

UFRRJ
INSTITUTO DE AGRONOMIA
DEPARTAMENTO DE PETROLOGIA E GEOTECTÔNICA
CURSO DE PÓS-GRADUAÇÃO EM MODELAGEM E EVOLUÇÃO
GEOLÓGICA

DISSERTAÇÃO

MAPEAMENTO E MODELAGEM DIGITAL DA VARIABILIDADE
TRIDIMENSIONAL DE ATRIBUTOS FÍSICO-HÍDRICOS DOS SOLOS
DA BACIA DO RIO GUAPI-MACACU - RJ, POR ESTATÍSTICA
MULTIVARIADA E ALGORITMOS

PRISCILLA AZEVEDO DOS SANTOS

2021



UNIVERSIDADE FEDERAL RURAL DO RIO DE JANEIRO
INSTITUTO DE AGRONOMIA
DEPARTAMENTO DE PETROLOGIA E GEOTECTÔNICA
CURSO DE PÓS-GRADUAÇÃO EM MODELAGEM E EVOLUÇÃO GEOLÓGICA

**MAPEAMENTO E MODELAGEM DIGITAL DA VARIABILIDADE
TRIDIMENSIONAL DE ATRIBUTOS FÍSICO-HÍDRICOS DOS SOLOS DA BACIA
DO RIO GUAPI-MACACU - RJ, POR ESTATÍSTICA MULTIVARIADA E
ALGORITMOS**

PRISCILLA AZEVEDO DOS SANTOS

Sob orientação da Professora

Dra. Helena Saraiva Koenow Pinheiro

e co-orientação de

Dr. Waldir de Carvalho Júnior

Dissertação submetida como requisito parcial para obtenção do grau de **Mestre em Geociências**, no Curso de Pós-Graduação em Modelagem e Evolução Geológica, Área de concentração de Modelagem de Processos Geológicos e Geodinâmicos, na linha de pesquisa de Processos Exógenos e Geoambientais.

Seropédica, RJ

Julho de 2021

Universidade Federal Rural do Rio de Janeiro
Biblioteca Central / Seção de Processamento Técnico

Ficha catalográfica elaborada
com os dados fornecidos pelo(a) autor(a)

S237m Santos, Priscilla Azevedo dos, 1991-
Mapeamento e Modelagem Digital da Variabilidade Tridimensional de Atributos Físico-Hídricos dos Solos da Bacia do Rio Guapi-Macacu - RJ, por Estatística Multivariada e Algoritmos / Priscilla Azevedo dos Santos. - Seropédica, Rio de Janeiro, 2021. 179 f.: il.

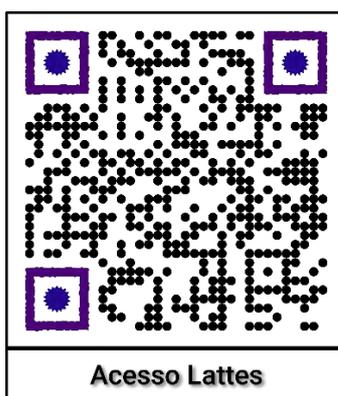
Orientadora: Helena Saraiva Koenow Pinheiro.
Coorientador: Waldir de Carvalho Júnior.
Dissertação(Mestrado). -- Universidade Federal Rural do Rio de Janeiro, Programa de Pós-Graduação em Modelagem e Evolução Geológica - PPGMEG, 2021.

1. Aprendizado de Máquina e AQP . 2. Estatística Multivariada. 3. Geoprocessamento. 4. Hidropedologia. 5. Sensoriamento Remoto. I. Pinheiro, Helena Saraiva Koenow, 1985-, orient. II. Júnior, Waldir de Carvalho, 1960-, coorient. III Universidade Federal Rural do Rio de Janeiro. Programa de Pós-Graduação em Modelagem e Evolução Geológica - PPGMEG. IV. Título. Mestre(a) em Geociências.

© 2021

Todos os direitos autorais reservados à Priscilla Azevedo dos Santos. A reprodução deste documento, em partes ou completo, só poderá ser feita mediante a citação correta da fonte.

E-mail: priscillaas@ufrj.br; priscilla.agricart@hotmail.com.



O formato deste documento segue as recomendações contidas na terceira edição do Manual de Instruções para Organização e Apresentação de Dissertações e Teses da Universidade Federal Rural do Rio de Janeiro (UFRRJ), publicado no ano de 2006.

O presente trabalho foi realizado com o apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – CAPES, Brasil, sob o código de financiamento 001. *This study was supported in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – CAPES, Brazil, by the 001 finance code.*

UNIVERSIDADE FEDERAL RURAL DO RIO DE JANEIRO
INSTITUTO DE AGRONOMIA
DEPARTAMENTO DE PETROLOGIA E GEOTECTÔNICA
CURSO DE PÓS-GRADUAÇÃO EM MODELAGEM E EVOLUÇÃO GEOLÓGICA

PRISCILLA AZEVEDO DOS SANTOS

Dissertação submetida como requisito parcial para obtenção do grau de **Mestre em Geociências**, no curso de Pós-graduação em Modelagem e Evolução Geológica, área de concentração de Modelagem de Processos Geológicos e Geodinâmicos, na linha de pesquisa de Processos Exógenos e Geoambientais.

DISSERTAÇÃO APROVADA EM: 12/07/2021.

Prof. Dra. Helena Saraiva Koenow Pinheiro, UFRRJ
(Orientador)

Prof. Dr. Marcos Bacis Ceddia, IA (UFRRJ)

Dr. Silvio Barge Bhering, CNPS (Embrapa Solos)

UNIVERSIDADE FEDERAL RURAL DO RIO DE JANEIRO
INSTITUTO DE AGRONOMIA
CURSO DE PÓS-GRADUAÇÃO EM MODELAGEM E EVOLUÇÃO
GEOLÓGICA

NOME DO PÓS-GRADUANDO: **PRISCILLA AZEVEDO DOS SANTOS**

Dissertação de mestrado submetida como requisito parcial para obtenção do grau de Mestre em Geociências no Curso de Pós-Graduação em Modelagem e Evolução Geológica, área de Concentração em Modelagem de Processos Geológicos e Geodinâmicos.

DISSERTAÇÃO APROVADA EM 12/07/2021

Conforme deliberação número 001/2020 da PROPPG, de 30/06/2020, tendo em vista a implementação de trabalho remoto e durante a vigência do período de suspensão das atividades acadêmicas presenciais, em virtude das medidas adotadas para reduzir a propagação da pandemia de Covid-19, nas versões finais das teses e dissertações as assinaturas originais dos membros da banca examinadora poderão ser substituídas por documento(s) com assinaturas eletrônicas. Estas devem ser feitas na própria folha de assinaturas, através do SIPAC, ou do Sistema Eletrônico de Informações (SEI) e neste caso a folha com a assinatura deve constar como anexo ao final da tese / dissertação.

BANCA EXAMINADORA:

Dra. Helena Saraiva Koenow Pinheiro (orientadora e presidenta da banca; DS-UFRRJ)

Dr. Marcos Bacis Ceddia (membro titular interno; DS-UFRRJ)

Dr. Sílvio Barge Bhering (membro titular externo, EMBRAPA)



Emitido em 12/07/2021

DOCUMENTOS COMPROBATÓRIOS Nº 11196/2021 - DPETRO (11.39.40)

(Nº do Protocolo: NÃO PROTOCOLADO)

(Assinado digitalmente em 27/08/2021 18:30)

HELENA SARAIVA KOENOW PINHEIRO

PROFESSOR DO MAGISTERIO SUPERIOR

DeptS (12.28.01.00.00.00.33)

Matrícula: 2223668

(Assinado digitalmente em 29/08/2021 18:43)

MARCOS BACIS CEDDIA

PROFESSOR DO MAGISTERIO SUPERIOR

DeptS (12.28.01.00.00.00.33)

Matrícula: 1220296

(Assinado digitalmente em 27/08/2021 17:01)

SILVIO BARGE BHERING

ASSINANTE EXTERNO

CPF: 706.481.827-20

Para verificar a autenticidade deste documento entre em <https://sipac.ufrrj.br/documentos/> informando seu número:
11196, ano: **2021**, tipo: **DOCUMENTOS COMPROBATÓRIOS**, data de emissão: **27/08/2021** e o código de
verificação: **fd9e97f1da**

DEDICATÓRIA

Aos meus avós Eduardo Mesquita (*in memoriam*) e Maria de Lourdes Costa

Ao meu companheiro Igor Leite da Silva

À minha gatinha Fifi (*in memoriam*)

Com muito amor e carinho

Dedico

AGRADECIMENTOS

Ao Divino, pela força e energia positiva diária.

Ao meu falecido avô, Eduardo Mesquita dos Santos, por ter sido pai e avô ao mesmo tempo, sendo um modelo de homem justo, amoroso e dedicado na minha vida. Obrigada por ter sido meu primeiro professor, ao me ensinar caligrafia e tabuada, e por estimular o meu progresso até o último suspiro. Você será sempre a minha inspiração. Te amo eternamente.

À minha avó, Maria de Lourdes Costa dos Santos, por todo amor, carinho, dedicação, força e crença no meu potencial, tornando possível o ingresso, manutenção e conclusão dos meus estudos no ensino superior na UFRRJ, tanto na graduação quanto no mestrado. Obrigada pelo amor maior.

Ao meu companheiro Igor Leite da Silva, por todo apoio, amor, dedicação, paciência e carinho demonstrado ao longo dessa intensa jornada acadêmica. Obrigada por ser o meu ponto de equilíbrio e acreditar em mim todos os dias.

À Profa. Dra. Helena Saraiva Koenow, por me acolher enquanto orientada, dedicando seu tempo e conhecimento para que eu alcançasse a excelência nesta pesquisa e por contribuir diretamente na minha formação profissional. Agradeço a confiança e carinho a mim depositados. Agradeço ainda por compor o corpo de pesquisadores do laboratório LAAI.

Ao Dr. Waldir de Carvalho Júnior, pela coorientação, pelo compartilhamento de sua experiência na área de Mapeamento Digital de Solos e pelas boas risadas regadas à rapadura nas idas à campo.

Ao Prof. Dr. Marcos Bacis Ceddia, por partilhar conhecimentos na área de Física dos Solos e por me acolher no laboratório LASAGEO.

Aos professores que compõe o quadro docente do PPGMEG, por lecionarem as disciplinas de forma excepcional e se tornarem motivo de exemplo educacional e profissional para nós discentes em um país que carece de recursos e valorização da pesquisa científica. Agradeço a oportunidade em compor o corpo de pesquisadores do laboratório LabMeg.

Aos pesquisadores da Embrapa Solos, em especial o MSc. Nilson Rendeiro Pereira e o Dr. Waldir de Carvalho Júnior, por participarem ativamente nos levantamentos hidropedológicos, compartilhando conhecimentos e experiências sobre o tema da dissertação. Obrigada pelas campanhas de campo ricas em boas risadas e trabalho árduo.

Aos colegas do curso de geologia, Theresa Rocco e José Mateus Matos, por compor a equipe de campo e atuar na coleta dos dados. Muito obrigada!

Ao pesquisador Dr. Silvio Barge Bhering, pela concessão do Permeâmetro de Guelph usado nas campanhas de campo e por sanar dúvidas sob o método de coleta hidropedológica.

Aos professores participantes da banca examinadora: Dr. Silvio Barge Bhering e Dr. Marcos Bacis Ceddia (titulares); Dr. Cesar da Silva Chagas, Dra. Olga Venimar de Oliveira Gomes e Dr. Waldir de Carvalho Júnior (suplentes); pelo convite aceito, tempo dedicado e pelas valiosas colaborações e sugestões que contribuíram para o enriquecimento do conteúdo deste trabalho.

À Embrapa Solos do Rio de Janeiro, por disponibilizar o transporte, os equipamentos e as ferramentas usadas no levantamento de campo.

À Universidade Federal Rural do Rio de Janeiro, agradeço por me proporcionar um ensino público, gratuito e de qualidade, se tornando a morada do meu conhecimento.

À CAPES, pela concessão da bolsa de estudos, garantindo a minha subsistência no mestrado.

Aos órgãos governamentais FAPERJ e PETROBRÁS, pelo financiamento do projeto associado a este estudo; e à DRM, por ceder as cartas geológicas do projeto 1:50.000 as quais fundamentaram parte desta pesquisa.

Ao Serviço Geológico do Brasil - CPRM e à *European Space Agency - ESA*, pela política de dados abertos ao público e disponibilização dos dados aerogeofísicos e imagens de satélites em seus repositórios online.

Aos meus professores do corpo docente do departamento de matemática (DeptM) da UFRRJ, à saber: Dra. Adriana Oliveira Andrade, Dr. Diógenes Ferreira Filho, Dr. Renato Nunes Pereira e Dr. Felipe Leite Coelho da Silva; pelo ensino e conhecimentos adquiridos em estatística aplicada que auxiliaram na interpretação dos dados e resultados da pesquisa.

O presente trabalho foi realizado com o apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - CAPES, Brasil, sob o código de financiamento 001. *This study was supported in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - CAPES, Brazil, by the 001 finance code.*

“A flor que desabrocha na adversidade é a mais rara e mais bela de todas.”
(Mulan)

BIOGRAFIA

Priscilla Azevedo dos Santos nasceu no dia 27 de maio de 1991, na cidade de Engenho de Dentro, município do Rio de Janeiro. Viveu desde seu nascimento até os 28 anos de idade com seus avós, Maria de Lourdes Costa dos Santos e Eduardo Mesquita dos Santos (falecido), na cidade de Nilópolis, baixada fluminense do Rio de Janeiro. Estudou até 2005 no colégio ABEU, ingressando no colégio Equipe Grau no ano seguinte, onde cursou o ensino médio com bolsa (2006-2008) e, em 2018, foi contemplada com bolsa de 100% para cursar o pré-vestibular no mesmo colégio, mediante processo seletivo interno (simulado). Em 2013, foi aprovada em primeiro lugar no processo seletivo do curso técnico/profissionalizante em Edificações da Fundação de Apoio à Escola Técnica do Estado do Rio de Janeiro – FAETEC, formando-se técnica no ano posterior (2013-2014). Ingressou na Universidade Federal Rural do Rio de Janeiro em novembro de 2013, em meio a uma greve geral, para cursar Engenharia de Agrimensura e Cartográfica. Na vivência acadêmica, atuou como Assessora de Projetos e Diretora Financeira na empresa Mensurar Júnior (2015-2017) e desenvolveu pesquisa de Iniciação Científica na área de Geociências com ênfase em Meteorologia Atmosférica financiada pelo CNPq (2015-2017) sob orientação do Prof. Dr. Gustavo Bastos Lyra (IF/UFRRJ). Ainda, realizou estágio externo no laboratório de Geoprocessamento LIS da FIOCRUZ campus Manginhos (2017-2018) sob supervisão das pesquisadoras Dr.^a Mônica de Avelar Figueiredo Mafra Magalhães e Dr.^a Renata de Saldanha da Gama Gracie Carrijo; e estágio interno no laboratório de Geomática do Colégio Técnico da Universidade Rural – CTUR (2018), sob orientação do Prof. MSc. Alexandre Araújo Ribeiro Freire. Atuou ainda como monitora voluntária na disciplina de Sistemas de Informações Geográficas (SIG) Avançado no laboratório de Sensoriamento Remoto e Análise Espacial do Instituto de Tecnologia da UFRRJ (2018), sob orientação da Prof.^a Dr.^a Alessandra Carreiro Baptista. Desenvolveu monografia no Laboratório de Fotogrametria (IT/UFRRJ), voltada para as áreas de Fotogrametria e Sensoriamento Remoto do Meio Ambiente usando imagens coletadas por sensor orbital e sensor embarcado em Aeronave Remotamente Pilotada (ARP). Graduiu-se em dezembro de 2018, com colação de grau realizada no mês seguinte (janeiro de 2019). Em março de 2019, foi contemplada com bolsa CAPES no processo seletivo do curso de Pós-graduação em Modelagem e Evolução Geológica (PPGMEG/UFRRJ), na área de concentração de Processos Geológicos e Geodinâmicos e linha de pesquisa associada à Processos Exógenos e Geoambientais, sob orientação da Prof.^a Dr.^a Helena Saraiva Koenow Pinheiro (IA/UFRRJ) e coorientação do pesquisador Dr. Waldir de Carvalho Júnior (CNPS/Embrapa Solos). Desenvolveu a atual pesquisa no Laboratório de Análise Ambiental Integrada (LAAI) e no Laboratório de Modelagem e Evolução Geológica (LabMeg). Atuou concomitantemente com pesquisa de apoio no Laboratório de Geoprocessamento aplicado a Água e Solos em Agroecossistemas (LASAGEO), onde ganhou conhecimentos relativos à sua dissertação junto ao Prof. Dr. Marcos Bacis Ceddia (IA/UFRRJ). Atualmente, mora em Seropédica, região metropolitana do Rio de Janeiro e cidade da amada UFRRJ.

RESUMO GERAL

SANTOS, Priscilla Azevedo dos. **Mapeamento e Modelagem Digital da Variabilidade Tridimensional de Atributos Físico-Hídricos dos Solos da Bacia do Rio Guapi-Macacu - RJ, por Estatística Multivariada e Algoritmos**. 2021. 182f Dissertação (Mestrado em Modelagem e Evolução Geológica, Geociências). Instituto de Agronomia, Departamento de Petrologia e Geotectônica, Universidade Federal Rural do Rio de Janeiro, Seropédica, RJ, 2021.

O conhecimento acerca dos atributos físico-hídricos dos solos é importante para estudos voltados a compreensão do regime hídrico e monitoramento do fluxo de água, principalmente em bacias hidrográficas, onde o conteúdo de água armazenado e disponibilizado afeta tanto as funções ambientais dos solos, quanto a biodiversidade e a sustentabilidade desse recurso natural. No Brasil, os bancos de dados de solos possuem poucas informações coletadas acerca de parâmetros hídricos dos solos tais como a velocidade de infiltração básica (vib) e a condutividade hidráulica saturada (Ksat), devido à não realização sistemática de testes de infiltração ao se executar os levantamentos pedométricos e a dificuldade de mensuração de tais parâmetros nas camadas mais profundas da pedosfera. Neste contexto, torna-se passível a estimativa da vib e da Ksat associando-se as propriedades granulométricas e físico-químicas dos solos coletadas em campo por meio de algoritmos para pedologia quantitativa (do inglês, *Algorithms for Quantitative Pedology* - AQP) e implementação de funções de pedotransferência usando análise regressiva multivariada e algoritmos de *machine learning* baseados em árvores, capazes de modelá-los vertical (em perfil) e espacialmente sob a área de estudo. Ainda, como forma de ampliar as informações sobre a área estudada e garantir uma modelagem mais fidedigna e robusta, é desejável associar parâmetros mensuráveis em campo e laboratório com demais informações relevantes que ajudem a análise de bacias hidrográficas compondo assim as variáveis de entrada nos modelos citados. Este estudo sugere a aplicação de variáveis oriundas de modelagem numérica do terreno, obtidas através de Modelo Digital de Elevação (MDE), e dados radiométricos, derivados aerogeofísica ambiental (aeromagnetometria e aerogamaespectrometria) e análise espectral sob índices relativos à vegetação, solo e água utilizando imagens do sensor Sentinel-2A (índices espectrais) por meio de Sensoriamento Remoto. Para a análise quantitativa dos dados e seleção de covariáveis dos modelos, foram abordados métodos estatístico-descritivo e multivariado, visando o entendimento interrelacional das variáveis preditoras e a redução de dimensionalidades e/ou multicolinearidade nas variáveis de entrada nos modelos. Pelos resultados obtidos, os modelos baseados em árvores (*Random Forest* – RF e Árvores de Regressão - AR) apresentaram melhor desempenho na modelagem dos atributos físico-hídricos frente ao modelo regressivo na estimativa das funções de pedotransferência. A abordagem multivariada usando os métodos de seleção e redução de dimensionalidade permitiram a escolha otimizada das variáveis de entrada na modelagem, eliminação de problemas de multicolinearidade dos dados e redução do conjunto de dados, obtendo diversificada resposta para as camadas de solos avaliadas. O estudo mostra o potencial de integração de dados topográficos, pedológicos e radiométricos e sua contribuição no mapeamento e modelagem digital de solos, visando a compreensão da variabilidade dos atributos físico-hídricos na bacia hidrográfica estudada.

Palavras-chave: Aprendizado de máquina. AQP. Estatística Multivariada. Geoprocessamento. Hidropedologia. Sensoriamento Remoto.

GENERAL ABSTRACT

SANTOS, Priscilla Azevedo dos. **Digital Mapping and Modeling of Three-dimensional Soil Physical-hydric Attributes Variability in the Guapi-Macacu Watershed – RJ, Using Multivariate Statistics and Algorithms.** 2021. 182p Dissertation (Master Science in Modeling and Geological Evolution, Geoscience). Agronomy Institute, Petrology and Geotectonics Department, Federal Rural University of Rio de Janeiro, Seropédica, RJ, 2021.

The knowledge about soil attributes and hydrogeology is important for studies that aim understand the hydrological regime and monitoring water flow, especially in watersheds, where the stored and available water content affects both soils environmental functions, as well as the biodiversity and the sustainability of this natural resource. In Brazil, soil databases have few information collected about soil water parameters such as basic infiltration speed ratio (bir) and saturated hydraulic conductivity (Ksat), due to the non-systematic performance of infiltration tests at soil surveys and the difficulty of measuring such parameters in the pedosphere's deepest layers. In this context, it is possible to estimate the bir and Ksat by associating the granulometric and physicochemical properties of the soil collected in the field, by means of algorithms for quantitative pedology (Algorithms for Quantitative Pedology - AQP) and implementation of pedotransfer functions using multivariate regressive analysis and tree-based machine learning algorithms, capable of modeling them vertically (in profile) and spatially at the study area. Still, as a way of expanding the information about the studied area and guaranteeing a more reliable and robust modeling, it is desirable to associate measurable parameters in the field and laboratory with other relevant information that help the analysis of hydrographic basins, thus composing the input variables in the models mentioned. This study suggests the application of variables from numerical modeling of the terrain, obtained through the Digital Elevation Model (DEM), radiometric data, derived from environmental aerogeophysics (aeromagnetometry and aerogamaespectrometry) and spectral analysis through indices related to vegetation, soil and water using images from the Sentinel-2A sensor (spectral indexes). The quantitative analysis and covariates selection, statistical-descriptive and multivariate methods were applied, aiming to select potential covariates and the reduction of dimensionalities and / or multicollinearity in the input variables in the models. Based on the results obtained, the tree-based models (Random Forest - RF and Regression Trees - RT) showed better performance in modeling the physical-hydric attributes when compared to the regressive model to build pedotransfer functions. The multivariate approach using covariates selection and dimensionality reduction methods allowed the optimized choice of input variables in the modeling, elimination of multicollinearity problems, obtaining a diversified response to the evaluated soil layers. The study shows the potential for integrating topographic, pedological and radiometric data and their contribution to digital soil mapping and modeling, aiming at understanding the variability of physical-hydric attributes in the studied watershed.

Keywords: Machine Learning. AQP. Multivariate Statistics. Hydrogeology. Digital Mapping. Remote Sensing.

LISTAS DE FIGURAS

Figura 1. Fluxograma da metodologia adotada na pesquisa.....	27
Figura 2. Funções do Solo. Fonte: FAO (2015).	28
Figura 3. Ciclo hidrológico. Fonte: adaptado de Brasil Escola (2020).	29
Figura 4. Movimentos e processos internos do solo. Fonte: VEREECKEN et. al (2016).	31
Figura 5. Zonas de saturação do solo e a porosidade dos solos, onde: no círculo à direita estão discriminadas as composições dos poros (em cor azul) e da parte sólida (em cor verde) dos solos. Fonte: adaptado de DOS SANTOS et al. (2018) & VEREECKEN et. al (2016).	33
Figura 6. Fluxo da água em solo saturado: diagrama de dinâmica dos fluidos segundo a lei de Darcy (1856) (arranjo experimental). Fonte: adaptado de LIBARDI (2012).....	36
Figura 7. Fluxo da água em solo não saturado: diagrama da equação de Darcy-Buckingham (arranjo experimental). Fonte: LIBARDI (2012).	37
Figura 8. Elemento de volume de solo, através do qual a água está fluindo. Fonte: LIBARDI (2012).	38
Figura 9. Mapa de distribuição dos pontos amostrais de vib e Ksat na bacia.	52
Figura 10. Fluxograma da metodologia da pesquisa.	62
Figura 11. Fluxo de procedimentos adotado na estimativa das funções de pedotransferência para o conjunto de dados.	63
Figura 12. Sub bacias e Morfometria da bacia hidrográfica do Rio Guapi-Macacu, Rio de Janeiro, contendo as seguintes informações: (a) Delimitação das unidades hidrológicas de proteção (sub bacias); (b) Mapa de declividade (%); (c) Mapa de índice topográfico de umidade (do inglês, <i>Topographic Wetness Index</i> , TWI); (d) Mapa de curvatura classificada (Planar + Perfil); (e) Mapa de Aspecto da vertente (pontos cardeais e orientação).	64
Figura 13. Gráfico de autocorrelação dos atributos dos solos da bacia hidrográfica do Rio Guapi-macacu, segundo análise do coeficiente de correlação de Karl Pearson (ρ ; Pearson, 1904; Mukaka, 2012).....	72
Figura 14. Coleção de perfis de solos da bacia hidrográfica do Rio Guapi-Macacu, Rio de Janeiro. O identificador do perfil corresponde às informações contidas na Tabela 1.	75
Figura 15. Composição e variabilidade da areia (total) atributos nos perfis dos solos, na unidade $g.kg^{-1}$. O identificador do perfil corresponde às informações contidas na Tabela 1.	77
Figura 16. Composição e variabilidade do silte nos perfis dos solos, na unidade $g.kg^{-1}$. O identificador do perfil corresponde às informações contidas na Tabela 1.	78

Figura 17. Composição e variabilidade da argila nos perfis dos solos, na unidade g.kg^{-1} . O identificador do perfil corresponde às informações contidas na Tabela 1	79
Figura 18. Composição e variabilidade do valor T nos perfis de solo da bacia, na unidade $\text{cmol}_c.\text{kg}^{-1}$. O identificador do perfil corresponde às informações contidas na Tabela 1.	81
Figura 19. Composição e variabilidade da areia fina na coleção de perfis, na unidade g.kg^{-1} . O identificador do perfil corresponde às informações contidas na Tabela 1.	82
Figura 20. Composição e variabilidade da porosidade na coleção de perfis, na unidade porcentagem (%). O identificador do perfil corresponde às informações contidas na Tabela 1.	83
Figura 21. Composição da areia grossa na coleção de perfis, na unidade g.kg^{-1} . O identificador do perfil corresponde às informações contidas na Tabela 1.	84
Figura 22. Composição da Argila dispersa em água na coleção de perfis, na unidade g/kg . O identificador do perfil corresponde às informações contidas na Tabela 1.	85
Figura 23. Composição e variabilidade do carbono orgânico nos perfis dos solos, na unidade g.kg^{-1} . O identificador do perfil corresponde às informações contidas na Tabela 1.	87
Figura 24. Composição e variabilidade do carbono orgânico (g.kg^{-1}) na bacia segundo (a) triângulo textural da classificação SiBICS e gráfico de bolhas (<i>bubble plot</i>) da variação do carbono orgânico conforme a elevação (altimetria, Z), considerando na legenda os quantis da distribuição (0, 25,50,75 e 100). A variável Corgânico é plotada como um fator de expansão considerando Z; (b) Mapa de calor (do inglês, <i>heatmap</i>) interpolado usando o IDW (<i>Inverse Distance Weighted</i>) e contendo as isoietas de elevação representadas pelas linhas contínuas (em azul royal).	88
Figura 25. Composição e variabilidade da densidade do solo (d_s) nos perfis dos solos, em g.cm^{-3} . O identificador do perfil corresponde às informações contidas na Tabela 1.	90
Figura 26. Composição e variabilidade da densidade da partícula (d_p) nos perfis dos solos, em g.cm^{-3} . O identificador do perfil corresponde às informações contidas na Tabela 1.	91
Figura 27. Gráficos de importância das variáveis preditoras na estimativa da condutividade hidráulica saturada (K_{sat}) por modelagem <i>Random Forest</i>	95
Figura 28. Gráficos de importância das variáveis preditoras na estimativa da velocidade de infiltração básica (vib) por modelagem <i>Random Forest</i>	96
Figura 29. Gráficos de Árvores de Regressão (<i>Regression Trees</i> – RT) e iterações na calibração das funções de pedotransferência de modelos baseados em árvores de regressão, onde: (a,b,c,d) condutividade hidráulica saturada (K_{sat}); (e,f,g,h) velocidade de infiltração básica (vib). Nota: <i>yes</i> = booleano para “sim”; <i>no</i> = booleano para “não”	97

Figura 30. Mapa de localização da área de estudo e distribuição espacial dos 122 pontos associados às informações hidropedológicas do conjunto de dados analisado na pesquisa...	117
Figura 31. Funções de Resposta Espectral (FRE) do satélite Sentinel-2A (S2A) e transmissão devido à coluna vertical de absorção de Vapor D'água (VD) para a atmosfera subártica de inverno (azul claro, menor teor de VD) e atmosfera tropical (azul mais escuro, maior teor de VD). Os dados de transmissão VD são oriundos do programa computacional MODTRAN (do inglês, MODerate resolution atmospheric TRANsmission). O eixo x refere-se ao comprimento de onda do espectro eletromagnético (Wavelength) e, o eixo y, ao valor de transmissão e resposta espectral normalizada. Fonte: ESA (2021).....	121
Figura 32. Fluxograma da metodologia proposta.....	124
Figura 33 . Mapa aerogeofísico de gamaespectrometria dos elementos K:Th:U.....	126
Figura 34. Mapa aerogeofísico de magnetometria do sinal analítico em.....	126
Figura 35. Matriz de correlação de Pearson envolvendo o conjunto de covariáveis estudo e a variável Ksat para a camada de 0-20cm.	129
Figura 36. Matriz de correlação de Pearson envolvendo o conjunto de covariáveis estudo e a variável Ksat para a camada de 20-40cm.	130
Figura 37. Matriz de correlação de Pearson envolvendo o conjunto de covariáveis estudo e a variável vib para a camada de 0-20cm.	131
Figura 38. Matriz de correlação de Pearson envolvendo o conjunto de covariáveis estudo e a variável vib para a camada de 20-40cm.	132
Figura 39. Variáveis espectrais representadas segundo as componentes principais explicativas do modelo ACP, onde: (a) importância dos componentes principais na ACP; (b) separação das componentes principais na ACP.....	132
Figura 40. Gráficos do método RFE (quantidade de variáveis explicativas versus RMSE do ajuste) para os atributos físico-hídricos estudados e seus respectivos intervalos de profundidade: (a) Ksat _{0-20 cm} ; (b) Ksat _{20-40 cm} ; (c) vib _{0-20 cm} ; (d) vib _{20-40 cm} . A linha tracejada indica o alcance da quantidade ótima de variáveis selecionadas pelo RFE.	136
Figura 41. Gráficos de ajuste do modelo <i>Random Forest</i> , evidenciando a quantidade de árvores versus a minimização do erro para as variáveis Ksat e vib, segundo as camadas avaliadas: (a-c) Ksat _{0-20cm} ; (d-f) Ksat _{20-40cm} ; (g-i) vib _{0-20cm} ; (j-l) vib _{20-40cm} ; e os três métodos analisados: (a,d,g,j) VIF; (b,e,h,k) RFE; (c,f,i,l) <i>Stepwise AIC</i>	139
Figura 42. Mapas da variabilidade dos atributos vib e Ksat na bacia hidrográfica do Rio Guapi-Macacu, Rio de Janeiro, estimados através do modelo <i>Random Forest</i> , segundo as respectivas	

profundidades e métodos de seleção: **(a)** vib_{0-20cm} - RFE; **(b)** vib_{20-40cm} - StepwiseAIC; **(c)** Ksat_{0-20cm} - RFE; **(d)** Ksat_{20-40cm} - StepwiseAIC..... 141

LISTA DE TABELAS

Tabela 1. Valores de velocidade de infiltração básica (vib) e condutividade hidráulica saturada (Ksat) medidos nos perfis de solos da bacia hidrográfica do rio Guapi-Macacu.	54
Tabela 2. Parâmetros geométricos da bacia.	65
Tabela 3. Estatística exploratória dos atributos físico-hídricos, da composição granulométrica e da argila dispersa em água na bacia do Rio Guapi-Macacu.	68
Tabela 4. Estatística exploratória dos atributos da bacia do Rio Guapi-Macacu (propriedades químicas pH e Valor T; densidades da partícula e do solo; porosidade e carbono orgânico. ..	69
Tabela 5. Equações pedotransferência ajustadas por meio de Regressão linear múltipla.	94
Tabela 6. Testes estatísticos para pressupostos em regressão.	94
Tabela 7. Parâmetros estatísticos de qualidade dos modelos treinados após validação cruzada (LOOCV aplicado aos 36 pontos amostrais).	98
Tabela 8. Parâmetros estatísticos da totalidade dos dados (preditos x estimados) obtidos na validação cruzada (LOOCV aplicado aos 122 pontos amostrais).	99
Tabela 9. Covariáveis selecionadas para modelagem da variável Ksat após aplicação dos métodos propostos.	135
Tabela 10. Qualidade do modelo <i>Random Forest</i> no treino dos dados físico-hídricos a partir dos métodos de seleção aplicados (70% do banco de dados).	138
Tabela 12. Qualidade do Modelo <i>Random Forest</i> na Validação dos dados físico-hídricos de teste (30% do banco de dados).	140

LISTA DE QUADROS

Quadro 1. Características físico-químicas da água.	30
Quadro 2. Forças potenciais do Fluxo de água no solo.....	34
Quadro 3. Classificação e condição hidrológica dos solos da BHRGM.....	56
Quadro 4. Organização e composição do banco de dados pedológico analítico da dissertação, com as variáveis de estudo e suas respectivas características.	67
Quadro 5. Análise de adequação de modelos regressivos.....	93
Quadro 6. Covariáveis do terreno relativas à atributos topográficos físico-hídricos e temodinâmicos primários e secundários do MDE.....	166
Quadro 7. Características do Projeto Aerogeofísico do Estado do Rio de Janeiro.	168
Quadro 8. Características das Aeronaves e do processo de aquisição de dados do Projeto Aerogeofísico do Estado do Rio de Janeiro.	169
Quadro 9. Características espectrais do sensor MSI da missão Sentinel-2A.	170
Quadro 10. Índices de vegetação derivados das bandas espectrais do sensor Multiespectral da missão Sentinel-2.....	171

LISTA DE SIGLAS E SÍMBOLOS

ACP	Análise de Componentes Principais
AIC	<i>Akaike Information Criterion</i>
ANOVA	<i>Analysis Of Variance</i>
AQP	<i>Algorithms for Quantitative Pedology</i>
AR	Árvores de Regressão
ASA	Amplitude do Sinal Analítico
BHRGM	Bacia Hidrográfica do Rio Guapi-Macacu
BIC	<i>Bayesian Information Criterion</i>
BIR	<i>Basic Infiltration Ratio</i>
BP	<i>Breusch-Pagan Test</i>
cLHS	<i>Conditioned Latin Hypercube Sample</i>
CMA	Campo Magnético Anômalo
CPRM	Companhia de Pesquisa de Recursos Minerais
CV	<i>Cross Validation</i>
DEM	<i>Digital Elevation Model</i>
DRM	Diretoria de Recursos Minerais
DW	<i>Durbin-Watson Test</i>
ESA	<i>European Space Agency</i>
FA	Floresta Aleatória
FAPED	Fundação de Apoio à Pesquisa e ao Desenvolvimento
FPT	Função de Pedotransferência
FRE	Funções de Resposta Espectral
FTPs	Funções de Pedotransferência
FWHM	<i>Full Width Half Maximum</i>
GeoINEA	Portal de Geociências do Instituto Estadual do Ambiente
GIS	<i>Geographic Information Systems</i>
GMES	<i>Global Monitoring for Environment and Security</i>
GEOBIA	<i>Geographic Object-Based Image Analysis</i>
GeoSGB	Sistema de Geociências do Serviço Geológico do Brasil
IBGE	Instituto Brasileiro de Geografia e Estatística
IDW	<i>Inverse Distance Weighted</i>
IGRF	<i>International Geomagnetic Reference Field</i>

INEA	Instituto Estadual do Ambiente
KFCV	<i>K-fold cross-Validation</i>
Ksat	Condutividade Hidráulica Saturada ou, do inglês, <i>Saturated Hydraulic Conductivity</i>
LOOCV	<i>Leave One Out Cross-Validation</i>
MAE	<i>Mean Absolute Error</i>
MDE	Modelo Digital de Elevação
MDEHC	Modelo Digital Hidrologicamente Consistido
MDS	Mapeamento Digital de Solos
MFA	<i>Multiple Factor Analysis</i>
ML	<i>Machine Learning</i>
MLR	<i>Multiple Linear Regression</i>
MNT	Modelo Numérico do Terreno
MODTRAN	MODerate resolution atmospheric TRANsmission
NRCS	<i>Natural Resources Conservation Service</i>
MSI	<i>Multispectral Sensor Instrument</i>
OOB	<i>Out-Of-Bag</i>
PCs	<i>Principal Components</i>
PRESS	<i>Predicted Residual Error Sum of Squares</i>
PTF	<i>Pedotransfer Functions</i>
RF	<i>Random Forest</i>
RFE	<i>Recursive Feature Elimination</i>
RG	Referência Gravitacional
RH	Região Hidrográfica
RKFCV	<i>Repeated K-fold cross-validation</i>
RLM	Regressão Linear Múltipla
RMSE	<i>Root Squared Mean Error</i>
RT	<i>Regression Trees</i>
RTP	<i>Reduction To Pole</i>
SDF	<i>Soil-Depth Functions</i>
SEA	Secretaria Estadual do Ambiente
SGN	Sistema Cartográfico Nacional
SiBCS	Sistema Brasileiro de Classificação dos Solos
SIG	Sistemas de Informações Geográficas

SQE	Soma dos Quadrados dos Erros
SQR	Soma de Quadrados da Regressão
SQT	Soma de Quadrados Total
SRF	<i>Spectral Response Functions</i>
StepAIC	<i>Stepwise AKAIKE Criterion</i>
SW	<i>Shapiro-Wilk Test</i>
TWI	<i>Topographic Wetness Index</i>
UFRRJ	Universidade Federal Rural do Rio de Janeiro
USDA	<i>United States Department of Agriculture</i>
UTM	<i>Universal Transversa de Mercator</i>
VD	Vapor D'água
Vib	Velocidade de Infiltração Básica dos Solos
Vif	<i>Variance Inflation Factor</i>
VIS-IR	<i>Visible-Infrared Bands</i>

SUMÁRIO

1 INTRODUÇÃO GERAL	24
1.1 Contextualização	24
1.2 Visão Geral da Pesquisa	25
2 REVISÃO DE LITERATURA.....	28
2.1 O Fluxo de Água e as Propriedades Físico-hídricas dos Solos	28
2.2 Hidropedologia: o Teste de Infiltração de Guelph e a Modelagem de Atributos Físico-hídricos dos Solos.....	42
CAPÍTULO I	47
MODELAGEM VERTICAL DOS ATRIBUTOS FÍSICO-HÍDRICOS DOS SOLOS...47	
1. INTRODUÇÃO	50
2. MATERIAL E MÉTODOS	50
2.1. Área de Estudo e Caracterização Hidrológica	50
2.2. Banco de Dados de Solos e Testes de Infiltração	51
2.3. Modelagem por Funções de Pedotransferência para a Determinação de Atributos Físico-Hídricos dos Solos.....	57
2.4. Procedimentos Adotados na Modelagem Vertical de Atributos Hidropedológicos	59
3. RESULTADOS E DISCUSSÃO	63
3.1. Caracterização Hidrológica da Bacia do Rio Guapi-Macacu	63
3.2. Análise Exploratória dos Dados.....	66
3.3. Análise por Algoritmos para Pedologia Quantitativa (AQP).....	73
3.4. Modelagem dos Atributos Físico-Hídricos	92
4. CONCLUSÕES	102
CAPÍTULO II.....	103
MODELAGEM ESPACIAL DOS ATRIBUTOS FÍSICO-HÍDRICOS DOS SOLOS..103	
1. INTRODUÇÃO	106
1.1. Contextualização	106
1.2. Revisão: Métodos de Redução de Dimensionalidade e Seleção na Modelagem Digital via Classificação Supervisionada	108
2. MATERIAL E MÉTODOS	115
2.1. Área de Estudo	115
2.2. Banco de Dados e Testes Hidropedológicos.....	116
2.3. Modelagem Numérica do Terreno: Extração de Covariáveis e suas Relações com a Paisagem na Bacia	117
2.4. Aquisição e Processamento dos Dados Aerogeofísicos e Imagens Óticas Sentinel-2A	118
2.5. Modelagem: Métodos de Seleção e Redução de Dimensionalidade dos Dados, <i>Random Forest</i> e Validação.....	122
2.6. Softwares e Pacotes.....	125
3. RESULTADOS E DISCUSSÃO	125
3.1. Composição das Informações Aerogeofísicas	125
3.2. Análises de Autocorrelação e Componentes Principais.....	128
3.3. Modelagem Espacial	133

3.4. O Método de Seleção Ideal Para a Modelagem: Ponderações e Sugestões	142
4. CONCLUSÕES	144
3 CONCLUSÕES GERAIS	147
4 CONSIDERAÇÕES FINAIS.....	148
5 RECOMENDAÇÕES E SUGESTÕES PARA TRABALHOS FUTUROS.....	148
6 REFERÊNCIAS BIBLIOGRÁFICAS	150
ANEXOS	164
A – PLANILHA DE PARÂMETROS E CÁLCULOS (PERMEÂMETRO DE GUELPH)	164
B – QUADROS DO CAPÍTULO II	167
C - CARTAS GEOLÓGICAS CEDIDAS PELO DRM.....	172

1 INTRODUÇÃO GERAL

1.1 Contextualização

Os recursos naturais são fundamentais para a subsistência da vida na Terra e atuam diretamente na criação de uma série de bens e serviços associados ao ecossistema e bem-estar humano. A gestão cuidadosa desses é essencial para o uso sustentável da terra e um caminho para salvaguardar os serviços ecossistêmicos e a biodiversidade. Porém, o uso desses recursos nem sempre é feito de forma sustentável (FAO, 2015). O desequilíbrio provocado pela exploração incorreta da terra tem causado pressões sobre os solos, levando-os a limites críticos. A exaustão e a perda por erosão são algumas das consequências desse desequilíbrio e afetam as funções ambientais como a pureza, disponibilização e o armazenamento de água e controle de inundações (FAO, 2017).

Não obstante, a complexidade na movimentação do fluxo de água nos solos e a grande variabilidade espacial dos atributos químicos, físicos, hídricos e biológicos dos solos (mineralogia, textura, matéria orgânica e outros) faz com que seja necessária a aplicação de técnicas capazes de adquirir informações com maior precisão e acurácia geoespacial para mapear e modelar os recursos hídricos (RAMÍREZ et al., 2019). Neste sentido, o Mapeamento Digital de Solos (MDS) surge como uma solução para facilitar a coleta, armazenamento, análise e interpretação de dados de solos e, quando aliados a novas tecnologias (sensoriamento remoto, sistemas de informações geográficas, geotecnologias, modernos instrumentos de coleta in situ e georreferenciamento de informações, algoritmos e software de programação), o seu potencial é amplificado, permitindo que haja a quantificação e representação de diversos atributos e classes do solo através de métodos matemáticos computacionais e estatísticos, com acurácia e precisão conhecidas (SANTOS & CATEN., 2015; DA SILVA NUNES & UAGODA, 2020). O MDS é capaz ainda de atender à demanda dos levantamentos pedológicos, obtendo mapas cartográficos em escalas de maior detalhe (semidetalhado e detalhado) àquelas obtidas pelos levantamentos sistemáticos de solos existentes, preenchendo lacunas na ausência de dados medidos através da predição de seus valores pela correlação com as variáveis ambientais determinantes da formação dos solos (ARRUDA et al. ,2013; DA SILVA NUNES & UAGODA, 2020).

Assim, torna-se importante o uso de técnicas de mapeamento e modelagem digital de solos que trazem uma maior elucidação acerca dos atributos químicos, físicos e hídricos do solo que influenciam os processos ambientais em uma dada área de estudo

(RAMÍREZ et al., 2019), como no caso de uma bacia hidrográfica, permitindo entender a dinâmica hídrica e garantindo a manutenção deste recurso no meio ambiente, prevenindo desastres e processos degradantes do solo, da flora e fauna em seu entorno.

O presente estudo tem por objetivo modelar os atributos físico-hídricos dos solos da bacia hidrográfica do rio Guapi-Macacu, no estado do Rio de Janeiro, a saber: a velocidade de infiltração básica de água nos solos (v_{ib}) e condutividade hidráulica saturada (K_{sat}); por meio da integração de técnicas computacionais e estatísticas modernas, gerando análises e produtos cartográficos que permitam a caracterização dos fenômenos hídricos ocorridos da região estudada. A pesquisa serve como apoio na tomada de decisões e conscientização de órgãos públicos e da população acadêmico-científica acerca do manejo e conservação das funções ambientais dos solos em bacias hidrográficas; bem como traz uma abordagem multidisciplinar na modelagem de atributos físico-hídricos dos solos por MDS, integrando técnicas estatísticas, computacionais, sensoriamento remoto e geofísica ambiental.

1.2 Visão Geral da Pesquisa

O presente estudo é estruturado em dois capítulos. O primeiro capítulo (I) trata da modelagem vertical dos atributos físico-hídricos dos solos da bacia de estudo, tendo como base os algoritmos para pedologia quantitativa AQP, para harmonização e descrição de dados, e implementação de Funções de Pedotransferência (FPTs), visando estimar a variabilidade desses atributos em profundidade. Os dados físico-hídricos de velocidade de infiltração básica dos solos (v_{ib}) foram mensurados *in situ* pelo uso de permeâmetro de Guelph no levantamento hidropedológico executado na região, em dois níveis de profundidade – componente Z (camadas entre 0-20 e 20-40 cm) e, posteriormente, foram calculados os valores de condutividade hidráulica saturada (K_{sat}). Complementarmente a estes, a pesquisa contou com um banco de dados legado da região contendo parâmetros químicos e físicos dos solos.

Sob o banco de dados da pesquisa, realizou-se a análise exploratória para identificar características do comportamento dos atributos na relação solo-paisagem da área de estudo e avaliar quais os dados pedológicos que representam melhor as variáveis de estudo. Associado à esta, foi feito um estudo de caracterização hidrológica da bacia hidrográfica através da extração de parâmetros topográficos (morfométricos) e geométricos (índices) que avaliaram a distribuição hidrológica na região, caracterizando o hidromorfismo (ou não) dos solos.

Destaca-se neste capítulo a metodologia adotada na separação do banco de dados, fundamentada no hidromorfismo dos solos e na profundidade das camadas analisadas. As FPTs implementadas basearam-se no ajuste de equações regressivas (regressão linear múltipla) e modelos machine learning usando algoritmos baseados em árvores (Árvores de Regressão e *Random Forest*). A modelagem dos atributos por pedotransferência serviu como base para o segundo capítulo (II), permitindo que haja densificação de pontos amostrais a partir das funções de pedotransferência estimadas e aplicadas sob dados que carecem de valores medidos desses atributos (predição), mas que possuem informações de demais atributos passíveis de modelagem (análises granulométricas, químicas e descritivas dos perfis de solos).

O segundo capítulo trata da modelagem espacial dos atributos físico-hídricos, tendo como dados de entrada os pontos preditos pelas FPTs, associados ainda à variáveis oriundas de análises de modelagem numérica do terreno, tendo como alvo a avaliação de Modelo Digital de Elevação (MDE), e radiometria da superfície terrestre, quantificada por dados aerogeofísicos (aeromagnetometria e aerogamaespectrometria) processados do levantamento geológico feito pela CPRM e índices espectrais do sensor Multispectral Instrument (MSI) da missão Sentinel-2A. Agregou-se ainda os valores analíticos de propriedades dos solos relativas aos dados hidropedológicos, definidas no capítulo I, à etapa de modelagem espacial.

Na fase pré-modelagem, foram implementadas técnicas multivariadas de seleção de variáveis, à saber: *Variation Inflation Fator* (VIF), *Stepwise Akaike Information Criterion* (StepAIC) e *Recursive Feature Elimination* (RFE); visando otimizar a escolha das variáveis predictoras (covariáveis). O banco de dados hidropedológicos oriundo das FPTs foi separado na porção 70% e 30%, para treinamento e validação dos modelos *Random Forest*, respectivamente. Os valores de vib e Ksat foram preditos com base no modelo *Random Forest* segundo duas camadas de profundidade dos solos (0-20cm e 0-40cm) e sua variabilidade espacial (componentes X e Y) representada através de mapas da extensão da bacia hidrográfica. A partir dessa modelagem foram obtidos quatro mapas como produtos cartográficos consistentes, precisos e acurados representativos do comportamento espacial dos atributos físico-hídricos dos solos na bacia hidrográfica do Rio Guapi-Macacu.

O processamento de todas as etapas do estudo foi realizado em ambiente RStudio v. 1.3.959, através da criação de rotinas estatísticas (*Scripts*). Para extração de variáveis

analíticas do terreno usou-se o software SAGA GIS v. (Conrad, 2015) e, para análises geoespaciais e geração de mapas, usou-se os softwares QGIS v.3.14 e ArcGIS v.10.6, respectivamente.

Para a análise criteriosa do modelo predito e seu grau de acerto (validação) foram usadas métricas estatísticas de acurácia e precisão dos mapas produzidos (índices estatísticos: coeficiente de determinação - R^2 , *Root Squared Mean Error* – RMSE e *Mean Absolute Error* - MAE).

O fluxograma a seguir ilustra o esquema metodológico resumido adotado na pesquisa (Figura 1).



Figura 1. Fluxograma da metodologia adotada na pesquisa.

2 REVISÃO DE LITERATURA

2.1 O Fluxo de Água e as Propriedades Físico-hídricas dos Solos

É sabido que os solos são corpos naturais que recobrem a superfície terrestre, formados a partir do intemperismo de rochas e seus sedimentos, servindo como meio de desenvolvimento para plantas e subsistência da vida humana e animal. Entretanto, tal material consiste num sistema aberto, ou seja, se desenvolvem e evoluem continuamente, interagindo com o ambiente ao redor (SiBCS, 2018). Segundo a Organização das Nações Unidas para Agricultura e Alimentação – FAO (2015), em função desta natureza dinâmica, o solo assume diversas funções, como mostrado na Figura 2 abaixo:



Figura 2. Funções do Solo. Fonte: FAO (2015).

Dentre as funções citadas, destacam-se aquelas relacionadas ao manejo e preservação dos recursos hídricos pois é notável a essencialidade da água na manutenção da vida terrestre, sendo este um recurso finito.

Um dos objetivos da FAO é a sustentabilidade da água, pois através da sua infiltração e armazenamento de forma eficiente esta é capaz de garantir a sobrevivência de toda a vida na superfície terrestre (ex. irrigação das plantas para fornecimento de

alimentos, consumo humano e animal) e, ainda, regulamentar os fenômenos ambientais (ex. regulação do clima para evitar aquecimento global, manutenção do ciclo da água para garantir o fluxo do sistema hídrico do solo e da atmosfera, drenagem correta da água da chuva para evitar inundações).

Logo, compreender a água e os seus movimentos no solo (fluxo de água) é importante para garantir a existência e a manutenção dessas funções do solo. E para entendê-lo, precisamos primeiro conhecer as propriedades e o papel desempenhado por esse recurso tão essencial para a vida no planeta, que é a água.

O processo de infiltração e armazenamento de água tem origem no ciclo hidrológico, o qual é um processo contínuo de transformação e circulação da água na natureza, passando de um estado para outro, apresentando-se nos estados líquido, sólido ou gasoso ao longo dessa dinâmica, conforme a Figura 3.



Figura 3. Ciclo hidrológico. Fonte: adaptado de Brasil Escola (2020).

Graças ao ciclo hidrológico, a água infiltrada nas camadas superficiais do solo (processo chamado infiltração) e armazenada em seguida no subsolo por meio dos aquíferos (reservatórios subterrâneos de água) compõe o processo total chamado de "fluxo de água nos solos". Esse fluxo é caracterizado pelos movimentos da água internamente no solo e, ainda, pela estrutura do solo. A ocorrência do fluxo de água no solo ocorre preliminarmente graças às características físico-químicas intrínsecas da molécula da água, descritas no Quadro 1 a seguir.

Quadro 1. Características físico-químicas da água.

Propriedade	Descrição
Estrutura Molecular	A distribuição de cargas e as ligações da molécula da água permitem que ela aja como um solvente e que adsorva em superfícies sólidas e hidrate os íons oriundos dos minerais e coloides do solo.
Ângulo da Molécula	Por possuir um ângulo de 104,5° entre os átomos de hidrogênio (estrutura tetraédrica dos orbitais sp ³ do oxigênio), a molécula de água é formada por ligações de pontes de hidrogênio (força de Van der Waals) que garantem o equilíbrio dinâmico da água entre os estados líquido, sólido e gasoso. A característica de um líquido polar permite que esta interaja com demais elementos disponíveis no solo (cátion, íons, partículas) e garantem a movimentação da água em múltiplas direções.
Ângulo de contato	Um ângulo de 180° significaria um completo não molhamento entre a superfície sólida (solo) e a gota d'água, mantendo esta última na sua forma esférica (não sendo possivelmente considerada líquida à temperatura ambiente). Logo, o ângulo de contato de 0° entre a partícula de água e o solo garante uma tensão superficial ideal para que haja o completo molhamento (equilibrado) da superfície das partículas no solo.
Adsorção de água nas superfícies sólidas	Fenômeno oriundo das diferentes forças de atração e repulsão que ocorrem entre as moléculas das diferentes fases (sólida, líquida e gasosa) na sua superfície de contato. Isto gera diferença na densidade do material frente à sua zona (ou fase), permitindo o molhamento das partículas.
Capilaridade	As moléculas de água aderem às superfícies do caule da planta (por exemplo) como resultado de forças elétricas (oriunda da sua polaridade), e essa propriedade, associada à alta tensão superficial, causa a capilaridade. Logo, ela consegue se movimentar pelo solo ao gerar pressão hidrostática nos coloides.
Tensão superficial	Associada com o fenômeno de capilaridade (ascensão), garantindo a diferença de pressão da água e permitindo que ela se mova dentro do solo.
Viscosidade	Força necessária para o deslocamento do líquido nas camadas do solo, sendo proporcional à velocidade de deslocamento das camadas. Sua viscosidade permite a presença de fricção interna nas camadas do solo, deslocando-se fluidamente ideal por este.
Curvatura de superfície e pressão hidrostática	A curvatura da água em uma superfície permite que ela tenha uma viscosidade menor, causando uma menor tensão superficial e menor força para deslocar a água no interior do solo.

Fonte: adaptado de HILLEL (1980); KOOREVAAR (1983); CEDDIA (2020).

No interior dos solos a água se move em função da diferença de potenciais entre pontos dentro do perfil. Esses movimentos se relacionam diretamente com a natureza da molécula da água (propriedades físico-químicas) e com características do solo (propriedades morfológicas e de gênese), os quais podem ser explicados pelo fenômeno hidráulico descrito por equações desenvolvidas por três grandes pesquisadores: Darcy (1856) - solos saturados, posteriormente por Buckingham (1907) - solos não-saturados e, por fim, aprimorado por RICHARDS (1931), os quais regem os princípios da hidropedologia (item 2.2 desta pesquisa).

A água que infiltra no solo se move segundo essas equações, associados ainda aos movimentos exercidos internamente nos solos, conforme mostrado na Figura 4.

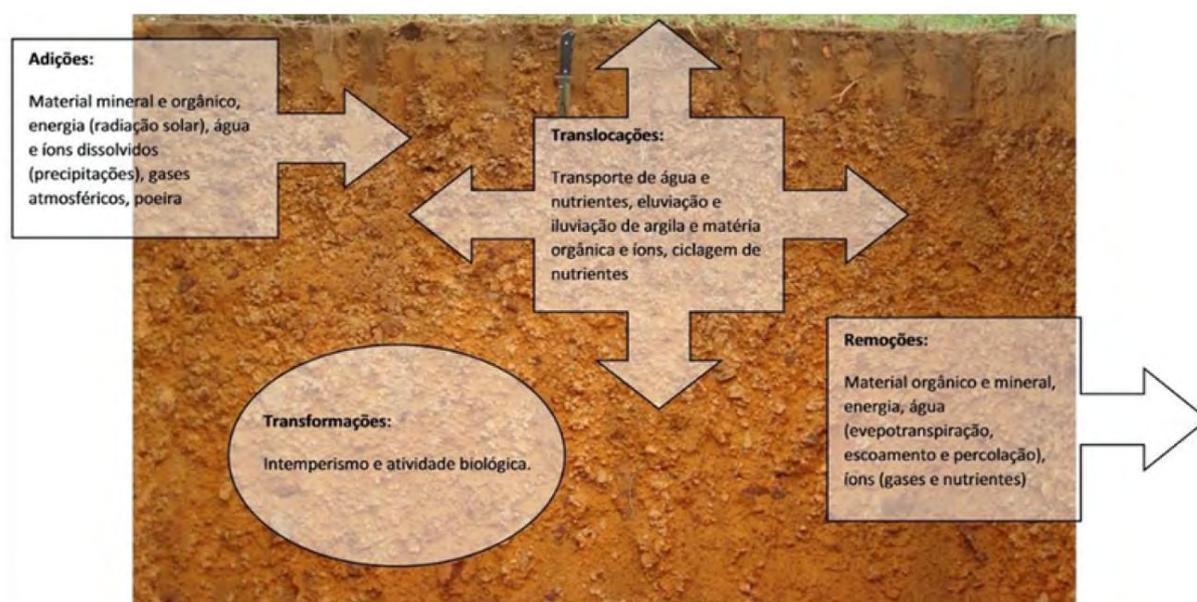


Figura 4. Movimentos e processos internos do solo. Fonte: VEREECKEN et. al (2016).

Complementando as atividades descritas na Figura 4, o processo translocação envolve principalmente a movimentação de colóides dentro do perfil do solo, que são partículas de natureza mineral ou orgânica com diâmetros menores que $1\mu\text{m}$, apresentando cargas quando em superfície e alta superfície específica (relação de razão entre área e massa), associada a sua forma e tamanho das partículas. A movimentação de nutrientes e íons ao longo do perfil do solo decorrem do processo de lixiviação (remoções/perdas de solo) e ciclagem de nutrientes. Já nos processos de transformação, ocorrem sumariamente reações de oxirredução nas quais modificam a matéria orgânica bruta em compostos mais estáveis (fração húmica) por atividade biótica e, ainda,

modificações nas rochas e materiais pedológicos (intemperismo). No processo de remoção, há a ocorrência ainda de erosão, capaz de deteriorar os solos e rochas pela ação de fatores naturais (chuva, água, vento, gelo, clima, outros), gerando deslocamentos desse material para as áreas mais baixas do relevo (altimetria).

O nível de saturação do solo também influencia na movimentação da água, pois os horizontes/camadas saturados (onde a água preenche todos os espaços porosos) já alcançaram a capacidade de armazenamento e tendem a tornar os solos mais vulneráveis a problemas ambientais como deslizamentos de massa. Solos não-saturados possuem seus espaços porosos ocupados por ar e água (50% do sistema trifásico do solo), o que permite que a água flua (movimente-se) devido às pressões hidrostáticas entre os horizontes, porém tal fluxo insaturado não ocorre facilmente (Figura 4). O fluxo de água decorrido de solos insaturados depende de fatores como o potencial total da água e da condutividade hidráulica (no qual, ambos reduzem quanto mais drenado estiver o solo), os quais possuem uma relação intrínseca com a pedogênese.

Tal fenômeno é diretamente associado com as características morfológicas de cada solo, como a textura (composição granulométrica do solo) e a porosidade. Estas propriedades influenciarão no fluxo ao longo do perfil do solo (variação vertical).

No estudo dos aspectos físicos do solo, a dinâmica da água dentro deste sistema é de fundamental importância para seu pleno entendimento. Sob esse ponto de vista, os solos são considerados sistemas trifásicos, com as fases sólidas (minerais e matéria orgânica), líquida (solução do solo: água, nutrientes e demais solutos) e gasosa (ar do solo). Dependendo do arranjo das partículas sólidas (estrutura), o solo terá mais ou menos espaços vazios (poros), que serão eventualmente ocupados ou por água ou por ar. No geral os poros de maior tamanho (macroporos) perdem água mais facilmente, sendo preenchidos pelos gases do solo, enquanto os de menor tamanho (microporos) retêm mais água.

A preservação de agregados formados pelas partículas sólidas é fundamental para manter uma porosidade total do solo próxima do ideal, como apresentado na Figura 5.

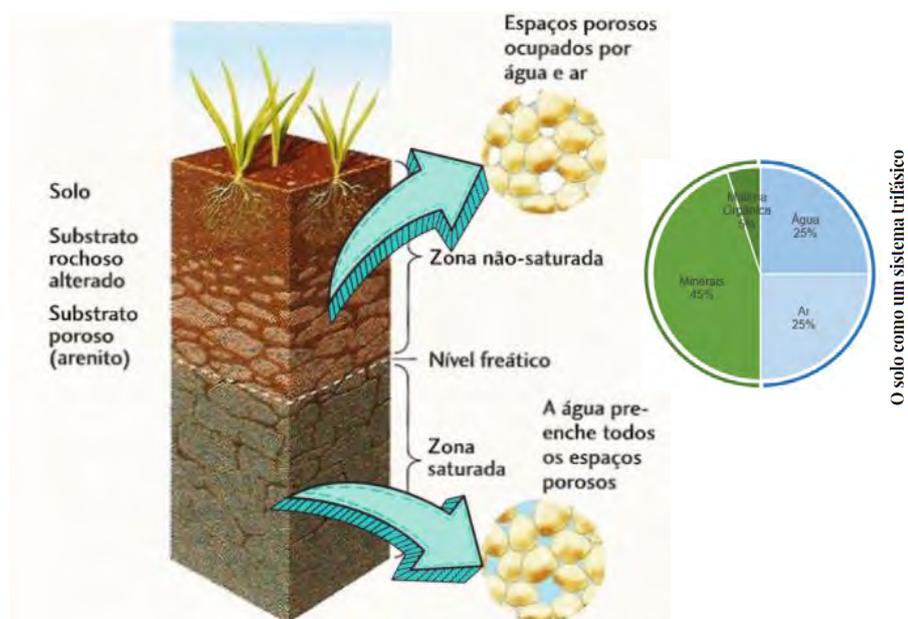


Figura 5. Zonas de saturação do solo e a porosidade dos solos, onde: no círculo à direita estão discriminadas as composições dos poros (em cor azul) e da parte sólida (em cor verde) dos solos. Fonte: adaptado de DOS SANTOS et al. (2018) & VEREECKEN et. al (2016).

Considerando o sistema trifásico do solo (Figura 5), uma distribuição ideal seria 50% de fase sólida (cerca de 45% mineral e 5% orgânica) e 50% poros (cerca de 25% preenchidos com água e outros 25% com ar), garantindo assim uma disponibilidade de água adequada à demanda das plantas, além de uma boa aeração. Entretanto, para que esta água presente no solo esteja realmente disponível para ser absorvida pelas raízes, é necessário que ocorra o fluxo de água do solo para a planta, ou seja, a transferência do material (água) entre diferentes pontos do sistema (solo) em uma unidade de tempo. Esta característica varia bastante de acordo com o tipo de solo. Em solos onde há o predomínio de partículas de areia (mais grosseiras), haverá mais macroporos, que, como visto, perdem a água mais rapidamente pela força da gravidade. Logo, nestas condições mais arenosas, a tendência é haver um fluxo maior de água. Já em solos mais argilosos (partículas sólidas menores), há mais microporos, que tendem a reter mais a água, reduzindo seu fluxo. Porém, tal afirmativa para texturas arenosas são verificadas em condições de saturação do solo, já que à medida que o solo seca, o fluxo cai abruptamente até sua estabilização (nulidade). Por outro lado, em solos mais argilosos, a ocorrência de elevação do fluxo é diretamente dependente da distribuição dos poros (composição granulométrica, tamanhos). Solos com maior fração de argila em sua composição possuem melhor

agregação e podem atingir boa distribuição de macro e microporos em suas camadas em perfil.

Para que haja fluxo de água, necessita-se que haja diferença de energia entre dois pontos quaisquer. Como a água no solo se move de maneira muito lenta, sua energia cinética (associada ao movimento) é desprezível. Por outro lado, a energia potencial da água (relacionada à posição) varia bastante ao longo do perfil do solo, sendo esta variação que possibilita que o líquido flua dentro do sistema solo. A água no solo é submetida a diversas forças que ocorrem no campo (gravidade, atração pela matriz sólida, presença de solutos, gases, pressão da coluna d'água) e alteram seu potencial. Nesse sentido, sua energia potencial total pode ser entendida como o somatório dos potenciais correspondentes a essas forças (Quadro 2).

