

**UFRRJ**  
**INSTITUTO DE AGRONOMIA**  
**PROGRAMA DE PÓS-GRADUAÇÃO EM AGRONOMIA**  
**CIÊNCIA DO SOLO**

**DISSERTAÇÃO**

**Mapeamento Digital de Solos no Estado do Mato  
Grosso do Sul a partir de Dados Legados e Modelos  
Preditivos Baseados em Árvores**

**Pedro Armentano Mudado Xavier**

**2019**



**UNIVERSIDADE FEDERAL RURAL DO RIO DE JANEIRO  
INSTITUTO DE AGRONOMIA  
PROGRAMA DE PÓS-GRADUAÇÃO EM AGRONOMIA  
CIÊNCIA DO SOLO**

**MAPEAMENTO DIGITAL DE SOLOS NO ESTADO DO MATO  
GROSSO DO SUL A PARTIR DE DADOS LEGADOS E MODELOS  
PREDITIVOS BASEADOS EM ÁRVORES**

**PEDRO ARMENTANO MUDADO XAVIER**

*Sob a orientação da Professora*  
**Lúcia Helena Cunha dos Anjos**

*e Coorientação da Professora*  
**Helena Saraiva Koenow Pinheiro**

Dissertação submetida como requisito parcial para obtenção do grau de **Mestre** no Programa de Pós-Graduação em Agronomia, Área de Concentração em Ciência do Solo

Seropédica, RJ  
Fevereiro de 2019

X3m

Xavier, Pedro Armentano Mudado, 1985 - Mapeamento Digital de Solos no Estado do Mato Grosso do Sul a partir de Dados Legados e Modelos Preditivos Baseados em Árvores / Pedro Armentano Mudado Xavier. – Seropédica - RJ, 2019. 77 f.: il.

Orientadora: Lúcia Helena Cunha dos Anjos.  
Coorientadora: Helena Saraiva Koenow Pinhiro.  
Dissertação (Mestrado). -- Universidade Federal Rural do Rio de Janeiro, PPGACS, 2019.

1. Levantamento de Solos. 2. Mapeamento de Solos.  
3. Sensoriamento Remoto. 4. Ciencia do Solo. 5. Mapeamento Digital. I. Anjos, Lúcia Helena Cunha dos, 1957, orient. II. Pinhiro, Helena Saraiva Koenow, 1985, coorient. III Universidade Federal Rural do Rio de Janeiro. PPGACS. IV. Título.

É permitida a cópia parcial ou total desta dissertação, desde que seja citada a fonte.

**O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.**

**UNIVERSIDADE FEDERAL RURAL DO RIO DE JANEIRO**  
**INSTITUTO DE AGRONOMIA**  
**PROGRAMA DE PÓS-GRADUAÇÃO EM AGRONOMIA - CIÊNCIA DO SOLO**

**PEDRO ARMENTANO MUDADO XAVIER**

Dissertação submetida como requisito parcial para obtenção do grau de **Mestre**, no Programa de Pós-Graduação em Agronomia, área de Concentração em Ciência do Solo.

DISSERTAÇÃO APROVADA EM 18/02/2019

---

Helena Saraiva Koenow Pinheiro. Dra. UFRRJ  
(Co-orientadora)

---

César da Silva Chagas. Dr. Embrapa Solos

---

Marcos Bacis Ceddia. Dr. UFRRJ

## DEDICATÓRIA

*Aos meus ancestrais,  
familiares,  
amigos,  
tutores.*

*Dedico*

## AGRADECIMENTOS

Primeiramente, agradeço ao comitê de orientação, as professoras Lúcia Helena Cunha dos Anjos e Helena Saraiva Koenow Pinheiro, estendendo aos pesquisadores Waldir de Carvalho Júnior e César da Silva Chagas (os quais também considero como tal), por todos os ensinamentos e conhecimentos que tornaram possível o desenvolvimento deste trabalho, e que inspiram minha carreira profissional e formação pessoal. A vocês em especial, muito obrigado pela oportunidade, amizade, confiança e paciência.

Agradeço a Embrapa Solos em especial ao Cesão, Didi, Silvinho e Nilsão, pela confiança depositada, e oportunidade de trabalharmos em parceria, e é claro, pelos dados gentilmente cedidos e sem os quais esse trabalho não seria possível.

Agradeço a minha família e amigos pelo apoio neste momento tão importante para meu amadurecimento pessoal e profissional; principalmente a minha mãe Tereza, pela compreensão e apoio incondicional mesmo a distância durante todos os momentos, inclusive os mais difíceis.

Agradeço também aos mestres, Marcos Gervasio, Marcos Bacis e Mauro Homem, pelos preciosos ensinamentos que tiveram grande importância na minha formação, além de muito me inspirarem como profissionais e como seres humanos.

À Universidade Federal Rural do Rio de Janeiro (UFRRJ), em especial à equipe do Programa de Pós-Graduação em Agronomia - Ciência do Solo (PPGA-CS) e do Laboratório de Gênese e Classificação de Solos (LGCS).

Aos órgãos financiadores da pesquisa: CAPES, ao PPGA-CS/UFRRJ e a Embrapa Solos.

## RESUMO

XAVIER, Pedro Armentano Mudado. **Mapeamento digital de solos no estado do Mato Grosso do Sul a partir de dados legados e modelos preditivos baseados em árvores. 77f.** Dissertação (Mestrado em Agronomia, Ciência do Solo). Instituto de Agronomia, Universidade Federal Rural do Rio de Janeiro, Seropédica, RJ, 2019.

