RL1 não são independentes, já que o valor-p é menor que o nível de significância de 5%.

Considerando o número de 111 dados e 3 termos na equação, o valor de  $d_w$  deve estar contido no intervalo  $1,74 \leq d_w < 2,26$ . Para os resultados da RL1,  $d_w$  apresentou um valor de 1,61, portanto, rejeita-se a hipótese nula (os resíduos não são independentes). De fato, está de acordo com a análise visual da distribuição dos resíduos, que indicou uma tendência nos valores. O valor de VIF, que indica a existência ou não de colinearidade/multicolinearidade, comprovou que as variáveis não são correlacionadas.

A ausência da independência dos resíduos não atende aos requisitos para uma regressão estatisticamente significativa. O procedimento seguinte seria identificar os dados que apresentaram resíduos elevados e verificar se há algum problema com as informações daquela amostra para, então, decidir removê-la ou não do conjunto de dados. E, assim, estabelecer uma nova regressão.

Todavia, o objetivo dessa pesquisa é comparar resultados entre técnicas para predição de  $k_{sat}$  e, como as redes neurais, são treinadas com todos os 111 dados pré-processados, são comparados apenas o desempenho das equações de regressão desenvolvidas a partir do mesmo conjunto de amostras que as redes neurais, sem a remoção dos dados com resíduos elevados. A Tabela 12 apresenta o resumo dos resultados dos testes de hipóteses realizados sobre a equação RNL1.

Tabela 12 - Resumo dos resultados dos testes estatísticos para a equação RNL1

Objetivo	Teste	valor-p	Conclusão
Teste dos termos da regressão	Teste t (intercepto)	$2,00 \times 10^{-16}$	Parâmetro significativa
	Teste t (LL)	$2,00 \times 10^{-16}$	Parâmetro significativa
	Teste t (Finos)	$1,11 \times 10^{-01}$	Parâmetro <b>não</b> significativa
	Teste t (LL <sup>2</sup> )	$1,63 \times 10^{-02}$	Parâmetro significativa
	Teste t (Finos <sup>2</sup> )	$8,3 \times 10^{-01}$	Parâmetro <b>não</b> significativa
	Teste t (LL*Finos)	$2,05 \times 10^{-04}$	Parâmetro significativa
Teste da regressão	Teste F	$2,20 \times 10^{-16}$	Parâmetro significativa
Normalidade dos resíduos	Teste de LF	$7,37 \times 10^{-02}$	Distribuição Normal dos resíduos
	Teste AD	$3,44 \times 10^{-01}$	Distribuição Normal dos resíduos
	Teste SF	$9,89 \times 10^{-02}$	Distribuição Normal dos resíduos
Independência dos resíduos	Teste DW	$3,59 \times 10^{-04}$	Resíduos <b>não</b> são independentes
Multicolinearidade	VIF_LL	22,63	Há multicolinearidade
	VIF_Finos	19,29	Há multicolinearidade
	VIF_LL <sup>2</sup>	19,81	Há multicolinearidade
	VIF_Finos <sup>2</sup>	24,66	Há multicolinearidade
	VIF_Finos*LL	34,63	Há multicolinearidade

Fonte: Autora (2022)

Os resultados dos testes F e t indicam que os parâmetros de regressão são significantes em relação aos dados observados, uma vez que o valor-p obtido é menor do que a significância de 5%, exceto apenas os termos Finos e Finos<sup>2</sup>, o que indica que a equação carece de ajustes.

O resultado do teste de aderência de Anderson-Darling (AD), Shapiro-Francia (SF) e Lilliefors (LF) apontam para a distribuição normal dos resíduos, com valor-p superior a significância de 5%. Já, o resultado da estatística de Durbin-Watson (DW) indica que os resíduos da equação RNL1 também não são independentes, já que o valor-p é menor que o nível de significância de 5%. Neste caso, o valor de dw foi de 1,40; novamente fora do intervalo de aceitação da independência dos resíduos que, para esta quantidade de dados e variáveis independentes, é  $1,81 \leq dw < 2,19$ .

O valor de VIF indicou a existência de correlação para todos os termos da regressão, mas, como abordado no Capítulo 2, isso se dá pela existência de produto entre as variáveis e de variáveis elevada à potência. Conclui-se, entretanto, que a ausência da

independência dos resíduos não atende aos requisitos para uma regressão estatisticamente significativa e que ajustes devem ser realizados na equação.

### 5.2.2 Resultados das redes neurais

O conjunto de 111 amostras foi utilizado para o treinamento de 25 arquiteturas de redes neurais. Um resumo dos resultados das medidas estatísticas de desempenho, para cada fase de calibração das redes, é mostrado na Tabela 13.

Tabela 13 – Análise 1 – Resumo das arquiteturas de RNA treinadas

RNA	Nº DE NEURÔNIOS			PARÂMETROS ESTATÍSTICOS								
	CAMADA			TREINO			TESTE			VALIDAÇÃO		
	Nº 1	Nº 2	Nº 3	R <sup>2</sup>	RMSE	MAE	R <sup>2</sup>	RMSE	MAE	R <sup>2</sup>	RMSE	MAE
RNA101	2			0,834	0,840	0,745	0,930	0,686	0,571	0,917	0,467	0,357
RNA102	4			0,815	0,866	0,700	0,922	0,648	0,522	0,919	0,437	0,346
RNA103	6			0,827	0,784	0,602	0,929	0,605	0,484	0,915	0,449	0,349
<b>RNA104</b>	<b>8</b>			<b>0,808</b>	<b>0,837</b>	<b>0,625</b>	<b>0,931</b>	<b>0,599</b>	<b>0,494</b>	<b>0,926</b>	<b>0,437</b>	<b>0,351</b>
RNA105	10			0,817	0,821	0,620	0,934	0,695	0,584	0,914	0,520	0,398
RNA106	2	1		0,837	0,786	0,593	0,933	0,791	0,670	0,915	0,598	0,478
RNA107	4	2		0,831	0,758	0,581	0,930	0,601	0,487	0,916	0,447	0,351
RNA108	6	3		0,821	0,781	0,587	0,924	1,024	0,928	0,907	0,870	0,732
RNA109	8	4		0,815	0,910	0,752	0,924	0,619	0,445	0,890	0,571	0,474
RNA110	10	5		0,821	0,800	0,612	0,938	0,580	0,440	0,885	0,533	0,426
RNA111	1	2		0,711	1,054	0,757	0,842	0,861	0,714	0,858	0,628	0,489
RNA112	2	4		0,826	0,847	0,679	0,932	0,845	0,727	0,913	0,643	0,519
RNA113	3	6		0,827	0,790	0,616	0,930	0,718	0,603	0,919	0,515	0,412
RNA114	4	8		0,816	0,805	0,624	0,929	0,749	0,626	0,914	0,558	0,445
RNA115	5	10		0,831	0,763	0,601	0,927	0,662	0,552	0,905	0,525	0,407
RNA116	2	1	1	0,000	2,070	1,587	0,725	2,491	2,259	0,597	1,589	1,237
RNA117	4	2	2	0,719	1,261	1,024	0,925	0,684	0,582	0,901	0,540	0,415
RNA118	6	3	3	0,796	0,830	0,627	0,927	0,703	0,598	0,905	0,521	0,402
RNA119	8	4	4	0,810	0,819	0,627	0,927	0,622	0,494	0,899	0,506	0,393

Fonte: Autora (2022).

Tabela 14 – Análise 1 – Resumo das arquiteturas de RNA treinadas – cont.

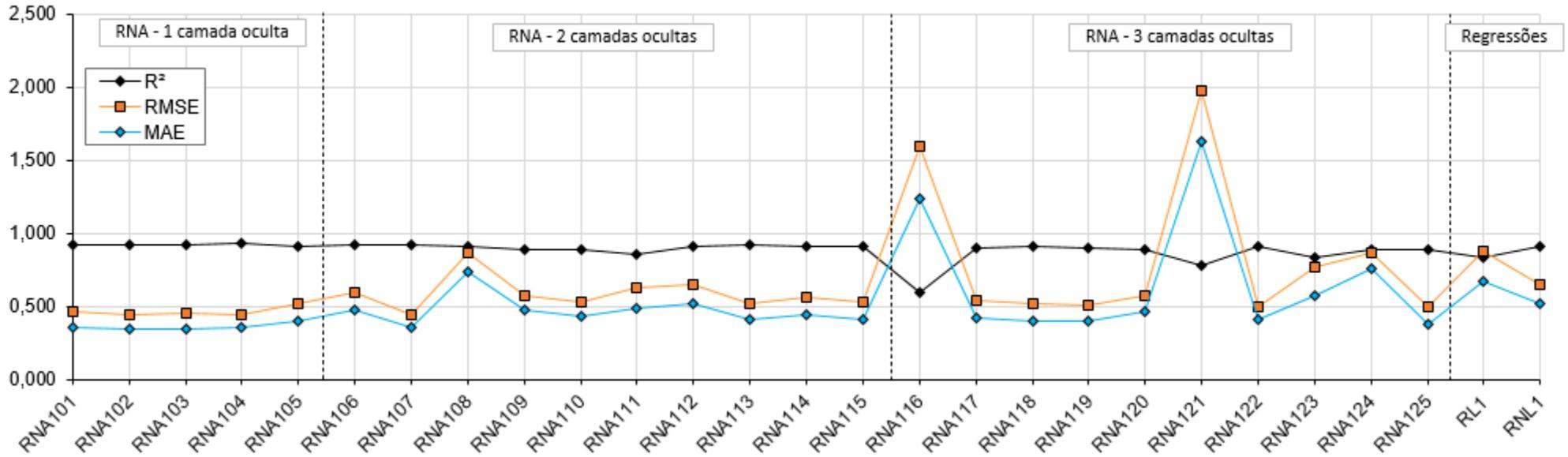
RNA	Nº DE NEURÔNIOS			PARÂMETROS ESTATÍSTICOS								
	CAMADA			TREINO			TESTE			VALIDAÇÃO		
	Nº 1	Nº 2	Nº 3	R <sup>2</sup>	RMSE	MAE	R <sup>2</sup>	RMSE	MAE	R <sup>2</sup>	RMSE	MAE
RNA120	10	5	5	0,795	0,862	0,668	0,929	0,630	0,466	0,886	0,575	0,461
RNA121	2	2	1	0,207	1,823	1,509	0,769	2,979	2,652	0,779	1,970	1,627
RNA122	4	4	2	0,717	1,178	0,913	0,902	0,739	0,619	0,909	0,500	0,405
RNA123	6	6	3	0,812	0,831	0,658	0,814	1,085	0,875	0,830	0,770	0,575
RNA124	8	8	4	0,808	0,824	0,640	0,931	0,894	0,791	0,889	0,870	0,756
RNA125	10	10	5	0,808	0,807	0,614	0,925	0,656	0,508	0,892	0,499	0,377

Fonte: Autora (2022).

Considerando os resultados da fase de validação cruzada, observa-se que, entre as redes com 1 camada oculta (RNA101 a RNA105) apresentaram melhor desempenho, isto é, maior ajuste e menores erros na predição. Numericamente, as redes com 2 camadas, com número de neurônios da 2ª camada igual à metade dos neurônios da 1ª camada (RNA106 a RNA110), apresentaram desempenho similar às arquiteturas de 1 camada, mas com desempenho ligeiramente inferior, exceto a RNA108, que apresentou maiores RMSE e MAE. As redes com 03 camadas ocultas apresentaram maior instabilidade dos resultados e menor desempenho dentre as arquiteturas treinadas.

Avaliando apenas as métricas R<sup>2</sup>, RMSE e MAE da validação cruzada, as redes apresentaram desempenho equivalente à regressão polinomial, para o banco de dados utilizado. Considerando as métricas estatísticas utilizadas, a arquitetura que apresentou o melhor desempenho relativo, na fase de validação cruzada, foi a rede com uma camada oculta, com 8 neurônios (RNA104), embora pode-se verificar pouca alteração do desempenho nas redes de 1 camada. O Gráfico 5 apresenta graficamente a comparação entre os desempenhos das RNA e equações de regressão para a Análise 1.

Gráfico 5 - Análise 1 - Resultados estatísticos das RNA (fase de validação cruzada) e das equações de regressão propostas



Fonte: Autora (2022)

O uso de três camadas intermediárias prejudicou substancialmente o desempenho das arquiteturas treinadas. E as arquiteturas com duas camadas intermediárias apresentaram resultados semelhantes quando o número de neurônios da segunda camada foi igual a metade do número de neurônios da primeira e vice-versa.

Por esse motivo, para as próximas redes modeladas, descartou-se o uso de três camadas intermediárias e apenas considerou, quando mais de uma camada oculta, o número de neurônios da segunda ser a metade dos neurônios da primeira camada intermediária.

### 5.3 Análise 2

Na Análise 2, objetivou-se prever o  $\log(k_{\text{sat}})$  em função de, agora, três variáveis de entrada: LL, %Finos e  $d_{10}$ . Os 111 conjuntos de dados selecionados foram submetidos à análise estatística para a proposição de equações de regressão múltipla e ao treinamento de redes neurais. Posteriormente, os resultados são comparados com os fornecidos por correlações da literatura.

#### 5.3.1 Resultados das análises estatísticas

Os dados foram submetidos ao programa de análise estatística para o desenvolvimento e análise de equações de regressão. Novamente, uma equação linear e uma equação polinomial foram testadas e avaliadas. A Tabela 15 resume as equações e os respectivos valores das medidas estatísticas de desempenho:  $R^2$ ,  $R^2_{\text{ajust}}$ , RMSE e MAE.

Tabela 15 - Análise 2 – Resumo das equações de regressão propostas

Análise	Variáveis de entrada	Equação de regressão		$R^2$	RMSE	MAE
2	Finos – LL – $d_{10}$	RL2	$\log(k_{\text{sat}}) = -7,574 - 0,010831 \text{ LL} + 0,000102 \text{ } d_{10} + 0,00359 \text{ Finos}$	0,834	0,855	0,653
		RNL2	$\log(k_{\text{sat}}) = -8,399 - 0,01916 \text{ LL} + 0,003161 \text{ } d_{10} + 0,0350 \text{ Finos} + 0,000007 \text{ LL}^2 - 0,00000016 \text{ } d_{10}^2 - 0,000294 \text{ Finos}^2 + 0,000023 \text{ LL} * d_{10} + 0,000077 \text{ LL} * \text{Finos} - 0,000027 \text{ } d_{10} * \text{Finos}$	0,909	0,632	0,514

Onde:

LL – Limite de liquidez (%)

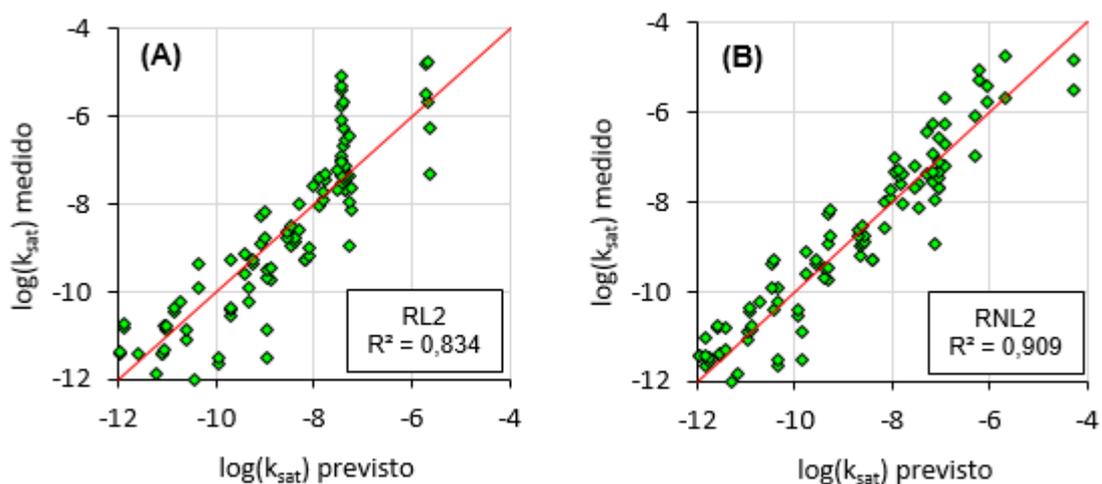
Finos – Percentual de finos (%)

$d_{10}$  – Diâmetro efetivo ( $10^{-5}$  mm)

Fonte: Autora (2022)

Analisando unicamente os valores de  $R^2$ , RMSE e MAE, o acréscimo da variável  $d_{10}$  melhorou o desempenho quanto aos erros da RL2, comparando à equação RL1, mas provocou uma pequena redução do  $R^2$ . Porém, quando avaliada a equação de regressão polinomial, o acréscimo de  $d_{10}$  melhorou o desempenho, estatisticamente, quando comparado aos resultados da RNL1, isto é, aumentou o  $R^2$  e diminuiu os erros da equação. O Gráfico 6, (a) e (b), mostra a dispersão entre valores previstos e medidos de  $\log(k_{sat})$  para RL2 e RNL2, respectivamente.

Gráfico 6 - Análise 2 – Valores medidos e previstos de  $\log(k_{sat})$



Fonte: Autora (2022).

Os testes para verificar a significância estatística e o atendimento aos requisitos para as regressões foram conduzidos e os resultados para a equação RL2 são apresentados na Tabela 16.

Tabela 16 – Análise 2 - Resumo dos resultados dos testes estatísticos para a equação RL2

Objetivo	Teste	valor-p	Conclusão
Teste dos termos da regressão	Teste t (intercepto)	$2,00 \times 10^{-16}$	Parâmetro significativo
	Teste t (LL)	$2,00 \times 10^{-16}$	Parâmetro significativo
	Teste t (Finos)	$3,45 \times 10^{-01}$	Parâmetro <b>não</b> significativo
	Teste t (d10)	$3,32 \times 10^{-05}$	Parâmetro significativo
Teste da regressão	Teste F	$2,20 \times 10^{-16}$	Parâmetro significativo
Normalidade dos resíduos	Teste AD	$1,18 \times 10^{-01}$	Distribuição Normal dos resíduos
	Teste SF	$1,49 \times 10^{-01}$	Distribuição Normal dos resíduos
	Teste de LF	$1,81 \times 10^{-01}$	Distribuição Normal dos resíduos
Independência dos resíduos	Teste DW	$1,18 \times 10^{-01}$	Resíduos independentes
Multicolinearidade	VIF_LL	1,26	Não há multicolinearidade
	VIF_Finos	1,64	Não há multicolinearidade
	VIF_d10	1,41	Não há multicolinearidade

Fonte: Autora (2022).

Os resultados dos testes de hipóteses (Testes F e t) indicam que os parâmetros de regressão são significantes em relação aos dados observados, uma vez que o valor-p obtido é menor do que a significância de 5%, exceto apenas o termo %Finos que não contribui significativamente para o ajuste da equação, neste caso.

O resultado do teste de aderência de Anderson-Darling (AD), Shapiro-Francia (SF) e Lilliefors (LF) apontam para a distribuição normal dos resíduos, com valor-p superior a significância de 5%. Já, o resultado da estatística de Durbin-Watson (DW) indica que os resíduos da equação RL2 são independentes, já que o valor-p é maior que o nível de significância de 5%. Com o acréscimo de uma variável, o intervalo para aceitação da hipótese nula (os resíduos são independentes) na estatística de Durbin-Watson é  $1,76 \leq d_w < 2,24$ . Neste caso, o valor de  $d_w$  foi de 1,80; dentro do intervalo de aceitação da independência dos resíduos.

O valor de VIF não indicou a existência de correlação entre as variáveis. Conclui-se, entretanto, que a equação atende aos requisitos de normalidade e independência dos resíduos, além da ausência de multicolinearidade. Entretanto, um refinamento na escolha dos termos da equação pode melhorar seu desempenho estatístico. Os testes realizados sobre a regressão polinomial RNL2 são resumidos e apresentados na Tabela 17.

Tabela 17 - Análise 2 - Resumo dos resultados dos testes estatísticos para a equação RNL2

Objetivo	Teste	valor-p	Conclusão
Teste dos termos da regressão	Teste t (intercepto)	$2,00 \times 10^{-16}$	Parâmetro significativa
	Teste t (LL)	$9,96 \times 10^{-11}$	Parâmetro significativa
	Teste t (Finos)	$3,62 \times 10^{-02}$	Parâmetro significativa
	Teste t (LL <sup>2</sup> )	$1,68 \times 10^{-01}$	Parâmetro <b>não</b> significativa
	Teste t (Finos <sup>2</sup> )	$3,75 \times 10^{-02}$	Parâmetro significativa
	Teste t (d <sub>10</sub> <sup>2</sup> )	$1,87 \times 10^{-06}$	Parâmetro significativa
	Teste t (LL*Finos)	$6,93 \times 10^{-03}$	Parâmetro significativa
	Teste t (d <sub>10</sub> *Finos)	$3,85 \times 10^{-01}$	Parâmetro <b>não</b> significativa
	Teste t (d <sub>10</sub> *LL)	$4,60 \times 10^{-01}$	Parâmetro <b>não</b> significativa
Teste da regressão	Teste F	$2,20 \times 10^{-16}$	Parâmetro significativa
Normalidade dos resíduos	Teste AD	$5,59 \times 10^{-01}$	Distribuição Normal dos resíduos
	Teste SF	$3,57 \times 10^{-01}$	Distribuição Normal dos resíduos
	Teste de LF	$3,28 \times 10^{-01}$	Distribuição Normal dos resíduos
Independência dos resíduos	Teste DW	$2,59 \times 10^{-03}$	Resíduos <b>não</b> são independentes
Multicolinearidade	VIF_LL	49,78	Há multicolinearidade
	VIF_Finos	52,94	Há multicolinearidade
	VIF_d <sub>10</sub>	1644,29	Há multicolinearidade
	VIF_LL <sup>2</sup>	40,96	Há multicolinearidade
	VIF_Finos <sup>2</sup>	45,63	Há multicolinearidade
	VIF_(d <sub>10</sub> <sup>2</sup> )	1575,48	Há multicolinearidade
	VIF_(LL*Finos)	40,08	Há multicolinearidade
	VIF_(d <sub>10</sub> *Finos)	6,58	Não há multicolinearidade
	VIF_(d <sub>10</sub> *LL)	1,74	Não há multicolinearidade

Fonte: Autora (2022).

Os resultados dos testes estatísticos indicaram, resumidamente, que esta equação também precisa de ajustes quanto aos resíduos, considerados não independentes. As variáveis escolhidas contribuem para a regressão, mas os termos LL<sup>2</sup>, d<sub>10</sub>\*Finos e d<sub>10</sub>\*LL não estão significativamente relacionados com a variável resposta nesta equação.

A distribuição normal dos resíduos foi constatada em todos os testes realizados para esta finalidade e o valor de VIF indicou a existência de multicolinearidade, que se atribui a existência de produto entre variáveis e variável elevada a potência. Sem realizar a exclusão dos resíduos elevados e um novo ajuste, esta equação de regressão, não atendeu os critérios necessários para sua significância estatística.

### 5.3.2 Resultados das redes neurais

As redes neurais se mostraram promissoras para a tarefa de predição de  $k_{sat}$  quando comparadas às regressões realizadas anteriormente. A autonomia dos algoritmos de aprendizagem representa uma vantagem em relação às análises estatísticas, que solicitam demasiada intervenção do pesquisador para o ajuste da equação, a começar pela necessidade do conhecimento prévio da relação matemática entre as variáveis do problema. A Tabela 18 resume os resultados das arquiteturas de RNA treinadas na Análise 2, isto é, considerando também o  $d_{10}$  como variável de entrada.

Tabela 18 - Análise 2 – Resumo das arquiteturas de RNA treinadas

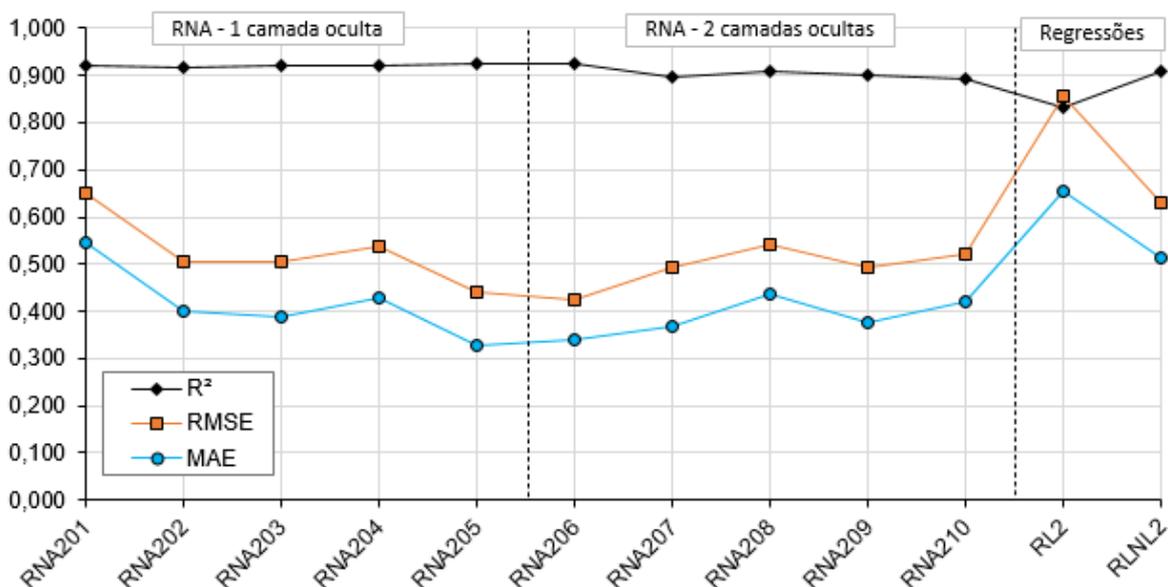
RNA	Nº DE NEURÔNIOS		PARÂMETROS ESTATÍSTICOS								
			TREINO			TESTE			VALIDAÇÃO		
	CAMADA 1	CAMADA 2	R <sup>2</sup>	RMSE	MAE	R <sup>2</sup>	RMSE	MAE	R <sup>2</sup>	RMSE	MAE
RNA201	2		0,839	0,796	0,590	0,933	0,802	0,688	0,920	0,650	0,544
RNA202	4		0,841	0,726	0,538	0,925	0,703	0,579	0,919	0,506	0,400
RNA203	6		0,823	0,963	0,836	0,933	0,678	0,560	0,921	0,504	0,389
RNA204	8		0,832	0,806	0,605	0,935	0,709	0,581	0,920	0,536	0,430
<b>RNA205</b>	<b>10</b>		<b>0,838</b>	<b>0,795</b>	<b>0,608</b>	<b>0,934</b>	<b>0,620</b>	<b>0,507</b>	<b>0,925</b>	<b>0,440</b>	<b>0,329</b>
RNA206	2	1	0,818	0,796	0,629	0,920	0,690	0,572	0,924	0,426	0,341
RNA207	4	2	0,839	0,773	0,611	0,933	0,583	0,471	0,896	0,494	0,369
RNA208	6	3	0,792	0,840	0,631	0,928	0,707	0,590	0,910	0,543	0,436
RNA209	8	4	0,821	0,779	0,568	0,938	0,601	0,484	0,902	0,492	0,378
RNA210	10	5	0,813	0,881	0,720	0,932	0,574	0,439	0,893	0,520	0,422

Fonte: Autora (2022).

Novamente, as arquiteturas de RNA tiveram desempenho superior às equações de regressões. Embora a diferença nos valores de  $R^2$  seja pequena, as redes apresentaram menor desvio entre o valor previsto e medido de  $k_{sat}$ . Dentre as arquiteturas treinadas, a arquitetura com uma camada intermediária com 10 neurônios, isto é, a rede RNA205, foi a que apresentou o melhor desempenho relativo, embora todas as arquiteturas tenham fornecido resultados parecidos.

Arquiteturas com mais de duas camadas, como dito anteriormente, não foram testadas a partir da Análise 2. O Gráfico 7 mostra os resultados entre os desempenhos das redes neurais e das regressões para a Análise 2.

Gráfico 7 - Análise 2 - Resultados estatísticos das RNA (fase de validação cruzada) e das equações de regressão propostas



Fonte: Autora (2022).

Identifica-se, de fato, uma similaridade nos resultados de  $R^2$  entre as redes e as regressões, com ligeira vantagem para as redes. Já os valores RMSE e MAE foram sensivelmente menores que os apresentados pelas regressões. De forma geral, observa-se pouca alteração no desempenho estatístico das redes com o acréscimo de camadas e neurônios.

## 5.4 Análise 3

Na Análise 3, uma quarta variável foi adicionada às equações para predição de  $k_{sat}$ : o coeficiente de uniformidade. Assim, as equações utilizaram as variáveis de entrada: LL, %Finos,  $d_{10}$  e  $C_u$ . Os 111 conjuntos de dados selecionados foram submetidos à análise estatística para a proposição de equações de regressão múltipla e ao treinamento de redes neurais.

### 5.4.1 Resultados das análises estatísticas

A Tabela 19 resume as equações e os respectivos valores das medidas estatísticas de desempenho ( $R^2$ , RMSE e MAE) para as equações linear e polinomial propostas.

Tabela 19 - Análise 3 – Resumo das equações de regressão propostas

Análise	Variáveis de entrada	Equação de regressão	R <sup>2</sup>	RMSE	MAE	
3	Finos – LL – d <sub>10</sub> – C <sub>u</sub>	RL3	$\log(k_{sat}) = -7,214 - 0,010277 LL + 0,000082 d_{10} - 0,000264 C_u - 0,00179 \text{ Finos}$	0,845	0,828	0,624
		RNL3	$\log(k_{sat}) = -10,19 - 0,01670 LL + 0,002058 d_{10} - 0,001517 C_u + 0,0727 \text{ Finos} + 0,000004 LL^2 - 0,000000106 d_{10}^2 + 0,0000000153 C_u^2 - 0,000481 \text{ Finos}^2 + 0,000011 LL * d_{10} + 0,000004 LL * C_u + 0,000064 LL * \text{Finos} + 0,000034 d_{10} * C_u - 0,000024 d_{10} * \text{Finos} + 0,000001 C_u * \text{Finos}$	0,920	0,593	0,461

Onde:

LL – Limite de liquidez (%)

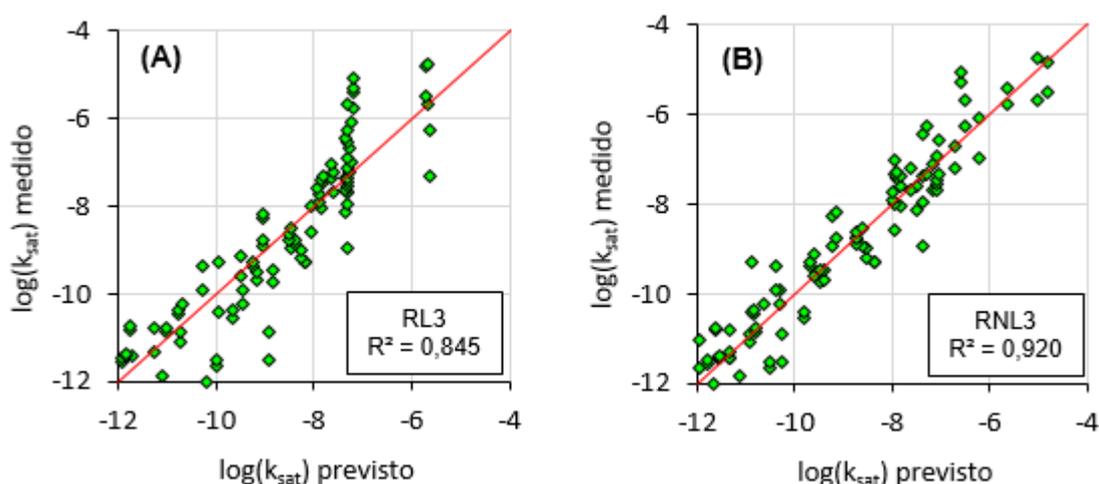
Finos – Percentual de finos (%)

d<sub>10</sub> – Diâmetro efetivo (10<sup>-5</sup> mm)

C<sub>u</sub> – Coeficiente de uniformidade (adimensional)

Fonte: Autora (2022)

A dispersão entre valores medidos e previsto foi analisada e o resultado gráfico é mostrado no Gráfico 8, (a) e (b), para os resultados de RL3 e RNL3, respectivamente.

Gráfico 8 - Análise 3 – Valores medidos e previstos de log(k<sub>sat</sub>)

Fonte: Autora (2022).

A inclusão da quarta variável não melhorou o desempenho da equação linear; pois houve um ligeiro aumento nos erros. A regressão polinomial aumentou o ajuste e reduziu os erros, mas a equação é sobrecarregada em comparação com as regressões das análises 1 e 2 (ao todo, 16 termos, incluindo a constante) e sem ganhos expressivos na qualidade do ajuste. Os coeficientes dos termos d<sub>10</sub><sup>2</sup> e C<sub>u</sub><sup>2</sup> são da ordem de 10<sup>-7</sup> e 10<sup>-8</sup>, isto é, matematicamente, pouco contribuem para o resultado de log(k<sub>sat</sub>). Já os maiores coeficientes foram para os termos LL e Finos. Além disso, há de destacar que C<sub>u</sub> e d<sub>10</sub> são correlacionados, o que justifica os resultados dos testes de hipóteses.

As equações foram submetidas aos testes de hipóteses para conferência do atendimento aos requisitos exigidos para as regressões e os resultados da regressão linear são sumarizados na Tabela 20.