Quadro 2. Forças potenciais do Fluxo de água no solo.

Forças potenciais	Descrição
Potencial Gravitacional (Ψ_g)	Relacionado à força gravitacional no campo, determinado pela posição da água num determinado ponto em relação a um referencial (geralmente no nível inferior do perfil do solo). Neste caso, qualquer ponto dentro do perfil apresentará potencial gravitacional positivo.
Potencial de Pressão (Ψ_p)	Resultante de uma pressão diferente da pressão de referência. Quando a pressão da água do solo é maior que a atmosférica, seu potencial de pressão é positivo. Os solos apresentam pressão hidrostática positiva em condições de saturação (carga ou lâmina de água). Quando a pressão da água no solo for menor que a atmosférica, seu potencial será negativo. Neste último caso, tem-se o potencial mátrico, um dos componentes do potencial de pressão, detalhado abaixo.
Potencial Mátrico (Ψ_m)	Recebe este nome por sua relação com a matriz sólida do solo, aparecendo em condições insaturadas (maioria dos casos em campo). Quanto mais seco estiver um solo, a água remanescente nele ficará retida nos microporos (por capilaridade) e nas superfícies das partículas (por adsorção) com maior intensidade, exigindo um maior gasto energético para retirá-la deste estado de equilíbrio. Esta atração da água com a matriz do solo origina o potencial mátrico, que é menor que o da água livre.
Potencial Osmótico (Ψ_o)	Componente do potencial de água oriundo do efeito das espécies de solutos presentes na solução do solo. A presença destes elementos na água do solo reduz a pressão de vapor e o potencial energético.

Fonte: adaptado de HILLEL (1980); KOOREVAAR (1983).

Na maioria das condições de campo, os solos encontram-se insaturados, sendo desprezível o potencial de pressão. Da mesma forma, o potencial osmótico só é relevante para o fluxo de água quando há uma membrana seletiva ou barreira de difusão (mais visto na interface solo/raiz), sendo na maior parte do sistema negligenciado. Por isso, na prática os potenciais que influenciam o fluxo são: gravitacional e mátrico. Só haverá o fluxo se houver também diferença entre as energias potenciais totais de diferentes pontos no perfil do solo perfil do solo.

Entendido os conceitos fundamentadores do fluxo de água, podemos então defini-lo como a quantidade de material ou energia que é transferida através de um sistema ou porção do sistema em uma unidade de tempo (taxa). Quando o fluxo se refere à transferência de fluido, este é medido através do volume de fluxo (volume de fluido por unidade de tempo). Quando o fluxo é medido por unidade de área, passa a se chamar densidade de fluxo.

O fluxo de água no solo depende, ainda, das condições de umidade, ou seja, varia para solos saturados e não saturados e, ao longo do tempo, vários modelos matemáticos foram propostos para explicar este fenômeno. Aqui, serão destacados 3 destes modelos, que levam os nomes de seus pesquisadores: Darcy, Buckingham e Richards. O estudo deste movimento de fluidos iniciou-se há mais de um século e um dos trabalhos pioneiros nessa área foi do engenheiro francês Henry Darcy, que em 1856 descreveu o fluxo em experimento com camada de areia saturada. Assim, foi estabelecida a chamada "Lei de Darcy", descrita abaixo (Figura 6 e eq. (1)).

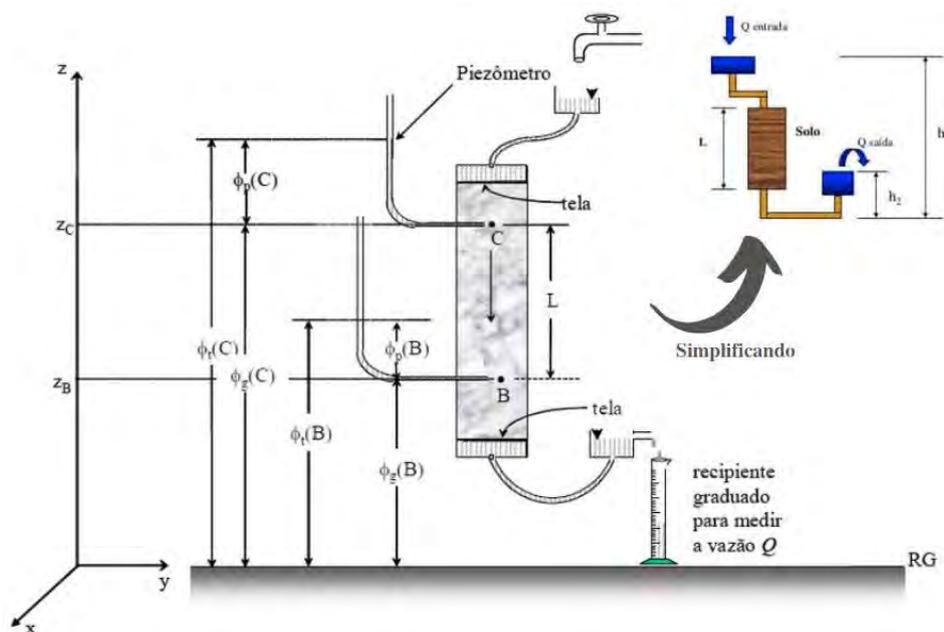


Figura 6. Fluxo da água em solo saturado: diagrama de dinâmica dos fluidos segundo a lei de Darcy (1856) (arranjo experimental). Fonte: adaptado de LIBARDI (2012).

$$q = \frac{Q}{A} = -k_{\theta} \cdot \left(\frac{\Delta H}{L} \right) \quad (1)$$

Onde: Q/A é a densidade de fluxo de água (vazão pela área); q é o fluxo de água (vazão ou taxa); k_{θ} é a condutividade hidráulica; $\Delta H/L$ é o gradiente hidráulico; ΔH é a diferença de energia potencial entre os pontos (h_1-h_2) ou potencial total ($\Delta\Psi_t$); L é a distância entre pontos. Unidade(s): $\text{cm}\cdot\text{s}^{-1}$ ou $\text{mm}\cdot\text{h}^{-1}$ ou $\text{mm}\cdot\text{dia}^{-1}$.

Porém, a Lei de Darcy não previa situações, tais como: (1) condutividade hidráulica deve ser constante; (2) o gradiente hidráulico deve ser constante e (3) o fluxo deve ser laminar, pois caso contrário o fluxo perde sua proporcionalidade com o gradiente hidráulico; tornando-a limitada para alguns tipos de situações (insaturação, por exemplo) pois, como bem sabemos, solos mais arenosos apresentam poros de diâmetro maior, logo são os mais suscetíveis a apresentar o fluxo não laminar (LIBARDI, 2012).

Assim, ao longo dos anos este modelo matemático foi aprimorado para se adequar às condições de insaturação (como dito anteriormente) predominantes em situações de campo.

Em 1907, Buckingham propôs uma extensão da lei de Darcy, considerando um fluxo de água insaturado. Nesta equação, conhecida como Darcy-Buckingham, a

condutividade hidráulica (K) é definida em função da umidade volumétrica (θ) (que varia em solos insaturados) (Figura 7 e eq. 2).

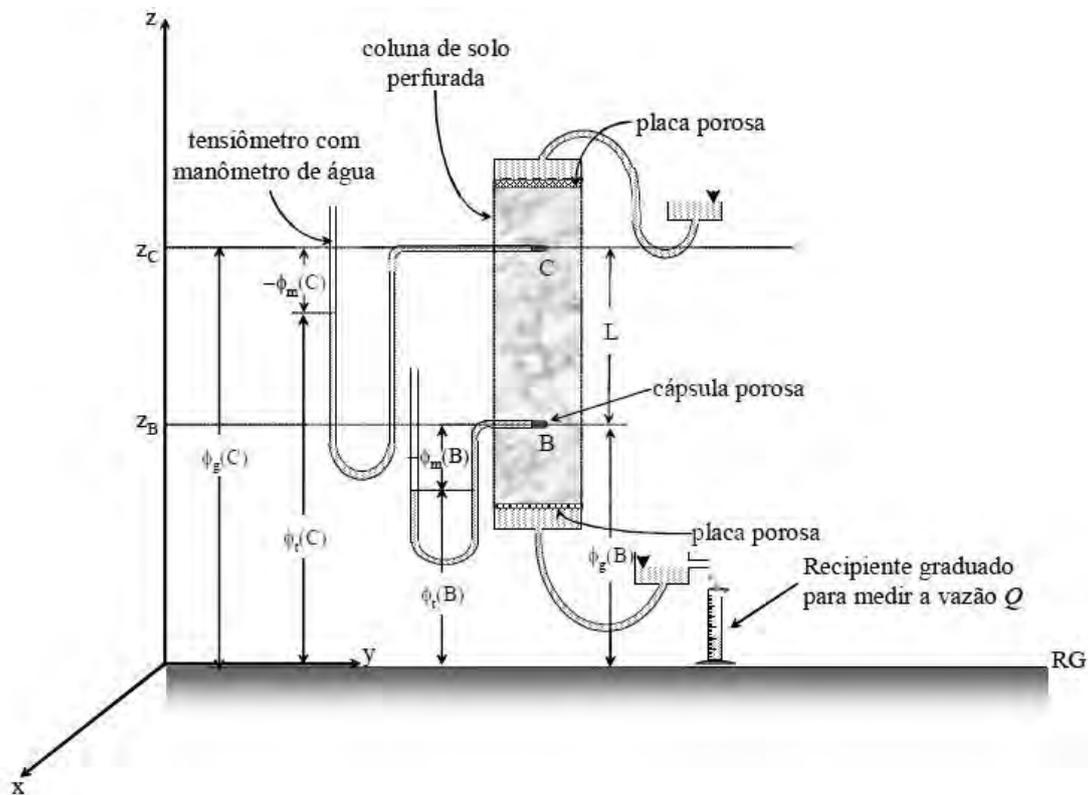


Figura 7. Fluxo da água em solo não saturado: diagrama da equação de Darcy-Buckingham (arranjo experimental). Fonte: LIBARDI (2012).

$$q = -k_{\theta} \cdot \left[\frac{\partial(\psi + z)}{\partial z} \right] = -k_{\theta} \cdot \frac{\Delta\Psi_t}{L} \quad (2)$$

Onde: k_{θ} é a função condutividade hidráulica e $\Psi_t = \Psi_m + \Psi_g$ (potencial total em função dos potenciais mátrico e gravitacional); θ é a umidade volumétrica do solo; L representa sempre o comprimento de solo ao longo da direção do movimento de água. Cálculos feitos sob a coluna de solo na horizontal e inclinada.

Vale ressaltar que esta fórmula é válida para fluxo horizontal ou para quando o efeito da gravidade é desprezível. Sob condição de saturação, todos os poros estão preenchidos com água e não existem meniscos (interface água/ar) e a adsorção é considerada nula. Nesta situação o componente matricial é igual a zero e o Potencial Total da Água do Solo (ou carga piezométrica, como chamava Darcy) é determinado por outros componentes como potencial osmótico e o gravitacional (em função de z). Logo, tem-se que: $\theta = \theta_s$ (conteúdo de água de saturação), $K(\theta_s) = K_0$ (condutividade hidráulica do

solo saturado) e $\Psi_t = \Psi_{gz} + \Psi_o$ (potencial total em função dos potenciais gravitacional – piezométrico - e osmótico). Na ocorrência da saturação, o potencial total acaba por se igualar ao potencial gravitacional, já que o potencial matricial é anulado e o osmótico é apenas considerado quando não há membrana semipermeável dentro do solo.

Posteriormente, Richards em 1931 adaptou o modelo Darcy-Buckingham e associou-o à equação de continuidade, gerando uma equação que descreve o movimento de água em solos não saturados, considerando um fluxo vertical. Isto porque, para descrever uma situação de regime transiente (ou não estacionária) de um fluido, tem-se que levar em consideração a equação da continuidade (eq. 3) que, matematicamente falando, estabelece o princípio de que não pode haver nem criação nem destruição de massa, isto é, deve haver a conservação de massa no volume de solo (Figura 8) (RICHARDS, 1931; BACCHI, 1988).

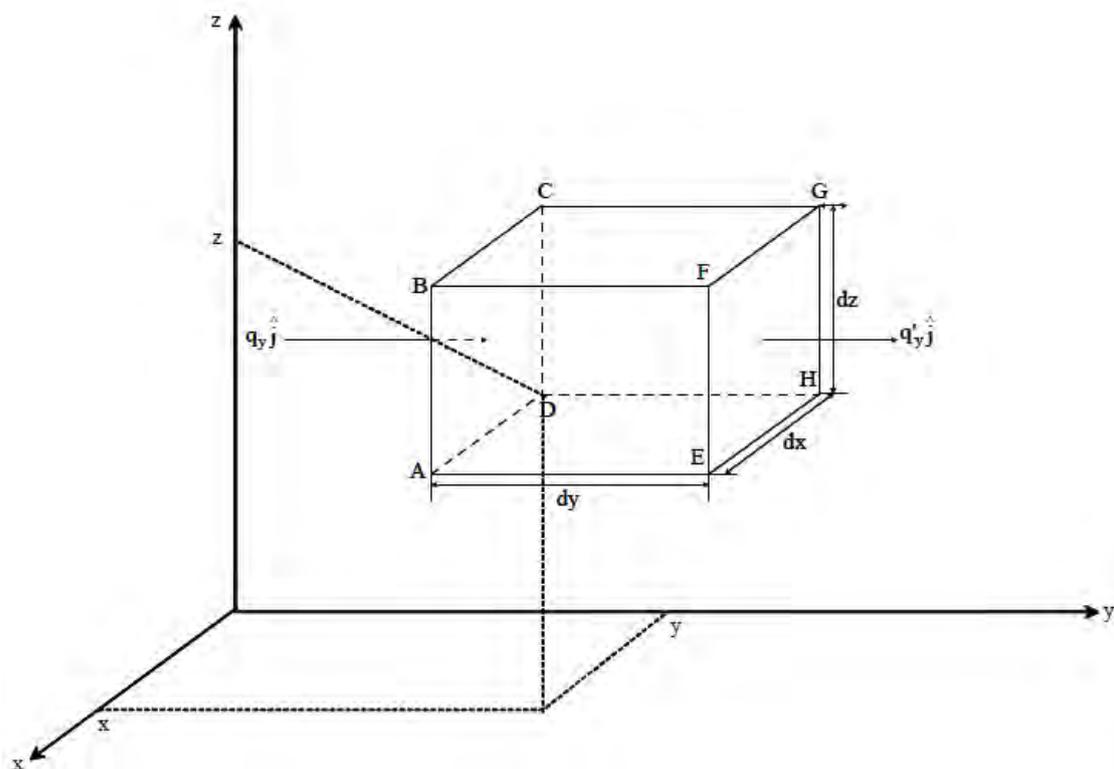


Figura 8. Elemento de volume de solo, através do qual a água está fluindo. Fonte: LIBARDI (2012).

$$\frac{\partial \theta}{\partial t} = - \left(\frac{\partial q_x}{\partial x} + \frac{\partial q_y}{\partial y} + \frac{\partial q_z}{\partial z} \right) = - \frac{\partial q}{\partial s} \quad (3)$$

Em que: $\partial\theta/\partial t$ é a taxa de variação instantânea do conteúdo de água à base de volume θ no volume de solo ∂V ; S é o comprimento da coluna de solo posicionada numa direção qualquer no espaço; q é a vazão ou densidade de fluxo.

A equação anterior (eq. 3) é denominada de “equação da continuidade para a densidade de fluxo da água no solo”, calculada a partir da diferencial parcial de θ com relação a t (tempo), porque o elemento é de volume e, portanto, as coordenadas x , y e z , permanecem fixas no espaço enquanto t varia (LIBARDI, 2012; QUEIRÓZ, 2017).

Logo, combinando-a com a equação de Darcy-Buckingham, temos a equação de Richards, dada pela fórmula abaixo (eq. 4, 5 e 6).

$$\frac{\partial\theta}{\partial t} = \frac{\partial}{\partial x} \left[k_{\theta} \cdot \left(\frac{\partial\psi_t}{\partial z} \right) \right] \quad (4)$$

(caso horizontal, onde $x=y$ já que $\partial\Psi_g/\partial x = \partial\Psi_g/\partial y = 0$)

$$\frac{\partial\theta}{\partial t} = \frac{\partial}{\partial z} \left[k_{\theta} \cdot \left(\frac{\partial\psi_t}{\partial z} \right) \right] \quad (5)$$

(caso vertical - z)

ou, ainda,

$$\frac{\partial\theta}{\partial t} = \frac{\partial}{\partial z} \left[k_{\theta} \cdot \left(\frac{d\psi_m}{dz} + 1 \right) \right] \quad (6)$$

(caso vertical, quando o fluido é a água)

Onde: $\partial\Psi_g/\partial z = \mu_{\text{água}}=1$ (o fluido é a água, logo sua massa específica vale 1g/cm^3). A equação é expressa em potencial J.N^{-1} ; $d\psi_m$ é o potencial mátrico e $\partial\Psi_g$ é o potencial gravitacional da água; k_{θ} é a condutividade hidráulica no regime transiente, em função da umidade volumétrica θ .

O fluxo de água se dá sempre do maior para o menor potencial e depende da condutividade hidráulica e do gradiente hidráulico. A condutividade hidráulica (k_{θ}) se refere à facilidade de um solo transmitir água e pode ser defini-la como um volume de água (cm^3) que passa numa unidade de área (cm^2) numa unidade de tempo (s) sob um gradiente de potencial. Tal parâmetro é dependente da umidade do solo, de modo que quanto maior a umidade, maior será a condutividade, logo esta será máxima quando o

solo estiver em condição de saturação, com todos os poros ocupados por água (umidade igual à porosidade total). A condutividade depende basicamente de dois fatores: permeabilidade intrínseca do solo e fluidicidade, onde:

(1) Permeabilidade intrínseca do solo: é uma propriedade relativa aos poros do solo e descreve a condição na qual o meio poroso transmite água e outros fluidos;

(2) Fluidicidade: depende exclusivamente dos fluidos do solo; é inversamente proporcional à viscosidade (quanto mais viscoso for o fluido, menor será a sua fluidicidade).

A condutividade hidráulica (k ou K_{sat}) é influenciada ainda por características físicas do solo como: textura, estrutura, estado de umidade, além das características inerentes ao fluido (LAWALL, 2010).

Considerando um estado de saturação, um solo arenoso (textura) ou com bom desenvolvimento de agregados (estrutura) apresentará maior condutividade em relação a um solo argiloso ou com fraco desenvolvimento de agregados. Isso se dá porque solos com textura arenosa ou com boa agregação apresentam grandes quantidades de poros de maior diâmetro (macroporos), que, como já visto, favorecem um maior fluxo de água.

Ao contrário, em condições de insaturação (predominantes em campo), os solos mais argilosos ou com agregação menos desenvolvida, apresentarão maior quantidade de água retida nos poros de menor diâmetro (microporos), enquanto os solos mais arenosos ou com boa agregação terão boa parte dos seus espaços porosos ocupados por ar, o que causa uma descontinuidade no fluxo, reduzindo a condutividade hidráulica.

O outro fator que influencia o fluxo de água no solo é o gradiente hidráulico, que se refere à diferença de energia potencial total entre dois pontos (ΔH) dividida pela distância entre estes pontos (L) - $\Delta H/L$. Este parâmetro é considerado a força motriz do fluxo, uma vez que sem gradiente não há movimento de água no solo, lembrando que tal movimento se dá sempre do maior para o menor potencial total, em busca do equilíbrio.

A condutividade hidráulica pode ser obtida de forma direta ou indireta em laboratório:

(1) O método indireto tem como base a correlação entre o conteúdo das partículas finas do solo e a condutividade hidráulica do solo, plotando-se os dados de % de argila + silte em um gráfico que relaciona estes com os valores de condutividade hidráulica;

(2) O método direto pode ser feito: (a) no laboratório, onde amostras de solo coletadas em campo (de preferência indeformadas) são acondicionadas em um recipiente

de dimensões conhecidas, saturadas com água e colocadas em um sistema denominado Permeâmetro; (b) No campo, onde podem ser aplicados os métodos Piezômetro ou Poço Raso à depender do nível do lençol freático (preservam as características naturais do solo) ou o uso do Permeâmetro de Guelph (carga constante ou variável) diretamente sob o solo (preservação das características naturais do solo amostrado, requer maior amostragem e tempo de trabalho).

Outro atributo físico-hídrico importante na estimativa do fluxo de água é a velocidade de infiltração do solo (vib), determinada pela medição de uma dada quantidade de água infiltrada na superfície do solo por um intervalo de tempo, percorrida ao longo do perfil do solo, na qual decresce com o tempo e alcança um valor constante (POTT & DE MARIA, 2003). A vib depende diretamente do umedecimento do perfil, ou seja, em perfis mais secos poderão ser mensurados valores de vib maiores nos instantes iniciais da aplicação de um teste de infiltração, do que em perfis já umedecidos. Existem inúmeros métodos capazes de determinar a vib do solo, como, por exemplo, testes de infiltração usando infiltrometro, que podem ser de aspersão (simulador de chuva) (ROTH et al., 1985; CHAVES et al., 1993; ALVES SOBRINHO, 1997), de tensão (PERROUX & WHITE, 1988; REYNOLDS & ELRICK, 1991; COOK & BROEREN, 1994; BORGES et al., 1999) ou de pressão (REYNOLDS & ELRICK, 1990; ELRICK & REYNOLDS, 1992) ou permeâmetro (ELRICK et al., 1989; REYNOLDS et al., 1992). Este atributo torna-se um valioso aliado nas análises de drenagem e conservação do solo e da água, auxiliando tomadores de decisão em projetos conservacionistas associados à susceptibilidade à erosão hídrica, mau dimensionamento de projetos de irrigação e drenagem na agricultura de precisão e deslocamentos de massa em terrenos acidentados (POTT & DE MARIA, 2003).

É importante frisar a correta aplicação dos métodos de estimativa da vib visando a aplicação na qual se destina o estudo, pois para análises como as citadas acima, por exemplo, o emprego de métodos que não consideram o impacto da gota água da chuva podem superestimar os valores de vib, sendo essencial o uso da técnica em condições de precipitação (SIDIRAS & ROTH, 1987).

Assim como visto para a Ksat, a vib também é influenciada por atributos do solo associados à morfologia, textura e física do solo, como por exemplo: tamanho e grau de agregação das partículas, porosidade total, densidade do solo, tipo de cobertura da superfície do solo, umidade do perfil, quantidade de matéria orgânica, entre outros

(BERTONI & LOMBARDI NETO, 1990; EVERTS & KANWAR, 1992; REICHERT et al., 1992); sendo esses dois parâmetros físico-hídricos influenciados, ainda, pela variabilidade espacial da área de estudo (KLAR, 1984).

Neste estudo, é feita a análise hidropedológica dos solos à partir da mensuração das propriedades físico-hídricas do solo, a saber: velocidade de infiltração básica e condutividade hidráulica saturada, seguindo os princípios do fluxo de água para a modelagem vertical e espacial da água nos solos.

2.2 Hidropedologia: o Teste de Infiltração de Guelph e a Modelagem de Atributos

Físico-hídricos dos Solos

As simulações de fluxo de água em solo não saturado (equação de Richards) descrevem o movimento da água no solo ocorrido pelo processo chamado de infiltração. Este processo, como citado anteriormente, inicia-se quando a água entra no solo e perdura enquanto houver disponibilidade da mesma na sua superfície terrestre (REICHARDT & TIMM, 2004), ou seja, enquanto houver água disponível para a infiltração na camada superficial do solo (interface pedosfera-hidrosfera) e quando tal não atingiu a sua saturação. Lawall (2010) descreve esse estudo hidrológico dos solos como atribuição da hidropedologia:

“A pedologia, ramo da ciência do solo, integra e quantifica a formação, a distribuição, a morfologia e a classificação dos solos; enquanto a hidrologia trata da ocorrência, distribuição, circulação e propriedades da água nas camadas da superfície terrestre (pedosfera). A interação água-solo age, em múltiplas escalas, no desenvolvimento do solo, resultando em grande variabilidade espacial de dados na pedologia. Essas interações, por sua vez, controlam a qualidade e a quantidade de água nos sistemas superficiais e subsuperficiais estudados na hidrologia. (LAWALL, 2010)”

De forma a estudar o potencial de um solo em movimentar e armazenar água (hidropedologia), torna-se necessário o uso de ferramentais capazes de mensurar as variáveis físico-hídricas dos solos (DOS SANTOS et. al, 2008). O Permeâmetro de Guelph destaca-se como um desses ferramentais, tratando-se de um equipamento desenvolvido por REYNOLDS & ELRICK (1983) e aperfeiçoado posteriormente pela University of Guelph, no Canadá em 1985 (ELRICK, 1985), usado para medição da

velocidade de infiltração básica dos solos (v_{ib}) e, indiretamente através da função matemática de solução da equação de Darcy e suas evoluções (DARCY, 1856; REYNOLDS & ELDRICK., 1985; RICHARDS, 1931), pode-se obter a condutividade hidráulica saturada do solo (K_{sat} ou K_{θ}). Este possui diversas aplicações, tais quais em sistemas de irrigação, de drenagem e estudos hidropedológicos em geral (NISHIYAMA et. al, 2007).

A partir do método do permeâmetro de Guelph é possível mensurar ainda o coeficiente de armazenamento (adimensional) (α) que representa a percentagem de água gravítica que coexiste em um determinado volume de água nos solos, onde o potencial matriz do fluxo da água é a capacidade com que o solo tem de absorver a água através do efeito de capilaridade sob as colunas d'água do equipamento. Os três fatores citados (v_{ib} , α e a capilaridade) regem a movimentação de um fluido por meio de um meio poroso não saturado, enquanto a K_{sat} equivale a medida de alcance de saturação do solo, ou seja, quando o fluxo se torna constante (REYNOLDS et. al., 1986).

O ensaio pode ser feito com aplicação de carga constante (solos com $K_{\theta} > 0,60$ cm/h) ou variável (solos com $K_{\theta} < 0,60$ cm/h). O equipamento conta com duas colunas d'água, de diâmetros diferentes, de forma a permitir medições em qualquer posição do perfil do solo. Nele é aplicado inicialmente um volume constante de água, o qual é comandado pela equação da vazão do regime permanente (vazão de água do ensaio vezes a área do permeâmetro).

O referido equipamento constitui uma ferramenta válida para medições de permeabilidade saturada de solos argilosos e arenosos, tendo como vantagem a sua rapidez na determinação de tal medida e a facilidade de operação. Porém, tal equipamento alcança apenas até as camadas de profundidade menor que 60 cm abaixo do nível do terreno (referência de nível). A partir deste nível, é necessário a estimativa do fluxo através de outras abordagens.

Entretanto, CEDDIA (2020) ressalta um ponto importante acerca deste tipo de mensuração:

“Outro fator que influencia de forma marcante os valores da condutividade hidráulica é a alta variabilidade espacial desta propriedade no campo, sendo necessário um grande número de amostragens para a obtenção de valores mais precisos nas medições laboratoriais.”
(CEDDIA, 2020)

Uma estratégia para a ampliação do conjunto amostral na estimativa do fluxo de água nos solos em profundidades abaixo de 60 cm é a implementação de Funções de Pedotransferência (FPTs), cuja fundamentação é pautada na calibração de equações matemáticas baseadas em princípios estatísticos e de programação, conhecidos como algoritmos, capazes de estimar as propriedades físico-hídricas ao longo do perfil do solo (estimativa de transferência vertical do fluxo de água). É importante destacar também a importância da harmonização dos dados como etapa de pré-processamento na aplicação das FPTs, que pode ser feita através de pacote estatístico que reúne Algoritmos para Pedologia Quantitativa (AQP) (BEAUDETTE et al., 2013).

O pacote AQP mostra-se uma ferramenta poderosa para a análise vertical de amostras de solo (perfis), permitindo verificar a relação entre as propriedades dos solos e seus parâmetros físico-hídricos e, ainda, a sua variabilidade no terreno, desempenhando funções de análise em topo-sequências, harmonização de grandes conjuntos de dados em profundidade nos perfis, fatiamento de horizontes, observações qualitativas visuais de classes e atributos dos solos, entre outros.

No entanto, PINHEIRO et al. (2016) e MCBRATNEY et al. (2011) afirmam que no tangível à variação vertical da composição granulométrica, são necessários estudos mais aprofundados que amplifiquem o entendimento desta propriedade nos perfis dos solos. Então, para o desenvolvimento da modelagem vertical em mapeamento digital de solos, torna-se importante entender a relação entre essas propriedades em profundidade para que sejam escolhidas as variáveis que melhor representem a variabilidade da área estudada (MCBRATNEY, 2002; MCBRATNEY, 2003; LAGACHERIE et al., 2006; CEDDIA et al, 2009). Neste contexto, o AQP surge como uma alternativa pré-modelagem viável para estes fins, empregando esforços pedométricos que propõem a aplicação de funções de profundidade para representação da variabilidade de propriedades do solo em perfil, conforme análises mostradas em PONCE-HERNANDEZ et al. (1986), MINASNY et al. (2006), MISHRA et al. (2009) e WIESE et al. (2016).

Ainda, a ferramenta serve de apoio para a harmonização dos horizontes do solo, etapa preliminar importante na estimativa e calibração de equações denominadas de funções de pedotransferência (FPTs). No Brasil, autores como ZINN et al. (2005), CARVALHO JUNIOR et al. (2015), PÁDUA et al. (2015), PINHEIRO (2015) e PINHEIRO (2016) estudaram as relações entre a composição granulométrica, a densidade

do solo e o estoque de carbono orgânico para o fatiamento, harmonização e estimativa de funções de profundidade nas camadas superficiais do solo.

As funções de pedotransferência ou FTPs são equações aplicadas na estimativa de propriedades do solo as quais carecem de informações nos levantamentos pedológicos devido à natureza morosa e custosa desses levantamentos (alta demanda de equipamento, tempo, mão-de-obra especializada, transporte e custos associados a coleta em campo e análise laboratorial), como a densidade aparente (do inglês, *bulk density* - ρ_b) (NANKO et al., 2014; SOUZA et al., 2016), retenção de água (TOMASELLA et al., 2003; BOTULA et al., 2012; BOTULA et al., 2014), capacidade de campo – CC- e ponto de murcha permanente – PMP- (OTTONI FILHO et al, 2007; ANDRADE & STONE, 2011; RODRIGUES et al., 2011), estoque de carbono (ZINN et al, 2005; DON et al., 2007; JEUNE, 2012; GOMES et al, 2017; CIDIN et al, 2019) água disponível total (ADT) (van den BERG, 1996) e condutividade hidráulica (van GENUCHTEN, 1980; ABDELBAKI et al., 2009; SILVA, 2018). Como os dados sobre solos são provenientes da leitura de relatórios de levantamentos e de repositórios online (banco de dados) e as informações dos atributos físico-hídricos são escassas, transfigura-se imprescindível a implementação de FPTs para modelagem e análise pedométrica (SOUZA et al., 2016).

As FPTs são funções preditivas que relacionam mais facilmente os dados mensurados de solos como resposta ao preenchimento de lacunas geradas pela inexistência de informações de propriedades do solo (não mensuração), que ocorre devido a diversas questões intrínsecas da região de estudo, sendo amplamente usada por cientistas no passar das últimas três décadas (BOTULA et al., 2014). Conforme descrito por BOUMA (1989), o termo “funções de pedotransferência” descreve a intenção de “traduzir os dados que temos para o que precisamos”

As FPTs podem ser obtidas a partir de dois tipos de modelos. A primeira possibilidade é o emprego de modelos empíricos baseados em equações clássicas baseadas no fluxo de água (DARCY, 1856; REYNOLDS & ELDRICK., 1985; RICHARDS, 1931) e em estimativa de curvas de retenção de água (SCHEINOST et al., 1997; ALMEIDA et al., 2008; SOUZA et al, 2017; HAGHVERDI et al, 2018; BRUNING et al., 2019).

A segunda, trata da implementação de modelos paramétricos ou não paramétricos de classificação baseados em regressão estatística (VERECKEN et al., 1990; RAWLS & BRAKENSIEK, 1998; WÖSTEN et al., 1999; VERECKEN & HERBST, 2004),

geoestatística (WILDING & DREES, 1993; ROMANDO & SANTINI, 1997; HERBST et al, 2006; AQUINO et al., 2009; MISHRA et al, 2009) e aprendizado de máquinas (SCHAAP, 1998; SCHAAP et al., 2001; PACHEPSKY & RAWLS, 2004; NEMES et al., 2006a,b; 2008; TWARAKAVI et al., 2009). Neste último, a classificação (extração de informação em imagens para reconhecer padrões e objetos homogêneos) pode ser feita de forma supervisionada ou não supervisionada, a depender do conhecimento que o analista possui sobre a área estudada. Ainda, essas funções podem ser baseadas em amostragem pontual, categorizada (determinação de classes de solo) ou contínua (estimativa de funções sem agrupamento de dados e usando o banco de dados completo para derivar as equações).

Segundo WÖSTEN et al. (2001), modelos paramétricos baseados em Regressão Linear Múltipla (RLM) são considerados os mais simples e são mais comumente usados para estimativa de FPTS em ciência do solo. Dentre os diversos modelos já implementados na comunidade científica, JALABERT et al. (2010) & TRANTER et al. (2007) citam as Florestas Aleatórias (do inglês, *Random Forest*) como o modelo não paramétrico que ainda não possui tanta abrangência de uso na predição de parâmetros como a densidade do solo e a condutividade hidráulica saturada e que apontam significativas contribuições para o mapeamento de solos em profundidade.

CAPÍTULO I

Modelagem Vertical dos atributos Físico-Hídricos dos Solos

Este capítulo deu origem ao artigo nomeado “Predição de atributos físico-hídricos dos solos através de funções de pedotransferência e algoritmos para pedologia quantitativa” ou, do inglês, *Predicting Soil Physical-Hydric Attributes Based on Pedotransfer Functions and Algorithms for Quantitative Pedology*, fruto desta dissertação de mestrado.

RESUMO

O conhecimento acerca dos atributos físico-hídricos dos solos torna-se um fator importante para estudos voltados a compreensão hídrica em bacias hidrográficas, onde o conteúdo de água armazenado e disponibilizado afeta tanto as funções ambientais dos solos, quanto a biodiversidade e a sustentabilidade desse recurso natural. No Brasil, os bancos de dados de solos possuem poucas informações acerca de parâmetros hídricos dos solos tais como a velocidade de infiltração básica (vib) e a condutividade hidráulica saturada (Ksat), devido à não realização sistemática de testes de infiltração ao se executar os levantamentos pedométricos e a dificuldade de mensuração de tais parâmetros nas camadas mais profundas. Neste contexto, o objetivo tange a estimativa da vib e da Ksat associadas as propriedades granulométricas e físico-químicas dos solos, por meio da implementação de funções de pedotransferência, análise geoespacial e algoritmos de machine learning capazes de modelá-los verticalmente. Para isto, foi feita uma análise quantitativa dos perfis dos solos por meio do uso de algoritmo para pedologia quantitativa (AQP) e, em seguida, implementadas três funções de pedotransferência usando Regressão Linear Múltipla, Árvores de Regressão e *Random Forest*. O AQP mostrou-se uma ferramenta potencial para estudos em profundidade dos atributos físico-hídricos, evidenciando a variabilidade dos parâmetros do solo e sua implicação sob o comportamento da vib e Ksat. Pelas métricas estatísticas avaliadoras dos modelos implementados, o modelo que mostrou melhor desempenho na estimativa dos atributos foi o *Random Forest* ($R^2=0.9409$, MAE= 0.0015 e RMSE= 0.0033 – Ksat; $R^2= 0.9466$; MAE= 0.0902; RMSE=0,1922 - vib), observado na validação. Conclui-se que modelos baseados em árvores são capazes de modelar atributos físico-hídricos dos solos e, neste estudo específico, modelos regressivos perdem acurácia na estimativa devido à incapacidade de abranger a alta variabilidade destes atributos na interface pedosfera-hidrosfera e por dependerem de pressupostos estatísticos mais rígidos, diferentemente dos modelos baseados em árvores.

Palavras-chave: Geoprocessamento. Hidrologia dos Solos. Aprendizagem de Máquinas.

ABSTRACT

Understanding soils physical-hydric attributes is an important factor for studies focused on investigating how the water content stored and available in hydrographic basins affects soils environmental functions, biodiversity and sustainability of this natural resource. Brazil's available soils databases has little information about soil-water parameters, such as basic soil infiltration rate (*bir*) and saturated hydraulic conductivity (*Ksat*), due to non-systematic performance of soil infiltration tests during pedometric surveys and due to difficulty in measuring such parameters in the deepest soil layers. Thus, the current study aims to estimate *bir* and *Ksat* associated with soil granulometric and physicochemical properties, based on pedotransfer functions implementation, geospatial analysis and machine learning algorithms capable of vertically modeling them. Soil profiles quantitative analysis was carried out by using algorithms for quantitative pedology (AQP) and, next, three pedotransfer functions were implemented using multiple linear regression, regression trees and Random Forest. Results showed AQP as a potential tool for physical-hydric soil attributes in-depth studies, as well as highlighted soils parameters variability and their implication on *bir* and *Ksat* behavior. Based on statistical metrics, validation process showed Random Forest as the best performance model at *Ksat/bir* estimation ($R^2 = 0.9409$, MAE = 0.0015 and RMSE = 0.0033 - *Ksat*; $R^2 = 0.9466$; MAE = 0.0902; RMSE = 0.1922 - *bir*). Models tree-based were capable of modeling soils physical-hydric attributes, whereas regressive models presented low estimate accuracy due to their inability to cover the observed high variability in these attributes, as well as, unlike tree-based models, they depend on stricter statistical assumptions.

keywords: Geoprocessing. Soil Hydrology. Machine Learning.

1. INTRODUÇÃO

As constantes mudanças na dinâmica de uso e cobertura do solo enfatizam a necessidade de uma maior compreensão dos fenômenos relacionados à superfície terrestre, mediante a análise espacial de dados e monitoramento contínuo, procedimentos dos quais são possíveis graças a aplicação do mapeamento digital e da modelagem, a partir de covariáveis ambientais e dados de campo (SANTOS et al., 2018). O conhecimento aprofundado acerca das variáveis ambientais de entrada em modelos preditivos torna mais fidedignas e robustas as informações produzidas, garantindo que os produtos cartográficos sejam representativos à realidade da área estudada (KRAEMER, 2007).

Na intenção de se conhecer a dinâmica hídrica de uma bacia hidrográfica, García-SINOVAS et. al. (2001) destacam a velocidade de infiltração básica e a condutividade hidráulica dos solos como os atributos do solo mais relevantes para entendimento dos processos de infiltração e movimentação do fluxo d'água nos solos, pois são parâmetros requeridos para a predição do fluxo de água nos solos, apresentando elevada variabilidade espacial e dependência da estrutura do solo, propriedades as quais condicionam tal variável. Desta forma, é essencial a determinação destes para auxiliar o estudo hidropedológico em bacias hidrográficas.

O presente estudo tem como proposta modelar os atributos físico-hídricos dos solos da bacia hidrográfica do rio Guapi-Macacu, no estado do Rio de Janeiro, a saber: a velocidade de infiltração básica de água nos solos (v_{ib}) e condutividade hidráulica saturada (K_{sat}); através da implementação de funções de pedotransferência utilizando modelos regressivos e de aprendizado de máquina baseado em algoritmos de árvores.

2. MATERIAL E MÉTODOS

2.1. Área de Estudo e Caracterização Hidrológica

A bacia hidrográfica do Rio Guapi-Macacu (BHRGM) situa-se na região metropolitana do Rio de Janeiro, compondo a área conhecida como Baía de Guanabara, definida pelo Instituto Estadual do Ambiente – INEA (INEA, 2019) como integrante da Região Hidrográfica V (RH-V). Localizada no Sudeste do estado, a bacia interconecta os municípios de Guapimirim, Cachoeiras de Macacu e Itaboraí; e povoado de Guapiaçu. A bacia possui área de captação de aproximadamente 1.250 km², delimitada por um perímetro de 199,2 km. O clima na região é tropical chuvoso com invernos secos (Aw) - pela classificação de Köppen-Geiger (PEEL et al., 2007), com precipitação média anual

variando de 1.200-2.600 (contrafortes oriundos da Serra Verde Imperial) e temperatura média anual em torno de 23°C (HORA et al., 2010). A bacia tem sua origem datada por atividade tectônica iniciada no tempo geológico Terciário, ocorrendo ao processo de sedimentação com deposição aluvial, fluvial e lacustrina – Formação Gráben da Guanabara (FERRARI, 2001).

A vegetação original pertence ao bioma Mata Atlântica, com porções de floresta tropical e vegetação esparsa característica da transição do ambiente de Mar de Morros para a Baixada Litorânea do estado (PINHEIRO et. al., 2016). Conforme SANTOS et al. (2020), a análise sob a caracterização hidrológica da bacia mostra implicações da geometria e morfometria da área sob as medições da velocidade de infiltração básica (vib) e da condutividade hidráulica saturada (Ksat), associado principalmente a presença (ou ausência) de hidromorfismo dos solos. Tais noções podem ser vistas na etapa de resultados e discussões deste trabalho.

2.2. Banco de Dados de Solos e Testes de Infiltração

O banco de dados pedológico é composto por dados físicos e químicos da bacia hidrográfica do Rio Guapimirim-Macacu coletados por levantamento de solos realizado pela Embrapa Solos em parceria com a Fundação de Apoio à Pesquisa e ao Desenvolvimento (FAPED) e a Universidade Federal Rural do Rio de Janeiro (UFRRJ), através do documento “Descrição e Resultados das Análises dos Perfis de Solo Coletados nas Bacias Hidrográficas dos Rios Guapi-Macacu E Caceribu”, Contrato N° 6000.00419115.08.2 (CHAGAS et al., 2011).

As medições físico-hídricas (vib e Ksat) estão associadas aos perfis descritos em CHAGAS et al. (2011) e serviram de complemento para o banco de dados pedológico da região. Tem-se um total de 36 pontos de perfis de solos contendo análise descritiva, morfologia, análises físico-químicas e dados de teste de infiltração. A distribuição dos pontos pode ser observada na Figura 9.

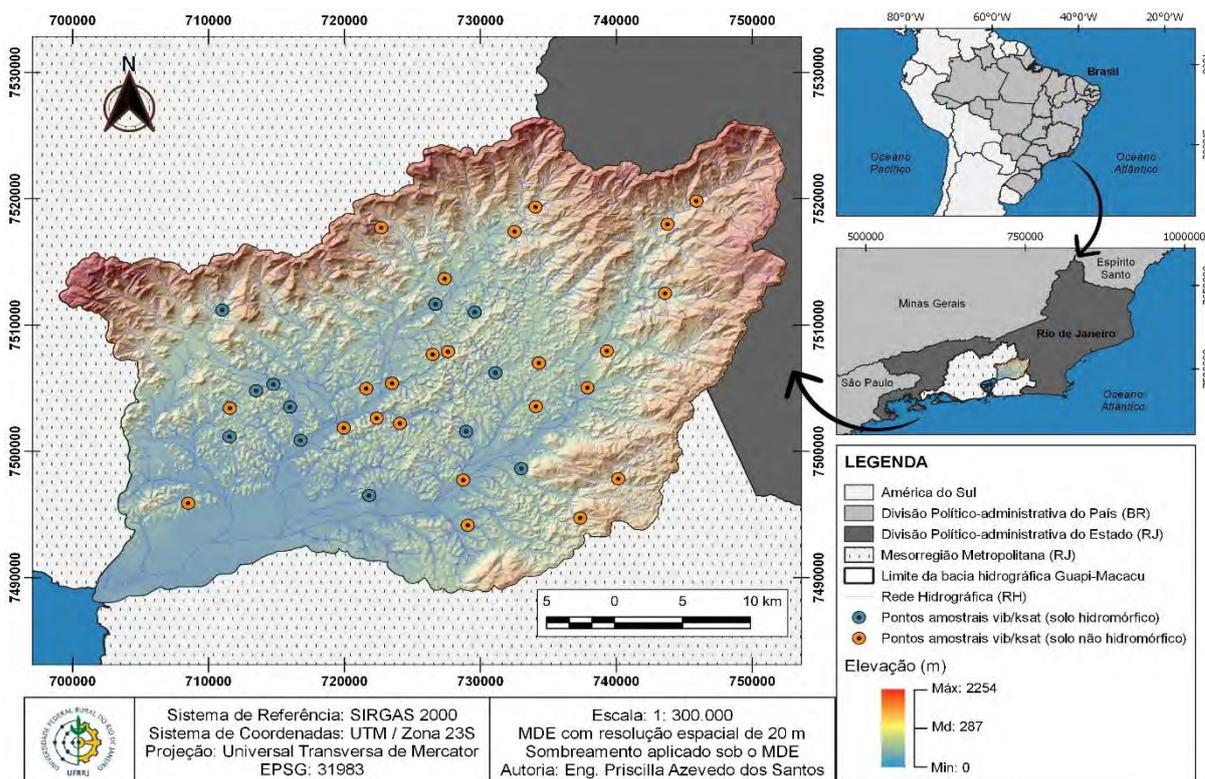


Figura 9. Mapa de distribuição dos pontos amostrais de vib e Ksat na bacia.

Os dados hidropedológicos de vib foram coletados com o auxílio de um permeâmetro de Guelph (SOILMOISTURE, 2012), sendo medidos em intervalos de 1 a 5 minutos dependendo da vazão do reservatório e da área da seção transversal (A) o qual a medição estava sendo monitorada. A área da seção transversal do reservatório considerada baseia-se nos princípios do permeâmetro e é padronizada, sendo igual a $A=35.22\text{cm}$ para o caso de combinação dos dois reservatórios do equipamento; ou $A=2.16\text{cm}$ para utilização apenas do reservatório interno. A carga hidráulica aplicada foi constante e padronizada para toda a coleta, com o valor igual a 10cm.

Conforme variação do tipo de solo e sua característica hidráulica (solos mais drenáveis ou pouco drenáveis), aplicou-se o método 1 ou 2 (ver planilha do anexo A), com o intuito de evitar constantes preenchimentos dos reservatórios pelo operador. Esta escolha foi necessária porque solos muito drenáveis acabam por absorver a água muito rapidamente e impossibilitam a medição da vib em tempo hábil (medição do ponto de saturação), sendo necessário alimentar o sistema permeâmetro em menor intervalos de tempo, limitando a coleta em campo, já que para isto é necessário interromper a medição. Já os solos pouco drenáveis requerem menos alimentação do permeâmetro com água.

De qualquer forma, a escolha do método 1 ou 2 não altera o valor final obtido para de vib ou Ksat, já que em ambos os casos os parâmetros de cálculo são compensados nas equações de cada um dos métodos, resultando em valores iguais para cada atributo físico-hídrico medido (anexo A).

Para cada ponto amostral foram coletadas duas medidas de vib, separadas por camada do solo (superficial: 0-20 cm e subsuperficial: 20-40 cm), de forma a obter valores representativos das camadas pertinentes aos solos amostrados e permitir uma modelagem mais fidedigna da variabilidade espacial desses atributos nos solos. As medidas foram tomadas no valor médio de profundidade das camadas superficiais e subsuperficiais dos solos de forma a caracterizar tais profundidades, assim como o cálculo da Ksat.

A partir da sua mensuração, os valores de vib foram inseridos na planilha de transformação do fabricante do permeâmetro (anexo A) para o cálculo da Ksat (baseada nas equações de RICHARDS, 1931). Os parâmetros requeridos para o cálculo da Ksat foram: a coluna d'água, o raio do fundo dos buracos tradados no solo (diferentes para as medições superficiais e subsuperficiais), textura do solo testado (variável conforme classificação do solo) e o valor da vib medida, na unidade $\text{cm}\cdot\text{min}^{-1}$.

Os valores obtidos para vib e Ksat podem ser analisados na Tabela 1 a seguir.

Tabela 1. Valores de velocidade de infiltração básica (vib) e condutividade hidráulica saturada (Ksat) medidos nos perfis de solos da bacia hidrográfica do rio Guapi-Macacu.

ID	Coordenadas*		Código do Perfil**	Classe***	Superficial		Subsuperficial	
	X (m)	Y (m)			vib (cm/min)	Ksat (cm/min)	vib (cm/min)	Ksat (cm/min)
1	743585	7512512	AE01	CXbd	0.82	1.54×10^{-3}	1.34	1.67×10^{-3}
2	713496	7504795	P04	GXbd	0.01	3.07×10^{-4}	0.02	4.07×10^{-4}
3	721821	7496505	P05	SXd	0.27	5.08×10^{-4}	0.95	1.18×10^{-3}
4	711558	7501163	P10	SXd	0.01	1.25×10^{-5}	0.01	9.22×10^{-6}
5	716776	7500876	P14	GXbd	0.04	4.99×10^{-5}	0.11	1.01×10^{-4}
6	728741	7497726	P24	LAd	0.26	4.89×10^{-4}	0.38	4.74×10^{-4}
7	745893	7519822	P25	LAd	1.15	2.16×10^{-3}	0.49	6.11×10^{-4}
8	723508	7505385	P42	OXfi	0.04	7.53×10^{-5}	0.23	2.12×10^{-4}
9	729591	7511023	P48	GXbd	0.00	0.00	0.00	0.00
10	711576	7503423	AE49	FXd	0.20	3.76×10^{-4}	0.29	3.62×10^{-4}
11	728958	7501584	AE53	SXd	0.88	1.66×10^{-3}	0.88	1.66×10^{-3}
12	737351	7494728	P57	LVAd	0.21	3.95×10^{-4}	0.56	6.98×10^{-4}
13	729077	7494151	P72	LAd	0.38	1.17×10^{-2}	0.14	2.85×10^{-3}
14	743779	7517979	P73	CXbd	0.74	2.27×10^{-2}	1.10	2.24×10^{-2}
15	715994	7503513	AE95	GXve	0.02	2.49×10^{-5}	0.04	3.69×10^{-5}
16	712708	7517713	ExtraRio16	CXba	2.47	7.58×10^{-2}	1.44	2.93×10^{-2}
17	734090	7503557	ExtraRio519	PA	0.13	3.99×10^{-3}	0.40	8.13×10^{-3}
18	727379	7513670	P84	LVAd	0.06	1.22×10^{-3}	0.39	7.93×10^{-3}
19	740155	7497830	AE03	PA	3.50	4.36×10^{-3}	0.17	2.12×10^{-4}
20	733013	7498627	AE13	SXd	0.00	0.00	1.35	2.74×10^{-2}
21	724080	752215	P77	LAd	0.08	1.63×10^{-3}	0.70	9.73×10^{-4}
22	719950	7501854	P33	LAd	0.25	5.08×10^{-3}	0.35	7.11×10^{-3}

Continua

Continua

Tabela 1. Valores de velocidade de infiltração básica (vib) e condutividade hidráulica saturada (Ksat) medidos nos perfis de solos da bacia hidrográfica do rio Guapi-Macacu.

ID	Coordenadas*		Código do Perfil**	Classe***	Superficial		Subsuperficial	
	X (m)	Y (m)			vib (cm/min)	Ksat (cm/min)	vib (cm/min)	Ksat (cm/min)
23	722380	7502621	P97	LVAd	1.70	3.46×10^{-2}	4.10	5.11×10^{-3}
24	739315	7507953	AE28	PAd	0.06	1.22×10^{-3}	0.44	5.49×10^{-4}
25	734058	7519344	P74	LAd	3.70	7.52×10^{-2}	0.08	9.97×10^{-5}
26	732526	7517407	P18/AE18	LAd	1.15	2.34×10^{-2}	2.58	5.24×10^{-2}
27	734308	7506991	AE71	NXd	0.30	6.10×10^{-3}	0.62	1.26×10^{-2}
28	737880	7505036	T3P3	LVAdc	0.04	8.13×10^{-4}	0.47	9.55×10^{-3}
29	711018	7511184	P51/AE51	RQg	0.35	7.11×10^{-3}	0.05	6.23×10^{-5}
30	726700	7511655	P60	RYbd	0.44	5.49×10^{-4}	0.50	1.02×10^{-2}
31	714773	7505296	P93	RYq	1.50	3.05×10^{-2}	0.75	1.52×10^{-2}
32	721600	7504974	P43	LAd	1.15	2.34×10^{-2}	0.00	0.00
33	726499	7507676	AE94	PAd	0.25	5.08×10^{-3}	0.00	0.00
34	727620	7507898	AE58	LVAd	0.13	2.64×10^{-3}	0.16	3.25×10^{-3}
35	731112	7506225	AE15	GXbd	0.04	8.13×10^{-4}	0.00	0.00
36	708507	7495899	AE81	LVAd	0.50	1.02×10^{-2}	0.00	0.00

*Segundo o Sistema de Coordenadas Cartesianas Bidimensional – Projeção Universal Transversa de *Mercator* (UTM); **Segundo nomenclatura de descrição dos perfis feita por CHAGAS et al. (2011). Para melhor entendimento, ver o Quadro 3; ***Segundo o Sistema Brasileiro de Classificação de Solos (SiBCS).

Fonte: dados de campo.

Aliado às informações das medições pontuais (Tabela 1), tem-se a classificação dos tipos de solos analisados na bacia discriminados no Quadro 3 abaixo.

Quadro 3. Classificação e condição hidrológica dos solos da BHRGM.

Nomenclatura	Classificação do Solo (SiBCS)	Condição hidrológica
CXbd	Cambissolo Háptico Tb Distrófico Típico	Não-hidromórfico
GXbd	Gleissolo Háptico Tb Distrófico Típico	Hidromórfico
SXd	Planossolo Háptico Distrófico Típico	Hidromórfico
LAd	Latossolo Amarelo Distrófico Típico	Não-hidromórfico
OXfi	Organossolo Háptico Fíbrico Tétrico	Não-hidromórfico
FXd	Plintossolo Háptico Distrófico Típico	Não-hidromórfico
LVAd	Latossolo Vermelho-Amarelo Distrófico Típico	Não-hidromórfico
GXve	Gleissolo Háptico Ta Eutrófico Típico	Hidromórfico
CXba	Cambissolo Háptico Tb Álico	Não-hidromórfico
PAd	Argissolo Amarelo Distrófico Típico	Não-hidromórfico
NXd	Nitossolo Háptico Distrófico Típico	Não-hidromórfico
LVAdc	Latossolo Vermelho-Amarelo Distrófico Cambissólico	Não-hidromórfico
RQg	Neossolo Quartzarênico Hidromórfico Típico	Hidromórfico
RYbd	Neossolo Flúvico Tb Distrófico Gleissólico	Hidromórfico
RYq	Neossolo Flúvico Psamítico Típico	Hidromórfico

Fonte: adaptado de CHAGAS et al. (2011).

Conforme a Tabela 1, solos com valores de vib ou Ksat resultantes iguais a zero são devido à limitações associadas à condições de clima e tempo no dia anterior e na data da coleta em campo (chuvas intensas e alagamentos condicionam solos à saturação, levando variável intervalo de tempo até que retornem às condições normais/naturais) e características hidrológicas dos solos (pouca capacidade de drenagem e, conseqüentemente, saturação muito rápida). O Quadro 3 mostra a classificação e condição hidrológica dos solos analisados na BHRGM que, quando associadas à Tabela 1, descrevem o comportamento dinâmico dos valores mensurados dos atributos físico-hídricos nas duas camadas de análise no perfil (superficial e subsuperficial).

Verificou-se que solos já úmidos por ação natural da chuva tiveram decréscimo no valor da Ksat devido ao alcance do ponto de saturação, apresentando assim medições influenciáveis e, conseqüentemente, valores baixos para medições em seu horizonte superficial e nulos para a medição na porção subsuperficial.

2.3. Modelagem por Funções de Pedotransferência para a Determinação de Atributos Físico-Hídricos dos Solos

As Funções de Pedotransferência (FPTs) visam a suprir a ausência de informações contínuas e de mensurações dos atributos ao longo dos perfis dos solos. Sua fundamentação pauta-se na calibração de equações matemáticas baseadas em métodos estatísticos, capazes de estimar as propriedades físico-hídricas ao longo do perfil do solo (estimativa de transferência vertical do fluxo de água), as quais carecem de informações nos levantamentos pedológicos (WADOUX et al., 2020). Desta forma, para o desenvolvimento de funções de profundidade em mapeamento digital de solos, torna-se importante entender a relação entre as propriedades dos solos e suas camadas para que sejam escolhidas as variáveis que melhor representem a variabilidade da área estudada (LAGACHERIE et al., 2006; MCBRATNEY et al., 2011). Neste contexto, é importante destacar também a harmonização dos dados como etapa de pré-processamento na aplicação das FPTs, que pode ser feita através de Algoritmos para Pedologia Quantitativa (AQP) (BEAUDETTE et al., 2013; PINHEIRO et al., 2018; XAVIER et al., 2019).

Segundo Wösten et al. (2001), modelos paramétricos baseados em Regressão Linear Múltipla (RLM) são considerados os mais simples e são mais comumente usados em FPTs. Dentre os diversos modelos já implementados na comunidade científica, Tranter et al. (2007) citam os modelos baseados em Árvores (*Regression Trees e Random Forest*) como métricas não paramétricas de caráter abrangente e positivo na predição de parâmetros como a densidade aparente do solo e a condutividade hidráulica saturada e que apontam significativas contribuições para o mapeamento de solos em profundidade.

Um ferramental capaz de auxiliar a calibração de modelos preditivos (como as FPTs) é a Eliminação Recursiva de Feições (do inglês, *Recursive Feature Elimination - RFE*) (GUYON et al., 2002), cujo método visa a escolha das variáveis preditoras que melhor ajustam o modelo desejado, seja ele regressivo, multivariado ou *machine learning*, removendo os preditores que sejam menos relevantes na predição da variável dependente (variável predita) (JEBARA & JAAKKOLA, 2000). Desta forma, o resultado do RFE é a escolha das variáveis preditoras que otimizam o modelo que se deseja estabelecer (SVETNIK et al., 2004), ao eliminar recursivamente um determinado número de variáveis preditoras por iteração, eliminando ainda dependências (principalmente entre a variável predita e suas variáveis preditoras) e colinearidade (multicolineridade) que podem existir entre as variáveis e que afetam de forma negativa a modelagem

(CARVALHO et al., 1999; KUHN & JOHNSON, 2013). Como métrica, o RFE é capaz de avaliar e minimizar/maximizar um dos três parâmetros estatísticos de avaliação da acurácia e precisão da modelagem por tentativa, sendo estes: *Root Mean Squared Error* (RMSE), *Mean Absolute Error* (MAE) e o R^2 (coeficiente de determinação) (SCHAAP & LEIJ, 1998; SCHAAP, 2004); podendo ainda associar-se ao *Random Forest* com uma etapa preliminar a este, visando otimizar essas estatísticas na modelagem (DARST et al., 2018; GREGORUTTI et al., 2017).

A acurácia e precisão de modelos baseados em regressão e aprendizado de máquinas, são comumente avaliados segundo três métricas básicas: RMSE, MAE e R^2 . A Raiz quadrada do Erro médio (do inglês, *Root Mean Squared Error* - RMSE) (eq. (7)), como o próprio nome sugere, é a raiz quadrada da diferença quadrática média entre o valor real (valor observado) e o valor estimado (valor predito) da variável estudada. Associando tal informação à validação cruzada para cálculo do índice, este é capaz de fornecer objetivamente uma avaliação dos modelos e propiciar a escolha do melhor modelo a ser usado na FPT.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (7)$$

Onde: \hat{y}_i é o valor predito ou estimado para a variável estudada; y_i é o valor real ou observado para a variável estudada; n é o número total de observações na amostra; $y_i - \hat{y}_i = \varepsilon_i$ é o resíduo do modelo.

O Erro Absoluto Médio (do inglês, *Mean Absolute Error* - MAE) (eq. (8)) é uma métrica que fornece a média da diferença absoluta entre os valores reais (observados) e os valores previstos (estimados ou preditos) pelo modelo para a variável estudada. Se não há a presença de outliers ou os seus valores não influenciarem diretamente a precisão do modelo, então o MAE pode ser usado para avaliar o desempenho do modelo. Quanto menores forem os valores para o MAE, maior é a acurácia do modelo ajustado.

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (8)$$

O valor da métrica do coeficiente de determinação (R^2) nos fornece uma ideia sobre o quanto a porcentagem da variabilidade da variável dependente é explicada através das variáveis independentes. Em outras palavras, ele reflete a força do relacionamento

entre a variável predita e o modelo em uma escala de 0 a 1, mediante a avaliação da covariância dos preditores. Portanto, é desejável que um modelo com melhor ajuste possua um valor de R^2 elevado (acima de 70%). A qualidade do ajuste de um modelo de regressão é medida pela estatística R^2 ao se comparar a Soma de Quadrados da Regressão (SQR), que representa a quantidade de variabilidade da variável resposta (y_i) explicada pela reta de regressão, e a Soma de Quadrados Total (SQT), que representa a variabilidade dos valores observados (y_i) ao redor da sua média; obtendo-se, desta forma, a seguinte equação (eq. (9)):

$$R^2 = \frac{SQR}{SQT} = \frac{\sum_{i=1}^n |\hat{y}_i - \bar{y}_i|}{\sum_{i=1}^n |y_i - \bar{y}_i|} \quad (9)$$

A regressão torna-se perfeita quando os valores observados e estimados são iguais, ou seja, quando a Soma dos Quadrados dos Erros (SQE) do modelo é zero. Logo, R^2 se aproxima do valor 1 (um). Já com relação a não significância do modelo, esta é verificada quando os valores estimados para a variável resposta são iguais à média dos valores observados para as mesmas, logo SQR é equivalente a zero e o coeficiente de determinação do modelo se aproxima do valor 0 (zero). O coeficiente de determinação fornece uma métrica de qualidade do ajuste de modelos de regressão linear (simples ou múltipla) e atualmente adotado na avaliação da qualidade de classificadores quantitativos em aprendizado de máquinas.

2.4. Procedimentos Adotados na Modelagem Vertical de Atributos Hidropedológicos

Inicialmente, foi realizada a análise estatística descritiva aplicada aos dados para identificar quais atributos dos solos podem atuar como potenciais variáveis de entrada (potenciais preditores) nos modelos de pedotransferência e, ainda, quais dessas apresentam uma relação intrínseca com os atributos físico-hídricos dos solos (K_{sat} e vib) segundo a separação por hidromorfismo.

O banco de dados com as variáveis independentes foi obtido a partir de CHAGAS et al. (2011) e é composto pelos seguintes atributos físico-químicos e informações: composição granulométrica (Frações de Terra, Calhaus, Cascalho), textura (Areia, Silte, Argila), Argila dispersa em água, porosidade, densidade da partícula, densidade do solo, carbono orgânico, cor do solo (matriz, croma e valor), valor T, pH do solo em água, identificação do perfil (ID), geolocalização (coordenadas UTM X, Y e Z) e classificação dos solos segundo o Sistema Brasileiro de Classificação de Solos (SiBCS) (DOS SANTOS et al., 2018).

Como visto em CHAGAS (2006) e PINHEIRO (2012), a textura (composição granulométrica), a argila dispersa, a porosidade e a quantidade de matéria orgânica (carbono orgânico) nos solos são atributos importantes nos estudos hidrológicos pois estão associados diretamente à infiltração e ao fluxo de água no solo. Outros atributos foram analisados como potenciais preditores importantes para a modelagem, como a densidade, o pH e o valor T dos solos. Os atributos citados foram submetidos à análise exploratória prévia visando avaliar o comportamento destes com as variáveis físico-hídricas, mantendo variáveis cuja correlação de Pearson (R) fosse superior a 0.5.

Ao final do processo, sob o banco de dados pedológico contendo as variáveis preditoras selecionadas foram realizados testes de hipótese para verificação dos pressupostos teóricos assumidos pela análise regressiva multivariável (FERRARI, 1989) que são: comportamento linear entre as variáveis e significância da regressão (teste F de *Fisher-Snedecor*, também conhecido como *ANOVA – Analysis Of Variance*), normalidade dos resíduos (avaliada pelo teste de *Shapiro-Wilk*), homoscedasticidade das variâncias (avaliada pelo teste de Breusch-Pagan) e independência dos resíduos (avaliada pelo teste de *Durbin-Waltson*); e que são importantes para as funções de pedotransferência. Os testes foram avaliados a um nível de significância de $\alpha=5\%$ (nível de confiança de 95%) e os dados faltantes foram removidos.

Os preditores selecionados para compor o banco de dados final, e que se associam teoricamente as variáveis hídricas de velocidade de infiltração de água (vib) e condutividade hidráulica saturada (Ksat), são 10 (dez) atributos do solo, a saber: areia, silte, argila, argila dispersa em água, valor T, potencial hidrogeniônico em água (pH), densidade do solo (ds), densidade da partícula (dp), Porosidade e Carbono Orgânico.

Findada a análise exploratória, prosseguiu-se com a implementação das ferramentas AQP para harmonização, fatiamento e homogeneização dos dados, conforme métricas informadas em BEAUDETTE et al. (2013). Para isto, foram aplicadas as seguintes funções de profundidade (*soil-depht functions - SDF*): (a) função *Spline* – para uniformização dos dados; (b) técnica *Slice-Wise* – para fatiamento e harmonização dos horizontes dos solos (intervalos de 1 em 1 cm) e (c) função *Slab-Wise* – para agregação dos dados mediante os intervalos de profundidade especificados (segundo dois intervalos de 20 cm de profundidade) (Figura 10). Após a harmonização, o resultado são quatro tabelas usadas na modelagem das funções de pedotransferência (Figura 11).

Conforme a caracterização hidrológica da bacia, o banco de dados foi separado conforme sua condição de hidromorfismo. Desta forma, o banco de dados foi separado em solos hidromórficos (12 amostras) e solos não-hidromórficos (24 amostras) e, ainda, por profundidade. Em seguida, foi aplicado *Recursive Feature Elimination* (RFE) (GUYON et al., 2002), com o objetivo de selecionar atributos que mantiveram uma relação mais relevante entre os parâmetros dos solos e os atributos físico-hídricos.

Após a aplicação do RFE, foram implementadas e calibradas as funções de pedotransferência com base em modelos regressivos (Regressão Linear Múltipla - RLM) e árvores (*Regression Trees* – RT e *Random Forest* – RF), sob os dados de vib e Ksat separados por camada de profundidade (0-20cm e 20-40cm) e hidromorfismo do solo.

Os parâmetros e hiperparâmetros abordados em cada método de modelagem são distintos e encontram-se listados nos itens a seguir:

- para a implementação do RFE, usando o banco de dados hidropedológico previamente analisado, através da aplicação da função *rfe* em linguagem R, com os seguintes parâmetros: controle de seleção das variáveis sob a implementação da função *rfeControl*; a métrica usada para minimização do erro foi o RMSE. No controle do *rfe*, os parâmetros aplicados foram: método de validação *Leave-One-Out Cross-Validation* (LOOCV); número de dobras ou número de iterações de reamostragem igual a dez (10) repetições;
- para implementação dos modelos baseados em RLM, implementou-se a função *lm* em linguagem R usado como parâmetro o padrão (*default*) dessa função;
- para implementação do modelo *Random Forest* foi usada a função *randomForest* em linguagem R, com os hiperparâmetros conforme o descrito a seguir: número de variáveis randomicamente amostradas como candidato de cada nó/separação (*mtry*) igual $p/3$, onde p é o número de variáveis observadas em análise regressiva; número de árvores para crescimento (*ntree*) igual a 500; critérios de divisão em cada nó foi o MSE; a profundidade máxima das árvores individuais (*max nodes*) e as amostras mínimas para dividir em um nó interno (*sample size*) foram usados o *default* da função; tamanho mínimo dos nós terminais (*node size*) igual a 5 (adequado para regressão). A importância dos preditores foi habilitada e avaliada na função. Não houve aplicação de ajuste (*tune*) para otimização do *mtry*;

- para modelagem usando Árvore de Regressão (AR) ou *Regression Trees* (RT), implementou-se a função *rpart* em linguagem R usado como parâmetro o método ANOVA, no qual é adequado para avaliar a regressão.

Com os modelos implementados, foram estimados os valores dos atributos alvo para os horizontes dos oitenta e seis (86) pontos restantes do banco de dados total de solos da bacia, sendo considerado para fins desta pesquisa as duas camadas de análise (0-20cm e 20-40cm). Os modelos preditos foram avaliados segundo métricas estatísticas de qualidade e acurácia (MAE, RMSE e R^2), usando a validação cruzada.

Com os resultados da validação, fez-se a escolha dos modelos finais para a estimativa da variabilidade dos atributos físico-hídricos. O processo metodológico adotado pode ser visto a seguir (Figura 10). Todos os procedimentos de processamento, análise estatística e elaboração das funções de pedotransferência foram executados no software R v.4.0.1 e RStudio v. 1.3.959 (RSTUDIO TEAM, 2020).

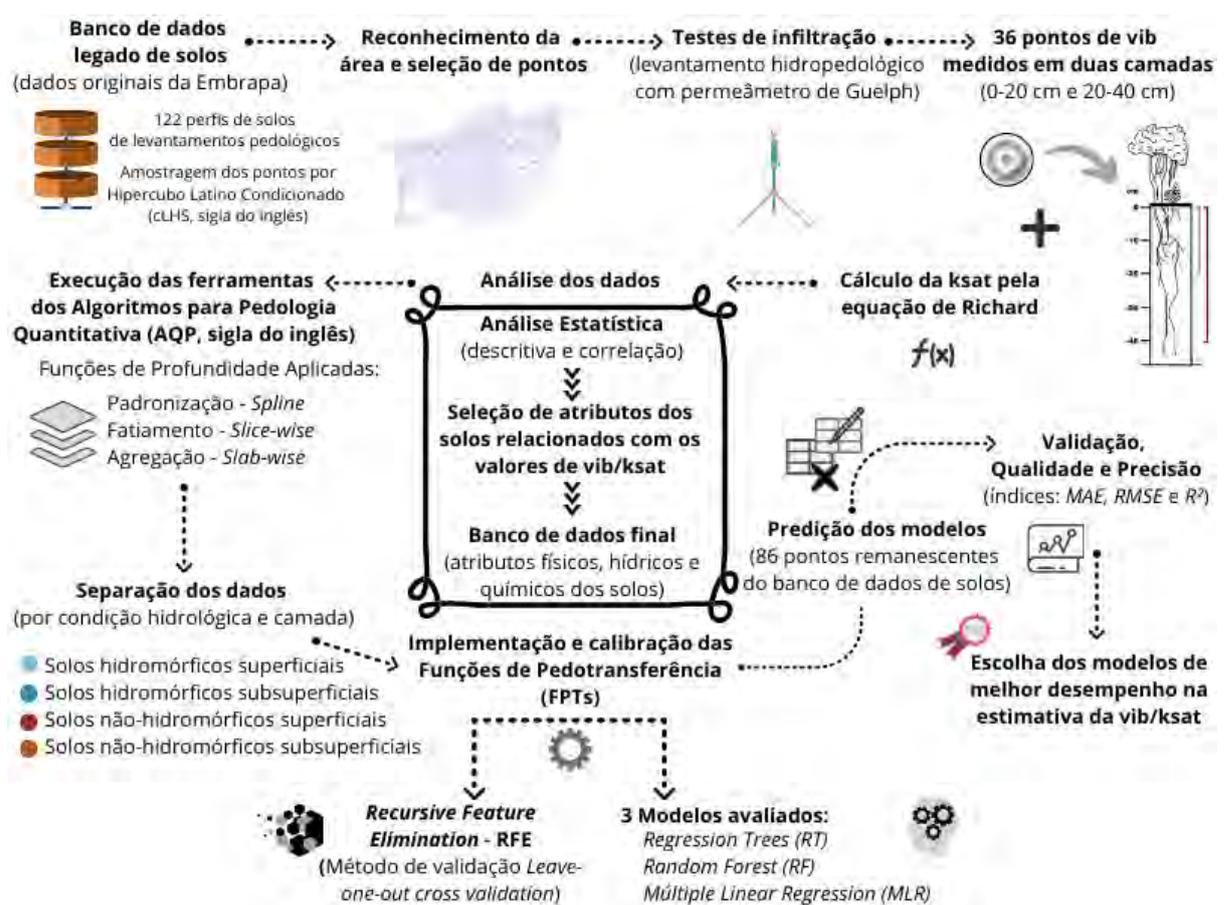


Figura 10. Fluxograma da metodologia da pesquisa.

Diante da característica do banco de dados, relativas à natureza hidromórfica dos solos, e da análise feita para as diferentes camadas, decidiu-se por treinar os modelos separadamente, gerando assim um total de 4 (quatro) funções para cada um dos três modelos propostos (Regressão Linear Múltipla - RLM, Árvores de Regressão - AR e *Random Forest* - RF), que podem ser vistos com detalhes na Figura 11.

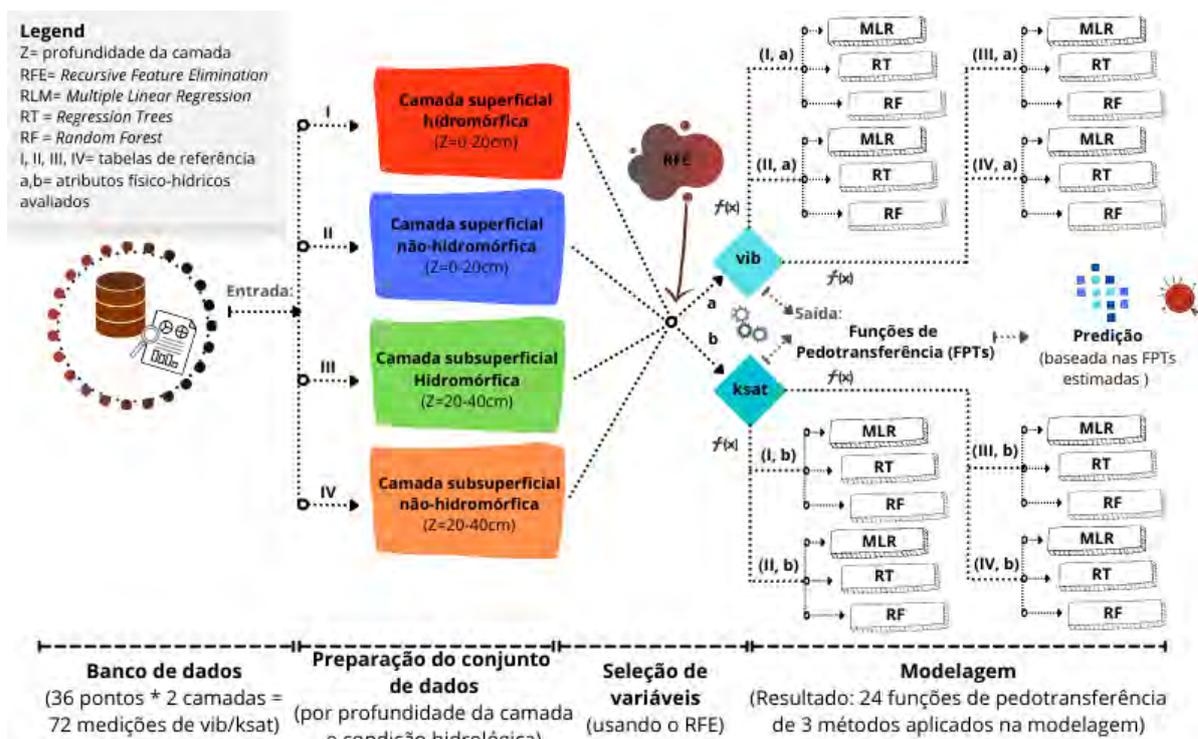


Figura 11. Fluxo de procedimentos adotado na estimativa das funções de pedotransferência para o conjunto de dados.

3. RESULTADOS E DISCUSSÃO

3.1. Caracterização Hidrológica da Bacia do Rio Guapi-Macacu

A bacia hidrográfica é caracterizada hidrológicamente pela existência de três potenciais unidades hidrológicas de proteção, constituída pelas suas três sub bacias, impostas segundo a topografia, rede de drenagem e geomorfologia, conforme o mapa a seguir (Figura 12).

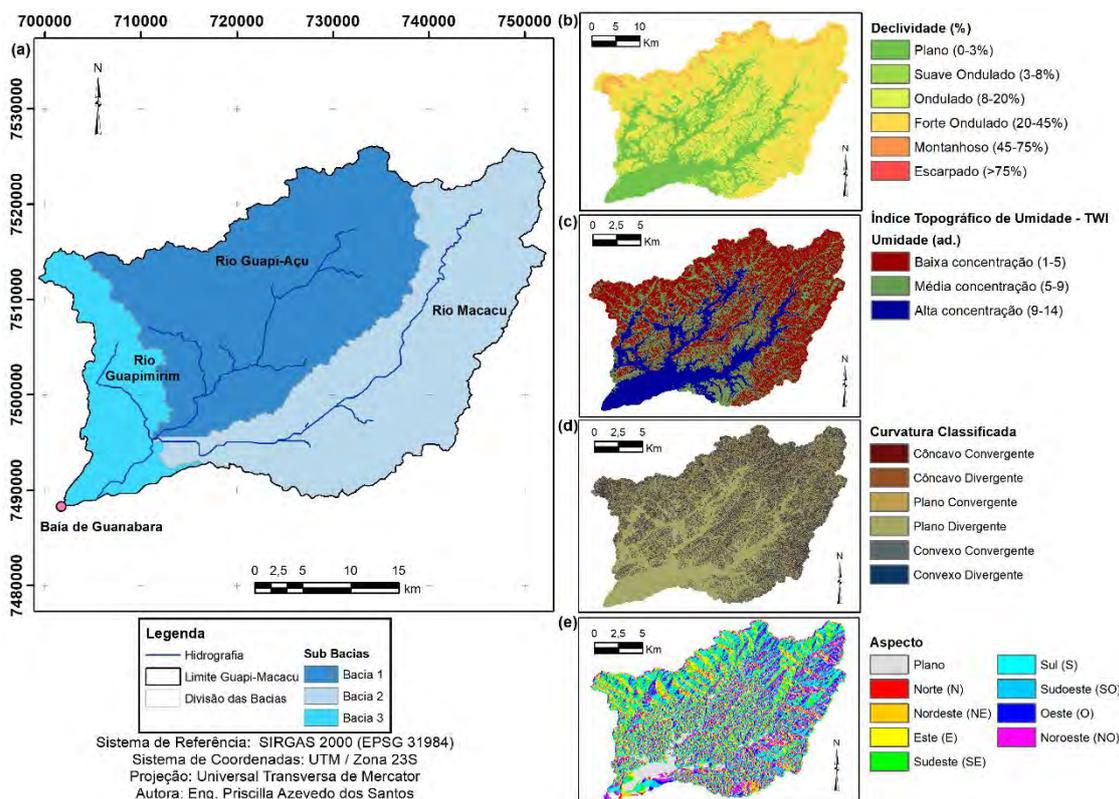


Figura 12. Sub bacias e Morfometria da bacia hidrográfica do Rio Guapi-Macacu, Rio de Janeiro, contendo as seguintes informações: **(a)** Delimitação das unidades hidrológicas de proteção (sub bacias); **(b)** Mapa de declividade (%); **(c)** Mapa de índice topográfico de umidade (do inglês, *Topographic Wetness Index*, TWI); **(d)** Mapa de curvatura classificada (Planar + Perfil); **(e)** Mapa de Aspecto da vertente (pontos cardeais e orientação).

A classificação das variáveis morfométricas que dão origem aos mapas representados na Figura 12 foi feita usando as classes de declividade (fases de relevo) descritas pela EMBRAPA (DOS SANTOS et al., 2018), que subdivide a declividade em seis classes, à saber: Plano (0-3%), Suave Ondulado (3-8%), Ondulado (8-20%), Forte Ondulado (20-45%), Montanhoso (45-75%) e Escarpado (>75%). Para o TWI, classificou-se a concentração de umidade em três classes distintas: Baixa concentração (1-5 ad.), Média concentração (5-9 ad.) e Alta concentração (9-14 ad.) (Prates et al., 2010). A curvatura classificada baseou-se nos mapas de curvatura planar (classes côncavo, planar e convexo) e perfil (classes divergente e convergente) para a bacia do rio Guapi-Macacu, resultando em seis classes distintas (SANTOS et al., 2019). Ainda, a classificação do aspecto da vertente segundo a orientação e posição geográfica resultou em nove classes conforme analisado em SANTOS et al. (2019).

As três sub-bacias classificadas como potenciais unidades de proteção (Figura 12) possuem cada uma as seguintes dimensões de área: 561,05 km² (bacia 1), 528,41 km²

(bacia 2) e 175,57 km² (bacias 3) (Tabela 2). Os parâmetros geométricos para as três sub-bacias e para extensão total da bacia do Guapi-Macacu, avaliados segundo as definições de VILLELA & MATTOS (1975), permitem inferir que a bacia 3 é mais irregular (maior valor de F) e mais susceptível a eventos de inundações com relação as demais devido ao maior valor de coeficiente de compacidade e índice de circularidade em relação às demais bacias analisadas (Tabela 2).

Tabela 2. Parâmetros geométricos da bacia.

Parâmetros	Guapi-Macacu	SB1	SB2	SB3
Padrão de drenagem	D	D	D	D
Área (km ²)	1265,15	561,05	528,41	175,57
Perímetro (km)	203,99	160,72	176,29	112,72
CC ¹	1,61	1,90	2,15	2,38
IC ¹	0,38	0,27	0,21	0,17
Dd (km/km ²)	0,00	0,12	0,14	0,14
IR ¹	0,30	0,28	0,31	0,32
CM (m ² /m)	7,57	8,13	7,25	6,97
L (km)	167,16	69,03	72,93	25,20
F ¹	0,04	0,12	0,10	0,28

¹ad = adimensional; D= dendrítico; CC= Coeficiente de Compacidade; IC= Índice de Circularidade; Dd= densidade de drenagem; IR= Índice de Rugosidade; CM= Coeficiente de Manutenção; L= somatório do comprimento (extensão) de drenagens da bacia ; F= Fator de forma; Sb= Sub-Bacias.

Assim como evidenciado por SANTOS et al. (2012) e HORA et al. (2010), a BHRGM possui, em condições pluviométricas normais, pouca susceptibilidade a ocorrência enchentes. Isto pode ser observado ao comparar os baixos valores encontrados no coeficiente de compacidade (CC) versus os valores atingidos para os fatores formas das sub-bacias (F). Ainda, os resultados obtidos para o Coeficiente de manutenção (Cm) nas sub-bacias indicam que para manter a constância do canal, são necessários 7,57 km² de área de canal disponível para a passagem de água no Guapi-Macacu, 8,13 km² para a bacia 1, 7,25 km² para a bacia 2 e 6,97 km² para a bacia 3, em condições normais de precipitação.

O afastamento das unidades do índice de circularidade (IC) em relação aos fatores de forma (F) (diferença entre valores) nas bacias e sub-bacias, indicam que elas não tendem à forma circular, ou seja, possuem forma mais alongada, o que implica em uma

menor concentração de deflúvio. Esta afirmativa atenta para a espacialidade da distribuição hídrica da bacia na região, caracterizada por uma rede de drenagem bem distribuída.

Ainda, tomando como base STRAHLER (1952) e LOLLO (1995), a densidade de drenagem (Dd) na região é pobre ($<0,5 \text{ km/km}^2$) dada a característica alongada da bacia do Guapi-Macacu frente a sua grande extensão, evidenciados nos mapas de curvatura (plano convergente) e declividade (plana) (Figura 12). Tal característica é corroborada ao se comparar os valores dos índices geométricos de circularidade e rugosidade calculados (IC e IR, respectivamente) para a bacia do Guapi-Macacu com os mapas morfométrico do índice topográfico de umidade, evidenciando a presença de áreas onde os solos têm características hidromórficas na bacia, localizados principalmente nas áreas de baixada da bacia e ao redor da rede de drenagem.

Solos hidromórficos podem influenciar as medições da condutividade hidráulica saturada (SANTOS et al., 2020), tornando-se importante a investigação e calibração das FPTs separadamente para os tipos e condições hidrológicas dos solos da região. Desta maneira, pela característica da bacia, torna-se essencial a separação prévia do banco de dados conforme as condições de hidromorfismo dos solos, procedimentos estes considerados na etapa a seguir.

3.2. Análise Exploratória dos dados

A intenção da realização de análise em estatística descritiva aplicada ao MDS é identificar quais atributos dos solos podem atuar como potenciais variáveis de entrada (potenciais preditores) nos modelos de pedotransferência do solo e, ainda, quais dessas apresentam uma relação intrínseca com os atributos físico-hídricos dos solos (Ksat e vib). Sendo assim, a quantificação de medidas de posição (média, mediana e separatrizes) e dispersão (amplitude, desvio-padrão, variância e coeficiente de variação), assimetria e curtose; e, ainda, a avaliação gráfica sob os dados (histogramas de frequência e gráfico boxplot) tornam mais robustas as informações acerca dessa interrelação entre os as propriedades do solo (composição granulométrica, densidade do solo, porosidade, densidade da partícula e outros) e os atributos físico-hídricos estudados. Junto a estas, uma análise bivariada (autocorrelação) auxilia preliminarmente a visualização das interrelações medidas desses atributos segundo o coeficiente de correlação de Pearson, identificando problemas de multicolinearidade, por exemplo.

Partindo-se da tabela geral dos perfis de Macacu organizada segundo os dados do levantamento pedológico (CHAGAS et. al., 2006), na qual estes estão organizados em horizontes de solos por perfis. Os dados foram filtrados e extraídas as variáveis de interesse. A organização do banco de dados e a descrição dos dados podem ser visualizadas no Quadro 4 a seguir.

Quadro 4. Organização e composição do banco de dados pedológico analítico da dissertação, com as variáveis de estudo e suas respectivas características.

Variáveis	Descrição	Formato
<i>Condutividade Hidráulica Saturada (Ksat)</i>	Atributo que mede a facilidade com que o solo transmite água.	Quantitativa
<i>Velocidade de Infiltração básica (vib)</i>	Processo pelo qual ocorre a entrada de água no solo através da sua superfície. A infiltração decresce com o tempo, dependendo do umedecimento do perfil, e assume um valor constante chamado vib.	Quantitativa
<i>Terra Grossa</i>	Partículas sólidas do sistema trifásico com tamanho maior que 2 mm (esqueleto)	Quantitativa
<i>Terra Fina</i>	Partículas sólidas do sistema trifásico com tamanho variando de 0,002 mm a 2 mm.	Quantitativa
<i>Calhaus</i>	Fração sólida do solo com diâmetro variando de 20 a 200 mm.	Quantitativa
<i>Cascalho</i>	Fração sólida do solo com diâmetro variando de 2 a 20 mm.	Quantitativa
<i>Areia</i>	Fração sólida do solo com diâmetro variando de 0,05 a 2 mm.	Quantitativa
<i>Silte</i>	Fração sólida do solo com diâmetro variando de 0,05 a 0,002 mm.	Quantitativa
<i>Argila</i>	Fração sólida do solo com diâmetro menor que 2 mm.	Quantitativa
<i>Argila Dispersa em água</i>	Desagregação da argila em meio aquoso, na qual afeta a compactação dos solos e no grau de floculação (importante para estudos de erosão e manejo dos solos).	Quantitativa
<i>Porosidade</i>	Volume de vazios ou ainda o espaço do solo não ocupado pela matriz	Quantitativa
<i>Densidade da partícula</i>	Relação entre a massa das partículas e o volume que estas ocupam.	Quantitativa
<i>Densidade do solo</i>	Relação entre a massa do solo e o volume que este ocupa.	Quantitativa
<i>Carbono Orgânico</i>	Quantidade de carbono orgânico existente no solo.	Quantitativa
<i>Cor do solo</i>	Impressão que a luz refletida pelos solos produz no órgão de visão (olhos), traduzido em Valores de Chroma, Hue e Value - Onda refletida do espectro eletromagnético.	Qualitativa
<i>Valor T</i>	Indicador da atividade da argila do solo.	Quantitativo
<i>pH solo (em água)</i>	Correspondente ao potencial hidrogeniônico de uma solução (concentração de íons de hidrogênio - H ⁺). Mede o grau de acidez, neutralidade ou alcalinidade do solo na solução (água).	Quantitativo
<i>Identificador (ID)</i>	Ordem crescente de organização dos dados (Geocódigo).	Qualitativa
<i>Geolocalização (X, Y e Z)</i>	Coordenadas geoespaciais dos pontos amostrais na bacia, no sistema UTM.	Quantitativa
<i>Classificação dos solos (SiBCS)</i>	Classificação dos solos segundo o sistema taxonômico de solos SiBCS (Sistema Brasileiro de Classificação de Solos).	Qualitativa

Fonte: Hillel (1980); Khiel (1979); Kohnke (1968); Libardi (1995); Pott & Maria (2003); Santos et. al. (2010); Sobral et al. (2015); Dos Santos et al. (2018); Barreto et. al. (2019).

Como abordado em CHAGAS et. al. (2011) e PINHEIRO et. a. (2012), a textura, a argila dispersa, a porosidade e a quantidade de matéria orgânica (carbono orgânico) nos

solos são atributos interessantes para estudos hidrológicos pois estão associados diretamente à infiltração e ao fluxo de água no solo. CEDDIA (2020) afirma também que a densidade atua sobre o comportamento dos atributos físico-hídricos, tornando importante a investigação dessas propriedades neste estudo. Desta maneira, estes serão os dados escolhidos para compor o banco de dados para este presente trabalho, associando-se as variáveis hídricas de velocidade de infiltração de água (vib) e condutividade hidráulica saturada (Ksat). No total, são 12 atributos de estudo, à saber: vib, Ksat, areia total, silte, argila, argila dispersa em água, porosidade, carbono orgânico, valor T, pH em água, densidade da partícula e densidade do solo.

Partindo-se do banco de dados de Macacu e, tendo em vista os dados de 36 amostras de propriedades físico-hídricas dos solos obtidos no levantamento hidropedológico realizado na bacia, procedeu-se com a análise exploratória dos dados, envolvendo a avaliação de medidas de posição e dispersão para os dados. O resultado obtido após a elaboração do script deste trabalho pode ser visto na Tabela 3 e Tabela 4 a seguir. A avaliação dos atributos foi feita separadamente por grupamentos (químicos, físico-hídricos e granulométricos) de forma analisar as grandezas em iguais unidades de medida, quando avaliadas as medidas de posição. Já para as análises sob as medidas de dispersão, avaliou-se os atributos de maneira geral.

Tabela 3. Estatística exploratória dos atributos físico-hídricos, da composição granulométrica e da argila dispersa em água na bacia do Rio Guapi-Macacu.

Medidas / Variável	Vib (cm.min ⁻¹)	Ksat (cm.min ⁻¹)	Areia Total (g.kg ⁻¹)	Silte (g.kg ⁻¹)	Argila (g.kg ⁻¹)	Argila dispersa em água (g.kg ⁻¹)
<i>Média</i>	0,668	0,008	545,515	139,641	314,854	88,238
<i>Desvio-padrão</i>	0,917	0,014	173,881	77,645	158,162	114,759
<i>Mediana</i>	0,380	0,002	547,000	124,000	304,000	20,000
<i>Mín.</i>	0,000	0,000	46,000	29,000	40,000	0,000
<i>Máx.</i>	4,100	0,076	931,000	385,000	687,000	407,000
<i>Amplitude Total</i>	4,100	0,076	885,000	356,000	647,000	407,000
<i>Assimetria</i>	2,361	2,806	-0,091	1,358	0,250	1,166
<i>Curtose</i>	5,462	8,923	0,084	1,665	-0,804	0,291
<i>Variância</i>	0,841	~0,000	30234,742	6028,723	25015,302	13169,683
<i>Coefficiente de Variação (%)</i>	137,280	173,574	31,875	55,603	50,233	130,057

Tabela 4. Estatística exploratória dos atributos da bacia do Rio Guapi-Macacu (propriedades químicas pH e Valor T; densidades da partícula e do solo; porosidade e carbono orgânico).

Medidas / Variável	pH (ad)	Valor T ($\text{cmol}_c.\text{kg}^{-1}$)	ds ($\text{g}.\text{cm}^{-3}$)	dp ($\text{g}.\text{cm}^{-3}$)	Porosidade* (%)	Corgânico** ($\text{g}.\text{kg}^{-1}$)
<i>Média</i>	5,042	6,442	1,357	2,523	46,161	9,611
<i>Desvio-padrão</i>	0,446	3,034	0,194	0,090	7,266	9,840
<i>Mediana</i>	4,900	5,500	1,355	2,540	46,815	7,400
<i>Mín.</i>	4,200	2,200	0,620	2,080	24,390	0,900
<i>Máx.</i>	6,500	23,800	1,860	2,640	70,120	80,400
<i>Amplitude Total</i>	2,300	21,600	1,240	0,560	45,730	79,500
<i>Assimetria</i>	1,084	2,460	-0,474	-2,293	-0,059	4,600
<i>Curtose</i>	1,352	9,845	1,939	7,699	1,501	27,795
<i>Variância</i>	0,199	9,206	0,038	0,008	52,788	96,833
<i>Coefficiente de Variação (%)</i>	8,849	47,100	14,330	3,581	15,739	102,383

pH= potencial hidrogeniônico (H^+); ad= adimensional; ds=densidade do solo; dp= densidade da partícula; *relação poro-solo ($\text{cm}^3.100\text{cm}^{-3}$), expressa na unidade porcentagem; **C é abreviação para “carbono”.

Avaliando-se a Tabela 3 e Tabela 4, tem-se que, em relação à média, o conjunto de dados aponta alta dispersão nas variáveis granulométricas (areia, silte e argila) e na argila dispersa devido à valores grandes de desvio padrão, o que nos informa que provavelmente há a presença de valores atípicos observados e, ainda, por ser uma medida que depende de todos os valores observados, a média pode indicar que há outliers (valores atípicos) no conjunto estudado.

Com relação a variância (Tabela 3 e Tabela 4), os dados de textura (composição granulométrica) e argila dispersa em água apontam valores muito elevados, o que pode indicar que o conjunto está com presença de dados discrepantes do normal, elevando tanto a medida do desvio-padrão quanto da variância dos dados, já que um é complementar a outro. Já o atributo Ksat é o que apresentou menor variância, sendo tão pequena que se aproxima de zero ($0,0002034 \text{ cm}.\text{min}^{-1} \sim 0 \text{ cm}.\text{min}^{-1}$), o que indica que os dados estão bem próximos dos valores da média e mediana do conjunto avaliado.

Os atributos que mais se afastam de sua média são os relacionados à granulometria e a argila dispersa em água, sendo o principal deles a areia, na qual possui valor de desvio padrão igual a $173,881 \text{ g}.\text{kg}^{-1}$. O atributo que menos se afasta da média é a Ksat, com valor de desvio padrão igual a $0.014 \text{ cm}.\text{min}^{-1}$.

Com relação à amplitude total dos dados (Tabela 3 e Tabela 4), os atributos apresentaram no geral uma razoável variabilidade, indicando que os diferentes solos que compõe a bacia afetaram de maneira distinta a composição dos dados, mostrando a importância em separar o conjunto de dados na etapa de pré-modelagem, levando consideração critérios intrínsecos destes. Conforme mostrado na seção anterior, um dos critérios importantes para separação desse banco de dados é a condição hidrológica dos solos da bacia e a profundidade (camada) na qual estão sendo avaliados os atributos dos solos, já que a ocorrência e distribuição dos atributos tratados na Tabela 3 e Tabela 4 dependem diretamente da caracterização do solo e da profundidade da camada analisada em perfil do solo.

Com relação a assimetria (Tabela 3 e Tabela 4), os atributos vib, Ksat, silte, argila, argila dispersa em água, pH, valor T e carbono orgânico apresentaram valores acima de zero ($AS > 0$), logo apresentando uma distribuição é assimétrica positiva (à direita). Os atributos areia total, ds, dp e porosidade apresentaram distribuição assimétrica negativa, devido aos seus valores negativos relativos à distribuição de seus conjuntos de dados ($AS < 0$). O ideal é que os dados sejam simétricos com relação as suas medidas de posição (média, mediana e moda), porém no nosso caso observa-se esse afastamento para todos os dados. Não existem dados com distribuição simétrica neste estudo ($AS=0$). Embora os valores mostrem uma assimetria no conjunto de dados dos atributos avaliados, estes não se afastam significativamente na distribuição, porém podem indicar não-normalidade dos dados, que deve ser testada com a aplicação de testes de hipóteses específicos (teste de *Shapiro-Wilk*, por exemplo). O atributo que apresentou maior assimetria foi o carbono orgânico (Corgânico), com $AS=4,600 \text{ g.kg}^{-1}$.

A curtose é a medida que define o grau de achatamento de uma distribuição de dados quando comparada ao de uma distribuição conhecida como distribuição normal. Observando-se os dados (Tabela 3 e Tabela 4), tem-se que a maioria dos atributos (exceto os atributos físico-hídricos, valor T, dp e Corgânico) comportam-se de maneira platicúrtica em sua distribuição, ou seja, os valores de curtose são menores que o valor 3 ($km < 3$), logo a curva ou distribuição é platicúrtica. Já os atributos físico-hídricos (vib e Ksat), o valor T, a densidade da partícula (dp) e o carbono orgânico (Corgânico) possuem valores maiores do que 3 para a curtose ($km > 3$), logo a curva ou distribuição desses

dados é leptocúrtica. Não há dados com curva ou distribuição mesocúrtica ($km = 3$) no banco de dados estudado.

Sobre o coeficiente de variação, medida na qual expressa a variabilidade dos dados de uma variável de modo independente da unidade de medida utilizada, retirando a influência da ordem de grandeza da variável, pode-se inferir que dos atributos estudados, tem-se as seguintes características para os conjuntos de dados: baixa dispersão ($cv < 15\%$) para os atributos pH, ds e dp e, portanto, apresentam maior homogeneidade em sua distribuição; média dispersão ($15\% < cv < 30\%$) evidenciada pela porosidade apenas; e alta dispersão ($cv > 30\%$) evidenciada pelos demais atributos (vib, Ksat, areia total, silte, argila, argila dispersa em água, valor T e Corgânico) e, portanto, apresentam maior heterogeneidade em sua distribuição.

Em seguida, procedeu-se com a análise de correlação dos dados (análise bivariada), tendo como cálculos os dados da matriz de variância-covariância para os atributos dos solos. O resultado da correlação é dado pelo gráfico de autocorrelação a seguir (Figura 13).

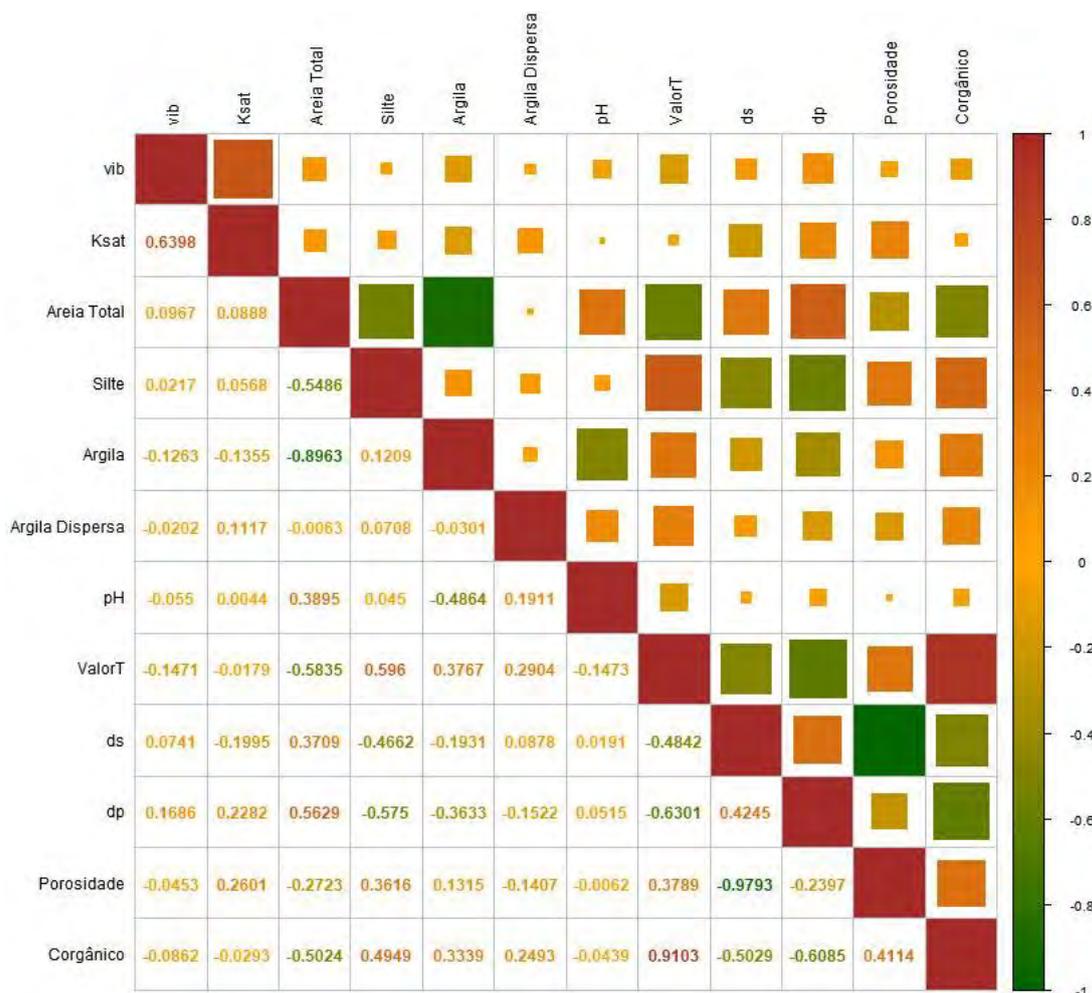


Figura 13. Gráfico de autocorrelação dos atributos dos solos da bacia hidrográfica do Rio Guapi-macacu, segundo análise do coeficiente de correlação de Karl Pearson (ρ ; Pearson, 1904; Mukaka, 2012).

Para a análise da matriz de correlação (Figura 13), usou-se os valores determinados por MUKAKA (2012), os quais associam o grau de correlação à sua interpretação como sendo: correlação insignificante ($0,00 < \rho < 0,30$ ou $0,00 < \rho < -0,30$); baixa correlação positiva (negativa) ($0,30 < \rho < 0,50$ ou $-0,30 < \rho < -0,50$); correlação moderada positiva (negativa) ($0,50 < \rho < 0,70$ ou $-0,50 < \rho < -0,70$); alta correlação positiva (negativa) ($0,70 < \rho < 0,90$ ou $-0,70 < \rho < -0,90$); e correlação positiva (negativa) muito alta ($0,90 < \rho < 1,00$ ou $-0,90 < \rho < -1,00$). Para valores iguais a 1,00, a correlação é dita perfeita positiva com relação direta entre as variáveis, ou seja, quando uma variável aumenta (diminui), outra também aumenta (diminui). Já uma correlação igual a -1,00, esta é chamada de perfeita negativa com relação inversa entre as variáveis, ou seja, quando uma aumenta, a outra diminui seu valor (e vice-versa). A correlação pode ser expressa ainda

em porcentagem de explicação da variável (%), bastando multiplicar o valor por cem (100).

Uma correlação igual a 0,00 indica independência linear entre as variáveis, porém não implica em independência não-linear, devendo esta ser avaliada por outros meios como, por exemplo, o coeficiente de Spearman (SPEARMAN, 1904; BONETT & WRIGHT, 2000).

A Figura 13 indica que há uma correlação de média a forte direta ($40\% < \rho < 80\%$ - cor laranja-avermelhado) entre a variável hídricas (vib e Ksat) e os atributos porosidade e densidade da partícula, o que já era esperado já que tais atributos são citados nas bibliografias como sendo determinantes do regime hídrico dos solos, apresentando-se como atributos de potencial expressividade das constantes hídricas (vib e Ksat). Observa-se a correlação elevada ente os atributos argila e areia total ($\rho > 80\%$ - cor vermelha), porém com comportamento inverso evidenciado. Ou seja, a argila é inversamente proporcional a areia total, o que também é esperado já que estes atributos texturais são complementares na análise de solos. A argila possui menos poros grandes, capazes de reter mais água em sua estrutura e impedir o fluxo de água no solo (infiltração e transmissão). Há ainda uma correlação dita insignificante entre o silte e as variáveis físico-hídricas diretamente (ρ), devido principalmente ao menor tamanho das suas partículas e a baixa coesão entre suas partículas (bem menor que a argila, por exemplo), o que causa instabilidade a longo prazo nos solos, e estes acabam por apresentar facilidade à erosão e desagregação natural.

Há um indício de multicolinearidade entre os valores de areia total e argila e, ainda, de densidade do solo (ds) e porosidade, que podem vir a influenciar as modelagens posteriormente. Desta forma, avaliou-se a distribuição dessas e das demais propriedades através da aplicação da ferramenta AQP visando obter mais informações acerca do comportamento dos dados antes de fazer a seleção das variáveis mais representativas para modelagem.

3.3. Análise por Algoritmos para Pedologia Quantitativa (AQP)

A visualização da variabilidade vertical dos atributos físico-hídricos dos solos da bacia do Rio Guapi-Macacu foi possível a partir do emprego das ferramentas AQP, as quais mostram a distribuição desses atributos e outras propriedades do solo, como a composição granulométrica (areia, silte e argila), argila dispersa em água, valor T, potencial hidrogeniônico em água (pH), densidade do solo (ds), densidade da partícula

(dp), porosidade e o conteúdo de carbono orgânico ao variar a profundidade. Esta etapa se deu após a análise exploratória dos dados para a observação à priori do comportamento das potenciais variáveis preditoras das funções de pedotransferência. A coleção de perfis de solos (Figura 14) analisada pelo pacote AQP apresentou as seguintes características: total de 36 perfis com valores de mínimo 30 cm e máximo de 1,15 cm de profundidade e com variações de altimetria no intervalo de 0 a 900 metros de altitude.

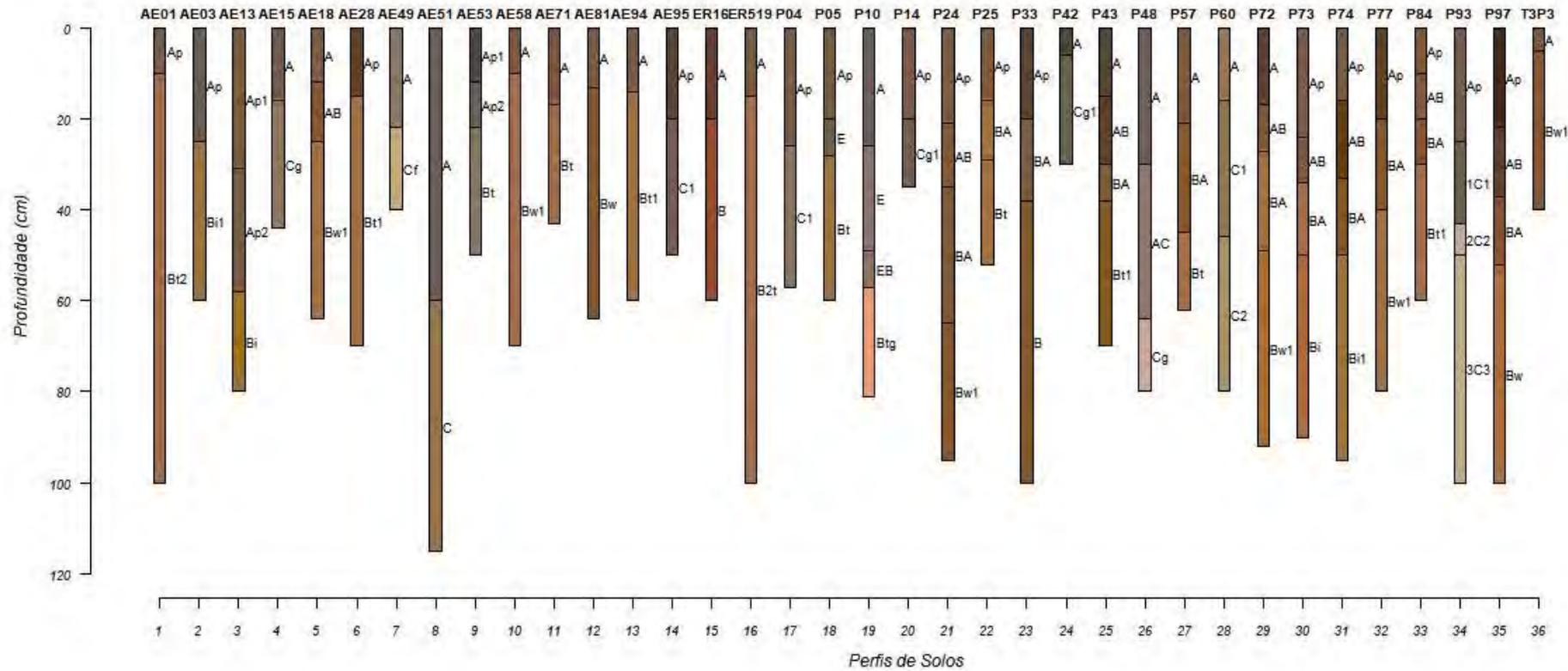


Figura 14. Coleção de perfis de solos da bacia hidrográfica do Rio Guapi-Macacu, Rio de Janeiro. O identificador do perfil corresponde às informações contidas na Tabela 1.

A areia (Figura 15) distribuiu-se homogênea nas camadas superficiais e subsuperficiais ao longo dos perfis, variando de 400-800 g.kg⁻¹. Já o silte (Figura 16) varia aproximadamente entre 150-300 g.kg⁻¹ na passagem das camadas superficiais para as subsuperficiais, atingindo o seu valor mais elevado (400g.kg⁻¹) na camada superficial da amostra P14 (solo hidromórfico, classe GXbd). O conteúdo de argila (Figura 17) varia de 400 a 700 g.kg⁻¹ e tende a aumentar em profundidade. A análise granulométrica desses perfis (Figura 15, Figura 16 e Figura 17) associam o comportamento diferencial das partículas quanto à facilidade de movimentação e fluxo de água no solo (percolação), maiores para argila e silte do que para areia, por maior tamanho de sua partícula e peso associado, nas camadas dos solos indicando assim que a velocidade de infiltração básica (vib) e, conseqüentemente, a Ksat é maior nos horizontes superficiais e subsuperficiais onde a porção granulométrica de areia fina está presente em maior quantidade em relação a areia grossa e a porosidade do solo também cresce nesses horizontes saturados. Observou-se, ainda, que a Ksat aumentou principalmente no limite entre os horizontes superficiais e o subsuperficiais (p~20-35 cm), onde a fração de argila e silte decrescem em proporção ao aumento de areia fina, dando mais mobilidade para a água fluir entre os poros. Verificou-se que solos já úmidos por ação natural da chuva tiveram decréscimo no valor da Ksat devido ao alcance do ponto de saturação, apresentando assim medições influenciáveis e, conseqüentemente, valores baixos para medições em seu horizonte superficial e nulos para a medição na porção subsuperficial.

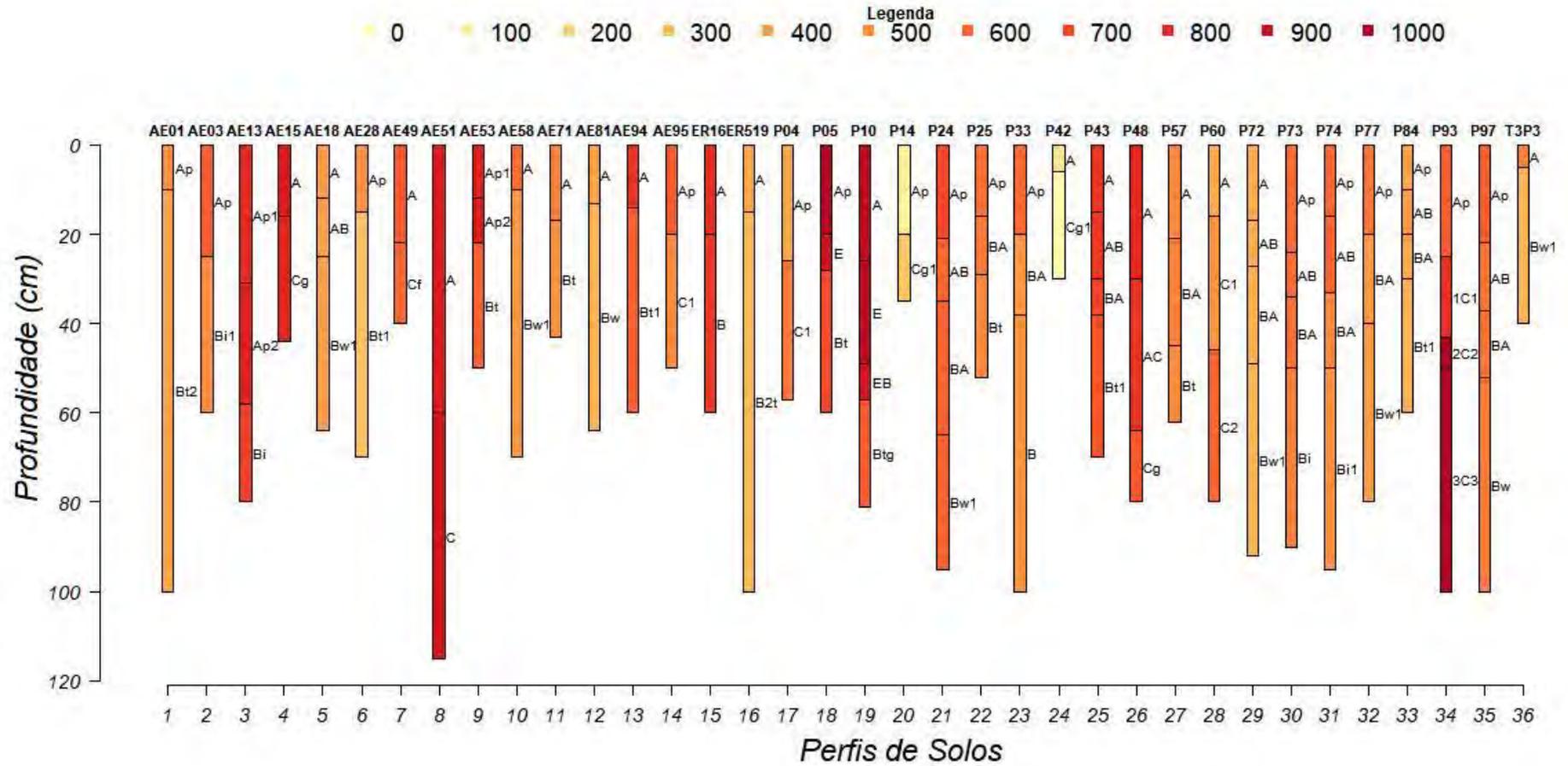


Figura 15. Composição e variabilidade da areia (total) atributos nos perfis dos solos, na unidade $g.kg^{-1}$. O identificador do perfil corresponde às informações contidas na Tabela 1.

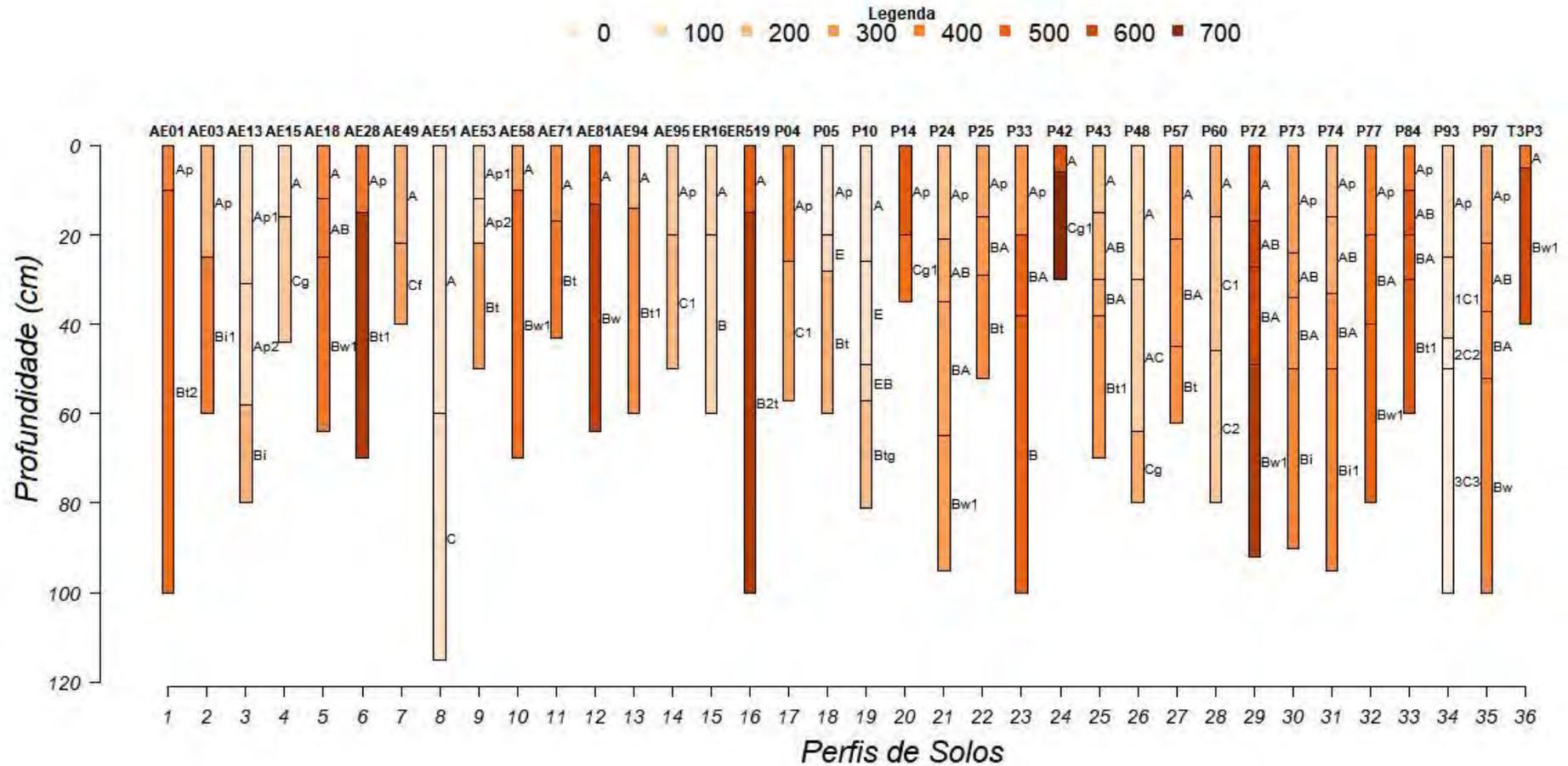


Figura 17. Composição e variabilidade da argila nos perfis dos solos, na unidade g.kg⁻¹. O identificador do perfil corresponde às informações contidas na Tabela 1

Analisando o valor T os perfis do solo (Figura 18), observamos que o valor T aumenta para solos com horizontes orgânicos subsuperficiais e diminui para os demais. Com relação ao valor T ou CTC (Capacidade de Troca de Cátions), o aumento da capacidade do solo em trocar cátions eleva o fluxo de água no solo por capilaridade (verticalmente) e, conseqüentemente, aumenta a infiltração no solo (aumenta a velocidade de infiltração do solo e a condutividade hidráulica do solo). Entretanto, tal afirmativa não é uma regra, pois a presença de argila de alta atividade (2:1) nos solos e, principalmente, com alta CTC possuem baixa vib e K_{sat} .

No geral, há uma diminuição do valor T (Figura 18) na transição entre os horizontes superficiais e subsuperficiais, indicando que provavelmente a porção de água infiltrada nos solos é reduzida quantitativamente a partir da profundidade (Z) igual a 40 cm. O valor T apresentou-se na análise exploratória com correlação de baixa à moderada em relação às variáveis físico-hídricas, porém manteve correlação de moderada à alta, mesmo que inversa, com a granulometria dos solos, principalmente dos conteúdos de areia total e silte. Como a análise de correlação é pareada, vale a pena explorar a potencialidade deste atributo na calibração de FPTs.

A distribuição das frações de areia fina, porosidade, areia grossa e argila dispersa em água podem ser vistas na Figura 19, Figura 20, Figura 21 e Figura 22 a seguir. A análise desses atributos dos solos nos permite inferir a variação da porosidade e compará-la com a granulometria do solo e com a dispersão da argila em meio aquoso, principalmente da areia que influencia na quantidade de água transmitida ou movimentada ao longo do perfil dos solos.

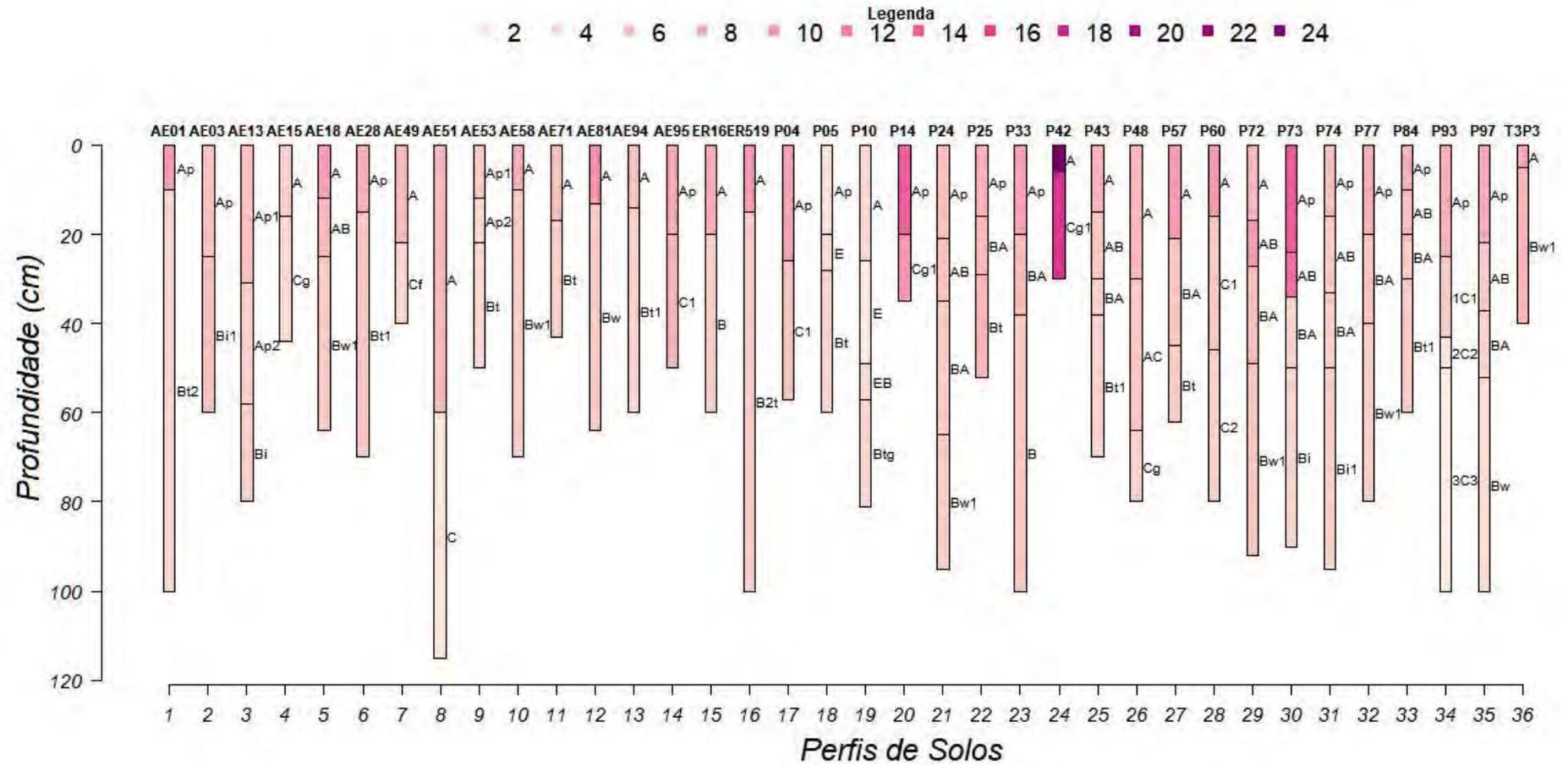


Figura 18. Composição e variabilidade do valor T nos perfis de solo da bacia, na unidade $\text{cmol}_c.\text{kg}^{-1}$. O identificador do perfil corresponde às informações contidas na Tabela 1.

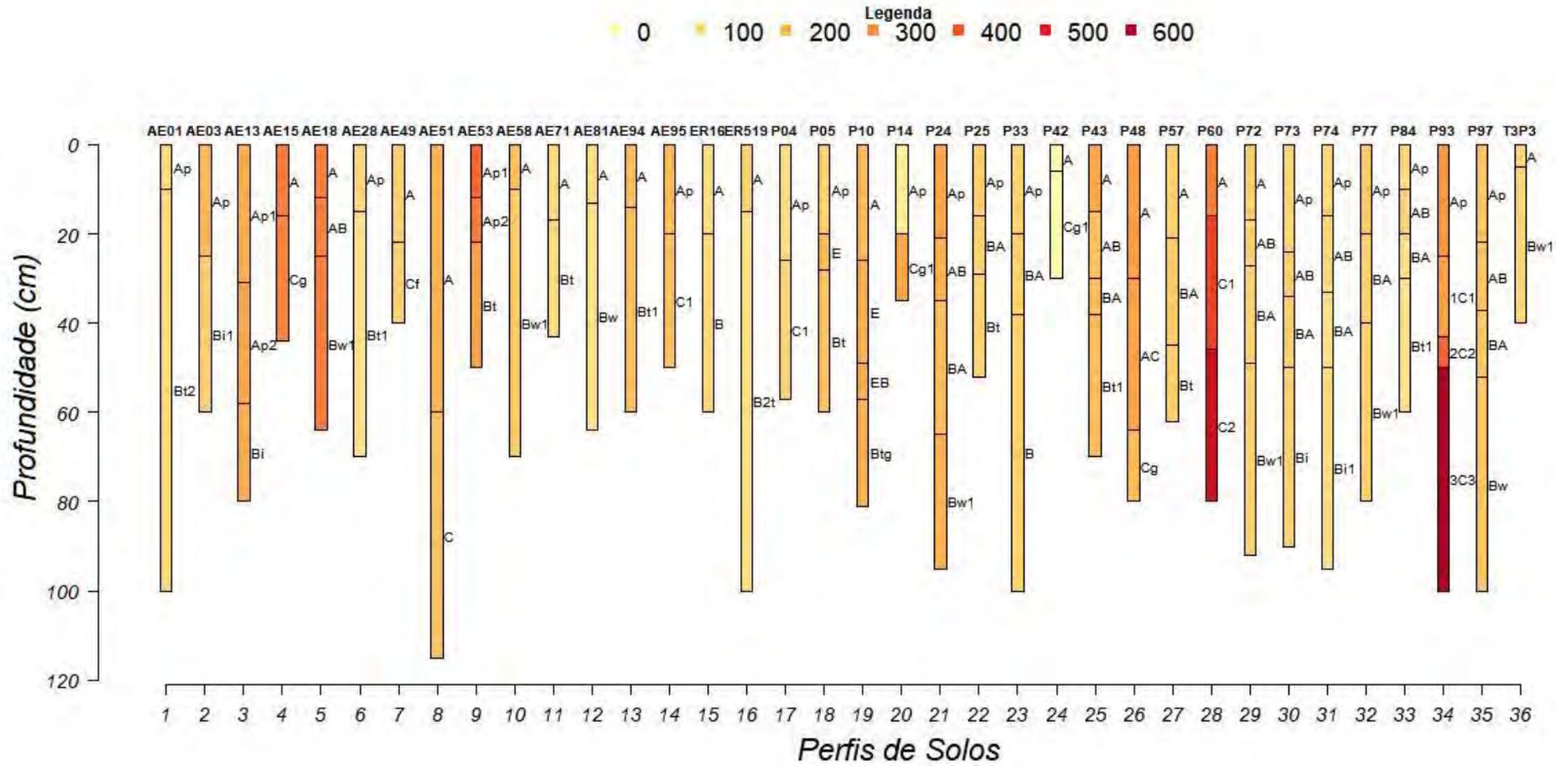


Figura 19. Composição e variabilidade da areia fina na coleção de perfis, na unidade g.kg⁻¹. O identificador do perfil corresponde às informações contidas na Tabela 1.