Tabela 20 - Análise 3 - Resumo dos resultados dos testes estatísticos para a equação RL3

Objetivo	Teste	valor-p	Conclusão
Teste dos termos da regressão	Teste t (intercepto)	$2,00 \times 10^{-16}$	Parâmetro significativa
	Teste t (LL)	$2,00 \times 10^{-16}$	Parâmetro significativa
	Teste t (Finos)	$6,73 \times 10^{-01}$	Parâmetro <b>não</b> significativa
	Teste t ( $d_{10}$ )	$9,27 \times 10^{-04}$	Parâmetro significativa
	Teste t ( $C_u$ )	$9,33 \times 10^{-03}$	Parâmetro significativa
Teste da regressão	Teste F	$2,20 \times 10^{-16}$	Parâmetro significativa
Normalidade dos resíduos	Teste AD	$6,18 \times 10^{-02}$	Distribuição Normal dos resíduos
	Teste SF	$7,59 \times 10^{-02}$	Distribuição Normal dos resíduos
	Teste de LF	$2,41 \times 10^{-01}$	Distribuição Normal dos resíduos
Independência dos resíduos	Teste DW	$1,13 \times 10^{-02}$	Resíduos <b>não</b> são independentes
Multicolinearidade	VIF_LL	1,44	Não há multicolinearidade
	VIF_Finos	1,56	Não há multicolinearidade
	VIF_ $d_{10}$	1,36	Não há multicolinearidade
	VIF_ $C_u$	2,13	Não há multicolinearidade

Fonte: Autora (2022).

Os resultados mostram que o parâmetro Finos não foi estatisticamente significativo para a previsão de  $k_{sat}$ . A normalidade dos resíduos foi confirmada, embora não haja independência, pela estatística de Durbin-Watson.

Para cinco termos e 111 dados, o intervalo de  $d_w$  para aceitação da hipótese nula passa a ser  $1,78 \leq d_w < 2,22$ . Para a equação RL3, o valor de  $d_w$  foi de 1,68, portanto, rejeita-se a hipótese nula e os resíduos são considerados não independentes. Não foi detectada multicolinearidade entre as variáveis independentes, isto é, a presença dessas variáveis não inflou a variância da equação de forma significativa.

Os testes realizados sobre a regressão polinomial RNL3 são resumidos e apresentados na Tabela 21.

Tabela 21 - Análise 3 - Resumo dos resultados dos testes estatísticos para a equação RNL3

Objetivo	Teste	valor-p	Conclusão
Teste dos termos da regressão	Teste t (intercepto)	$3,09 \times 10^{-16}$	Parâmetro significativo
	Teste t (LL)	$3,10 \times 10^{-05}$	Parâmetro significativo
	Teste t (Finos)	$5,98 \times 10^{-03}$	Parâmetro significativo
	Teste t (Cu)	$3,96 \times 10^{-03}$	Parâmetro significativo
	Teste t (d <sub>10</sub> )	$1,97 \times 10^{-02}$	Parâmetro significativo
	Teste t (Finos <sup>2</sup> )	$4,51 \times 10^{-01}$	Parâmetro <b>não</b> significativo
	Teste t (LL <sup>2</sup> )	$1,38 \times 10^{-02}$	Parâmetro significativo
	Teste t (Cu <sup>2</sup> )	$4,54 \times 10^{-03}$	Parâmetro significativo
	Teste t (d <sub>10</sub> <sup>2</sup> )	$8,34 \times 10^{-01}$	Parâmetro <b>não</b> significativo
	Teste t (LL*d <sub>10</sub> )	$7,16 \times 10^{-01}$	Parâmetro <b>não</b> significativo
	Teste t (LL*Cu)	$1,22 \times 10^{-01}$	Parâmetro <b>não</b> significativo
	Teste t (LL*Finos)	$1,25 \times 10^{-01}$	Parâmetro <b>não</b> significativo
	Teste t (d <sub>10</sub> *Cu)	$3,02 \times 10^{-03}$	Parâmetro significativo
	Teste t (d <sub>10</sub> *Finos)	$4,31 \times 10^{-01}$	Parâmetro <b>não</b> significativo
	Teste t (Cu*Finos)	$9,42 \times 10^{-01}$	Parâmetro <b>não</b> significativo
Teste da regressão	Teste F	$2,20 \times 10^{-16}$	Parâmetro significativo
Normalidade dos resíduos	Teste AD	$4,15 \times 10^{-01}$	Distribuição Normal dos resíduos
	Teste SF	$6,43 \times 10^{-01}$	Distribuição Normal dos resíduos
	Teste de LF	$5,15 \times 10^{-01}$	Distribuição Normal dos resíduos
Independência dos resíduos	Teste DW	$1,10 \times 10^{-03}$	Resíduos <b>não</b> são independentes
Multicolinearidade	VIF_(LL)	111,43	Há multicolinearidade
	VIF_(Finos)	141,05	Há multicolinearidade
	VIF_(Cu)	98,28	Há multicolinearidade
	VIF_(d <sub>10</sub> )	2309,28	Há multicolinearidade
	VIF_(Finos <sup>2</sup> )	93,46	Há multicolinearidade
	VIF_(LL <sup>2</sup> )	44,97	Há multicolinearidade
	VIF_(Cu <sup>2</sup> )	37,51	Há multicolinearidade
	VIF_(d <sub>10</sub> <sup>2</sup> )	2172,16	Há multicolinearidade
	VIF_(LL*d <sub>10</sub> )	1,98	Não há multicolinearidade
	VIF_(LL*Cu)	147,79	Há multicolinearidade
	VIF_(LL*Finos)	93,13	Há multicolinearidade
	VIF_(d <sub>10</sub> *Cu)	21,13	Há multicolinearidade
	VIF_(d <sub>10</sub> *Finos)	6,74	Há multicolinearidade
VIF_(Cu*Finos)	103,44	Há multicolinearidade	

Fonte: Autora (2022).

Os resultados dos testes estatísticos indicaram que este formato de equação polinomial não produziu bom desempenho quando utilizadas as quatro variáveis. A equação resultante foi sobrecarregada de termos que não são fortemente significativos para a previsão de  $k_{sat}$ . Além da análise dos resíduos, cujo teste de hipótese verificou a não independência.

Embora a distribuição normal dos resíduos tenha sido constatada em todos os testes realizados para esta finalidade, o valor de VIF indicou a existência de multicolinearidade entre praticamente todos os termos, que se atribui à escolha de interações entre variáveis e variáveis elevadas à potência. Pode-se concluir que esta equação não atendeu aos critérios necessários para sua significância estatística. Nestes casos, técnicas de seleção de termos podem auxiliar para um modelo mais significativo e a inclusão de termos que, de fato, contribuem para a previsão da variável resposta. Montgomery e Runger (2012) e Maroco (2007) apresentam algumas técnicas de seleção de variáveis que podem ser aplicadas em regressões polinomiais deste tipo.

#### 5.4.2 Resultados das redes neurais

As redes neurais foram treinadas seguindo os mesmos critérios que a regressões: mesmos 111 dados e mesmas variáveis de entrada. Os resultados para as 10 arquiteturas são resumidos na Tabela 22.

Tabela 22 - Análise 3 – Resumo das arquiteturas de RNA treinadas

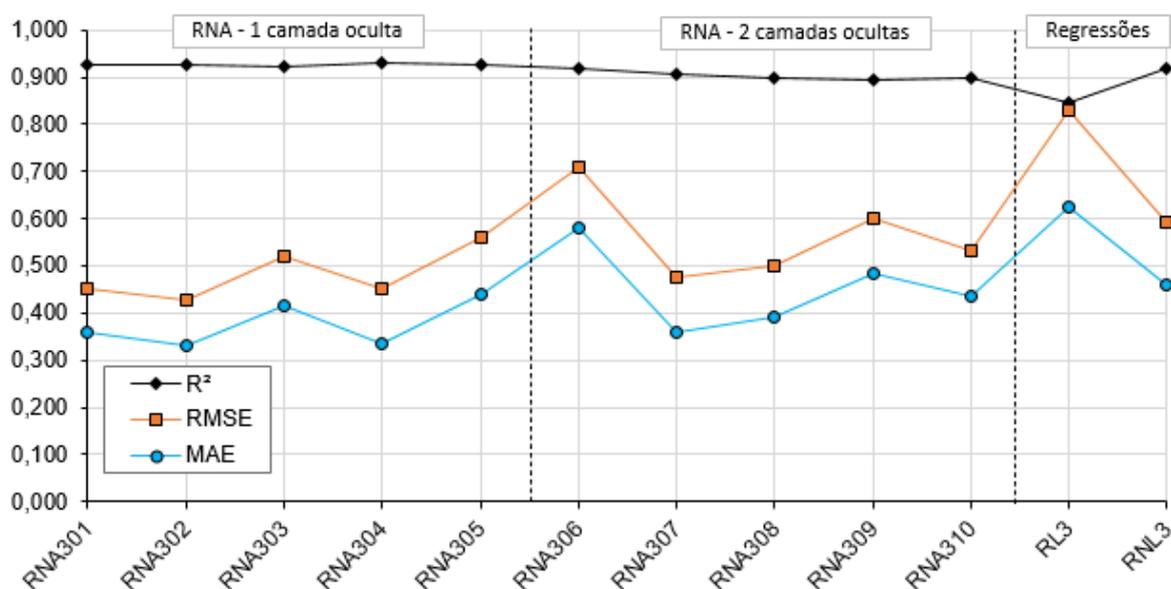
RNA	Nº DE NEURÔNIOS		PARÂMETROS ESTATÍSTICOS								
			TREINO			TESTE			VALIDAÇÃO		
	CAMADA 1	CAMADA 2	R <sup>2</sup>	RMSE	MAE	R <sup>2</sup>	RMSE	MAE	R <sup>2</sup>	RMSE	MAE
RNA301	2		0,830	0,770	0,569	0,925	0,687	0,575	0,925	0,451	0,359
RNA302	4		0,834	0,757	0,592	0,929	0,609	0,512	0,925	0,429	0,333
RNA303	6		0,822	0,813	0,585	0,938	0,723	0,588	0,922	0,521	0,417
RNA304	8		0,815	0,805	0,625	0,930	0,649	0,548	0,929	0,451	0,334
RNA305	10		0,811	0,812	0,561	0,940	0,754	0,634	0,926	0,561	0,441
RNA306	2	1	0,812	0,797	0,608	0,929	0,957	0,804	0,917	0,711	0,580
<b>RNA307</b>	<b>4</b>	<b>2</b>	<b>0,796</b>	<b>0,894</b>	<b>0,659</b>	<b>0,940</b>	<b>0,573</b>	<b>0,471</b>	<b>0,905</b>	<b>0,474</b>	<b>0,361</b>
RNA308	6	3	0,788	0,837	0,641	0,927	0,644	0,528	0,897	0,502	0,392
RNA309	8	4	0,809	0,918	0,731	0,939	0,654	0,548	0,893	0,599	0,482
RNA310	10	5	0,805	0,837	0,712	0,935	0,577	0,440	0,900	0,534	0,436

Fonte: Autora (2022).

As redes neurais apresentaram resultados relativamente parecidos entre as arquiteturas. A rede RNA302, com uma camada intermediária com 4 neurônios foi a rede com melhores resultados relativos de  $R^2$ , RMSE e MAE.

Comparando o  $R^2$  também com as regressões, todas as redes apresentaram resultados similares entre as técnicas. Porém, as regressões apresentaram maiores erros na predição e, considerando a necessidade de ajuste adicional das equações de regressão, devido aos resíduos não independentes e elevados, pode-se concluir que as RNA foram relativamente mais promissoras. O Gráfico 9 mostra os resultados entre as 10 arquiteturas de RNA treinadas.

Gráfico 9 - Análise 3 - Resultados estatísticos das RNA (fase de validação cruzada) e das equações de regressão propostas



Fonte: Autora (2022).

O gráfico reforça que os menores desempenhos estatísticos foram apresentados por RNA de 2 camadas intermediárias, especialmente a RNA306. Observou-se um decréscimo de  $R^2$  e um aumento da RMSE e do MAE à medida que se aumentou o número de camadas e neurônios por camada.

## 5.5 Análise 4

Na Análise 4, optou-se por remover o  $d_{10}$  e utilizaras variáveis de entrada: LL, %Finos e  $C_u$ , uma vez que os resultados da Análise 3 não foram satisfatórios quando comparados aos resultados das análises anteriores e parte disso deu-se devido ao

fato de  $d_{10}$  e  $C_u$  serem correlacionados. A partir do mesmo conjunto de 111 dados, foram desenvolvidas as equações de regressão no Minitab e treinadas as arquiteturas de RNA no RapidMiner. Os resultados são apresentados a seguir.

### 5.5.1 Resultados das análises estatísticas

Assim, utilizando a base de dados dessa pesquisa, as equações de regressão foram ajustadas no Minitab e os resultados são apresentados na Tabela 23.

Tabela 23 - Análise 4 – Resumo das equações de regressão propostos

Análise	Variáveis de entrada	Equação de regressão	R <sup>2</sup>	RMSE	MAE	
4	Finos – LL - $C_u$	RL4	$\log(k_{sat}) = -6,624 - 0,010127 LL - 0,000369 Cu - 0,00972 Finos$	0,828	0,872	0,669
		RNL4	$\log(k_{sat}) = -5,772 - 0,02235 LL - 0,001353 Cu - 0,0223 Finos + 0,000006 LL^2 + 0,000000 Cu^2 + 0,000028 Finos^2 + 0,000002 LL * Cu + 0,000119 LL * Finos + 0,000013 Cu * Finos$	0,895	0,681	0,549

Onde:

LL – Limite de liquidez (%)

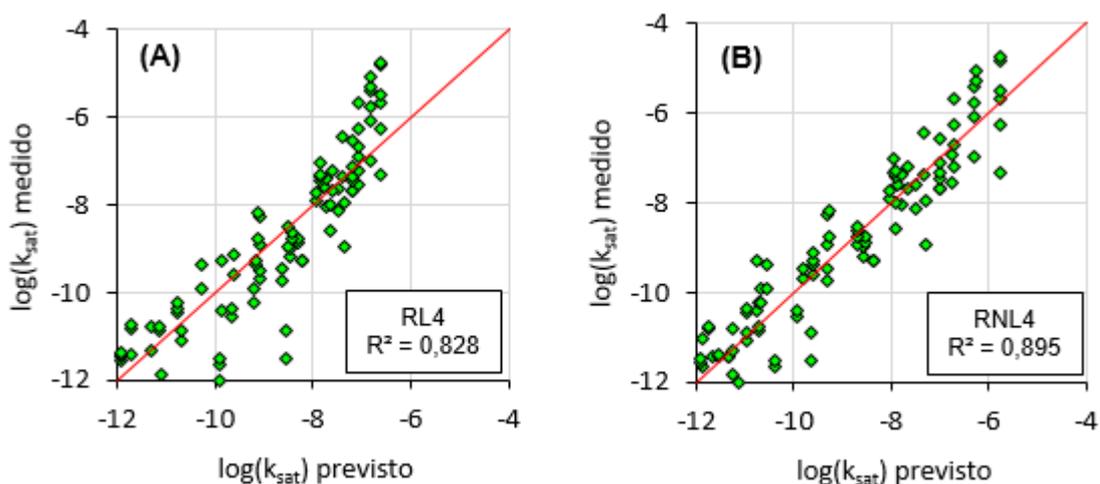
Finos – Percentual de finos (%)

$C_u$  – Coeficiente de uniformidade (adimensional)

Fonte: Autora (2022).

Nestas novas equações, verificou-se pouca alteração no desempenho das regressões, quando comparadas à Análise 3, considerando as medidas estatísticas R<sup>2</sup>, RMSE e MAE. Na Análise 4, houve um desempenho menor que na Análise 3, ainda que a diferença tenha sido pequena. O Gráfico 10 mostra a dispersão entre valores medidos e previstos pelas equações RL4 e RLN4.

Gráfico 10 - Análise 4 – Valores medidos e previstos de  $\log(k_{sat})$



Fonte: Autora (2022).

Os testes estatísticos foram realizados e os resultados para a equação RL4 são apresentados na Tabela 23. Assim como na Análise 1, todos os termos da RL4 estão significativamente relacionados à variável resposta. Verificou-se que os parâmetros utilizados foram significantes para a equação e que os resíduos são normalmente distribuídos, em dois testes de hipóteses. O teste de Shapiro-Francia indicou a não normalidade dos resíduos. Além disso, também não foi verificada a independência dos resíduos pela estatística de Durbin-Watson:  $d_w$  calculado foi de 1,51, portanto fora do intervalo de aceitação da independência dos resíduos ( $1,78 \leq d_w < 2,22$ ). Assim como na Análise 1, todos os termos da RL4 estão significativamente relacionados à variável resposta.

Tabela 24 - Análise 4 - Resumo dos resultados dos testes estatísticos para a equação RL4

Objetivo	Teste	valor-p	Conclusão
Teste dos termos da regressão	Teste t (intercepto)	$2,00 \times 10^{-16}$	Parâmetro significativo
	Teste t (LL)	$2,00 \times 10^{-16}$	Parâmetro significativo
	Teste t (Finos)	$9,56 \times 10^{-03}$	Parâmetro significativo
	Teste t (Cu)	$3,16 \times 10^{-04}$	Parâmetro significativo
Teste da regressão	Teste F	$2,20 \times 10^{-16}$	Parâmetro significativo
Normalidade dos resíduos	Teste AD	$1,44 \times 10^{-01}$	Distribuição Normal dos resíduos
	Teste SF	$3,89 \times 10^{-02}$	Distribuição dos resíduos não é normal
	Teste de LF	$5,67 \times 10^{-01}$	Distribuição Normal dos resíduos
Independência dos resíduos	Teste DW	$2,90 \times 10^{-03}$	Resíduos não são independentes
Multicolinearidade	VIF_LL	1,44	Não há multicolinearidade
	VIF_Finos	1,48	Não há multicolinearidade
	VIF_Cu	1,23	Não há multicolinearidade

Fonte: Autora (2022).

O cálculo do valor de VIF não indicou a existência de multicolinearidade entre as variáveis. Considerando os requisitos que uma regressão deve atender para ter sua significância estatística garantida, pode-se concluir que a equação RL4 carece de melhores ajustes e, principalmente, uma análise dos resíduos para melhoria do desempenho estatístico.

A equação RNL4, apesar de atender aos critérios de resíduos normalmente distribuídos e independentes, mostrou-se uma equação sobreajustada, com a inclusão de termos que não se mostraram significativamente relacionados com a predição de  $k_{sat}$ . A aplicação de técnicas estatísticas que auxiliem na definição das

variáveis que, de fato comporão, a equação pode ajudar a refinar a RNL4. Os resultados dos testes estatísticos estão sumarizados na Tabela 25.

Tabela 25 - Análise 4 - Resumo dos resultados dos testes estatísticos para a equação RNL4

Objetivo	Teste	valor-p	Conclusão
Teste dos termos da regressão	Teste t (intercepto)	$2,20 \times 10^{-16}$	Parâmetro significativa
	Teste t (LL)	$2,51 \times 10^{-11}$	Parâmetro significativa
	Teste t (Finos)	$8,07 \times 10^{-02}$	Parâmetro <b>não</b> significativa
	Teste t ( $C_u$ )	$3,64 \times 10^{-03}$	Parâmetro significativa
	Teste t ( $C_u^2$ )	$6,51 \times 10^{-01}$	Parâmetro <b>não</b> significativa
	Teste t ( $LL^2$ )	$2,01 \times 10^{-01}$	Parâmetro <b>não</b> significativa
	Teste t ( $Finos^2$ )	$8,32 \times 10^{-01}$	Parâmetro <b>não</b> significativa
	Teste t ( $LL * C_u$ )	$4,33 \times 10^{-01}$	Parâmetro <b>não</b> significativa
	Teste t ( $LL * Finos$ )	$4,53 \times 10^{-03}$	Parâmetro significativa
	Teste t ( $C_u * Finos$ )	$3,87 \times 10^{-01}$	Parâmetro <b>não</b> significativa
Teste da regressão	Teste F	$2,20 \times 10^{-16}$	Parâmetro significativa
Normalidade dos resíduos	Teste AD	$1,95 \times 10^{-01}$	Distribuição Normal dos resíduos
	Teste SF	$1,39 \times 10^{-01}$	Distribuição Normal dos resíduos
	Teste de LF	$2,19 \times 10^{-01}$	Distribuição Normal dos resíduos
Independência dos resíduos	Teste DW	$2,90 \times 10^{-03}$	Resíduos <b>não</b> independentes
Multicolinearidade	VIF_(LL)	53,97	Há multicolinearidade
	VIF_(Finos)	26,83	Há multicolinearidade
	VIF_( $C_u$ )	39,72	Há multicolinearidade
	VIF_( $C_u^2$ )	33,04	Há multicolinearidade
	VIF_( $LL^2$ )	25,39	Há multicolinearidade
	VIF_( $Finos^2$ )	35,21	Há multicolinearidade
	VIF_( $LL * C_u$ )	129,42	Há multicolinearidade
	VIF_( $LL * Finos$ )	73,83	Há multicolinearidade
VIF_( $C_u * Finos$ )	80,74	Há multicolinearidade	

Fonte: Autora (2022).

A quantidade de termos da equação não estatisticamente significativos evidencia que o modelo de polinômio investigado por ser refinado. Embora, nesta Análise, os resíduos apresentaram distribuição normal, não foi verificada a independência ( $d_w = 1,51$ , fora do intervalo  $1,90 \leq d_w < 2,10$ ). Os valores de VIF ficaram acima de 10 para a maioria dos termos, mas, considerando o formato de polinômio investigado, a existência de produtos entre variáveis e variáveis elevadas ao quadrado, era de se esperar a existência de multicolinearidade.

### 5.5.2 Resultados das redes neurais

Ao todo, as mesmas 10 arquiteturas propostas foram treinadas considerando as variáveis de entrada da Análise 4 (LL, %Finos e  $C_u$ ). Os resultados de  $R^2$ , RMSE e MAE foram listados e apresentados na Tabela 26.

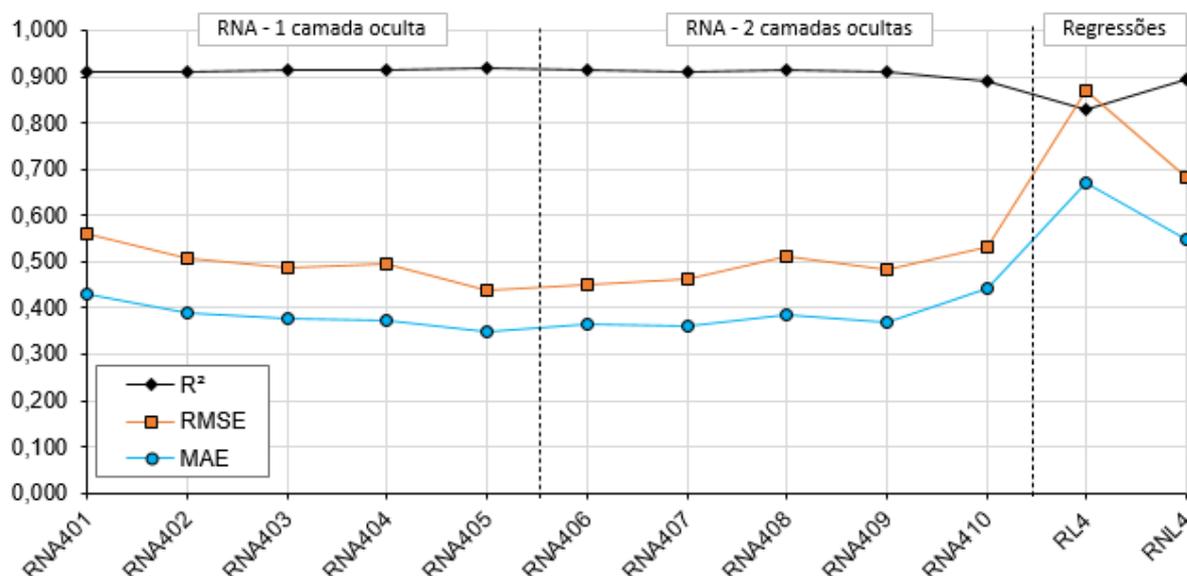
Tabela 26 - Análise 4 – Resumo das arquiteturas de RNA treinadas

RNA	Nº DE NEURÔNIOS		PARÂMETROS ESTATÍSTICOS								
			TREINO			TESTE			VALIDAÇÃO		
	CAMADA 1	CAMADA 2	$R^2$	RMSE	MAE	$R^2$	RMSE	MAE	$R^2$	RMSE	MAE
RNA401	2		0,836	0,792	0,605	0,927	0,751	0,636	0,909	0,562	0,428
RNA402	4		0,853	0,691	0,540	0,931	0,671	0,554	0,912	0,507	0,389
RNA403	6		0,818	0,998	0,860	0,932	0,657	0,550	0,913	0,488	0,376
RNA404	8		0,825	0,801	0,586	0,936	0,688	0,569	0,916	0,494	0,374
<b>RNA405</b>	<b>10</b>		<b>0,827</b>	<b>0,769</b>	<b>0,610</b>	<b>0,932</b>	<b>0,609</b>	<b>0,496</b>	<b>0,918</b>	<b>0,439</b>	<b>0,347</b>
RNA406	2	1	0,833	0,765	0,588	0,921	0,704	0,577	0,916	0,450	0,364
RNA407	4	2	0,837	0,755	0,567	0,935	0,569	0,467	0,910	0,464	0,359
RNA408	6	3	0,808	0,802	0,613	0,929	0,712	0,586	0,916	0,510	0,387
RNA409	8	4	0,822	0,774	0,587	0,938	0,616	0,502	0,909	0,481	0,370
RNA410	10	5	0,830	0,826	0,650	0,931	0,578	0,437	0,889	0,531	0,444

Fonte: Autora (2022).

Analisando os resultados das medidas estatísticas de desempenho, a rede RNA405, cuja arquitetura foi construída com uma camada intermediária (10 neurônios), foi a que apresentou os melhores índices. Comparadas com as redes treinadas nas Análises 1 a 3, as redes RNA401 a RNA410 apresentaram resultados superiores, com RMSE e MAE sempre abaixo de 0,600 e 0,500, respectivamente. Dentre todas as arquiteturas da Análise 4, observou-se também que os valores de RMSE e MAE foram mais próximos um do outro, sugerindo maior uniformidade e a ausência de resíduos elevados e discrepantes. O Gráfico 11 mostra os resultados de maneira gráfica e o comparativo com as equações de regressão linear e polinomial.

Gráfico 11 - Análise 4 - Resultados estatísticos das RNA (fase de validação cruzada) e das equações de regressão propostas



Fonte: Autora (2022).

Dentre as arquiteturas de RNA, especificamente nesta Análise 4, verificou-se que as arquiteturas com 1 ou 2 camadas intermediárias apresentaram menores diferenças de desempenho do que nas Análises anteriores. E, novamente, o uso de RNA mostrou-se mais indicado para a tarefa de predição do que as regressões estatísticas.

## 5.6 Análise 5

A última análise realizada com o banco de dados de 111 amostras de Wang e Huang (1984) testou a inclusão da propriedade índice de vazios no conjunto de variáveis de entrada. Sendo assim, estudou-se a predição de  $\log(k_{sat})$  a partir de LL, %Finos,  $C_u$  e  $e$ , através das técnicas de regressão e redes neurais.

### 5.6.1 Resultados das análises estatísticas

As equações linear e polinomial foram desenvolvidas na ferramenta computacional Minitab e os testes estatísticos foram aplicados no RStudio. A Tabela 27 apresenta o equacionamento das duas regressões propostas e os resultados de  $R^2$ , RMSE e MAE.

Tabela 27 - Análise 5 – Resumo das equações de regressão propostas

Análise	Variáveis de entrada	Equação de regressão	R <sup>2</sup>	RMSE	MAE	
5	Finos – LL - C <sub>u</sub> - e	RL5	$\log(k_{sat}) = -6,567 - 0,010061 LL - 0,000374 C_u - 0,156 e - 0,00875 \text{ Finos}$	0,828	0,871	0,669
		RNL5	$\log(k_{sat}) = -5,69 - 0,02207 LL - 0,001577 C_u + 2,44 e - 0,0556 \text{ Finos} + 0,000010 LL^2 + 0,000000 C_u^2 - 5,48 e^2 - 0,000317 \text{ Finos}^2 + 0,000001 LL * C_u - 0,00839 LL * e + 0,000203 LL * \text{Finos} + 0,00121 C_u * e + 0,000008 C_u * \text{Finos} + 0,0978 e * \text{Finos}$	0,906	0,644	0,513

Onde:

LL – Limite de liquidez (%)

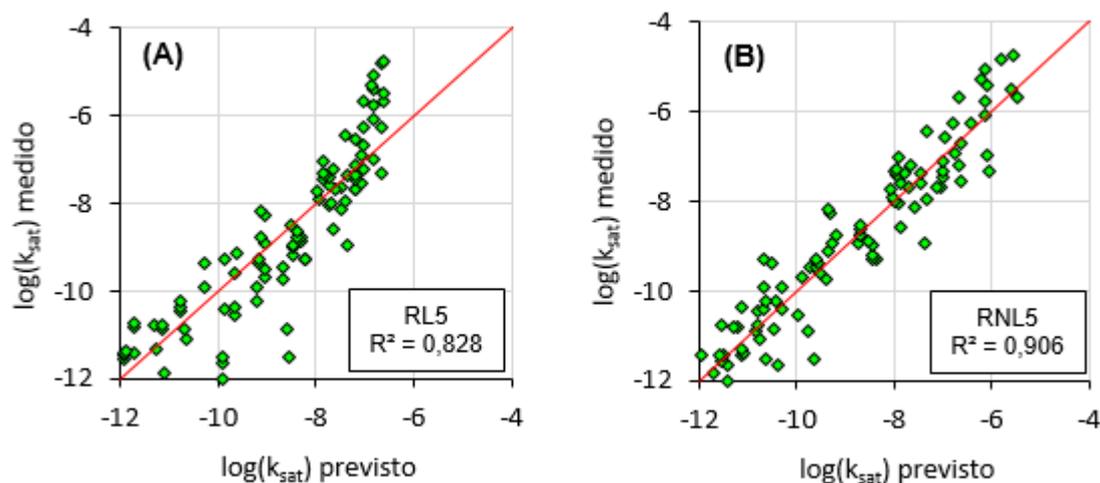
Finos – Percentual de finos (%)

C<sub>u</sub> – Coeficiente de uniformidade (adimensional)

e – Índice de vazios (adimensional)

Fonte: Autora (2022).

O Gráfico 12 mostra a dispersão de valores medidos e previstos de  $\log(k_{sat})$  para as duas regressões da Análise 5.

Gráfico 12 - Análise 5 – Valores medidos e previstos de  $\log(k_{sat})$ 

Fonte: Autora (2022).

Os resultados mostram que o acréscimo da variável não melhorou significativamente as estatísticas de desempenho das equações de regressão; o desempenho de RL5 em relação a RL4 (que não considera o índice de vazios na equação) foi idêntico, exceto pelo valor de RMSE, que diferiu em 0,001 entre estas regressões. Como não é recomendado que os valores de R<sup>2</sup>, RMSE e MAE sejam avaliados de forma isolada, os testes de aderência foram conduzidos para verificar os demais requisitos que devem ser atendidos pelas equações de regressão, assim como foi realizado em todas as análises até aqui. A Tabela 28 resume os resultados dos testes de cada requisito para a equação RL5.

Tabela 28 - Análise 5 - Resumo dos resultados dos testes estatísticos para a equação RL5

Objetivo	Teste	valor-p	Conclusão
Teste dos termos da regressão	Teste t (intercepto)	$2,00 \times 10^{-16}$	Parâmetro significativo
	Teste t (LL)	$2,00 \times 10^{-16}$	Parâmetro significativo
	Teste t (Finos)	$1,97 \times 10^{-01}$	Parâmetro <b>não</b> significativo
	Teste t (Cu)	$4,28 \times 10^{-04}$	Parâmetro significativo
	Teste t (e)	$8,68 \times 10^{-01}$	Parâmetro <b>não</b> significativo
Teste da regressão	Teste F	$2,20 \times 10^{-16}$	Parâmetro significativo
Normalidade dos resíduos	Teste AD	$1,32 \times 10^{-01}$	Distribuição Normal dos resíduos
	Teste SF	$3,63 \times 10^{-02}$	Distribuição dos resíduos <b>não</b> é normal
	Teste de LF	$4,77 \times 10^{-01}$	Distribuição Normal dos resíduos
Independência dos resíduos	Teste DW	$2,69 \times 10^{-03}$	Resíduos <b>não</b> são independentes
Multicolinearidade	VIF_LL	2,01	Não há multicolinearidade
	VIF_Finos	4,94	Não há multicolinearidade
	VIF_Cu	1,30	Não há multicolinearidade
	VIF_e	6,56	Não há multicolinearidade

Fonte: Autora (2022).

De fato, analisando os resultados dos testes estatísticos, observa-se que dois termos não contribuíram significativamente para a equação: %Finos e e. Apesar da normalidade dos resíduos em dois testes, Shapiro-Francia novamente indicou a não normalidade dos resíduos. Também não se verificou a independência entre eles, pela estatística de Durbin-Watson ( $d_w = 1,50$ , fora do intervalo  $1,78 \leq d_w < 2,22$ ).

A equação polinomial RNL5 também foi testada com o objetivo de verificar o atendimento aos requisitos necessários para as regressões. Os resultados foram sumarizados na Tabela 29.