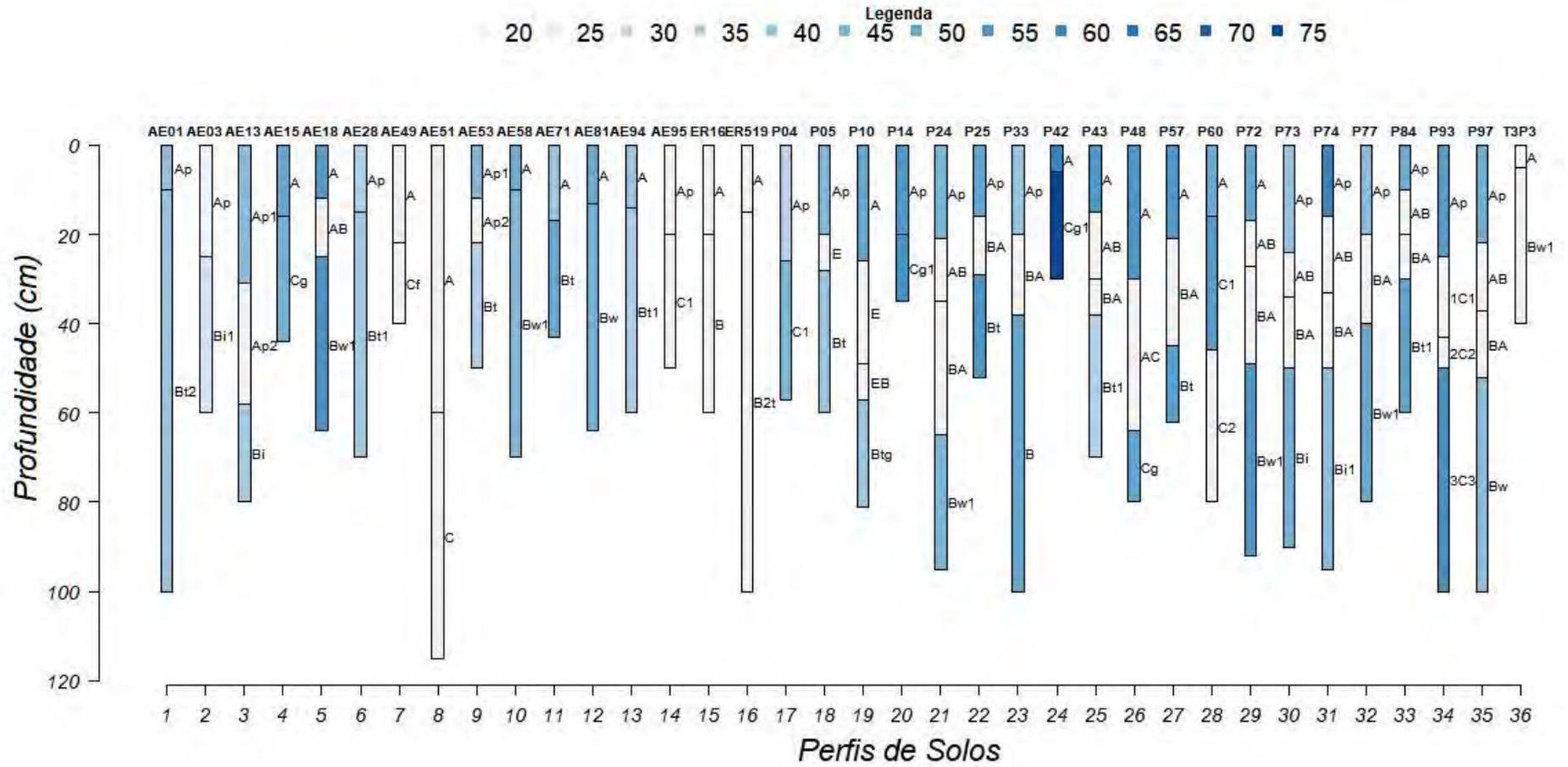


Figura 20. Composição e variabilidade da porosidade na coleção de perfis, na unidade porcentagem (%). O identificador do perfil corresponde às informações contidas na Tabela 1.

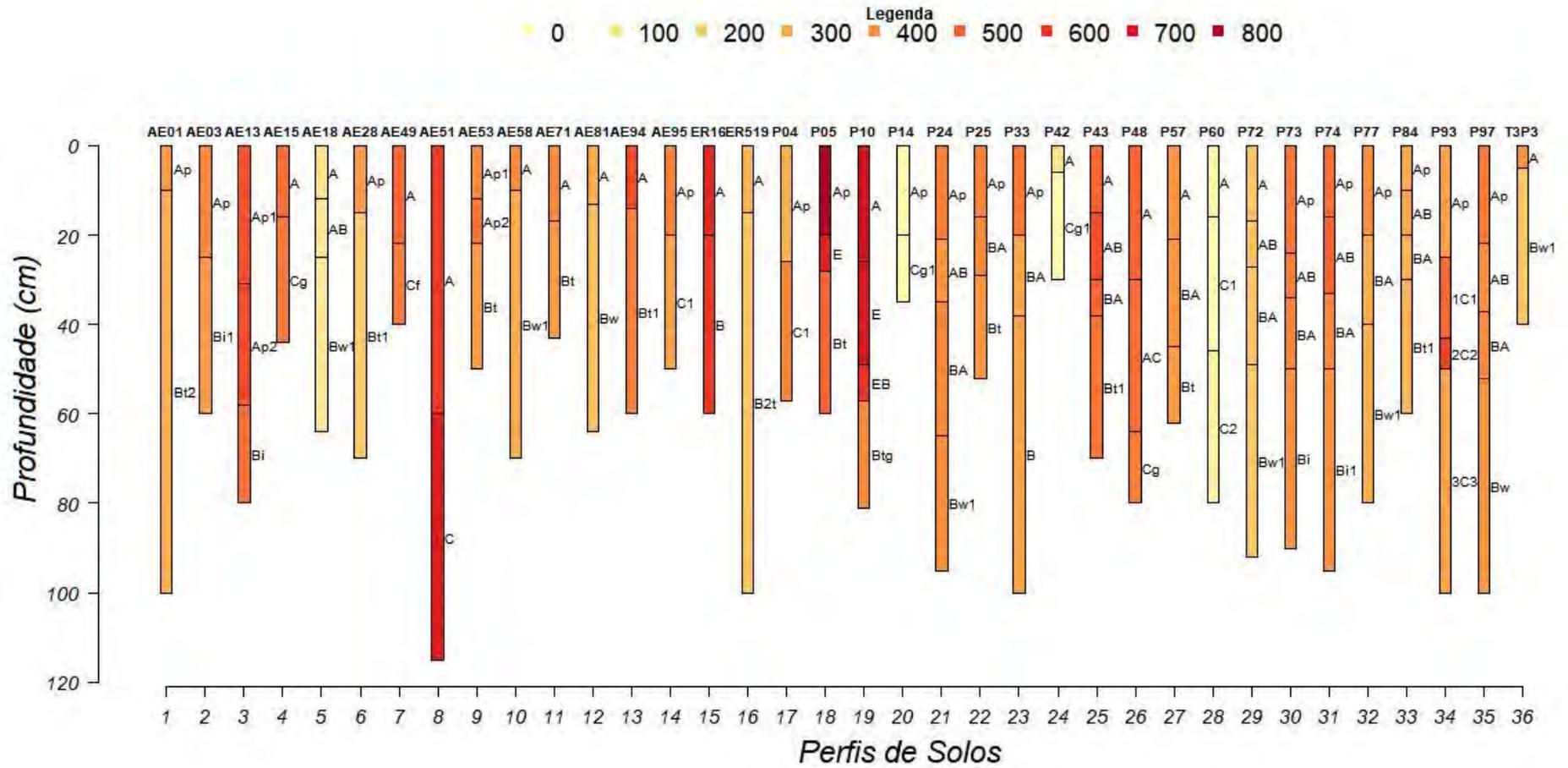


Figura 21. Composição da areia grossa na coleção de perfis, na unidade g.kg⁻¹. O identificador do perfil corresponde às informações contidas na Tabela 1.

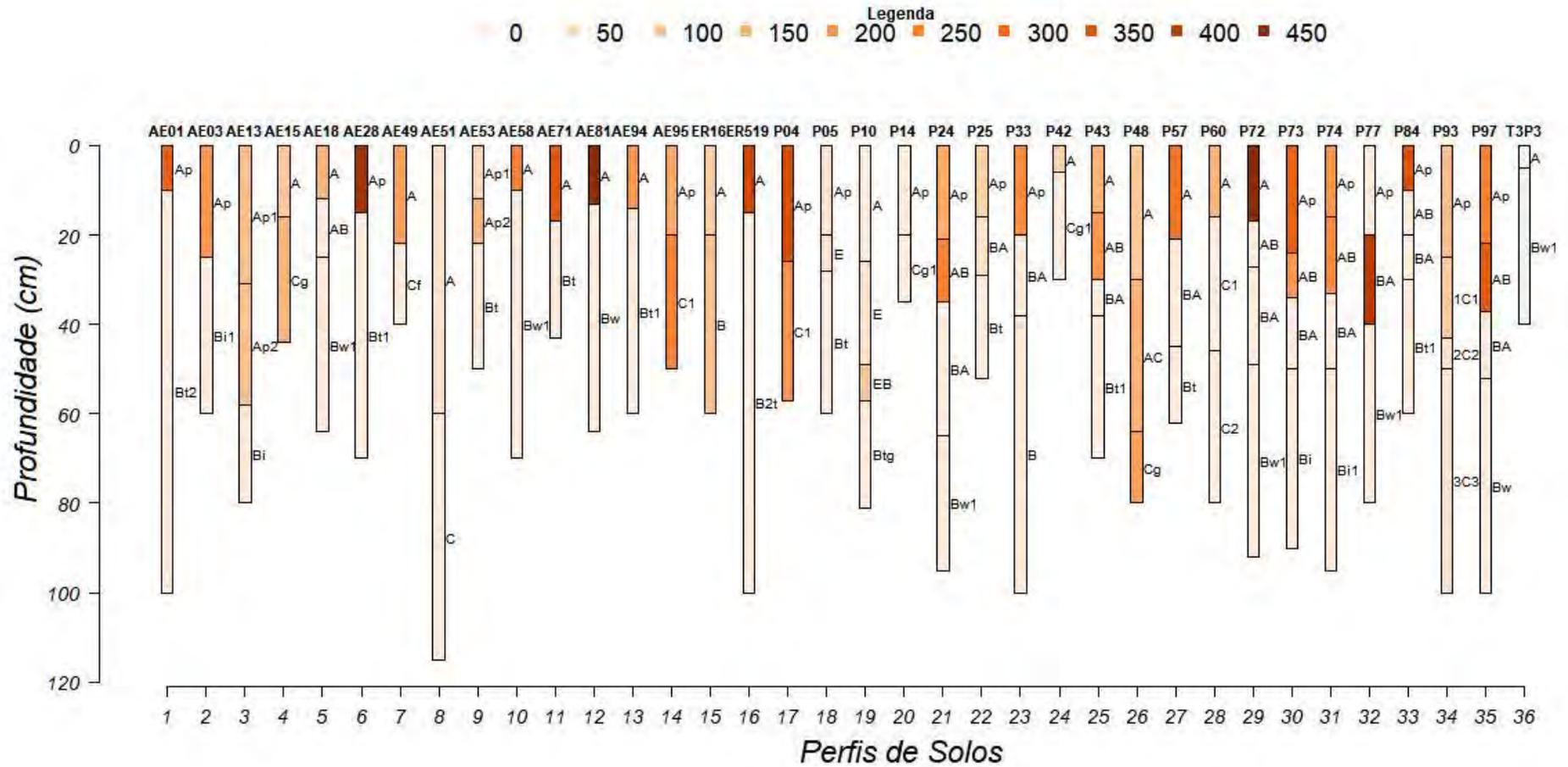


Figura 22. Composição da Argila dispersa em água na coleção de perfis, na unidade g.kg^{-1} . O identificador do perfil corresponde às informações contidas na Tabela 1.

Observou-se pelos dados originais que a K_{sat} aumentou principalmente no limite entre os horizontes superficiais e o subsuperficiais ($Z \sim 20-35$ cm), onde a fração de argila e silte decrescem em proporção ao aumento de areia, dando mais mobilidade para a água fluir entre os poros.

Analisando o carbono orgânico em relação à composição granulométrica (Figura 23), a classificação dos solos (Figura 24a) e a elevação da área (atributo correlacionado positivamente com o valor de cota) (Figura 24b), observa-se que os valores mais elevados de carbono orgânico estão relacionados as frações franco-arenosas e os valores médios-baixos associados a horizontes arenosos, de textura franca e argilo-arenosa.

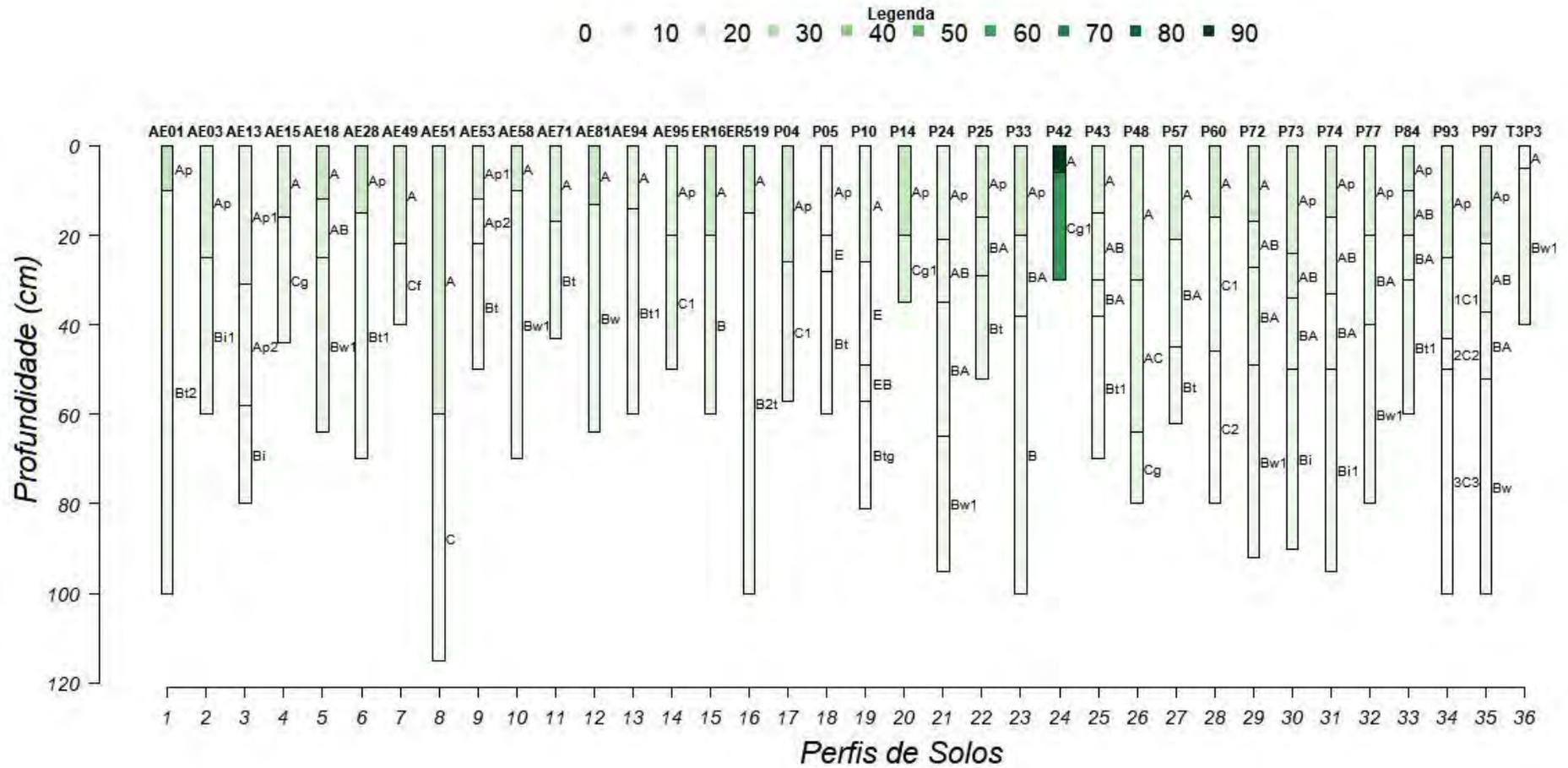


Figura 23. Composição e variabilidade do carbono orgânico nos perfis dos solos, na unidade g.kg^{-1} . O identificador do perfil corresponde às informações contidas na Tabela 1.

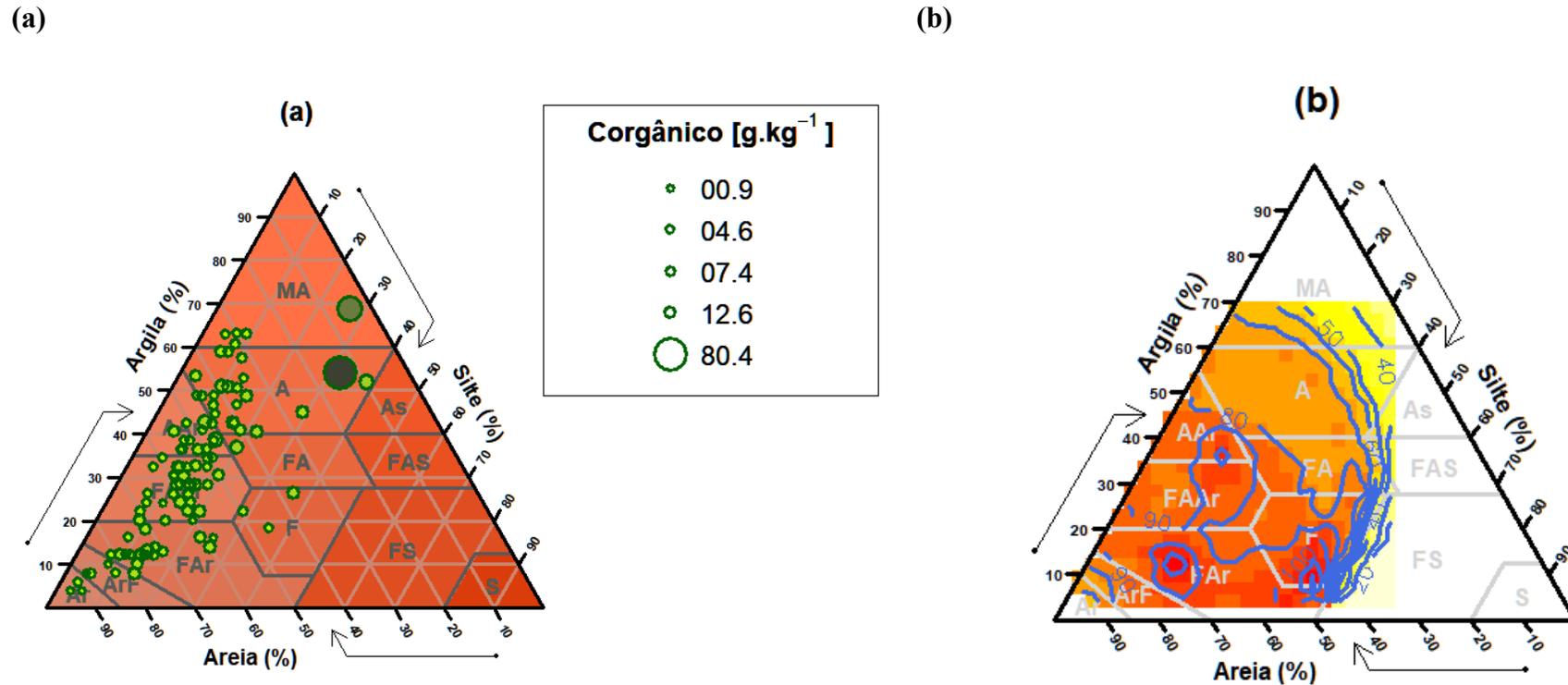


Figura 24. Composição e variabilidade do carbono orgânico (g.kg^{-1}) na bacia segundo (a) triângulo textural da classificação SiBICS e gráfico de bolhas (*bubble plot*) da variação do carbono orgânico conforme a elevação (altimetria, Z), considerando na legenda os quantis da distribuição (0, 25, 50, 75 e 100). A variável Corgânico é plotada como um fator de expansão considerando Z; (b) Mapa de calor (do inglês, *heatmap*) interpolado usando o IDW (*Inverse Distance Weighted*) e contendo as isoietas de elevação representadas pelas linhas contínuas (em azul royal).

Outra análise interessante é o gráfico de mapa de calor (*heat map*) (Figura 24b), muito útil para identificar padrões e diferenciar a ocorrência de determinadas tendências, distinguindo-os por meio de intensidades de cor e isolinhas. Para este estudo, fez-se a associação das isolinhas de altimetria com as informações interpoladas do triângulo textural, gerando o mapa de calor. Ao comparar a distribuição do carbono orgânico (Figura 23) com o *heat map* (Figura 24b) e, associando-se à classificação dos solos, é possível verificar que os valores mais altos de carbono orgânico (representado por círculos maiores) são verificados nas elevações de 80-100 m (cores alaranjadas) e onde há a ocorrência de altos conteúdos de areia e baixo de argila. Segundo o Sistema Brasileiro de Classificação de Solos (SiBCS), os solos com conteúdo de Corgânico foram classificados, em sua grande maioria, como argilo-arenosos, embora tendo ainda a coleção de perfis da bacia constituição de porção franca em sua composição granulométrica nos solos.

Por fim, o último atributo analisado foi a densidade (Figura 25 e Figura 26), o qual é um importante parâmetro nas análises hidropedológicas, tanto pelo ponto de vista do solo (densidade do solo, Figura 25) quanto da partícula (densidade da partícula, Figura 26). Os parâmetros de densidade tendem a manter uma relação direta com a permeabilidade da água, sendo maior ou menor a movimentação da água nos espaços intra e extra poros. A permeabilidade depende ainda de outros atributos do solo, sendo os principais deles a densidade, a porosidade, a macro e microporosidade (MESQUITA & MORAES, 2004; SAMPAIO et al., 2006). A granulometria e a estrutura apresentam influência sobre o espaço poroso do solo, sua porosidade total e distribuição de poros, dificultando em diferentes níveis de intensidade o movimento da água nos solos (MESQUITA, 2001).

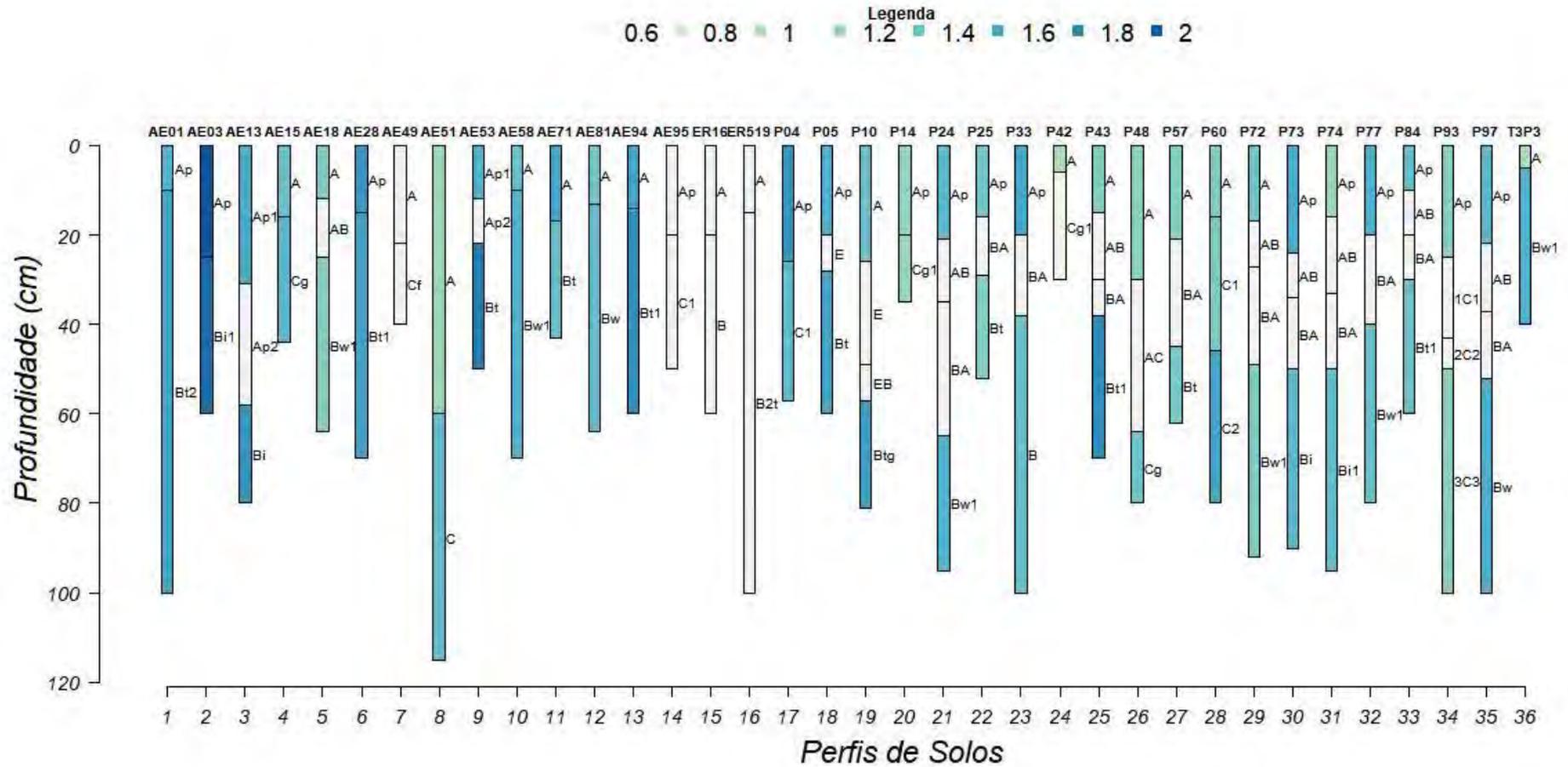


Figura 25. Composição e variabilidade da densidade do solo (ds) nos perfis dos solos, em g.cm⁻³. O identificador do perfil corresponde às informações contidas na Tabela 1.

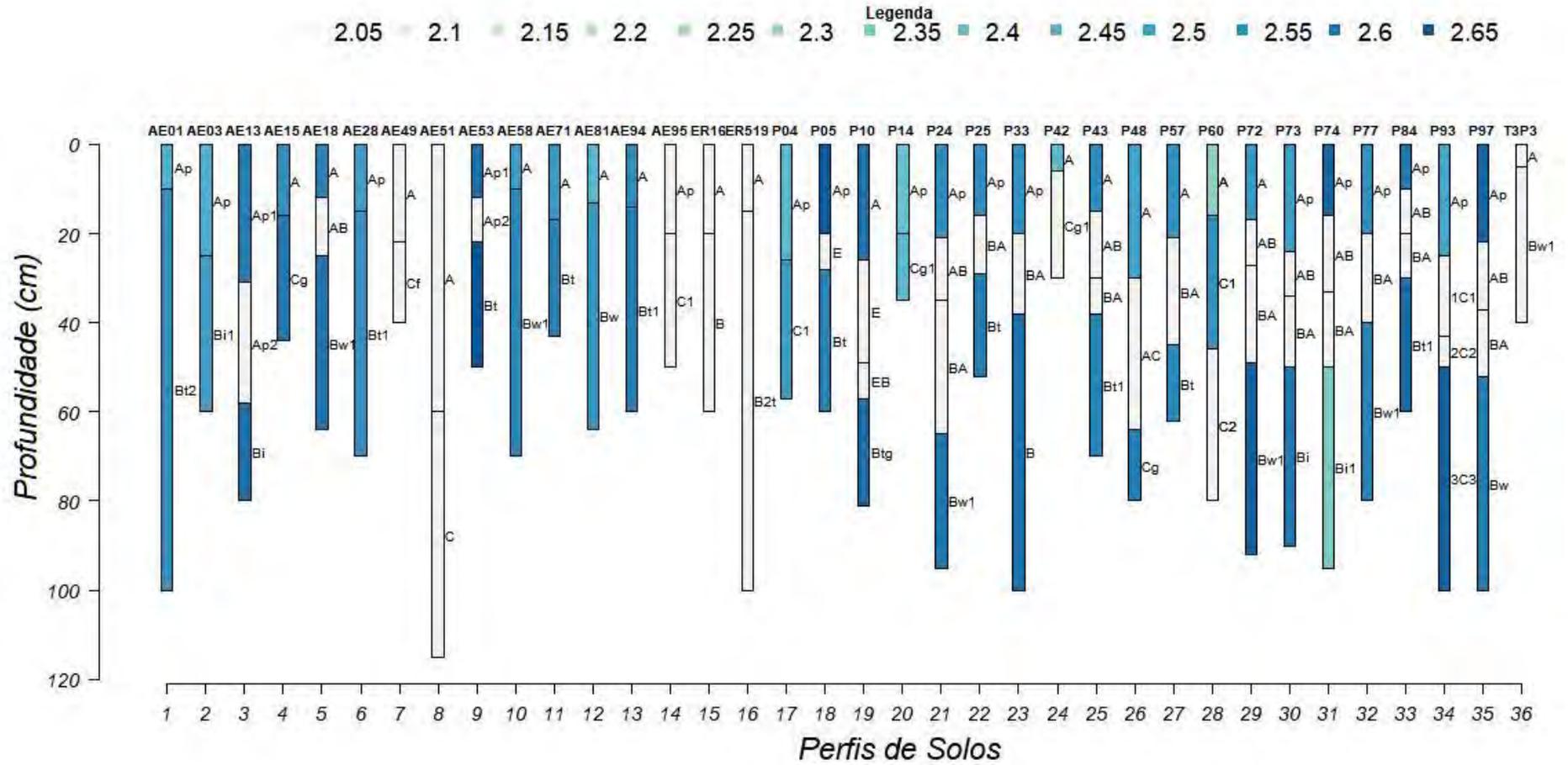


Figura 26. Composição e variabilidade da densidade da partícula (dp) nos perfis dos solos, em g.cm^{-3} . O identificador do perfil corresponde às informações contidas na Tabela 1.

A densidade do solo (Figura 25), no geral, aumenta com a profundidade dos solos da Bacia. Os solos orgânicos apresentam baixos valores de densidade do solo, no entorno de 0-0.6 g.cm⁻³. Nos solos minerais (presença de horizonte AB, BA e C), a densidade do solo diminui, porém, a densidade dos demais solos aumenta em profundidade. O comportamento é bastante diferente para a densidade da partícula (Figura 26), que afeta a agregação e o comportamento do fluxo de água e está associada à dimensão da partícula (granulometria do solo). O valor de densidade da partícula chega ser a maior do que a do solo, alcançando um máximo de 2.65 g.cm⁻³ nos solos da bacia.

Ainda, os solos da bacia distribuem-se num padrão bem expressivo indicando composição em proporções variadas de seus atributos ao longo das camadas dos perfis dos 36 pontos avaliados. Uma razão para a ocorrência deste comportamento é, certamente, a grande variabilidade das características físicas e químicas dos solos estudados para a bacia do Guapi-Macacu.

Finalizada a análise qualitativa, deu-se início aos procedimentos de organização dos dados mediante a aplicação de funções de profundidade (*soil-depth functions*), por meio do método *Slice-Wise*, *Slab* e *Spline* em ambiente RStudio (BEAUDETTE et al., 2021), utilizando o pacote AQP. O fatiamento dos perfis de solos deu-se em intervalos de 1 em 1 cm para a harmonização e agregação do conjunto de dados segundo seis intervalos de 20 cm de profundidade. Ainda, os dados foram filtrados segundo o hidromorfismo (solos hidromórficos e não-hidromórficos) e profundidade (0-20cm para solos superficiais; 20-40 cm para solos subsuperficiais) de análise dos perfis. Após a harmonização, o resultado são quatro tabelas usadas na modelagem das funções de pedotransferência.

3.4. Modelagem dos Atributos Físico-Hídricos

Os primeiros modelos a ser implementados foram os baseados em Regressão Linear Múltipla (RLM), aplicados sob as variáveis hidropedológicas avaliadas anteriormente segundo análise do RFE. Para implementação e análise de modelos RLM, é necessário verificar as hipóteses estatísticas de normalidade, independência, significância dos regressores e autocorrelação dos resíduos. Para isto, levou-se em consideração os testes propostos pelos pesquisadores indicados no Quadro 5 a seguir.

Quadro 5. Análise de adequação de modelos regressivos.

Teste	Hipóteses *
<i>ANOVA</i> (F de Fisher - <i>Snedecor</i>)	H ₀ : O modelo ajustado não possui significância nos regressores estimados ($\beta_1 = \beta_2 = \dots = \beta_n = 0$). H ₁ : O modelo ajustado possui significância em seus regressores estimados (pelo menos um deles) ($\beta_1 \neq \beta_2 \neq \dots \neq \beta_n \neq 0$).
<i>Shapiro-Wilk</i>	H ₀ : Há normalidade nos resíduos. H ₁ : Não há normalidade nos resíduos.
<i>Breusch-Pagan</i>	H ₀ : variâncias são iguais (homoscedasticidade). H ₁ : variâncias são diferentes (heterocedasticidade).
<i>Durbin-Watson</i>	H ₀ : a autocorrelação entre os resíduos é igual a zero (independência dos resíduos). H ₁ : a autocorrelação entre os resíduos é diferente de zero (dependência dos resíduos).
Interpretação: A um nível de confiança β^{**} , se $p\text{-valor} < \alpha^{**}$, rejeita-se a hipótese nula (H ₀) e, conseqüentemente, aceita-se a hipótese alternativa (H ₁). Se $p\text{-valor} > \alpha^{**}$, a hipótese nula (H ₀) é aceita.	

* H₀ e H₁ são as hipóteses nula e alternativa dos testes, respectivamente; ** α é o nível de significância adotado nos testes; $\beta = 1 - \alpha$ é o nível de confiança para o respectivo α adotado nos testes; $\beta_1, \beta_2, \dots, \beta_n$ são os parâmetros de inclinação da reta do modelo regressivo ajustado ($Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n$).

Fonte: BREUSCH & PAGAN (1979); DURBIN & WATSON (1951); SHAPIRO & WILK (1965); SNEDECOR (1934).

Após implementação da regressão linear múltipla usando as variáveis do solo (Tabela 5), os resíduos dos modelos RLM foram testados a um nível de significância de 5% ($\alpha = 5\%$ ou 0,05) (Tabela 6), obtendo-se resultados de rejeição da hipótese nula (H₀), indicando que os mesmos não podem ser aplicados para modelagem dos dados físico-hídricos, conforme a formulação das hipóteses (Quadro 5). O teste SW (Tabela 6) apresentou resultados abaixo do nível de significância adotado ($\alpha = 0,05$) para análise dos modelos MLR, tanto para Ksat quanto para vib, indicando que os dados não seguem uma distribuição normal de probabilidade. O teste BP (Tabela 6) evidenciou a heterocedasticidades das variâncias do conjunto de dados modelado, o que leva à ineficiência e tendenciosidade dos estimadores do modelo adotado e tendenciosidade (apresentam viés), já que os modelos testados apresentam valores abaixo de $\alpha = 0,05$, levando à rejeição da hipótese nula (H₀). Ainda, para o teste DW (Tabela 6), os valores apresentados para os modelos MLR aplicados na modelagem da Ksat e vib refletem uma autocorrelação dos resíduos dos modelos (rejeição de H₀ à $\alpha = 0,05$), que causam omissão de variáveis explicativas e especificação deficiente do modelo matemático.

Os resultados dos testes indicam que os dados analisados não são indicados para o desenvolvimento de FPT voltados à modelagem de atributos físico-hídricos dos solos como a vib e a Ksat baseando-se em princípios do MLR, já que os requisitos teóricos não

foram atingidos. Segundo Izbicki & Dos Santos (2019), para a calibração de funções baseadas em árvores (*Regression Trees e Random Forest*), não é necessário utilizar tais critérios.

Tabela 5. Equações pedotransferência ajustadas por meio de Regressão linear múltipla.

Parâmetros*	Equação ajustada (funções de pedotransferência)
Z= 0-20cm, H	$y_{ksat} = -0,8919 + 0,0093 p + 0,0001 ard + 0,0001 s + 0,3015 ds + 0,00003 a + ar$
Z= 20-40cm, NH	$y_{ksat} = -0,7095 + 0,00006 s + 0,0013 vt - 0,00009 ar + 0,6907 ds + 0,0168 a + 0,00001 ard - 0,0059 ph - 0,3849 dp$
Z= 0-20cm, H	$y_{ksat} = -0,5853 - 0,0283 dp + 0,0076 p - 0,00006 a - 0,00009 ar + s + 0,2831 ds$
Z= 20-40cm, NH	$y_{ksat} = -1,0141 + -0,0422 dp + 0,0107 p + 0,0002 s + 0,4293 ds - 0,0005 c + 0,0023 valorT + 0,00002 ar + 0,000006 ard$
Z= 0-20cm, H	$y_{vib} = 0,8262 + 0,0036 s - 0,0033 ar - 0,0116 p + 0,0004 ard + a$
Z= 20-40cm, NH	$y_{vib} = -54,8694 + 53,9466 ds + 1,2915 p + 0,0065 s - 0,0002 ard - 0,1088 c + 0,0054 a + 0,2670 vt - 32,0423 dp - 0,3905 ph + ar$
Z= 0-20cm, H	$y_{vib} = 12,9848 + 3,2636 dp - 0,2188 p + 0,0122 silte - 0,0026 ar - 7,52921 ds + a + 0,0063 c - 0,0008 ard - 0,0988 vt$
Z= 20-40cm, NH	$y_{vib} = -40,7948 + 0,0159 s + 0,4234 pH + 15,5754 dp - 0,1144 vt + 0,1461 c - 0,0540 p$

*Camada e condição hidrológica: H= solos hidromórficos; NH= solos não-hidromórficos; p= porosidade; ds= densidade do solo; dp= densidade da partícula; a=areia; s=silte; ar=argila; c=carbono orgânico; ard = argila dispersa; vt= valor T; ph= potencial hidrogeniônico. Em azul estão destacados os regressores de cada equação.

Tabela 6. Testes estatísticos para pressupostos em regressão.

TE	Modelos			
	Ksat			
	MLR ₁	MLR ₂	MLR ₃	MLR ₄
SW	8,89x10 ⁻³	7,81x10 ⁻⁵	1,35x10 ⁻⁴	2,91x10 ⁻⁴
BP	1,47x10 ⁻⁶	4,35x10 ⁻⁴	4,72x10 ⁻⁶	3,43x10 ⁻³
DW	2,20x10 ⁻¹⁶	2,20x10 ⁻¹⁶	2,20x10 ⁻¹⁶	2,82x10 ⁻⁷
TE	vib			
	MLR ₅	MLR ₆	MLR ₇	MLR ₈
SW	1,32x10 ⁻⁴	2,95x10 ⁻⁶	6,55x10 ⁻³	1,44 x10 ⁻²
BP	1,02x10 ⁻³	1,43x10 ⁻²	1,76x10 ⁻⁴	5,63x10 ⁻³
DW	2,20x10 ⁻¹⁶	2,20x10 ⁻¹⁶	7,65x10 ⁻¹	3,40x10 ⁻⁴

TE= Testes Estatísticos; SW= Shapiro-Wilk; BP= Breusch-Pagan; DW= Durbin-Watson; MLR= Multiple Linear Regression. MLR estimado considerando: 1;5= solos hidromórficos superficiais; 2;6= solos hidromórficos subsuperficiais; 3;7= solos não-hidromórficos superficiais; 4;8= solos não-hidromórficos subsuperficiais.

Para contornar tal situação, buscou-se a melhor transformação para os dados segundo o pacote “*best Normalize*” (PETERSON, 2019). Os resultados apontaram

normalidade dos resíduos dos modelos para os dados transformados, porém mantiveram-se os resultados de heterocedasticidade e dependência dos resíduos, conforme os testes do Quadro 5. Outra alternativa seria proceder com a remoção de possíveis candidatos a outliers, porém esta solução foi descartada devido à natureza dos dados (grande variabilidade de ocorrência espacial). Os resultados obtidos nestes testes indicam, por razões intrínsecas ao fenômeno em que o dado é descrito (movimentação da água no solo), que a variabilidade dos atributos físico-hídricos não consegue ser explicada através de modelos de regressão múltipla, dada a heterogeneidade dos dados, mesmo sendo aplicadas transformações estatísticas sob estes.

As funções estimadas para Ksat pelo modelo *Random Forest* atingiram a estabilização do RMSE (minimização) a partir de uma quantidade de aproximadamente 250 árvores. Já para a vib, a começar pela quantidade de 150 árvores as funções já apresentaram estabilização no RMSE para todos os horizontes avaliados. A Figura 27 e Figura 28 mostram os resultados da modelagem *Random Forest* para a Ksat e vib, respectivamente.

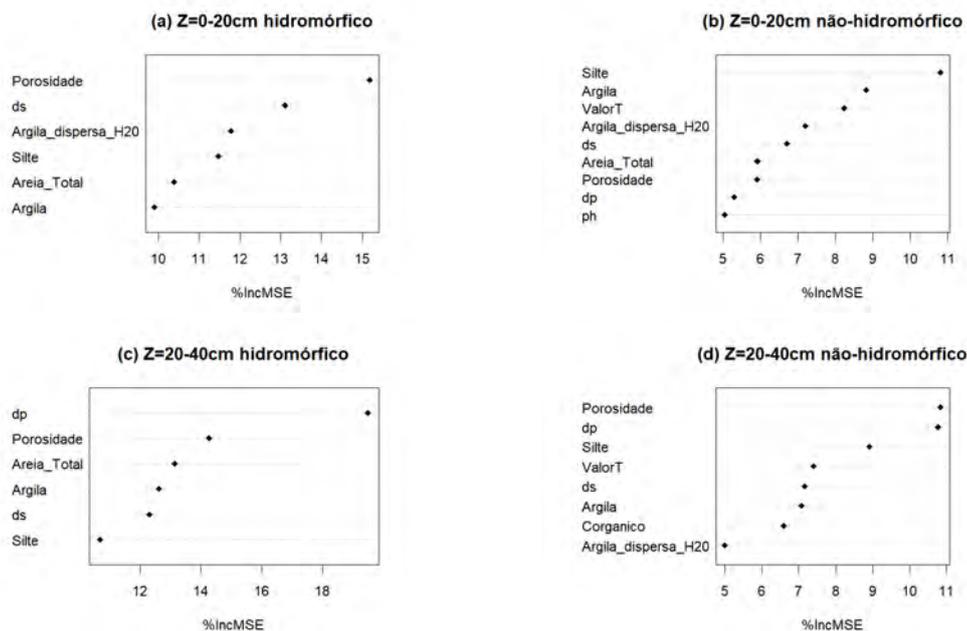


Figura 27. Gráficos de importância das variáveis preditoras na estimativa da condutividade hidráulica saturada (Ksat) por modelagem *Random Forest*.

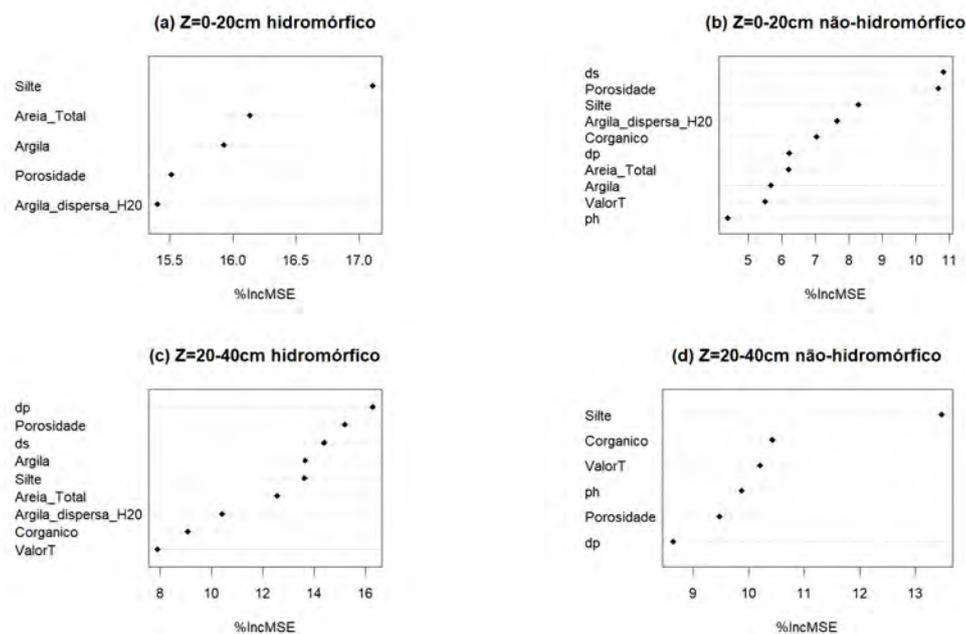


Figura 28. Gráficos de importância das variáveis predictoras na estimativa da velocidade de infiltração básica (vib) por modelagem *Random Forest*.

Na estimativa da K_{sat} (Figura 27), os modelos utilizaram em média 6 variáveis de entrada para a modelagem, selecionadas pelo RFE. A maior importância no rank das variáveis são: porosidade (Figura 27a), silte (Figura 27b), e densidade da partícula (dp) (Figura 27c e Figura 27d), respectivamente. Nas FTPs estimadoras da vib para o modelo *Random Forest* (Figura 28), o RFE selecionou em média 5 variáveis. No rank de importância da vib, as variáveis mais explicativas dos modelos foram: silte (Figura 28a e Figura 28d), porosidade (Figura 28b) e densidade da partícula (Pd) (Figura 28c). As equações ajustadas pelos modelos de regressão linear múltipla são mostradas no Tabela 5.

Na análise das FTPs resultantes da modelagem por árvores de regressão, observa-se que na predição da K_{sat} (Figura 29 – a,b,c,d) o modelo teve a progressão e subdivisão dos nós diversificadamente para cada tipo de camada estimada, alcançando 1 nó para $Z=0-20\text{cm}$ hidromórfico e entre 2 e 4 nós para as demais camadas. A camada $Z=0-20\text{cm}$ hidromórfico destaca-se por apresentar convergência rápida na estimativa da FTP associada, com estruturação da árvore diretamente a partir da porosidade (64% dos dados separados a partir da porosidade $<50\%$ e, 36%, a partir da porosidade $>50\%$ em função dos hiperparâmetros associados ao modelo, o que indica um ajuste baixo na estimativa do modelo. Analisando-se agora a vib (Figura 29 – e,f,g,h), as funções convergiram majoritariamente entre 2 e 4 nós, partindo de diferentes atributos.

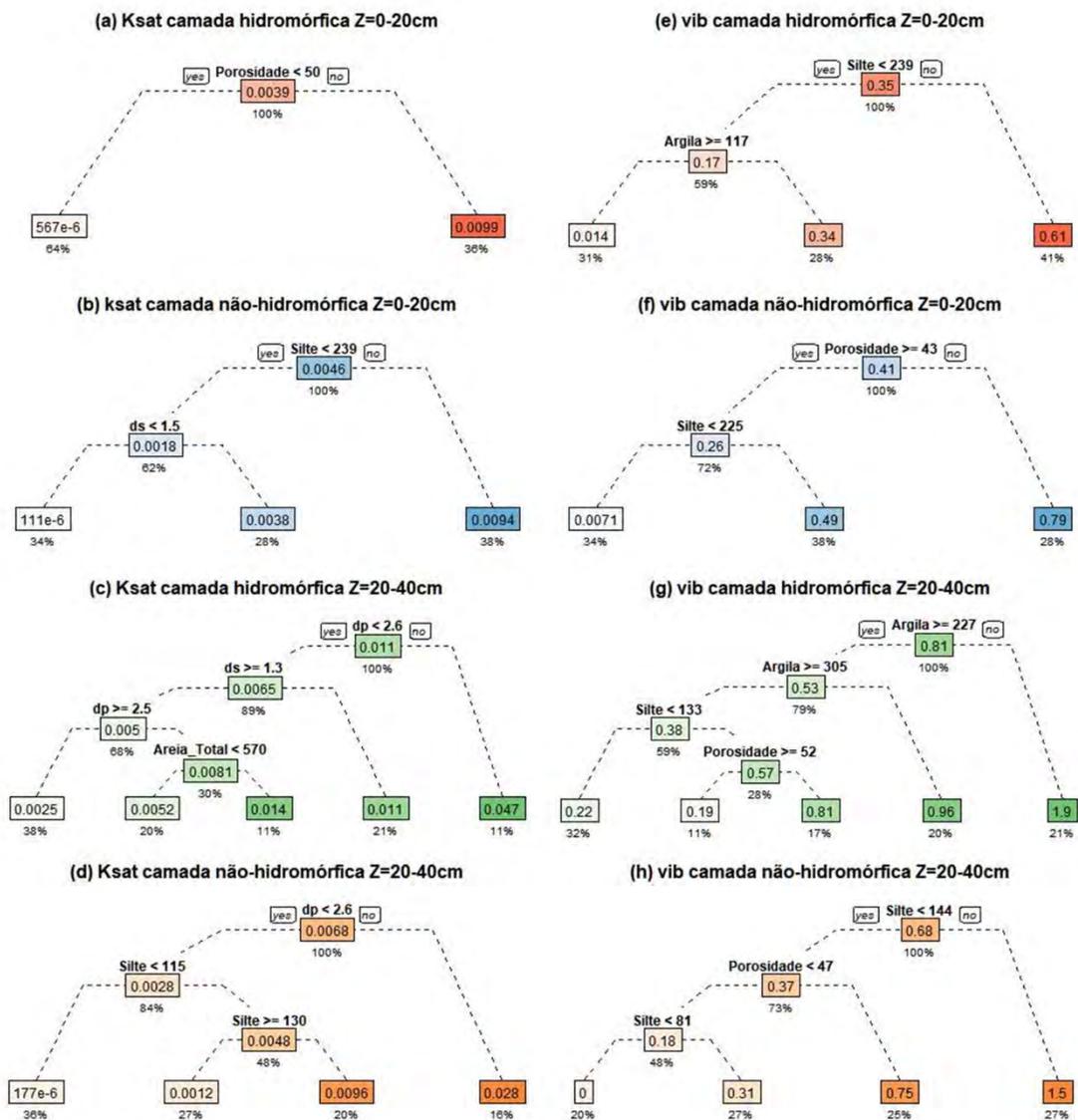


Figura 29. Gráficos de Árvores de Regressão (*Regression Trees – RT*) e iterações na calibração das funções de pedotransferência de modelos baseados em árvores de regressão, onde: (a,b,c,d) condutividade hidráulica saturada (Ksat); (e,f,g,h) velocidade de infiltração básica (vib). Nota: *yes*= booleano para “sim”; *no*= booleano para “não”.

As FTPs foram aplicadas aos dados dos perfis dos solos predizendo os valores de vib e Ksat para 86 pontos amostrais distribuídos na extensão da área da bacia (existentes no conjunto de dados do levantamento legado), totalizando-se 394 horizontes com suas respectivas camadas associadas. O conjunto de dados preditos foram reunidos em uma tabela de validação para verificar, em seguida, a qualidade dos modelos na estimativa das variáveis de estudo. Os valores foram preditos conforme as funções de calibração da pedotransferência estimadas para cada tipo de camada (0-20cm e 20-40cm), condição hidrológica (hidromorfismo e não-hidromorfismo) e atributo associado (vib e Ksat) (Figura 11).

A fase de validação dos modelos baseou-se na validação cruzada *Leave-One-Out Cross-Validation* – LOOCV (SAMMUT & WEBB, 2010). Os três modelos foram avaliados conforme a validação LOOCV, usando os dados preditos segundo cada função de pedotransferência ajustada. Os parâmetros estatísticos de qualidade dos modelos podem ser observados nas Tabela 7 e Tabela 8 a seguir.

Tabela 7. Parâmetros estatísticos de qualidade dos modelos treinados após validação cruzada (LOOCV aplicado aos 36 pontos amostrais).

Variável	Parâmetros*	FPT	Métricas Estatísticas		
		Modelo	MAE**	RMSE**	R ²
Ksat	Z= 0-20cm; H	MLR	0,0037	0,0052	0,6727
		RT	0,0044	0,0080	0,2394
		RF	0,0004	0,0006	0,9951
	Z= 20-40cm; H	MLR	0,0028	0,0045	0,6512
		RT	0,0033	0,0065	0,2722
		RF	0,0015	0,0033	0,8139
	Z= 0-20cm; NH	MLR	0,0096	0,0131	0,3749
		RT	0,0069	0,0100	0,6344
		RF	0,0019	0,0040	0,9416
	Z= 20-40cm; NH	MLR	0,0068	0,0102	0,4169
		RT	0,0048	0,0090	0,5511
		RF	0,0020	0,0036	0,9260
Vib	Z= 0-20cm; H	MLR	0,2685	0,3423	0,4591
		RT	0,2692	0,3924	0,2894
		RF	0,0393	0,0668	0,9794
	Z= 20-40cm; H	MLR	0,1171	0,2045	0,7624
		RT	0,1379	0,2780	0,5610
		RF	0,0818	0,1539	0,8655
	Z= 0-20cm; NH	MLR	0,6007	0,7490	0,4469
		RT	0,5292	0,7852	0,3921
		RF	0,0796	0,1586	0,9752
	Z= 20-40cm; NH	MLR	0,4703	0,5762	0,5677
		RT	0,3948	0,6746	0,4074
		RF	0,1521	0,3115	0,8737

*Z= camada do solo; *H= solos hidromórficos; *NH= solos não-hidromórficos; **unidade= cm.min⁻¹; FPT= Funções de Pedotransferência; ID= Identificador; MLR= *Multiple Linear Regression*; RT= *Regression Trees*; RF= *Random Forest*; MAE= *Mean Absolute Error*; RMSE= *Root Mean Squared Error*; R²= coeficiente de determinação.

Tabela 8. Parâmetros estatísticos da totalidade dos dados (preditos x estimados) obtidos na validação cruzada (LOOCV aplicado aos 122 pontos amostrais).

Variáveis	Modelos	Métricas Estatísticas		
		MAE*	RMSE*	R ²
Ksat	MLR	0,0066	0,0101	0,4543
	RT	0,0055	0,0091	0,5578
	RF	0,0015	0,0033	0,9409
Vib	MLR	0,4218	0,5750	0,5223
	RT	0,3796	0,6281	0,5300
	RF	0,0902	0,1922	0,9466

*unidade= cm.min⁻¹. MLR= *Multiple Linear Regression*; RT= *Regression Trees*; RF= *Random Forest*; MAE= *Mean Absolute Error*; RMSE= *Root Mean Squared Error*; R²= coeficiente de determinação.

Analisando os parâmetros métricos resultantes das FPTs estimadas para cada particularidade do solo (hidromorfismo e profundidade da camada) (Tabela 7), o modelo *Random Forest* foi o que obteve qualidade superior na validação ($R^2 > 0,80$ ou 80%) para as FPT's estimadas tanto para a vib quanto para a Ksat. Já o modelo de árvores de regressão teve desempenho mediano ($R^2 > 0,50$ ou 50%) na estima da Ksat para as camadas não-hidromórficas superficiais ($R^2 = 0,6344$) e subsuperficiais ($R^2 = 0,5511$) dos solos, sendo ambas árvores determinadas primariamente pela densidade da partícula do solo (dp); enquanto para a vib apenas a camada hidromórfica subsuperficial do solo ($R^2 = 0,5610$) apresentou tal desempenho, tendo como atributo principal a porosidade. As demais FPTs estimadas pelo modelo de árvores de regressão (Tabela 7) segundo suas respectivas camadas e condições hidrológicas, tiveram baixo desempenho ($R^2 < 0,50$ ou 50%), mostrando a incapacidade de generalização dos atributos físico-hídricos pelos modelos implementados, os quais informam que menos de 50% da variabilidade dos atributos Ksat e vib podem ser explicadas pelas variáveis de entrada dos modelos (atributos do solo).

Com relação à MLR, na estimativa da Ksat as FPTs estimadas para as camadas hidromórficas superficiais e subsuperficiais dos solos alcançaram um desempenho acima de 60% mensurado pelo R², indicando 60% de explicabilidade da Ksat pelos atributos do solo avaliados (regressores da Tabela 5). Na avaliação da vib (Tabela 7), apenas a FPT estimada para camada hidromórfica subsuperficial do solo ($R^2 = 0,7624$) apresentou qualidade de média à alta ($0,70 < R^2 < 0,80$), apresentando ainda os menores erros na validação do modelo (MAE=0,1171 e RMSE=0,2045). A vib estimada camada não-hidromórfica subsuperficial teve desempenho mediano ($R^2 = 0,5677 > 0,50$), porém valores altos de erro de estimativa da FPT (MAE= 0,4703 e RMSE=0,5762).

No geral, as FTP's estimadas para Ksat apresentaram menores erros de validação (preponderantemente erros milimétricos) (Tabela 7), evidenciados pelas métricas MAE e RMSE, as quais indicam que os valores preditos se mostram precisos, porém não acurados, com relação às observações feitas em campo (valores mensurados in situ). Para a vib, as métricas MAE e RMSE mostram o baixo ajuste das FPT's estimadas pelos três métodos (Tabela 7), sendo o modelo MLR estimado para a camada não-hidromórfica superficial do solo a que apresenta maiores erros na predição da FPT (MAE=0,6007 e RMSE=0,7490).

O modelo *Random Forest*, no geral, foi o que obteve resultados mais explicativos para a modelagem das funções de pedotransferência com as variáveis investigadas frente aos demais modelos (Tabela 8), obtendo valores elevados nos indicadores estatísticos de predição ($R^2 > 0,80$ ou 80%) e reduzidos para os erros nos resíduos (MAE < 0,01 ou 1% e RMSE < 0,02 ou 2%).

Os indicadores estatísticos da validação (Tabela 8) apontaram o modelo RF como o de melhor desempenho na calibração das FPTs frente à MLR e RT (valores observados versus valores preditos), para as variáveis preditas e respectivas camadas. O baixo ajuste do R^2 nos modelos de RLM podem ser explicados pelo não cumprimento dos pressupostos teóricos, reafirmando a necessidade de se avaliar os dados previamente à modelagem. Verifica-se que, ao comparar os resultados dos três modelos mostrados na Tabela 8, um aumento na performance dos modelos MLR quando associados ao aprendizado de máquinas (modelos baseados em árvores, neste caso), pois os modelos de predição baseados em árvores usam uma estrutura computacional capaz de armazenar um modelo de regressão em cada folha da árvore e prover uma partição do conjunto de dados utilizado para induzir o modelo final à melhor qualidade de estimativa dos atributos físico-hídricos usando as variáveis de entrada informadas (atributos do solo).

Os modelos estimados por meio de árvores conseguem atender aos requisitos da mineração de dados físico-hídricos pois, como avaliado por HASTIE et al. (2009), são invariáveis em escala e várias outras transformações de valores dos recursos de entrada (variáveis preditoras), agindo de forma robusta para a inclusão de recursos relevantes à modelagem, além da produção de modelos inspecionáveis. Ainda são precisos, principalmente na análise de variáveis de conjunto de dados com padrões altamente irregulares (alta variabilidade), como é o caso dos atributos físico-hídricos, os quais tendem a apresentar árvores mais “profundas” no treinamento dos modelos. Os modelos

calculam a média de várias árvores de decisão aprofundadas e ramificadas, em diferentes partes do mesmo conjunto de treinamento, com o objetivo de reduzir a variância. Conforme afirmado por BREIMAN (2001), isso ocorre às custas de um pequeno aumento no viés e alguma perda de interpretabilidade, mas em contrapartida, conseguem geralmente aumentar o desempenho no modelo final.

4. CONCLUSÕES

O uso do AQP permitiu a análise da variabilidade dos atributos dos solos em profundidade, permitindo correlacionar as características dos solos com os atributos físico-hídricos estudados, atuando como apoio ao analista na tomada de decisão sob as variáveis de entrada em modelos e possibilitando a harmonização dos dados de entrada. Quando associado aos modelos de aprendizado de máquinas, principalmente o *Random Forest*, o AQP mostrou-se uma ferramenta potencial para estudos preliminares à pedotransferência, sendo recomendado o seu uso em pesquisas voltadas às funções hídricas dos solos.

Os modelos baseados em árvores tiveram desempenho melhor na validação dos dados preditos em contraposição à regressão múltipla, indicando que os atributos na bacia hidrográfica, embora adequados aos pressupostos estatísticos de linearidade comportamental, apresentam uma variabilidade tão grande que os modelos regressivos não são capazes de estimar valores precisos para os dados levantados e, ainda, os modelos de aprendizado de máquinas são capazes de melhor compreender tal interação através de seus algoritmos de ajuste devido a sua capacidade de realizar previsões, memorizar os dados e reproduzir padrões de maneira simultânea, iterativa e cuja resposta da estimação seja a mais otimizada possível. É válido ressaltar que embora as FPTs sejam capazes de prever verticalmente (análise em profundidade) a variação de atributos nos solos como a composição granulométrica, por exemplo, os parâmetros físico-hídricos podem não acompanhar tal lógica devido a sua natureza altamente dinâmica. Ainda, a presença de solos enxarcados e inundações, são capazes de influenciar os valores medidos das variáveis estudadas e gerar informações incorretas no momento da coleta. Embora se tenha tido o cuidado de verificar o clima e tempo na região durante as campanhas e campo, é sugerido que os levantamentos executados para coleta de dados de natureza físico-hídrica sejam parametrizados por estação do ano, condições do solo, profundidade e demais fatores ambientais que sejam relevantes para estudos de hidrologia dos solos.

Os resultados apontam a relevância da associação entre as variáveis físico-hídricas e outras propriedades do solo. Sugere-se o estudo interrelacionando as variáveis citadas com as seguintes propriedades: porosidade, densidade da partícula, densidade do solo e argila dispersa em água; com a associação da modelagem preditiva segundo as curvas de retenção e umidade do solo para análises mais minuciosas e aprofundadas.

CAPÍTULO II

Modelagem Espacial dos Atributos Físico-Hídricos dos Solos

Este capítulo deu origem ao artigo nomeado “Aprendizado de máquina aplicado n redução de Dimensionalidade de Dados Espectrais VIS-IR e Radiométricos para Mapeamento Digital de propriedades hídricas dos solos” ou, do inglês, *Machine Learning Applied at Spectral VIS-IR and Radiometric Data Dimensionality Reducing for Soil’s Hydric Properties Digital Mapping*.

RESUMO

Existe uma relação da hidrologia dos solos com a paisagem na interface pedosfera-hidrosfera que pode ser representada com o suporte e integração de dados topográficos e de Sensoriamento Remoto. O objetivo deste estudo foi analisar diferentes métodos estatísticos para seleção e redução de dimensionalidade de dados radiométricos e espectrais relacionados ao meio ambiente que auxiliem na caracterização de atributos físico-hídricos dos solos, como a velocidade de infiltração básica dos solos (vib) e a condutividade hidráulica saturada (K_{sat}), preditas a parte de técnicas de mineração de dados aplicados ao mapeamento digital de propriedades hidropedológicas. Para tal, a pesquisa integrou informações de índices e bandas espectrais do Visível ao Infravermelho ($VIS-IR$) do sensor *Multispectral Instrument* (MSI) da missão Sentinel-2A, modelagem numérica do terreno e aerogeofísica para a modelagem de água nos solos em duas camadas (0-20cm e 20-40cm). Os dados pré-processados foram submetidos a análises estatísticas (multivariada e testes de hipóteses) e, posteriormente, aplicados métodos (*variation inflation factor*, *stepwise Akaike information criterion* e *recursive feature elimination*) para mineração das covariáveis a ser utilizadas na predição de atributos hidropedológicos via *Random Forest*. Os resultados apontaram singularidades e distinção na escolha dos dados por cada método, variando-se o grau de importância e contribuição para predição das propriedades físico-hídricas. Segundo métricas estatísticas aplicadas e critérios de tomada de decisão (maior R^2 e menor $RMSE/MAE$), os métodos escolhidos foram: RFE (camadas de 0-20cm) e *StepwiseAIC* (camadas de 20-40cm), envolvendo as variáveis estudadas (vib e K_{sat}). A abordagem foi capaz de capturar a importância das variáveis ambientais, destacando o seu potencial uso no mapeamento digital hidropedológico.

Palavras-chave: Geoprocessamento. Hidropedologia. Estatística Aplicada. Radiometria. Mineração de Dados. Sensoriamento Remoto.

ABSTRACT

There is a relationship between soil hydrology and the landscape in the pedosphere-hydrosphere interface that can be represented with the support and integration of topographic and remote sensing data. Thus, the aim of this study was to analyze different statistical methods for radiometric and spectral data selection and dimensionality reduction Environment-related that can assist the classification of soil physical-hydric attributes, such as soils basic infiltration rate (bir) and saturated hydraulic conductivity (Ksat); and act in data mining process applied in digital mapping of hydrogeological properties. In order to achieve the goal, the research gathered information from Visible to Infrared (VIS-IR) spectral indexes and bands of Sentinel's 2A mission Multispectral Instrument (MSI) sensor, terrain numerical modeling and aerogeophysical data to for modeling soil-water content in two layers (0-20cm and 20-40cm). The pre-processed data were subjected into statistical analyzes (multivariate and hypothesis tests) and, subsequently, methods were applied (variation inflation factor, stepwise Akaike information criterion and recursive feature elimination) in order to mine covariates used at the Random Forest modeling procedure. The results showed distinction and singularities at choosing spectral and radiometric data for each method, varying the importance degrees and contribution of each one in relation to soils physical-hydric properties. According to the applied statistical metrics and decision-making criteria (highest R^2 and lowest RMSE / MAE), the chosen methods were: RFE (0-20cm layers) and StepwiseAIC (20-40cm layers), involving both studied variables (bir and Ksat). The approach was able to capture environmental variables importance, highlighting its potential use in hydrogeological digital mapping at Guapi-Macacu watershed.