Tabela 29 - Análise 5 - Resumo dos resultados dos testes estatísticos para a equação RNL5

Objetivo	Teste	valor-p	Conclusão
Teste dos termos da regressão	Teste t (intercepto)	$1,05 \times 10^{-05}$	Parâmetro significativa
	Teste t (LL)	$2,91 \times 10^{-07}$	Parâmetro significativa
	Teste t (Finos)	$1,02 \times 10^{-01}$	Parâmetro não significativa
	Teste t ( $C_u$ )	$2,60 \times 10^{-02}$	Parâmetro significativa
	Teste t (e)	$6,50 \times 10^{-01}$	Parâmetro não significativa
	Teste t ( $LL^2$ )	$4,26 \times 10^{-02}$	Parâmetro significativa
	Teste t ( $Finos^2$ )	$3,09 \times 10^{-01}$	Parâmetro não significativa
	Teste t ( $C_u^2$ )	$6,65 \times 10^{-01}$	Parâmetro não significativa
	Teste t ( $e^2$ )	$3,82 \times 10^{-01}$	Parâmetro não significativa
	Teste t ( $LL \cdot C_u$ )	$7,14 \times 10^{-01}$	Parâmetro não significativa
	Teste t ( $LL \cdot Finos$ )	$6,11 \times 10^{-03}$	Parâmetro significativa
	Teste t ( $LL \cdot e$ )	$2,97 \times 10^{-01}$	Parâmetro não significativa
	Teste t ( $Finos \cdot C_u$ )	$6,51 \times 10^{-01}$	Parâmetro não significativa
	Teste t ( $Finos \cdot e$ )	$2,14 \times 10^{-01}$	Parâmetro não significativa
	Teste t ( $C_u \cdot e$ )	$4,04 \times 10^{-01}$	Parâmetro não significativa
Teste da regressão	Teste F	$2,20 \times 10^{-16}$	Parâmetro significativa
Normalidade dos resíduos	Teste AD	$6,13 \times 10^{-01}$	Distribuição Normal dos resíduos
	Teste SF	$4,93 \times 10^{-01}$	Distribuição Normal dos resíduos
	Teste de LF	$8,21 \times 10^{-01}$	Distribuição Normal dos resíduos
Independência dos resíduos	Teste DW	$3,61 \times 10^{-04}$	Resíduos não são independentes
Multicolinearidade	VIF_(LL)	103,48	Há multicolinearidade
	VIF_(Finos)	203,07	Há multicolinearidade
	VIF_( $C_u$ )	99,62	Há multicolinearidade
	VIF_(e)	380,58	Há multicolinearidade
	VIF_( $LL^2$ )	31,04	Há multicolinearidade
	VIF_( $Finos^2$ )	208,42	Há multicolinearidade
	VIF_( $C_u^2$ )	45,25	Há multicolinearidade
	VIF_( $e^2$ )	1398,89	Há multicolinearidade
	VIF_( $LL \cdot C_u$ )	174,95	Há multicolinearidade
	VIF_( $LL \cdot Finos$ )	245,09	Há multicolinearidade
	VIF_( $LL \cdot e$ )	488,86	Há multicolinearidade
	VIF_( $Finos \cdot C_u$ )	104,65	Há multicolinearidade
	VIF_( $Finos \cdot e$ )	1725,13	Há multicolinearidade
	VIF_( $C_u \cdot e$ )	179,15	Há multicolinearidade

Fonte: Autora (2022).

Os testes t e F confirmaram o sobreajuste da equação, pela quantidade de termos sem significância estatística com a variável dependente. Quanto à normalidade dos

resíduos, esse critério foi atendido em todos os testes realizados. Mas resíduos não foram identificados como independentes.

O produto entre variáveis e variáveis elevadas à potência 2 contribuíram para os resultados elevados de VIF, como já discutido anteriormente. Conclui-se que o acréscimo do índice de vazios aumentou a complexidade da equação sem ganhos tão significativos de qualidade do ajuste.

### 5.6.2 Resultados das redes neurais

O desempenho das arquiteturas de RNA também foi verificado para o mesmo conjunto de variáveis de entrada. Os resultados estão apresentados na Tabela 30.

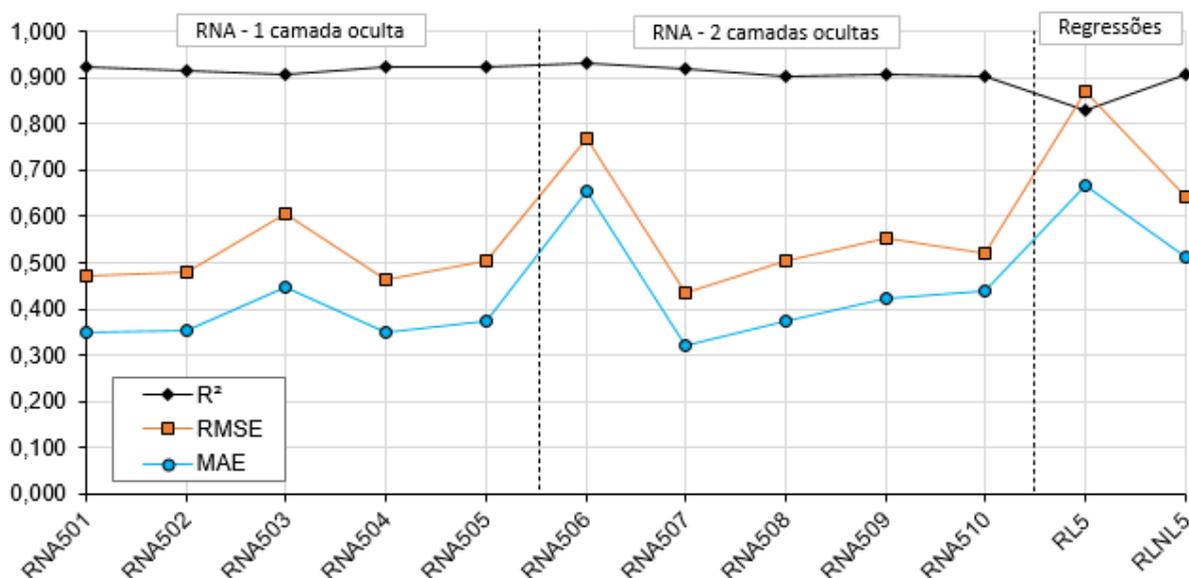
Tabela 30 - Análise 5 – Resumo das arquiteturas de RNA treinadas

RNA	Nº DE NEURÔNIOS		PARÂMETROS ESTATÍSTICOS								
			TREINO			TESTE			VALIDAÇÃO		
	CAMADA 1	CAMADA 2	R <sup>2</sup>	RMSE	MAE	R <sup>2</sup>	RMSE	MAE	R <sup>2</sup>	RMSE	MAE
RNA501	2		0,829	0,756	0,541	0,927	0,674	0,543	0,923	0,471	0,351
RNA502	4		0,823	0,790	0,627	0,925	0,606	0,502	0,916	0,478	0,353
RNA503	6		0,829	0,760	0,535	0,932	0,777	0,672	0,909	0,605	0,447
<b>RNA504</b>	<b>8</b>		<b>0,832</b>	<b>0,754</b>	<b>0,559</b>	<b>0,930</b>	<b>0,622</b>	<b>0,526</b>	<b>0,924</b>	<b>0,464</b>	<b>0,351</b>
RNA505	10		0,812	0,848	0,651	0,938	0,633	0,527	0,922	0,506	0,376
RNA506	2	1	0,842	0,734	0,545	0,935	1,016	0,869	0,930	0,769	0,656
RNA507	4	2	0,826	0,836	0,636	0,941	0,569	0,461	0,919	0,436	0,323
RNA508	6	3	0,819	0,793	0,615	0,935	0,624	0,509	0,903	0,505	0,376
RNA509	8	4	0,817	0,919	0,743	0,940	0,630	0,524	0,906	0,555	0,424
RNA510	10	5	0,832	0,765	0,610	0,937	0,556	0,431	0,904	0,522	0,440

Fonte: Autora (2022).

Observa-se que, comparado ao desempenho das RNA da Análise 4, houve uma pequena redução da qualidade do desempenho das redes com a inclusão da variável índice de vazios. A arquitetura que apresentou os melhores índices, considerando as fases de treinamento, teste e validação, foi a rede RNA504, composta por 1 camada com 8 neurônios. O resumo gráfico dos resultados é apresentado no Gráfico 13.

Gráfico 13 - Análise 5 - Resultados estatísticos das RNA (fase de validação cruzada) e das equações de regressão propostas



Fonte: Autora (2022).

Novamente, os menores RMSE e MAE e maior  $R^2$  ocorreram nas RNA, superando o desempenho das regressões. Em parte, isso se deve ao formato, principalmente de regressão polinomial, escolhido.

## 5.7 Conclusões parciais das equações de regressão e redes neurais

Neste item, são resumidas algumas conclusões sobre os resultados das regressões e redes neurais desenvolvidas para o banco de dados experimentais dessa pesquisa.

- 1) Em geral, as medidas estatísticas de desempenho de todas as técnicas, em cada análise, não apresentaram diferenças muito significativas entre si;
- 2) Entre as análises, a escolha das variáveis de entrada na Análise 4 se mostrou mais assertiva, pois apresentou menores erros;
- 3) As equações de regressão, em sua maioria, apresentaram problemas com de resíduos muito grandes e modelos sobreajustados, com termos que não foram significativamente correlacionados com a variável resposta, o que comprometeu o atendimento ao requisito de independência dos resíduos e significância estatística;
- 4) A verificação dos dados que apresentaram resíduos grandes, a fim de decidir por sua exclusão ou não da base de dados, deveria ser realizada para cada

equação de regressão em que isso tenha sido constatado. Todavia, para manter a paridade na comparação com o desempenho das redes neurais, optou-se por utilizar a mesma base de dados (resultante da fase de pré-tratamento) para todas as técnicas de predição investigadas;

- 5) As redes neurais, como já afirmado, apresentam a vantagem da autonomia dos algoritmos de aprendizagem, uma vez que não é necessário estabelecer previamente a relação matemática entre variáveis independentes e dependente;
- 6) As arquiteturas que apresentaram os melhores desempenhos em cada análise foram: RNA104, RNA205, RNA302, RNA405 e RNA504, nas análises 1 a 5, respectivamente. Estas arquiteturas foram comparadas, juntamente com as regressões, com os resultados de correlações disponíveis na literatura.

## **5.8 Resultados de correlações empíricas e semiempíricas**

O item 2.3 apresentou um conjunto de equações empíricas e semiempíricas reunidas da literatura para previsão de  $k_{sat}$  a partir de propriedades índices. Os critérios de validade de cada correlação foram avaliados a fim de selecionar as equações para esta análise comparativa.

O atendimento aos critérios de definição da correlação mostrou-se uma dificuldade para utilização nos dados experimentais, pois, a depender da correlação, apenas uma parte dos dados foi possível de ser aplicada e tal correlação não considerada nessa pesquisa.

A equação de Hazen (1911) é aplicável somente em 06 amostras de Wang e Huang (1984), considerando os critérios de validade da correlação, isto é, apenas 06 amostras apresentaram  $d_{10}$  entre 0,1 mm e 3,0 mm e  $C_u \leq 5$ . Kozeny-Carman (1927, 1937), na sua forma original, é aplicável apenas para areias e é necessário o conhecimento da superfície específica (ausente em todos os bancos de dados investigados nessa pesquisa). Já a versão modificada por Chapuis e Aubertin (2003), aplicável para solos plásticos e não plásticos, necessita ainda da superfície específica, motivo pelo qual não foi possível aplicá-la aos dados experimentais.

A equação apresentada no NAVFAC DM7 (1974) é aplicável somente às areias e misturas de areias e pedregulhos, o que por si só já restringiria a uma quantidade pequena de amostras do banco de dados. Todavia, um dos critérios de aplicação da correlação exige o conhecimento da razão  $d_{10}/d_5$ , não disponível nos bancos de dados. Chapuis (2004) foi descartada, pois é aplicável apenas às areias e pedregulhos. Sivappullaiah *et al.* (2000) é válida apenas para 37 amostras, que apresentam  $LL > 50\%$ . Shahabi *et al.* (1984) é aplicável apenas a 06 amostras, pois, além de ser restrita às areias, o coeficiente de uniformidade deve ser no máximo 8, o que descarta grande parte dos dados.

Mbonimpa *et al.* (2002) desenvolveram uma correlação que é aplicável a solos plásticos e não plásticos (C8 e C9). Por não exigir dados como a superfície específica e nem outras restrições, foi possível aplicá-la a todos os dados experimentais. Pucket *et al.* (1985) e Jabro (1992) apresentaram correlações sem restrições explícitas e válidas para solos plásticos e não plásticos.

A equação de Pucket *et al.* (1985) foi investigada para o conjunto de dados experimentais, apesar da faixa de  $k_{sat}$  utilizada para o desenvolvimento da equação ter sido diferente da faixa de valores presentes no banco de dados dessa dissertação; a saber, Pucket *et al.* (1985) desenvolveram a correlação utilizando  $k_{sat}$  entre  $8,33 \times 10^{-10}$  e  $3,94 \times 10^{-05}$  m/s. Já Jabro (1992) não foi utilizada nessa pesquisa, pois, apesar de não apresentar critérios específicos de validade, a equação envolve o cálculo do logaritmo do percentual de silte e argila, ou seja, amostras sem silte e/ou argila não podem ter o  $k_{sat}$  estimado por essa correlação.

Também Salarashayeri e Siosemade (2012) apresentaram uma correlação que não foi possível aplicar aos dados experimentais (apenas em 02 amostras), pois é restrita a areias com  $k_{sat}$  entre  $1,00 \times 10^{-02}$  e  $4,19 \times 10^{-05}$  m/s. Por último, Carrier (2003) apresentou uma modificação da equação de Kozeny-Carman (1927, 1937) a partir de uma relação entre as frações granulométricas e a superfície específica. A partir da proposta de Carrier (2003) é possível a aplicação da equação baseada nos princípios de Kozeny-Carman (1927, 1937) mesmo sem o conhecimento da superfície específica. Entretanto, sua proposição, por ter sido apenas e diretamente derivada de Kozeny-Carman (1927, 1937), é válida para as mesmas premissas (SANZENI;

COLLESELLI; GRAZIOLI, 2013). Na Tabela 31, são apresentadas as equações selecionadas.

Tabela 31 - Equações de predição de  $k_{sat}$  selecionadas para o estudo comparativo

Nº	Equação	Equação	Comentários
C8	Mbonimpa <i>et al.</i> (2002)	$k_{sat} = C_G \times \frac{\gamma_w}{\mu_w} \times C_u^{1/3} \times d_{10}^2 \times \frac{e^{3+x}}{1+e}$	Válida para solos não plásticos.
C9		$k_{sat} = C_P \times \frac{\gamma_w}{\mu_w} \times \frac{e^{3+x}}{1+e} \times \frac{1}{\rho_s^2 \times LL^{2x}}$	Válida para solos plásticos.
C10	Pucket <i>et al.</i> (1985)	$k_{sat} = 4,36 \times 10^{-5} \times e^{-0,1975 \times Arg}$	Não são apresentadas restrições explícitas. Condições de desenvolvimento da PTF*: $8.33e-10 \leq k_{sat} \leq 3.94e-05$ m/s. Solos plásticos e não plásticos.

Fonte: Autora (2022).

O objetivo deste item é observar a precisão da previsão de  $k_{sat}$  por correlações publicadas na literatura e comparar o desempenho dessas previsões com os resultados das equações propostas de regressão e de redes neurais. A Tabela 32 resume os resultados dessa etapa, a saber: a correlação considerada e as medidas estatísticas de desempenho do ajuste ( $R^2$ , RMSE e MAE).

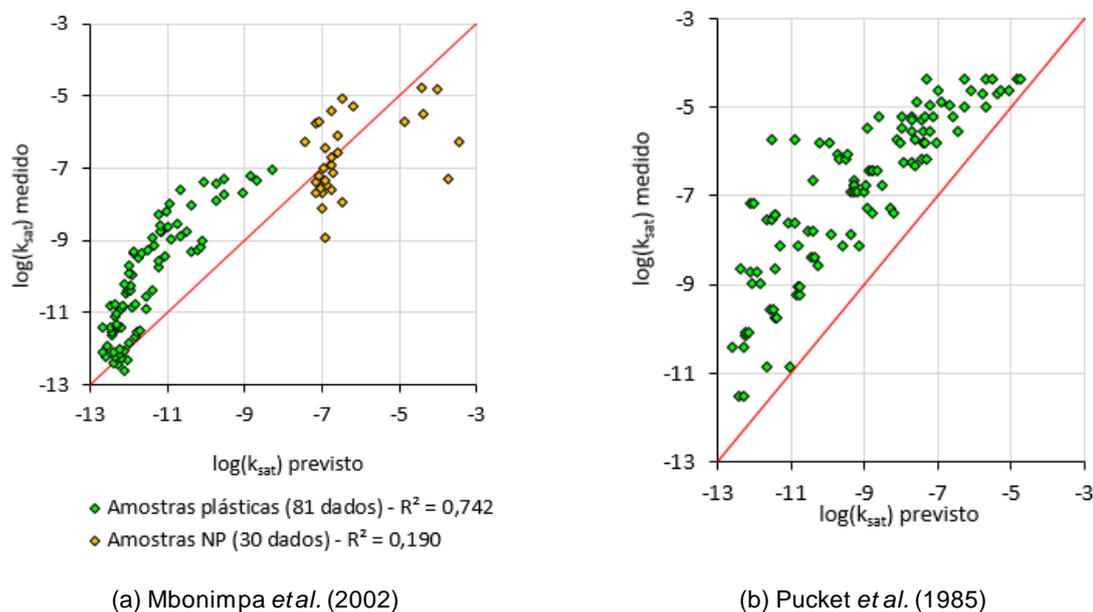
Tabela 32 – Resultados de correlações selecionadas da literatura aplicadas ao conjunto de 111 dados experimentais

Nº	Equação	$R^2$	RMSE	MAE
C8, C9	Mbonimpa <i>et al.</i> (2002)	0,764	1,531	1,275
C10	Pucket <i>et al.</i> (1985)	0,732	2,486	2,236

Fonte: Autora (2022).

O Gráfico 14 mostra os diagramas de dispersão de correlações publicadas na literatura. Fica evidente a dificuldade de um melhor ajuste entre os valores medidos e valores previstos pelas correlações empíricas, quando comparada às outras técnicas utilizadas nessa pesquisa, mesmo obedecendo aos critérios de validade de cada equação. Tal resultado demonstra a necessidade, por vezes, de estabelecer novas equações de previsão de  $k_{sat}$ , seja por uso de análise estatísticas ou, como já mostrado, pela técnica de redes neurais artificiais.

Gráfico 14 - Valores medidos e previstos de  $\log(k_{sat})$  pelas correlações selecionadas da literatura



Fonte: Autora (2022).

Considerando o resultado geral da previsão de  $k_{sat}$  por Mbonimpa *et al.* (2002), verifica-se um  $R^2$  similar às técnicas de regressão e RNA aplicadas nessa pesquisa. Todavia, a previsão pela correlação empírica apresentou maiores desvios em relação aos dados amostrais (RMSE e MAE elevados) que os resultados das regressões e redes neurais. Quando avaliado o desempenho para previsão de  $k_{sat}$ , considerando a plasticidade da amostra, observa-se que Mbonimpa *et al.* (2002) se ajustou melhor às amostras plásticas de Wang e Huang (1984) ( $R^2 = 0,742$ ) e não apresentou um ajuste satisfatório para amostras não plásticas ( $R^2 = 0,190$ ), com maior erro entre o valor medido e valor previsto pela correlação.

Salienta-se que a equação de Mbonimpa *et al.* (2002) foi desenvolvida para um conjunto de dados diferentes dos dados experimentais desta pesquisa e alguns desvios nos resultados previstos são esperados. Contudo, trata-se de uma correlação que considera as propriedades do fluido percolante e faz a distinção do solo pela plasticidade. A equação C8 (válida para solos não plásticos) considera na sua formulação propriedades da distribuição granulométrica, como o  $C_u$  e o  $d_{10}$ , além de incluir o índice de vazios.

Já na equação C9 (válida para solos plásticos), propriedades da fração fina são consideradas e, principalmente, o LL, que permite inferir acerca da natureza da fração

fina e apresenta relação, em parte, com a superfície específica. Ou seja, apesar dos valores elevados de RMSE e MAE, para dos dados experimentais desta pesquisa, o desenvolvimento e as propriedades consideradas na formulação estão de acordo com os apontamentos realizados a partir da revisão bibliográfica desta pesquisa.

Pucket *et al.* (1985) apresentou um  $R^2 = 0,732$ . Entretanto, observa-se nos resultados da correlação, que o menor valor de  $\log(k_{sat})$  previsto pela correlação foi de -11,543, que equivale ao  $k_{sat} = 2,86 \times 10^{-12}$  m/s. Dados experimentais com  $k_{sat}$  menor que esse valor apresentaram, conseqüentemente, um desvio grande entre o valor medido e o previsto pela correlação.

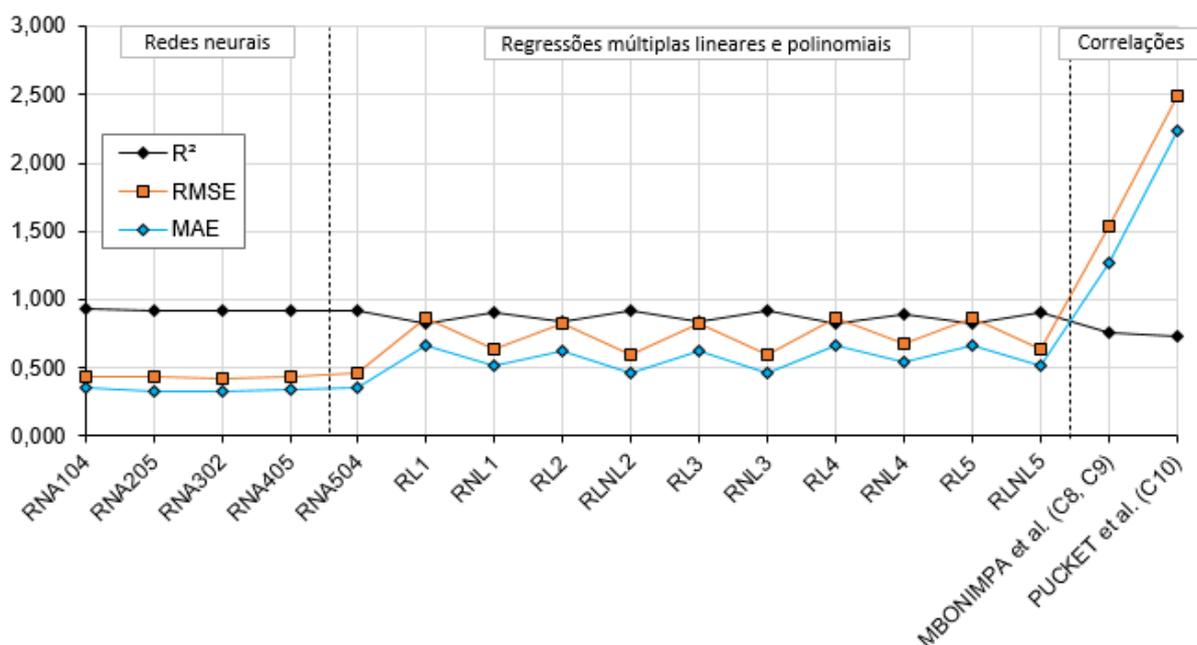
Pucket *et al.* (1985) desenvolveram a correlação considerando uma relação exponencial entre o  $k_{sat}$  e o percentual de argila. Sendo assim, um solo com %argila de 100%, o  $k_{sat}$  mínimo capaz de ser previsto por Pucket *et al.* (1985) será, portanto, de  $1,23 \times 10^{-13}$  m/s, apesar de não ser uma restrição explícita apresentada pelos pesquisadores.

## **5.9 Análise comparativa dos resultados das regressões, RNA e correlações empíricas**

Neste item, são comparados graficamente os desempenhos das equações propostas de regressões e das arquiteturas de RNA que apresentaram o melhor resultado em cada Análise com os resultados de correlações empíricas previamente selecionadas da literatura.

O Gráfico 15 mostra o comparativo entre  $R^2$ , RMSE e MAE apresentados pelas arquiteturas de redes que apresentaram o melhor desempenho relativo em cada análise (RNA104, RNA205, RNA302, RNA405 e RNA504) com as equações de regressão linear e polinomial resultantes das Análises 1 a 5.

Gráfico 15 – Comparação de  $R^2$ , RMSE e MAE entre as equações de melhor desempenho das regressões e redes neurais



Fonte: Autora (2022).

Observa-se do Gráfico 15 que, de forma geral, o desempenho das redes neurais superou as regressões e as correlações previamente selecionadas da literatura, considerando apenas os valores de  $R^2$ , RMSE e MAE. De todos os resultados das análises realizadas, a RNA302, que considerou como variáveis independentes o %Finos, LL,  $C_u$  e  $d_{10}$ , apresentou, numericamente, os melhores resultados, quando comparada às demais técnicas; apresentou os menores erros (RMSE = 0,429 e MAE = 0,333) e o melhor ajuste entre todas as técnicas ( $R^2 = 0,925$ ).

Entretanto, entre estas redes com melhores resultados em cada análise, verificou-se diferenças sutis de desempenho, mesmo variando a arquitetura entre elas e as variáveis de entrada.

Comparando os resultados das equações de regressão linear e polinomial e as redes neurais, obtidos nesta pesquisa a partir dos dados experimentais, com os resultados das equações empíricas de Mbonimpa *et al.* (2002) ( $R^2 = 0,764$ , RMSE = 1,531, MAE = 1,275) e Pucket *et al.* (1985) ( $R^2 = 0,732$ , RMSE = 2,486, MAE = 2,236), observa-se que as equações propostas de RNA e regressões múltiplas melhor se adequaram ao banco de dados experimentais.

Em suma, mesmo com as limitações das regressões lineares e polinomiais desta pesquisa, com relação à presença dos resíduos, elas se mostraram relativamente

mais ajustadas aos dados que as correlações empíricas. As redes neurais, entretanto, além de melhores parâmetros estatísticos, apresentaram ainda a vantagem de não ser necessário o conhecimento prévio da relação matemática entre as variáveis do problema, sendo uma técnica relativamente mais autônoma do que as regressões.

Em problemas de predição com a complexidade do  $k_{sat}$ , as redes neurais são potencialmente interessantes pela capacidade de estabelecer e definir o grau de relações matemáticas complexas.

## 6 CONSIDERAÇÕES FINAIS E SUGESTÕES

A permeabilidade do solo depende da distribuição e geometria dos poros do solo e, no caso do solo saturado, é caracterizada pelo  $k_{sat}$ , um parâmetro influenciado por diversas propriedades do solo sensíveis à variação do teor da fração de partículas finas ( $\emptyset \leq 0,075$  mm). O  $k_{sat}$  pode ser determinado experimentalmente ou estimado por correlações empíricas, análises estatísticas ou técnicas de inteligência artificial.

O objetivo geral desta pesquisa foi analisar e comparar o desempenho de RNAs e de equações de regressão, desenvolvidas a partir de dados experimentais disponibilizados na literatura, com os resultados obtidos de correlações empíricas selecionadas previamente na previsão de  $k_{sat}$  de solos compostos por grãos grossos e finos. Para o desenvolvimento das equações, foram consideradas as seguintes variáveis: %finos, LL,  $d_{10}$ ,  $C_u$  e  $e$ . Diversas análises foram conduzidas, incluindo o treinamento de redes neurais e ajustes de equações de regressão linear múltipla e polinomial para cinco diferentes combinações das variáveis de entrada selecionadas (%finos, LL,  $d_{10}$ ,  $C_u$  e  $e$ ). Dos resultados desta pesquisa, conclui-se:

- 1) O  $k_{sat}$  é uma propriedade de natureza complexa e altamente variável, pois é afetada por diversas propriedades do solo, tais como a densidade do solo, a porosidade, as características do fluido percolante e a natureza dos minerais que compõem o solo, e algumas delas sob influência da fração e do tipo de fino presente na amostra, como a geometria e distribuição dos poros interconectados;
- 2) O teor de finos, isto é, o percentual de partículas presentes com tamanho menor de  $75 \mu\text{m}$ , exerce acentuada influência no  $k_{sat}$  até um percentual, próximo a 20% de finos; a partir desse teor, o  $k_{sat}$  dos solos compostos por grãos grossos e finos se aproxima do valor do fino puro. Todavia, pequenos teores de finos de uma argila muito plástica, na ordem de 5%, já são responsáveis por uma significativa redução de  $k_{sat}$  quando adicionados a um solo granular puro;
- 3) O aumento do teor de finos provoca variações no índice de vazios, no diâmetro efetivo e, conseqüentemente, no coeficiente de uniformidade. A influência da fração fina ainda é dependente da plasticidade dessa partícula;

- 4) Para caracterização de  $k_{sat}$ , é importante incluir propriedades índices que estejam relacionadas à distribuição e proporção de poros, além da característica mineralógica de partículas finas.
- 5) Para solos não plásticos, características da distribuição granulométrica e da geometria dos poros devem ser consideradas nas estimativas de  $k_{sat}$ , por exemplo, o índice de vazios e o coeficiente de uniformidade. Já para solos plásticos, observou-se que as características da fração fina são mais determinantes da estimativa de  $k_{sat}$  do que a granulometria do solo; neste caso, as propriedades que caracterizem a natureza do fino devem ser consideradas, como o LL. Sendo assim, as variáveis %finos, LL,  $d_{10}$ ,  $C_u$  e  $e$ , nesse sentido, se mostraram promissoras para descrição de problemas de determinação da permeabilidade saturada do solo;
- 6) A análise de bancos de dados evidenciou que, em solos com predominância de finos (ou com o comportamento hidráulico controlado pelos finos), o  $k_{sat}$  não é caracterizado de forma adequada somente com a informação da fração granulométrica desses solos;
- 7) O levantamento de bancos de dados de  $k_{sat}$  da literatura evidenciou uma carência de informações importantes para o estudo da permeabilidade do solo saturado. Muitos bancos, apesar de informar o  $k_{sat}$ , não apresentavam informações complementares para o desenvolvimento de correlações que caracterizassem adequadamente o problema;
- 8) É importante que a estruturação de bancos de dados de  $k_{sat}$  incluam informações como distribuição granulométrica,  $d_{10}$ ,  $C_u$ ,  $e$ , limites de consistência e superfície específica (ou outra propriedade que permita inferir sobre essa variável);
- 9) A ausência de informações complementares sobre as amostras, em bancos de dados, dificulta a fase de tratamento dos dados, pois a identificação e decisão sobre a exclusão de valores estatisticamente discrepantes é dificultada;
- 10) Mesmo com as limitações das regressões lineares e polinomiais desta pesquisa, com relação à presença dos resíduos, as regressões se mostraram relativamente mais ajustadas aos dados que as correlações previamente selecionadas da literatura, em termos dos valores de  $R^2$ , RMSE e MAE;
- 11) Especificamente sobre a análise estatística, é muito importante a seleção correta dos testes de hipóteses, baseados no número de amostras e variáveis.

- Tais testes para verificação da significância estatística da equação são imprescindíveis e devem complementar a avaliação dos parâmetros  $R^2$ , RMSE e MAE, os quais, isoladamente, dizem pouco sobre o desempenho da equação;
- 11) Em problemas de predição com a complexidade do  $k_{sat}$ , as correlações empíricas apresentaram dificuldades de aplicação aos dados experimentais, principalmente, pelos critérios restritivos, normalmente, de cada equação empírico;
  - 12) De correlações empíricas, a que estimou valores de  $k_{sat}$  mais próximos aos valores medidos foi a equação de Mbonimpa *et al.* (2002) ( $R^2 = 0,764$ , RMSE = 1,531, MAE = 1,275) e, mesmo assim, os valores de RMSE e MAE foram maiores que nos equações de regressão e redes neurais. Todavia, se tratando de uma equação desenvolvida a partir de outros dados, considera-se que se ajustou satisfatoriamente aos dados experimentais;
  - 13) A equação empírica de Mbonimpa *et al.* (2002) ratifica a importância de se considerar propriedades que caracterizem a fração grossa e a fração fina do solo, como os dados granulométricos e os limites de consistência. Além disso, a correlação diferencia o comportamento hidráulico de solos plásticos e não plásticos, estando de acordo com os apontamentos resultantes desta pesquisa;
  - 14) Para um problema que envolve múltiplas variáveis e de equacionamento complexo, como a previsão do  $\log(k_{sat})$ , conclui-se que as redes neurais se mostraram mais viáveis a este tipo de problema pela capacidade de estabelecer relações matemáticas complexas entre as variáveis e pela possibilidade de ser atualizada com a apresentação de novos dados à rede, característica deste tipo de técnica. Em contrapartida, é complexa a conversão da arquitetura de uma rede em uma função algébrica, e que poderia auxiliar na interpretação de um problema como a previsão do  $\log(k_{sat})$  e da relação existente entre as variáveis de entrada e saída;
  - 15) Das análises realizadas, considerando somente os resultados de  $R^2$ , RMSE e MAE, observou-se uma variação pouco significativa com a inclusão de variáveis de entrada adicionais, embora o uso de arquiteturas de RNA com mais de duas camadas intermediárias piorou consideravelmente o desempenho estatístico das RNA;
  - 16) Ainda sobre as análises realizadas, considerando os valores de  $R^2$ , RMSE e MAE, os melhores resultados foram apresentados na Análise 3, que considerou

%finos, LL,  $C_u$  e  $d_{10}$  como variáveis de entrada para previsão do  $\log(k_{sat})$ , ainda que  $d_{10}$  e  $C_u$  sejam correlacionadas. Das técnicas empregadas, a previsão melhor ajustada foi realizada por uma rede neural com uma camada oculta com 4 neurônios. A referida rede foi a RNA302 ( $R^2 = 0,925$ , RMSE = 0,429 e MAE = 0,333).

## 6.1 Sugestões para trabalhos futuros

Para trabalhos futuros que complementem essa pesquisa, sugere-se:

- 1) A verificação de equações utilizando banco com maior número de dados experimentais, considerando inclusive a segmentação do banco baseada no teor de finos;
- 2) A investigação de técnicas de regressão por etapas, aplicada aos polinômios propostos, para seleção dos termos com melhor significância estatística com  $\log(k_{sat})$ ;
- 3) A investigação dos efeitos do aumento do número de camadas intermediárias no desempenho das redes neurais artificiais;
- 4) O uso de outras ferramentas computacionais para o treinamento das RNA que permitam maior liberdade na escolha das funções de ativação e algoritmo de aprendizagem;
- 5) A complementação do trabalho utilizando novas técnicas de predição, principalmente redes neurais otimizadas, máquinas de vetores de regressão e outras técnicas de inteligência artificial.

## REFERÊNCIAS BIBLIOGRÁFICAS

- AKBULUT, S. Artificial neural networks for predicting the hydraulic conductivity of coarse-grained soils. **Eurasian Soil Science**, v. 38, n. 4, p. 392-398, 2005.
- ALAKAYLEH, Z.; CLEMENT, T.; FANG, X. Understanding the Changes in Hydraulic Conductivity Values of Coarse and Fine-Grained Porous Media Mixtures. **Water**, v. 10, n. 3, p. 313, 13 mar. 2018.
- ALDAOOD, A. Impact of fine materials on the saturated and unsaturated behavior of silty sand soil. **Ain Shams Engineering Journal**, v. 11, n. 3, p. 717–725, set. 2020.
- ASTM INTERNATIONAL. ASTM D 653-22 - Standard terminology relating to soil, rock, and contained fluids. ASTM International, Estados Unidos, 2022. p. 1–50.
- BANDINI, P.; SATHISKUMAR, S. Effects of Silt Content and Void Ratio on the Saturated Hydraulic Conductivity and Compressibility of Sand-Silt Mixtures. **Journal of Geotechnical and Geoenvironmental Engineering**, v. 135, n. 12, p. 1976–1980, dez. 2009.
- BARBER; E. S.; SAWYER; C. L. Highway subdrainage. In: ANNUAL MEETING OF THE HIGHWAY RESEARCH BOARD, 31, 1952, Washington, D.C. *Proceedings*, Highway Research Board Proceedings, p. 648-666, 1952.
- BELKHATIR, M.; ARAB, A.; DELLA, N.; SCHANZ, T. Laboratory Study on the Hydraulic Conductivity and Pore Pressure of Sand-Silt Mixtures. **Marine Georesources & Geotechnology**, v. 32, n. 2, p. 106–122, 3 abr. 2014.
- BENSON, C. H.; ZHAI, H.; WANG, X. Estimating Hydraulic Conductivity of Compacted Clay Liners. **Journal of Geotechnical Engineering**, v. 120, n. 2, p. 366–387, fev. 1994.
- BICALHO, K. V.; MACHADO, S. L.; GERSCOVICH, D. M. S. Condutividade hidráulica e fluxo em solos não saturados. In: CARVALHO, J. C. de; *et al.* (Orgs.). **Solos não saturados no contexto geotécnico**. São Paulo: Associação Brasileira de Mecânica dos Solos e Engenharia Geotécnica, 2015.
- BILARDI, S.; IELO, D.; MORACI, N. Predicting the Saturated Hydraulic Conductivity of Clayey Soils and Clayey or Silty Sands. **Geosciences**, v. 10, n. 393, p. 1-16, out. 2020.
- BOADU, F. Hydraulic Conductivity of Soils from Grain-Size Distribution: New Models. **Journal Of Geotechnical And Geoenvironmental Engineering**, v. 126, n. 8, p. 739-746, 2000.
- BORGES, F. A. S. Extração de características combinadas com árvore de decisão para detecção e classificação dos distúrbios de qualidade da energia elétrica. Dissertação (Mestrado em Ciências) - Escola de Engenharia de São Carlos, Universidade de São Paulo. São Paulo, 2013. 118 f.
- CARRARO, J. A. H.; BANDINI, P.; SALGADO, R. Liquefaction Resistance of Clean and Nonplastic Silty Sands Based on Cone Penetration Resistance. **Journal of Geotechnical and Geoenvironmental Engineering**, v. 129, n. 11, p. 965–976, nov. 2003.

CARRIER, W. D. Goodbye, Hazen; Hello, Kozeny-Carman. **Journal of Geotechnical and Geoenvironmental Engineering**, n. 129, p. 1054-1056, 2003.

CHAPUIS, R. P. Predicting the saturated hydraulic conductivity of sand and gravel using effective diameter and void ratio. **Canadian Geotechnical Journal**, n. 41, p. 787-795, 2004.

\_\_\_\_\_. Predicting the saturated hydraulic conductivity of soils: a review. **Bulletin of Engineering Geology and the Environment**, v. 71, n. 3, p. 401–434, ago. 2012.

\_\_\_\_\_; AUBERTIN, M. Predicting the coefficient of permeability of soils using the Kozeny-Carman equation. *École Polytechnique Montréal, EPM-RT-2003-03*, 2003.

\_\_\_\_\_; LÉGARÉ, P. P. A Simple Method for Determining the Surface Area of Fine Aggregates and Fillers in Bituminous Mixtures. In: **ASTM STP**, v. 1147, p. 177-186, 1992.

CHEMMAM, M., ARAB, A.; BELKHATIR, M.; BOUFERRA, R. Behavior of Loose Silty Sand of Chlef River: Effect of Low Plastic Fine Contents and Other Parameters. **Marine Georesources & Geotechnology**, v. 34, n. 4, p. 384–394, 18 mai. 2016.

CREVELIN, L. G. Comparação entre os valores de limite de liquidez obtidos pelos métodos de Casagrande e cone para diferentes solos argilosos. Dissertação (Mestrado em Engenharia Civil) - Universidade Federal do Espírito Santo, Centro Tecnológico, 2018, 162 p.

DASH, H.K; SITHARAM, T.G; BAUDET, B. A. Influence of non-plastic fines on the response of a silty sand to cyclic loading. **Soils and Foundations**, v. 50, n. 5, p. 695-704, 2010.

DENSON, K. H.; SHINDALA, A.; FENN, C. D. Permeability of Sand with Dispersed Clay Particles. **Water Resources Research**, v. 4, n. 6, p. 1275–1276, dez. 1968.

DIMINSKY, A. S. Análise de problemas geotécnicos através de redes neurais. Tese (Doutorado em Engenharia Civil) - Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2000, 193 f.

DOLINAR, B. Predicting the hydraulic conductivity of saturated clays using plasticity-value correlations. **Applied Clay Science**, v. 45, n. 1–2, p. 90–94, jun. 2009.

EKICI, A.; HUVAJ, N.; AKGÜNER, C. Effects of Fines Content and Plasticity on Index Properties and Hydraulic Conductivity of Coarse-Fine Mixtures. **Journal of Materials in Civil Engineering**, v. 34, n. 5, p. 1-15, mai. 2022.

FAIR, G. M.; HATCH, L. P. Fundamental factors governing the stream-line flow of water through sand. **Journal - American Water Works Association**, v. 25, n. 11, p. 1551–1563, 1933.

GLEASON, M. H.; DANIEL, D. E.; EYKHOLT, G. R. Calcium and Sodium Bentonite for Hydraulic Containment Applications. **Journal of Geotechnical and Geoenvironmental Engineering**, v. 123, n. 5, p. 438–445, mai. 1997.

GOLDSCHMIDT, R.; PASSOS, E. **Data mining: um guia prático**. Rio de Janeiro: Elsevier, 2005.

GOOTMAN, K. S.; KELLNER, E.; HUBBART, J. A. A comparison and validation of saturated hydraulic conductivity models. **Water**, v. 12, n. 7, p. 1-19, 2020.

JABRO, J. D. Estimation of saturated hydraulic conductivity of soils from particle size distribution and bulk density data. **Transactions of the ASAE**. American Society of Agricultural Engineers, v.35, n. 2, 1992.

JOHNSTON, R.; JONES, K.; MANLEY, D. Confounding, and collinearity in regression analysis: a cautionary tale and an alternative procedure, illustrated by studies of British voting behaviour. **Quality & Quantity**, v. 52, n. 4, p. 1957-1976, 2018.

HAZEN, A. Discussion of dams on sand foundations by A. C. Koenig. **Transactions American Society Civil Engineering**, v. 73, n.3, p. 199-203, 1911.

HAYKIN, S. **Redes neurais: princípios e prática**. Porto Alegre: Bookman, 2001.

KASHANI, M. H.; *et al.* Multiple AI model integration strategy – Application to saturated hydraulic conductivity prediction from easily available soil properties. **Soil & Tillage Research**, v. 196, p. 1-10, 2020.

KHAN, T. A.; FAROOQ, K.; MUHAMMAD, M.; KHAN, M. M.; SHAH, S. A. R.; SHOAIB, M.; ASLAM, M. A.; RAZA, S. S. The Effect of Fines on Hydraulic Conductivity of Lawrencepur, Chenab and Ravi Sand. **Processes**, v. 7, n. 11, p. 1-13, 2 nov. 2019.

KOLTERMANN, C. E.; GORELICK, S. M. Fractional packing model for hydraulic conductivity derived from sediment mixtures. **Water Resources Research**, v. 31, n. 12, p. 3283–3297, dez. 1995.

LAMBE, W. T. The permeability of fine-grained soils. In: SYMPOSIUM ON PERMEABILITY OF SOILS, Annual Meeting American Society for Testing Materials, 57, 1954, Chicago, *Proceedings*, ASTM, p. 56-67, 1954.

LAMBE, T. W.; WHITMAN, R. V. **Soil Mechanics**. Nova York: John Wiley & Sons, 1969

LEE, P. Y.; SUEDKAMP, R. J. Characteristics of irregularly shaped compaction curves of soils. **Highway Research Record**. n. 381, p.1–9, 1972.

LIU, Y. F.; JENG, D.-S. Pore scale study of the influence of particle geometry on soil permeability. **Advances in Water Resources**, v. 129, p. 232–249, jul. 2019.

LU, N.; LIKOS, W. J. **Unsaturated Soil Mechanics**. Hoboken, NJ, USA: John Wiley & Sons, Inc., 2004.

MAROCO, J. **Análise estatística com utilização do SPSS**. 3 ed. Lisboa (Portugal): Edições Sílabo, 2007.

MASROURI, F.; BICALHO, K. V.; KAWAI, K. Hydraulic testing in unsaturated soils. **Geotechnical and Geological Engineering**, v. 26, p. 691-704, 2008.

MCCULLOCH, W. S.; PITTS, W. A logical calculus of the ideas immanent in nervous activity. **Bulletin of Mathematical Biology**, v. 52, n. 1/2, p. 99-115, 1990.

MITCHELL, J. K.; SOGA, K. **Fundamentals of soil behavior**. 3 ed. Nova Jersey: John Wiley & Sons, 2005.

MONTGOMERY, D. C.; PECK, E. A.; VINING, G. G. **Introduction to linear regression analysis**. 5 ed. Hoboken, Nova Jersey (EUA): John Wiley & Sons, 2012.

\_\_\_\_\_; RUNGER, G. C. **Estatística aplicada e probabilidade para engenheiros**. 5 ed. Rio de Janeiro: LTC Editora, 2012.

MUHUNTHAN, B. Liquid limit and surface area of clays. **Geotechnique**, v. 41, n. 1, p. 135-138, 1991.

NEMES, A.; SCHAAP, M. G.; LEIJ, F. J.; WÖSTEN, J.H. M. Description of the unsaturated soil hydraulic database UNSODA version 2.0. **Journal of Hydrology**, v. 251, n. 3–4, p. 151–162, out. 2001.

OLIVEIRA FILHO, A. G. de. Análise comparativa da estimativa do índice de compressão de argilas por redes neurais artificiais e correlações empíricas. Dissertação (Mestrado em Engenharia Civil) – Universidade Federal do Espírito Santo, Centro Tecnológico, 2019, 204 f.

OTTONI, M. V.; OTTONI FILHO, M. G.; 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.

PAPADOPOULOU, A; TIKA, T. The effect of fines on critical state and liquefaction resistance characteristics of non-plastic silty sands. **Soils and Foundations**, v. 48, n. 5, p. 713-725, 2008.

PARK, H. I. Development of neural network model to estimate the permeability coefficient of soil. **Marine Georesources & Geotechnology**, v. 29, n. 4, p. 267-278, 2011.

PUCKETT, W. E.; DANE, J. H.; HAJEK, B. F. Physical and Mineralogical Data to Determine Soil Hydraulic Properties. **Soil Science Society of America Journal**, v. 49, n. 4, p. 831–836, jul. 1985.

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.

RODRIGUES, S. C. A. Equação de regressão linear e suas aplicações. Relatório de estágio (Mestrado em Ensino de Matemática no 3º Ciclo do Ensino Básico e no Ensino Secundário) - Universidade da Beira Interior, Covilhã (Portugal), 2012, 94 p.

ROGIERS *et al.* Estimation of hydraulic conductivity and its uncertainty from grain-size data using GLUE and artificial neural networks. **Mathematical Geosciences**, v. 44, n. 6, 2012.

ROSENBLATT, F. The perceptron: a probabilistic model for information storage and organization in the brain. **Psychological Review**, v. 65, n. 6, p. 386-408, 1958.

SALARASHAYERI, A. F.; SIOSEMARDE, M. Prediction of soil hydraulic conductivity from particle-size distribution. **International Journal of Geological and Environmental Engineering**, v. 6, n.1, 2012.

SANTOS, M. D. dos. Correlações entre sondagem de simples reconhecimento e resultados de ensaios de campo (SPT, CPT, DP) para diferentes solos arenosos. Dissertação (Mestrado em Engenharia Civil) - Universidade Federal do Espírito Santo, Centro Tecnológico, 2017, 175 p.

SANZENI, A.; COLLESELLI, F.; GRAZIOLI, D. Specific surface and hydraulic conductivity of fine-grained soils. **Journal of Geotechnical and Geoenvironmental Engineering**, v. 13, n. 10, p. 1828-1832, 2013.

SEDAGHAT, A.; BAYAT, H.; SINAGANI, A. A. S. Estimation of soil saturated hydraulic conductivity by artificial neural networks ensemble in smectitic soil. **Eurasian Soil Science**, v. 49, n. 3, p. 347-357, 2016.

SHAHABI, A.A.; DAS, B.M.; TARQUIN, A.J. An empirical relation for coefficient of permeability of sand. **Fourth Australia - New Zealand Conference on Geomechanics**, 84(2), p. 54-57, 1984.

SHAHIN, M. A.; JAKSA, M. B.; MAIER, H. R. Artificial neural network applications in geotechnical engineering. **Australian Geomechanics Journal**, v. 36, n. 1, p. 49-62, 2001.

SHAKER, A. A.; DAFALLA, M.; AL-MAHBASHI, A. M.; AL-SHAMRANI, M. A. Predicting Hydraulic Conductivity for Flexible Wall Conditions Using Rigid Wall Permeameter. **Water**, v. 14, n. 3, p. 286, 19 jan. 2022.

SINGH, V. K.; *et al.* Modelling of soil permeability using different data driven algorithms based on physical properties of soil. **Journal of Hydrology**, v. 580, p. 1-10, 2020.

SIVAPULLAIAH, P. V.; SRIDHARAN, A.; STALIN, V. K. Hydraulic conductivity of bentonite-sand mixtures. **Canadian Geotechnical Journal**, 37, p. 406-413, 2000.

TAYLOR, D. W. **Fundamentals of soil mechanics**. Nova York: John Wiley & Sons, 1948.

TIZPA, P. *et al.* ANN prediction of some geotechnical properties of soil from their index parameters. **Arabian Journal of Geosciences**, v. 8, n. 5, n.p., 2014.

TORMAN, V. B. L.; COSTER, R.; RIBOLDI, J. Normalidade de variáveis: métodos de verificação e comparação de alguns testes não-paramétricos por simulação. **Revista do Hospital das Clínicas de Porto Alegre**, v. 32, n. 2, p. 227-234, 2012.

TREJO-ALONSO, J. *et al.* Saturated hydraulic conductivity estimation using artificial neural networks. **Water**, v. 13, n. 705, p. 1-15, 2021.

U.S. Department of Agriculture (USDA). **U.S. Soil survey manual**, v. 18, cap. 3. Department of Agriculture Soil Survey Division Staff, Soil Conservation Service, 1993.

U.S. Department of Agriculture (USDA). **Measurement and Estimation of Permeability of Soils for Animal Waste Storage Facility Design**, Series n. 717,

South National Technical Center, Soil Conservation Service, Fort Worth, Texas, 1991.

WANG, M. C.; HUANG, C. C. Soil Compaction and Permeability Prediction Models. **Journal of Environmental Engineering**, v. 110, n. 6, p. 1063–1083, dez. 1984.

WIEBENGA, W. A. *et al.* Empirical relations in properties of unconsolidated quartz sands and silts pertaining to water flow. **Water Resources Research**, v. 6, p. 1154-1161, 1970.

WILLIAMS, C. G.; OJURI, O. O. Predictive modelling of soil's hydraulic conductivity using artificial neural network and multiple linear regression. **SN Applied Sciences**, v. 3, n. 152, n.p., 2021

YANG, S.; LACASSE, S.; SANDVEN, R. Determination of the Transitional Fines Content of Mixtures of Sand and Non-plastic Fines. **Geotechnical Testing Journal**, v. 29, n. 2, p. 1-7, 2006.

YILMAZ *et al.* Neural computing models for prediction of permeability coefficient of coarse-grained soils. **Neural Computing and Applications**, v. 21, n. 5, p. 1-12, 2011

YUKSELEN, Y.; KAYA, A. Comparison of Methods for Determining Specific Surface Area of Soils. **Journal of Geotechnical and Geoenvironmental Engineering**, v. 132, n. 7, p. 931–936, jul. 2006.

ZIVKOVIC, P.; SAFRAN, M. B.; ZELIC, B. K. Comparison of measured and estimated permeability for artificially prepared coarse-grained soil samples. **The Mining Geology Petroleum Engineering Bulletin**, v.36, n. 3, p. 167-178, 2021.

**ANEXO 1 – Durbin-Watson: Valores críticos de  $d_L$  e  $d_U$  para  $\alpha = 0,05$** 

		Valores críticos de $d_L$ e $d_U$ para $\alpha = 0.05$													
$n \backslash p$		1		2		3		4		5		10		15	
		$d_L$	$d_U$	$d_L$	$d_U$	$d_L$	$d_U$	$d_L$	$d_U$	$d_L$	$d_U$	$d_L$	$d_U$	$d_L$	$d_U$
6		0.61	1.40												
10		0.88	1.32	0.70	1.64	0.53	2.02	0.38	2.41	0.24	2.82				
20		1.20	1.41	1.10	1.54	1.00	1.68	0.89	1.83	0.79	1.99	0.34	2.89	0.06	3.68
30		1.35	1.49	1.28	1.57	1.21	1.65	1.14	1.74	1.07	1.83	0.71	2.36	0.39	2.94
40		1.44	1.54	1.39	1.60	1.34	1.66	1.29	1.72	1.23	1.79	0.95	2.15	0.68	2.56
50		1.50	1.59	1.46	1.63	1.42	1.67	1.38	1.72	1.34	1.77	1.11	2.05	0.88	2.35
60		1.55	1.62	1.51	1.65	1.48	1.69	1.44	1.73	1.41	1.77	1.22	1.98	1.03	2.28
70		1.58	1.64	1.55	1.67	1.53	1.70	1.49	1.74	1.46	1.77	1.31	1.95	1.14	2.15
80		1.61	1.66	1.59	1.69	1.56	1.72	1.53	1.74	1.51	1.77	1.37	1.93	1.22	2.09
90		1.64	1.68	1.61	1.70	1.59	1.73	1.57	1.75	1.54	1.78	1.42	1.91	1.29	2.06
100		1.65	1.69	1.63	1.72	1.61	1.74	1.59	1.76	1.57	1.78	1.46	1.90	1.35	2.03
200		1.76	1.78	1.75	1.79	1.74	1.80	1.73	1.81	1.72	1.82	1.67	1.87	1.61	1.93

Fonte: Moroco (2007).

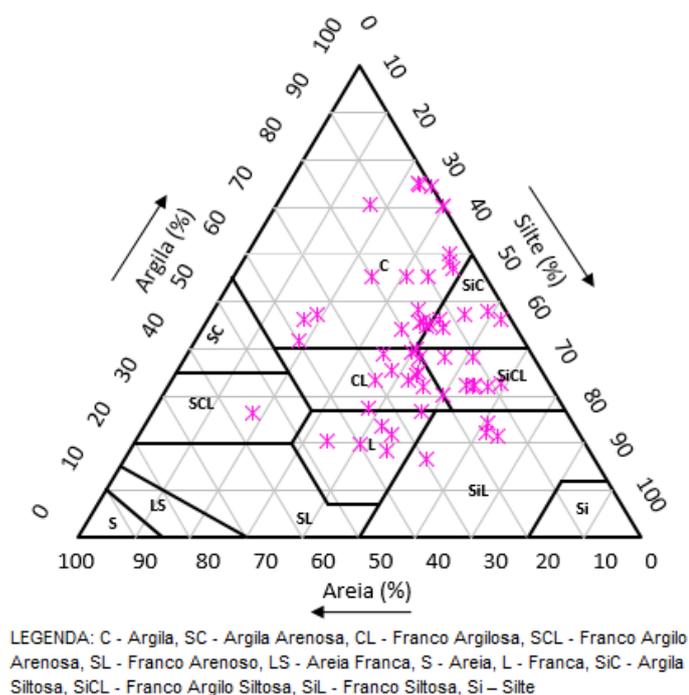
## APÊNDICE 1 – Descrição e detalhamento dos bancos de dados obtidos da literatura

### Dados disponibilizados em Benson *et al.* (1984)

Um banco de dados, originalmente, com informações de 67 amostras de solos utilizados em barreiras impermeáveis de aterros sanitários na América do Norte foi compilado e apresentado por Benson *et al.* (1984). Do total, 56 amostras apresentam informações sobre as frações granulométricas, conjunto este avaliado para esta pesquisa. As amostras de Benson *et al.* (1984) são compostas por solos argilosos compactados; corpos de provas indeformados foram retirados por tubo *Shelby* ou extração de blocos de um aterro com solo natural argiloso.

As amostras eram compostas por dois tipos de finos: um silte, como fino não plástico, e uma bentonita (montmorilonita sódica), como fino plástico. A determinação do coeficiente de permeabilidade foi realizada em permeômetros de parede flexível, diretamente no tubo *Shelby* ou em células de adensamento equipadas com medidor de permeabilidade.

O banco de dados contém as seguintes informações sobre as amostras: LL, IP, atividade, %pedregulho, %areia, %finos, %argila,  $w$  natural,  $\gamma_d$  e parâmetros de compactação. A distribuição dos solos no triângulo textural é mostrada na Figura 15.

Figura 15 – Triângulo textural – 56 amostras de Benson *et al.* (1984)

Fonte: Autora (2022).

A disposição das amostras no triângulo textural confirma também que se trata de amostras predominantemente argilosas e praticamente ausência de materiais de granulação grossa, como na Tabela 33.

Tabela 33 - Classificação relativa da permeabilidade das 56 amostras de Benson *et al.* (1984)

Classificação relativa da permeabilidade	Intervalo de referência (m/s)	Amostras plásticas nº de amostras
Extremamente alta	$k_{sat} \geq 10^{-2}$	-
Muito alta	$3,53 \times 10^{-4} \leq k_{sat} < 10^{-2}$	-
Alta	$3,53 \times 10^{-6} \leq k_{sat} < 3,53 \times 10^{-4}$	-
Média	$3,53 \times 10^{-7} \leq k_{sat} < 3,53 \times 10^{-6}$	-
Baixa	$3,53 \times 10^{-8} \leq k_{sat} < 3,53 \times 10^{-7}$	-
Muito baixa	$3,53 \times 10^{-9} \leq k_{sat} < 3,53 \times 10^{-8}$	2
Praticamente impermeável	$k_{sat} < 3,53 \times 10^{-9}$	54
TOTAL DE AMOSTRAS		56

Fonte: Autora (2022).

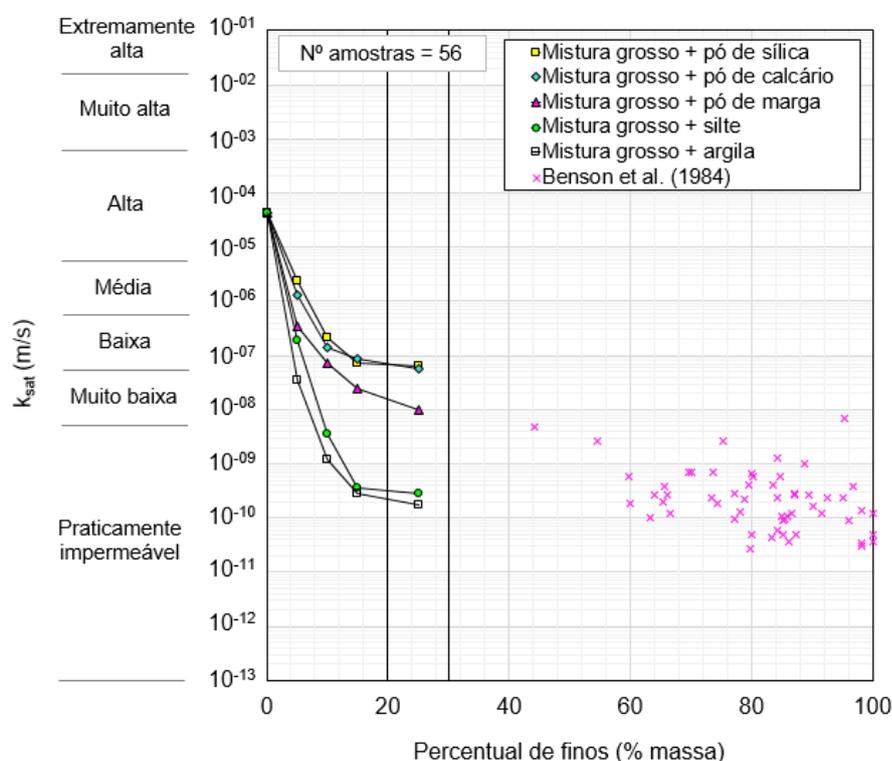
Como o critério de seleção das amostras por Benson *et al.* (1984) contemplava solos utilizados para barreiras sanitárias em aterros compactados com argilas naturais, a maioria das amostras apresenta teor de finos acima de 60%. E assim 96% das amostras apresentam  $k_{sat}$  equivalente a solo praticamente impermeável, o que é coerente com os critérios de seleção de corpos de prova adotados pelos pesquisadores. Foi possível realizar a classificação das amostras no SUCS:

- Argila de alta plasticidade (CH): 18 amostras
- Argila de baixa plasticidade (CL): 35 amostras
- Silte de alta plasticidade (MH): 2 amostras
- Areia argilosa (SC): 1 amostra

Tanto a classificação pelo triângulo textural quanto pelo SUCS estão coerentes com a origem das amostras de solos utilizadas pelos pesquisadores Benson *et al.* (1984). Os valores de  $k_{sat}$ , para o conjunto de dados considerado, variam de  $2,70 \times 10^{-11}$  m/s a  $6,70 \times 10^{-09}$  m/s.

Os resultados do banco de dados foram comparados às curvas de  $k_{sat}$  em função do teor de finos obtidas por Barber e Sawyer (1952) para misturas controladas em que uma mesma matriz granular pura para evidenciar a influência da plasticidade do fino. Pela composição predominante de finos plásticos, as amostras de Benson *et al.* (1984) estão próximas às curvas de Barber e Sawyer (1952) que apresentaram plasticidade, como mostrado no Gráfico 16.

Gráfico 16 - Variação de  $k_{sat}$  com o teor de finos para diferentes misturas controladas (Dados de: BARBER; SAWYER, 1952; BENSON *et al.*, 1984)



Fonte: Autora (2022).

As amostras de Benson *et al.* (1984) são compostas por três frações granulométricas: areia, silte e argila, indicada pelos pesquisadores como bentonita (não indicam a categoria da montmorilonita). Os percentuais entre cada fração são variáveis, por se tratarem se amostras naturais. A predominância é de solos com teor de finos acima de 60% e, dessas amostras, 26 apresentam teor de argila maior que o teor de silte.

Ao contrário das amostras de Wang e Huang (1984), as amostras de Benson *et al.* (1984) apresentam um IP médio de  $27 \pm 15$  (%) e um IP variando de 7% a 62%. Considerando a teoria apresentada e discutida, essa plasticidade tem efeito sobre os valores de  $k_{sat}$  desse banco de dados, principalmente, em comparação com os dados de Wang e Huang (1984).

O banco de dados de Benson *et al.* (1984) não apresenta informações sobre índice de vazios nem diâmetro efetivo das amostras e, portanto, análises da influência do índice de vazios não podem ser realizadas. Entretanto, apresenta os valores de  $\rho_t$  e  $\rho_d$ , mas, considerando que são amostras indeformadas provenientes de 67 aterros localizados na América do Norte, não é possível inferir com exatidão sobre o índice de vazios e o  $G_s$  das amostras, pois são provenientes de origens e processos de compactação possivelmente distintos.

### **Dados disponibilizados em Ottoni *et al.* (2018)**

Ottoni *et al.* (2018) buscaram consolidar dados de retenção de água e coeficiente de permeabilidade com dados de propriedades índices básicas do solo, bem como identificando o método de determinação dessas características, com o objetivo de fornecer um banco de dados de solos brasileiros para o desenvolvimento de funções de pedotransferência mais representativas para essa região.

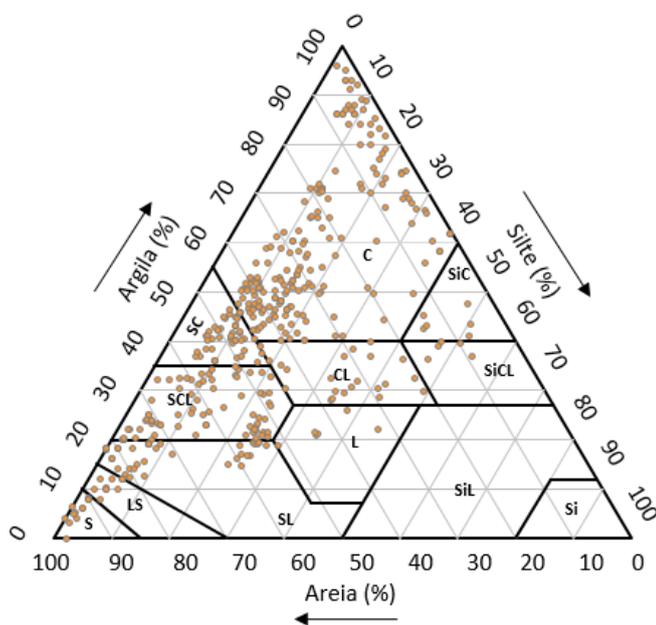
Ao todo, existem dados de 1075 amostras de solos, distribuídas em 15 unidades federativas, e que formam o Banco de Dados Hidrofísicos de Solos no Brasil (HYBRAS). O banco fornece informações de propriedades como frações granulométricas (%argila, %silte e %areia), massa específica do solo, massa específica das partículas, teor de matéria orgânica, coeficiente de permeabilidade, valores de sucção, dados de retenção de água etc.

Os tipos de solos predominantes nas amostras totais do HYBRAS são solos típicos de ambientes tropicais intemperizados, com teores de silte menores que 20% e baixa representativa de solos de textura grossa. Solos com alto teor de silte (acima de 50%), com elevado teor de matéria orgânica (acima de 6%) e de baixa massa específica (menor que 0,8g/cm<sup>3</sup>) foram pouco representados no banco de dados.

Todavia, os conjuntos de dados que apresentam valores de  $k_{sat}$  e valores de outras propriedades índices (%argila, %silte, %areia,  $e$ ,  $\rho_d$  e  $G_s$ ) de potencial interesse para essa pesquisa totalizam 419 amostras distribuídas em 10 estados brasileiros.

A disposição das amostras selecionadas no triângulo textural é fornecida na Figura 16 e confirma uma predominância, após a seleção das 419 amostras, de solos com teores de argila ou silte abaixo de 50%, embora as amostras apresentem teor de finos superior acima de 50% (275 amostras ao todo). Não foi possível realizar a classificação pelo SUCS das amostras, pois não são disponibilizadas informações sobre os limites de consistência,  $C_u$  e  $C_c$ .

Figura 16 – Triângulo textural – 419 amostras de Ottoni *et al.* (2018)



LEGENDA: C - Argila, SC - Argila Arenosa, CL - Franco Argilosa, SCL - Franco Argilo Arenosa, SL - Franco Arenoso, LS - Areia Franca, S - Areia, L - Franca, SiC - Argila Siltosa, SiCL - Franco Argilo Siltosa, SiL - Franco Siltosa, Si - Silte

Fonte: Autora (2022).

Considerando os critérios de classificação de  $k_{sat}$  da Nota Técnica N° 717 do *Soil Conservation Service* dos Estados Unidos da América (USDA, 1991), as amostras de Ottoni *et al.* (2018) estão distribuídas segundo a Tabela 34.