Keywords: Geoprocessing. Hydrogeology. Applied Statistics. Radiometry. Data Mining. Remote Sensing.

1. INTRODUÇÃO

1.1. Contextualização

As técnicas de sensoriamento remoto têm contribuído significativamente para a compreensão e suporte a estudos de fenômenos ambientais que ocorrem na superfície terrestre. Isto se deve a capacidade de coleta de informações para diversos comprimentos de onda do espectro eletromagnético (multibandas) adquirida pelos sensores remotos, diferenciando variados alvos terrestres e com cobertura de grandes extensões de área. Tais técnicas tem avançado significativamente visando principalmente estudos voltados para o manejo correto dos recursos ambientais e da sustentabilidade no planeta Terra (BRENNER, 2015). Visando tais propósitos, torna-se de suma importância a aplicação das técnicas de imageamento por satélite para obtenção da resposta espectral dos alvos de interesse para que, desta forma, variáveis ambientais de difícil quantificação como a infiltração de água nos solos (vib) e aquelas decorrentes diretamente da sua medição (condutividade hidráulica saturada - Ksat) possam ser mapeadas e modeladas com base nas relações espectrais e topográficas.

Nota-se, ainda, um avanço significativo na área de geofísica ambiental voltada à estudos hidrogeológicos no qual alia dados aerolevados por sensores geofísicos com dados de campo associados à litologia, mineralogia, física e química dos solos e rochas, sendo capaz de avaliar e modelar a distribuição dos recursos hídricos em aquíferos em bacias sedimentares, por exemplo (KIRSCH, 2006; NOVAKOWSKI et al, 2006; MADRUCCI et al, 2008; LEE et al, 2012; LEE & LEE, 2015; DE ALMEIDA PIRES & MIRANDA, 2017). Entretanto, carecem estudos que tratam do potencial de integração dos mapas geofísicos em análises de favorabilidade hidrogeológica (determinação de áreas de maior potencial à ocorrência de água subterrânea) e dinâmica hidropedológica e que tratam esses dados como variáveis de entrada em Mapeamento Digital de Solos (MDS), inspirando assim a busca por aprofundamento avaliação desse potencial ferramental nos estudos de modelagem hídrica dos solos.

Existe várias referências na literatura as quais tratam da seleção e classificação de atributos topográficos visando mapear propriedades dos solos, os quais se baseiam na análise geomorfométrica de covariáveis oriundas do Modelo Digital de Elevação (MDE) para representar a relação solo-paisagem em uma dada área de estudo (BÖHNER, 2016; MCKENZIE, 1993; OLIVEIRA, 2017; SANTOS, 2019; WILSON & GALLANT, 2000). Atualmente, o mapeamento digital de atributos dos solos ainda se baseia prioritariamente

nos parâmetros morfométricos dos solos. Porém, com o avanço do sensoriamento remoto em termos de resoluções espaciais-espectrais e disponibilidade de dados, o emprego de índices baseados na relação de bandas espectrais em técnicas de modelagem espacial preditiva, como o aprendizado de máquinas (*machine learning*), tem se tornado cada vez mais presente na quantificação de variáveis de interesse (CHAGAS, 2006; PINHEIRO, 2012; CUNHA, 2013; ZHANG, 2017). Neste contexto, embora tenha crescido o uso do sensoriamento remoto em modelagem digital, verifica-se que ainda é pouco explorada a influência individual de índices espectrais na modelagem de atributos físico-hídricos do solo, sendo mais comumente verificados estudos voltados ao mapeamento de atributos físico-químicos do solo, regionalização climática e diferenciação de espécies florestais usando índices espectrais (CARVALHO JUNIOR et al., 2011; PINHEIRO et al., 2019; RAJAH et al., 2019).

A tendência no avanço do sensoriamento remoto do solo, geologia e da vegetação por análises multiespectrais trazem contribuição na caracterização da geomorfologia da paisagem e tornam possíveis análises mais robustas no que tange o comportamento espectral dos alvos imageados. Dessa forma, em locais onde coexistem uma heterogeneidade de padrões vegetativos e grandes variações climatológicas e topográficas, torna-se possível a identificação e quantificação das relações sinérgicas entre as variáveis ambientais citadas e a disponibilidade hídrica, por exemplo, por meio do conhecimento do analista (classificação supervisionada) (PIZARRO et al, 2001; BLASCHKEL & KUX, 2005; LATORRE et al, 2007; LOPES et al, 2014; ANJOS et al, 2017; PINHEIRO et al., 2019). Neste contexto, destacam-se os espectros VIS-IR do sensor multiespectral (*MSI*) da missão Sentinel-2A capazes de quantificar e reconhecer padrões sinérgicos com as variáveis ambientais estudadas, permitindo-se que estas sejam estratificadas e tenham sua dimensionalidade reduzida a ponto de garantir que a informação de saída exprima intrinsecamente a relação da paisagem com a dinâmica hidrológica, quantificada por modelos de aprendizado de máquinas (e.g. *Random Forest*) (VAN DER MEER et al, 2014; BELGIU & CSILLIK, 2018; WESSEL et al, 2018; BANGIRA et al, 2019; PETERSON et al, 2020).

Arelado a isto, fatores como a dificuldade em mensurar variáveis físico-hídricas em campo, a ausência de dados dessas variáveis em levantamento técnicos tradicionais e as restrições impostas naturalmente pela região de estudo como, por exemplo, a inviabilidade de acesso a regiões com grandes desníveis nas encostas e áreas com mata

densa, limitam o mapeamento dessas propriedades e, por consequência, a produção de pesquisas visando tal abordagem (OLIVEIRA, 2017; OTTONI, 2005).

Levando-se em consideração que os parâmetros físico-hídricos do solos são influenciados pelas propriedades genéticas do solo associados à morfologia, textura e física do solo, como por exemplo: tamanho e grau de agregação das partículas, porosidade total, densidade do solo, tipo de cobertura da superfície do solo, umidade do perfil, quantidade de matéria orgânica, entre outros (BERTONI & LOMBARDI NETO, 1990; EVERTS & KANWAR, 1992; REICHERT et al., 1992); e, sendo eles influenciados, ainda, pela variabilidade espacial da área de estudo (KLAR, 1984), avaliar os potenciais dados espectrais e radiométricos de entrada que contribuem na quantificação dessas propriedades na pré-modelagem digital torna o processo de mapeamento mais robusto e fidedigno na área de estudo.

Inspirado nesse cenário, o presente estudo tem por objetivos: (1) analisar diferentes métodos estatísticos baseados em análise bivariada, multivariada e supervisionada para seleção e redução de dimensionalidade de dados topográficos e de sensoriamento remoto (aerogeofísica e satelital) relacionados à vegetação, solo e geologia, os quais possam auxiliar na caracterização do conteúdo de água infiltrado nas camadas superficiais dos solos, através da velocidade de infiltração básica (v_{ib}), e transmitido nas camadas subsuperficiais do solo, pela condutividade hidráulica saturada (K_{sat}); (2) mapear a variabilidade espacial de propriedades hídricas dos solos da bacia hidrográfica do Rio Guapi-Macacu, estado do Rio de Janeiro, através do algoritmo *Random Forest* e utilizando as variáveis mais relevantes selecionadas na etapa anterior de mineração de dados.

1.2. Revisão: métodos de redução de dimensionalidade e seleção na modelagem digital via classificação supervisionada

A Estatística Multivariada (ou Análise Multivariada) é um ramo da estatística na qual aborda métodos de investigação das relações entre múltiplas variáveis dependentes e/ou independentes, sendo evidenciadas ou não relações de causa/efeito entre estes dois grupos. É verificado, ainda, as relações entre indivíduos caracterizados por duas ou mais variáveis. Os métodos multivariados permitem a exploração da performance conjunta das variáveis alvo, sendo capaz de determinar a influência ou importância de cada uma.

Segundo ANDREWS (1991), a autocorrelação é uma medida de similaridade na qual informa o quanto o valor de uma realização de uma variável aleatória é capaz de

influenciar seus vizinhos, medida pela distância entre os dados. Segundo as diretrizes estatísticas, o valor da autocorrelação está entre 1 (correlação perfeita) e -1 (anticorrelação perfeita ou correlação inversa); já o valor 0 (zero), significa a total ausência de correlação. Os vizinhos os quais estão sendo tratados neste estudo são as variáveis independentes, ou seja, os índices e bandas espectrais, os quais podem se correlacionar ou não com a variáveis dependente, que é a condutividade hidráulica saturada. Essa se faz uma análise interessante pois evidenciará, por exemplo, a presença de um valor mais alto dentre o conjunto total de variáveis no qual condiciona valores também altos de seus vizinhos (análise de matriz de correlação).

Outra abordagem é avaliar a multicolinearidade, que consiste em um problema comum em regressões, no qual as variáveis independentes possuem relações exatas ou aproximadamente exatas linearmente entre si, sendo verificada a sua presença mais claramente quando o coeficiente de determinação (R^2) é bastante alto, porém nenhum dos coeficientes da regressão é estatisticamente significativo segundo a estatística t de Student convencional (O'HAGAN et al., 1975). As consequências desta análise em uma regressão vão de erros-padrão elevados no caso de multicolinearidade moderada ou severa e até mesmo a impossibilidade de qualquer estimação se a multicolinearidade for perfeita (MIHOLA, 2014). O mais importante neste método, e que será tratado neste trabalho, é o caso em que ocorre um R^2 estimado for alto e as correlações parciais das variáveis independentes são baixas, inferindo a possibilidade de multicolinearidade nos dados e, quando isso ocorre, uma ou mais variáveis podem ser desnecessárias no modelo, sendo necessária a redução de variáveis independentes no modelo (BREUSCH et al., 1978; JING et al., 2007). Os efeitos negativos da multicolinearidade na modelagem são ocasionados não somente pela sua presença, mais sim pelo grau com que se manifesta. Dentre os efeitos causados pelo elevado grau de multicolinearidade, podem ser citados as estimativas inconsistentes dos modelos e a superestimação dos efeitos diretos das variáveis explicativas sobre a variável predita, o que pode levar à interpretação equivocada (FERRARIA, 1989; MONTGOMERY & PECK, 1981).

O grau de multicolinearidade não constitui problemas para a modelagem quando este é considerado fraco (CARVALHO et al., 1999). Ademais, quando identificado um grau de moderado à forte na multicolinearidade, deve ser analisado cautelosamente o problema e métricas estatísticas devem ser aplicadas para atenuar o problema. Uma forma de controlar tal problema é identificar quais variáveis causam o inflacionamento da

multicolinearidade no modelo regressivo. Em ambiente de programação, a identificação e seleção das variáveis que maximizam a estimativa do coeficiente de correlação, que apresentam a menor interferência de multicolinearidade possível, podem ser realizadas por meio do procedimento conhecido como *Stepwise* (BELSLEY et al., 1980).

O *stepwise* é um método de ajuste de modelos de regressão em que a escolha de variáveis preditivas é realizada por um procedimento automático e iterativo, em que a cada etapa (iteração), uma variável é considerada para adição ou subtração do conjunto de variáveis explicativas com base em um critério pré-especificado. Normalmente, assume-se uma sequência de testes F-Snedecor (análise de variância) ou testes *t-Student*, mas outras técnicas são possíveis, como R^2 ajustado, critério de informação de *Akaike* (do inglês, *Akaike Information Criterion* - AIC), critério de informação Bayesiano (do inglês, *Bayesian information criterion* – BIC), estatística C_p de Mallows, Soma Dos Quadrados Do Erro Residual Previsto (do inglês, *Predicted Residual Error Sum of Squares* – PRESS) ou taxa de descoberta falsa.

As principais abordagens do método *stepwise* são:

- Seleção direta (*Forward Selection*), que envolve começar sem variáveis no modelo e ir testando a adição de cada variável usando um critério de ajuste do modelo escolhido. A inclusão garante uma significativa melhoria estatística no ajuste. O processo é repetido até que não haja modificação significativa no modelo;
- Eliminação reversa (*Backward elimination*), que envolve iniciar o teste com todas as variáveis candidatas e ir excluindo cada variável, uma a uma, usando um critério de ajuste do modelo escolhido. A variável é excluída se houver degradação do modelo, o que causa insignificância no ajuste. O processo é repetido até que não haja outras variáveis a serem excluídas sem uma perda de ajuste estatisticamente insignificante;
- Eliminação bidirecional (*Bidirectional Elimination*), cuja ideia é combinação das opções acima, testando em cada etapa as variáveis a serem incluídas ou excluídas.

Outra abordagem de multicolinearidade é fator de inflação de variância (do inglês, *Variance Inflation Factor* – VIF), o qual trata-se do quociente entre a como variância de um modelo com vários termos pela variância de um modelo com um único termo. Ele quantifica a gravidade desta em uma análise de regressão de mínimos quadrados ordinária. Ele fornece um índice que mede o quanto a variância (o quadrado do desvio

padrão da estimativa) de um coeficiente de regressão estimado é aumentada devido à colinearidade (DANIEL & WOOD, 1980).

ADBI et al. (2010) define a análise de componentes principais (ACP) como sendo uma técnica multivariada que analisa uma tabela de dados na qual as observações são descritas por várias variáveis dependentes quantitativas inter-correlacionadas cujo objetivo é extrair as informações importantes da tabela, representá-las como um conjunto de novas variáveis ortogonais chamadas componentes principais e exibir o padrão de similaridade das observações e das variáveis como vetores multidimensionais. A qualidade do modelo ACP pode ser avaliada usando técnicas de validação cruzada, como o *bootstrap* e o *Jackknife*. Ao lidar com conjuntos heterogêneos de variáveis quantitativas, é indicado o uso de ACP por análise de múltiplos fatores (*Multiple Factor Analysis - MFA*) para redução de dimensionalidade de dados. A padronização dos dados na ACP por este método é dada pela matriz de correlação (R) e o resultado é a quantidade de componentes principais necessárias para subdivisão dos dados em grupos menores e definição das variáveis mais explicativas a um determinado estudo, com base na decisão de análise do operador (JOLLIFFE, 1973).

A eliminação recursiva de feições (do inglês, *Recursive Feature Elimination - RFE*) é um dos métodos de seleção existentes no universo do *Data Mining* o qual atua na etapa pré-modelagem (modelagem regressiva ou classificação supervisionada/ não-supervisionada) para aplicação de modelos de aprendizado de máquinas (BLUM & LANGLEY, 1997; BRADLEY et al., 1998). Este método visa a escolha das variáveis preditoras que melhor ajustam o modelo desejado, seja ele regressivo, multivariado ou ML, e removendo, ainda, os “recursos” (preditores) usados no banco de dados de treinamento do modelo que sejam menos relevantes na predição da variável almejada (variável predita) até que o número especificado de recursos seja alcançado (JEBARA & JAAKKOLA, 2000; WESTON et al., 2000). Desta forma, o resultado do RFE é a escolha das variáveis preditoras que otimizam melhor o modelo que se deseja estabelecer (SVETNIK et al., 2004).

O RFE (GUYON et al., 2002) é basicamente uma seleção retroativa dos preditores de um modelo aplicado. Essa técnica inicia-se pela construção de um modelo em todo o conjunto de preditores e calculando uma pontuação de importância (ranqueamento) para cada preditor. Em seguida, os preditores menos importantes são removidos baseado em uma análise de importância dos dados visando atingir um modelo ajustado com menor

valor medido do erro quadrático médio (RMSE). Em seguida, o modelo é reconstruído, gerando novas pontuações de importância computadas aos preditores. Em suma, o analista especifica o número de subconjuntos preditores a serem avaliados, bem como o tamanho de cada subconjunto. Portanto, o tamanho do subconjunto é um parâmetro de ajuste para o RFE. O tamanho do subconjunto que otimiza os critérios de desempenho é usado para selecionar os preditores com base nas classificações de importância. O subconjunto ideal de preditores é então usado para treinar o modelo final, sendo este o objetivo pelo qual busca-se a presente pesquisa. Esse tipo de seleção é conhecido como *Backward Selection* e é frequentemente usada com modelos de floresta aleatórias - FA (ou, do inglês, *Random Forest* - RF), pois o ato de restringir o número de preditores utiliza-se dos princípios de multicolinearidade para mensurar a importância das árvores de decisão e, assim, selecionar as variáveis que serão usadas no modelo final, já que reduz a redundância dos preditores nas pontuações de importância.

No RFE, existem duas opções de configuração importantes: a primeira, baseada na escolha do número de “recursos” a serem selecionados e, o segundo, a definição do algoritmo usado para auxiliar a escolha desses recursos. Ambas as configurações podem ser exploradas, embora o desempenho do método não dependa fortemente desses parâmetros serem bem configurados (SVETNIK et al., 2004). O procedimento de seleção é bastante simples: os recursos (variáveis preditoras) são classificados pelo modelo escolhido e, ao eliminar recursivamente um pequeno número de recursos por iteração (*loop*), o RFE tenta eliminar dependências (principalmente entre a variável predita e suas variáveis preditoras) e colinearidade (multicolineridade) que podem existir entre as variáveis e que afetam de forma negativa a modelagem (Kuhn & Johnson, 2013).

O RFE requer um número específico de recursos para avançar no ajuste do modelo, porém a quantidade ideal demandada é obtida através da implementação do algoritmo junto a um teste de validação cruzada (do inglês, *Cross Validation* - CV), para pontuar diferentes subconjuntos de recursos e selecionar a melhor pontuação para a coleção de recursos analisada. O sucesso na remoção das quantidades de variáveis desejadas é obtido ajustando o algoritmo de aprendizado de máquina fornecido usado no núcleo do modelo, classificando os recursos por importância, descartando os recursos menos importantes e reajustando o modelo. Este processo é repetido até que um determinado número de recursos permaneça (KUHNS & JOHNSON, 2013).

A seleção de feições ou variáveis pode ser feita por meio de cálculos de importância podem ser baseados em modelo (por exemplo, o critério de importância do *Random Forest*) ou usando uma abordagem estatística mais geral que é independente do modelo completo (KUHNS & JOHNSON, 2013; EVANS & CUSHMAN, 2009; EVANS et al., 2011). As funções aplicadas no RFE podem ser baseadas em modelos de regressão linear (*Linear Regression*), árvores de regressão encapsuladas (*Bagged Trees*) e florestas aleatórias (*Random Forest*), enquanto as validações cruzadas podem basear-se nos métodos: *Cross Validation* (CV) tradicional, *Leave One Out Cross-Validation* (LOOCV), *K-fold cross-Validation* (KFCV) e *Repeated K-fold cross-validation* (RKFCV), e outros. Cada tipo de validação citada possui suas vantagens e desvantagens, sendo sugerida a escolha do melhor tipo para os dados treinados visando o objetivo do analista na modelagem. Além disso, a existência de alguns dos métodos de validação citados depende diretamente do tipo de pacote usado para implementar o RFE e da linguagem escolhida para o processamento (por exemplo, o pacote *caret* do RStudio e o pacote *scikit learn* do Python). É requerido ainda o parâmetro de sintonização (*tuning*) que determina o ajuste do modelo a uma certa quantidade de iteração determinada pelo analista, nos casos em que a função implementada no algoritmo é baseada em árvores. Como métrica, o RFE é capaz de avaliar e minimizar um dos três parâmetros estatísticos de avaliação da acurácia e precisão da modelagem por tentativa, sendo estes: *Root Mean Squared Error* (RMSE), *Mean Absolute Error* (MAE) e o R^2 (coeficiente de determinação) (SCHAAP & LEIJ, 1998; SCHAAP, 2004).

A utilização de *machine learning* na área de solos tem crescido substancialmente nos últimos anos. Pesquisadores da área de mapeamento digital tem utilizado cada vez mais de técnicas do chamado aprendizado de máquina em suas pesquisas (PADARIAN et al, 2020). Com a popularização da técnica, diversos trabalhos com a temática vêm surgindo nos últimos anos. Vale ressaltar, porém, que por não ser uma técnica desenvolvida por para aplicações de solo, diversas abordagens e definições podem ser vistas. Quando tratado puramente de aprendizado de máquinas, o uso de todas as variáveis disponíveis é recomendado, a fim de que o aprendizado do algoritmo será otimizado (ROSSITER, 2018). Tal fato não é visto no MDS, já que, do ponto de vista da ciência de solos, nem sempre uma variável pode ser representativa ou explicativa de um determinado fenômeno. Assim, o conceito de *machine learning* (ML) deve ser visto e aplicado com cautela (PADARIAN et al., 2020). Alguns modelos de aprendizado de máquina se

popularizaram nas últimas décadas no mapeamento digital de solos, sobretudo os modelos derivados de árvore, como as árvores de decisão e regressão. Abaixo, será abordado os conceitos básicos do *Random Forest*, modelo de *machine learning* utilizado durante a pesquisa.

Os modelos ML de árvores baseiam-se na criação de uma hierarquia de nós internos e externos conectados por ramos, sendo o nó interno ou raiz da unidade a tomada de decisão, que avalia qual será o próximo nó descendente ou filho realizando um teste lógico. O nó externo não possui descendente, e é conhecido também como folha ou nó terminal, sendo associado a um rótulo ou um valor (COLIN et al., 2017). Tratando-se do modelo *Random Forest* (RF) sua forma de aprendizado e hierarquização baseia-se no conjunto de árvores de classificação (para variáveis categóricas) ou regressão (variáveis contínuas) randomizados (BREIMAN, 2001). Ambos os casos um grande número de árvores é gerado dentro do algoritmo e então são agregados para se obter um único valor ou classe de predição. O RF é frequentemente uma coleção de centenas de milhares de árvores, onde cada árvore é cultivada usando uma amostra de *bootstrap* dos dados originais. Agregação de *bootstrap* é um procedimento geral que pode ser usado para reduzir a variação para aqueles algoritmos que possuem alta variância.

O RF depende somente de três parâmetros definidos pelos usuários: o número de árvores (*ntree*) na floresta, o número mínimo de pontos de dados em cada nó terminal (*node size*), e o número de variáveis usadas para produzir cada árvore (*mtry*). Geralmente os valores indicados pela literatura são: *ntree*=500, *nodesize*=5 e *mtry*= um terço do número total de preditores, para predição de dados contínuos. No entanto, podem ser testados outros valores objetivando o menor erro, para isso o RF fornece estimativas de erro, usando o chamado *Out-Of-Bag* (OOB) de dados (que é uma porção dos dados não utilizada no subconjunto de *bootstrap*) (GRIMM et al., 2008; DUBE & MUTANGA, 2014; TAGHIZADEH-MEHRJARDI et al., 2015).

O critério de importância gerado pelo modelo *Random Forest* (*Importance Rank*) é um dos parâmetros de qualidade usados para verificar a porcentagem explicativa das variáveis preditas (vib/Ksat) (KUHN & JOHNSON, 2013; EVANS & CUSHMAN, 2009; EVANS et al., 2011). O RF pode ainda, avaliar e minimizar um dos três parâmetros estatísticos de avaliação da acurácia e precisão da modelagem por tentativa, sendo estes: *Root Mean Squared Error* (RMSE), *Mean Absolute Error* (MAE) e o R^2 (coeficiente de determinação) (SCHAAP & LEIJ, 1998; SCHAAP, 2004). Além disso, o RF informa a

variação e minimização dos resíduos do ajuste do modelo à medida que a iteração das árvores é executada (análise gráfica do número de árvores versus erro residual).

2. MATERIAL E MÉTODOS

2.1. Área de Estudo

A área selecionada para este estudo é composta pela Bacia Hidrográfica do Rio Guapimirim-Macacu, localizada no domínio da Região Hidrográfica Baía de Guanabara (RH-V), região metropolitana estado do Rio de Janeiro (Figura 30). Integrando os limites políticos-administrativos de Itaboraí, Guapimirim e Cachoeiras de Macacu, a bacia possui dimensões de área de captação e perímetro de 1,250.78 km² e 199.2 km, respectivamente, e o relevo caracteriza-se pela elevação variante de 0 a 2254 m, verificados através do Modelo Digital de Elevação (MDE) da região (resolução espacial de 20 m) (Figura 30).

O clima predominante é tropical chuvoso com invernos secos (Projeto Macacu, 2010), com temperatura média anual variando em torno de 23°C e precipitação média anual entre 1.200-2.600 mm devido a contrafortes serranos. A vegetação se insere no bioma Mata Atlântica, com porções de floresta tropical e vegetação característico do ambiente de Mar de Morros com transição para Baixada Litorânea (PINHEIRO et. al., 2016).

Geologicamente, a região é caracterizada pelo Gráben da Guanabara, no qual discrimina um vale alongado com fundo plano, onde ocorreram falhas geológicas que deram origem a uma sequência de deposição sedimentar oriunda de atividade tectônica iniciada no Terciário onde deu-se a formação da Bacia Sedimentar de Macacu (FERRARI, 2001). As feições geomorfológicas relacionam-se com eventos deposicionais ocorridos na região, de características aluviais, fluviais e lacustrinos (PINHEIRO et. al., 2016).

Os solos da região apresentam características intrínsecas de ambientes de deposição sedimentar devido à influência lacustre e fluvial da geomorfologia local. O ambiente de transição de mar de morros para campos de altitude faz com que a bacia seja circundada por vales e montanhas, alcançando altitudes de até 2254 metros, o que leva à formação e caracterização de solos de grande variabilidade pedogenética e diversidade taxonômica, como: Argissolos, Cambissolos, Gelissolos, Latossolos e Neossolos (PINHEIRO et al, 2015).

2.2. Banco de Dados e Testes Hidropedológicos

O banco de dados pedológico contém as descrições morfológicas e as análises físicas e químicas das amostras coletadas, como parte importante do estudo de caracterização dos solos da bacia. Os dados foram coletados na área da bacia segundo a técnica do Hipercubo Latino (do inglês, *Conditioned Latin Hypercube Sample* ou cLHS), que realiza a inferência de pontos selecionados na área por meio de recursos computacionais que apresentem uma maior representatividade das características ambientais condicionadas das bacias do rio Guapi-Macacu (MCKAY et al., 1979; MINASNY & MCBRATNEY, 2006; CARVALHO JÚNIOR et al., 2014).

Os dados de velocidade de infiltração básica dos solos (vib) foram obtidos por meio de campanhas de levantamento hidropedológico realizado *in situ* na área da bacia hidrográfica, com o uso do Permeâmetro de Guelph, modelo 2800K1, da empresa Soil Moisture, cujo método de medição associado foi elaborado por REYNOLDS & ELRICK (1985) e aperfeiçoado posteriormente pela *University of Guelph*, no Canadá em 1985 (ELRICK, 1985). Através da aplicação de funções matemáticas de solução da equação de Darcy e suas evoluções (DARCY, 1856; REYNOLDS & ELDRICK., 1985; RICHARDS, 1931) torna-se possível obter os valores de condutividade hidráulica saturada (Ksat) estimada a partir da vib, usando fórmulas com parâmetros conhecidos e estimados em laboratório.

A partir dos dados vib mensurados, pode-se obter os valores de Ksat por transformação em função dos parâmetros do equipamento (e.g. coluna d'água, carga hidráulica e diâmetro da coluna d'água), características do solo (porosidade, gênese e coeficiente de armazenamento) e funções matemáticas de solução da equação de Darcy e suas evoluções (DARCY, 1856; REYNOLDS & ELDRICK., 1985; RICHARDS, 1931). No total, foram obtidos de trinta e seis (36) dados mensurados para ambos os atributos físico-hídricos na área de estudo. Os dados mensurados foram separados em quatro níveis, sendo eles segundo a condição hidrológica dos solos (hidromórfico e não-hidromórfico) e tipo de profundidade (superficial ou subsuperficial) e, por fim, submetidos à calibração de funções de pedotransferência junto com propriedades intrínsecas dos solos (e.g. composição granulométrica, densidade do solo, ph da água, argila dispersa em água, porosidade, densidade da partícula, carbono orgânico e Valor T) para estimativa do montante de oitenta e seis (86) pontos restantes na bacia, resultando num total de 122 pontos contendo informações superficiais e subsuperficiais das propriedades analisadas

(Figura 30), os quais compõem o conjunto de dados analisado nesta pesquisa. As etapas descritas anteriormente foram executadas pelos autores em análise prévia a esta pesquisa.

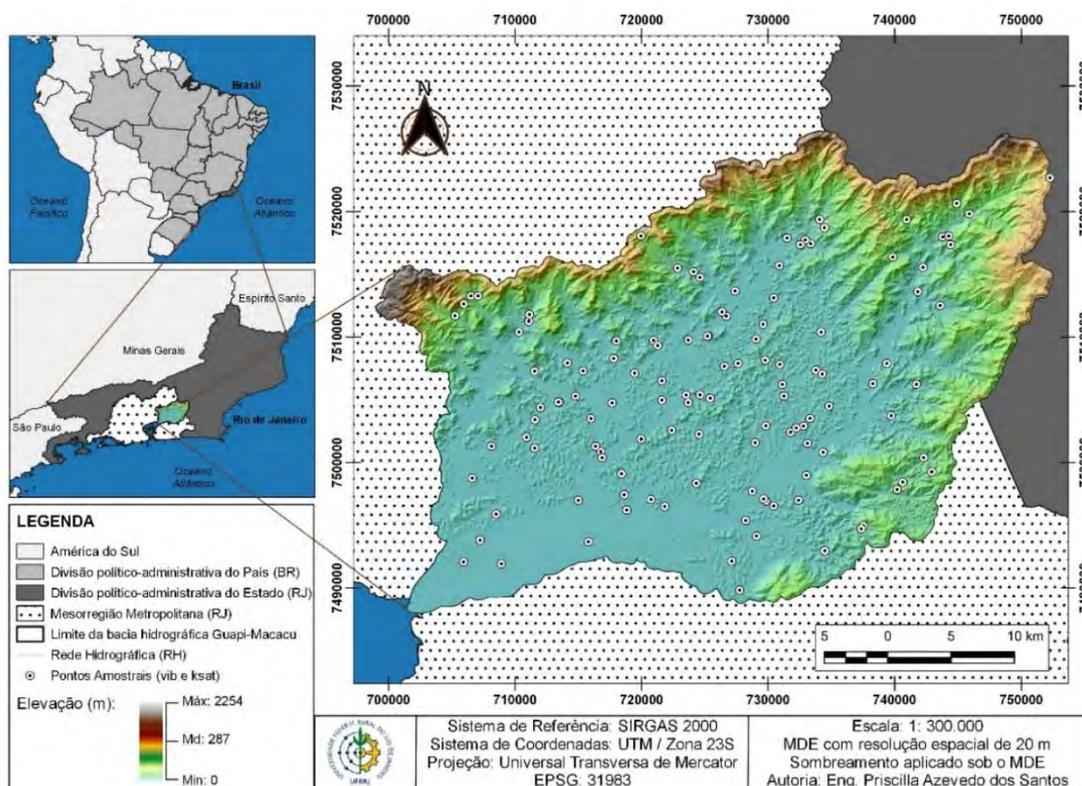


Figura 30. Mapa de localização da área de estudo e distribuição espacial dos 122 pontos associados às informações hidropedológicas do conjunto de dados analisado na pesquisa.

2.3. Modelagem numérica do terreno: extração de covariáveis e suas relações com a paisagem na bacia

O SAGA GIS, Sistema de Informações Geográficas (SIG) voltado para análises geocientíficas, foi usado para obter os mapas topográficos representativos dos parâmetros morfométricos da superfície, derivados de um modelo digital de elevação hidrologicamente consistido representativo das altitudes ortométricas (altimetria) da superfície topográfica agregada aos elementos geográficos existentes sobre a bacia, como por exemplo a cobertura vegetal. Os dados vetoriais usados para gerar o modelo digital de elevação são oriundos do banco de dados cartográfico do repositório de Geociências do Instituto Brasileiro de Geografia e Estatística (IBGE) em parceria com a Secretaria Estadual do Ambiente (SEA), constituindo a base cartográfica contínua na escala 1:25.000 do levantamento planialtimétrico do estado do Rio de Janeiro.

Para o recorte dos dados na área de interesse usou-se como máscara o limite da região político-administrativa da bacia obtidos na base de dados geoespaciais do portal de Geociências do Instituto Estadual do Ambiente (GeoINEA), delimitadas a partir de dados compilados da série de cartas topográficas na escala 1:50.000 que compõe o Sistema Cartográfico Nacional (SCN). O limite da bacia também foi padronizado segundo o datum e sistema de coordenadas admitidos no tratamento dos dados vetoriais do IBGE. A partir do recorte dos dados para a área de referência, prosseguiu-se com a extração do MDE baseando-se no interpolador *TopoToRaster*, em ambiente ArcGIS 10.6, ferramenta essa projetada para a criação de modelos digitais de elevação com consistência hidrológica. Os dados vetoriais foram interpolados usando uma resolução espacial de 20m, dando origem ao MDE para a região. Em seguida, passou por análise de consistência hidrológica através do uso das ferramentas *Flow Accumulation*, *Flow Direction* e *Fill*, tornando-se um Modelo Digital de Elevação Hidrologicamente Consistido (MDEHC). Como resultado, obteve-se o mapa de elevação da região estudada, conforme mostrado na Figura 30.

As covariáveis do terreno foram selecionadas criteriosamente de forma a compor a paisagem da bacia, levando em consideração os elementos ambientais da pedopaisagem (solo, vegetação, hidrologia e geologia) nela coexistentes. Os dados matriciais radiométricos (*rasters*) e suas características são vistos no Quadro 6 (anexo B).

Os trinta e seis (36) atributos primários e secundários selecionados apresentam grande expressividade na representação das propriedades ambientais e processos ativos da paisagem, caracterizando os fatores de formação do solo, principalmente o relevo. Portanto, a prioridade na escolha das covariáveis representativas da bacia envolveu a extração de superfícies contínuas a partir da aplicação das ferramentas *Hydrology*, *Lighting*, *Visibility*, *Morphometry*, *Slope Stability* e *Channels*; módulo *Terrain Analysis* do SAGA GIS, capazes de quantificar as variáveis ecossistêmicas de montanhas (campos de altitude) e bioma nativo da região (bioma de Mata Atlântica).

2.4. Aquisição e Processamento dos Dados Aerogeofísicos e Imagens Óticas Sentinel-2A

Os dados aerogeofísicos de magnetometria e gamaespectrometria empregados no estudo são originados do repositório online do sistema de geociências do Serviço Geológico do Brasil (GeoSGB) da Companhia de Pesquisa de Recursos Minerais (CPRM), onde encontram-se as bases aéreas dos Projetos de Levantamentos

Aerogeofísicos executados pela CPRM e outras instituições. Nesta base encontram-se dados de sensores aerotransportados para determinação de valores de parâmetros magnetométricos, gamaespectrométricos, gravimétricos, radiométricos, dentre outros, do território nacional. As informações referentes ao projeto de levantamento de dados aerogeofísicos estão informadas no Quadro 7 (anexo B).

Além das informações acerca dos projetos aerogeofísicos, tem-se ainda as características técnicas das aeronaves usadas na execução do projeto e o processo de coleta e processamento dos dados adquiridos (Quadro 8, anexo B).

Os dados digitais aeromagnetométricos e gamaespectrométrico brutos foram obtidos da base de projetos aerogeofísicos do sistema de Geociências do Serviço Geológico do Brasil (GeoSGB), com resolução espacial de 500m, e processados usando o *software* Geosoft Oasis Montaj v. 9.0. Os dados aeromagnetométricos foram tratados para correção de problemas intrínsecos à natureza do aerolevante e dos equipamentos usados para a execução do mesmo (correção de erro de paralaxe, remoção da variação diurna, nivelamento e micronivelamento dos perfis e definição da superfície de tendência do campo geomagnético – IGRF), procedimentos estes executados pela CPRM no seu relatório final de processamento (CPRM, 2012). Os dados tabulados já corrigidos foram geoespacializados no software para o georreferenciamento e criação do grid radiométrico bi-direcional do Campo Magnético Anômalo (CMA) usando como canais principais as coordenadas (x, y, z) e o valor magnetométrico medido pelo International Geomagnetic Reference Field (IGRF), interpolado em um máximo de 1/5 das linhas de vôo para que não haja perdas nas assinaturas magnetométricas pós-processamento (Módulo *MagMap*), preenchendo-se ainda a ausência de valores no grid através de interpolação *dummy* usando o método *Square* (módulo *Grid and Image*). O mapa do campo magnético anômalo (CMA) gerado pelo grid radiométrico, com resolução espacial de 100 m foi gerado a partir da correção do IGRF (cálculo da transformada rápida de Fourier mudando para o domínio da frequência). Em seguida, aplicou-se um filtro ASA (Amplitude do Sinal Analítico), na qual busca a taxa de variação (gradiente) nos três eixos (x, y e z) e reduz interferências do CMA na identificação de domínios magnetométricos que podem estar associados às rochas aquíferas e mineralogia.

Como o Rio de Janeiro encontra-se numa região de baixa latitude, simulamos o RTP (*Reduction to Pole*) com derivada em relação ao eixo x, através da geração do mapa da Amplitude do Sinal Analítico (ASA). Essa simulação permite ressaltar as bordas das

feições geológicas (mineralizações mais superficiais). Por final, têm-se taxas de variação das assinaturas espectrais em comprimentos de onda mais curtos, visando destacar os corpos menores (mineralogia e rochas menores), representados pelo mapa ASA. Os parâmetros necessários do IGRF para a redução foram: a inclinação (-35.64), a declinação (-21.69), a amplitude corrigida (-54.36) para a data do aerolevanteamento (08/01/2012) – separador “ponto” para entrada dos dados. Os dois primeiros parâmetros são calculados automaticamente pelo Oasis Montaj a partir da entrada da data do aerolevanteamento.

Os dados passaram por tratamento e correção, procedimentos estes executados pela CPRM, levando em consideração as seguintes operações sequenciadas: correção do tempo morto; filtragem de altura, radiação cósmica e canal de Urânio *upward* (mudança da contagem para segundos); correção de paralaxe; remoção de background da aeronave e cósmica (CPRM, 2012); usando o módulo *Radiometric Processing System*. O resultado obtido é dado pelas colunas de contagem dos elementos corrigidos na tabela de entrada dos dados, os quais tiveram suas coordenadas tridimensionais (x, y e z) geoespacializadas no software Oasis Montaj para criação do *grid* radiométrico. Para a criação do grid dos elementos Urânio equivalente (eU), Tório equivalente (eTh) e Potássio (K), foi usado o interpolador de mínima curvatura, indicado para tais dados conforme orientações da CPRM (2012), com base em 1/5 das linhas de vôo para se equiparar à resolução espacial dos dados aeromagnetométricos (100 m), corrigindo-se ainda a ausência de informações do grid através de interpolação *dummy* usando o método Square (módulo *Grid and Image*). Por fim, gerou-se o mapa ternário radiométrico em composição RGB (vermelho, verde, azul) relativo às bandas dos radioelementos (R-potássio, G-tório, B-urânio) representadas pelo triângulo ternário, usando o módulo *Grid and Image*.

As imagens multiespectrais do sensor *Multispectral Instrument* (MSI) da missão Sentinel-2 foram selecionadas mediante a proximidade de datas e disponibilidade no repositório do *Copernicus Sentinel Hub gerenciado pela European Space Agency* (ESA) através do programa GMES (*Global Monitoring for Environment and Security*), de forma a coincidir com a época de coleta dos dados hidropedológicos. O tratamento e processamento realizado baseou-se na transformação das informações de radiância para reflectância de superfície, além de permitir a coincidência de resoluções espaciais entre as bandas do sensor, transformando as de 10 para 20 m, visando compatibilizar os pixels para execução de álgebra de bandas (cálculo dos índices). As bandas espectrais usadas na análise foram de B2 até B8A, a B11 e a B12 do satélite 2A da missão Sentinel-2 para

identificar possíveis padrões relacionais dessas com os atributos físico-hídricos, cujas características estão descritas no Quadro 9 (anexo B). As Funções de Resposta Espectral – FRE (do inglês, *Spectral Response Functions*) para cada faixa do espectro eletromagnético imageada pelo sensor MSI do satélite Sentinel 2-A é mostrada na Figura 31.

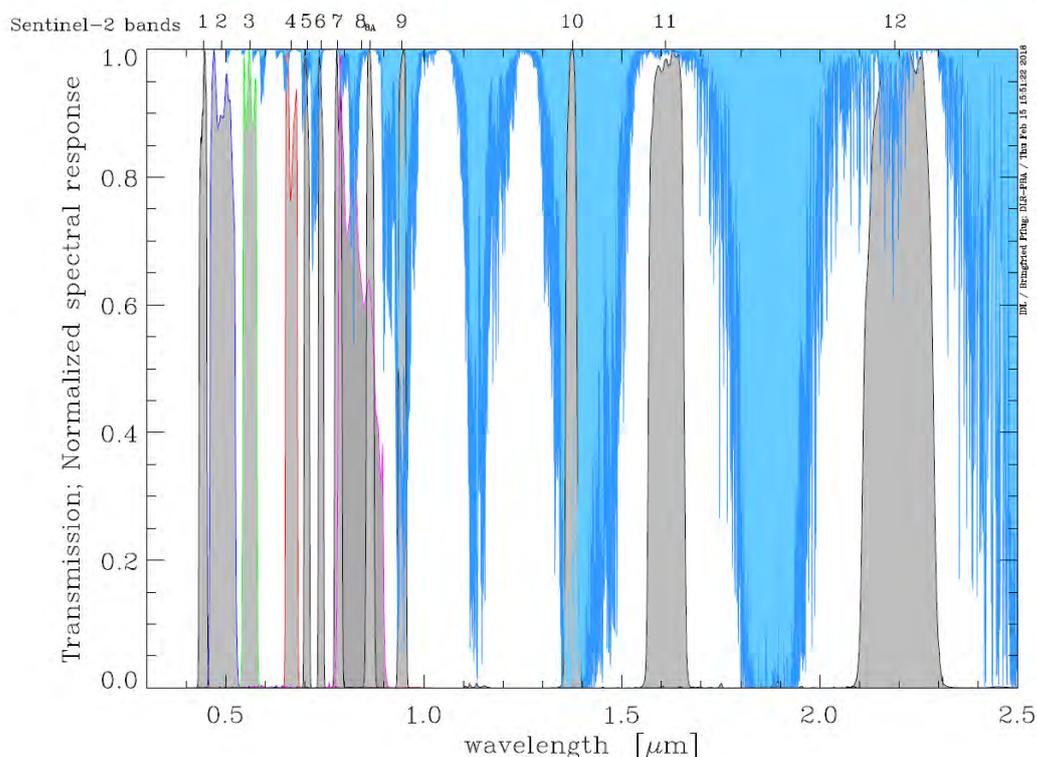


Figura 31. Funções de Resposta Espectral (FRE) do satélite Sentinel-2A (S2A) e transmissão devido à coluna vertical de absorção de Vapor D'água (VD) para a atmosfera subártica de inverno (azul claro, menor teor de VD) e atmosfera tropical (azul mais escuro, maior teor de VD). Os dados de transmissão VD são oriundos do programa computacional MODTRAN (do inglês, MODerate resolution atmospheric TRANsmission). O eixo x refere-se ao comprimento de onda do espectro eletromagnético (Wavelength) e, o eixo y, ao valor de transmissão e resposta espectral normalizada. Fonte: ESA (2021).

Em posse das bandas VIS-IR, foram calculados os índices espectrais. A escolha de índices relacionados ao espectro da vegetação, solo e geologia foi feita criteriosamente pensando na possível relação destes com as variáveis de estudo, os quais se encontram listados no Quadro 10 (anexo B). Índices espectrais são interessantes para estudos radiométricos ambientais por ressaltar feições de absorção/reflectância da Radiação Eletromagnética (REM) que contribuem para o entendimento do comportamento espectral desses materiais na composição da superfície terrestre.

Todos os procedimentos de tratamento e processamento dos dados espectrais foram realizados através da implementação de rotina estatística em ambiente RStudio v.1.2.1335 (R CORE TEAM, 2020), com uso específico do pacote *sen2R* para o processamento das imagens Sentinel-2A. Ainda, obteve-se os índices usando imagens de 9 bandas espectrais (*VIS-IR*) da missão Sentinel 2A através de matemática de bandas neste ambiente.

2.5. Modelagem: Métodos de Seleção e Redução de Dimensionalidade dos dados, *Random Forest* e Validação

O uso de métodos distintos de seleção de variáveis para modelagem (predição) ambiental visou reduzir a dimensionalidade dos dados a partir da eliminação de parâmetros que possuem correlação alta e/ou baixa entre si ou que, ainda, não possuem correlação com as variáveis de estudo (vib e Ksat). Ademais, os métodos propostos abrangem perspectivas distintas de separação de dados, o que nos leva a discussão sobre a importância desta como uma etapa decisiva na pré-modelagem, dada que a resposta obtida por cada método influenciará posteriormente o uso de algoritmos de aprendizado de máquinas e outros modelos existentes na literatura para mapear atributos ambientais tais como os analisados neste estudo.

Primariamente, os dados de entrada foram analisados segundo quatro análises estatísticas de adequação para aplicação nos métodos e, ainda, minimização de erros e perdas de informação (análise de autocorrelação, multicolinearidade, componentes principais e testes de hipóteses). Posteriormente, para a seleção e redução da dimensionalidade dos dados foram aplicados métodos estatísticos baseados em análises estatísticas regressiva, multivariada e de aprendizado de máquinas (*variation inflation factor*, *stepwise Akaike information criterion* e *recursive feature elimination*), implementados por meio de rotina construída em ambiente RStudio.

Os parâmetros e hiper parâmetros associados aos métodos de seleção e ao modelo ao *Random Forest*, respectivamente, são distintos e foram ajustados conforme as informações indicadas nos itens a seguir:

- para implementação do vif, usou-se a função *vifstep* no modo padrão (*default*), alterando apenas o limite de variáveis (*threshold*) igual a 10;
- para a implementação do *StepwiseAIC*, foi aplicada a função *stepAIC* para calibração das equações geradas pela função *lm* (ajuste e composição da RLM à priori). Aplicou-se os parâmetros do *default* da função;

- para implementação do RFE aplicou-se a função *rfe* em linguagem R, com os seguintes parâmetros: controle de seleção das variáveis pelo uso da função *rfeControl*; a métrica usada para minimização do erro foi o RMSE. No controle do *rfe*, os parâmetros aplicados foram: método de validação Leave-One-Out Cross-Validation (LOOCV); número de dobras ou número de iterações de reamostragem igual a 5 repetições;
- para implementação do modelo *Random Forest* foi usada a função *randomForest* em linguagem R, com os hiperparâmetros conforme o descrito a seguir: número de variáveis randomicamente amostradas como candidato de cada nó/separação (*mtry*) igual $P/3$, onde p é o número de variáveis observadas em análise regressiva; número de árvores para crescimento (*ntree*) igual a 500; critérios de divisão em cada nó foi o MSE; a profundidade máxima das árvores individuais (*max nodes*) e as amostras mínimas para dividir em um nó interno (*sample size*) foram usados o *default* da função; tamanho mínimo dos nós terminais (*node size*) igual a 5 (adequado para regressão). A importância dos preditores foi habilitada e avaliada na função. Houve aplicação de ajuste (*tune*) para otimização do *mtry* através da aplicação da função *tuneRF*, com os seguintes parâmetros associados: número de árvores usadas na etapa de ajuste (*ntreeTry*) igual a 500; valor de inflação (ou esvaziamento) do *mtry* a cada iteração (*stepFactor*) igual a 2; demais parâmetros usados em *default*.

A determinação do método de melhor desempenho na estimativa dos atributos físico-hídricos foi dada segundo os parâmetros de qualidade fornecidos pelos valores dessas propriedades preditas via algoritmo *Random Forest*, sendo este o modelo escolhido para a classificação dos recursos hídricos devido a sua fácil implementação e robustez (BLUM & LANGLEY, 1997; BRADLEY et al., 1998).

Para implementação do RF, os dados de entrada foram separados na porção 70% e 30%, para treinamento e validação dos métodos avaliados nesta pesquisa, respectivamente. O critério de qualidade dos modelos compostos pelas variáveis selecionadas e/ou reduzidas do banco de dados através dos métodos implementados se deu por meio de análise estatística criteriosas sob os dados verificando a acurácia e precisão (índices estatísticos) dos valores de vib/Ksat estimados através dos modelos, à saber: *Mean Absolut Error* (MAE), *Root Mean Squared Error* (RMSE) e coeficiente de determinação (R^2); e, ainda, pela avaliação de importância do modelo *Random Forest*.

Atrelada a isto, a escolha do modelo de seleção no qual apresentou melhor performance de predição e maior grau de acerto foi feita mediante a implementação de validação utilizando 30% dos dados amostrais não-treinados.

A partir da definição do melhor método de seleção e de redução de dimensionalidade dos dados, fez-se a predição dos valores de vib e Ksat para todo o grid da área da bacia de Macacu e, por final, foram obtidas as superfícies mapeadas para ambos os atributos mediante o critério de separação dos dados (profundidades de 0-20 cm e 20-40 cm). Todas as análises foram empregadas por meio de rotinas estatísticas no ambiente computacional RStudio.

Os produtos cartográficos finais são quatro mapas de estimativa da vib e da Ksat a partir da classificação *Random Forest*, contendo informações desses atributos para duas profundidades (superficial: 0-20 cm e subsuperficial: 20-40 cm) dos solos da bacia. O resumo metodológico pode ser visto na Figura 32 a seguir.

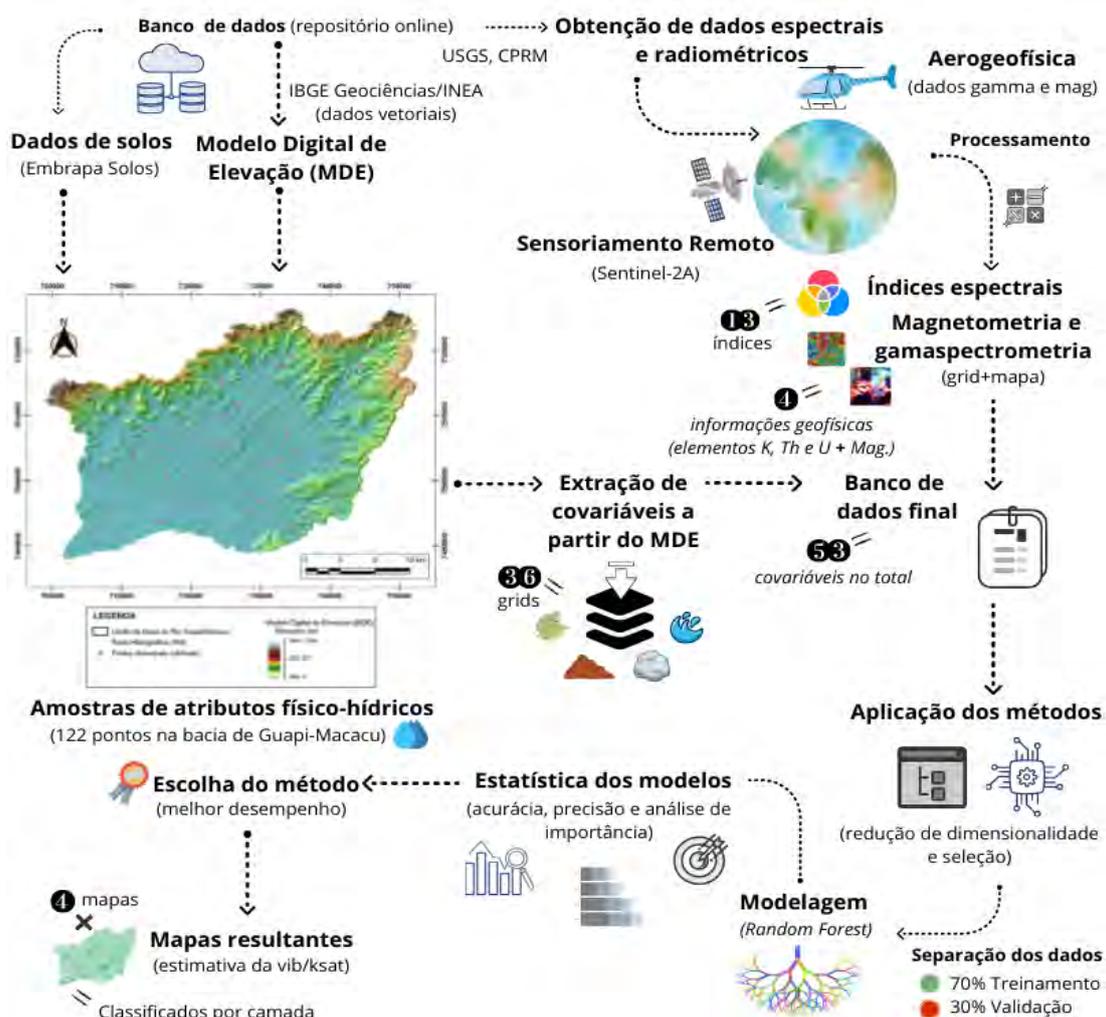


Figura 32. Fluxograma da metodologia proposta.

2.6. Softwares e Pacotes

Como forma de reduzir e processar os dados da pesquisa, foram utilizados os seguintes softwares listados a seguir: *Microsoft Office Excel* v.365 licenciado, para a redução, tabulação e compilação dos dados de perfis de solos contendo as análises físico-químicas e informações de classificação segundo a Sistema Brasileiro de Classificação dos Solos (SIBCS); Software livre de ambiente de desenvolvimento integrado para linguagem R, RStudio v.1.3.959 e R v.4.0.1 (R CORE TEAM, 2020), para a implementação de rotinas dos algoritmos de predição *Random Forest* aplicados na modelagem dos dados geoespaciais, tratamento/processamento dos dados em geral e análise estatística dos produtos obtidos na pesquisa; Software licenciado ESRI ArcGIS v.10.6, para a geração do Modelo Digital de Elevação Hidrológicamente Consistido na área de estudo; Software livre de Sistema para Análises Geocientíficas Automatizadas, SAGA GIS v. 7.9.0, para a geração das covariáveis primárias e secundárias do terreno; Software licenciado Geosoft Oasis Montaj v.9.0 da empresa Geosoft, para tratamento e processamento dos dados aerogeofísicos; Software livre Quantum GIS (QGIS) v. 3.14.0-Pi, da empresa QGIS *Development Team*, para a geração de layouts dos mapas produzidos.

Os principais pacotes listados a seguir compuseram o processamento dos dados da pesquisa e desenvolvimento dos métodos de seleção, sendo de suma importância na etapa de implementação de rotinas de análises da pesquisa no software RStudio: *Mass*, *Raster*, *sp*, *sf*, *shapefiles* e *openxlsx* para leitura e manipulação de dados geoespaciais em formatos tabular, matricial (*raster*) e vetorial; *Sen2R*, para o tratamento e processamento das imagens Sentinel-2A e geração dos índices de vegetação, solo e geologia; *Caret*, *corrplot*, *usdm*, *FactoMineR* e *randomForest* para implementação dos quatro métodos analisados e para classificação supervisionada; *Ggplot2*, para plotagem gráfica.

3. RESULTADOS E DISCUSSÃO

3.1. Composição das informações aerogeofísicas

Na Figura 33 e Figura 34 a seguir pode ser visto o resultado do processamento dos dados aerogeofísicos, no qual foi obtido dois mapas RGB para o Estado do Rio de Janeiro: mapa de aerogramaspectrometria e mapa de aeromagnetometria.

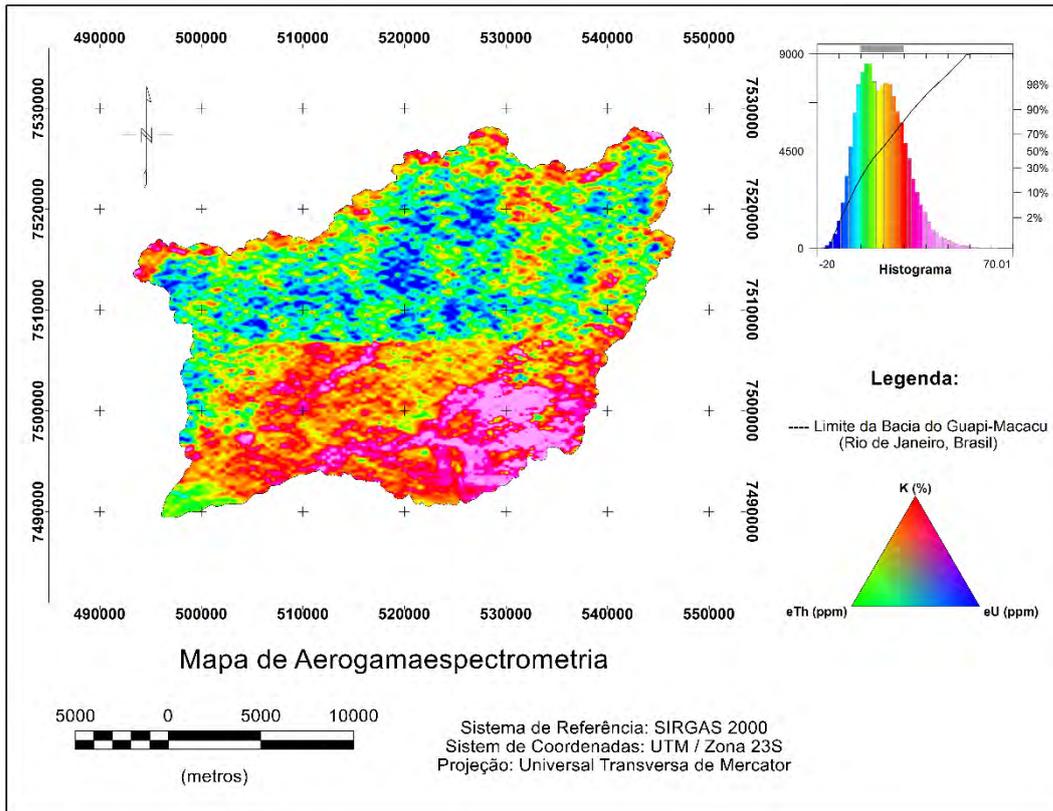


Figura 33 . Mapa aerogeofísico de gamaespectrometria dos elementos K:Th:U (R- potássio, G- tório, B - Urânio) em composição ternária RGB.

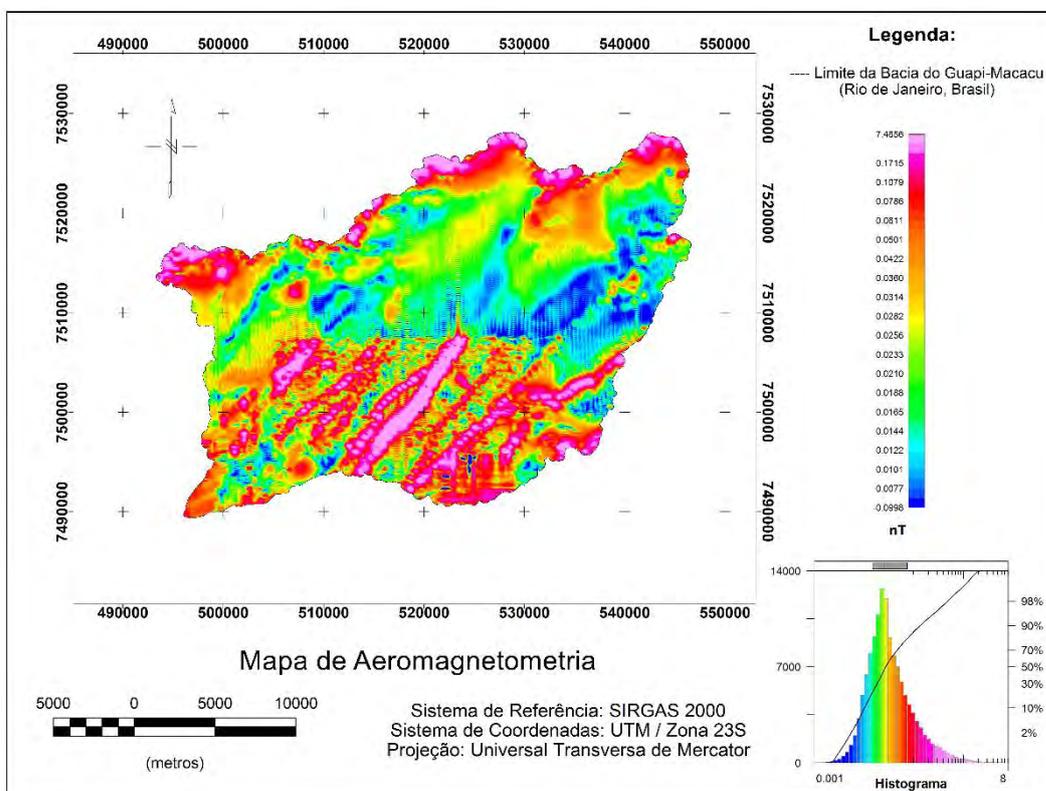


Figura 34. Mapa aerogeofísico de magnetometria do sinal analítico em composição RGB.

O grid dos mapas (Figura 33 e Figura 34) foram inseridos no script de processamento dos dados da pesquisa, com suas bandas RGB separadas para facilitar a análise e recortados para a área de estudo. Para fins de padronização da resolução espacial, os dados geofísicos foram reamostrados para 20 m.

Observou-se, inicialmente, que os dados magnetométricos tendem a perder a qualidade de suas informações devido a transformação do grid para outros formatos (principalmente imagem), devido a características do próprio dado, que são: (a) o dado possui uma tabela de atributos com muitas linhas e colunas, inviabilizando a abertura e espacialização em um SIG “comum”, ou seja, que não seja específico para a tratamento de informações geofísicas. Logo, os dados do grid foram exportados primariamente em formato *float* para que, no software SIG (ArcGIS), pudesse ser devidamente exportado para o formato “.tif” sem perdas de informação; (b) por apresentarem informações da resposta espectral de magnetização das rochas, os dados magnetométricos originais tendem a perder sua qualidade ao serem exportados na forma de arquivo geoespacial “.tiff” através do software Oasis Montaj, devido à perda das componentes direcionais de inclinação e declinação, restando apenas a componente intensidade. Como forma de resolver parte do problema citado em (b), os dados exportados para o processamento são resultantes da aplicação de filtros e derivadas ressaltando a resposta magnetométrica sob a ótica do analista (visualização de corpos mais superficiais ou subpuperficiais sob o eixo z de profundidade) e, ainda, precedeu-se com a exportação do dado em formato *float* para, no SIG (ArcGIS), ser convertido para formato “.tiff”. Tal problemática ocorre ainda com os dados aerogamaespectrométricos, porém sem a necessidade de filtragem já que a análise (potencialidade) gama é dada majoritariamente sob a parte superficial do terreno e os dados exportados apresentam as componentes RGB relativas diretamente aos elementos analisados (potássio, tório e urânio, respectivamente).

Devido à geologia da região, a porção centro-noroeste da bacia destaca-se pela presença de rochas mais profundas e indicativo de maior concentração hídrica no mapa de aeromagnetometria (Figura 34), região onde se contra o Complexo Petroquímico do Rio de Janeiro (COMPERJ), a maior densificação urbana da região e as áreas alagadiças de baixada da bacia. Logo, tais regiões obtêm respostas espectrais diferenciadas. Já no mapa de aerogamaespectrometria (Figura 33), a porção centro-noroeste e centro-sudeste da bacia aponta concentrações maiores de potássio (tons magenta/vermelho) e da mistura dos elementos potássio e tório (tons amarelados) na superfície, destacando-se das demais

regiões da bacia. O efeito da correção *dummy* sob a área restrita do COMPERJ, levantada pelo voo executado pela Companhia de Pesquisa de Recursos Minerais (CPRM), afetou diretamente a interpolação dos dados, evidenciando um efeito de ruído na porção central da bacia (tênue lineamento na direção leste-oeste da bacia). Porém, tal procedimento foi importante para viabilizar a modelagem devido a necessidade de uma estrutura matricial completa para a predição dos valores via *Random Forest*.

3.2. Análises de Autocorrelação e Componentes Principais

A autocorrelação é um método capaz de inspecionar preliminarmente, o comportamento dos dados através de uma análise criteriosa do cientista (e conhecimento prévio de seu banco de dados) sob a matriz de correlação de Pearson, onde os valores do coeficiente de determinação (R) que alcançam valores acima de 0,6 são inferidos em INMAN et al. (1994) como indicativos de correlação moderada à forte (inversa ou direta); valores inferiores a este, indicam correlação fraca (inversa ou direta) e, valores nulos (zero), evidenciam ausência de correlação. Observando a Figura 35, Figura 36, Figura 37 e Figura 38, temos a matriz de correlação de Pearson para os dados estudados.