Tabela 34 – Classificação relativa da permeabilidade das amostras de Ottoni *et al.* (2018)

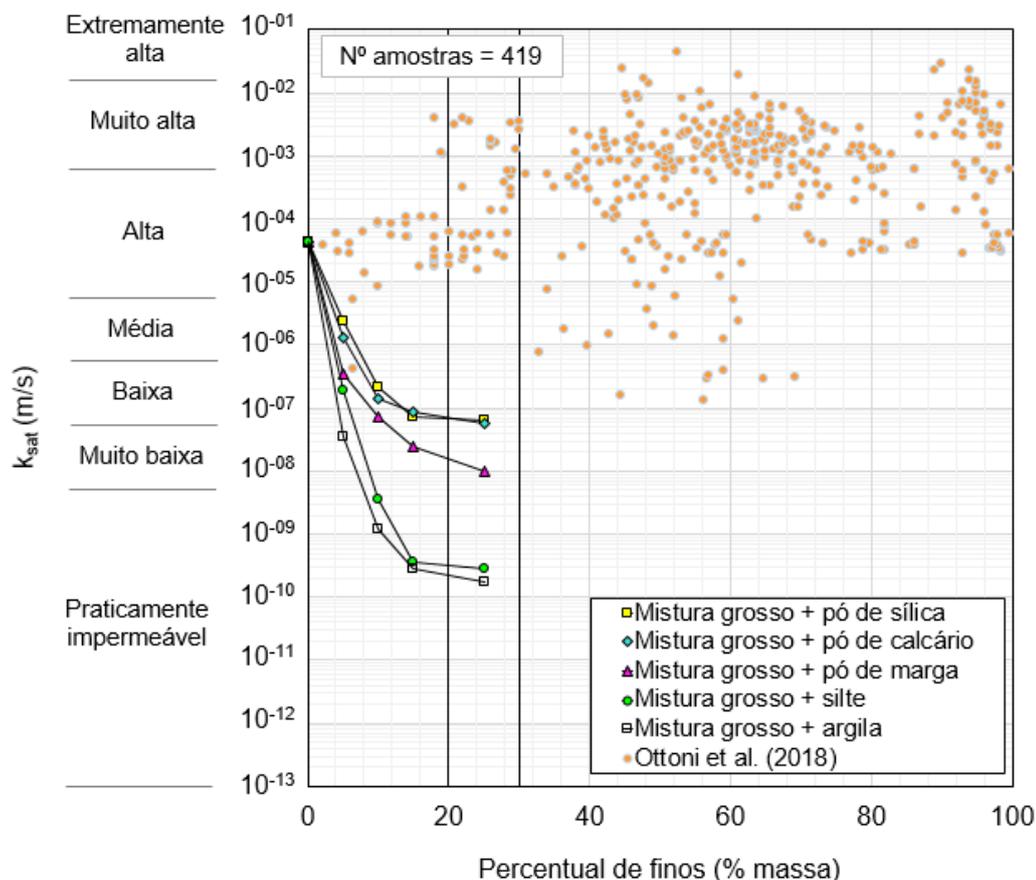
Classificação relativa da permeabilidade	Intervalo de referência (m/s)	Nº de amostras*
Extremamente alta	$k_{sat} \geq 10^{-2}$	-
Muito alta	$3,53 \times 10^{-4} \leq k_{sat} < 10^{-2}$	256
Alta	$3,53 \times 10^{-6} \leq k_{sat} < 3,53 \times 10^{-4}$	147
Média	$3,53 \times 10^{-7} \leq k_{sat} < 3,53 \times 10^{-6}$	10
Baixa	$3,53 \times 10^{-8} \leq k_{sat} < 3,53 \times 10^{-7}$	6
Muito baixa	$3,53 \times 10^{-9} \leq k_{sat} < 3,53 \times 10^{-8}$	-
Praticamente impermeável	$k_{sat} < 3,53 \times 10^{-9}$	-
<b>TOTAL DE AMOSTRAS</b>		<b>419</b>

\* Sem informação sobre a plasticidade das amostras.

Fonte: Autora (2022).

As 419 amostras selecionadas apresentam características típicas de permeabilidade alta e muito alta, uma vez que 96,2% das amostras apresentam  $k_{sat}$  entre superior a  $3,53 \times 10^{-6}$  m/s. A variação de  $k_{sat}$  com o teor de finos em amostras selecionadas de Ottoni *et al.* (2018) também foi analisada e apresentada no Gráfico 17.

Gráfico 17 - Variação de  $k_{sat}$  com o teor de finos para as 419 amostras de diferentes solos de Ottoni *et al.* (2018) (Dados de: BARBER; SAWYER, 1952; OTTONI *et al.*, 2018)



Fonte: Autora (2022).

Nas amostras selecionadas, observa-se uma predominância de solos com permeabilidade relativa alta (acima de  $10^{-5}$  m/s) ou muito alta (acima de  $10^{-4}$  m/s), independentemente do teor de finos. No Gráfico 17, também estão apresentadas as curvas de Barber e Sawyer (1952) para comparação de  $k_{sat}$  com a plasticidade.

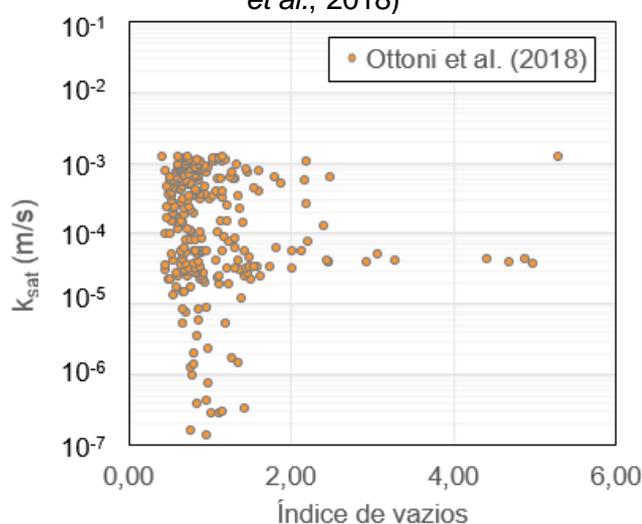
As amostras selecionadas de Ottoni *et al.* (2018) são principalmente de solos típicos de ambientes tropicais intemperizados. O banco de dados não indica o mineral que compõem a fração argila e que poderia ajudar a compreender os valores de  $k_{sat}$  e o posicionamento acima da curva de baixa plasticidade dos resultados de Barber e Sawyer (1952) (indicativo da plasticidade do fino).

As amostras com teor de finos menor que 5% apresentam  $k_{sat}$  da ordem de  $10^{-5}$  m/s, ou seja, com valores relativamente considerados de permeabilidade alta. Já as amostras com teor de finos acima de 95% (41 amostras) apresentam valores de  $k_{sat}$  que vão de  $3,07 \times 10^{-5}$  m/s a  $1,5 \times 10^{-2}$  m/s. Ou seja, no banco de dados, existem

amostras com teor de finos próximo a 100% com o  $k_{sat}$  da mesma ordem de grandeza que amostras com 100% de grossos.

No banco de dados, são disponibilizadas informações sobre índice de vazios. A relação  $k_{sat}$  vs índice de vazios foi investigada e apresentada no Gráfico 18. Todavia, pouco se esclarece sobre a relação de  $k_{sat}$  com o índice de vazios das amostras de Ottoni *et al.* (2018), indicando que há outras propriedades interferindo nos valores de  $k_{sat}$  ou que auxiliariam a interpretar melhor o comportamento das amostras.

Gráfico 18 - Variação de  $k_{sat}$  com índice de vazios para diferentes solos (Dados de: OTTONI *et al.*, 2018)



Fonte: Autora (2022).

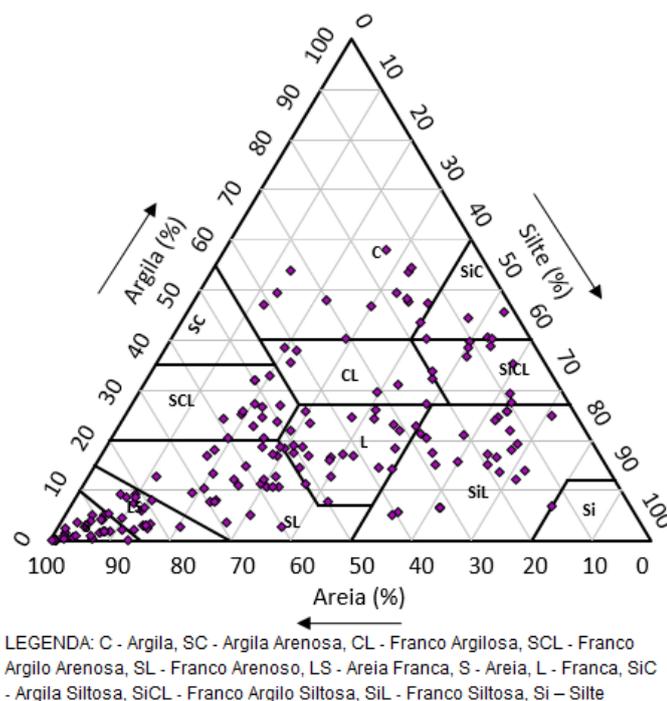
A disponibilidade de valores para propriedades índices que exercem influência sobre  $k_{sat}$  poderiam auxiliar na interpretação dos dados. Percebe-se que, à medida que os bancos de dados vão apresentando menos informações sobre as amostras, principalmente as informações sobre a característica dos finos menos é possível inferir sobre as amostras e sobre o comportamento dos valores  $k_{sat}$ .

### Dados disponibilizados em Nemes *et al.* (2001)

Nemes *et al.* (2001) compilaram o chamado *Unsaturated Soil Hydraulic Database* (UNSODA), que é um banco de dados com amostras de várias localidades na Ásia, Europa e América do Norte. De todo o banco de dados, 172 conjuntos de dados foram selecionados como potenciais para serem utilizados nesta pesquisa, por apresentarem dados completos de  $k_{sat}$ , %argila, %silte, %areia,  $\rho_t$ ,  $G_s$  e  $e$ .

A distribuição das amostras no triângulo textural, na Figura 17, confirma se tratar de amostras predominantemente de características granulares, especialmente, silto arenosas.

Figura 17 – Triângulo textural – amostras de 172 amostras de Nemes *et al.* (2001)



Fonte: Autora (2022).

As amostras de Nemes *et al.* (2001) não puderam ser classificadas no SUCS devido ausência de informações necessárias. Além disso, as amostras estão concentradas em solos que apresentam  $k_{sat}$  entre  $10^{-9}$  m/s e  $10^{-3}$  m/s, como mostrado na Tabela 35.

Tabela 35 - Classificação relativa da permeabilidade das amostras de Nemes *et al.* (2001)

Classificação relativa da permeabilidade	Intervalo de referência (m/s)	Nº de amostras*
Extremamente alta	$k_{sat} \geq 10^{-2}$	-
Muito alta	$3,53 \times 10^{-4} \leq k_{sat} < 10^{-2}$	7
Alta	$3,53 \times 10^{-6} \leq k_{sat} < 3,53 \times 10^{-4}$	111
Média	$3,53 \times 10^{-7} \leq k_{sat} < 3,53 \times 10^{-6}$	38
Baixa	$3,53 \times 10^{-8} \leq k_{sat} < 3,53 \times 10^{-7}$	13
Muito baixa	$3,53 \times 10^{-9} \leq k_{sat} < 3,53 \times 10^{-8}$	3
Praticamente impermeável	$k_{sat} < 3,53 \times 10^{-9}$	-
<b>TOTAL DE AMOSTRAS</b>		<b>172</b>

\* Sem informação sobre a plasticidade das amostras.

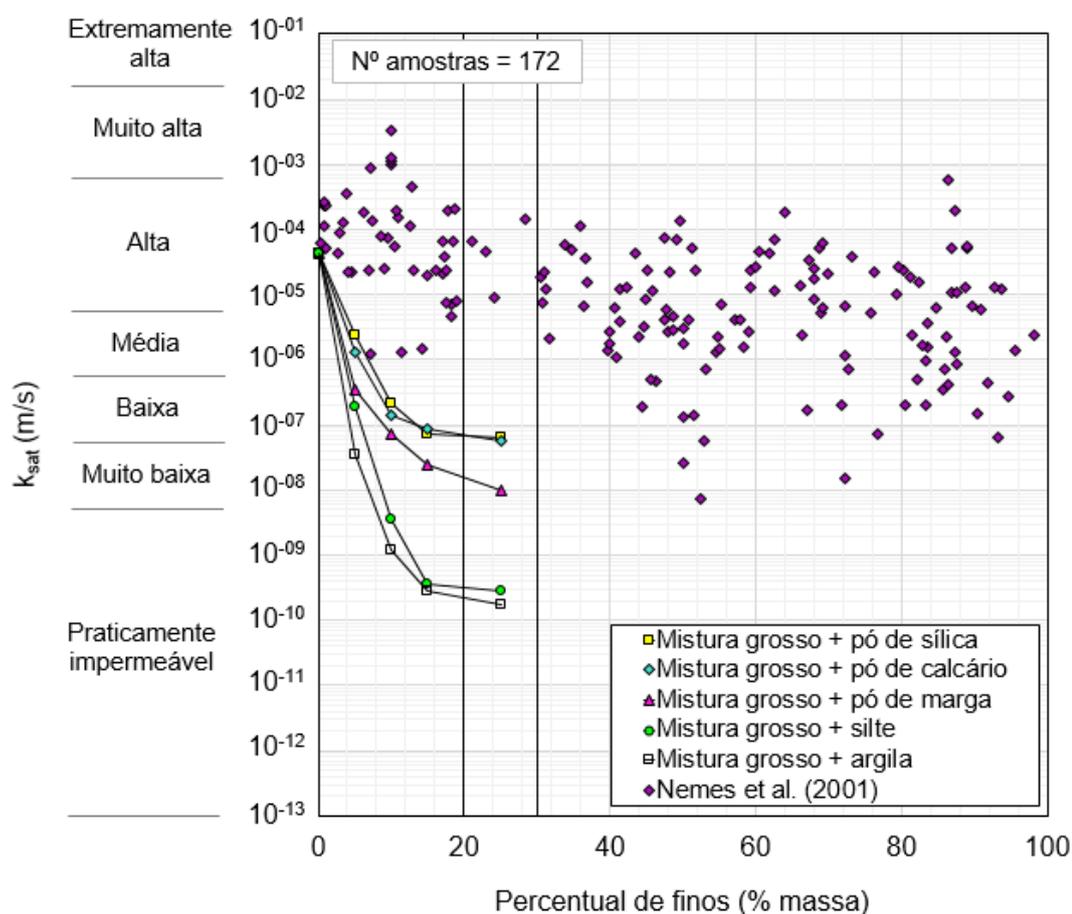
Fonte: Autora (2022).

A faixa de valores de  $k_{sat}$  compreendida nas amostras selecionadas do UNSODA indica a predominância de amostras com permeabilidade média a alta, segundo

critérios da Nota Técnica Nº 717 do *Soil Conservation Service* dos Estados Unidos da América (*USDA*, 1991).

As amostras são compostas basicamente por solos com  $k_{sat}$  variando de  $7,52 \times 10^{-9}$  m/s a  $3,20 \times 10^{-3}$  m/s e, portanto, posicionadas principalmente acima da curva de mistura não plástica dos resultados de Barber e Sawyer (1952), que também são mostrados no Gráfico 19. Também não são fornecidas informações sobre a mineralogia da fração fina que compõem as amostras, nem limites de consistência, que poderiam auxiliar a inferir a respeito.

Gráfico 19 - Variação de  $k_{sat}$  com o teor de finos para 172 amostras de diferentes solos de Nemes *et al.* (2001) (Dados de: BARBER; SAWYER, 1952; NEMES *et al.*, 2001)



Fonte: Autora (2022).

As amostras com teor de finos menor que 5% apresentam  $k_{sat}$  que varia de  $2,20 \times 10^{-5}$  m/s a  $3,20 \times 10^{-3}$  m/s, ou seja, com valores relativamente considerados de permeabilidade alta a permeabilidade muito alta. Já as amostras com teor de finos acima de 80% (37 amostras) apresentam valores de  $k_{sat}$  que vão de  $5,56 \times 10^{-8}$  m/s a

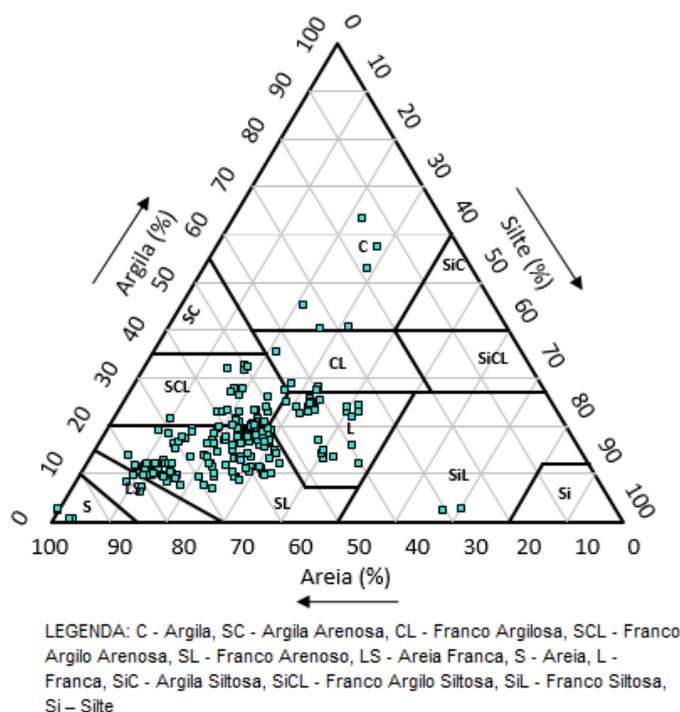
$4,45 \times 10^{-4}$  m/s. O fornecimento de maiores informações sobre as amostras poderia auxiliar a interpretar melhor a variabilidade nos valores de  $k_{sat}$ .

### **Dados disponibilizados em Rahmati *et al.* (2018)**

Rahamati *et al.* (2018) organizaram um banco com dados de curvas de infiltração de amostras ao longo de todo o globo que constituiu o *Soil Water Infiltration Global* (SWIG). O conjunto de dados inclui informações dos locais de medida de infiltração, propriedades diversas das amostras e uso do solo no local de coleta.

O banco de dados foi coletado a partir de publicações na literatura e em contato direto com pesquisadores, detentores de dados, para colaborarem com o fornecimento de informações. Ao todo, 5023 curvas de infiltração foram organizadas, originadas de 54 países. Deste total de 5023 dados, 196 conjuntos de dados foram selecionados contendo: %argila, %silte, %areia, %MO,  $d_{50}$ ,  $\rho_t$ ,  $\rho_s$  e  $k_{sat}$ . Essas 196 amostras são provenientes do Irã e da França, mas não é mencionada a natureza das amostras: se naturais ou misturas controladas.

O triângulo textural das amostras foi construído e é apresentado na Figura 18. Pode-se verificar e confirmar a presença de solos classificados, em sua maioria, como de matriz granular.

Figura 18 – Triângulo textural – 196 amostras de Rahmati *et al.* (2018)

Fonte: Autora (2022).

A distribuição das amostras mostra uma predominância de solos com teor de areia acima de 50% e poucas amostras de solos argilosos ou siltosos. Na Tabela 36, é mostrada a distribuição relativa dos valores de  $k_{sat}$ , segundo os intervalos da *USDA* (1991) de permeabilidade.

Tabela 36 – Classificação relativa da permeabilidade das amostras de Rahmati *et al.* (2018)

Classificação relativa da permeabilidade	Intervalo de referência (m/s)	Nº de amostras*
Extremamente alta	$k_{sat} \geq 10^{-2}$	-
Muito alta	$3,53 \times 10^{-4} \leq k_{sat} < 10^{-2}$	6
Alta	$3,53 \times 10^{-6} \leq k_{sat} < 3,53 \times 10^{-4}$	182
Média	$3,53 \times 10^{-7} \leq k_{sat} < 3,53 \times 10^{-6}$	7
Baixa	$3,53 \times 10^{-8} \leq k_{sat} < 3,53 \times 10^{-7}$	1
Muito baixa	$3,53 \times 10^{-9} \leq k_{sat} < 3,53 \times 10^{-8}$	-
Praticamente impermeável	$k_{sat} < 3,53 \times 10^{-9}$	-
TOTAL DE AMOSTRAS		196

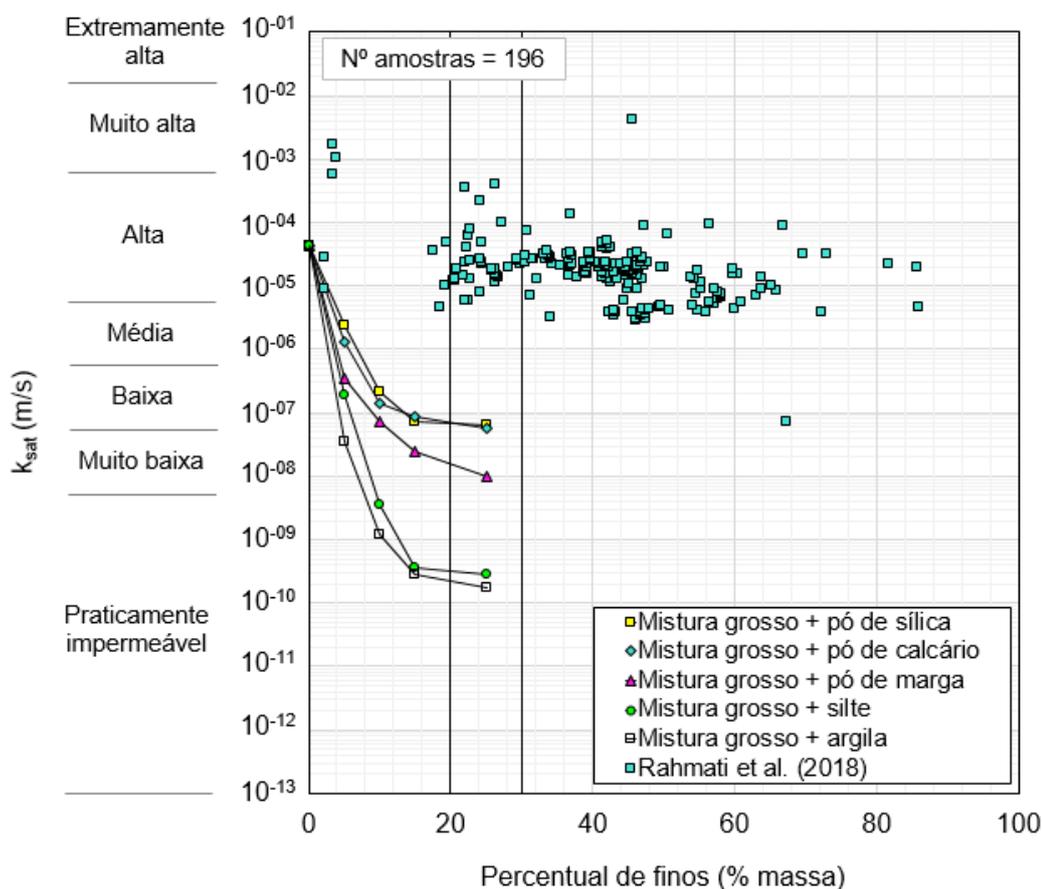
\* Sem informação sobre a plasticidade das amostras.

Fonte: Autora (2022).

De fato, é possível verificar que a maioria dos valores de  $k_{sat}$  estão localizados na faixa considerada de permeabilidade alta. As amostras de Rahmati *et al.* (2018) não puderam ser classificadas no SUCS devido ausência de informações necessárias. Os

valores de  $k_{sat}$  em função do teor de finos foram avaliados e apresentados com os resultados de Barber e Sawyer (1952) no Gráfico 20.

Gráfico 20 - Variação de  $k_{sat}$  com o teor de finos para 196 amostras de diferentes solos de Rahmati *et al.* (2018) (Dados de: BARBER; SAWYER, 1952; RAHMATI *et al.*, 2018)



Fonte: Autora (2022).

Coerentemente posicionadas principalmente acima da curva de mistura não plástica, as amostras de Rahmati *et al.* (2018) são, de fato, compostas basicamente por solos de matriz arenosa (157 amostras apresentam teor de areia maior que 50%), com  $k_{sat}$  variando de  $7,18 \times 10^{-8}$  m/s a  $4,15 \times 10^{-3}$  m/s. É possível perceber também que as amostras selecionadas estão concentradas após a faixa do teor de finos de transição.

As amostras com teor de finos menor que 5% apresentam  $k_{sat}$  que varia de  $9,10 \times 10^{-6}$  m/s a  $1,70 \times 10^{-3}$  m/s, ou seja, com valores relativamente considerados de permeabilidade alta a permeabilidade muito alta. Já as amostras com teor de finos acima de 80% (03 amostras) apresentam valores de  $k_{sat}$  que vão de  $4,51 \times 10^{-6}$  m/s a  $2,2 \times 10^{-5}$  m/s, como mostrado na Tabela 37.

Tabela 37 - Valores de  $k_{sat}$  para as amostras com maiores e menores teores de finos (Dados de: RAHMATI *et al.*, 2018)

Amostras com teor de finos < 5%						
%Argila	%Silte	%Areia	%Finos	$k_{sat}$ (m/s)	$\rho_t$ (g/cm <sup>3</sup> )	$G_s$
0	4	96	4	$1,07 \times 10^{-03}$	0,84	2,65
0	3	97	3	$5,83 \times 10^{-04}$	0,84	2,65
0	3	97	3	$1,70 \times 10^{-03}$	0,84	2,65
2	0	98	2	$2,81 \times 10^{-05}$	1,42	2,65
2	0	98	2	$9,10 \times 10^{-06}$	1,24	2,65
Amostras com teor de finos > 80%						
%Argila	%Silte	%Areia	%Finos	$k_{sat}$ (m/s)	$\rho_t$ (g/cm <sup>3</sup> )	$G_s$
63	23	14	86	$4,51 \times 10^{-06}$	1,18	2,65
57	29	14	86	$1,99 \times 10^{-05}$	1,16	2,65
53	29	18	82	$2,27 \times 10^{-05}$	1,36	2,65

Fonte: Autora (2022).

Além dos dados de  $k_{sat}$  e das frações granulométricas, Rahmati *et al.* (2018) apresentam  $\rho_t$  e  $G_s$  para as amostras selecionadas. Todavia, dados adicionais como  $d_{10}$ , índice de vazios e limites de consistência, não estão disponibilizados e uma avaliação mais conclusiva sobre as propriedades das amostras que justifique os valores de  $k_{sat}$  não é possível de ser realizada. Os valores de  $k_{sat}$  apresentam menor dispersão do que dos demais bancos de dados discutidos até o momento, mas pouca informação é disponibilizada que permitiria uma melhor análise desses valores.

### Dados disponibilizados em Alyamani e Sen (1993)

Um total de 32 conjuntos de dados foram apresentados por Alyamani e Sen (1993). Os pesquisadores reuniram dados de  $d_{10}$ ,  $d_{50}$ ,  $d_{60}$  e  $k_{sat}$  de solos da Arábia Saudita (22 amostras) e Austrália (10 amostras). Apenas as amostras australianas também tinham informações sobre a porosidade e, dos dados fornecidos,  $C_u$  foi inferido.

Como dados sobre  $d_{10}$ ,  $d_{50}$  e  $d_{60}$  são escassos nos bancos de dados levantados na literatura, as amostras apresentadas por Alyamani e Sen (1993) mostraram-se potenciais para estudos de correlações empíricas, haja vista que a maioria das equações utilizam dados como o diâmetro efetivo.

Os pesquisadores utilizaram amostras deformadas, coletadas em diversos locais da Arábia Saudita, para realizarem ensaios de permeabilidade em permeâmetro de carga

constante. Os dados de amostras australianas foram obtidos de Wiebenga *et al.* (1970). O objetivo dos pesquisadores era correlacionar as características de distribuição granulométrica e coeficiente de permeabilidade e uma equação empírica foi proposta, que foi apresentada no Capítulo 2 desta dissertação.

Embora não foi possível construir o triângulo textural das amostras de Alyamani e Sen (1993), por ausência dos percentuais das frações granulométricas, todas as amostras de Alyamani e Sen (1993) são classificadas com valores de  $k_{sat}$  típico de solos de permeabilidade alta a extremamente alta, de acordo com os critérios apresentados na *USDA* (1991), como mostrado na Tabela 38.

Tabela 38 - Classificação relativa da permeabilidade das 32 amostras de Alyamani e Sen (1993)

Classificação relativa da permeabilidade	Intervalo de referência (m/s)	Nº de amostras*
Extremamente alta	$k_{sat} \geq 10^{-2}$	1
Muito alta	$3,53 \times 10^{-4} \leq k_{sat} < 10^{-2}$	12
Alta	$3,53 \times 10^{-6} \leq k_{sat} < 3,53 \times 10^{-4}$	19
Média	$3,53 \times 10^{-7} \leq k_{sat} < 3,53 \times 10^{-6}$	-
Baixa	$3,53 \times 10^{-8} \leq k_{sat} < 3,53 \times 10^{-7}$	-
Muito baixa	$3,53 \times 10^{-9} \leq k_{sat} < 3,53 \times 10^{-8}$	-
Praticamente impermeável	$k_{sat} < 3,53 \times 10^{-9}$	-
TOTAL DE AMOSTRAS		32

\* Sem informação sobre a plasticidade das amostras.

Fonte: Autora (2022).

Por ausência de informações como os limites de consistência e a distribuição das frações granulométricas, não foi possível realizar a classificação no SUCS para as amostras de Alyamani e Sen (1993). Também a relação entre  $k_{sat}$  e o teor de finos não foi possível de ser realizada.

### Dados disponibilizados em Gupta *et al.* (2020)

Gupta *et al.* (2020), com o objetivo de criar um banco de dados global com dados de coeficiente de permeabilidade, coletaram, padronizaram e checaram 13267 medições de  $k_{sat}$ , a partir de uma busca em várias bases de dados, como *Scopus*, *Google Scholar* e *Science Direct*.

O banco de dados contém diversos parâmetros das amostras, como localização,  $\rho_t$ , umidade para diferentes valores de sucção, %argila, %silte, %areia, %CO,  $k_{sat}$  etc., e abrangem amostras localizadas ao longo dos cinco continentes.

Do total de dados, 7227 conjuntos de dados foram selecionados por apresentarem informações completas de %argila, %silte, %areia,  $\rho_t$  e  $k_{sat}$ . Desses dados selecionados, 6502 apresentaram também %CO.

As amostras selecionadas de Gupta *et al.* (2020), considerando somente os conjuntos que apresentam dados de %argila, %silte, %areia,  $\rho_t$  e  $k_{sat}$  (7227 amostras), estão classificadas, em sua maioria, com  $k_{sat}$  equivalente a permeabilidade média e alta. Apenas 31 amostras apresentaram coeficiente de permeabilidade menor que  $3,53 \times 10^{-9}$  m/s, como visto na Tabela 39.

Tabela 39 - Classificação relativa da permeabilidade das amostras de Gupta *et al.* (2020)

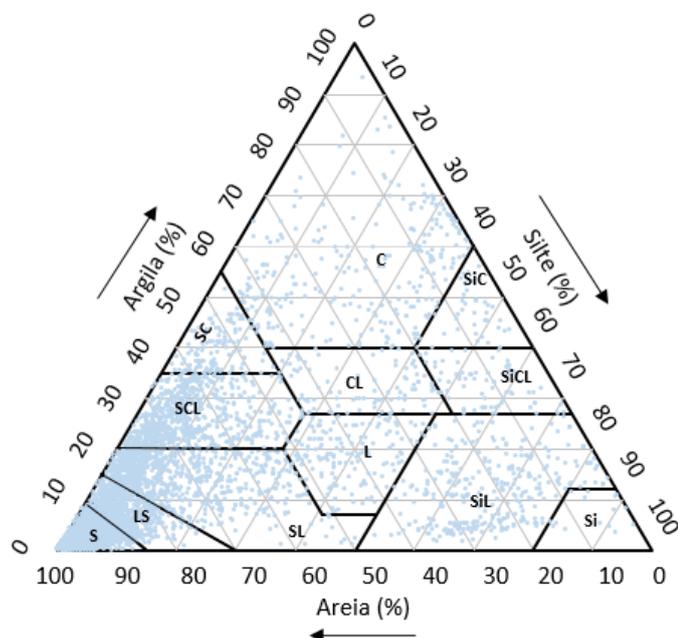
Classificação relativa da permeabilidade	Intervalo de referência (m/s)	Nº de amostras*
Extremamente alta	$k_{sat} \geq 10^{-2}$	-
Muito alta	$3,53 \times 10^{-4} \leq k_{sat} < 10^{-2}$	88
Alta	$3,53 \times 10^{-6} \leq k_{sat} < 3,53 \times 10^{-4}$	5378
Média	$3,53 \times 10^{-7} \leq k_{sat} < 3,53 \times 10^{-6}$	1224
Baixa	$3,53 \times 10^{-8} \leq k_{sat} < 3,53 \times 10^{-7}$	399
Muito baixa	$3,53 \times 10^{-9} \leq k_{sat} < 3,53 \times 10^{-8}$	107
Praticamente impermeável	$k_{sat} < 3,53 \times 10^{-9}$	31
<b>TOTAL DE AMOSTRAS</b>		<b>7227</b>

\* Sem informação sobre a plasticidade das amostras.

Fonte: Autora (2022).

As amostras foram classificadas no triângulo textural, como mostrado na Figura 19. Por ausência de informações, não foi possível classificá-las segundo o SUCS.

Figura 19 – Triângulo textural – 7227 amostras de Gupta *et al.* (2020)

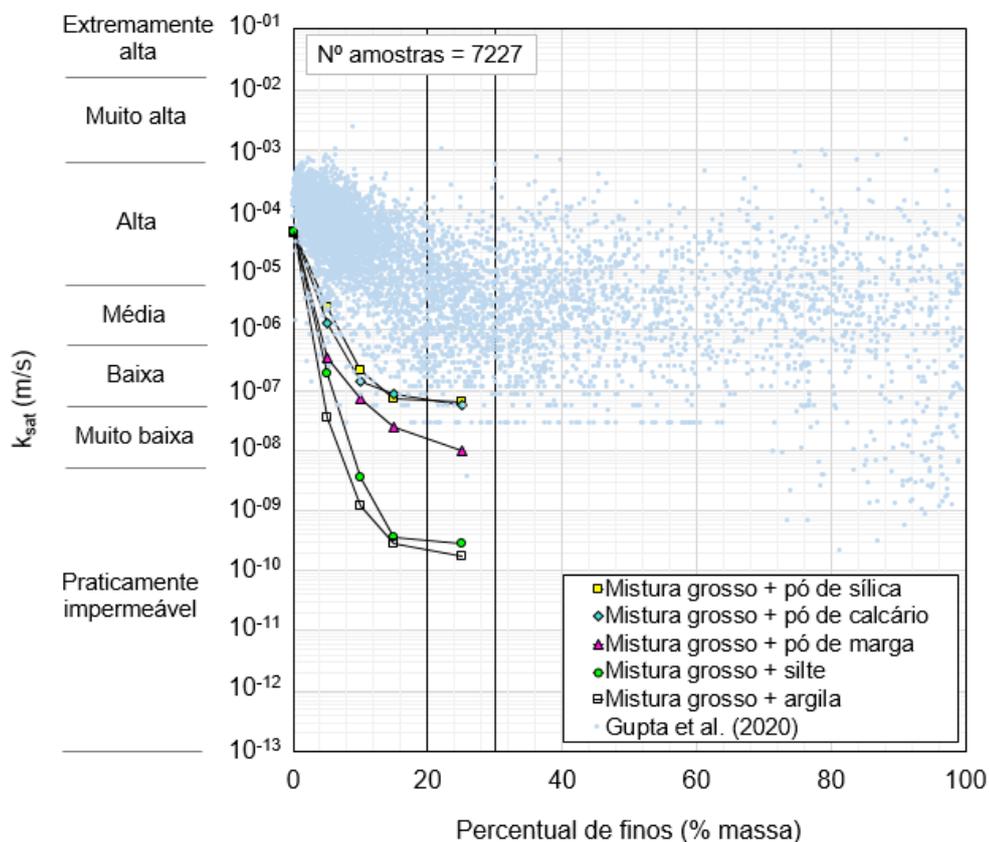


LEGENDA: C - Argila, SC - Argila Arenosa, CL - Franco Argilosa, SCL - Franco Argilo Arenosa, SL - Franco Arenoso, LS - Areia Franca, S - Areia, L - Franca, SiC - Argila Siltosa, SiCL - Franco Argilo Siltosa, SiL - Franco Siltosa, Si - Silte

Fonte: Autora (2022).

Analisando a distribuição das amostras no triângulo textural, apresentado na Figura 19, observa-se a presença de solos predominantemente de matriz granular. Mesmo assim, por se tratar de um número muito grande de dados, há amostras distribuídas ao longo de todo o triângulo textural. Os dados de  $k_{sat}$  vs teor de finos são apresentados no Gráfico 21, juntamente com as curvas dos resultados de Barber e Sawyer (1952).

Gráfico 21 - Variação de  $k_{sat}$  com o teor de finos para diferentes misturas controladas (Dados de: BARBER; SAWYER, 1952; GUPTA *et al.*, 2020)



Fonte: Autora (2022).

A classificação do conjunto de dados de Gupta *et al.* (2020) justifica o posicionamento da maioria das amostras acima da curva de misturas não plásticas de Barber e Sawyer (1952). Os valores de  $k_{sat}$  dos dados de Gupta *et al.* (2020) variam de  $2,15 \times 10^{-10}$  m/s a  $2,34 \times 10^{-3}$  m/s.

Apesar de ser um numeroso banco de dados, o que para análises de regressão pode ser atrativo, poucas informações são fornecidas para um estudo das amostras e compressão dos valores de  $k_{sat}$  e da relação deste parâmetro com outras propriedades dessas amostras. Apenas baseado no teor de finos, verifica-se uma redução mais acentuada de  $k_{sat}$  até o percentual de finos de, aproximadamente, 20%. A partir desse teor, não é possível identificar uma tendência de redução.

Quando analisadas as amostras que apresentaram  $k_{sat}$  da ordem de  $10^{-10}$  m/s comparadas com as amostras com maior teor de finos, pouco pode se concluir sobre os resultados apenas baseado na granulometria, conforme se apresenta na Tabela 40.

Tabela 40 - Permeabilidade de amostras menores  $k_{sat}$  e de amostras com maiores teores de finos (Dados de: GUPTA *et al.*, 2020)

Amostras com $k_{sat}$ na ordem de $10^{-10}$ m/s				
%Argila	%Silte	%Areia	%Finos	$k_{sat}$ (m/s)
69	26	6	94	$8,40 \times 10^{-10}$
48	25	26	74	$6,97 \times 10^{-10}$
58	19	23	77	$6,73 \times 10^{-10}$
60	31	9	91	$5,52 \times 10^{-10}$
74	13	13	87	$3,12 \times 10^{-10}$
43	39	19	81	$2,15 \times 10^{-10}$
Amostras com teor de finos maior que 99,5%				
%Argila	%Silte	%Areia	%Finos	$k_{sat}$ (m/s)
5,0	94,5	0,5	99,5	$7,50 \times 10^{-07}$
64,8	34,7	0,5	99,5	$3,64 \times 10^{-07}$
60,9	38,6	0,5	99,5	$1,03 \times 10^{-07}$
14,2	85,4	0,4	99,6	$1,17 \times 10^{-05}$
11,0	88,7	0,3	99,7	$6,85 \times 10^{-05}$
8,2	91,6	0,2	99,8	$8,83 \times 10^{-06}$

Fonte: Autora (2022).

A permeabilidade de solos com predominância de finos (ou com o comportamento hidráulico controlado pelos finos) não é caracterizada de forma adequada somente com a informação da fração granulométrica desses solos. É possível notar a existência de amostras com 64,8% de argila com  $k_{sat}$  de  $3,64 \times 10^{-07}$  m/s e amostras com 43% de argila com  $k_{sat}$  de  $2,15 \times 10^{-10}$  m/s. Caso informações complementares fossem fornecidas (como por exemplo a composição mineralógica dos solos finos), esses resultados poderiam ser melhor investigados e analisados.

A ausência de dados complementares não desqualifica o banco de dados, apenas não o viabiliza para um estudo mais específico sobre a permeabilidade saturada, pois, como discutido e apresentado, existem propriedades da fração fina que se relacionam e afetam os valores de  $k_{sat}$  e que são importantes para a previsão dessa propriedade hidráulica. Auxiliariam, inclusive, a explicar a variabilidade dos valores de  $k_{sat}$  das amostras reunidas por Gupta *et al.* (2020).

### Dados disponibilizados em Pucket *et al.* (1985)

Outro banco de dados levantado da literatura foi apresentado por Pucket *et al.* (1985). Os pesquisadores coletaram seis amostras para cada sete diferentes localidades no estado norte-americano do Alabama. Ao todo, 42 amostras de solos foram coletadas e informações sobre sua composição granulométrica (%argila, %silte, %areia e %areia\_fina),  $S_s$ , mineralogia (%caulinita, %gibsitita e %vermiculita),  $n$ ,  $\rho_d$ ,  $\theta_r$ ,  $\theta_s$  e parâmetros  $\alpha$  e  $n$  da equação de Van Genuchten.

Uma regressão não linear foi desenvolvida e uma equação resultante foi apresentada para estimar  $k_{sat}$  em função do teor de argila; a equação proposta pelos pesquisadores foi apresentada no Capítulo 2.

Os valores de  $k_{sat}$  das 42 amostras de Pucket *et al.* (1985) concentram-se, principalmente, nas faixas de baixa a alta permeabilidade, praticamente bem distribuídas entre esses intervalos, como mostra a Tabela 41.

Tabela 41 - Classificação relativa da permeabilidade das amostras de Pucket *et al.* (1985)

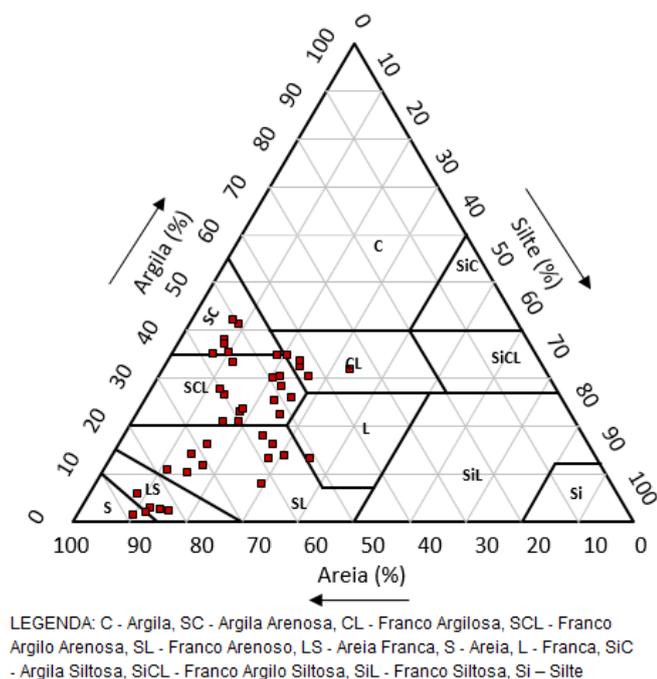
Classificação relativa da permeabilidade	Intervalo de referência (m/s)	Nº de amostras*
Extremamente alta	$k_{sat} \geq 10^{-2}$	-
Muito alta	$3,53 \times 10^{-4} \leq k_{sat} < 10^{-2}$	-
Alta	$3,53 \times 10^{-6} \leq k_{sat} < 3,53 \times 10^{-4}$	13
Média	$3,53 \times 10^{-7} \leq k_{sat} < 3,53 \times 10^{-6}$	11
Baixa	$3,53 \times 10^{-8} \leq k_{sat} < 3,53 \times 10^{-7}$	14
Muito baixa	$3,53 \times 10^{-9} \leq k_{sat} < 3,53 \times 10^{-8}$	2
Praticamente impermeável	$k_{sat} < 3,53 \times 10^{-9}$	2
<b>TOTAL DE AMOSTRAS</b>		<b>42</b>

\* Sem informação sobre a plasticidade das amostras.

Fonte: Autora (2022).

Apenas duas amostras apresentaram  $k_{sat}$  de solos praticamente impermeáveis. A Figura 20 apresenta as 42 amostras de Pucket *et al.* (1985) no triângulo textural.

Figura 20 – Triângulo textural – 42 amostras de Pucket *et al.* (1985)

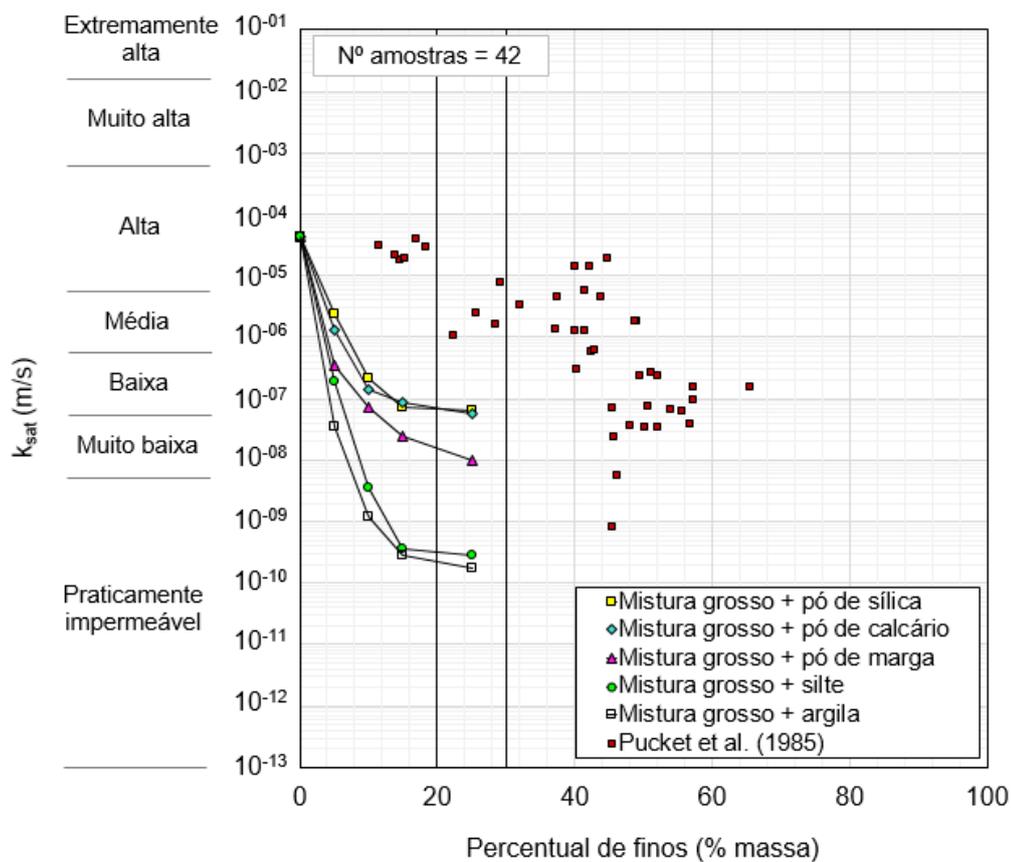


Fonte: Autora (2022).

O triângulo textural com a disposição das 42 amostras evidencia que os solos são, de fato, classificados como de matriz granular, segundo os critérios da *USDA* (1991). As amostras não puderam ser classificadas pelo SUCS por ausência de informações como limites de consistência e propriedades da distribuição granulométrica.

Ao todo, as 42 amostras são apresentadas e possuem  $k_{sat}$  que abrange o intervalo de  $8,33 \times 10^{-10}$  m/s a  $3,94 \times 10^{-5}$  m/s. A relação do teor de finos com  $k_{sat}$  foi avaliada no Gráfico 22. Novamente, os dados de  $k_{sat}$  vs teor de finos são apresentados, juntamente com as curvas dos resultados de Barber e Sawyer (1952). Por se tratar de amostras principalmente arenosas, a maior parte dos dados experimentais se localizou próxima às curvas de baixa plasticidade.

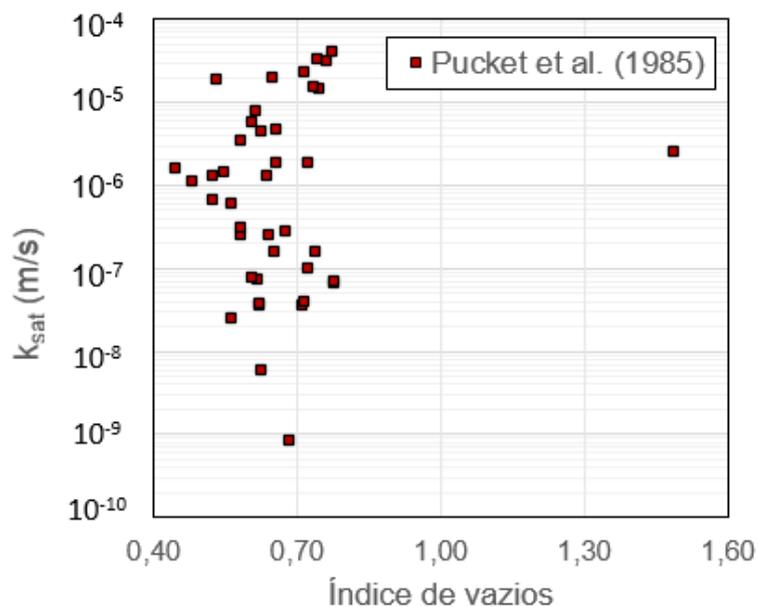
Gráfico 22 - Variação de  $k_{sat}$  com o teor de finos para diferentes solos naturais (Dados de: BARBER; SAWYER, 1952; PUCKET *et al.*, 1985)



Fonte: Autora (2022).

Observa-se a predominância de solos com teor de finos menor que 60%. No banco de dados, são disponibilizadas informações sobre índice de vazios e  $\rho_t$ . A relação  $k_{sat}$  vs índice de vazios foi investigada e apresentada no Gráfico 23.

Gráfico 23 - Variação de  $k_{sat}$  com índice de vazios para diferentes solos naturais (Dados de: PUCKET *et al.*, 1985)



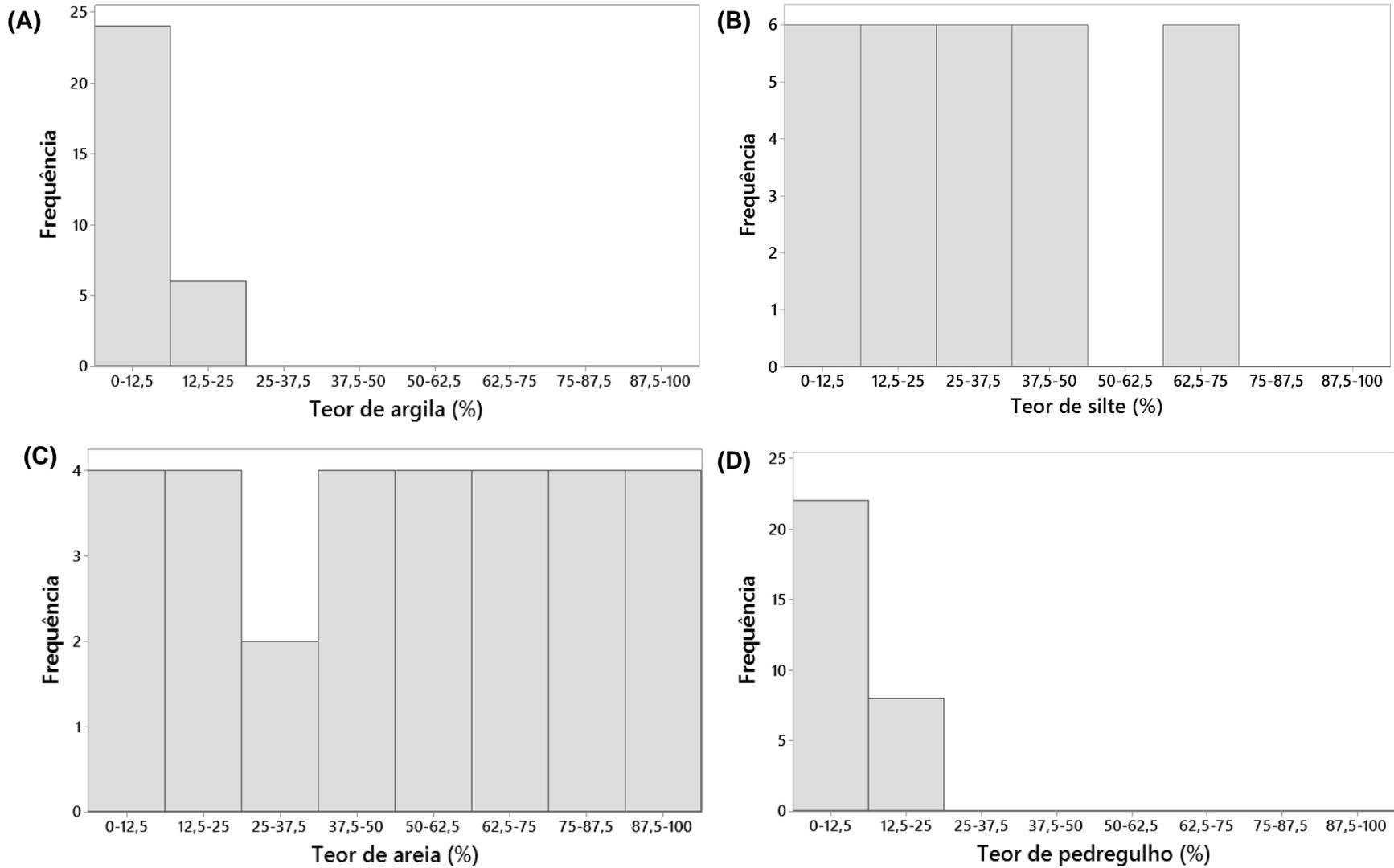
Fonte: Autora (2022).

Assim como nas amostras de Ottoni *et al.* (2018), o Gráfico 23 pouco esclarece sobre a relação de  $k_{sat}$  com o índice de vazios das amostras de Pucket *et al.* (1985), indicando que há outras propriedades interferindo nos valores de  $k_{sat}$ .

Pucket *et al.* (1985) apresentaram a composição da mineralogia das partículas de tamanho argila ( $< 2\mu\text{m}$ ) em g/kg de solo. A composição mineralógica das amostras não diferiram substancialmente e os três principais minerais presentes foram vermiculita hidrox-intercalada, caulinita e gibbsita. A disponibilidade dessas informações permitiu verificar que, em geral, os maiores valores de  $k_{sat}$  estão associados às amostras que apresentam maior quantidade de vermiculita hidrox-intercalada e/ou caulinita, com uma quantidade mínima de gibbsita. Já os menores valores de  $k_{sat}$  são associados às amostras que apresentam, proporcionalmente, as mesmas quantidades destes três minerais.

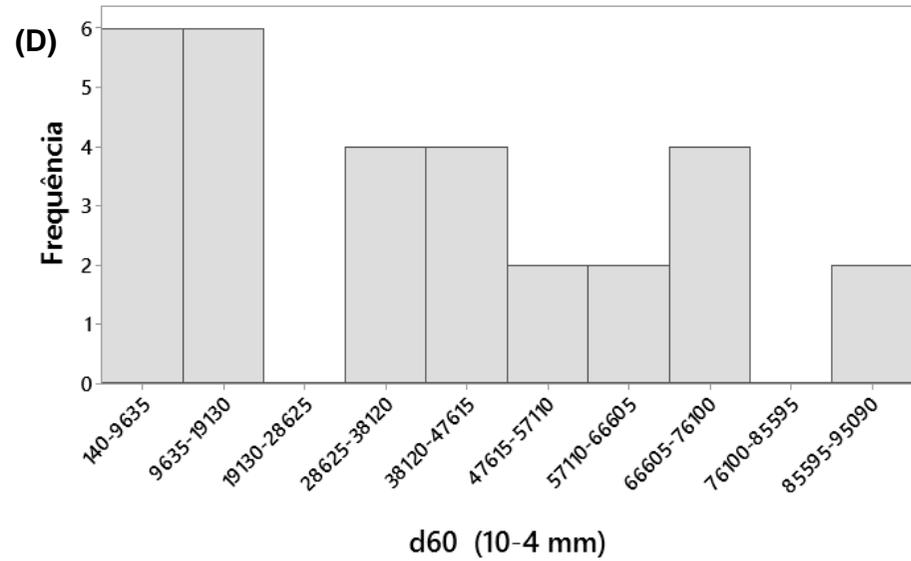
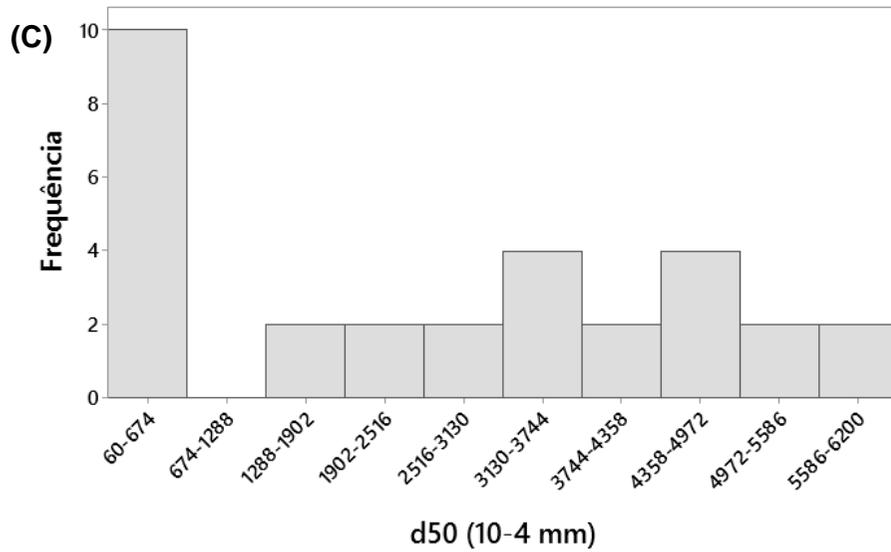
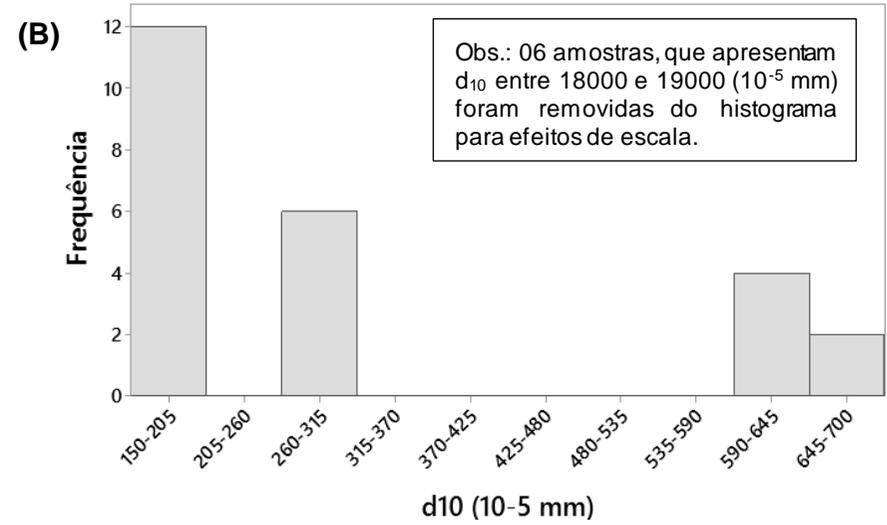
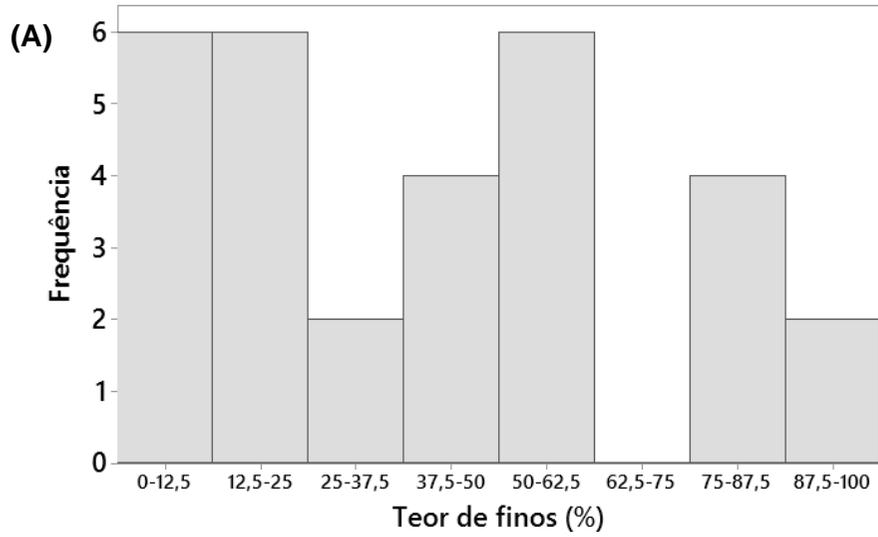
**APÊNDICE 2 – Histogramas dos dados experimentais (Dados de WANG; HUANG, 1984)**

Figura 21 –Histograma do (A) %argila, (B) %silte, (C) %areia e (D) %pedregulho das amostras não plásticas (NP) – 30 amostras



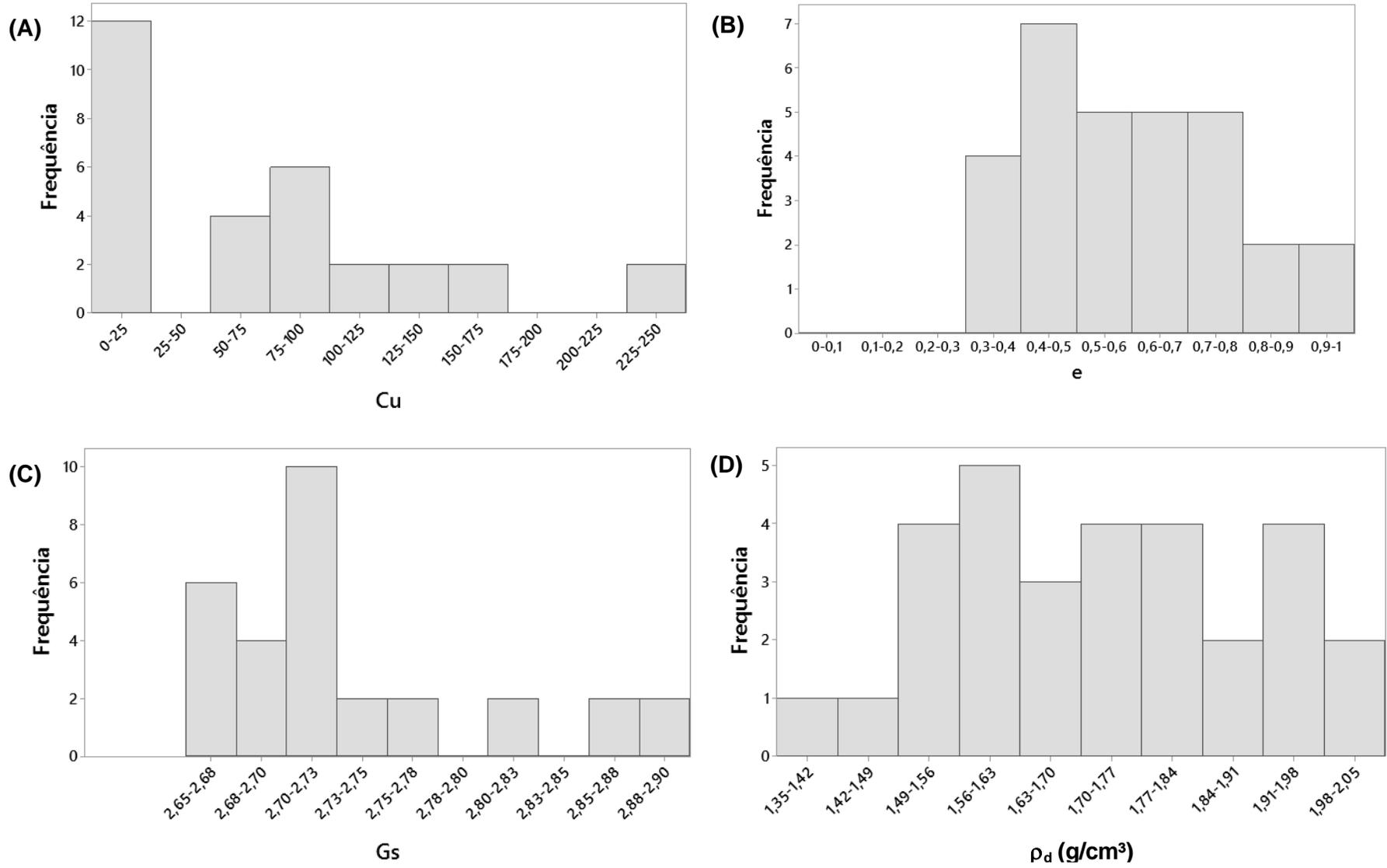
Fonte: Autora (2022).

Figura 22 –Histograma do (A) %finos, (B)  $d_{10}$ , (C)  $d_{50}$  e (D)  $d_{60}$  das amostras não plásticas (NP) – 30 amostras

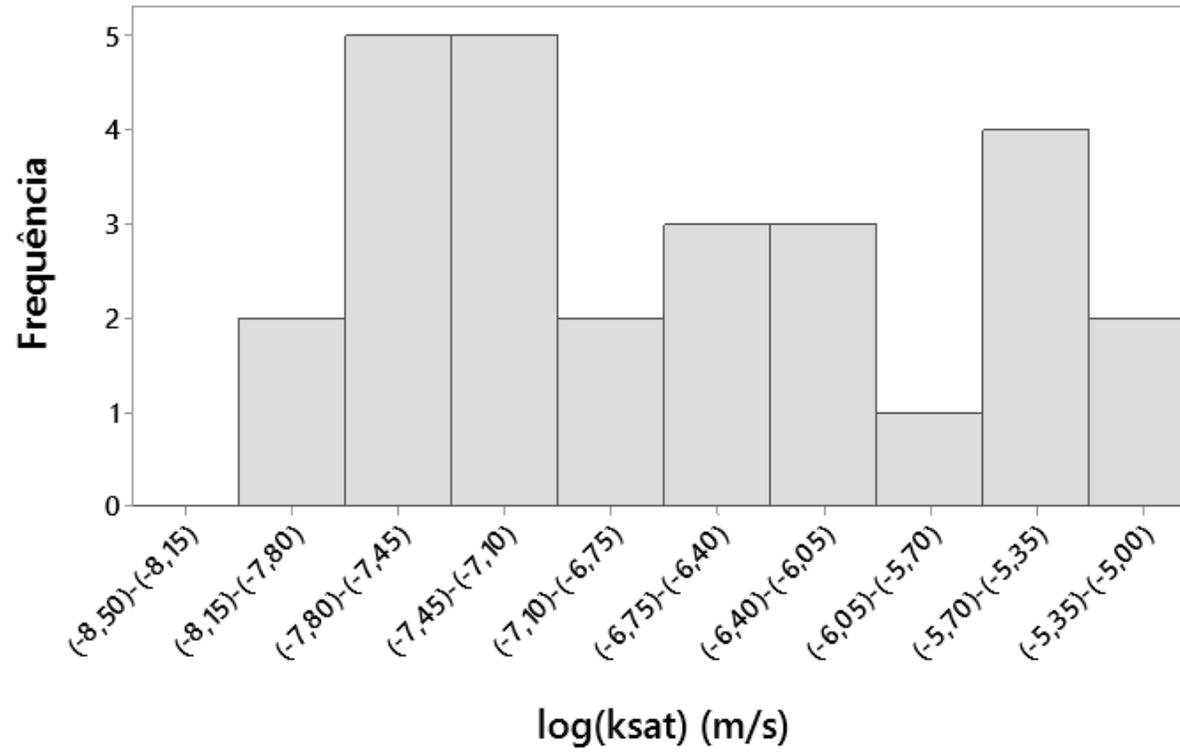


Fonte: Autora (2022).