A matriz de correlação infere que os índices espectrais possuem correlação de moderada com os atributos Ksat e vib, sendo esta relação ora positiva (*Clay Minerals*, EVI, *Ferruginous Regolith*, NDRE, NDVI, NLI, SAVI, VARI e TDVI), ora negativa (*Ferrous Minerals*, GSI e *Iron Oxide*). Já com relação as covariáveis do terreno, a morfometria e as variáveis primárias do MDE (*Slope*, *Aspect* e *Elevação*) apresentaram correlação positiva de fraca à moderada ($0,4 < R < 0,6$). De forma análoga, as covariáveis de canais de hidrologia (como por exemplo, MRRTF e MRVBF) apresentaram correlação de fraca à moderada, porém de forma inversa. Ou seja, à medida que os atributos físico-hídricos aumentam seu valor, essas variáveis têm seu valor diminuído. Esta análise é compatível com as características do relevo local, indicando que há uma drenagem efetiva da água espacialmente por escoamento superficial e/ou vertical ao longo da bacia, que deve ser mais bem averiguado no mapa de predição. As covariáveis de visibilidade, iluminação e hidrologia apresentaram correlação de fraca à moderada ($0,2 < R < 0,4$) em relação ao conjunto total de dados.

Nas matrizes (Figura 35, Figura 36, Figura 37 e Figura 38), os dados espectrais interrelacionam-se de forma a aparentar uma possível multicolinearidade ou até dimensionalidade (vários dados representando características iguais ou muito semelhantes). De modo a obter tais informações de contribuição dos dados espectrais,

fez-se a análise de componentes principais. A análise retornou uma explicação acima de 90% para duas componentes principais, podendo ser explicado o modelo através da análise bidimensional das variáveis espectrais (Figura 39a).

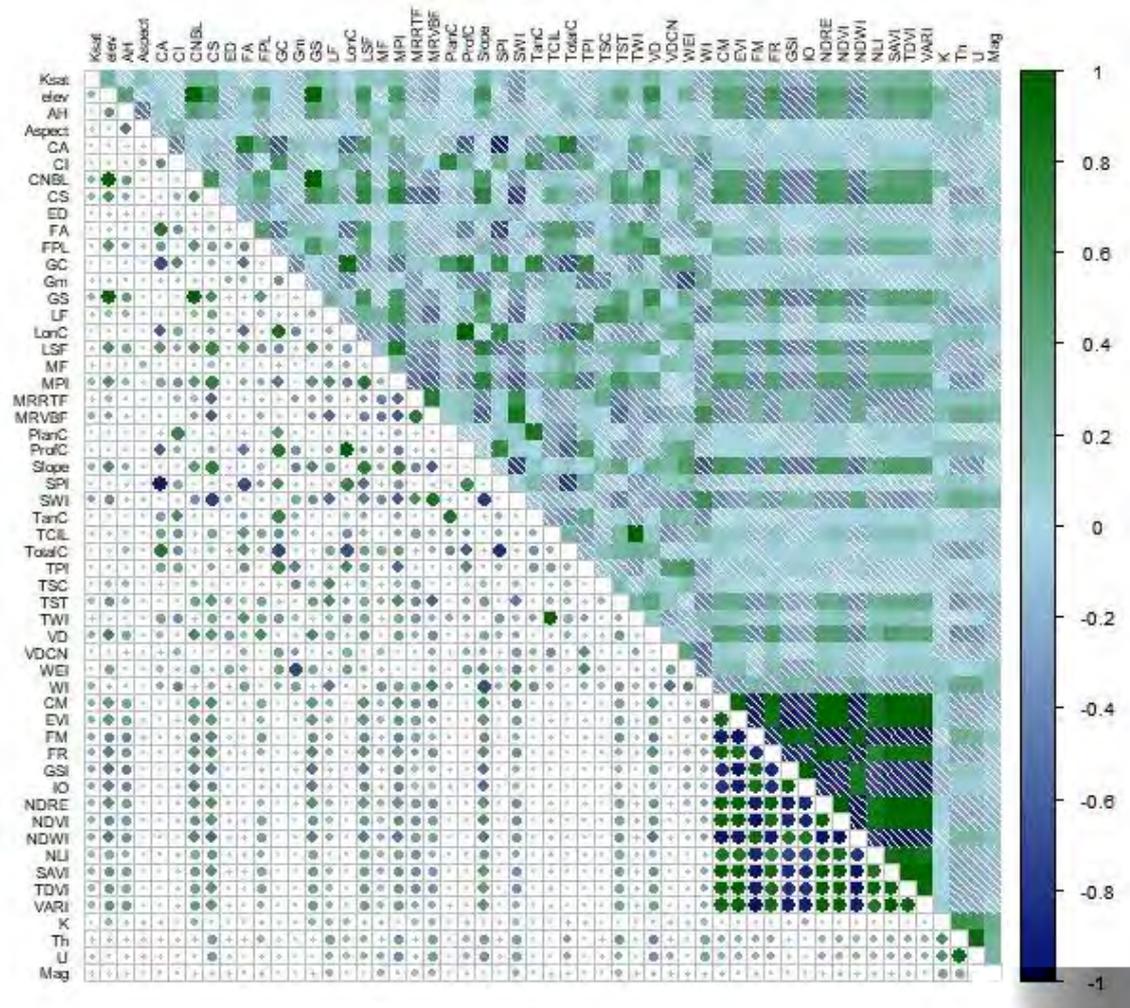


Figura 35. Matriz de correlação de Pearson envolvendo o conjunto de covariáveis estudo e a variável Ksat para a camada de 0-20cm.

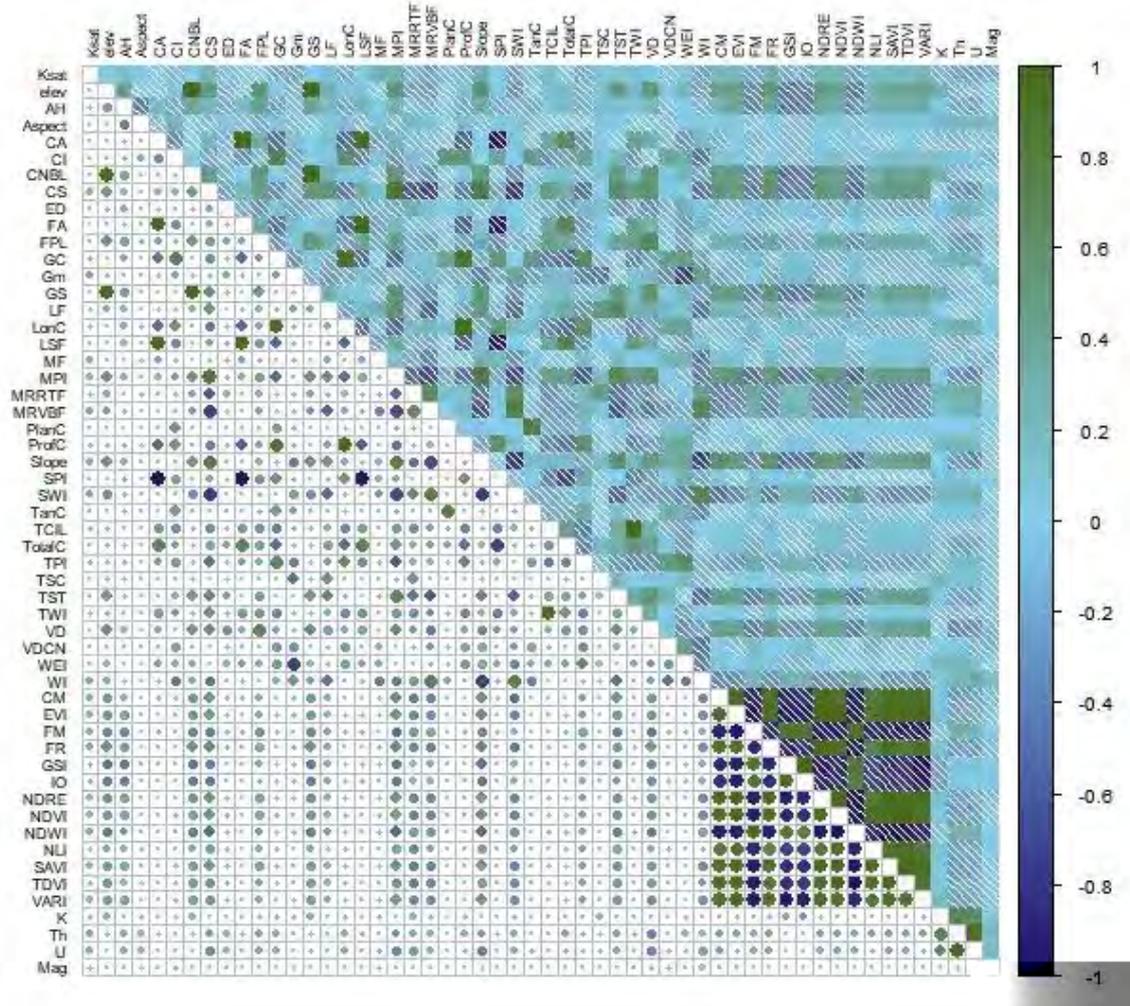


Figura 36. Matriz de correlação de Pearson envolvendo o conjunto de covariáveis estudo e a variável Ksat para a camada de 20-40cm.

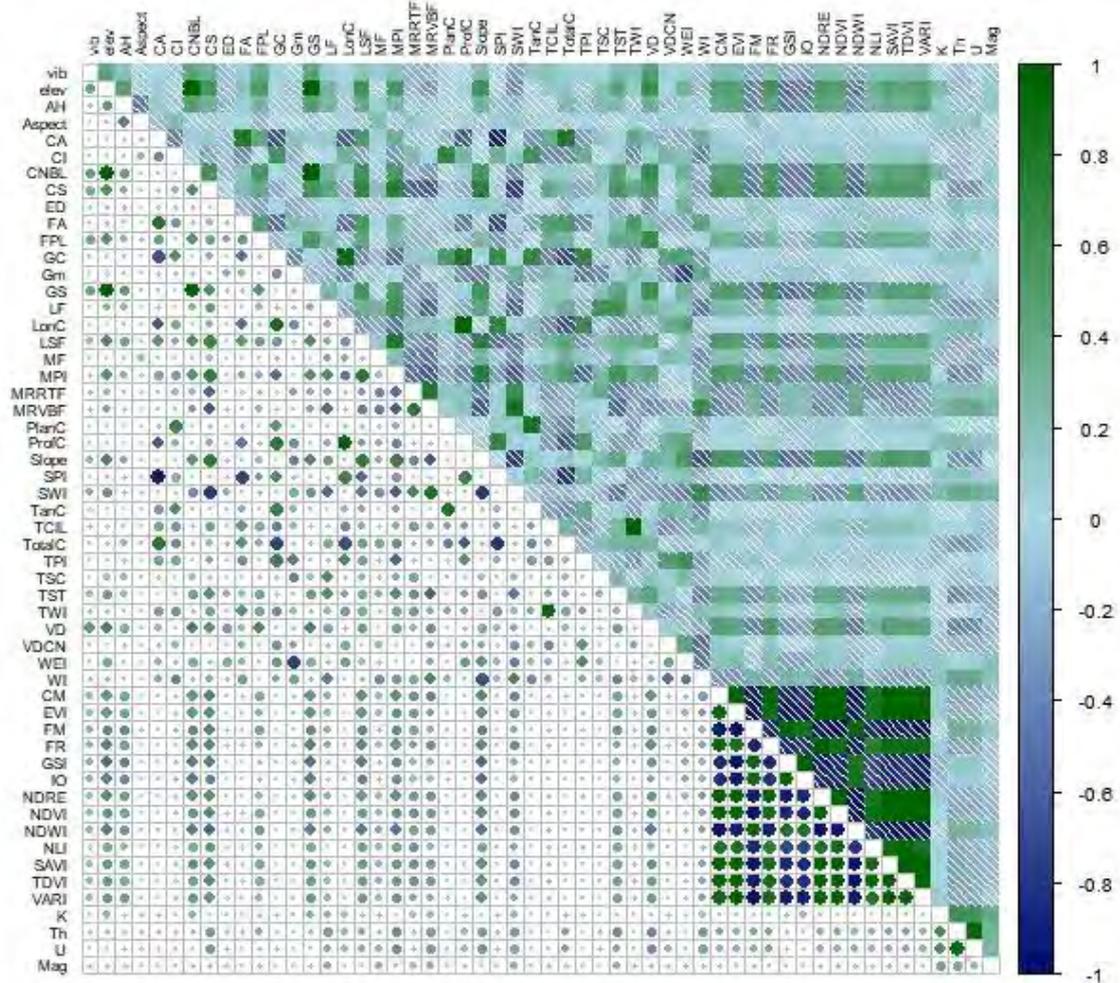


Figura 37. Matriz de correlação de Pearson envolvendo o conjunto de covariáveis estudo e a variável vib para a camada de 0-20cm.

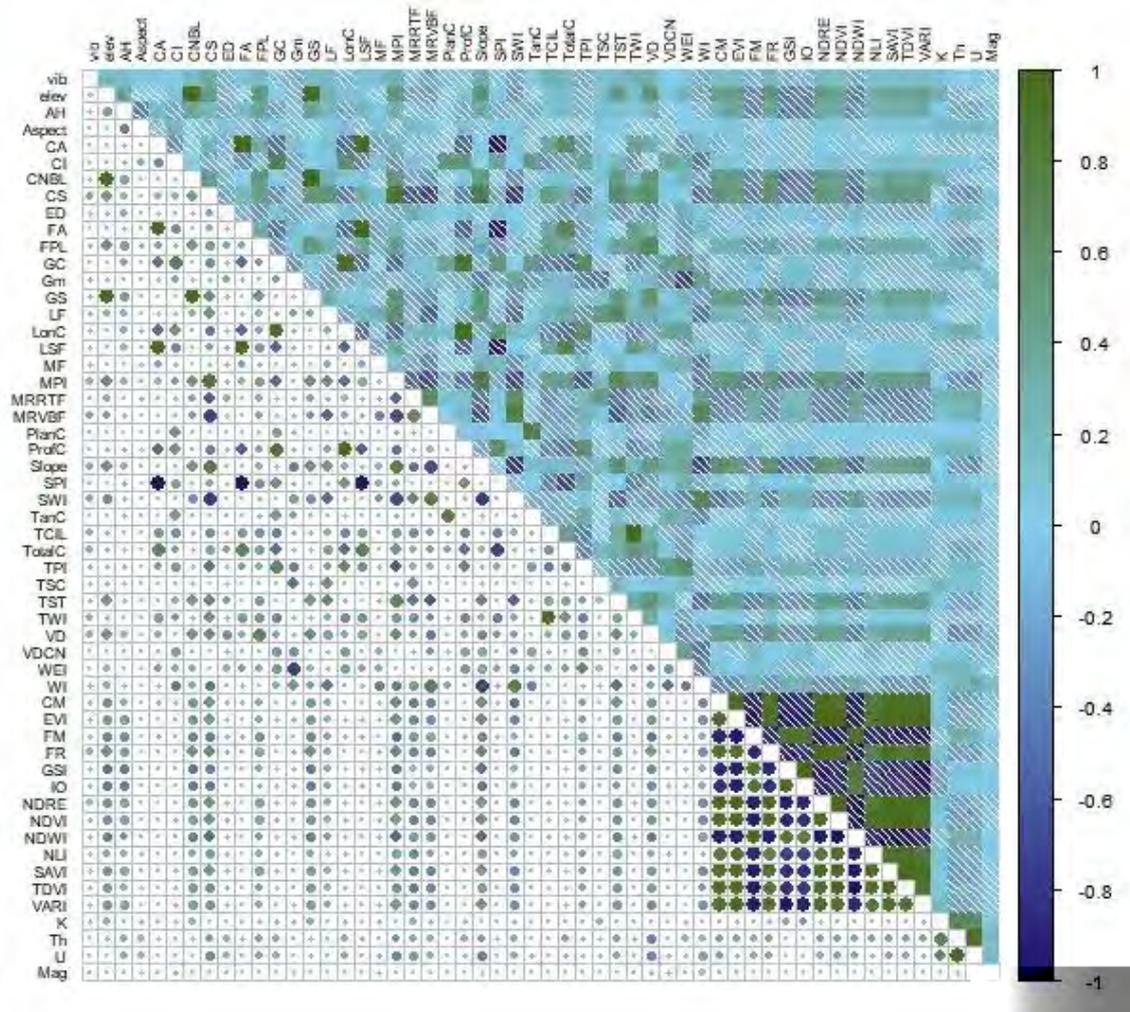


Figura 38. Matriz de correlação de Pearson envolvendo o conjunto de covariáveis estudo e a variável vib para a camada de 20-40cm.

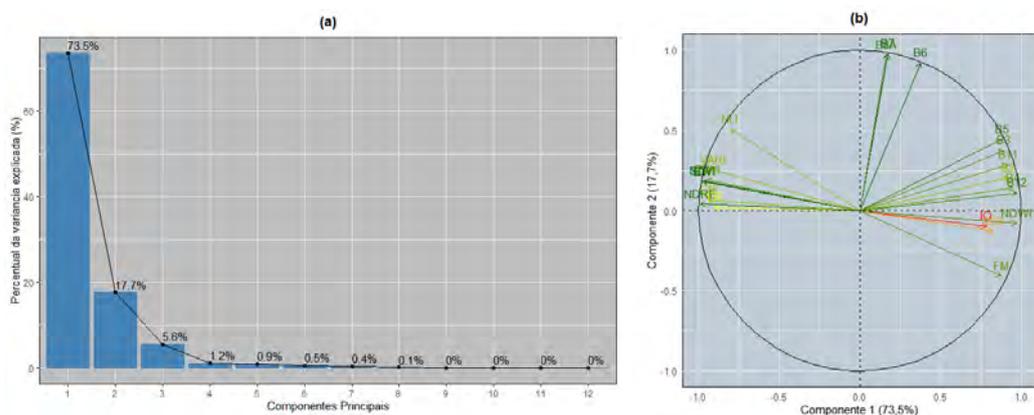


Figura 39. Variáveis espectrais representadas segundo as componentes principais explicativas do modelo ACP, onde: (a) importância dos componentes principais na ACP; (b) separação das componentes principais na ACP.

Observando-se as componentes espacialmente, tem-se que, dada a distribuição das variáveis (Figura 39b), há a separação do conjunto espectral em 3 grupos diferentes. Pode-se inferir ainda que os índices *Iron Oxide* (IO) e *Grain Size Index* (GSI) são os que percentualmente menos contribuem para o grupo 1 na explicação das variáveis físico-hídricas. Conforme a separação dos grupos mostradas na Figura 39b, foram selecionadas 10 variáveis compondo o grupo 1, 3 compondo o grupo 2 e 10 compondo o grupo 3, sendo duas variáveis no grupo 1 (índices IO e GSI) passíveis de análise mais criteriosa na etapa de modelagem. Com esses resultados, é possível inferir que os dados espectrais oriundos de sensoriamento remoto são importantes para a modelagem, pois explicam boa parte da sua variabilidade e, conforme a distribuição da Figura 39, o conjunto de dados pode ter sua dimensionalidade reduzida para evitar os problemas associados influência da multicolinearidade dessas variáveis na modelagem.

As análises de correlação e dimensionalidade são importantes para estudos de caráter multivariável, entretanto, atuam como uma etapa prévia à aplicação dos métodos, visando apenas conhecer os dados e identificar padrões na área de estudo, pois carecem de informações tais como a homocedasticidade (ausência de tendência nas variâncias dos erros), multicolinearidade e normalidade; o que pode afetar negativamente conclusões diretas sob os resultados. Tais critérios estatísticos podem ser mais bem avaliados por aplicação de testes de *Breusch-Pagan*, *Shapiro-Wilk* e *Dublin-Watson* sob o banco de dados de forma a adequá-los aos requisitos estatísticos exigidos pelos métodos adotados na pesquisa.

Segundo Mansfield et al. (1982), a multicolinearidade pode ter vários efeitos adversos nos coeficientes estimados em uma análise de regressão múltipla; conseqüentemente, é importante que os pesquisadores sejam treinados para detectar sua existência. Sendo assim, foram analisados três métodos distintos (VIF, RFE e *Stepwise AIC*) de forma a escolher as variáveis adequadas para cada atributo estudado, em suas respectivas profundidades determinadas (superficial e subsuperficial), e verificada a qualidade da estimativa desses atributos por meio da modelagem usando o classificador de aprendizado de máquinas *Random Forest*.

3.3. Modelagem Espacial

Com base na separação do banco de dados por profundidade (0-20cm e 20-40cm) e variáveis estudadas (vib e Ksat), a seleção das variáveis resposta deu-se através da calibração de três métodos distintos e a análise de redução de dimensionalidade dos dados

espectrais. Para cada método aplicado, observam-se diferentes respostas na definição das variáveis de saída (explicativas) (Tabela 9) as quais mantêm uma relação significativa (nível de significância $\alpha=5\%$), direta ou inversa, com as variáveis físico-hídricas.

Tabela 9. Covariáveis selecionadas para modelagem da variável Ksat após aplicação dos métodos propostos.

Método	Covariáveis selecionadas	Total
<i>Ksat_{0-20 cm}</i>		
VIF	<i>AH; Aspect; CI; CS; ED; FA; FPL; Gm; GS; LF; LSF; MF; MRRTF; PlanC; TanC; TotalC; TPI; TSC; TST; TWI; VD; VDCN; WEI; WI; FERRUGINOUSREGOLITH; IRONOXIDE; NLI; K; U; Mag</i>	30
RFE	<i>CS; GS; CNBL; NDWI; elev; LSF; FERRUGINOUSREGOLITH; SWI; CLAYMINERALS; TSC; SAVI</i>	11
Stepwise AIC	<i>elev; AH; Aspect; CA; CNBL; FPL; Gm; GS; LF; LonC; MF; MPI; MRVBF; ProfC; SWI; TanC; TCIL; TPI; TST; TWI; VD; WI; EVI; FERROUSMINERALS; GSI; IRONOXIDE; NDRE; NDVI; NDWI; SAVI; TDVI; VARI; K; Th; U; Mag</i>	36
<i>Ksat_{20-40 cm}</i>		
VIF	<i>MRRTF; MRVBF; PlanC; ProfC; Slope; TanC; TCIL; TotalC; TSC; VD; VDCN; WEI; IRONOXIDE; NLI; K; U; Mag</i>	17
RFE	<i>Mag; FERROUSMINERALS; VARI; MRRTF; K; LSF; MPI; elev; SWI; GSI; AH; IRONOXIDE; CNBL; GS; VD; EVI; NDVI; Slope; CLAYMINERALS; MF; GC; NLI; TDVI; SAVI; Aspect; WEI; NDRE</i>	27
Stepwise AIC	<i>elev; CA; FA; FPL; GC; GS; LonC; LSF; MF; MPI; ProfC; Slope; SPI; TCIL; TotalC; TPI; TSC; TST; TWI; VD; FERROUSMINERALS; FERRUGINOUSREGOLITH; GSI; IRONOXIDE; NDWI; NLI; SAVI; TDVI; VARI; Th; U</i>	31
<i>vib_{0-20 cm}</i>		
VIF	<i>AH; Aspect; CI; CS; ED; FA; FPL; Gm; GS; LF; LSF; MF; MRRTF; PlanC; TanC; TotalC; TPI; TSC; TST; TWI; VD; VDCN; WEI; WI; FERRUGINOUSREGOLITH; IRONOXIDE; NLI; K; U; Mag</i>	30
RFE	<i>CNBL; GS; elev; VD; NDWI; CS; NLI; CLAYMINERALS; TDVI; TanC; NDRE; FERRUGINOUSREGOLITH; SAVI; NDVI</i>	14
Stepwise AIC	<i>elev; CS; FA; FPL; GC; Gm; GS; LF; LSF; MF; MPI; MRRTF; MRVBF; ProfC; SWI; TanC; TotalC; TPI; TST; TWI; VD; WEI; WI; EVI; FERRUGINOUSREGOLITH; GSI; IRONOXIDE; NDRE; NDVI; NDWI; NLI; SAVI; VARI; K; Th; U</i>	36
<i>vib_{20-40 cm}</i>		
VIF	<i>MRRTF; MRVBF; PlanC; ProfC; Slope; TanC; TCIL; TotalC; TSC; VD; VDCN; WEI; IRONOXIDE; NLI; K; U; Mag</i>	17
RFE	<i>LSF; Mag; CS; VD; GS; CNBL; elev; MRRTF; TotalC; SWI; TCIL; U; TWI; GSI; Slope; WEI; MRVBF; AH; CLAYMINERALS; NLI; VARI; Th; NDVI; K; NDRE; TDVI; EVI; FPL; SAVI; SPI; TPI; LF; IRONOXIDE; MPI; Aspect; PlanC; FERROUSMINERALS</i>	37
Stepwise AIC	<i>elev; AH; CA; CI; FA; FPL; GS; LonC; LSF; MRRTF; MRVBF; PlanC; ProfC; Slope; TanC; TCIL; TotalC; TPI; TST; TWI; GSI; IRONOXIDE; NDVI; NDWI; TDVI; VARI</i>	26

VIF= Variance Inflation Factor; RFE= Recursive Feature Elimination; AIC= Akaike Information Criterion. *Siglas cujos significados estão apresentados no Quadro 6.

Analisando a Tabela 9, a maioria dos métodos resultaram em mais de 10 variáveis explicativas selecionadas, com ausência de multicolinearidade e interferências de correlações fortes nos dados. Na estimativa da Ksat e da vib, ambas para a profundidade entre 0-20 cm, o método RFE de seleção tiveram como resposta a menor quantidade de variáveis compondo o modelo RF, assumindo-se 11 e 14 variáveis no total, respectivamente. Observa-se, ainda, que nos quatro critérios adotados, os métodos selecionaram diversificadamente suas covariáveis, agregando tanto aspectos topográficos quanto radiométricos da área de estudo. Com base no método RFE, o alcance da quantidade ideal de variáveis que garantisse um menor resíduo no modelo *Random Forest* (menor RMSE) pode ser visto nas figuras de a-d (Figura 40).

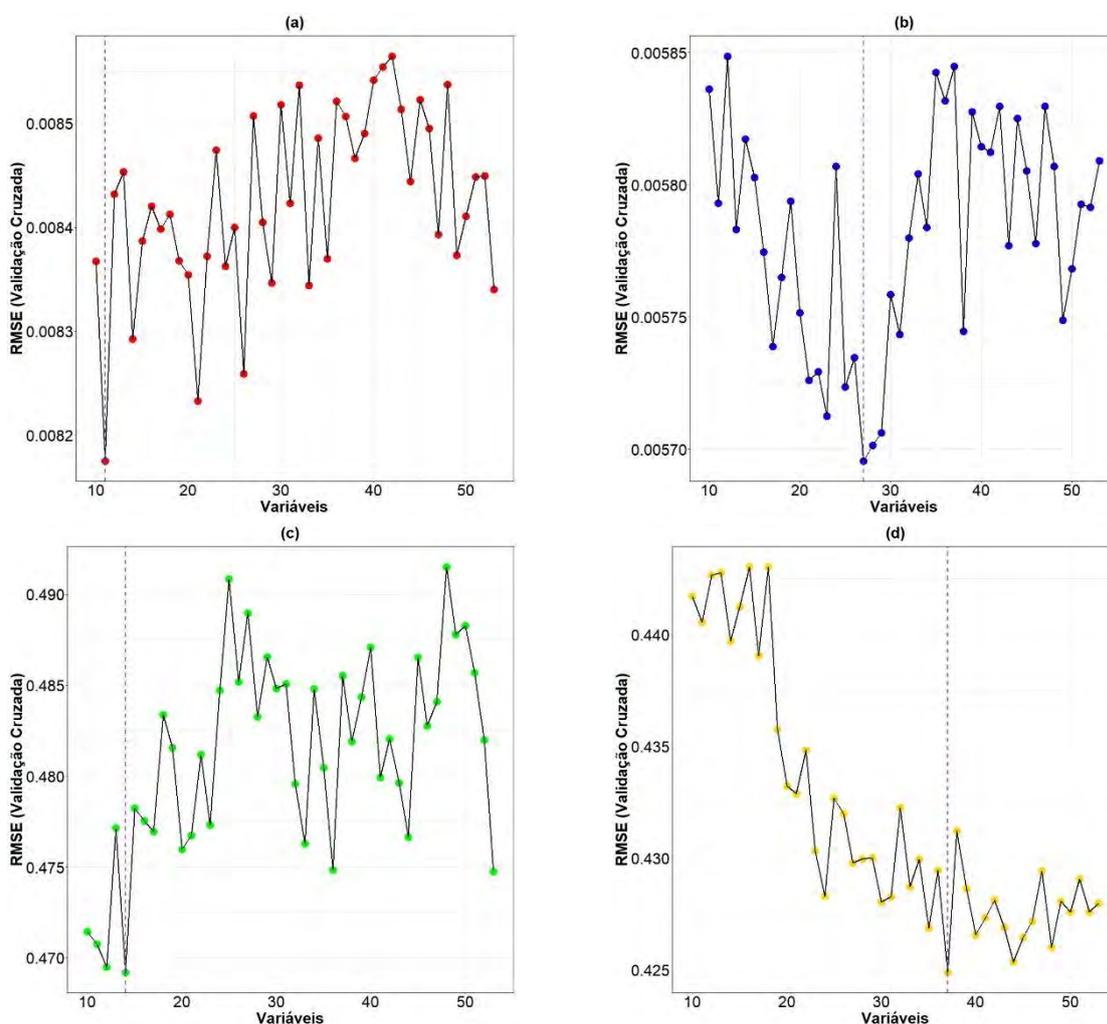


Figura 40. Gráficos do método RFE (quantidade de variáveis explicativas versus RMSE do ajuste) para os atributos físico-hídricos estudados e seus respectivos intervalos de profundidade: **(a)** Ksat₀₋₂₀ cm; **(b)** Ksat₂₀₋₄₀ cm; **(c)** vib₀₋₂₀ cm; **(d)** vib₂₀₋₄₀ cm. A linha tracejada indica o alcance da quantidade ótima de variáveis selecionadas pelo RFE.

A linha tracejada nos gráficos da Figura 40 (a, b, c e d) denota o alcance da quantidade ideal de variáveis ajustadas pelo modelo *Random Forest* para obter o menor valor de RMSE, critério este adotado para as análises, usando o método RFE. Verifica-se que apenas a modelagem aplicada sob a variável vib_{0-20cm} apresenta um alcance de um menor RMSE para uma menor quantidade de variáveis (11) após 99 iterações executadas pelo software (Figura 40a). Nos três outros critérios ($Ksat_{0-20cm}$, $Ksat_{20-40cm}$ e $vib_{20-40cm}$) (Figura 40 b-d), foram necessárias 27, 14 e 37 variáveis, respectivamente, para a obtenção do ajuste ideal do modelo RF. O método RFE para $vib_{20-40cm}$ foi o que exigiu a maior quantidade de variáveis na composição de seu conjunto explicativo em relação geral aos outros dois métodos implementados, com 1 variável explicativa a mais que o método *Stepwise AIC* avaliado para $Ksat_{0-20cm}$ e vib_{0-20cm} . Vale salientar que o modelo *Random Forest* otimizou o número de feições randômicas avaliadas (*mtry*) pelo uso da função *tuneRF*.

Os testes de *Breusch-Pagan*, *Dublin-Watson* e *Shapiro-Wilk* foram aplicados sob o banco de dados, retornando valores significativos para a aceitação da hipótese nula (H_0) frente à hipótese alternativa (H_1), a um nível de significância de 5% ($\alpha=5\%$, logo $\beta=95\%$ de nível de confiança). O teste de *Shapiro-Wilk* visou verificar a normalidade dos dados (H_0 : Há normalidade nos resíduos versus H_1 : Não há normalidade nos resíduos). Já o teste de *Breusch-Pagan* visou verificar a homoscedasticidade das variâncias dos resíduos (H_0 : variâncias são iguais - homoscedasticidade versus H_1 : variâncias são diferentes - heterocedasticidade). Por fim, o teste de *Durbin-Watson* buscou avaliar a existência ou não de correlação entre os resíduos, no qual é um indicativo de multicolinearidade (H_0 : a autocorrelação entre os resíduos é igual a zero - independências dos resíduos versus H_1 : a autocorrelação entre os resíduos é diferente de zero - dependência dos resíduos). Dado que os modelos *vif* e *Stepwise* partem de princípios de análise regressiva, a etapa de teste dos métodos realizados é essencial para garantir que sejam cumpridos os pressupostos estatísticos. Para todos os métodos, os valores p (*p-value*) dos testes estiveram dentro do intervalo exigido ($p\text{-value} < \alpha = 5\%$), sem que haja a necessidade de correções e tratamentos nos dados. A qualidade dos métodos treinados na porção de 70% do banco de dados, previamente separado, usando o modelo *Random Forest* é verificada na Tabela 10.

Tabela 10. Qualidade do modelo *Random Forest* no treino dos dados físico-hídricos a partir dos métodos de seleção aplicados (70% do banco de dados).

Variável estudada	Método aplicado	Métrica de qualidade avaliada para o <i>Random Forest</i>		
		RMSE _{treino}	MAE _{treino}	R ² _{treino}
Ksat Z= 0-20 cm	VIF	0,0054	0,0034	0,8755
	RFE	0,0049	0,0032	0,8731
	<i>Stepwise AIC</i>	0,0050	0,0035	0,8948
Ksat Z= 20-40 cm	VIF	0,0025	0,0019	0,8679
	RFE	0,0025	0,0019	0,8657
	<i>Stepwise AIC</i>	0,0025	0,0020	0,8663
Vib Z = 0-20 cm	VIF	0,2743	0,1834	0,9336
	RFE	0,3038	0,1980	0,8130
	<i>Stepwise AIC</i>	0,2733	0,1879	0,9004
Vib Z= 20-40cm	VIF	0,2713	0,1867	0,7276
	RFE	0,2710	0,1871	0,7306
	<i>Stepwise AIC</i>	0,2767	0,1962	0,7220

Valores baseados na validação cruzada do modelo *Random Forest* para os dados separados do treino.

Observa-se que todos os modelos atingiram acurácia superior a 70% ($R^2 > 0,70$) na predição dos dois atributos alvos deste estudo (Ksat e vib) para as profundidades de 0-20 cm e 20-40cm (Tabela 10), avaliados segundo o critério de validação cruzada. Isto indica bons ajustes dos modelos na calibração do método RF. Verifica-se que o ajuste do modelo *Random Forest*, para cada um dos três métodos propostos, obteve estabilização no erro (RMSE) entre, aproximadamente, 300-350 árvores aleatórias desenvolvidas para as variáveis físico-hídricas estudadas (Figura 41). Os modelos *Random Forest* que obtiveram melhor acurácia na predição das variáveis de estudo, segundo o método RFE, são indicados com coloração em destaque na Figura 41 a seguir, com estabilização do erro RMSE a partir de 300 árvores.

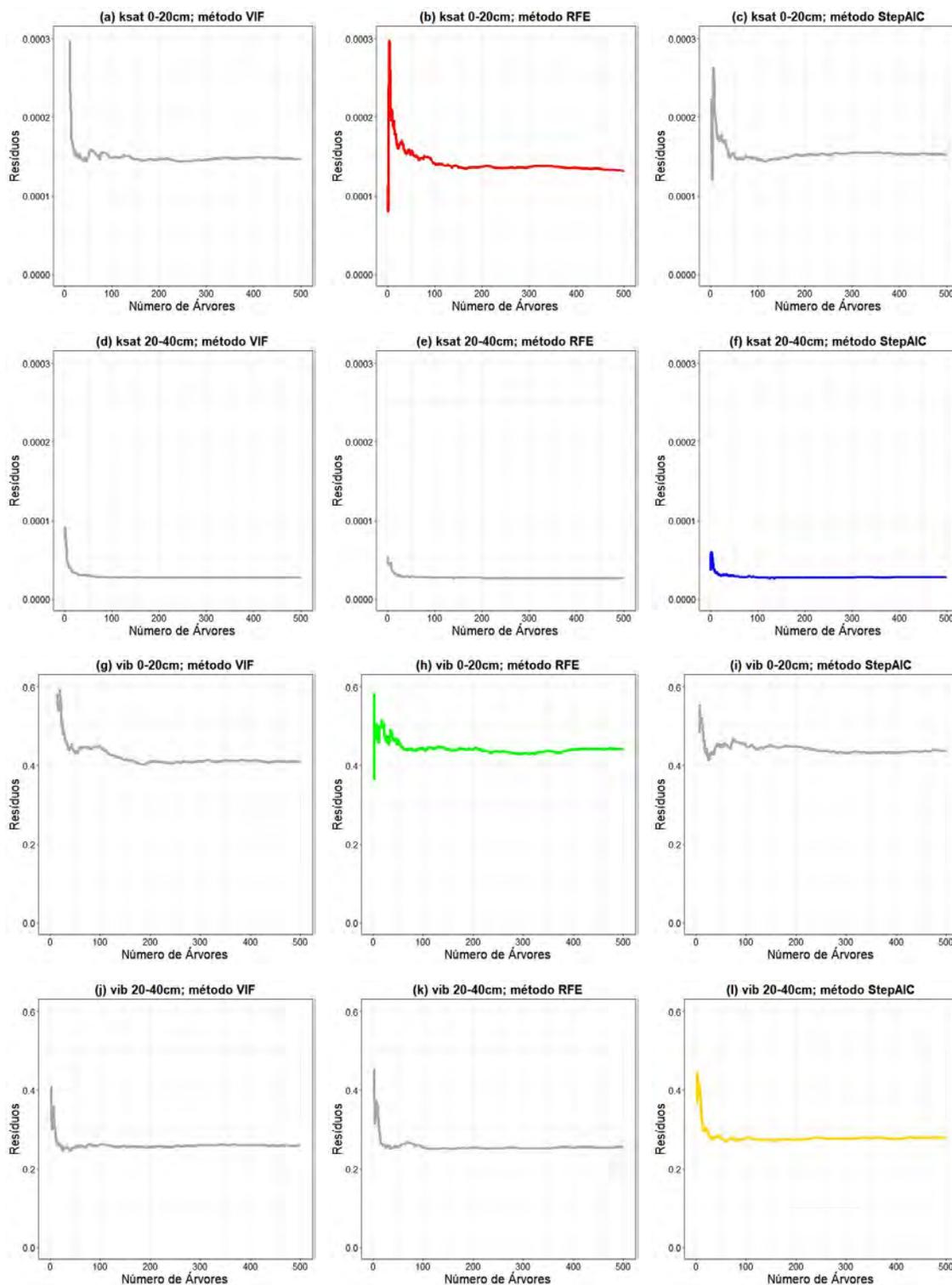


Figura 41. Gráficos de ajuste do modelo *Random Forest*, evidenciando a quantidade de árvores versus a minimização do erro para as variáveis Ksat e vib, segundo as camadas avaliadas: (a-c) Ksat_{0-20cm}; (d-f) Ksat_{20-40cm}; (g-i) vib_{0-20cm}; (j-l) vib_{20-40cm}; e os três métodos analisados: (a,d,g,j) VIF; (b,e,h,k) RFE; (c,f,i,l) *Stepwise AIC*.

Diferente do observado no treino, a resposta do modelo *Random Forest* para a validação ao serem avaliados os modelos treinados em relação aos valores de teste (30%

do banco de dados não usados para treinamento) foi de diferença acentuada entre os valores preditos (estimados pelo modelo) e os valores reais (mensurados in situ) evidenciada pelo R^2_{teste} inferior a 20% para os métodos implementados sob os critérios de análise (profundidades e variáveis independentes), excetuando-se a variável vib analisada para a camada 0-20cm, na qual obteve R^2 superior (entre 28-48%, aproximadamente), conforme visto na Tabela 11.

Tabela 11. Qualidade do Modelo *Random Forest* na Validação dos dados físico-hídricos de teste (30% do banco de dados).

Variável estudada	Método aplicado	Métrica de qualidade avaliada para o <i>Random Forest</i>		
		RMSE _{teste}	MAE _{teste}	R ² _{teste}
Ksat Z= 0-20 cm	VIF	0,0083	0,0071	0,0525
	RFE	0,0065	0,0056	0,1433
	Stepwise AIC	0,0084	0,0076	0,0461
Ksat Z= 20-40 cm	VIF	0,0077	0,0045	0,0199
	RFE	0,0074	0,0044	0,0375
	Stepwise AIC	0,0069	0,0042	0,1088
Vib Z = 0-20 cm	VIF	0,4586	0,3594	0,4057
	RFE	0,4171	0,3394	0,4735
	Stepwise AIC	0,4799	0,3910	0,2801
Vib Z= 20-40cm	VIF	0,3704	0,2771	0,1439
	RFE	0,3617	0,2748	0,1629
	Stepwise AIC	0,3531	0,2747	0,2003

Valores baseados na validação do modelo *Random Forest* para os dados separados no teste.

A qualidade dos modelos *Random Forest* verificada após a validação interna (Tabela 11) resultou valores baixos no geral para as métricas aplicadas (RMSE, MAE e R^2), inferindo que os modelos ajustados possuem menos de 20% de explicação da variabilidade dos atributos Ksat e vib na região estudada. A resposta para esta verificação pauta-se na ausência de valores ou da baixa significância (valores muito pequenos, tendendo a zero) dos valores de elevação na região de baixada da bacia, devido a sua curvatura planar desta região (Figura 42) e a sua proximidade com a baía da Guanabara, onde desaguam-se as águas da bacia (fóz), no qual é um local que sofre constantes processos de sedimentação. O resultado apresentado pode ainda ser associado a problemas de sobreajuste nos dados de treino (*overfitting*) ou de caráter randômico na escolha dos valores de validação. Uma solução para este problema é a amplificação na quantidade amostral das variáveis estimadas, levando em consideração critérios como classificação dos solos e mensuração de pontos homogeneamente na área de estudo.

Embora esses parâmetros tenham sido incluídos no planejamento do levantamento, a impossibilidade de coleta dos dados em regiões de intensa antropização e densidade florestal, as quais compõem a paisagem e a formação dos municípios de Cachoeiras de Macacu, Guapi-Açu, Guapimirim e Itaboraí (locais estes onde a bacia exerce influência); podem ser fatores relevantes quando avaliado o fato de que os valores de Ksat e vib foram estimados para 122 pontos segundo modelos de pedotransferência, como evidenciados na pesquisa realizada posteriormente, na qual este trabalho se integra. Portanto, podem ocorrer problemas de natureza randômica nos erros e fontes exteriores ao mapeamento, mesmo tendo-se estimados os 122 pontos com valores de acurácia e precisão relativamente elevados.

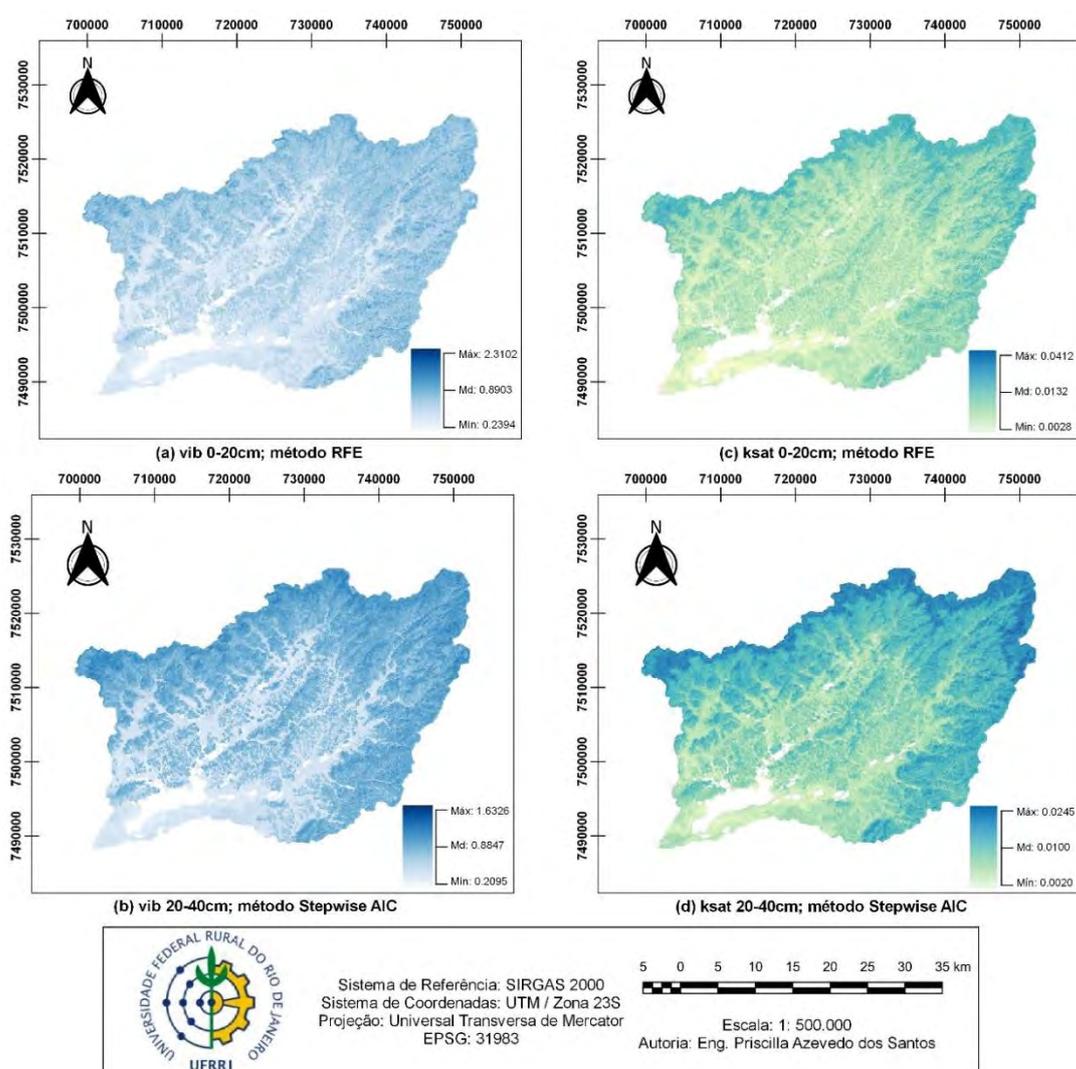


Figura 42. Mapas da variabilidade dos atributos vib e Ksat na bacia hidrográfica do Rio Guapi-Macacu, Rio de Janeiro, estimados através do modelo Random Forest, segundo as respectivas profundidades e métodos de seleção: (a) vib_{0-20cm} - RFE; (b) vib_{20-40cm} - StepwiseAIC; (c) Ksat_{0-20cm} - RFE; (d) Ksat_{20-40cm} - StepwiseAIC.

Se avaliarmos a composição sedimentar, é esperado que a região apresente valores variáveis nos atributos físico-hídricos (Figura 42), apresentando valores mais altos em regiões com maior existência de argilas - principalmente argilas dispersas e de alta atividade (devido a existência de equilíbrio entre micro e macroporos, que favorecem o processo de capilaridade e retenção de água) e matéria orgânica no solo (devido a formação de colóides que aumentam a superfície específica de contato das partículas com a água e, conseqüentemente, favorece a retenção de água); e menores valores em regiões com presença de areias e materiais mais grosseiros e mal estruturados (pois há maior presença de macroporos, reduzindo a capacidade do solo em transmitir a água por capilaridade e reter por tempo suficiente para que a água recarregue o aquífero). A região que apresenta solos minerais (porções de feldspato e quartzo associados), evidenciados pela carta geológica da região (DRM, 2019) (anexo C). na bacia apresenta maior densidade do solo e, conseqüentemente, redução nos processos de infiltração e capilaridade da água.

Os valores evidenciados no mapa são condizentes com a escala intervalar das variáveis vib e K_{sat} (Figura 42). Na transição entre os horizontes (de superficial para subsuperficial), ambas variáveis tiveram seus valores decrescidos, o que nos dá indício de baixa infiltração e escoamento capilar e, conseqüentemente, redução na retenção da água nos horizontes mais subsuperficiais da bacia. A região serrana circundante da bacia (municípios de Petrópolis, Teresópolis e Nova Friburgo), devido à presença de vegetação densa (bioma de Mata Atlântica), apresenta os maiores valores de vib e K_{sat} nos horizontes subsuperficiais, evidenciando essas áreas como prioritárias e potenciais para conservação dos recursos hídricos na bacia. A degradação dos solos e o desmatamento nestas áreas são práticas que podem levar à escassez dos recursos citados.

3.4. O método de seleção ideal para a modelagem: ponderações e sugestões

Vale salientar que a escolha das variáveis selecionadas vai depender diretamente da escolha do método de seleção o qual o operador adote na fase pré-modelagem, podendo ser tanto as variáveis resultantes da interseção dos métodos como as variáveis resultantes de um método apenas. A tomada de decisão quanto ao “melhor” método vai depender do objetivo do estudo e conhecimento do analista. Para este estudo em questão, foram avaliados os métodos segundo as respostas de qualidade dos modelos nas fases de treino e teste via *Random Forest*, pelas métricas estatísticas do RMSE e R^2 . A proposição do

modelo ideal para cada atributo estudado e sua referida profundidade resultou no mapeamento (modelagem espacial) da variabilidade hídrica nos solos da região.

Sobre outro ponto de vista, propõe-se a avaliação individual do comportamento das variáveis e a sua interseção (comparação) nos quatro métodos aplicados para definir quais variáveis são mais importantes de serem inseridas na fase de modelagem para classificação da Ksat na área de estudo. Não obstante, é considerável que as variáveis explicativas analisadas de forma individualizada em suas respectivas naturezas (topográfica, espectral, outras) podem não ser capaz de representar a Ksat ou a vib como um todo, como ocorre na avaliação em grupos, sendo importante então a definição de um método preferencial de seleção.

No geral, o ajuste dos modelos *Random Forest* para os métodos aplicados foi baixo, implicando na utilização de outras variáveis não aplicadas no processo de modelagem (PINHEIRO, 2012) ou, ainda, de implementação de outros algoritmos de aprendizado de máquinas para comparar os ajustes feitos por estes com os resultados do *Random Forest*. É válido comentar que Modelo Digital de Elevação extraído da topografia da bacia possui áreas planas com pouca ou nenhuma informação altimétrica na quantificação de valores de elevação e demais feições, principalmente em sua região de baixada, onde a curvatura planar dificulta a verificação da variabilidade dos atributos no terreno ao retornarem para os mapas valores muito pequenos ou constantes mensurados acerca destes. Conforme sugerido por PINHEIRO et al. (2012), o uso de modelo digital de elevação interpolado com informações de altimetria gerados por imagens satelitais (por exemplo, Aster ou SRTM) podem elevar a qualidade nos dados coletados nesta região e trazer soluções ao baixo ajuste dos modelos.

Aponta-se potencialidades nos dados aerogeofísicos quando analisados junto ao mapa geológico da região e aos mapas preditos, sugerindo áreas prioritárias para conservação e sustentabilidade dos recursos hídricos de forma que os municípios que compõem a área da bacia possam fazer o correto manejo e uso desses bens não-renováveis da natureza. Ainda, a garantia do prolongamento da vida ecossistêmica depende diretamente da água infiltrada, escoada e recarregada, respectivamente, no solo, na superfície terrestre e nos aquíferos da bacia hidrográfica. Sendo assim, estudos e produtos associados a geologia e a geofísica integradas ao mapeamento de solos demonstram potencialidades na avaliação e tomada de decisão dos órgãos públicos e particulares frente ao uso dos recursos hídricos, visando o bem-estar da humanidade, da flora e da fauna.

Sugere-se, ainda, o uso combinado de dados espectrais e covariáveis do terreno (derivadas do MDE) associados à atributos dos solos, como a composição granulométrica (areia, silte e argila), densidade do solo, carbono orgânico e outras, capazes de relacionar-se espacialmente com a Ksat e, assim, tornar o modelo mais robusto e explicativo da variável em estudo.

4. CONCLUSÕES

Os três métodos avaliados apresentaram diferentes respostas de seleção, havendo a redução dos dados de 53 variáveis de entrada iniciais (fase pré-modelagem) para menores quantidades, à saber: **30, 11 e 36** para Ksat com Z=0 -20 cm; **17, 27 e 31** para a Ksat com Z=20-40 cm; **30,14 e 36** para a vib com Z=0-20cm; **17, 37 e 26** para a vib com Z=20-40 cm, respectivamente. Dentre todas as variáveis do conjunto de dados analisado, o índice Iron Oxide (IO) foi majoritariamente escolhido pelos métodos aplicados, exceto na aplicação do RFE sob os atributos vib e Ksat nas camadas superficiais dos solos (0-20cm), no qual não retornou tal variável.

Os métodos vif e RFE foram os que apresentaram menor quantidade de variáveis explicativas para as camadas de 0-20cm ($RFE_{Ksat0-20cm}=11$ variáveis; $RFE_{vib0-20cm}=14$ variáveis) e 20-40cm ($VIF_{Ksat20-40cm}=VIF_{vib20-40cm}=17$ variáveis) das variáveis estudadas, respectivamente. Isto mostra a eficácia na redução da dimensionalidade dos dados (variáveis redundantes) e resultando em componentes mais explicativos para modelos preditivos em geral. Como visto, os modelos finais adotados na modelagem dos dados (espacialização - mapa), segundo a separação por horizontes do solo e variável de estudo, foram: modelos vif e RFE para a Ksat e modelos *Stepwise* AIC e RFE para a vib, nos horizontes superficiais e subsuperficiais dessas variáveis, respectivamente. Os resultados obtidos permitiram classificar tais variáveis e são imprescindíveis no mapeamento digital de recursos hídricos nos solos da bacia em estudo, principalmente na etapa de pré-modelagem (seleção e redução da dimensionalidade de dados de entrada).

No geral, o método RFE e *Stepwise*AIC foram os que retornaram melhor qualidade na modelagem *Random Forest* para as variáveis físico-hídricas, quando aplicados às camadas 0-20cm e 20-40cm, independentemente da variável de estudo ($RFE: Ksat_{0-20cm}=14,33\%$ e $vib_{0-20cm}=47,35\%$; *Stewise*AIC: $Ksat_{20-40cm}=10,88\%$ e $vib_{20-40cm}=20,03\%$), resultando em melhores explicações para estes modelos. Entretanto, no geral, o ajuste dos modelos RF foi menor para os horizontes superficiais e superficiais da

variável Ksat em relação aos mesmos da variável vib, provavelmente devido a sua alta variabilidade espacial afetar a modelagem pelo algoritmo.

A escolha do melhor(es) método(s) de seleção de variáveis deve ser cautelosamente pensada pautando-se em três propósitos principais: o objetivo do estudo, as características e comportamento da variável a ser caracterizada, o conhecimento do pesquisador sob a área de estudo e a qualidade atingida pelos modelos. Os dois últimos propósitos implicam diretamente sobre o primeiro, garantindo que a modelagem final seja mais fidedigna à realidade comportamental da variável em dada área de estudo.

Os produtos deste trabalho permitiram a compreensão da variabilidade da condutividade hidráulica saturada (Ksat) e da velocidade de infiltração básica dos solos (vib) para a área de estudo. A espectroradiometria associada aos produtos topográficos derivados do MDE resultaram em modelos mais robustos, evidenciado pela qualidade avaliada na pesquisa, garantindo resultados mais confiáveis ao mapeamento e modelagem digital de atributos físico-hídricos dos solos. A radiometria apresentou qualitativamente uma importância para análises da composição granulométrica dos solos à nível superficial, porém qualitativamente existem limitações que inviabilizam a análise crítica e indeterminam a potencialidade desses dados em modelagens numéricas via algoritmos *machine learning*.

Sugere-se, como forma de comparação e busca por consolidação e potencialidade dos dados, a investigação de outros modelos de aprendizado de máquinas e até mesmo de geoestatística para verificar se a estimativa dos atributos físico-hídricos sofre diferenciação ao ser modelado por diferentes algoritmos e dependência espacial (geoestatística). Ainda, é sugerida a atribuição de parâmetros químicos, biológicos que possam vir a apresentar diferentes interrelações com os atributos do estudo, e melhorar o entendimento destes, além de melhorar a qualidade do modelo final.

Os autores sugerem e incentivam o uso dos parâmetros e critérios adotados nesta pesquisa aplicado à outras áreas de estudo (ressalvando as características da região onde o estudo for aplicado) e com uso de diferentes modelos, de forma a obter uma resposta de como a pesquisa impacta a ciência do solo e se prova importante na análise de atributos físico-hídricos na área da Pedometria e, principalmente, no ramo do Mapeamento Digital de Solos.

A abordagem adotada foi potencialmente útil para compreender a relação intrínseca entre a espectroradiometria de superfície, a topografia, a composição dos solos

e a resposta hídrica nos solos superficiais ao envolver o banco de dados estudado; e, ainda, ressalta a crítica do quão tênue é a seleção de variáveis no processo de classificação da Ksat e vib no que tange o mapeamento digital de solos em geral.

3 CONCLUSÕES GERAIS

Com base nos estudos desenvolvidos nos capítulos I e II deste trabalho, a modelagem de atributos físico-hídricos através da implementação de algoritmos e conhecimentos de estatística multivariada mostra-se uma potencialidade no que tange ao Mapeamento Digital de Solos, principalmente pela capacidade de avaliar um grande volume de dados de diversas características e origens, agregando cada vez mais conhecimento sob os fenômenos ocorridos na superfície terrestre. É importante ressaltar que o uso de algoritmos computacionais em ambientes como o RStudio torna capaz a replicabilidade de parâmetros desse estudo e um maior controle sob todas as etapas da modelagem.

Os estudos preliminares desenvolvidos nesta pesquisa auxiliaram no desenvolvimento e na tomada de decisão acerca dos processos executados nos capítulos I e II, tratando da compreensão de três principais informações: característica do banco de dados pedológico usado (estudos em estatística descritiva e análise correlação), condição hidrológica dos solos (através da caracterização dos solos da bacia segundo técnicas e índices avaliados em hidrologia) e variabilidade de atributos dos solos ao longo do perfil (aplicação da ferramenta AQP). Esses estudos foram fundamentais para a compreensão do regime hídrico e das características particulares da bacia hidrográfica do Rio Guapi-Macacu, visando utilizar dados e ferramentas potencialmente aplicáveis na representatividade dos fenômenos a ela associados, avaliando-os da forma mais fidedigna possível. Dentre estes estudos preliminares, destaca-se as aplicações das ferramentas AQP nos estudos hidropedológicos, capazes de harmonizar, padronizar e analisar, tanto qualitativamente quanto quantitativamente, os perfis dos solos amostrados na bacia, inferindo informações importantes sobre a dispersão e distribuição dos atributos físicos e químicos dos solos ao longo das camadas.

Com relação aos resultados do Capítulo I, o *Random Forest* foi o modelo de melhor ajuste/predição do conjunto de dados físico-hídricos da bacia do Rio Guapi-Macacu frente aos demais (Árvores de Regressão e Regressão Linear Múltipla), sendo capaz de representar a variabilidade dos atributos físico-hídricos dos solos da bacia. No geral, os modelos baseados em árvores apresentaram melhor acurácia, indicada pelos índices estatísticos quantificados (*RMSE*, *MAE* e R^2), em contrapartida aos modelos regressivos avaliados devido possivelmente à sua incapacidade de abranger a alta variabilidade destes atributos na interface pedosfera-hidrosfera e por dependerem de pressupostos estatísticos.

Com relação aos resultados do Capítulo II, destaca-se a potencialidade da integração dos dados físico-hídricos (mensurados *in situ*) com multi-informações oriundas de dados

radiométricos, espectrais e pedológicos, na descrição de processos hidrológicos dos solos através de procedimentos de mapeamento e modelagem digital preditiva. Dentre as técnicas adotadas, o RFE e o *Stepwise AIC* mostraram-se mais aptos a reduzir e selecionar variáveis para modelagem preditiva (mineração de dados), conforme os dados aplicados (cinquenta e quatro covariáveis no total) e os critérios deste estudo. Verificou-se que os modelos ajustados segundo a profundidade da camada e respectivos atributos físico-hídricos, obtiveram respostas variadas na seleção de variáveis de entrada para as técnicas propostas (*vif*, RFE e *Stepwise AIC*) em decorrência das particularidades dos solos da bacia e da alta variabilidade dos atributos físico-hídricos ao longo das camadas do solo.

4 CONSIDERAÇÕES FINAIS

O presente estudo apresenta contribuições tecnológicas pautadas na compreensão multidisciplinar da dinâmica hídrica dos solos, visando a preservação e sustentabilidade dos recursos hídricos no país e serve de suporte para o desenvolvimento de outros estudos associados à prevenção a ocorrência de fenômenos como a erosão hídrica dos solos, os deslizamentos de massa, o desabastecimento e contaminação de aquíferos, a perda de biodiversidade e, ainda, a ocorrência e disseminação do fogo em vegetações ao regular a temperatura de superfície em sistemas ambientais. Logo, a posteriori, mapas interpretativos podem ser desenvolvidos pela consolidação desta pesquisa para várias finalidades.

O mapeamento e a modelagem digital de atributos dos solos, principalmente àqueles com alta variabilidade espacial como a velocidade de infiltração básica (*vib*) e da condutividade hidráulica saturada (*Ksat*), são processos que envolvem cenários complexos, demandando maiores estudos e conhecimentos acerca das variáveis condicionantes destes. Esta pesquisa trouxe uma metodologia concisa capaz de aproximar a modelagem espacial destes atributos à realidade de ocorrência do fenômeno hidrológico ocorrido nos solos existentes da bacia hidrográfica do Rio Guapi-Macacu, ressaltando a necessidade de maiores estudos acadêmicos sobre a hidropedologia.

5 RECOMENDAÇÕES E SUGESTÕES PARA TRABALHOS FUTUROS

A fim de amplificar os estudos desenvolvidos, sugere-se:

- aumento na quantidade de pontos amostrais de velocidade de infiltração básica (*vib*) e condutividade hidráulica saturada (*Ksat*) para melhor representação espacial dos atributos na área da bacia;

- aumento na quantidade de pontos amostrais dos atributos físico-hídricos estudados, segundo parâmetros de distribuição de classes de solos, visando homogeneizar a representatividade das condições hidrológicas dos solos e padronizar a classificação;
- implementação e avaliação de outros modelos de aprendizagem de máquina que não sejam baseados em árvores, como por exemplo *Support Vector Machines* e *Artificial Neural Networks*, buscando-se comparar mais métodos àqueles desenvolvidos nesta pesquisa e aprimorar assim a representação da variabilidade dos atributos físico-hídricos estudados;
- realizar validações de campo após a modelagem e mapeamento, com o objetivo de verificar se as predições realizadas são capazes de representar a distribuição dos atributos físico-hídricos segundo os critérios de condição hidrológica e profundidade da camada do solo, conforme avaliados nesta pesquisa;
- estudos aprofundados sobre a influência de dados aerogeofísicos na estimativa de atributos físico-hídricos, já que estes apresentam potencialidades para aplicações não somente na hidrogeologia, mas também na hidropedologia.

6 REFERÊNCIAS BIBLIOGRÁFICAS

- ABDELBAKI, A. M.; YOUSSEF, M. A.; NAGUIB, E. M.; KIWAN, M. E.; EL-GIDDAWY, E. I. Evaluation of pedotransfer functions for predicting saturated hydraulic conductivity for US soils. In: 2009 Reno, Nevada, June 21-June 24, 2009. **American Society of Agricultural and Biological Engineers**, p. 1., 2009.
- ABDI, H.; WILLIAMS, L. J. Principal component analysis. **Wiley interdisciplinary reviews: computational statistics**, v. 2, n. 4, p. 433-459, 2010.
- ALMEIDA, C. X. D.; CENTURION, J. F.; FREDDI, O. D. S.; JORGE, R. F.; BARBOSA, J. C. Funções de pedotransferência para a curva de resistência do solo à penetração. **Revista Brasileira de Ciência do Solo**, p. 2235-2243, 2008.
- ALVES SOBRINHO, T. Desenvolvimento de um infiltrômetro de aspersão portátil. **Tese de Doutorado**. Viçosa, Universidade Federal de Viçosa, 1997. 85p.
- ANDRADE, R. D. S.; STONE, L. F. Estimativa da umidade na capacidade de campo em solos sob Cerrado. **Revista Brasileira de Engenharia Agrícola e Ambiental**, v. 15, n. 2, p. 111-116, 2011.
- ANDREWS, D. W. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. **Econometrica: Journal of the Econometric Society**, p. 817-858, 1991.
- ANJOS, C. S. D.; ALMEIDA, C. M. D.; GALVÃO, L. S.; SOUZA FILHO, C. R.; LACERDA, M. G.; PRATI, R. C. Análise do nível de legenda de classificação de áreas urbanas empregando imagens multiespectrais e hiperespectrais com os métodos árvore de decisão C4. 5 e floresta randômica. **Boletim de Ciências Geodésicas**, v. 23, n. 2, p. 371-388, 2017.
- AQUINO, L. S.; TIMM, L. C.; NEBEL, A. L. C.; MIOLO, E. C. C.; PARFITT, J. M. B.; PAULETTO, E. A. Avaliação da capacidade de funções de pedotransferência em prever a estrutura de variabilidade espacial. **Simpósio de Geoestatística Aplicada em Ciências Agrárias**, ISBN: 978-85-98187-19-8, v. 4., Botucatu, São Paulo, 2009.
- ARRUDA, G. P. D.; DEMATTÊ, J. A. M.; CHAGAS, C. D. S. Mapeamento digital de solos por redes neurais artificiais com base na relação solo-paisagem. **Revista Brasileira de Ciência do Solo**, v. 37, n. 2, p. 327-338, 2013.
- BACCHI, O. Teoria dos fatores de escala na análise comparativa de métodos de determinação da condutividade hidráulica de um solo. Tese (Doutorado) — **Escola Superior de Agricultura**, USP, Piracicaba, 1988. Orientador: Klaus Reichardt.
- BANNARI, A.; ASALHI, H.; TEILLET, P. M. Transformed Difference Vegetation Index (TDVI) for Vegetation Cover Mapping. In **Proceedings of the Geoscience and Remote Sensing Symposium**, IGARSS '02, IEEE International, Volume 5 (2002).
- BANGIRA, T.; ALFIERI, S.M.; MENENTI, M.; VAN NIEKERK, A. Comparing thresholding with machine learning classifiers for mapping complex water. **Remote Sensing**, v. 11, n. 11, p. 1351, 2019.
- BEAUDETTE, D. E., ROUDIER, P., O'GEEN, A. T. Algorithms for quantitative pedology: a toolkit for soil scientists, **Computers & Geosciences**, vol. 52, pp. 258-268, 2013.
- BELGIU, M.; CSILLIK, O. Sentinel-2 cropland mapping using pixel-based and object-based time-weighted dynamic time warping analysis. **Remote sensing of environment**, v. 204, p. 509-523, 2018.
- BERTONI, J.; LOMBARDI NETO, F. Conservação do solo. Piracicaba, **Livro ceres**, 1990. 355p.
- BEVEN, K. J.; KIRKBY, M. J. A physically-based variable contributing area model of basin hydrology. **Hydrology Science Bulletin**, 24, 43-69, 1979.