Figura 23 –Histograma do (A)  $C_u$ , (B)  $e$ , (C)  $G_s$  e (D)  $\rho_d$  das amostras não plásticas (NP) – 30 amostras

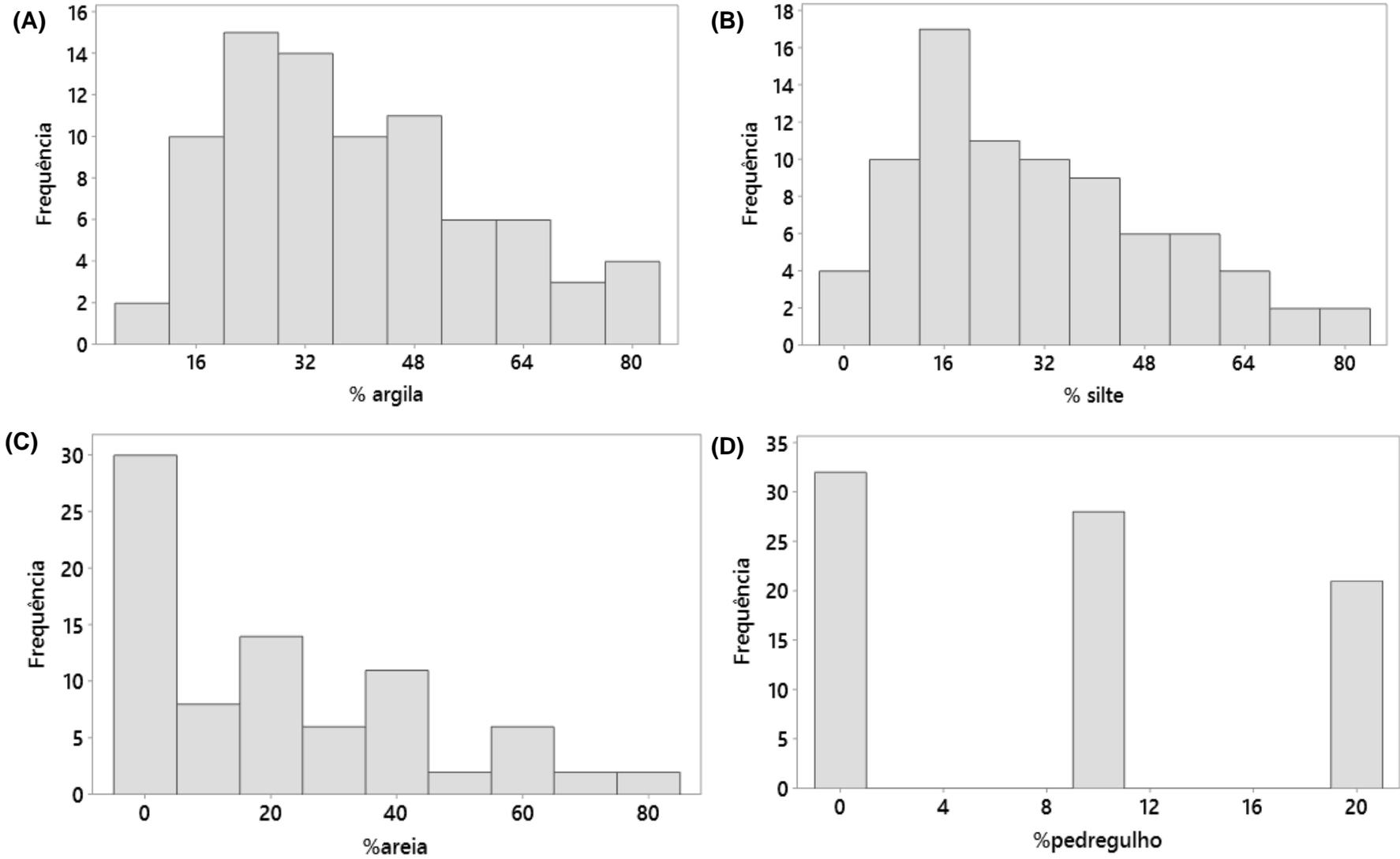


Fonte: Autora (2022).

Figura 24 –Histograma do  $\log(k_{sat})$  das amostras não plásticas (NP) – 30 amostras

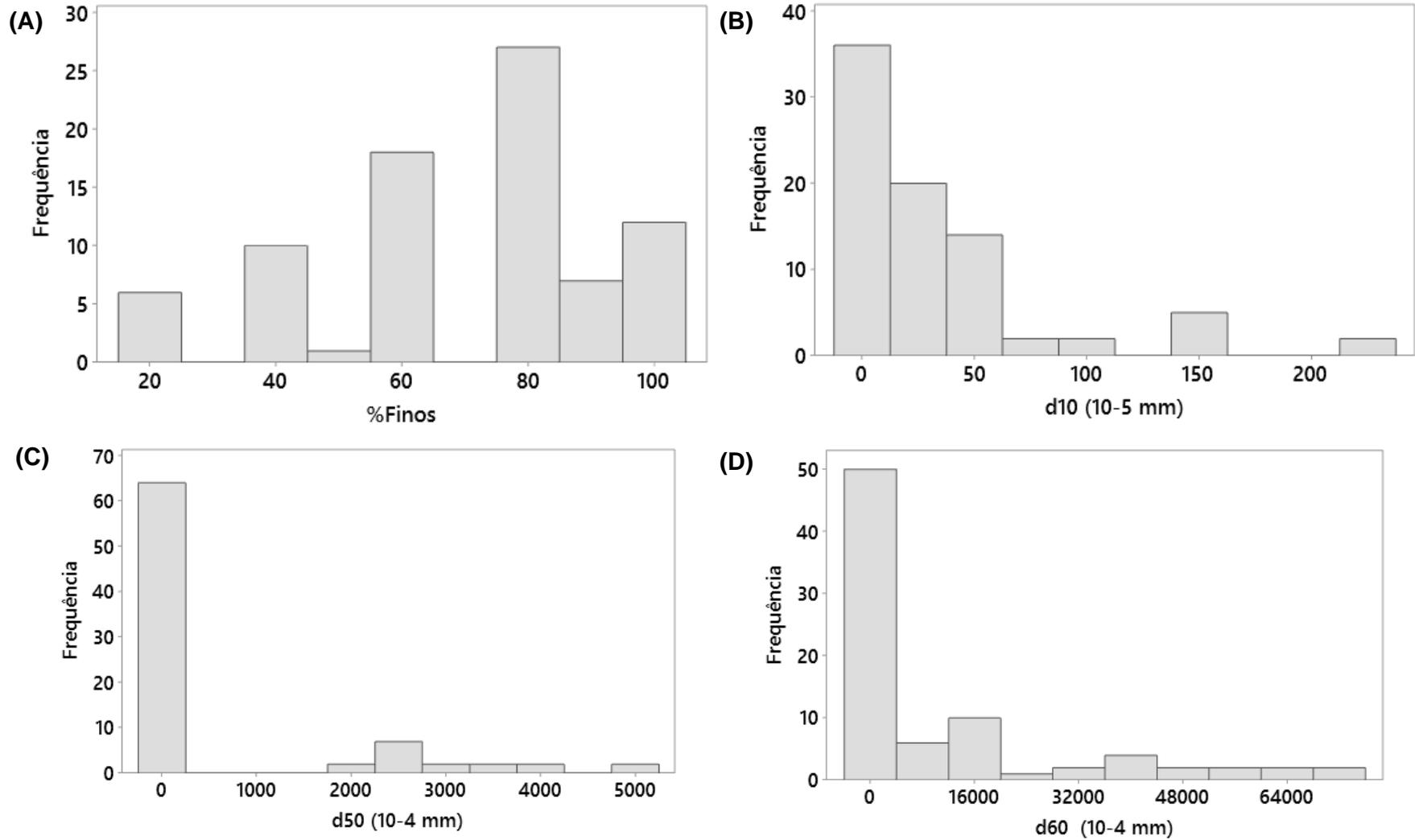
Fonte: Autora (2022).

Figura 25 –Histograma do (A) %argila, (B) %silte, (C) %areia e (D) %pedregulho das amostras plásticas – 81 amostras



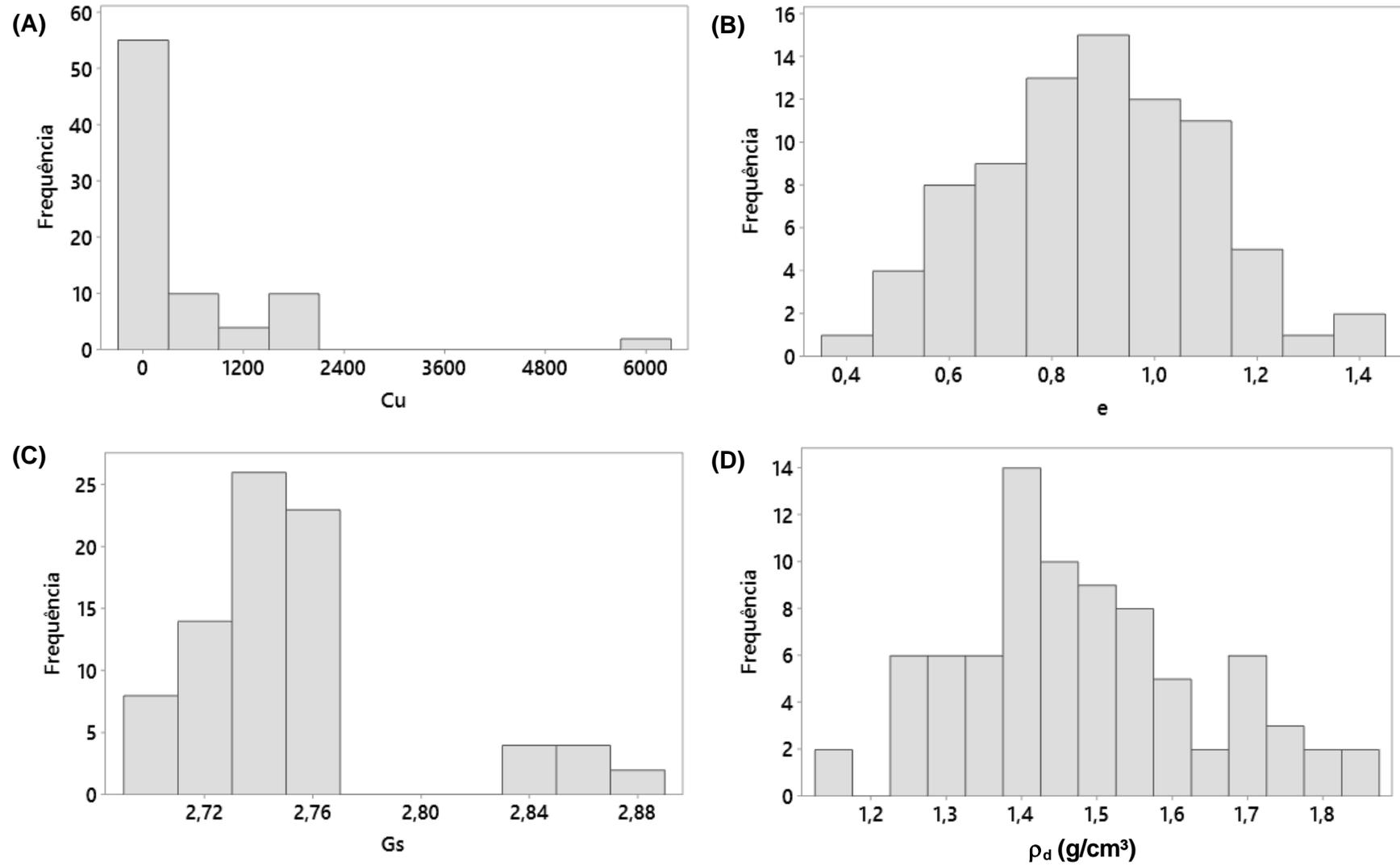
Fonte: Autora (2022).

Figura 26 – Histograma do (A) %finos, (B)  $d_{10}$ , (C)  $d_{50}$  e (D)  $d_{60}$  das amostras plásticas – 81 amostras



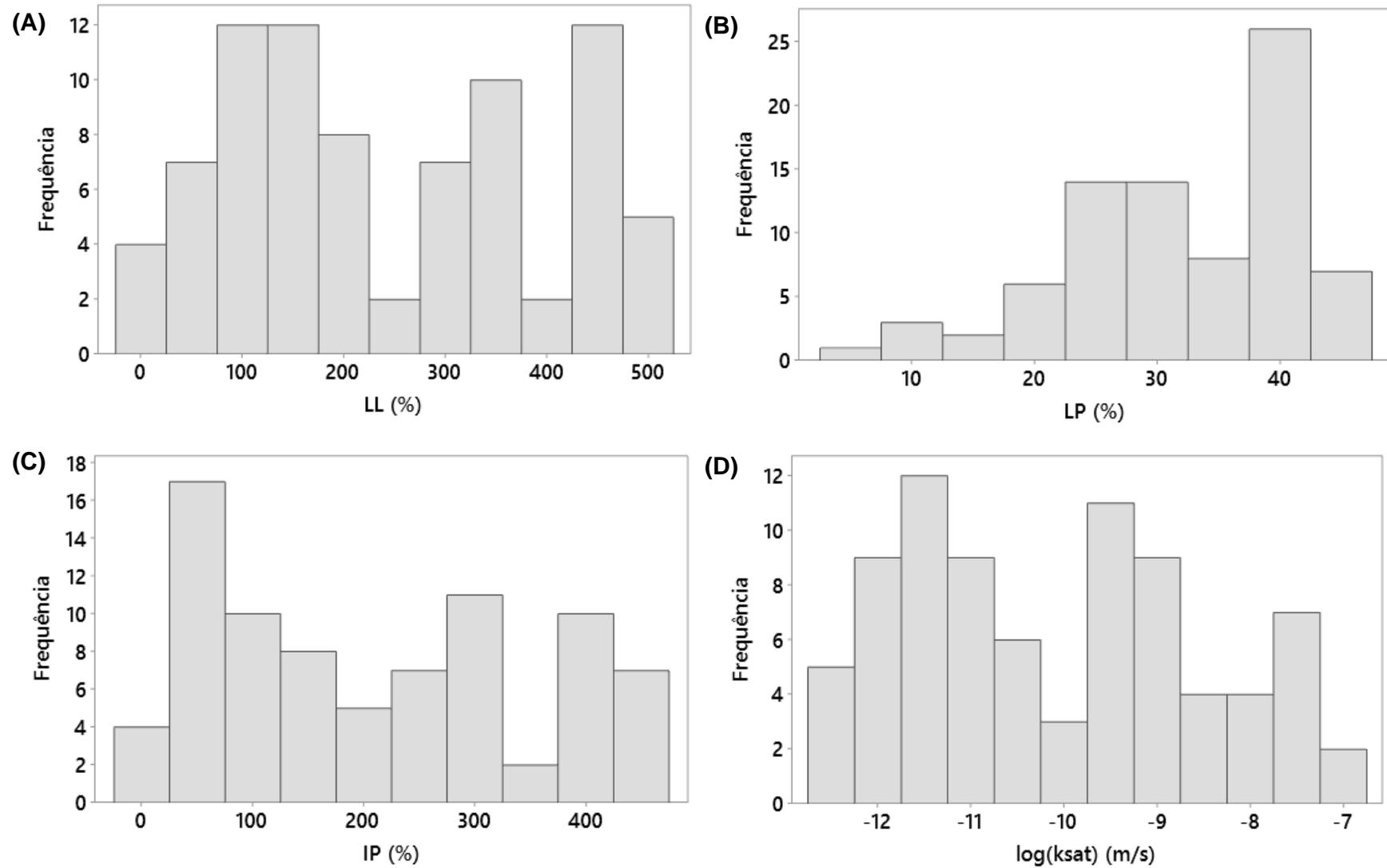
Fonte: Autora (2022).

Figura 27 – Histograma do (A)  $C_u$ , (B)  $e$ , (C)  $G_s$  e (D)  $\rho_d$  das amostras plásticas – 45 amostras



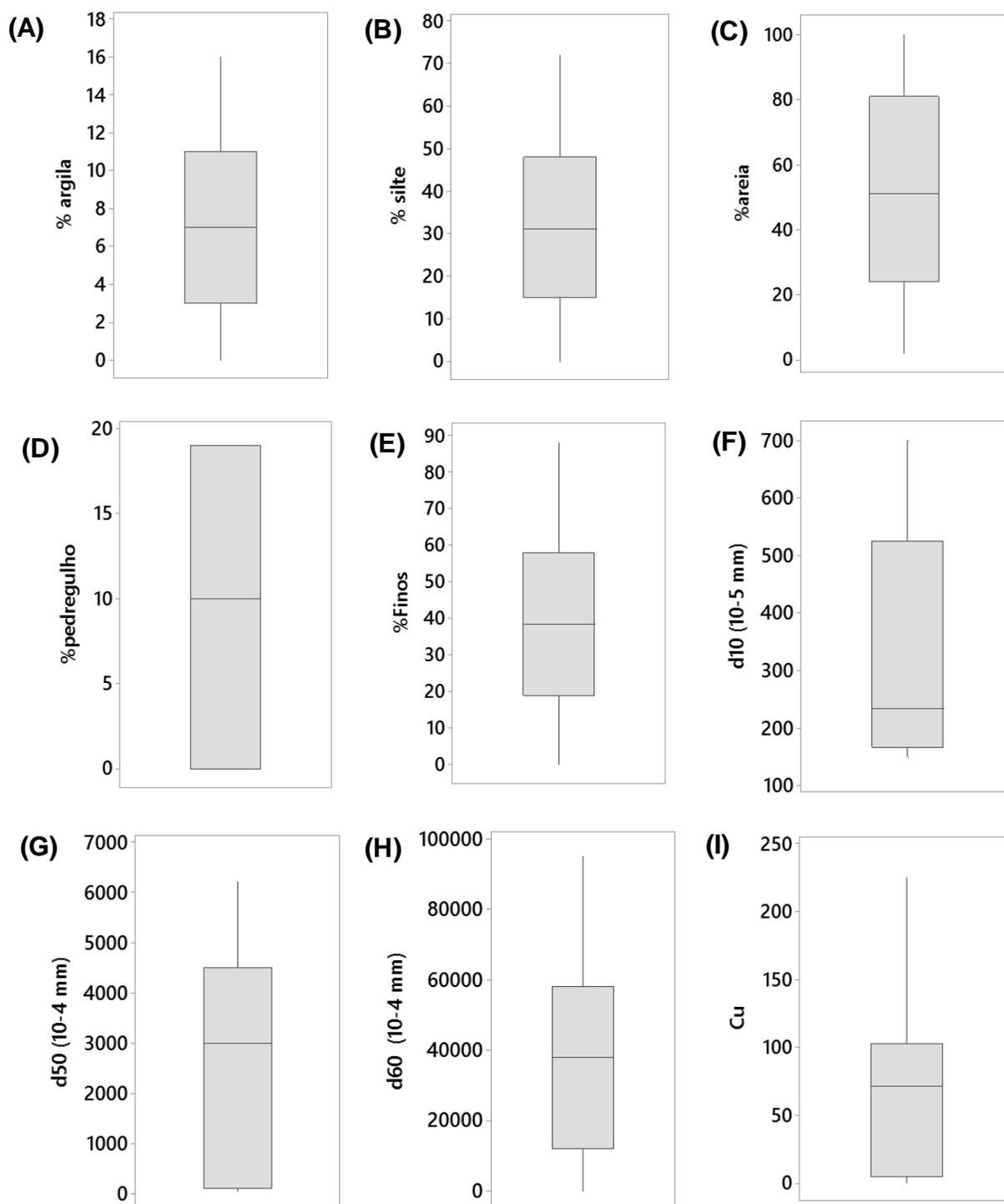
Fonte: Autora (2022).

Figura 28 – Histograma do (A) LL, (B) LP, (C) IP e (D)  $\log(k_{sat})$  das amostras plásticas – 81 amostras



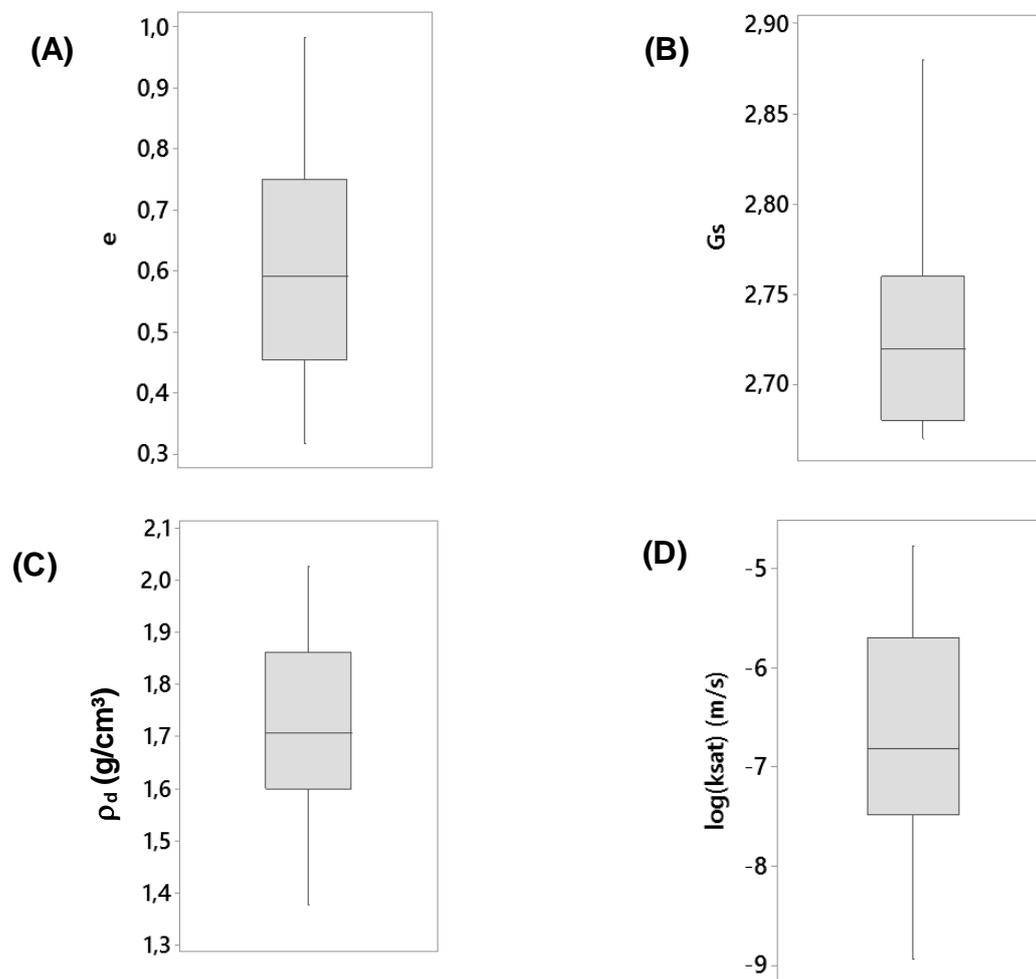
Fonte: Autora (2022).

Figura 29 – Diagramas de caixa do (A) %argila, (B) %silte, (C) %areia, (D) %pedregulho, (E) %finos, (F)  $d_{10}$ , (G)  $d_{50}$ , (H)  $d_{60}$  e (I)  $C_u$  das amostras não plásticas (NP) – 30 amostras



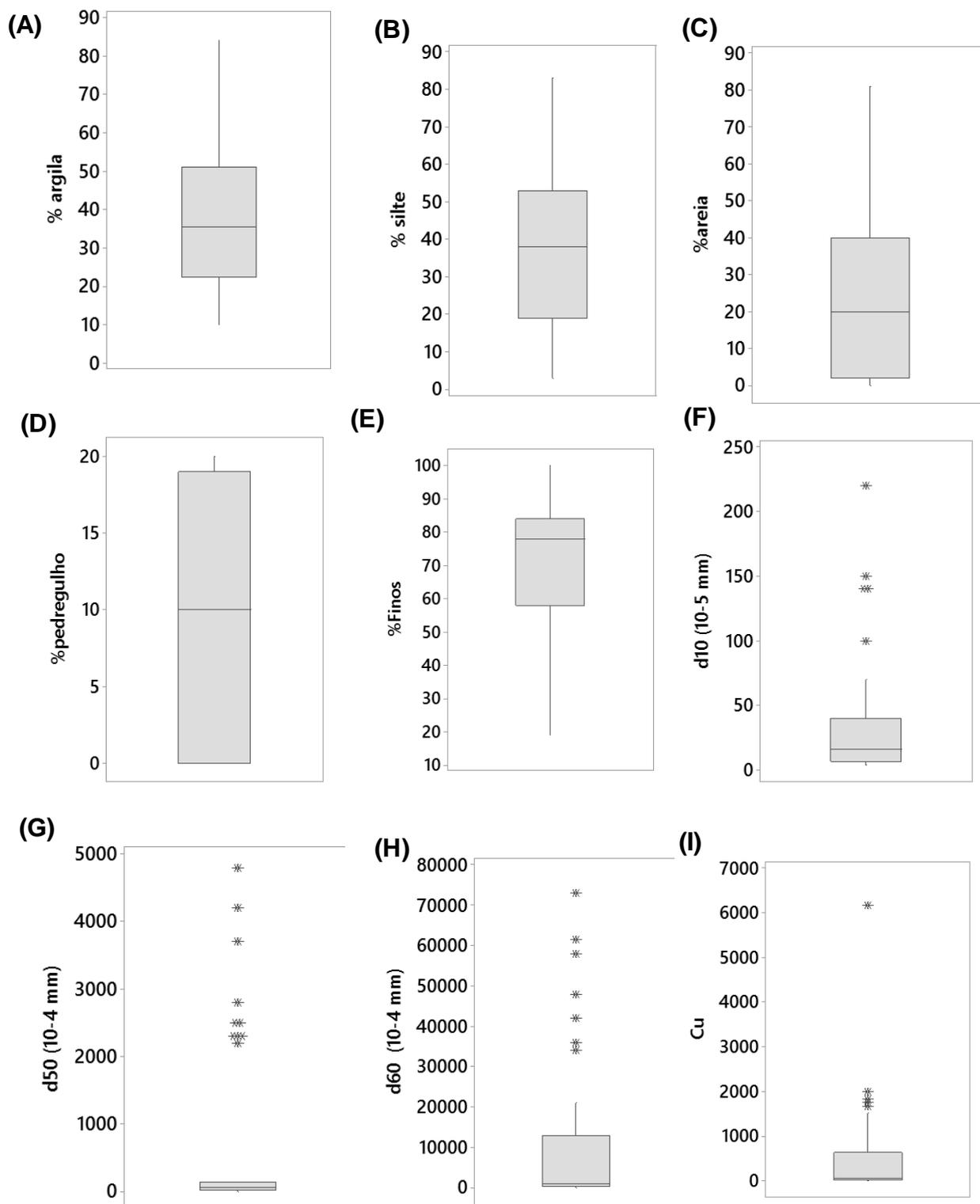
Fonte: Autora (2022).

Figura 30 – Diagramas de caixa do (A)  $e$ , (B)  $G_s$ , (C)  $\rho_d$ , (D)  $\log(k_{sat})$  das amostras não plásticas (NP) – 30 amostras



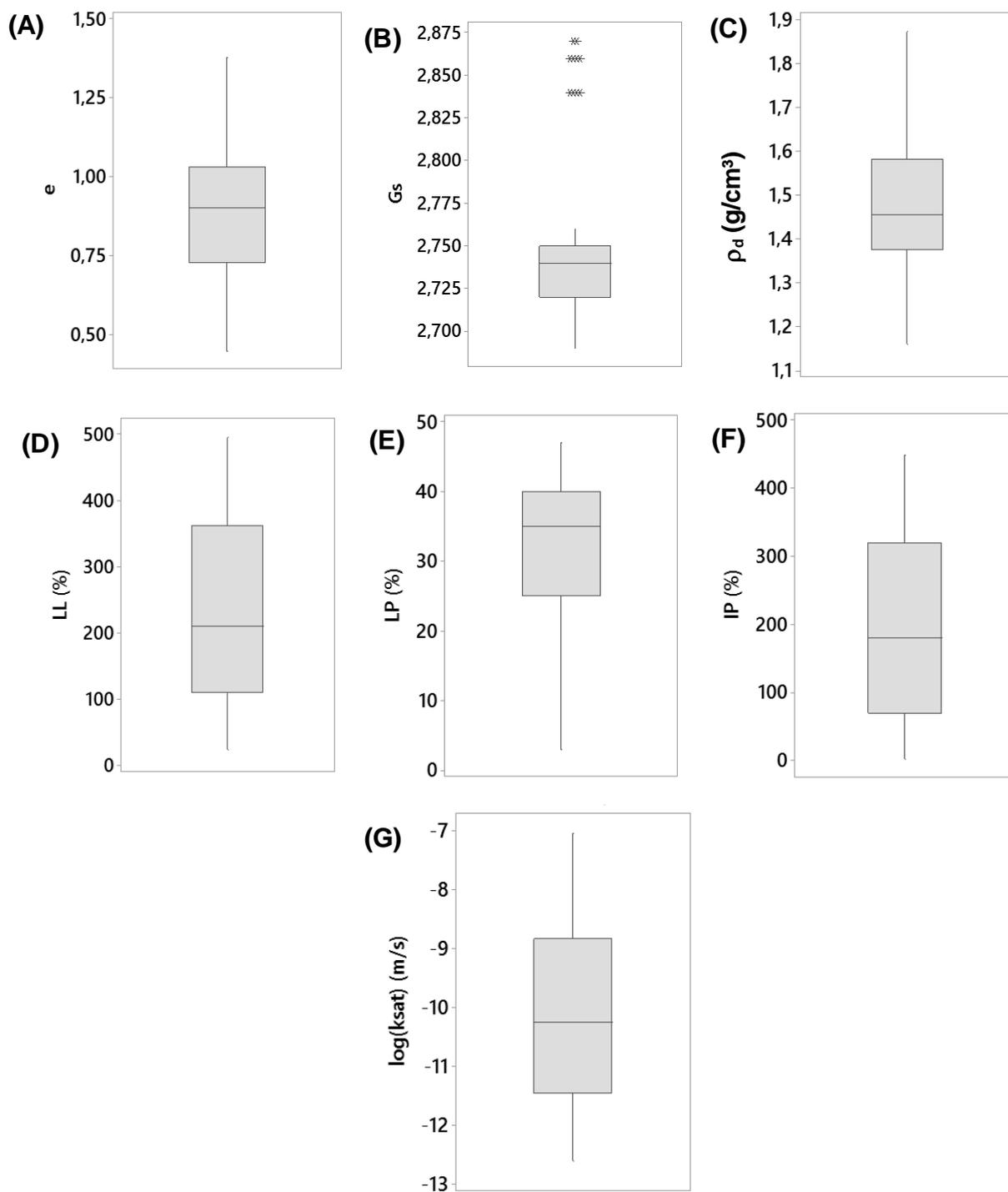
Fonte: Autora (2022).

Figura 31 – Diagramas de caixa do (A) %argila, (B) %silte, (C) %areia, (D) %pedregulho, (E) %finos, (F)  $d_{10}$ , (G)  $d_{50}$ , (H)  $d_{60}$  e (I)  $C_u$  das amostras plásticas – 81 amostras



Fonte: Autora (2022).

Figura 32 – Diagramas de caixa do (A)  $e$ , (B)  $G_s$ , (C)  $\rho_d$ , (D) LL, (E) LP, (F) IP e (G)  $\log(k_{sat})$  das amostras plásticas – 45 amostras



Fonte: Autora (2022).

### APÊNDICE 3 – Código em R para testes de hipóteses

#### #Leitura dos dados

```
read.delim2("NOMEARQUIVO.txt",header=FALSE) #importar tabela que deve estar
salva em Meus Documentos. Importante: usar vírgula como separador de decimal
name<-read.delim2("NOMEARQUIVO.txt",header=FALSE) #salvar tabela
```

#### #Atribuição das variáveis

```
x1<-name[,1] #atribui LL variável independente x1
x2<-name[,2] #atribui Finos variável independente x2
y<-name[,3] #atribui logKsat medido variável dependente y1
```

#### #Regressão Linear (ou não linear) Simples ou Múltipla

```
regressao<-lm(y ~ x1 + x2) #regressão linear múltipla feita pelo próprio R
regressao #mostra os valores previstos dos coeficientes pela equação acima
```

#### #Plotagem do gráfico de dispersão

```
plot(regressao,y,xlab="logKsat previsto",ylab="logKsat medido",main="RN1")
#plotagem gráfico de dispersão
abline(lm(y~regressao), col="blue") #plotagem da linha 1:1 no gráfico de dispersão
```

#### #Intervalo de confiança

```
confint(regressao) #intervalo de confiança
```

#### #ANOVA

```
anova(regressao)#tabela ANOVA
regressao_s=summary(regressao)
regressao_s$adj.r.squared #calcula o R2 ajustado
```

#### #Teste de Hipoteses

```
summary(regressao) #para verificar valores-p testes t e F (São as mesmas
observações da tabela ANOVA, pode usar aquele comando ou esse)
```

#### #Análise de resíduos - testes homocedasticidade

```
plot(fitted(regressao),residuals(regressao),xlab="logKsat previsto",ylab="resíduos")
abline(h=0)
```

#### #Análise de resíduos - testes normalidade

```
qqnorm(residuals(regressao),ylab="resíduos",xlab="quantis teoricos",main="")
#plotagem de gráfico de probabilidade normal de resíduos
qqline(residuals(regressao))
hist(residuals(regressao),freq=T,xlab="resíduos",ylab="frequencia") #plotagem de
histograma
```

```
install.packages("nortest") #instala pacote de testes de normalidade  
library(nortest)
```

```
shapiro.test(residuals(regressao)) #teste de Shapiro-Wilk (aplicável quando  
8<=n<=30)  
ks.test(residuals(regressao),pnorm) #teste de Kolmogorov-Sminorv  
lillie.test(residuals(regressao)) #teste de Lilliefors  
ad.test(residuals(regressao)) #teste de Anderson-Darling  
sf.test(residuals(regressao)) #teste de Shapiro-Francis
```

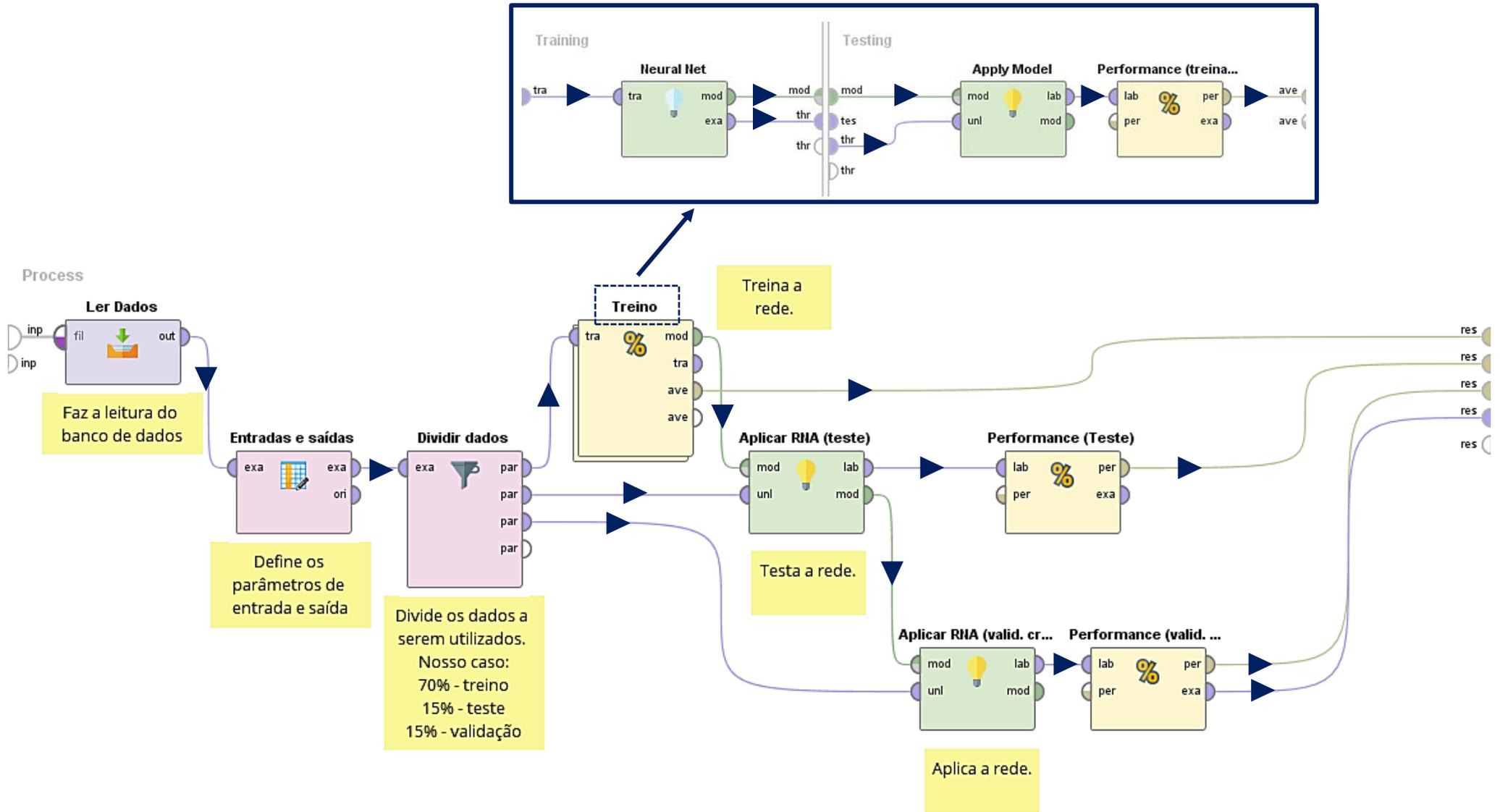
### #Análise de resíduos - teste de independencia

```
library(lmtest) #carregar pacote  
dwtest(regressao) #Estatística de Durbin-Watson
```

### #Teste de multicolinearidade VIF

```
install.packages("caTools") #Para regressões  
library(caTools)  
install.packages('car')  
library(car)  
vif(regressao) #calcula o valor VIF de cada variável  
vif_values <- vif(regressao) #cria vetor com os valores de VIF  
barplot(vif_values, main = "Valores de VIF", horiz = TRUE, col = "steelblue") #cria uma  
barra horizontal no gráfico de cada valor de VIF  
abline(v = 5, lwd = 3, lty = 2) #adiciona uma linha vertical de 5 em 5
```

### APÊNDICE 4 – Diagrama das etapas de treinamento de RNA no RapidMiner



**APÊNDICE 5 – Banco de dados experimentais – 111 dados (Dados de: WANG; HUANG, 1984)**

AMOSTRA	%Argila	%Silte	%Areia	%Pedreg	%Finos	$k_{sat}$ (m/s)	$\log(k_{sat})$ (m/s)	LL (%)	LP (%)	IP (%)	$d_{10}$ ( $10^{-5}$ mm)	$d_{60}$ ( $10^{-4}$ mm)	Cu	e	$\rho_d$ (g/cm <sup>3</sup> )	Gs
1	0	0	100	0	0	5.20E-07	-6.284	0	0	0	19000	56050	2.95	0.653	1.62	2.67
2	0	0	90	10	0	1.50E-05	-4.824	0	0	0	18000	75060	4.17	0.49	1.79	2.67
3	0	0	81	19	0	1.70E-05	-4.770	0	0	0	18500	95090	5.14	0.39	1.92	2.67
4	0	0	100	0	0	4.80E-08	-7.319	0	0	0	19000	56050	2.95	0.566	1.70	2.67
5	0	0	90	10	0	3.20E-06	-5.495	0	0	0	18000	75060	4.17	0.412	1.89	2.67
6	0	0	81	19	0	2.00E-06	-5.699	0	0	0	18500	95090	5.14	0.317	2.03	2.67
7	3	16	81	0	19	5.00E-06	-5.301	0	0	0	630	45000.9	71.43	0.568	1.71	2.68
8	16	3	81	0	19	1.30E-11	-10.886	136	25	111	46	48000.08	1043.48	0.594	1.69	2.69
9	3	16	71	10	19	8.00E-07	-6.097	0	0	0	600	58002	96.67	0.468	1.85	2.72
10	10	9	71	10	19	1.00E-08	-8.000	75	15	60	220	58000.8	263.64	0.561	1.72	2.69
11	4	15	62	19	19	4.00E-06	-5.398	0	0	0	700	72002	102.86	0.408	1.91	2.69
12	3	16	81	0	19	8.40E-06	-5.076	0	0	0	630	45000.9	71.43	0.497	1.79	2.68
13	16	3	81	0	19	3.00E-12	-11.523	136	25	111	46	48000.08	1043.48	0.51	1.78	2.69
14	3	16	71	10	19	1.00E-07	-7.000	0	0	0	600	58002	96.67	0.39	1.96	2.72
15	10	9	71	10	19	2.50E-09	-8.602	75	15	60	220	58000.8	263.64	0.479	1.82	2.69
16	4	15	62	19	19	1.70E-06	-5.770	0	0	0	700	72002	102.86	0.334	2.02	2.69
17	17	3	61	19	20	1.15E-10	-9.939	170	30	140	40	73000	1825	0.56	1.73	2.7
18	17	3	61	19	20	6.00E-11	-10.222	170	30	140	40	73000	1825	0.478	1.83	2.7
19	6	31	63	0	37	1.20E-07	-6.921	0	0	0	160	36000	225	0.717	1.57	2.7
20	6	31	63	0	37	2.70E-08	-7.569	0	0	0	160	36000	225	0.627	1.66	2.7
21	20	18	62	0	38	3.50E-10	-9.456	132	23	109	40	34000	850	0.817	1.49	2.7
22	20	18	62	0	38	1.80E-10	-9.745	132	23	109	40	34000	850	0.722	1.57	2.7
23	7.2	31.2	42.4	19	38.4	2.00E-06	-5.699	0	0	0	280	44998.8	160.71	0.492	1.82	2.71
24	7.2	31.2	42.4	19	38.4	5.40E-07	-6.268	0	0	0	280	44998.8	160.71	0.413	1.92	2.71
25	7	32	51	10	39	2.00E-07	-6.699	0	0	0	270	37999.8	140.74	0.594	1.73	2.76
26	27	12	51	10	39	5.00E-10	-9.301	210	30	180	18	36000	2000	0.7	1.59	2.71
27	7	32	51	10	39	6.00E-08	-7.222	0	0	0	270	37999.8	140.74	0.511	1.83	2.76
28	27	12	51	10	39	4.00E-11	-10.398	210	30	180	18	36000	2000	0.611	1.68	2.71
29	33	6	61	0	39	1.00E-12	-12.000	277	29	248	11	3200.01	290.91	0.678	1.62	2.71
30	33	6	61	0	39	8.10E-13	-12.092	277	29	248	11	3200.01	290.91	0.589	1.71	2.71

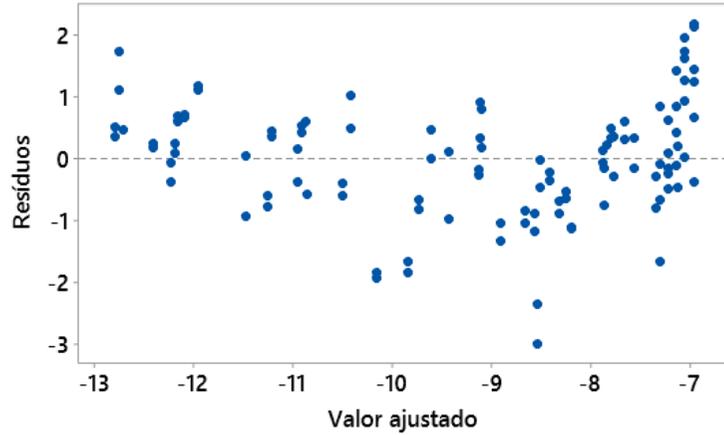
AMOSTRA	%Argila	%Silte	%Areia	%Pedreg	%Finos	$k_{sat}$ (m/s)	$\log(k_{sat})$ (m/s)	LL (%)	LP (%)	IP (%)	$d_{10}$ ( $10^{-5}$ mm)	$d_{60}$ ( $10^{-4}$ mm)	Cu	e	$\rho_d$ (g/cm <sup>3</sup> )	Gs
31	21	19	41	19	40	3.20E-10	-9.495	140	25	115	35	61600	1760	0.527	1.77	2.71
32	21	19	41	19	40	2.00E-10	-9.699	140	25	115	35	61600	1760	0.447	1.87	2.71
33	36	6	39	19	42	3.70E-12	-11.432	340	32	308	6.8	41999.996	6176.47	0.676	1.62	2.72
34	36	6	39	19	42	3.70E-12	-11.432	340	32	308	6.8	41999.996	6176.47	0.588	1.71	2.72
35	23	26	41	10	49	2.50E-08	-7.602	60	10	50	140	21000	150	0.575	1.75	2.75
36	10	47	43	0	57	2.70E-07	-6.569	0	0	0	200	16000	80	0.762	1.54	2.72
37	10	47	43	0	57	4.40E-08	-7.357	0	0	0	200	16000	80	0.654	1.64	2.72
38	24	34	42	0	58	1.70E-09	-8.770	94	25	69	30	12500.1	416.67	0.912	1.43	2.73
39	10	48	32	10	58	7.50E-08	-7.125	0	0	0	190	12000.4	63.16	0.747	1.61	2.81
40	30	28	32	10	58	5.20E-10	-9.284	175	35	140	23	13000.06	565.22	0.761	1.55	2.73
41	11	47	23	19	58	3.50E-08	-7.456	0	0	0	190	15000.5	78.95	0.676	1.62	2.72
42	24	34	22	19	58	2.30E-09	-8.638	110	25	85	44	12999.8	295.45	0.744	1.57	2.73
43	24	34	42	0	58	1.30E-09	-8.886	94	26	68	30	12500.1	416.67	0.811	1.51	2.73
44	10	48	32	10	58	2.00E-08	-7.699	0	0	0	190	12000.4	63.16	0.655	1.70	2.81
45	30	28	32	10	58	4.20E-10	-9.377	175	35	140	23	13000.06	565.22	0.688	1.62	2.73
46	11	47	23	19	58	2.00E-08	-7.699	0	0	0	190	15000.5	78.95	0.588	1.71	2.72
47	24	34	22	19	58	1.70E-09	-8.770	110	25	85	44	12999.8	295.45	0.653	1.65	2.73
48	37	22	41	0	59	3.10E-12	-11.509	240	35	205	12	8499.96	708.33	0.904	1.43	2.72
49	17	42	31	10	59	4.00E-08	-7.398	50	10	40	100	12800	128	0.733	1.57	2.72
50	37	22	41	0	59	2.15E-12	-11.668	240	3	237	12	8499.96	708.33	0.804	1.51	2.72
51	17	42	31	10	59	9.00E-09	-8.046	50	10	40	100	12800	128	0.595	1.71	2.72
52	38	22	21	19	60	1.30E-11	-10.886	300	40	260	11	13000.02	1181.82	0.763	1.54	2.72
53	38	22	21	19	60	8.00E-12	-11.097	300	40	260	11	13000.02	1181.82	0.67	1.63	2.72
54	50	10	40	0	60	3.80E-12	-11.420	389	32	357	5.6	8500.016	1517.86	0.951	1.40	2.73
55	44	16	30	10	60	1.60E-11	-10.796	342	32	310	6	10000.02	1666.67	0.902	1.44	2.74
56	50	10	40	0	60	4.00E-13	-12.398	389	32	357	5.6	8500.016	1517.86	0.849	1.48	2.73
57	44	16	30	10	60	4.80E-12	-11.319	342	32	310	6	10000.02	1666.67	0.802	1.52	2.74
58	51	10	20	19	61	1.15E-12	-11.939	455	40	415	6.5	3400.02	523.08	0.881	1.46	2.74
59	51	10	20	19	61	8.10E-13	-12.092	455	40	415	6.5	3400.02	523.08	0.782	1.54	2.74
60	13	63	24	0	76	1.10E-08	-7.959	0	0	0	300	141	0.47	0.983	1.38	2.73
61	13	63	24	0	76	1.15E-09	-8.939	0	0	0	300	141	0.47	0.793	1.52	2.73
62	28	50	22	0	78	5.10E-10	-9.292	81	27	54	25	1100	44	0.969	1.39	2.74
63	40	38	22	0	78	4.10E-11	-10.387	222	38	184	15	799.95	53.33	1.064	1.33	2.74

AMOSTRA	%Argila	%Silte	%Areia	%Pedreg	%Finos	$k_{sat}$ (m/s)	$\log(k_{sat})$ (m/s)	LL (%)	LP (%)	IP (%)	$d_{10}$ ( $10^{-5}$ mm)	$d_{60}$ ( $10^{-4}$ mm)	Cu	e	$\rho_d$ (g/cm <sup>3</sup> )	Gs
64	14	64	12	10	78	3.50E-07	-6.456	0	0	0	160	950.4	5.94	0.862	1.54	2.86
65	21	57	12	10	78	5.00E-08	-7.301	45	20	25	70	887.6	12.68	0.936	1.47	2.84
66	34	44	12	10	78	5.20E-09	-8.284	165	40	125	26	999.96	38.46	0.9	1.44	2.74
67	14	64	2	20	78	6.00E-08	-7.222	24	22	2	150	949.5	6.33	0.888	1.51	2.86
68	28	50	3	19	78	3.00E-09	-8.523	110	40	70	60	109.8	1.83	0.881	1.46	2.74
69	28	50	22	0	78	5.00E-10	-9.301	81	27	54	25	1100	44	0.865	1.47	2.74
70	40	38	22	0	78	2.90E-11	-10.538	222	38	184	15	799.95	53.33	0.956	1.40	2.74
71	14	64	12	10	78	4.20E-08	-7.377	0	0	0	160	950.4	5.94	0.764	1.62	2.86
72	21	57	12	10	78	3.60E-08	-7.444	45	20	25	70	887.6	12.68	0.834	1.55	2.84
73	34	44	12	10	78	1.20E-09	-8.921	165	40	125	26	999.96	38.46	0.8	1.52	2.74
74	14	64	2	20	78	2.00E-08	-7.699	24	22	2	150	949.5	6.33	0.789	1.60	2.86
75	28	50	3	19	78	1.10E-09	-8.959	110	40	70	60	109.8	1.83	0.782	1.54	2.74
76	41	38	2	19	79	4.20E-10	-9.377	285	40	245	16	800	50	0.945	1.41	2.75
77	41	38	2	19	79	1.20E-10	-9.921	285	40	245	16	800	50	0.843	1.49	2.75
78	54	25	21	0	79	1.40E-12	-11.854	362	42	320	7	380.03	54.29	1.099	1.31	2.74
79	47	32	11	10	79	4.30E-11	-10.367	330	40	290	8	600	75	1.032	1.35	2.74
80	54	25	21	0	79	9.10E-13	-12.041	362	42	320	7	380.03	54.29	0.989	1.38	2.74
81	47	32	11	10	79	3.40E-11	-10.469	330	40	290	8	600	75	0.925	1.42	2.74
82	61	19	10	10	80	3.40E-12	-11.469	445	40	405	5.5	199.98	36.36	1.065	1.33	2.74
83	55	25	1	19	80	1.75E-11	-10.757	425	40	385	7.5	339.975	45.33	0.987	1.38	2.75
84	61	19	10	10	80	2.80E-12	-11.553	445	40	405	5.5	199.98	36.36	0.956	1.40	2.74
85	55	25	1	19	80	1.50E-11	-10.824	425	40	385	7.5	339.975	45.33	0.882	1.46	2.75
86	68	13	0	19	81	5.80E-13	-12.237	495	46	449	4.4	125.004	28.41	0.942	1.42	2.75
87	22	62	2	10	84	1.90E-08	-7.721	50	25	25	60	720	12	1.013	1.42	2.86
88	22	62	2	10	84	1.20E-08	-7.921	50	25	25	60	720	12	0.907	1.50	2.86
89	67	13	20	0	80	7.00E-13	-12.155	467	39	428	5	120	24	1.204	1.25	2.76
90	67	13	20	0	80	6.00E-13	-12.222	467	39	428	5	120	24	1.088	1.32	2.76
91	16	72	2	10	88	2.40E-08	-7.620	0	0	0	150	769.5	5.13	0.977	1.46	2.88
92	16	72	2	10	88	7.50E-09	-8.125	0	0	0	150	769.5	5.13	0.873	1.54	2.88
93	35.6	53	1.8	9.6	88.6	6.30E-09	-8.201	162	47	115	30	600	20	1.068	1.37	2.84
94	35.6	53	1.8	9.6	88.6	1.70E-09	-8.770	162	47	115	30	600	20	0.959	1.45	2.84
95	71	28	1	0	99	5.00E-13	-12.301	443	36	407	5	110	22	1.377	1.16	2.76
96	63	26	1	10	89	3.80E-12	-11.420	435	40	395	5	170	34	1.143	1.28	2.75

AMOSTRA	%Argila	%Silte	%Areia	%Pedreg	%Finos	$k_{sat}$ (m/s)	$\log(k_{sat})$ (m/s)	LL (%)	LP (%)	IP (%)	$d_{10}$ ( $10^{-5}$ mm)	$d_{60}$ ( $10^{-4}$ mm)	Cu	e	$\rho_d$ (g/cm <sup>3</sup> )	Gs
<b>97</b>	71	28	1	0	99	2.50E-13	-12.602	443	36	407	5	110	22	1.252	1.23	2.76
<b>98</b>	63	26	1	10	89	4.10E-12	-11.387	435	40	395	5	170	34	1.03	1.35	2.75
<b>99</b>	49.2	40	1.2	9.6	89.2	5.60E-11	-10.252	322	37	285	11	419.98	38.18	0.996	1.38	2.75
<b>100</b>	76	14	0	10	90	9.50E-12	-11.022	495	46	449	3.8	79.99	21.05	1.191	1.26	2.75
<b>101</b>	76	14	0	10	90	2.30E-12	-11.638	495	46	449	3.8	79.99	21.05	1.076	1.32	2.75
<b>102</b>	44	53	3	0	97	2.58E-10	-9.588	203	38	165	13	489.97	37.69	1.107	1.31	2.75
<b>103</b>	44	53	3	0	97	7.50E-10	-9.125	203	38	165	13	489.97	37.69	0.996	1.38	2.75
<b>104</b>	57	41	2	0	98	1.70E-11	-10.770	351	36	315	10	250	25	1.225	1.24	2.76
<b>105</b>	57	41	2	0	98	1.40E-11	-10.854	351	36	315	10	250	25	1.133	1.29	2.76
<b>106</b>	30	70	0	0	100	1.00E-09	-9.000	84	31	53	40	560	14	1.095	1.37	2.87
<b>107</b>	17	83	0	0	100	9.00E-08	-7.046	24	22	2	140	670.6	4.79	1.174	1.26	2.75
<b>108</b>	30	70	0	0	100	6.40E-10	-9.194	84	31	53	40	560	14	1.072	1.39	2.87
<b>109</b>	17	83	0	0	100	4.50E-08	-7.347	24	22	2	140	670.6	4.79	0.942	1.42	2.75
<b>110</b>	84	16	0	0	100	5.20E-13	-12.284	495	46	449	5.6	58.016	10.36	1.362	1.17	2.76
<b>111</b>	84	16	0	0	100	3.60E-13	-12.444	495	46	449	5.6	58.016	10.36	1.238	1.23	2.76

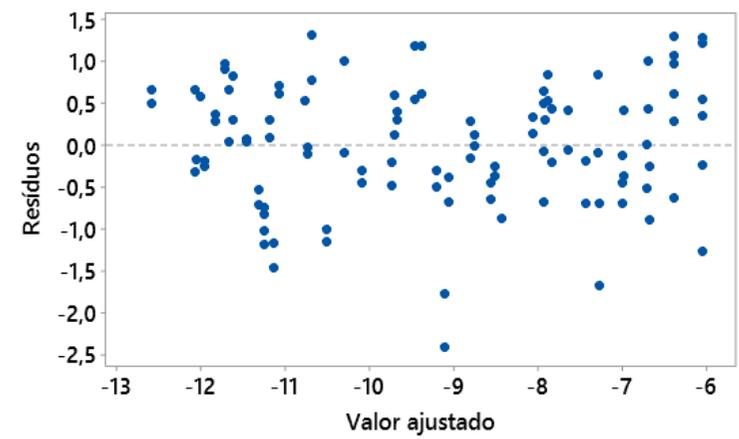
### APÊNDICE 6 – Diagrama de resíduos

Figura 33 – Análise 1 – Diagnóstico de resíduos – RL1



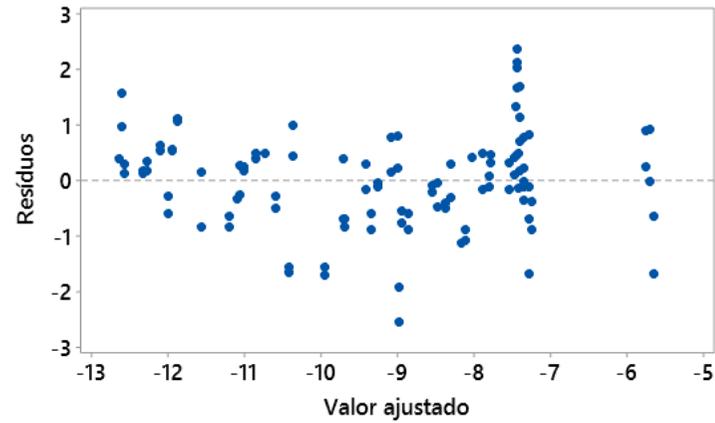
Fonte: Autora (2022).

Figura 34 - Análise 1 – Diagnóstico de resíduos – RNL1



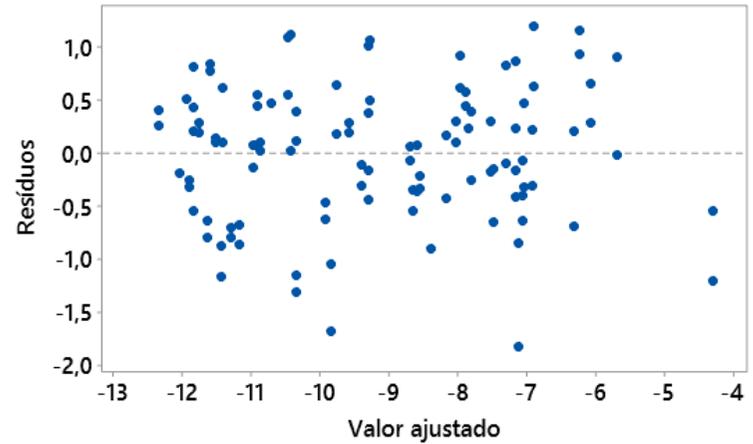
Fonte: Autora (2022).

Figura 35 - Análise 2 – Diagnóstico de resíduos – RL2



Fonte: Autora (2022).

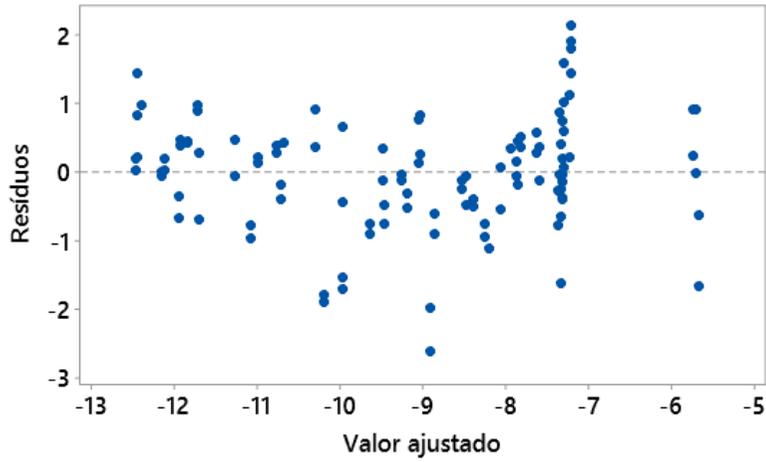
Figura 36 - Análise 2 – Diagnóstico de resíduos – RNL2



Fonte: Autora (2022).

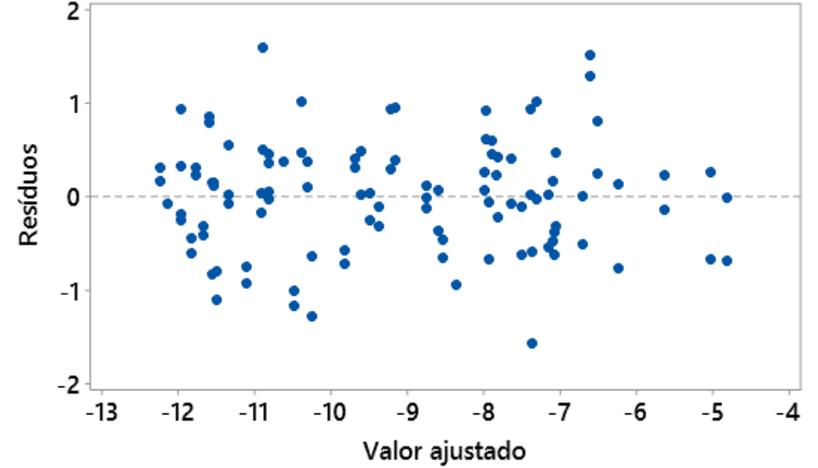
**APÊNDICE 6 – Diagrama de resíduos – continuação**

Figura 37 - Análise 3 – Diagnóstico de resíduos – RL3



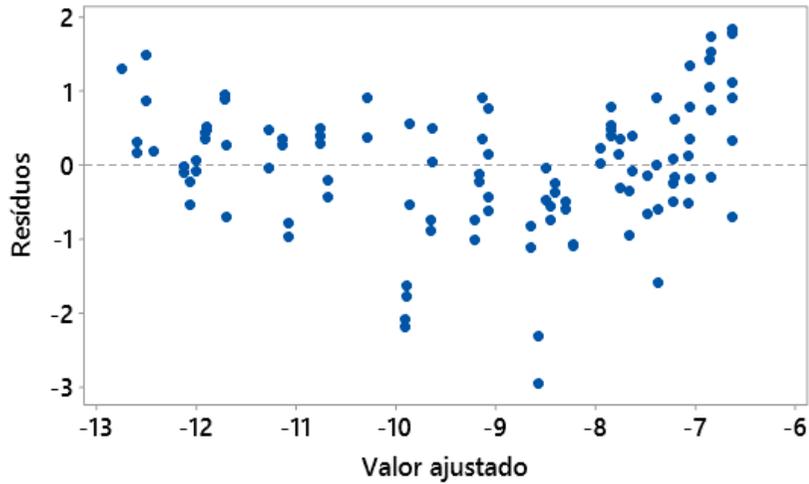
Fonte: Autora (2022).

Figura 38 - Análise 3 – Diagnóstico de resíduos – RNL3



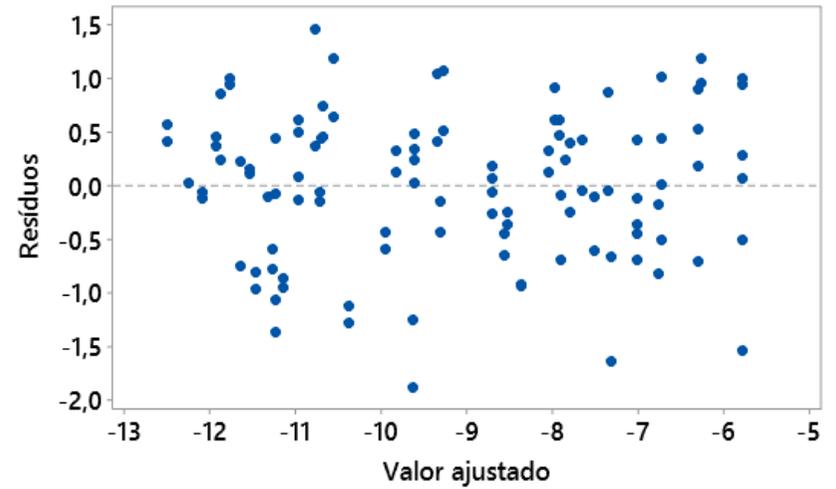
Fonte: Autora (2022).

Figura 39 - Análise 4 – Diagnóstico de resíduos – RL4



Fonte: Autora (2022).

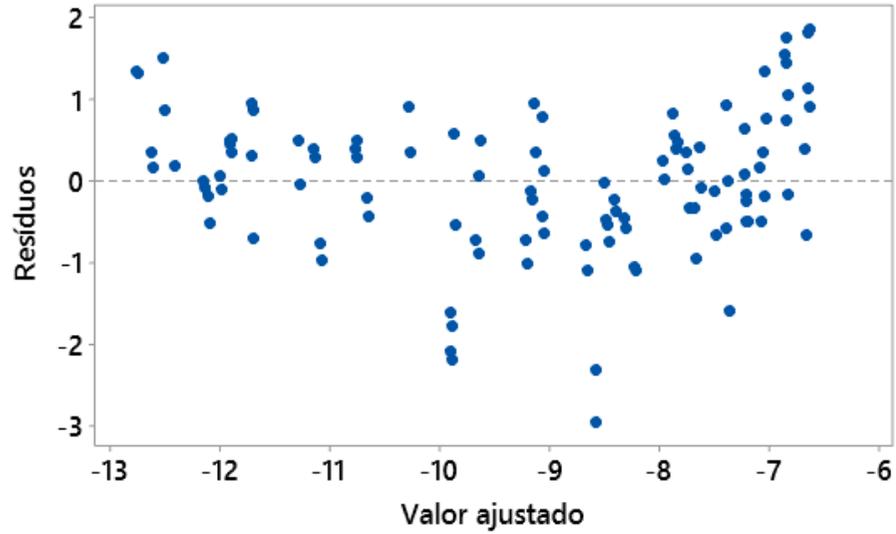
Figura 40 - Análise 4 – Diagnóstico de resíduos – RNL4



Fonte: Autora (2022).

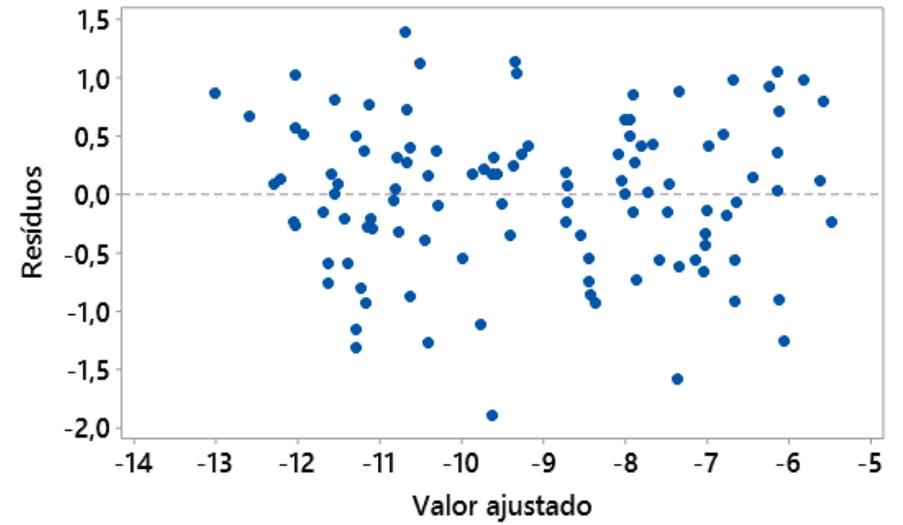
APÊNDICE 6 – Diagrama de resíduos – continuação

Figura 41 - Análise 5 – Diagnóstico de resíduos – RL5



Fonte: Autora (2022).

Figura 42 - Análise 5 – Diagnóstico de resíduos – RNL5



Fonte: Autora (2022).