- BLASCHKE, T., KUX, H. Sensoriamento remoto e SIG avançados: novos sistemas sensores, métodos inovadores. **Oficina de Textos**, 2005.
- BLUM, A. L.; LANGLEY, P. Selection of relevant features and examples in machine learning. **Artificial intelligence**, v. 97, n. 1-2, p. 245-271, 1997.
- BOCK, M.; BÖHNER, J.; CONRAD, O.; KOETHE, R.; RINGELER, A. Methods for creating Functional Soil Databases and applying Digital Soil Mapping with SAGA GIS. In: Hengl, T., Panagos, P., Jones, A., Toth, G. [Eds.]: **Status and prospect of soil information in south-eastern Europe: soil databases, projects and applications**. EUR 22646 EN Scientific and Technical Research series, Office for Official Publications of the European Communities, Luxemburg, p.149-162, 2007.
- BÖHNER, J.; KOETHE, R.; CONRAD, O.; GROSS, J.; RINGELER, A.; SELIGE, T. Soil Regionalisation by Means of Terrain Analysis and Process Parameterisation. In: Micheli, E., Nachtergaele, F., Montanarella, L. [Ed.]: **Soil Classification**, 2001. European Soil Bureau, Research Report No. 7, EUR 20398 EN, Luxembourg. pp.213-222.
- BÖHNER, J.; SELIGE, T. Spatial prediction of soil attributes using terrain analysis and climate regionalisation. In: Böhner, J., McCloy, K.R., Strobl, J. [Eds.]: **SAGA - Analysis and Modelling Applications**, Goettinger Geographische Abhandlungen, Goettingen, [S.L.], v. 15, p. 13-28, 2006. http://downloads.sourceforge.net/saga-gis/ggal15_02.pdf.
- BÖHNER, J.; ANTONIC, O. Land-surface parameters specific to topo-climatology. In: Hengl, T., Reuter, H. [Eds.]: **Geomorphometry - Concepts, Software, Applications. Developments in Soil Science**, Volume 33, p.195-226, Elsevier, 2009.
- BONETT, D.G.; WRIGHT, T. A. Sample size requirements for estimating Pearson, Kendall and Spearman correlations. **Psychometrika**, v. 65, n. 1, p. 23-28, 2000.
- BORGES, E.; ANTONINO, A.C.D.; DALL'OLIO, A.; AUDRY, P., CARNEIRO, C. J. G. Determinação da condutividade hidráulica e da sortividade de um solo não-saturado utilizando-se permeâmetro a disco. **Pesq. Agropec. Bras.**, 34:2083-2089, 1999.
- BOTULA, Y.D.; CORNELIS, W.M.; BAERT, G.; van RANST, E. Evaluation of pedotransfer functions for predicting water retention of soils in Lower Congo (D.R. Congo). **Agric. Water Manage.**, 111:1-10, 2012.
- BOTULA, Y. D.; VAN RANST, E.; CORNELIS, W. M. Pedotransfer functions to predict water retention for soils of the humid tropics: a review. **Revista Brasileira de Ciência do Solo**, v. 38, n. 3, p. 679-698, 2014.
- BOUMA, J. Using soil survey data for quantitative land evaluation. **Adv.Soil Sci.**, 9:177-213, 1989.
- BRADLEY, P.; MANGASARIAN, O.; STREET, W. Feature selection via mathematical programming. Technical Report. **Informs Journal on Computing**, 10, 209-217, 1998.
- BREIMAN, L. Random Forests. **Machine Learning**, v. 45, no. 1, pp. 5-32, 2001. ISSN 1573-0565. doi:10.1023/A:1010933404324, 2001.
- BRENNER, V.C.; GUASSELLI, L.A. Índice de diferença normalizada da água (NDWI) para identificação de meandros ativos no leito do canal do rio Gravataí/RS-Brasil. **Anais do XVII Simpósio Brasileiro De Sensoriamento Remoto-SBSR**, João Pessoa - PB, Brasil, v. 25, 2015.
- BREUSCH, T.S. Testing for Autocorrelation In Dynamic Linear Models. **Australian Economic Papers**, vol. 17, no. 31, pp. 334-355, 1978.
- BREUSCH, T. S. & PAGAN, A. R. 1979, A Simple Test for Heteroskedasticity and Random Coefficient Variation, **Econometrica**, vol.47, no. 5, pp. 1287-1294, 2002. doi:10.2307/1911963.
- BUSSAB, W. O.; MORETTIN, P. A. **Estatística Básica**, 5. Ed., Saraiva, São Paulo.
- CARVALHO JUNIOR, W. D.; CHAGAS, C. DA S.; FERNANDES FILHO, E.I.; VIEIRA, C.A.O.; SCHAEFER, C.E.G.; BHERING, S.B.; FRANCELINO, M.R. Digital soilscape

mapping of tropical hillslope areas by neural networks. *Scientia Agrícola*, Piracicaba, v. 68, n. 6, 691-696 p., nov./dec. 2011.

CARVALHO JUNIOR, W.D.; CHAGAS, C.D.S.; MUSELLI, A.; PINHEIRO, H.S.K.; PEREIRA, N.R.; BHERING, S.B. Método do hipercubo latino condicionado para a amostragem de solos na presença de covariáveis ambientais visando o mapeamento digital de solos. Embrapa Solos-**Artigo em periódico** indexado (ALICE), 2014.

CARVALHO JUNIOR; W., PEREIRA; N.R., CHAGAS; C.S.; BHERING, S.B.; CALDERANO FILHO, B. Pedologia quantitativa: o perfil mediano e o perfil médio. **Congresso Brasileiro de Ciência do Solo**, no 35, pp. 4, Sociedade Brasileira de Ciência do Solo, Natal, 2015.

CEDDIA, M.B.; VIEIRA, S.R.; VILLELA, A.L.O.; MOTA, L. DOS S.; ANJOS, L.H.C. DOS; CARVALHO, D.F. Topography and spatial variability of soil physical properties. *Sci. Agric.* 66, 338–352, 2009. <https://doi.org/10.1590/S0103-90162009000300009>.

CEDDIA, M. B. Apostila de Constituição Física do Solo e Água no Solo. **Curso de Pós-Graduação em Ciência do Solo, Disciplinas de Física do Solo**. Departamento de Solos da Universidade Federal Rural do Rio de Janeiro, Seropédica, RJ. Consultada em 15 de agosto de 2020.

CHAGAS, C. DA S. Mapeamento digital de solos por correlação ambiental e redes neurais em uma bacia hidrográfica no domínio de mar de morros. [**Tese de Doutorado**]. Universidade Federal de Viçosa, Viçosa, Minas Gerais, p.151, 223p, 2006.

CHAGAS, C. S.; JÚNIOR, W.C., PEREIRA, N. R.; BHERING, S. B.; FONSECA, O. O. M.; PINHEIRO, H. S. K.; MUSELLI, A.; JEUNE, W. Descrição e resultados das análises dos perfis de solo coletados nas bacias hidrográficas dos rios guapi-macacu e caceribu - contrato nº 6000.00419115.08.2, **Relatório Técnico**, Embrapa Solos, Rio de Janeiro, 2011. Acesso: 29 Abril 2021, <<https://ainfo.cnptia.embrapa.br>>.

CHAVES, H.M.L.; ORLOWISK, E.; ROLOFF, G. Previsão da infiltração sob condições dinâmicas de selamento superficial. *R. Bras. Ci. Solo*, 17:141-147, 1993.

CIDIN, A. C. M.; LOPES-ASSAD, M. L. R. C.; CRESTANA, S. Avaliação de atributos do solo para o desenvolvimento de função de pedotransferência estimativa de teores de carbono em solos brasileiros: análise exploratória. In: Embrapa Instrumentação-Artigo em anais de congresso (ALICE). In: **Simpósio Nacional de Instrumentação Agropecuária (IV SIAGRO)**, 4., 2019, São Carlos, SP. Anais... São Carlos, SP: Embrapa Instrumentação, p. 764., 2019.

CLEVERS, J.G.P.W.; GITELSON, A.A. Remote estimation of crop and grass chlorophyll and nitrogen content using red-edge bands on Sentinel-2 and 3. *Int. J. Appl. Earth Obs. Geoinf.* 2012, 23, 344–351, 2012.

CONRAD, O.; BECHTEL, B.; BOCK, M.; DIETRICH, H.; FISCHER, E.; GERLITZ, L.; WEHBERG, J.; WICHMANN, V.; AND BOEHNER, J. **System for Automated Geoscientific Analyses (SAGA) v. 2.1.4**. *Geosci. Model Dev.*, 8, 1991-2007, doi:10.5194/gmd-8-1991-2015, 2015.

COOK, F.J.; BROEREN, A. Six methods for determining sortivity and hydraulic conductivity with disc permeameters. *Soil Sci.*, 157:1-11; 1994.

COLIN, B., CLIFFORD, S.; WU, P. P.; RATHMANNER, S.; MENGERSEN, K. Using Boosted Regression Trees and Remotely Sensed Data to Drive Decision-Making. *Open Journal of Statistics*, v. 07, n. 05, p. 859–875, 2017. ISSN 2161-718X, 2161-7198.

CPRM - CENTRO DE PESQUISAS EM RECURSOS MINERAIS DO SERVIÇO GEOLÓGICO DO BRASIL (SBG). Relatório final do levantamento e processamento dos dados magnetométricos e gamaespectrométricos. Projeto Aerogeofísico Rio de Janeiro. Prospectores Aerolevantamentos e Sistemas Ltda, **Texto Técnico**, v.1, 2012.

- CUNHA, A. M. Seleção De Variáveis Ambientais E De Algoritmos De Classificação Paramapeamento Digital De Solos. **Tese De Doutorado**. Universidade Federal De Viçosa, Viçosa – Mg, P.32-33, 121p, 2013.
- DANIEL, C.; WOOD, FRED S. Fitting Equations To Data: Computer Analysis Of Multifactor Data. John Wiley & Sons, Inc., 1980. Isbn:978-0-471-05370-5.
- DARCY, H. P. G. Les Fontaines Publiques De La Ville De Dijon: Exposition Et Application Des Principes A Suivre Et Des Formulesa Employer Dans Les Questions De Distribution D'eau. **V. Dalmont**, Paris, 1986.
- DARST, B. F., MALECKI, K. C., & ENGELMAN, C. D. Using recursive feature elimination in random forest to account for correlated variables in high dimensional data, **BMC genetics**, vol. 19, no. 1, pp. 1-6, 2018.
- DA SILVA NUNES, J. G.; UAGODA, R. Análise da eficiência dos métodos indiretos para o mapeamento de solos frente às técnicas diretas, e suas possíveis associações: Uma revisão metodológica. **Revista Brasileira de Geografia Física**, v. 13, n. 02, p. 487-509, 2020.
- DE ALMEIDA PIRES, C.; MIRANDA, A.W.A. Análise Geométrica de Lineamentos e suas Relações com Águas Subterrâneas Associadas ao Aquífero Guaratiba-Região de Campo Grande e Guaratiba, RJ. **[Trabalho de conclusão de curso - Geologia]**. Departamento de Geociências, Universidade Federal Rural do Rio de Janeiro. Rio de Janeiro, 2017.
- DON, A.; SCHUMACHER, J.; SCHERER-LORENZEN, M.; SCHOLTEN, T.; SCHULZE, E. D. Spatial and vertical variation of soil carbon at two grassland sites—implications for measuring soil carbon stocks. *Geoderma*, v. 141, n. 3-4, p. 272-282, 2007.
- DOS SANTOS, J. C. P.; DE ARAUJO FILHO, J. C. Avaliação detalhada do potencial de terras para irrigação nas áreas de reassentamento de colonos do Projeto Jusante-área 3, Glória, BA. **Recife: Embrapa Solos**, 2008.
- DOS SANTOS, H. G.; JACOMINE, P. K. T.; DOS ANJOS, L. H. C.; DE OLIVEIRA, V. A.; LUMBRERAS, J. F.; COELHO, M. R.; DE ALMEIDA, J. A.; FILHO, J. C. A.; DE OLIVEIRA, J. B.; CUNHA, T. J. F. **Sistema brasileiro de classificação de solos**, Empresa Brasileira de Pesquisa Agropecuária Embrapa Solos, Ministério da Agricultura, Pecuária e Abastecimento, 5nd ed., Brasília, Distrito Federal, 2018.
- DRM – DIRETORIA DE RECURSOS MINERAIS DO ESTADO DO RIO DE JANEIRO. Carta geológica na escala 1:50.000 dos municípios de Itaboraí, Itaipava, Nova Friburgo, Teresópolis, Petrópolis e Rio Bonito. Governo do Estado do Rio de Janeiro, Secretaria de estado de energia, da indústria naval e do petróleo, Rio de Janeiro, Brasil, 2019. **Dados cedidos de requerimento do Ofício nº 07/2019 à UFRRJ**.
- DRURY, S. A. Image interpretation in geology. London: **Allen and Unwin**, 243 pp, 1987.
- DUBE, T.; MUTANGA, O. Estimating aboveground biomass in the uMgeni catchment, South Africa using medium-spatial resolution Landsat 8 multispectral sensor. **In: 2nd National Conference on Global Change at Nelson Mandela Metropolitan University**, Conference Paper, Port Elizabeth, South Africa, 2014.
- DURBIN, J., WATSON, G. S. Testing for Serial Correlation in Least Squares Regression, I and II, **Biometrika**, vol. 38, no. 1-2, pp. 159–179, 1951. doi:10.1093/biomet/38.1-2.159.
- FERRARI, A. L. Evolução tectônica do Gráben da Guanabara, DSc. **Thesis**, Universidade de São Paulo, São Paulo, 2001.
- ELRICK, D.E.; REYNOLDS, W.D.; TAN, K.A. Hydraulic conductivity measurements in the unsaturated zone using improved well permeameter analysis. **Ground Water Monit. Rev.** 9:184-193, 1989.
- ESA. Sentinel Online: Technical Guide for Sentinel-2 MSI Instrument: Overview. The European Space Agency. **Sítio online**. Acesso: 28 fevereiro 2020. <<https://earth.esa.int/web/sentinel/technical-guides/sentinel-2-msi/msi-instrument>>.

- ESA. Sentinel Online: User guide for Sentinel-2 MSI Processing Levels: Level-2 Products. The European Space Agency **Sítio online**. Acesso: 28 fevereiro 2020. <<https://sentinel.esa.int/web/sentinel/user-guides/sentinel-2-msi/processing-levels/level-2>>.
- ESA. Sentinel Online: MultiSpectral Instrument (MSI) Overview. The European Space Agency. **Sítio online**. Acesso: 05 julho 2021. <<https://sentinels.copernicus.eu/web/sentinel/technical-guides/sentinel-2-msi/msi-instrument>>.
- EVANS, J. S.; CUSHMAN, S. A. Gradient Modeling of Conifer Species Using Random Forest. **Landscape Ecology**, v5, pg.673-683, 2009.
- EVANS J. S.; MURPHY, M. A.; HOLDEN, Z. A.; CUSHMAN, S. A. Modeling species distribution and change using Random Forests. **In: Predictive Modeling in Landscape Ecology**. Springer, New York, NY, p. 139-159, 2011.
- EVERTS, C.J.; KANWAR, R.S. Interpreting tension-infiltrometer data for quantifying soil macropores: some practical considerations. **Trans. Am. Soc. Agric. Eng.**, 36:423-428, 1992.
- FAO. **Revised World Soil Charter**. Italy, p. 4-5, 2015.
- FAO. **Voluntary Guidelines for Sustainable Soil Management**. Rome, Italy. 2017.
- FERNÁNDEZ-MANSO, A.; FERNÁNDEZ-MANSO, O.; QUITANO, C. SENTINEL-2A red-edge spectral indices suitability for discriminating burn severity. **Int. J. Appl. Earth Obs. Geoinf.** 2016, 50, 170–175, 2016.
- FERRARI, F. Estimadores viesados para modelos de regressão em presença de multicolinearidade, DSc. **Thesis**, Escola Superior de Agricultura Luiz de Queiroz, Universidade de São Paulo, Piracicaba, São Paulo, 1989.
- FLORINSKY, I. V.; EILERS, R. G.; MANNING, G. R.; FULLER, L. G. Prediction of soil properties by digital terrain modeling. **Environmental Modelling & Software**, v. 17, 295-311 p., 2002.
- GALLANT, J.C.; DOWLING, T.I. A multiresolution index of valley bottom flatness for mapping depositional areas. **Water Resources Research**, 39/12:1347-1359, 2003.
- GARCÍA-SINOVAS, D.; REGALADO, C.; MUÑOZ-CARPENA, R.; ÁLVAREZ-BENEDÍ, J. Comparación de los permeámetros de Guelph y Philip-Dunne para la estimación de la conductividad hidráulica saturada del suelo, **Actas de las V Jornadas sobre Investigación en la Zona no Saturada**, pp. 31-36, 2001.
- GERLITZ, L.; CONRAD, O.; BÖHNER, J. Large scale atmospheric forcing and topographic modification of precipitation rates over High Asia – a neural network based approach. **Earth System Dynamics**, 6, 1-21. doi:10.5194/esd-6-61-2015, 2015.
- GITELSON, A. A.; MERZLYAK, M. N.; ZUR, Y.; STARK, R.; GRITZ, U. Non-destructive and remote sensing techniques for estimation of vegetation status. **In the Proceedings of the 3rd European Conference on Precision Agriculture**, Montpellier, France, 2001.
- GOEL, N.S.; QIN, W. Influences of canopy architecture on relationships between various vegetation indices and LAI and FPAR: A computer simulation. **Remote Sensing Reviews**, v. 10, n. 4, p. 309-347, 1994.
- GOMES, A. DA S.; FERREIRA, A.C. DE S.; PINHEIRO, É.F.M.; MENEZES, M.D. DE; CEDDIA, M.B.; GOMES, A. DA S.; FERREIRA, A.C. DE S.; PINHEIRO, É.F.M.; MENEZES, M.D. DE; CEDDIA, M.B.. The use of Pedotransfer functions and the estimation of carbon stock in the Central Amazon region. *Sci. Agric.* 74, 450–460, 2017. <https://doi.org/10.1590/1678-992x-2016-0310>.
- GREGORUTTI, B.; MICHEL, B.; SAINT-PIERRE, P. Correlation and variable importance in random forests, **Statistics and Computing**, vol.27, pp.659–78, 2017. <https://doi.org/10.1007/s11222-016-9646-1>.

- GRIMM, R.; BEHRENS, T.; MÄRKER, M.; ELSENBEEER, H. Soil organic carbon concentrations and stocks on Barro Colorado Island — Digital soil mapping using Random Forests analysis. **Geoderma**, v. 146, n. 1-2, p. 102–113, jul. 2008. ISSN 00167061
- GUISAN, A.; WEISS, S. B.; WEISS, A. D. GLM versus CCA spatial modeling of plant species distribution. **Plant Ecology**, v. 143, pg. 107-122, 1999.
- GUYON, I.; WESTON, J.; BARNHILL, S.; VAPNIK, V. Gene selection for cancer classification using support vector machines, **Machine Learning**, vol. 46, pp. 389–422. doi:10.1023/A:1012487302797, 2002.
- HAGHVERDI, A.; ÖZTÜRK, H. S.; DURNER, W. Measurement and estimation of the soil water retention curve using the evaporation method and the pseudo continuous pedotransfer function. **Journal of hydrology**, v. 563, p. 251-259, 2018.
- HAIR, J. F. JR.; ANDERSON, R. E.; TATHAM, R. L.; BLACK, W. C. *Multivariate Data Analysis* (3rd ed. Book). New York: **Macmillan**, 1995.
- HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. *The elements of statistical learning: data mining, inference, and prediction*, **Springer Science & Business Media**, 2nd ed. ISBN 0-387-95284-5, 2009.
- HERBST, M.; DIEKKRÜGER, B.; VERECKEN, H. Geostatistical co-regionalization of soil hydraulic properties in a micro-scale catchment using terrain attributes. **Geoderma**, v.132, p.206-221, 2006.
- HILLEL, D. *Fundamentals of Soil Physics*. **Academic Press**. 1980. London.
- HORA, A. F.; HWA, C. S.; HORA, M. A. G. M. Projeto Macacu: Planejamento Estratégico da Região Hidrográfica dos Rios Guapi-Macacu e Caceribu-Macacu. **Fundação Euclides da Cunha**, Universidade Federal Fluminense, Niterói, Rio de Janeiro, 2010.
- HUETE, A.R. A Soil-Adjusted Vegetation Index (SAVI). **Remote Sensing of Environment**, v. 25, p. 295-309, 1988.
- HUETE, A.; DIDAN, K.; MIURA, T.; RODRIGUEZ, E.P.; GAO, X.; FERREIRA, L.G. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. **Remote sensing of environment**, v. 83, n. 1-2, p. 195-213, 2002.
- HUNT, E.R.; DAUGHTRY, C.S.T.; EITEL, J. U.; LONG, D.S. Remote sensing leaf chlorophyll content using a visible band index. **Agronomy Journal**, v. 103, n. 4, p. 1090-1099, 2011.
- INEA - INSTITUTO ESTADUAL DO AMBIENTE. Região Hidrográfica Baía de Guanabara, 2018. Acesso: 23 Agosto 2019. <<http://www.inea.rj.gov.br/Portal/Agendas/GESTAODEAGUAS/instrumentosdegestoderechid/planoderecursoshidricos/baiadeguanabaraagendaazul/index.htm>>.
- INMAN, H. F. Karl Pearson and RA Fisher on statistical tests: a 1935 exchange from Nature. **The American Statistician**, v. 48, n. 1, p. 2-11, 1994.
- IWAHASHI, J.; PIKE, R. J. Automated classifications of topography from DEMs by an unsupervised nested-means algorithm and a three-part geometric signature. **Geomorphology**, Vol. 86, pp. 409–440, 2007.
- IZBICKI, R.; DOS SANTOS, T. M. Machine Learning sob a ótica estatística: Uma abordagem preditiva para a estatística com exemplos em R, **Departamento de Estatística**, Universidade Federal de Santa Catarina, Santa Catarina, Brasil, 2019. Acesso: 29 Abril 2021. <http://est.ufmg.br/~marcosop/est171-ML/MachineLearning_Izbicki.pdf>.
- JALABERT, S.S.M.; MARTIN, M.P.; RENAUD, J.P.; BOULONNE, L.; JOLIVET, C.; MONTANARELLA, L.; ARROUAYS, D. Estimating forest soil bulk density using boosted regression modelling. **Soil Use and Management**, v. 26, n. 4, p. 516-528, 2010.
- JASIEWICZ, J.; STEPINSKI, T. F. Geomorphons — a pattern recognition approach to classification and mapping of landforms. **Geomorphology**, 182, 147-156. ScienceDirect, 2013.

- JEBARA, T.; JAAKKOLA, T. Feature selection and dualities in maximum entropy discrimination. **16th Conference on Uncertainty in Artificial Intelligence (UAI2000)**, pp. 291-300, 2000.
- JEUNE, W. Espacialização da densidade do solo e do estoque de carbono da Bacia Guapi-Macacu/RJ, determinados por meio de funções de pedotransferência (FPTs). 2012. [52 f.]. Dissertação (Programa de Pós-Graduação em Agronomia - Ciência do Solo) - Universidade Federal Rural do Rio de Janeiro, [Seropédica-RJ].
- JING, J.; BAO, T.; GU, Y.; ZHU, Z. Analysis of Ridge Regression for the Multicollinearity of Monitoring Data. **Journal of Hydropower Automation and Dam Monitoring**, v. 3, 2007.
- JOLLIFFE, I.T. Discarding Variables in a Principal Component Analysis. II: Artificial Data. **Journal of the Royal Statistical Society: Series C (Applied Statistics)**, v. 22, n. 1, p. 21-31, 1973.
- KIRSCH, R. Groundwater geophysics. Berlin: **Springer**, Second Edition, 2006. ISBN: 978-3-540-88404-0. doi: 10.1007/978-3-540-88405-7.
- KLAR, A.E. A água no sistema solo-planta-atmosfera. São Paulo, **Nobel**, 1984. 408p.
- KOETHE, R.; LEHMEIER, F. SARA - System zur Automatischen Relief-Analyse. User Manual, 2. Edition. **Dept. of Geography, University of Goettingen, unpublished**, 1996.
- KRAEMER, G. B. Variabilidade espacial dos atributos do solo na delimitação das unidades de mapeamento, **MSc. Dissertation**, Universidade Federal do Paraná, Curitiba, 2007.
- KUHN, M., JOHNSON, K. 2013, Applied predictive modeling. **Springer**, 613p, New York, 2013.
- KOOREVAAR, P.; MENELIK, G.; DIRKSEN, C. Elements of Soil Physics. **Development in Soil Science**, Amsterdam, Elsevier, 13, 228p, 1983.
- LAGACHERIE, P.; MCBRATNEY, A.; VOLTZ, M. Digital soil mapping: an introductory perspective, **Elsevier Science**, 1st ed., vol. 31, pp. 1-600, 2006.
- LATORRE, M.L., CARVALHO JÚNIOR, O.A.D., SANTOS, J.R.D.; SHIMABUKURO, Y.E. Integração de dados de sensoriamento remoto multi resoluções para a representação da cobertura da terra utilizando campos contínuos de vegetação e classificação por árvores de decisão. **Revista Brasileira de Geofísica**, v. 25, n. 1, p. 63-74, 2007.
- LAWALL, S. Modificações na dinâmica hidrológica dos solos em resposta as alterações de uso e cobertura na Bacia do Bonfim, Região Serrana do Rio de Janeiro. **Dissertação** (Mestrado em Ciência, com ênfase em Geografia) – Instituto de Geociências, Universidade Federal do Rio de Janeiro, Rio de Janeiro. 2010.
- LEE, S.; SONG, K. Y.; KIM, Y.; PARK, I. Regional groundwater productivity potential mapping using a geographic information system (GIS) based artificial neural network model. **Hydrogeology Journal**, v. 20, n. 8, p. 1511-1527, 2012.
- LEE, S.; LEE, C.W. Application of decision-tree model to groundwater productivity-potential mapping. **Sustainability**, v. 7, n. 10, p. 13416-13432, 2015.
- LOPES, F.B.; BARBOSA, C.C.; NOVO, E.M.D.M.; ANDRADE, E.M.D.; CHAVES, L.C. Modelagem da qualidade das águas a partir de sensoriamento remoto hiperespectral. **Revista Brasileira de Engenharia Agrícola e Ambiental**, v. 18, p. 13-19, 2014.
- MADRUCCI, V.; TAIOLI, F.; DE ARAÚJO, C.C. Groundwater favorability map using GIS multicriteria data analysis on crystalline terrain, Sao Paulo State, Brazil. **Journal of Hydrology**, v. 357, n. 3-4, p. 153-173, 2008.
- MANSFIELD, E.R.; HELMS, B.P. Detecting multicollinearity. **The American Statistician**, v. 36, n. 3a, p. 158-160, 1982.
- MARQUARIDT, D.W. Generalized inverses, ridge regression, biased linear estimation, and nonlinear estimation. **Technometrics**, v. 12, n. 3, p. 591-612, 1970.

- MCBRATNEY, A.B.; MINASNY, B.; TRANTER, G. Necessary metadata for pedotransfer functions. **Geoderma**, v. 160, n. 3-4, p. 627-629, 2011.
- MCBRATNEY, A.B.; MINASNY, B.; CATTLE, S.R.; VERVOORT, R.W.. From pedotransfer functions to soil inference systems. **Geoderma**, v. 109, n. 1-2, p. 41-73, 2002.
- MCBRATNEY, A.B.; SANTOS, M.L.M.; MINASNY, B. On digital soil mapping. **Geoderma**, v. 117, n. 1-2, p. 3-52, 2003.
- MCKAY, M.D.; BECKMAN, R.J.; CONOVER, W.J. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. **Technometrics**, 21:239-245, 1979.
- MCKENZIE, N. J; AUSTIN, M. P. A quantitative Australian approach to medium and small scale surveys based on soil stratigraphy and environmental correlation. **Geoderma**, v. 57, pp. 329-355, 1993.
- MESQUITA, M. G. B. F. Caracterização estatística da condutividade hidráulica saturada do solo. 2001. 110 f. [Tese (Doutorado)]. Escola Superior de Agricultura de Luiz de Queiroz, Piracicaba, 2001.
- MESQUITA, M. G. B. F.; MORAES, S. O. A dependência entre a condutividade hidráulica saturada e atributos físicos do solo. **Ciência Rural**, Santa Maria, v. 34, n. 3, p. 963–969, maio/jun. 2004.
- MIHOLA, J.; BÍLKOVÁ, D. Measurement of Multicollinearity Using Determinants of Correlation Matrix. **International Journal of Mathematical Sciences**, 2014.
- MINASNY, B.; MCBRATNEY, A. B. A conditioned Latin hypercube method for sampling in the presence of ancillary information. **Computers and Geosciences**, v.32, p.1378-1388, 2006. DOI: 10.1016/j.cageo.2005.12.009.
- MINASNY, B.; MCBRATNEY, A. B. Digital soil mapping: A brief history and some lessons. **Geoderma**, v. 264, p. 301–311, fev. 2016. ISSN 00167061.
- MISHRA, U.; LAL, R.; SLATER, B.; CALHOUN, F.; LIU, D.S.; VAN MEIRVENNE, M. Predicting soil organic carbon stock using profile depth distribution functions and ordinary kriging. **Soil Science Society of America Journal**, v.73, p.614-621, 2009. DOI:10.2136/sssaj2007.0410.
- MONTGOMERY D. R.; DIETRICH, W. E. A physically based model for the topographic control on shallow landsliding. **Water Resources Research**, 30, 1153-1171, 1994.
- MOORE, I. D.; GRAYSON, R. B.; LADSON, A. R. Digital terrain modelling: a review of hydrological, geomorphological, and biological applications. **Hydrological Processes**, Vol.5, No.1, 1991.
- MUKAKA, M. M. A guide to appropriate use of correlation coefficient in medical research. **Malawi medical journal**, v. 24, n. 3, p. 69-71, 2012.
- NANKO, K.; UGAWA, S.; HASHIMOTO, S.; IMAHA, A.; KOBAYASHI, M.; SAKAI, H.; ISHIZUKA, S.; MIURA, S.; TANAKA, N. TAKAHASHI, M.; KANEKO, S. A pedotransfer function for estimating bulk density of forest soil in Japan affected by volcanic ash. **Geoderma**, v. 213, p. 36-45, 2014. <https://doi.org/10.1016/j.geoderma.2013.07.025>.
- NEMES, A.; RAWLS, W.J.; PACHEPSKY, Y.A. Use of the nonparametric nearest neighbor approach to estimate soil hydraulic properties. *Soil Sci. Soc. Am. J.*, 70:327-336, 2006a.
- NEMES, A.; RAWLS, W.J.; PACHEPSKY, Y.A.; van GENUCHTEN, M.TH. Sensitivity analysis of the nonparametric nearest neighbor technique to estimate soil water retention. *Vadose Zone J.*, 5:1222-1235, 2006b.

- NISHIYAMA, L.; COSTA, F. P. M. **Utilização do permeâmetro Guelph e penetrômetro de impacto em estudos de uso e ocupação dos solos em bacias hidrográficas.** *Caminhos de Geografia*, Uberlândia, v. 8, n. 24, pp.131 – 143, 2007.
- NOVAKOWSKI, K.; BICKERTON, G.; LAPCEVIC, P.; VORALEK, J.; ROSS, N. Measurements of groundwater velocity in discrete rock fractures. **Journal of contaminant hydrology**, v. 82, n. 1-2, p. 44-60, 2006.
- O'HAGAN, J.; MCCABE, B. Tests for the severity of multicollinearity in regression analysis: A comment, **The Review of Economics and Statistics**, vol. 57, no.3, pp. 368-370, 1975, 1975.
- OLIVEIRA, K.D.; KAPICHE, A.L.A.F.; COSTA, T.A.; SANCHES, I.D.A. Classificação de atributos topográficos para distinção de propriedades físico-hídricas e termodinâmicas do solo. **Anais do XVIII Simpósio Brasileiro de Sensoriamento Remoto – SBSR**, INPE, Santos – SP, Brasil, 2017. ISBN: 978-85-17-0088-1.
- OTTONI, M.V. Classificação físico-hídrica de solos e determinação da capacidade de campo in situ a partir de testes de infiltração. **[Dissertação de Mestrado em Ciências em Engenharia Civil]**. Universidade Federal do Rio de Janeiro, Ilha d Fundão - RJ, 141 p., 2005.
- OTTONI FILHO, T. B.; OTTONI, M. V.; OLIVEIRA, M. B.; MACEDO, J. R. Uma variação do enunciado de capacidade de campo (CC): I) uso de funções de pedotransferência (FPT) para determinar CC. **Sociedade Brasileira de Ciência do Solos (SBCS)**, Trabalho apresentado em evento, Repositório Institucional de Geociência da CPRM, 2007.
- PACHEPSKY, Y.; RAWLS, W. J. (Ed.). Development of pedotransfer functions in soil hydrology. Elsevier, Book: v.30, 510p., 2004.
- PADARIAN, J.; MINASNY, B.; MCBRATNEY, A. B. Machine learning and soil sciences: A review aided by machine learning tools. **Soil**, v. 6, n. 1, p. 35-52, 2020.
- PÁDUA, E.J. de; GUERRA, A.R.; ZINN, Y.L. Modelagem da densidade do solo em profundidade sob vegetação nativa em Minas Gerais. **Revista Brasileira de Ciência do Solo**, v.39, p.725-736, 2015. DOI: 10.1590/01000683rbcs20140028.
- PEARSON, K. On the theory of contingency and its relation to association and normal correlation. *Drapers' Company Research Memoirs*. **Biometric series I: Dulau and Co**, 1904.
- PEEL, M. C.; FINLAYSON, B. L.; MCMAHON, T. A. Updated world map of the Köppen-Geiger climate classification. **Hydrology and earth system sciences**, v. 11, n. 5, p. 1633-1644, 2007.
- PERERA, Y.Y.; ZAPATA, C.E.; HOUSTON, W.N.; HOUSTON, S.L. Prediction of the soil-water characteristic curve based on grain-size-distribution and index properties. **In Advances in Pavement Engineering**, p. 1-12, 2005.
- PERROUX, K.M.; WHITE, I. Designs for disc permeameters. **Soil Sci. Soc. Am. J.**, 52:1205-1215, 1988.
- PETERSON, K. T.; SAGAN, V.; SLOAN, J. J. Deep learning-based water quality estimation and anomaly detection using Landsat-8/Sentinel-2 virtual constellation and cloud computing. **GIScience & Remote Sensing**, v. 57, n. 4, p. 510-525, 2020.
- PETERSON, R. A. Ordered quantile normalization: a semiparametric transformation built for the cross-validation era. - Best Normalize package R, **Journal of Applied Statistics**, vol. 47, no. 13-15, pp. 1-16, 2019. doi: 10.1080/02664763.2019.1630372.
- PINHEIRO, H.S.K. Mapeamento digital de solos por redes neurais artificiais da bacia hidrográfica do rio Guapi-Macacu, Rio de Janeiro. **[Dissertação de Mestrado]**. Universidade Federal Rural do Rio de Janeiro, Seropédica - RJ, p.87, 175p, 2012.
- PINHEIRO, H.S.K. Métodos de mapeamento digital aplicados na predição de classes e atributos dos solos da Bacia Hidrográfica do Rio Guapi Macacu, RJ. 2015. 178p. **[Tese Doutorado]**. Universidade Federal do Rio de Janeiro, Rio de Janeiro.

- PINHEIRO, H.S.K.; ANJOS, L.H.C.; CHAGAS, C.S. Mapeamento digital de solos por redes neurais artificiais-estudo de caso: bacia hidrográfica do rio Guapi-Macacu, RJ. 2012. 138 f. **[Dissertação de Mestrado em Agronomia, Ciência do Solo]** - Instituto de Agronomia, Universidade Federal Rural do Rio de Janeiro, Seropédica - RJ, 2012.
- PINHEIRO, H.S.K.; BARBOSA, T.P.; ANTUNES, M.A.; CARVALHO, D.C.D.; NUMMER, A.R.; CARVALHO JUNIOR, W.D.; PEREIRA, M.G. Assessment of Phytoecological Variability by Red-Edge Spectral Indices and Soil-Landscape Relationships. **Remote Sensing**, v. 11, n. 20, p. 2448, 2019.
- PINHEIRO, H. S. K.; CHAGAS, C. D. S.; CARVALHO JÚNIOR, W. D.; ANJOS, L. H. C. D. Modelos de elevação para obtenção de atributos topográficos utilizados em mapeamento digital de solos. **Pesquisa Agropecuária Brasileira**, v. 47, n. 9, p. 1384-1394, 2012.
- PINHEIRO, H. S. K.; DA SILVA CHAGAS, C.; DE CARVALHO JÚNIOR, W.; DOS ANJOS, L. H. C. Ferramentas de pedometria para caracterização da composição granulométrica de perfis de solos hidromórficos. **Pesquisa Agropecuária Brasileira**, v. 51, n. 9, p. 1326-1338, 2016. doi: 10.1590/S0100-204X2016000900032.
- PINHEIRO, H. S.; DOS ANJOS, L. H. C.; XAVIER, P. A.; CHAGAS, C. S.; DE CARVALHO JUNIOR, W. Quantitative pedology to evaluate a soil profile collection from the Brazilian semi-arid region, **South African Journal of Plant and Soil**, vol. 35, no. 4, pp. 269-279, 2018.
- PIZARRO, M.A.; EPIPHANIO, J.C.N.; GALVÃO, L.S. Caracterização mineralógica de solos tropicais por sensoriamento remoto hiperespectral. **Pesquisa Agropecuária Brasileira**, v. 36, n. 10, p. 1277-1286, 2001.
- PONCE-HERNANDEZ, R.; MARRIOTT, F.H.C.; BECKETT, P.H.T. An improved method for reconstructing a soil profile from analyses of a small number of samples. **Journal of Soil Science**, v.37, p.455-467, 1986. DOI: 10.1111/j.1365-2389.1986.tb00377.x.
- POTT, C. A.; DE MARIA, I. C. Comparação de métodos de campo para determinação da velocidade de infiltração básica. **Revista Brasileira de Ciência do Solo**, v. 27, n. 1, p. 19-27, 2003.
- PRATES, V.; DE PAULA SOUZA, L. C., BRISKI, S. J.; DE OLIVEIRA JUNIOR, J. C. Utilização de índices para representação da paisagem como apoio para levantamento pedológicos em ambiente de geoprocessamento. **Simpósio Brasileiro De Sensoriamento Remoto**, v. 15, p. 4885-4892, 2010.
- QUEIRÓZ, B. F. D. Estudo de soluções numéricas da equação de Richards através do método de elementos finitos e diferenças finitas para simulação de fluxo unidimensional em solo não-saturado. **Dissertação de mestrado**. Programa de Pós-Graduação em Modelagem Matemática e Computacional, Universidade Federal Rural do Rio de Janeiro, UFRRJ, Seropédica, Rio de Janeiro, p.21-28, 2017.
- RAJAH, P.; ODINDI, J.; MUTANGA, O.; KIALA, Z. The utility of Sentinel-2 Vegetation Indices (VIs) and Sentinel-1 Synthetic Aperture Radar (SAR) for invasive alien species detection and mapping. **Nature Conservation**, v. 35, p. 41, 2019.
- RAMÍREZ, V. M. V.; ARAUJO, G. A.; CARRILLO, C. I. J. B.; ANDRÉS, D.; LATORRE, G.; CASALLAS, J. M. E.; MALDONADO, F. E. M. Mapeamento Digital de Atributos Físico-Químicos de Solos do Planalto Cundiboyacense na Colômbia. **XLVIII Congresso Brasileiro De Engenharia Agrícola - CONBEA 2019**. Centro de Convenções da Unicamp, Campinas, São Paulo, 17 a 19 de setembro, 2019.
- RAWLS, W. J.; GIMENEZ, D.; GROSSMAN, R. Use of soil texture, bulk density, and slope of the water retention curve to predict saturated hydraulic conductivity. **Transactions of the ASAE**, vol. 41, n.4, p.983, 1998.

- R CORE TEAM (2020). R: A language and environment for statistical computing. **R Foundation for Statistical Computing**, Vienna, Austria. URL <https://www.R-project.org/>.
- REICHARDT, K.; TIMM, L. C. Solo, Planta e Atmosfera – Conceitos, Processos e aplicações. **Barueri: Manole**, 1. ed., São Paulo, 2004.
- REICHERT, J.M.; VEIGA, M.; CABEDA, M.S.V. Selamento superficial e infiltração de água em solos do Rio Grande do Sul. **R. Bras. Ci. Solo**, 16:289-298, 1992.
- REYNOLDS, W. D.; ELRICK, D. E. In situ measurement of field-saturated hydraulic conductivity, sorptivity, and the α -parameter using the Guelph permeameter. **Soil Science**, v. 140, n. 4, p. 292-302, 1985. DOI:10.1097/00010694-198510000-00008.
- REYNOLDS, W.D.; ELRICK, D.E. Pondered infiltration from a single ring: I. Analysis of steady flow. **Soil Sci. Soc. Am. J.**, 54:1233-1241, 1990.
- REYNOLDS, W.D.; ELRICK, D.E. Determination of hydraulic conductivity using a tension infiltrometer. **Soil Sci. Soc. Am. J.**, 55:633-639, 1991.
- REYNOLDS, W.; ELRICK, D.; DAVID, E. A method for simultaneous in situ measurement in the vadose zone of field-saturated hydraulic conductivity, sorptivity and the conductivity-pressure head relationship. **Groundwater Monitoring & Remediation**, v. 6, n. 1, p. 84-95, 1986.
- REYNOLDS, W.D.; VIEIRA, S.R. & TOPP, G.C. An Assessment of the single-head analysis for the constant head wheel permeameter. **Can. J. Soil. Sci.**, 72:489-501, 1992.
- RICHARDS, L. A. Capillary conduction of liquids through porous mediums. **Physics**, v.1, n.5, p.318–333, 1931. doi: 10.1063/1.1745010. Disponível em: <<http://dx.doi.org/10.1063/1.1745010>>. Acesso em 04 junho 2020.
- RODRIGUES, L. N.; MAIA, A.; DA SILVA, R. N. Funções de pedotransferência para estimar capacidade de campo, ponto de murcha permanente e densidade global em solos de uma bacia hidrográfica do Bioma Cerrado. In: Embrapa Meio Ambiente-Artigo em anais de congresso (ALICE). In: **Congresso Brasileiro de Engenharia Agrícola**, 40., 2011, Cuiabá. Geração de tecnologias inovadoras e o desenvolvimento do cerrado brasileiro: anais. Cuiabá: SBEA, 2011.
- ROMANO, N.; CHIRICO, G. B. The role of Terrain Analysis in using and Developing Pedotransfer Functions. **Developments in Soil Science**, v. 30, 2004.
- ROSSITER, D. G. Past, present & future of information technology in pedometrics. **Geoderma**, v. 324, p. 131–137, ago. 2018. ISSN 00167061. Acesso: 20 janeiro 2021. <<https://linkinghub.elsevier.com/retrieve/pii/S0016706117318724>>.
- ROTH, C.H.; MEYER, B.; FREDE, H.G. A portable rainfall simulator for studying factors affecting runoff, infiltration and soil loss. **Catena**, 12:79-85, 1985.
- ROUSE, J.W.; HAAS, R.H.; SCHELL, J.A.; DEERING, D.W. Monitoring Vegetation Systems in the Great Plains with ERTS. **Third ERTS Symposium**, NASA special publication, NASA, v. 351, p. 309, 1974.
- ROWAN, L. C.; MARS, J. C. Lithologic mapping in the Mountain Pass, California area using Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) data. **Journal of Remote Sensing of Environment**, v. 84, p. 350-366, 2003.
- RSTUDIO TEAM. RStudio: Integrated Development Environment for R. **RStudio**, PBC, Boston, MA, 2020. Disponível em: <<http://www.rstudio.com/>>. Acesso em: 26 Abril 2021.
- SAMMUT, C.; WEBB, G. I. Encyclopedia of Machine Learning: Leave-One-Out Cross-Validation, **Springer Online Link**, Boston, Massachusetts, 2010. doi: https://doi.org/10.1007/978-0-387-30164-8_469.

SAMPAIO, F. D. M. T.; MENEZES, S. M.; FURTINI, M. B.; RIBEIRO, K. D. Correlações entre a permeabilidade e os atributos físicos do solo da sub-bacia do córrego centenário da cidade de Lavras-MG. **Ciência e Agrotecnologia**, v. 30, n. 4, p. 798-803, 2006.

SANTOS, M.; CATEN, A. ten. Mapeamento Digital de Solos (MDS): avanços e desafios. **Embrapa Solos-Artigo em periódico indexado (ALICE)**, 2015.

SANTOS, P. A.; PINHEIRO, H. S. K.; SILVA, I. L. Análise de Produtos Oriundos de MDE para Compreensão dos Recursos Hídricos na Bacia Hidrográfica do Rio Guapi-Macacu, RJ: Um Estudo em Ambiente SIG. **Anais do II Congresso Alagoano de Engenharia de Agrimensura - CONEAGRI**, Rio Largo, Alagoas, 2019. Publicado em 2021 pela Editora da Universidade Federal de Alagoas, EDUFAL.

SANTOS, P.A; PINHEIRO, H. S. K.; JÚNIOR, W. C.; PEREIRA, N. R. Aplicação de ferramentas SIG nas análises geométrica e morfométricas para caracterização hidrológica das bacias hidrográficas do Rio Guapi-Macacu, RJ. **Anais da V Jornada de geotecnologias do Estado do Rio de Janeiro (JGEOTEC 2020)**, Universidade Federal Fluminense, pp.1018-1021, 2020. ISBN: 978-65-00-18327-6.

SCHAAP, M.G. Accuracy and uncertainty in PTF predictions. **Developments in soil science**, v. 30, p. 33-43, 2004.

SCHAAP, M. G.; LEIJ, F. J. Database-related accuracy and uncertainty of pedotransfer functions. **Soil Science**, v. 163, n. 10, p. 765-779, 1998.

SCHAAP, M. G.; LEIJ, F. J. Using neural networks to predict soil water retention and soil hydraulic conductivity. **Soil Till. Res.**, 47:37-42, 1998.

SCHAAP, M. G.; LEIJ, F. J.; VAN GENUCHTEN, M. T. Rosetta: A computer program for estimating soil hydraulic parameters with hierarchical pedotransfer functions. **Journal of hydrology**, v. 251, n. 3-4, p. 163-176, 2001.

SCHEINOST, A. C.; SINOWSKI, W.; AUERSWALD, K. Regionalization of soil water retention curves in a highly variable soilscape, I. Developing a new pedotransfer function. **Geoderma**, v. 78, n. 3, p. 129-143, 1997.

SEGAL, D. Theoretical basis for differentiation of ferric-iron bearing minerals, using landsat MSS data. **In Proceedings of Symposium for Remote Sensing of Environment**, 2nd Thematic Conference on Remote Sensing for Exploratory Geology, Fort Worth, TX, pp. 949-951, 1982.

SEIBERT, J.; MCGLYNN, B. A new triangular multiple flow direction algorithm for computing upslope areas from gridded digital elevation models. **Water Resources Research**, Vol. 43, W04501, C++ implementation in SAGA by Thomas Grabs (c). Wiley, 2007.

SHAPIRO, S. S.; WILK, M. B. An analysis of variance test for normality (complete samples), **Biometrika**, vol. 52, no 3-4, pp. 591-611, 1965. doi:10.1093/biomet/52.3-4.591.

SILVA, B. K. Funções de pedotransferência para predição da condutividade hidráulica saturada de solos do Brasil. **Dissertação de Mestrado**. Programa de Pós-Graduação em Ciência do solo, Universidade Federal do Paraná, Curitiba, 2018.

SINERGISE. Sentinel-Hub Repository Satellite Indices. Index database for Sentinel-2 Satellite. **Sítio online**. Disponível online: <https://custom-scripts.sentinel-hub.com/custom-scripts/sentinel-2/indexdb/> (Acesso em 23 dezembro 2020).

SNEDECOR, G. W. Calculation and interpretation of analysis of variance and covariance, Iowa State College Division of Industrial Science monographs, **Collegiate Press**, vol.1, 1934. <https://doi.org/10.1037/13308-000>.

SOILMOISTURE. Operating Instructions: Guelph Permeameter 2800. **SoilMoisture Equipment Corp**. S. Kellogg Ave., n. 801, Goleta, California, P.O. Box 30025, Santa Barbara, Califônia, USA, 2012. Disponível em:

- <https://www.soilmoisture.com/pdfs/Resource_Instructions_0898-2800_2800K1%20Guelph%20Permeameter%20.pdf>. Acesso: 18 Março 2021.
- SOUZA, E. D.; FERNANDES FILHO, E. I.; SCHAEFER, C. E. G. R.; BATJES, N. H.; SANTOS, G. R. D.; PONTES, L. M. Pedotransfer functions to estimate bulk density from soil properties and environmental covariates: Rio Doce basin. **Scientia Agricola**, v. 73, n. 6, p. 525-534, 2016.
- SOUZA, P. G. C.; JÚNIOR, J. C. F. B.; VIANA, J. H. M. Curva de Retenção e Disponibilidade Total de Água em Latossolos Empregando Funções de Pedotransferência. **Revista Engenharia na Agricultura**, v. 25, n. 5, p. 417-427, 2017.
- SPEARMAN, C. The Proof and Measurement of Association between Two Things. **The American Journal of Psychology**, v.15, n. 1, p.72–101, 1904. doi:10.2307/1412159.
- STEPINSKI, T.F.; JASIEWICZ, J. Geomorphons — a pattern recognition approach to classification and mapping of landforms. In: Hengl, T. / Evans, I.S. / Wilson, J.P. / Gould, M. [Eds.]: **Proceedings of Geomorphometry 2011**, Redlands, 109-112. pdf at geomorphometry.org, 2011.
- SVETNIK, V.; LIAW, A.; TONG, C.; WANG, T. Application of Breiman's random forest to modeling structure-activity relationships of pharmaceutical molecules. **In: International Workshop on Multiple Classifier Systems**. Springer, Berlin, Heidelberg, p.334-343,2004.
- TAGHIZADEH-MEHRJARDI, R.; NABIOLLAHI, K.; MINASNY, B.; TRIANTAFILIS, J. Comparing data mining classifiers to predict spatial distribution of USDA-family soil groups in Baneh region, Iran. **Geoderma**, v. 253, p. 67-77, 2015.
- TARINI, M.; CIGNONI, P.; MONTANI, C. Ambient Occlusion and Edge Cueing to Enhance Real Time Molecular Visualization. **IEEE Transactions on Visualization and Computer Graphics**, Vol. 12, No. 5, pp. 1237-1244, 2006.
- TOMASELLA, J.; PACHEPSKY, Y.; CRESTANA, S.; RAWLS, W.J. Comparison of two techniques to develop pedotransfer functions for water retention. **Soil Sci. Soc. Am. J.**, 67:1085-1092, 2003.
- TRANter, G.; MINASNY, B.; MCBRATNEY, A.B.; MURPHY, B.; MCKENZIE, N.J.; GRUNDY, M.; BROUGH, D.. Building and testing conceptual and empirical models for predicting soil bulk density. **Soil Use and Management**, v. 23, n. 4, p. 437-443, 2007.
- TWARAKAVI, N.K.C.; SIMUNEK, J.; SCHAAP, M.G. Development of Pedotransfer functions for estimation of soil hydraulic parameters using support vector machines. **Soil Sci. Soc. Am. J.**, 73:1443-1452, 2009.
- VAN DEN BERG, M. Available water capacity in strongly weathered soils of South East and Southern Brazil. In: Solo Suelo 96, Congresso LatinoAmericano de Ciencia do Solo, 8., Águas de Lindoia, 1996. **Proceedings... Águas de Lindoia**, 1996.
- VAN DER MEER, F. D.; VAN DER WERFF, H. M. A.; VAN RUITENBEEK, F. J. A. Potential of ESA's Sentinel-2 for geological applications. **Remote sensing of environment**, v. 148, p. 124-133, 2014.
- VAN GENUCHTEN, M.TH. A closed-form equation for predicting the hydraulic conductivity of unsaturated soils. **Soil Sci. Soc. Am. J.**, 44:892-898, 1980.
- VEREecken, H.; MAES, J.; FEYEN, J. Estimating unsaturated hydraulic conductivity from easily measured soil properties. **Soil Science**, v. 149, n. 1, p. 1-12, 1990.
- VEREecken, H.; HERBST, M. Statistical regression. In: PACHEPSKY, Y.A. & RAWLS, W.J., ed. **Development of pedotransfer functions in soil hydrology**. Amsterdam, Elsevier, 2004. p.415-429.

- WADOUX, A. M. C.; MINASNY, B.; MCBRATNEY, A. B. Machine learning for digital soil mapping: applications, challenges and suggested solutions, **Earth-Science Reviews**, vol. 210, pp. 1-17, 2020.
- WEISS, A. D. Topographic Position and Landforms Analysis. **Congress Poster**, 2002. Acesso: 23 dezembro 2020. <http://www.jennessent.com/downloads/tpi-poster-tnc_18x22.pdf>.
- WESSEL, M.; BRANDMEIER, M.; TIEDE, D. Evaluation of different machine learning algorithms for scalable classification of tree types and tree species based on Sentinel-2 data. **Remote Sensing**, v. 10, n. 9, p. 1419, 2018.
- WIESE, L.; ROS, I.; ROZANOV, A.; BOSHOFF, A.; CLERCQ, W. de; SEIFERT, T. An approach to soil carbon accounting and mapping using vertical distribution functions for known soil types. **Geoderma**, v.263, p.264-273, 2016. DOI: 10.1016/j. geoderma.2015.07.012.
- WILDING, L.P.; DREES, L.R. Spatial variability and pedology. In: WILDING, L.P.; SMECK, N.E.; HALL, G.F. (Eds). *Pedogenesis and soil taxonomy: concepts and interactions*. New York: Elsevier, p.83-116, 1983.
- WILSON, J. P.; GALLANT, J. C. Primary Topographic Attributes. In: Wilson, J.P. & Gallant, J.C. [Eds.]: **Terrain Analysis: Principles and Applications**, John Wiley & Sons, p.51-85, 2000.
- WILSON, J. P.; GALLANT, J. C. *Terrain Analysis: Principles and Applications*. Nova Jersey: **Jonh Wiley & Sons**, Inc. jan./2000. ISBN 0-471-32188-5, 2000.
- WOOD, J. The Geomorphological characterisation of Digital Elevation Models. Diss., Department of Geography, **University of Leicester**, U.K, online, 1996.
- WOOD, J. Geomorphometry in LandSerf. In: Hengl, T. and Reuter, H.I. [Eds.]: **Geomorphometry: Concepts, Software, Applications**. Developments in Soil Science, Elsevier, Vol.33, 333-349, 2009.
- WÖSTEN, J. H. M.; LILLY, A., NEMES, A.; LE BAS, C. Development and use of a database of hydraulic properties of European soils. **Geoderma**, v. 90, n. 3-4, p. 169-185, 1999.
- WÖSTEN, J. H. M.; PACHEPSKY, Ya A.; RAWLS, W. J. Pedotransfer functions: bridging the gap between available basic soil data and missing soil hydraulic characteristics. **Journal of hydrology**, v. 251, n. 3-4, p. 123-150, 2001.
- XAVIER, P.; DOS ANJOS, L. H. C.; PINHEIRO, H.; CHAGAS, C. D. S.; CARVALHO JUNIOR, W. D. Usage of pedometrics for data evaluation and harmonization in soil profiles from Cerrado region, Mato Grosso do Sul. **World Congress of Soil Science**, Proceedings of the 21st WCSS, vol. 2, pp. 75, 2019.
- XIAO, J.; SHEN, Y.; TATEISHI, R.; BAYAER, W. Development of topsoil grain size index for monitoring desertification in arid land using remote sensing. **International Journal of Remote Sensing**, v. 27, n. 12, p. 2411-2422, 2006.
- YOKOYAMA, R.; SHIRASAWA, M.; PIKE, R. J. Visualizing topography by openness: A new application of image processing to digital elevation models. **Photogrammetric Engineering and Remote Sensing**, Vol.68, pp.251-266 pdf at ASPRS, 2002.
- ZEVENBERGEN, L.W., THORNE, C.R. Quantitative analysis of land surface topography. **Earth Surface Processes and Landforms**, 12: 47-56, 1987.
- ZHANG, T.; SU, J.; LIU, C.; CHEN, W.H.; LIU, H.; LIU, G. Band selection in Sentinel-2 satellite for agriculture applications. In **23rd International Conference on Automation and Computing (ICAC)**. IEEE, pp. 1-6, 2017.
- ZINN, Y.L.; LAL, R.; RESCK, D.V.S. Texture and organic carbon relations described by a profile pedotransfer function for Brazilian Cerrado soils. **Geoderma**, v.127, p.168-173, 2005. DOI: 10.1016/j. geoderma.2005.02.010.

ANEXOS**A – Planilha de Parâmetros e Cálculos (Permeâmetro de Guelph)**



Green cells are inputs.
Yellow cells are results

In Double Method:

H2 must always be greater than H1.

R2 must always be greater than R1.

Measurements where R1 is equal or greater than R2 are invalid (physically unrealistic). Such results must be omitted from the calculations.

If invalid (negative) values of α are obtained (either K_{fs} or Φ_m will also be negative in value) or when α values are obtained that lie outside the realistic range of $0.01 \leq \alpha \leq 0.5 \text{ cm}^{-1}$, then the Single-Head Analysis should be applied to each of the two heads and the resulting values for K_{fs} and Φ_m averaged (use K value in "Average" area).

This spreadsheet is for private use only and the creator is not responsible for correctness and accuracy of the calculations. For detailed instructions please refer to Guelph Permeameter Operating Instructions.

Last Update: 11/26/2012

Support: ali@soilmoisture.com (Ali Farsad)

SOILMOISTURE Guelph Permeameter Calculations

Input
Result

Single Head Method (1)

Reservoir Cross-sectional area in cm²
(enter "35.22" for Combined and "2.16" for Inner reservoir):

Enter water Head Height ("H" in cm):

Enter the Borehole Radius ("a" in cm):

Enter the soil texture-structure category (enter one of the below numbers):

1. Compacted, Structure-less, clayey or silty materials such as landfill caps and liners, lacustrine or marine sediments, etc.
2. Soils which are both fine textured (clayey or silty) and unstructured; may also include some fine sands.
3. Most structured soils from clays through loams; also includes unstructured medium and fine sands. The category most frequently applicable for agricultural soils.
4. Coarse and gravelly sands; may also include some highly structured soils with large and/or numerous cracks, macropores, etc.

Steady State Rate of Water Level Change ("R" in cm/min):

Res Type:

H:

a:

H/a:

a*:

C0.01:

C0.04:

C0.12:

C0.36:

C:

R:

Q:

pi: 3.142

$\alpha^* = 0 \text{ cm}^{-1}$

C = 0

Q = 0

$K_{fs} = \text{#DIV/0! cm/sec}$

$\Phi_m = \text{#DIV/0! cm}^2/\text{min}$

Single Head Method (2)

Reservoir Cross-sectional area in cm²
(enter "35.22" for Combined and "2.16" for Inner reservoir):

Enter water Head Height ("H" in cm):

Enter the Borehole Radius ("a" in cm):

Enter the soil texture-structure category (enter one of the below numbers):

1. Compacted, Structure-less, clayey or silty materials such as landfill caps and liners, lacustrine or marine sediments, etc.
2. Soils which are both fine textured (clayey or silty) and unstructured; may also include some fine sands.
3. Most structured soils from clays through loams; also includes unstructured medium and fine sands. The category most frequently applicable for agricultural soils.
4. Coarse and gravelly sands; may also include some highly structured soils with large and/or numerous cracks, macropores, etc.

Steady State Rate of Water Level Change ("R" in cm/min):

Res Type:

H:

a:

H/a:

a*:

C0.01:

C0.04:

C0.12:

C0.36:

C:

R:

Q:

pi: 3.1415

$\alpha^* = 0 \text{ cm}^{-1}$

C = 0

Q = 0

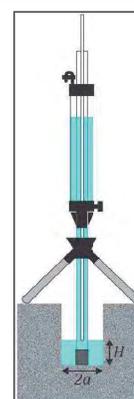
$K_{fs} = \text{#DIV/0! cm/sec}$

$\Phi_m = \text{#DIV/0! cm}^2/\text{min}$

Average

$K_{fs} = \text{#DIV/0! cm/sec}$

$\Phi_m = \text{#DIV/0! cm}^2/\text{min}$



Double Head Method

Reservoir Cross-sectional area in cm²
(enter "35.22" for Combined and "2.16" for Inner reservoir):

Enter the first water Head Height ("H1" in cm):

Enter the second water Head Height ("H2" in cm):

Enter the Borehole Radius ("a" in cm):

Enter the soil texture-structure category (enter one of the below numbers):

1. Compacted, Structure-less, clayey or silty materials such as landfill caps and liners, lacustrine or marine sediments, etc.
2. Soils which are both fine textured (clayey or silty) and unstructured; may also include some fine sands.
3. Most structured soils from clays through loams; also includes unstructured medium and fine sands. The category most frequently applicable for agricultural soils.
4. Coarse and gravelly sands; may also include some highly structured soils with large and/or numerous cracks, macropores, etc.

Steady State Rate of Water Level Change ("R1" in cm/min):

Steady State Rate of Water Level Change ("R2" in cm/min):

Res Type:

H1/a:

H2/a:

C1-0.01:

C2-0.01:

C1-0.04:

C2-0.04:

C1-0.12:

C2-0.12:

C1-0.36:

C2-0.36:

$\alpha^* = 0 \text{ cm}^{-1}$

$\alpha = \text{#DIV/0!}$

Q1 = 0

Q2 = 0

C1 = 0

C2 = 0

G1 = #DIV/0!

G2 = #DIV/0!

G3 = #DIV/0!

G4 = #DIV/0!

$K_{fs} = \text{#DIV/0! cm/sec}$

$\Phi_m = \text{#DIV/0! cm}^2/\text{min}$

$\Theta_{fs} = \text{#DIV/0! cm}^3/\text{cm}^3$

$\Theta_s = \text{#DIV/0! cm}^3/\text{cm}^3$

Sorptivity: (cm min^{-0.5})

Calculation formulas related to shape factors (C). Where H₁ is the first water head height (cm), H₂ is the second water head height (cm), a is borehole radius (cm) and a* is microscopic capillary length factor which is decided according to the soil texture-structure category. For one-head method, only C₁ needs to be calculated while for two-head method, C₁ and C₂ are calculated (Zang et al., 1998).

Soil Texture-Structure Category	$\alpha^*(\text{cm}^{-1})$	Shape Factor
Compacted, Structure-less, clayey or silty materials such as landfill caps and liners, lacustrine or marine sediments, etc.	0.01	$C_1 = \left(\frac{H_2/a}{2.081 + 0.121(H_2/a)} \right)^{0.672}$
Soils which are both fine textured (clayey or silty) and unstructured, may also include some fine sands.	0.04	$C_1 = \left(\frac{H_1/a}{1.992 + 0.091(H_1/a)} \right)^{0.683}$ $C_2 = \left(\frac{H_2/a}{1.992 + 0.091(H_2/a)} \right)^{0.683}$
Most structured soils from clays through loams; also includes unstructured medium and fine sands. The category most frequently applicable for agricultural soils.	0.12	$C_1 = \left(\frac{H_1/a}{2.074 + 0.093(H_1/a)} \right)^{0.754}$ $C_2 = \left(\frac{H_2/a}{2.074 + 0.093(H_2/a)} \right)^{0.754}$
Coarse and gravelly sands; may also include some highly structured soils with large and/or numerous cracks, macro pores, etc.	0.36	$C_1 = \left(\frac{H_1/a}{2.074 + 0.093(H_1/a)} \right)^{0.754}$ $C_2 = \left(\frac{H_2/a}{2.074 + 0.093(H_2/a)} \right)^{0.754}$

Calculation formulas related to one-head and two-head methods. Where R is steady-state rate of fall of water in reservoir (cm/s), K_{fs} is Soil saturated hydraulic conductivity (cm/s), Φ_m is Soil matric flux potential (cm²/s), a* is Macroscopic capillary length parameter (from Table 2), a is Borehole radius (cm), H₁ is the first head of water established in borehole (cm), H₂ is the second head of water established in borehole (cm) and C is Shape factor (from Table 2).

One Head, Combined Reservoir	$Q_1 = \bar{R}_1 \times 35.22$	$K_{fs} = \frac{C_1 \times Q_1}{2\pi H_1^2 + \pi a^2 C_1 + 2\pi \left(\frac{H_1}{a^2} \right)}$
One Head, Inner Reservoir	$Q_1 = \bar{R}_1 \times 2.15$	$\Phi_m = \frac{C_1 \times Q_1}{(2\pi H_1^2 + \pi a^2 C_1)a^* + 2\pi H_1}$
Two Head, Combined Reservoir	$Q_1 = \bar{R}_1 \times 35.22$ $Q_2 = \bar{R}_2 \times 35.22$	$G_1 = \frac{H_2 C_1}{\pi(2H_1 H_2 (H_2 - H_1) + a^2(H_1 C_2 - H_2 C_1))}$ $G_2 = \frac{H_1 C_2}{\pi(2H_1 H_2 (H_2 - H_1) + a^2(H_1 C_2 - H_2 C_1))}$ $K_{fs} = G_2 Q_2 - G_1 Q_1$ $G_3 = \frac{(2H_1^2 + a^2 C_1) C_1}{2\pi(2H_1 H_2 (H_2 - H_1) + a^2(H_1 C_2 - H_2 C_1))}$
Two Head, Inner Reservoir	$Q_1 = \bar{R}_1 \times 2.15$ $Q_2 = \bar{R}_2 \times 2.15$	$G_4 = \frac{(2H_1^2 + a^2 C_1) C_2}{2\pi(2H_1 H_2 (H_2 - H_1) + a^2(H_1 C_2 - H_2 C_1))}$ $\Phi_m = G_3 Q_1 - G_4 Q_2$

B – Quadros do Capítulo II

Quadro 6. Covariáveis do terreno relativas à atributos topográficos físico-hídricos e temodinâmicos primários e secundários do MDE.

Tipo	Atributo do Terreno	Processo físico	Descrição
Pr	<i>Digital Elevation Model (DEM) ou Elevation (elev)</i>	M	Dinâmica da paisagem e seus fenômenos.
P	<i>Slope, Aspect, Analytical HillShade (AH), Profile Curvature (ProfC); Planar Curvature (PlanC)</i>		Dinâmica dos fluxos de água; Vegetação; Geomorfologia; Intensidade da radiação solar; Falhamentos de estruturas geológicas; deposição de sedimentos e partículas do solo.
S	<i>Tangential Curvature (TanC), Total Curvature (TotalC), Morphometric Features (MF), Topographic Position Index (TPI), Wind Expositon Index (WEI), Terrain Surface Convexity (TSC), Longitudinal Curvature (LonC), General Curvature (GC), LandForms (LF), Terrain Surface Texture (TST), Convergence Index (CI), Generalized Surface (GS), Morphometric Protection Index (MPI)</i>		Taxa de acumulação lateral na paisagem; Retenção e armazenamento de água no solo; características do solo; Abordagem para derivação de parâmetros morfométricas da paisagem.
S	<i>Catchment Area (CA), Catchment Slope (CS), Multiresolution Index of Valley Bottom Flatness (MRVBF), Multiresolution Index of The Ridge Top Flatness (MRRTF), Terrain Classification Index for Lowlands (TCIL), Flow Accumulation (FA), Stream Power Index (SPI), Topographic Wetness Index (TWI), SAGA Wetness Index (SWI), Valley Depth (VD), Vertical Distance to Channel Network (VDCN), Channel Network Base Level (CNBL), Flow Path Length (FPL), Euclidian Distance (ED)</i>	CH	Identificação e caracterização de depósitos de sedimentos; Previsão de zonas de saturação; Estudos hidrológicos de fluxo de água e escoamento; Concentração hídrica e canais de conectividade.
S	<i>LS Factor (LSF), Wetness Index (WI)</i>	E	Estimativa da erosão de bacias hidrográficas; Transporte de sedimentos; Permeabilidade da rocha.
S	<i>Geomorphons (Gm)</i>	VI	Classificação da paisagem em pedoformas; Descrição da geomorfologia da região;

Siglas: Pr= Principal; P= Primário; S= Secundário; M= Morfometria; CH= Canais e Hidrologia; E= Estabilidade; VI= Visibilidade e Iluminação.

Fonte: adaptado de Beven & Kirkby (1979), Zevenbergen & Thorne (1987), Moore et al. (1991), Guisan et al. (1999), Montgomery & Dietrich (1994), Koethe & Lehmeier (1996), Wood (1996;2009), Wilson & Gallant (2000), Böhner et al. (2002), Weiss (2000), Wilson & Gallant (2000), Florinsky et al. (2002), Yokoyama et al. (2002), Gallant & Dowling (2003), Romano e Chirico (2004), Böhner & Selige (2006), Tarini et al. (2006), Bock et al. (2007), Iwarashi & Pike (2007), Seibert & McGlynn (2007), Bohner (2009), Stepinski & Jasiewicz (2011), Jasiewicz & Stepinski (2013), Gerlitz et al. (2015), Oliveira et al. (2017).

Quadro 7. Características do Projeto Aerogeofísico do Estado do Rio de Janeiro.

Item	Descrição
Código do Projeto	1117 (Rio de Janeiro)
Direção das linhas de vôo	N-S
Espaçamento entre as linhas de vôo	0,5 km
Direção das linhas de controle	E-W
Espaçamento entre as linhas de controle	10 km
Intervalo entre medições geofísicas consecutivas	0,1 s (magnetômetro) e 1,0 s (gamaespectrômetro)
Altura média de vôo	100 ± 15 m
Velocidade aproximada de vôo	270 km/h
Área de recobrimento	66.111,40 km
Sensores Aeromagnetômetros	Geometrics G822A de montagem Stinger com resolução de 0,001 nT, faixa de 20.000 a 95.000 nT.
Sensores Aerogamaespectrômetros	Radiations Solutions modelos RS-500 com 1024 canais espectrais de entrada, reduzidos para 256 canais espectrais de saída.

Fonte: adaptado de Relatório Final do Levantamento e processamento dos dados magnetométricos e gamaespectométricos, CPRM (2012).

Quadro 8. Características das Aeronaves e do processo de aquisição de dados do Projeto Aerogeofísico do Estado do Rio de Janeiro.

Item	Descrição
Aeronave Piper Navajo (PR-PEC)	Período: 13/08/2011 a 26/05/2012; Quantidade de linhas de vôo: 41.743,30 km; Base: Maricá
Aeronave Cessna Caravan (PT-MED)	Período: 13/08/2011 a 03/12/2011; Quantidade de linhas de vôo: 24.368,10 km; Base: Cabo Frio
Sistema de navegação e aquisição GPS	AG-NAV P151 GUIA "Real time" e TEEJET RX 410-P "Real time"
Radar Altímetro e Altímetro Barométrico	Free-Flight Systems TRA-3000; Honeywell PPT
Magnetômetro Terrestre a Vapor de Césio	GEM GSM-19
Sistema de Aquisição	RMS DAARC500 e DATA LOGGER IEI
Software de processamento dos Dados	Oasis Montaj versão 9.0 da empresa Geosoft
Aquisição dos dados do Projeto	Livre, no Portal GeoSGB, CPRM.

Fonte: adaptado de Relatório Final do Levantamento e processamento dos dados magnetométricos e gamaespectométricos, CPRM (2012).

Quadro 9. Características espectrais do sensor MSI da missão Sentinel-2A.

Número da Banda	Região Espectral	Comprimento de onda central (nm)	Largura da Banda (nm)	Resolução Espacial (m)
B1 ¹	<i>Coastal Aerosol</i>	442.7	21	60
B2	<i>Blue (B)</i>	492.4	66	10
B3	<i>Green (G)</i>	559.8	36	10
B4	<i>Red (R)</i>	664.6	31	10
B5	<i>Red-edge 1 (Re1)</i>	704.1	15	20
B6	<i>Red-edge 2 (Re2)</i>	740.5	15	20
B7	<i>Red-edge 3 (Re3)</i>	782.8	20	20
B8 ²	<i>Near Infrared (NIR)</i>	832.8	106	10
B8A	<i>Near Infrared narrow (NIRn) or Vegetation reledge</i>	864.7	21	20
B9 ¹	<i>Water Vapor</i>	945.1	20	60
B10 ¹	<i>Shortwave infrared (SWIR)/Cirrus</i>	1373.5	31	60
B11	<i>Shortwave infrared 1 (SWIR 1)</i>	1613.7	91	20
B12	<i>Shortwave infrared 2 (SWIR 2)</i>	2202.4	175	20

¹ Bandas de correção usadas apenas nas etapas de processamento.

² Banda não usada nas análises pois, ao reamostrar as imagens para 20 m, ela é omitida. Segundo a ESA (2020), produtos Level 2 corrigidos para reflectância de superfície e reamostrados para 20 m fornecem informação espectral mais precisa na região do infravermelho próximo ao se usar a banda B8A. O MSI mede a radiância refletida da Terra em 13 bandas espectrais, do VIS-NIR ao SWIR. Nota: A largura de banda (nm) é medida na metade da máxima largura total (do inglês, *Full Width Half Maximum - FWHM*).

Fonte: Adaptado de *MultiSpectral Instrument (MSI) Overview*, ESA (2020).

Quadro 10. Índices de vegetação derivados das bandas espectrais do sensor Multiespectral da missão Sentinel-2.

Índice	Fórmula dos Índices baseada nas bandas do sensor MSI Sentinel-2	Referências
<i>Enhanced Vegetation Index (EVI)</i>	$2.5 * \left[\frac{(\rho_{banda\ 8A} - \rho_{banda\ 4})}{(\rho_{banda\ 8A} + 6 * \rho_{banda\ 4} - 7.5 * \rho_{banda\ 2} + 1)} \right]$	Huete et al. (2002); Hunt et al. (2011)
<i>Clay Minerals Ratio (CM); Ferrous Minerals Ratio (FM); Ferruginous Regolith Ratio (FR); Iron Oxide Ratio (IO)</i>	$\frac{\rho_{banda\ 11}}{\rho_{banda\ 12}} ; \frac{\rho_{banda\ 11}}{\rho_{banda\ 8A}} ; \frac{\rho_{banda\ 8A}}{\rho_{banda\ 3}} ; \frac{\rho_{banda\ 4}}{\rho_{banda\ 2}}$	Drury (1987); Segal (1982); Rowan et al. (2003)
<i>Grain Size Index (GSI)</i>	$\frac{(\rho_{banda\ 4} - \rho_{banda\ 2})}{(\rho_{banda\ 4} + \rho_{banda\ 2} + \rho_{banda\ 3})}$	Perera et al. (2005); Xiao et al. (2006)
<i>Normalized Difference Red-Edge Index (NDRE)</i>	$\frac{(\rho_{banda\ 8A} - \rho_{banda\ 5})}{(\rho_{banda\ 8A} + \rho_{banda\ 5})}$	Cleavers et al. (2012); Férrandez-Manso et al. (2016)
<i>Normalized Difference Vegetation Index (NDVI)</i>	$\frac{(\rho_{banda\ 8A} - \rho_{banda\ 4})}{(\rho_{banda\ 8A} + \rho_{banda\ 4})}$	Rouse et al. (1973)
<i>Normalized Difference Water Index (NDWI)</i>	$\frac{(\rho_{banda\ 3} - \rho_{banda\ 8A})}{(\rho_{banda\ 3} + \rho_{banda\ 8A})}$	Brenner et al. (2015)
<i>Non-Linear Index (NLI)</i>	$\frac{(\rho_{banda\ 8A}^2 - \rho_{banda\ 4})}{(\rho_{banda\ 8A}^2 + \rho_{banda\ 4})}$	Goel et al. (1994)
<i>Soil Adjusted Vegetation Index (SAVI)</i>	$1.5 * \left[\frac{(\rho_{banda\ 8A} - \rho_{banda\ 4})}{(\rho_{banda\ 8A} + \rho_{banda\ 4} + 0.5)} \right]$	Huete (1988)
<i>Transformed Difference Vegetation Index (TDVI)</i>	$1.5 * \left[\frac{(\rho_{banda\ 8A} - \rho_{banda\ 4})}{\sqrt{(\rho_{banda\ 8A}^2 + \rho_{banda\ 4} + 0.5)}} \right]$	Bannari et al. (2002)
<i>Visible Atmospherically Resistant Index (VARI)</i>	$\frac{(\rho_{banda\ 3} - \rho_{banda\ 4})}{(\rho_{banda\ 3} + \rho_{banda\ 4} - \rho_{banda\ 2})}$	Gitelson et al. (2001); Hunt et al. (2011)

Fonte: adaptado de *Sentinel-Hub Repository Satellite Indices* (Sinergise, 2020).

C - Cartas Geológicas Cedidas pelo DRM

LEGENDA

Quaternário

- Qc
- Qp
- Ql
- Qm

Sedimentos Litorâneos: areias quartzosas marinhas, de cordões litorâneos e cúspides de lagunas, por vezes associadas a depósitos de conchas calcárias, relativamente bem selecionadas, com tonalidades esbranquiçadas, amareladas (onde aparecem traços de minerais pesados) e acinzentadas escuras (onde os Minerais Pesados aparecem em concentrações exploráveis)

Sedimentos fluviais: areias grosseiras, médias e finas, mal selecionadas, apresentando quartzo, feldspato e máficos, grãos subangulares a angulares, geralmente com matriz argilosa a silícia exibindo cores creme-amarelada, acastanhada, esverdeada e acinzentada, de leques aluviais. Argila cinza a cinza-amarelada, geralmente micacea de planície de inundação. Areias esbranquiçadas, com granulação variando de fina a conglomerática, composta principalmente de grãos de quartzo, podendo também apresentar feldspato, mica e máficos.

Sedimentos Paludiais: argilas de coloração cinza e depósitos de turfa de ambientes de lagos, pântanos ou brejos. Argilas de coloração negra, bioturbadas de mangues.

Depósitos de flúvio-marinhos

Terceário

- Tb
- Tybr
- Tybrf

Formação Barreiras: sedimentos constituídos por depósitos continentais, representados por níveis descontínuos alternados de material frívulo e mal selecionado, desde arenosos, areno-argilosos e argilosos, constituídos principalmente por grãos de quartzo

Microbrecha magmática de matriz fonolítica, cor cinza escura.

Mesozóico

- KB
- Ky

Diques básicos menores

Nefelina sienito, anfibólio sienitos, granulação grosseira, cor cinza-clara e, subordinadamente, micronefelina sienitos, microsienitos, fonólitos e traquitos, podendo conter pseudoleucita ou não.

Pré Cambiano

- Grad
- pCT
- pCqdg
- pCIcs
- pCIlse
- pCIIsf
- pCVsa

GRANITO ANDORINHA: biotita granito porfirítico, cor clara, matriz de grão médio com porfíros centimétricos de microclina, formando pequenos corpos irregulares. Presença de xenólitos digeridos ou não. Contatos intrusivos com a Unidade Santo Eduardo.

COMPLEXO BÁSICO DE GLEBA RIBEIRA: associação de olivina gabros, noritos e melanoritos grosseiros, com variações texturais para tipos de granulação mais fina e tipos gnáissificados. Localmente apresentando estrutura magmática (brecha magmática).

Quartzo dioritos e subordinadamente granodioritos, gnáissicos, homogêneos, granulação grosseira, localmente com textura ígnea preservada.

Unidade Cassoritiba: compreende (granada) - biotita - plagioclásio - K-feldspato gnáissicos homogêneos, predominantemente porfiroblásticos, com estruturas nebulítica e subordinadamente "schlieren". Matriz quartzodiorítica a granodiorítica, de granulação média.

Unidade Santo Eduardo: gnáissos listrados-lenticulares, com trama cataclástica blastomilonítica, mesocráticas, com intercalações discretas de anfibólito. O conjunto associa-se intimamente com rochas granulíticas foliadas, leuco a melanocráticas, com estrutura laminada (cataclástica) marcada pela presença de discóides de quartzo.

UNIDADE SÃO FIDELIS = sil-cord-(mo)-gd-bi-or-(mi)-pg-gnáissos e migmatitos, de coloração variável em tons de cinza, granulação variável, com estruturas flebítica e estromática; associados e leptinitos e a rochas calcossilicadas.

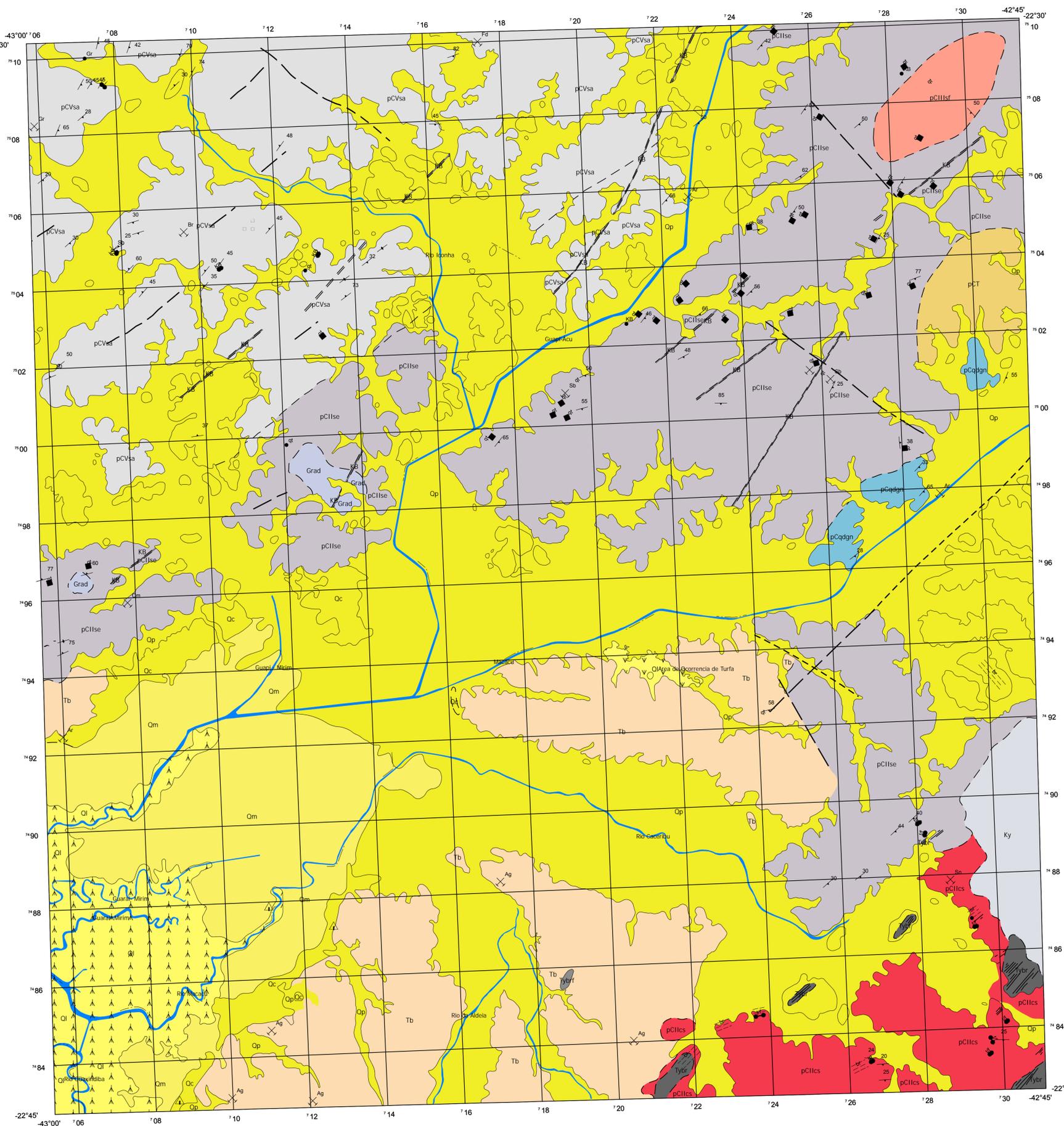
Unidade Santo Aleixo: diatexitos heterogêneos e homogêneos, representados predominantemente por (gd)-(hb)-bi-gnáissos graníticos a granodioríticos, cor cinza clara, granulação média a fina, excepcionalmente grosseira, com estruturas "schlieren", nebulítica e "schollen", podendo exibir texturas semiporfiroblástica e, excepcionalmente, porfiroblástica (p). Domínios heterogêneos com melanossoma constituído de (mi)-(hb)-bi-pg-gnáissos, bi-hb-pg-gnáissos e leucossoma de gnáissos quartzo-feldspático graníticos a pegmatóides. Domínios diatexiticos homogêneos, graníticos, leucocráticos. Contatos transicionais para Unidade Santo Eduardo.

ESTRUTURAS

- Contato Aproximado
- Contato Definido
- Contato Transicional
- Dique de Biotita Granito
- Dique de Diabásio
- Diques não Mapeáveis
- Sítio Arqueológico
- Direção de Folição Vertical
- Direção e Mergulho de Folição
- Mina Paralisada
- Mina em Atividade
- Faixa Aproximada
- Faixa Encoberta
- Faixa Provável
- Lineamentos Estruturais
- Zona de Brechiacao/Fraturamento Silicificado

OCCORRÊNCIAS

- Argila
- Granito para Cantaria
- Areia
- Sienito para Cantaria
- Ocorrência Mineral
- Granito pra Cantaria
- Dique de Diabásio
- Quartzo
- Caulim
- Birlita
- Argila
- Feldspato
- Salbreiro
- Presença de Manguezal
- Cataclásida
- Porfiroclástico
- Área de Ocorrência de Turfa



BASE CARTOGRÁFICA: Carta do Brasil, Escala 1:50.000 preparada pela Fundação IBGE em convênio com FUJINAS-CENTRAIS ELÉTRICAS S.A. Levantamento aerofotográfico de 1969; reambulação 1972; Restituição 1973; triangulação espacial. Origem da quilometragem: Equador e Meridiano 45°W. Gr. Acrescidas as constantes 10.000 Km e 500 Km, respectivamente. Declinação magnética no centro da folha 18°27' e cresce 8' minutos anualmente. Projeção Universal de Transversa de Mercator.

Mapa executado para o DRM - RJ pela GEOMITEC LTDA, em cumprimento ao contrato termo 06/80 de 11/12/80

Equipe Executora:
Geologia: Thomas L. Brenner, André L. Ferrari, Edson F. Mello, William M. Nassar

Petrografia: Maria do Carmo, C. B. Junho

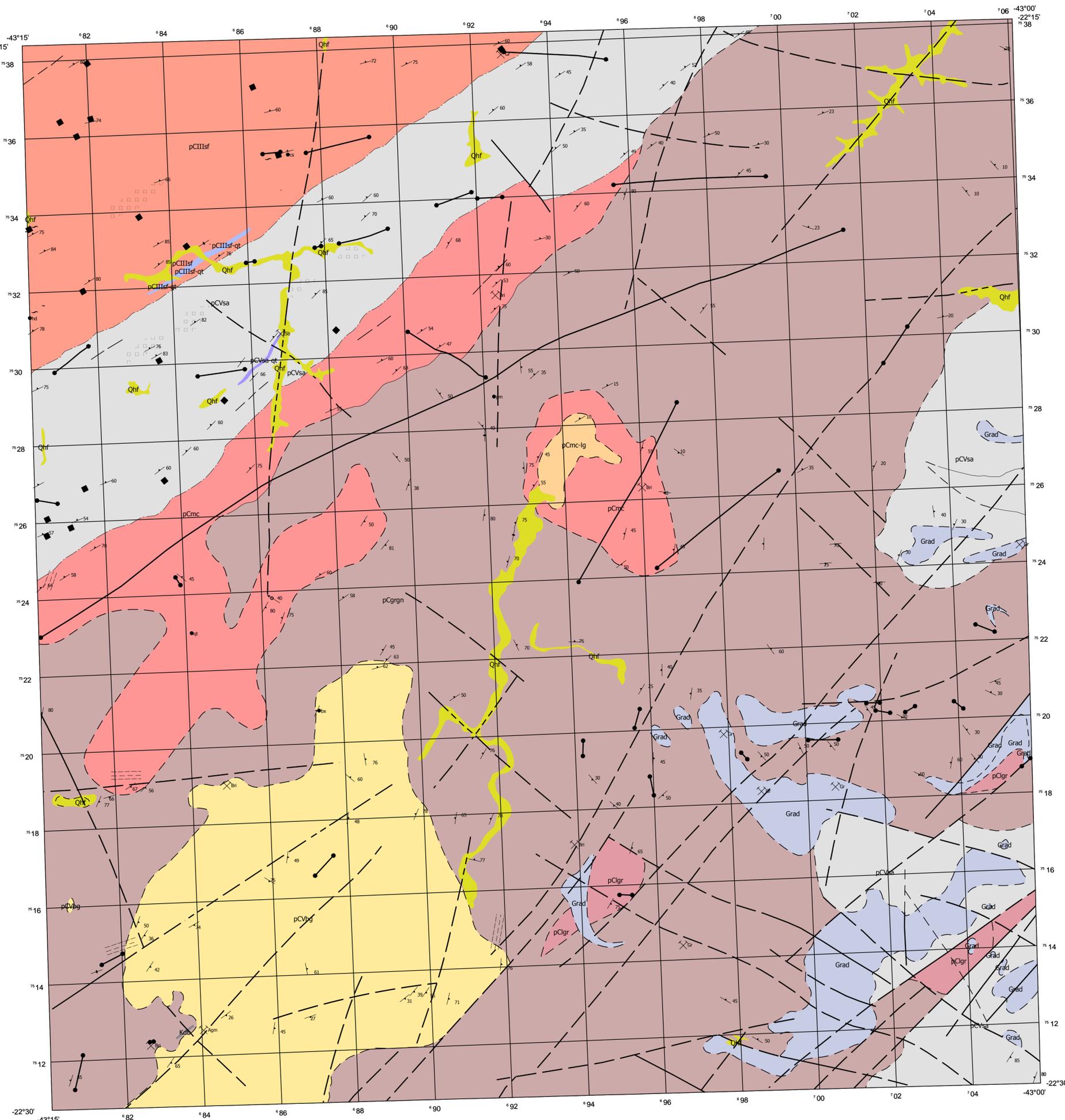
Consultoria: Helio Roncarati

Supervisão: Jaime S. A. Urdininea

Fiscalização DRM: Hernani R. Nunes

PROJETO CARTA GEOLÓGICA DO ESTADO DO RIO DE JANEIRO
FOLHA DE ITABORAÍ
SF-23-Z-B-V-1
ESCALA 1:50.000

Digitalização
Financiamento
Executor



LEGENDA

Quaternário

Qhf Depósitos de Várzea: areias feldspáticas estratificadas selecionadas e argilas silíticas orgânicas.

Mesozóico

Kdb Dique de Diabásio

Pré Cambiano

Grad GRANITO ANDORINHA: biotita granito porfírico, cor clara, matriz de grão médio com pórfiros centimétricos de microclina, formando pequenos corpos irregulares. Presença de xenólitos digeridos ou não. Contatos intrusivos com a Unidade Santo Eduardo.

pCmc Unidade Maria Comprida: migmatitos heterogêneos, bem diferenciados, frequentemente estromáticos, constituídos predominantemente de (hb) - bi-pg-mi gnaisses porfiroblásticos a pseudo-oftálmicos.

pCmc-lg Atravessados por granitos e leucogranitos gnaissicos, localmente mapeáveis (Lg), e subordinadamente biotita anfíbolitos. Contatos transicionais com a unidade Santo Aleixo.

pClgr Leucogranitos Gnaissicos, granulação grossiera a média, cor clara, gnaissificação moderada a ausente. Contatos intrusivos com demais litologias do batólito.

pCgrgn Compreende (gd) - (hb) - bi granitos a granodioritos gnaissicos, homogêneos, granulação grossiera a média, cor clara gnaissificação moderada. Presença restrita de enclaves xenolíticos migmatíticos - fácies local de granulação fina (f). Localmente injetados por veios leucograníticos gnaissicos com bandeamento de grão (ij). Contatos bruscos (intrusivos), com demais unidades exceto granito andorinha.

pCIIIsf UNIDADE SÃO FIDÉLIS = sill-cord-(mo)-gd-bi-or-(mi)-pg-gnaisses e migmatitos, de coloração variável em tons de cinza, granulação variável, compoendo estruturas felebitica e estromática; associados e leptinitos e a rochas calcossilicadas.

pCIIIsf-qt Gnaissio Associados a quartzitos (qt) e calcossilicatas, compoendo, estrutura estromática, localmente nebulítica.

pCVsa Unidade Santo Aleixo: diatexitos heterogêneos e homogêneos, representados dominantemente por (gd)-(hb)-bi-gnaisses graníticos a granodioríticos, cor cinza clara, granulação média a fina, excepcionalmente grossiera, com estruturas "schlieren", nebulítica e "schollen", podendo exibir texturas semiporfiroblástica e, excepcionalmente, porfiroblástica (p). Domínios heterogêneos com melanossoma constituído de (mi)-(hb)-bi-pg-gnaisses, bi-hb-pg-gnaisses e leucossoma de gnaisses quartzo-feldspático graníticos a pegmatídeos. Domínios diatexíticos homogêneos, graníticos, leucocráticos. Contatos transicionais para Unidade Santo Eduardo.

pCVsa-qt Intercalações quartzíticas. Localmente atravessados por leucogranitos gnaissicos. Contatos transicionais para as unidades Maria Comprida e São Fidélis.

pCVbg Unidade Bingen: compreende (granada) - (hornblenda) - biotita gnaisses graníticos, gnaissificação moderada ou ausente, granulação média a fina, cor clara, estrutura nebulítica a "schlieren". Domínios restritos migmatíticos, heterogêneos, não mapeáveis à escala. Migmatizado por injeção (mg). Contatos transicionais com os migmatitos heterogêneos da unidade santo Aleixo.

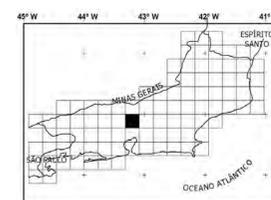
ESTRUTURAS

- Contato Aproximado
- Contato Definido
- - - Contato Encoberto
- · - Contato Transicional
- Dique de Diabásio
- Faixas não Mapeáveis
- ~ Cataclásada
- ↖ Rumo e Caimento de Eixo de Boudin
- ↑ Direção de Foliacao Vertical
- ↘ Direção e Mergulho de Foliacao
- ⊗ Mina Paralisada
- sb - Saibreira
- gr - Granito para Cantaria
- Falha Aproximada
- Falha Definida
- - - Falha Encoberta
- Lineamentos Estruturais
- ⊞ Zona de Brechiacao/Fratramento Silicificado
- ⊗ Mina em Atividade
- bri - Brita
- agm - Agua Mineral
- gr - Granito

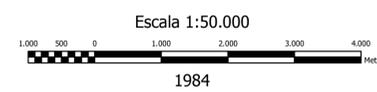
OCORRÊNCIAS

- Ocorrência Mineral
- bx - Biotita Xisto Melanossomatico
- cs - Calcossilicada
- hd - Hiperstenio Diorito
- pm - Protomilonito
- qt - Quartzito

LOCALIZAÇÃO NO ESTADO



ARTICULAÇÃO DA FOLHA



DIVISÃO MUNICIPAL



Mapeamento geológico executado pela UFRJ.
Geologia: Helio M. Penha
André Luiz Ferrari
Mário do Carmo B. Junho
Sergio Luiz A. de Souza
Thomas L. Bremer
Supervisão DRM
Geologia: Alfredo J. F. Puget
Cartografia: W. M. Sant'Anna

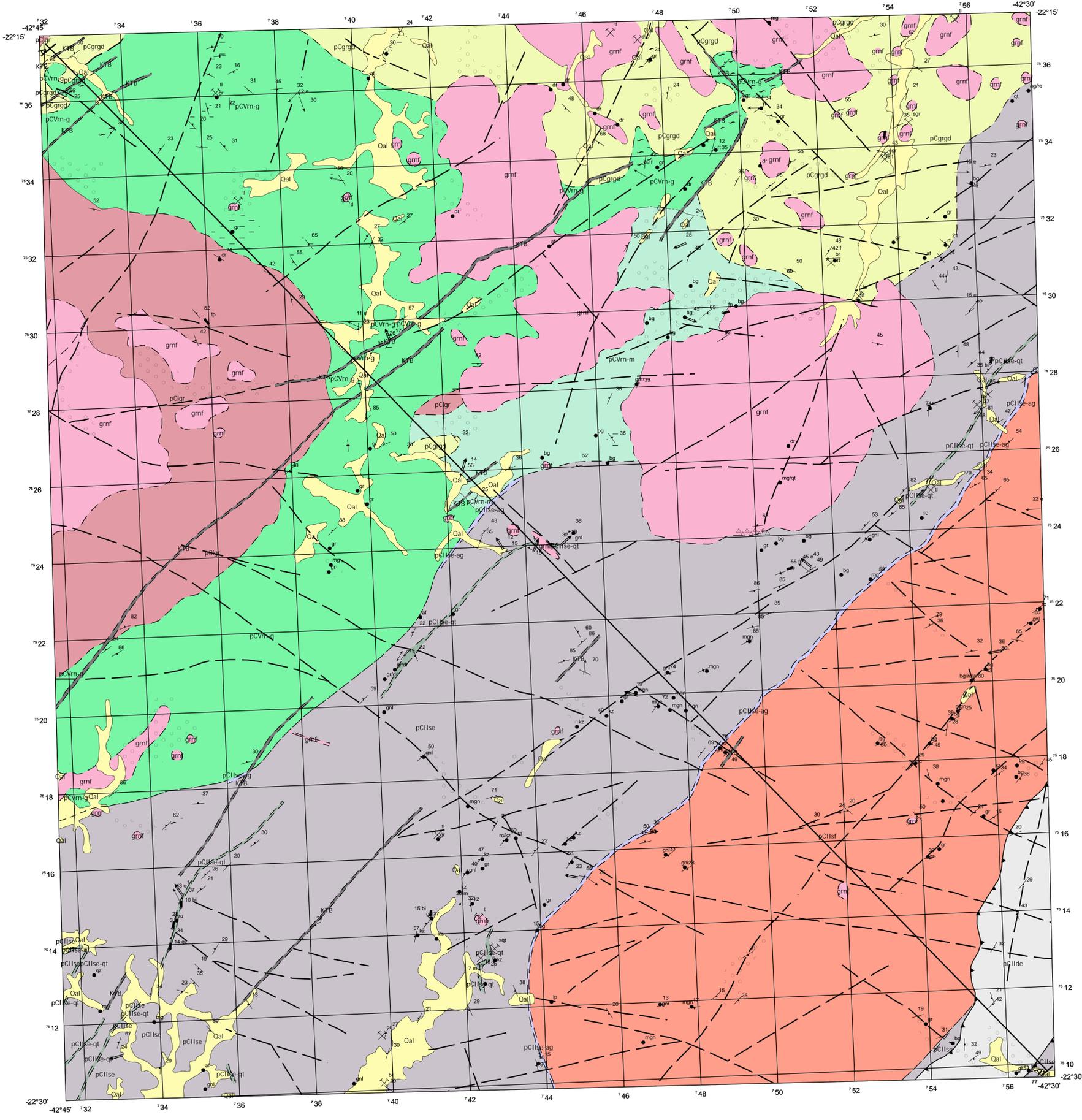
Nota: As diretrizes adotadas nesta folha as recomendações do I Seminário sobre critérios de mapeamento geológico e nomenclaturas de unidades do Pré-Cambriano no Estado do Rio de Janeiro e áreas limítrofes, realizado em Niterói - RJ em 1978.

PROJETO CARTA GEOLÓGICA DO ESTADO DO RIO DE JANEIRO
FOLHA ITAIPAUA
SF-23-Z-B-I-4
ESCALA 1:50.000

Digitalização: imagem
Financiamento: FAPERJ
Executor: DRM-RJ

GOVERNO DO RIO DE JANEIRO
SECRETARIA DE DESENVOLVIMENTO ECONÔMICO, ENERGIA, INDÚSTRIA E SERVIÇOS

Base compilada da Carta do Brasil, escala 1:50000 IBGE.
Edição em 1972, aerofotografias em 1966 AST-10/USAF.
Reambulação em 1974 e restituição fotogramétrica em 1972. Datum vertical: naegráfico/imbitubua; SC; Satum horizontal: Córrego Alegre, MG. Origem das coordenadas UTM: Equador e meridiano 45° W. Gr. acrescidas respectivamente, as constantes: 10000Km e 500Km. Projeção Universal transversa de Mercator. Equidistante das curvas de nível: 20m



Quaternário

Qal Depósitos aluvionares areno argilosos; cascalho portanto mutações (c).

Terceário

KTb Diabásio em Diques

Pré Cambiano

grnf Granito Nova Friburgo: rochas granitoides de estrutura homófona e fluidal, composição quartzodiorítica a granítica, constituindo maciços circunscritos, os tipos mais ácidos são, geralmente, porfíricos a porfíroides.

pCvrm-m Migmatitos variados, com palcosoma de biotita gnaissé, anfíbio gnaissé e neosoma de rocha granitóide fina e média: inclui-se biotita gnaissé fino(bg), laminado, da Unidade São Fidélis.

pCvrm-g Rocha granitóide serrogênica, fina a média, geralmente com moscovita.

pClgr *Leucogranitos Gnaissicos, granulação grossiera a média, cor clara, gnaissificação moderada a ausente. Contatos intrusivos com demais litologias do batólito.*

pCgrgd Plutonitos foliados primorogénicos, de composição granítica a granodiorítica, médios a grosseiros, com biotita-anfíbio (-granada), contendo segregações leucocráticas (lgr).

pClIse Unidade Santo Eduardo: gnaissés listrados-lenticulares, com trama cataclástica blastomilonítica, mesocráticas, com intercalações discretas de anfíbio. O conjunto associa-se intimamente com rochas granulíticas foliadas, leuco a melanocráticas, com estrutura laminada (cataclástica) marcada pela presença de escóides de quartzo.

pClIse-ag Unidade Santo Eduardo: milonito gnaissé e blastomilonito predominantes, geralmente tondeados, porfiroclásticos e porfiroblásticos, portando intercalações de variada litogia e com intercalações de anfíbio (-biotita) gnaissé migmatizado (ag).

pClIse-qt Importantes corpos de quartzo.

pClIse-ag Unidade São Fidélis: sillimanita - cordierita - (muscovita) - granada - biotita - ortocálcio - (microclina) - plagioclásio gnaissés e migmatitos, de coloração variável em tons de cinza, granulação variável, compondo estruturas febríticas, estromática e "schlieren" associados a leptinitos e a rochas calcossilicáticas. São transicionais para as rochas da suite charnockítica e incluem domínios mistos com os gnaissés semi-porfiroblásticos da Unidade Desengano - pm. Brecha Magmática: coloração roxa-acinzentada-clara; matriz afanítica de chert, envolvendo fragmentos angulosos de tamanho, cor e textura variáveis, de quartzo, calceldonia e de rocha - bm.

pClIde *Unidade Desengano: sillimanita) - (cordierita) - granada - biotita - ortocálcio - plagioclásio - microclina gnaissés lenticulares, homogêneos, desituidos de bandeamento, de coloração cinza clara, as vezes incipientemente foliados, granulação média a gr

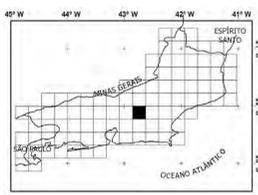
ESTRUTURAS

- Brecha Magmática
- Contato Aproximado
- Contato Definido
- Contato Encoberto
- Contato Provavel
- Contato Transicional
- Estrutura Anticinal Definida
- Estrutura Sincinal Definida
- Falha Aproximada
- Falha Definida
- Falha Encoberta
- Falha Inversa
- Falha Provavel
- Fratura Definida
- Lineamentos Estruturais
- Traco de Secao Geologica
- Cataclásada
- Brecha de Falha
- Zona Migmatizada
- Deposito de Talus e/ou Coluviais
- Direcao e Mergulho de Plano Axial de Dobra Aberta
- Direcao e Mergulho de Plano Axial de Dobra Similar
- Rumo Horizontalizado de Eixo de Dobra Similar
- Rumo e Caimento de Eixo de Dobra Similar e - Estrias m - Mullion
- F - Flexuramento
- Rumo e Caimento de Eixo de Dobra Paralela
- Direcao de Foliacao Vertical
- Direcao e Mergulho de Foliacao
- Orientacao Linear
- bi - Biotita
- fp - Fenocrístais de Feldspato
- Mina Paralisada
- br - Pedreira de Brita
- tl - Pedreira de Pedra Talhe
- mi - Mina em Atividade
- br - Pedreira de Brita
- tl - Pedreira de Pedra Talhe
- qz - Graos de Quartzo
- sgr - Saibreira de Rochas Granitoides
- sqz - Saibreira em Quartzo

OCORRÊNCIAS

- Arbóreo Mineral
- gnl - Milonito Gnaissé Leptinitica com Porfiroclastos
- qt - Quartzo
- rt - Restos de Teto nao Especificados
- ra - Rocha Alcalina
- rc - Rocha Calcossilicatica
- grm - Rocha Granitóide e Mascovita
- kz - Rocha Kinzigitica
- gr - Granito
- di - Diorito e Quartzo Diorito
- gr - Granito
- qz - Graos de Quartzo
- lp - Leptinito
- mg - Microgabro e Gabro

LOCALIZAÇÃO NO ESTADO



ARTICULAÇÃO DA FOLHA

ANTA	DUAS BARRAS	CORDEIRO
TERESÓPOLIS	NOVA FRIBURGO	QUARTEIS
ITABORAÍ	RIO BONITO	SILVA JARDIM



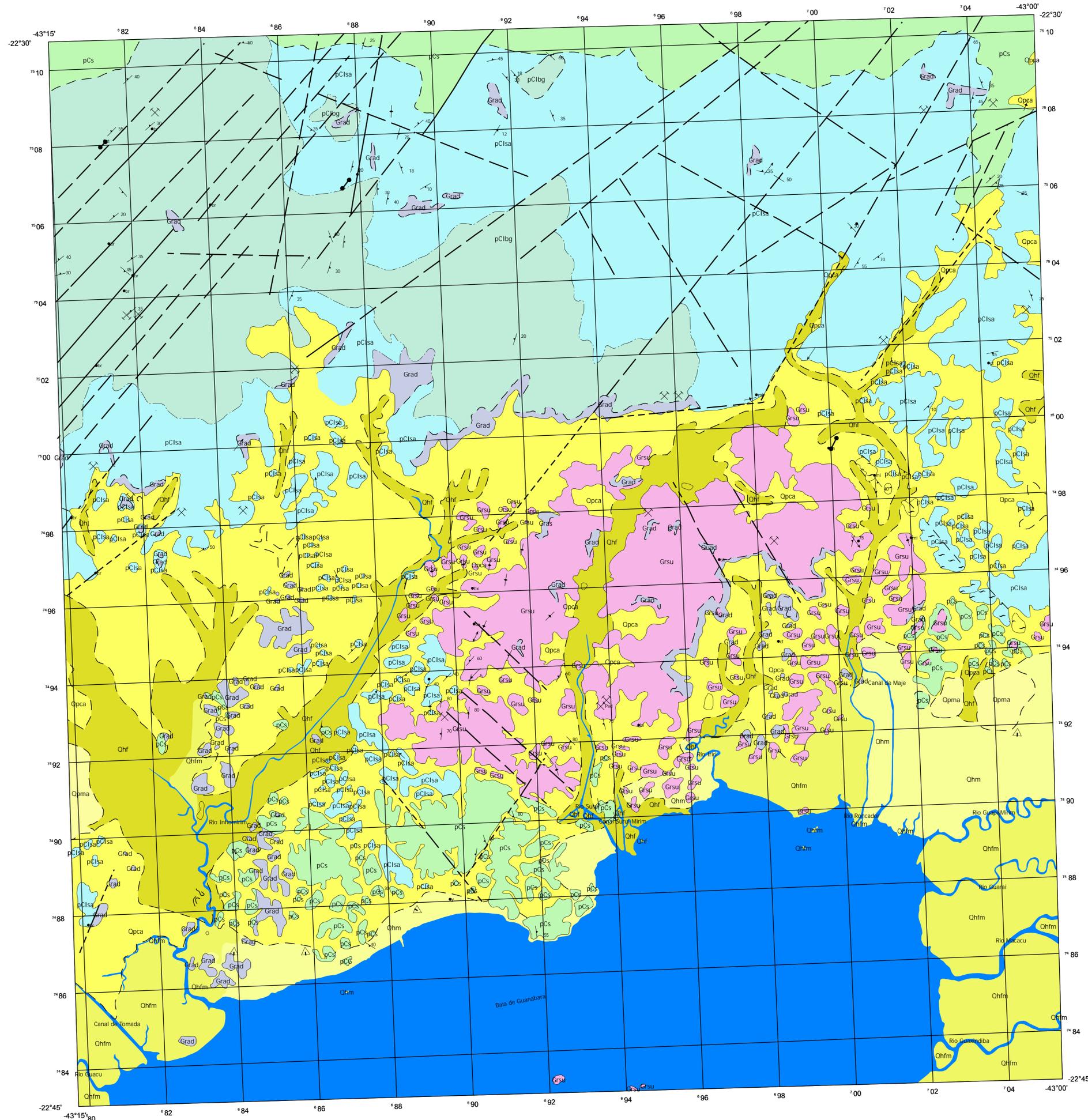
DIVISÃO MUNICIPAL



Serviços a cargo de GEOSOL - Geologia e Sondagens Ltda.
Chefe do Projeto: J. H. Grossi Sad
Coordenação: Aluizio Lício de Miranda Barbosa
Geologia por: Manfred Hettich
Supervisão DRM
NOTA: As diretrizes adotadas nesta folha seguem as recomendações do I Seminário sobre critérios de mapeamento geológico e nomenclatura de unidades do Pré - Cambriano no Estado do Rio de Janeiro e áreas limítrofes, realizado em Niterói-RJ, em 1978.

Bases compilada da Carta do Brasil, escala 1:50000, IBG - FIBG, Edição em 1974, aerofotografias, em 1965 YAST-10/USAF UTM, Equador e meridiano 45° W/G, acroscidas respectivamente, as constantes: 100000m e 500km, Projeção Universal Transversa de Mercator.

PROJETO CARTA GEOLÓGICA DO ESTADO DO RIO DE JANEIRO
FOLHA DE NOVA FRIBURGO
SF-23-Z-B-II-4
ESCALA 1:50.000



LEGENDA

Quaternário

- Ohf** Depósitos de Várzea: areias feldspáticas estratificadas selecionadas e argilas silíticas orgânicas.
- Ohfm** DEPÓSITOS FLUVIO MARINHOS: Sedimentos argilo - silíticos orgânicos.
- Ohm** DEPÓSITOS MARINHOS: Areias grosseiras granulosas, quartzosas, mal selecionadas.
- Opca** "FORMAÇÃO CACERIBU": Areias feldspáticas grosseiras, estratificadas, selecionadas; arenitos grosseiros com matriz caulinitica; conglomerados ortoquartzíticos e blocos e seixos de petrografia variada."
- Opma** "FORMAÇÃO MACACU": Camadas e lentes irregulares de areias arcozeanas, material arenogiloso, argilo arenoso de coloração variegada."

Pré Cambiano

- Grad** GRANITO ANDORINHA: biotita granito porfirítico, cor clara, matriz de grão médio com pórfiros centimétricos de microclina, formando pequenos corpos irregulares. Presença de xenólitos digeridos ou não. Contatos intrusivos com a Unidade Santo Eduardo.
- Grsu** "Granito Surui: róseo, granulação grosseira, até pegmatóide; megacristais centimétricos de microclina em matriz quartzo-diorítica de granulação média."
- pCs** "GRANITO: (gr) - (hb) - (bi) - Gnaise granítico a granodiorítico, de grão médio, cor clara, rosada, dobrado isoclinalmente foliação dada pelo alinhamento de aglomerados de máficos. Freqüentes diques de metabasitos deformados."
- pCibg** "UNIDADE BIGEN: Predominam biotita gnaisses graníticos, gnaissificação moderada ou ausente grão médio a fino, cor clara, estruturas nebulítica a schlieren localmente presente. Raros restos migmatíticos básicos. Localmente porfiroblástico."
- pCIsa** "UNIDADE SANTO ALEIXO: Predominam migmatitos com estruturas estromatolítica dobrada; localmente schollen, agmática, constituídos por leucossoma granítico a leplínico e melanossoma variando de biotita xistos a anfibolitos contatos transicionais com a unidade bigen. Tipos homogêneos, leuco a mesocráticos na mapeáveis a escala."

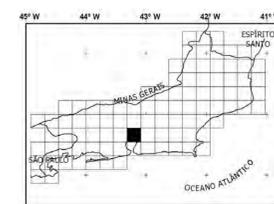
ESTRUTURAS

- Contato Aproximado
- Contato Definido
- - - Contato Transicional
- Dique de Diabasio
- △ Sítio Arqueológico
- ← Lineação Mineral Horizontal
- Lineação Mineral
- ↑ Direção de Foliacao Vertical
- ↖ Direção e Mergulho de Foliacao
- Falha Aproximada
- Falha Definida
- - - Falha Encoberta
- Cataclasada
- × Mina Paralisada
- ⊗ Mina em Atividade
- ped - Pedreira

OCORRÊNCIAS

- Ocorrência Litológica
- bi - Brecha de Injeção
- br - Brecha Tectônica
- bx - Biotita Anfíbolio Xisto Migmatítico
- ch - Charnockito
- cs - Calcossilicatada
- ct - Cataclasito
- mi - Milonito Gnaise
- qt - Quartzito
- y - Dique de Fonolito

LOCALIZAÇÃO NO ESTADO



ARTICULAÇÃO DA FOLHA



1972

DIVISÃO MUNICIPAL



MINISTERIO DO PLANEJAMENTO E COORDENAÇÃO GERAL
FUNDAÇÃO IBGE - INSTITUTO BRASILEIRO DE GEOGRAFIA
DEPARTAMENTO DE CARTOGRAFIA

EQUIPE TÉCNICA
COORDENADOR: Hélio Monteiro Penha
EQUIPE: André L. Ferrari
André Ribeiro
Elmo S. Amador
Fabio S. Pentagna
Marta do Carmo B. Junho
Thomas L. Brenner

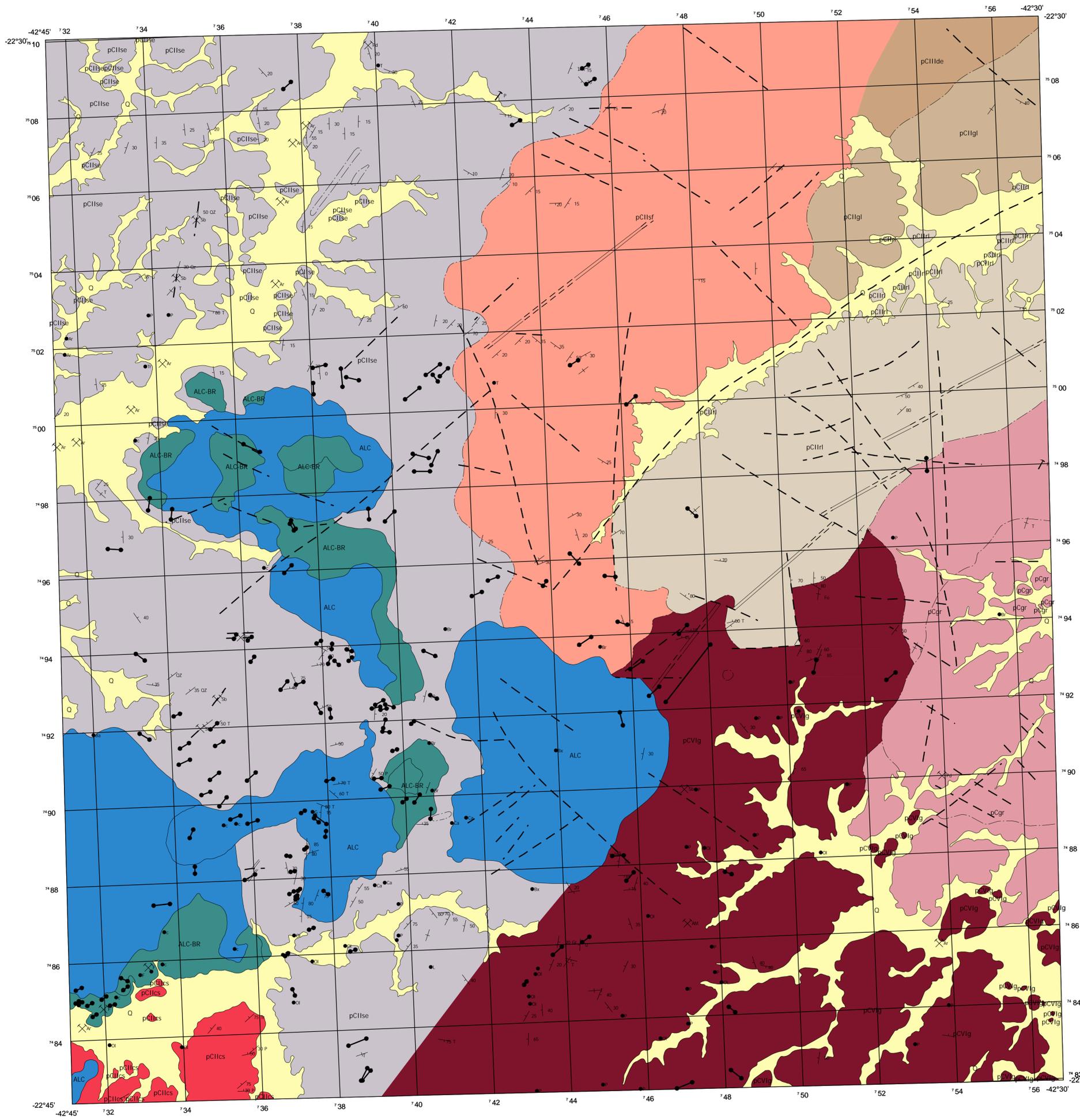
PROJETO CARTA GEOLÓGICA DO ESTADO DO RIO DE JANEIRO
FOLHA PETRÓPOLIS
SF-23-Z-B-IV-2
ESCALA 1:50.000

Digitalização Financiamento Executor

Origem da quilometragem: Equador e Meridiano 45°W.Gr.
acrescidas as constantes 10000km e 500km, respectivamente

Datum vertical: marégrafo Imbituba, SC
Datum horizontal: Corrego Alegre, MG

Levantamento estereofotogramétrico topográfico regular
Aerofotografias de 1966, reombinação 1971 e restrição 1972.
Triangulação especial



LEGENDA

Quaternário



Q

Pré Cambiano



ALC



pCgr



pCIcs



pCIrl



pCIse



pCIIsf



pCIIde



pCIlg



ALC-BR



pCVlg

ESTRUTURAS

- Contato Definido
- - - Contato Encoberto
- · - · Contato Transicional
- Dique de Diabasio
- Dique de Fonolito
- Dique de Lamprofiros
- Dique de Pegmatito
- Dique de Theralito
- Dique de Traquito
- - - Falha Provavel
- △△△ Brecha Tectonica

OCORRÊNCIAS

- Ocorrência Mineral
- Ar - Areia
- Ba - Brita
- Br - Brita
- Ca - Calcario
- F - Fonolito
- Bx - Biotita Xisto Melanossomatico
- L - Lamprofiros
- OI - Oligoclasio
- P - Pegmatito
- Tr - Theralito
- T - Traquito



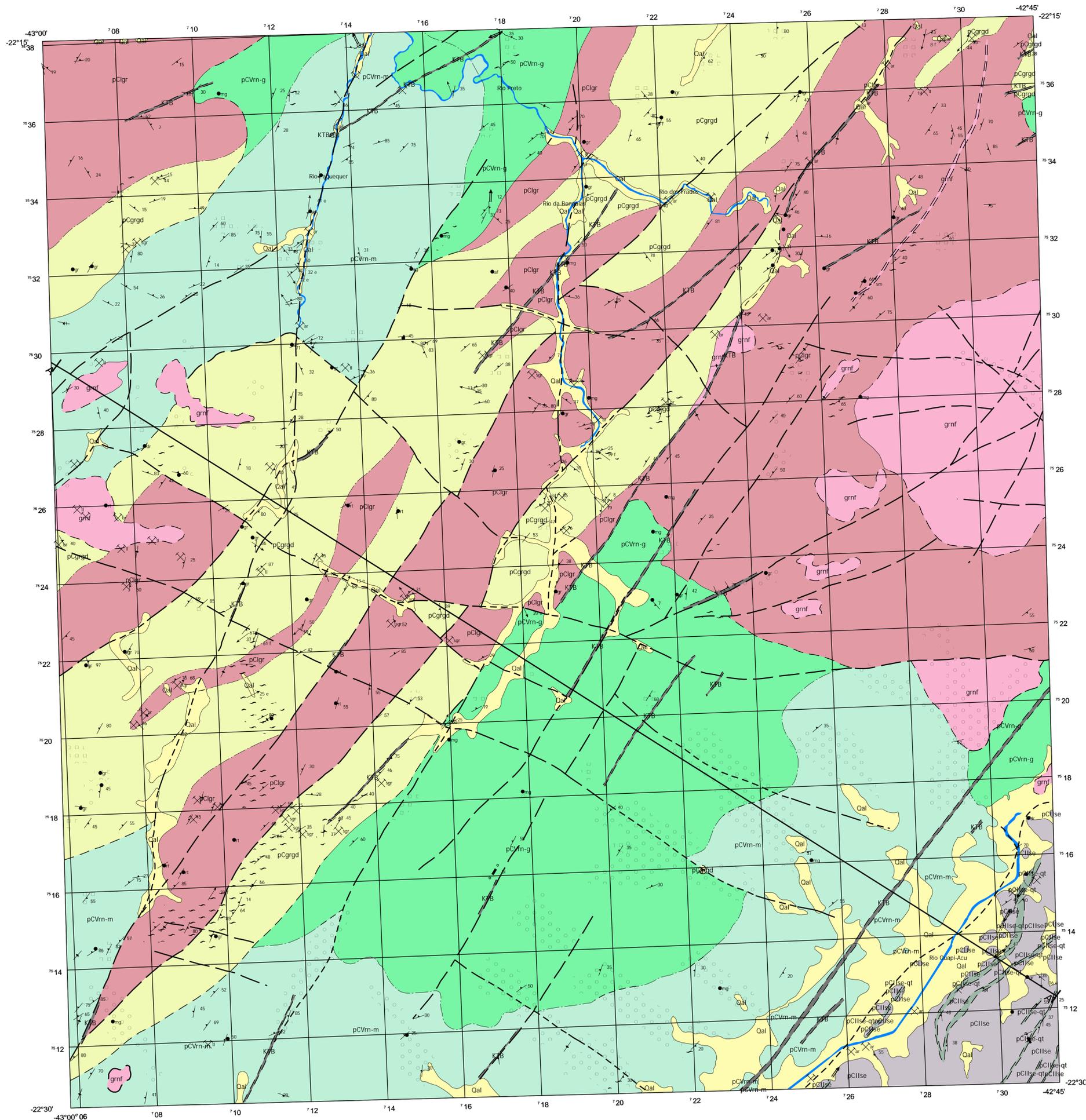
Escala 1:50.000

1978

DIVISÃO MUNICIPAL

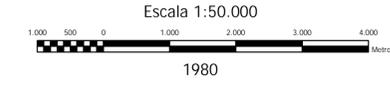


PROJETO CARTA GEOLÓGICA DO ESTADO DO RIO DE JANEIRO
 FOLHA RIO BONITO
 SF-23-Z-B-V-2
 ESCALA 1:50.000



- ### LEGENDA
- Quaternário**
- Qal Depósitos aluvionares areno argilosos; cascalho portante mutações (c).
- Terceário**
- KTB Diabásio em Diques
- Pré Cambiano**
- grnf Granito Nova Friburgo: rochas granitóides de estrutura homófona e fluidal, composição quartzodiorítica a granítica, constituindo maciços circunscritos, os tipos mais ácidos são, geralmente, porfiríticos a porfirídeos.
 - pCVrn-m Migmatitos variegados, com paleosoma de biotita gnaíse, anfíbolio gnaíse e neosoma de rocha granitóide fina a média com moscovita; assinalam-se zonas cataclásticas (ml) e porfiroclásticos (pf).
 - pCVrn-g Rocha granitóide serrogênica, fina a média, geralmente com moscovita.
 - pClgr "Leucogranitos Gnaíssicos, granulação grosseira a média, cor clara, gnaíssificação moderada a ausente. Contatos intrusivos com demais litologias do batólito."
 - pCgrgd Plutonitos foliados primogênicos, de composição granítica a granodiorítica, médios a grosseiros, com biotita-anfíbolio (-granada), contendo segregações leucocráticas (lgr).
 - pClise Unidade Santo Eduardo: gnaísses listrados-lenticulares, com trama cataclástica blastomilonítica, mesocráticas, com intercalações discretas de anfíbolio O conjunto associa-se intimamente com rochas granulíticas foliadas, leuco a melanocráticas, com estrutura laminada (cataclástica) marcada pela presença de discóides de quartzo.
 - pClise-qt Importantes corpos de quartzo.

- ### ESTRUTURAS
- AGULHAS_N-PT Contato Aproximado
 - Contato Definido
 - Contato Encoberto
 - Contato Suposto
 - Contato Transicional
 - Falha Aproximada
 - Falha Encoberta
 - Traco de Secao Geologica
- ### OCORRÊNCIAS
- AGULHAS_N-PT Depósito de Talus e/ou Colúviais
 - Porfiroclástico
 - Cataclásada
 - Direcao e Mergulho de Foliacao
 - Direcao de Foliacao Vertical
 - Xistosidade Horizontal
 - Orientacao Linear
 - Rumo Horizontalizado de Eixo de Dobra Paralela
 - Direcao de Plano Axial Verticalizado de Dobra Aberta
 - Direcao e Mergulho de Plano Axial de Dobra Aberta
 - Rumo Horizontalizado de Eixo de Dobra Similar
 - Rumo e Caimento de Eixo de Dobra Paralela
 - Orientacao Linear Horizontalizado
 - bi - Biotita
 - Rumo e Caimento de Eixo de Dobra Similar
 - e - Estrias
 - sc - Schlieren de Biotita
 - mi - Mioclivagem
 - f - Flexuramento
 - Mina em Atividade
 - ar - Areia
 - br - Pedreira de Brita
 - qt - Saibreira de Quartzo
 - lgr - Segregacoes Leucocráticas
 - mg - Microgabro
 - rc - Rocha Calcaissitica
 - rt - Restos de Teto nao Especificados
 - tr - Traquito



Serviços a cargo de GEOSOL - Geologia e Sondagens Ltda
Chefe do Projeto: J. H. Grossi Sad
Coordenação: Aluizio Licínio de Miranda Barbosa
Geologia por: Cláudio Pivo Pinto
Supervisão DRM
NOTA: As diretrizes adotadas nesta folha seguem as recomendações do I Seminário sobre critérios de mapeamento geológico e nomenclatura de unidades do Pré-Cambriano no Estado do Rio de Janeiro e áreas limítrofes, realizado em Niterói, em 1978.

PROJETO CARTA GEOLÓGICA DO ESTADO DO RIO DE JANEIRO
FOLHA TERESÓPOLIS
SF-23-Z-B-II-3
ESCALA 1:50.000

Base compilada da Carta do Brasil, escala 1:50000.
IBG - FIBG, Edição em 1974, aerofotografias em 1965/AST-10/USAF.
UTM: Equador e meridiano 45°W-Gr, acrescidas respectivamente, as constantes: 10000km e 500km.
Projeção Universal Transversa de Mercator.