Os solos constituem um recurso natural de grande relevância, sobretudo por sua importância na produção de alimentos, na sustentação dos biomas e no armazenamento de água, garantindo a reposição das nascentes e mananciais, além de outros serviços ambientais. Assim, o conhecimento sobre as propriedades dos solos e sua distribuição na paisagem é importante para o seu manejo e para o planejamento territorial. A principal forma de se obter informações sobre os solos e sua distribuição é através de levantamentos de solos. A hipótese principal do estudo é que a partir das técnicas de mapeamento digital por áreas de referência é possível prever a distribuição espacial das unidades de solos. Ainda, através da seleção de variáveis preditoras e da avaliação de métodos preditivos quantitativos pode-se aperfeiçoar o mapeamento, reduzindo o caráter subjetivo da interpretação, conferindo um caráter quantitativo ao produto final. Sendo assim, o estudo teve como objetivo avaliar a eficiência dos métodos preditivos baseados em árvores, *Random Forest* (RF) e Árvores de Decisão (AD), para a extrapolação de unidades de mapeamento de solo nos municípios de Nova Alvorada do Sul e Rio Brillhante, a partir de 46 perfis descritos em Sidrolândia e Campo Grande, ambos no Mato Grosso do Sul. A abordagem através de modelos baseados em árvores possibilitou uma avaliação quantitativa dos fatores envolvidos na pedogênese, o que contribuiu para a melhor compreensão sobre cada fator e sua contribuição direta na formação dos solos. A utilização dos dados legados mostrou-se promissora para o processo de aprendizado dos padrões morfométricos, bem como para a extrapolação das unidades de mapeamento para toda a área. Ambos os modelos preditivos testados se mostraram bastante eficientes na extração das informações a partir dos dados de entrada e possibilitaram extrapolação para áreas semelhantes ainda não mapeadas, sendo o modelo RF o que apresentou o melhor desempenho preditivo, em todos os índices estatísticos avaliados, em relação ao modelo AD. Os modelos baseados em árvores (AD e RF) podem contribuir para o conhecimento sobre os fatores de formação e para qualificar sua contribuição na pedogênese, bem como na compreensão dos pedoambientes e distribuição das classes de solo na paisagem, além de subsidiar o levantamento de áreas semelhantes ainda não mapeadas.

**Palavras-chave:** Mineração de dados. Pedometria. Dados legados. Levantamento de Solos. Landsat 8

## ABSTRACT

XAVIER, Pedro Armentano Mudado. **Digital mapping of soils in the Mato Grosso do Sul state by reference areas and tree-based predictive models.** 77p. Dissertation (Master Science in Agronomy Soil Science). Instituto de Agronomia, Universidade Federal Rural do Rio de Janeiro, Seropédica, RJ, 2019.

Soils are a natural resource of great relevance, mainly due to their importance in the production of food, the support of biomes and in the storage of water, assuring the replacement of springs and water sources, as well as other environmental services. Thus, the knowledge about the soil properties and their distribution in the landscape is important for their management and for the territorial planning. The principal way to obtain information about the soils and its distribution is through soil surveys. The main hypothesis of the study is that from the techniques of digital mapping by using reference areas it is possible to predict the spatial distribution of soil units. Also, through the selection of predictor variables and the evaluation of quantitative predictive methods, the soil survey can be improved, reducing the subjective character of the interpretation, besides giving a quantitative character to the final product. The objective of this study was to evaluate the efficiency of tree-based predictive methods, Random Forest and Decision Tree, for the extrapolation of soil mapping unities in the municipalities of Nova Alvorada do Sul and Rio Brilhante, from 46 profiles described in Sidrolândia and Campo Grande, both located in Mato Grosso do Sul state. The approach through tree-based models enabled a quantitative evaluation of the factors involved in pedogenesis, which contributed to a better understanding of each factor and its direct contribution to soil formation. The use of the reference area proved to be adequate for the process of learning the morphometric patterns, as well as for the extrapolation of the mapping units to the entire area. Both predictive models tested proved to be quite efficient in the extrapolation of the mapping units, and the Random Forest model presented the best predictive performance, in all statistical indices evaluated, in relation to the Decision Tree model. The models based on trees can contribute to the knowledge about the soil formation factors and to qualify their contribution in the pedogenesis, as well as in the understanding of the pedoambientes and distribution of the classes of soil in the landscape; also, to support the mapping of similar areas not yet surveyed.

**Key words:** Data mining. Pedometrics. Legacy data. Soil survey. Landsat 8

## SUMÁRIO

1	INTRODUÇÃO.....	1
2	REVISÃO DE LITERATURA .....	3
2.1	Levantamento de Solos no Brasil.....	3
2.2	Mapeamento Digital para Predição de Solos .....	4
2.2.1	Árvore de decisão (AD).....	6
2.2.2	<i>Random forest</i> (RF) .....	8
2.3	Métodos de Prospecção de Solos: Extração e Aproveitamento de Dados Legados.....	9
3	MATERIAL E MÉTODOS .....	11
3.1	Caracterização da Área.....	11
3.2	Fatores de Formação .....	11
3.2.1	Relevo .....	12
3.2.2	Material de origem e tempo (Litoestratigrafia) .....	13
3.2.3	Clima .....	15
3.2.4	Organismos (Vegetação) .....	16
3.3	Obtenção e Processamento de Dados.....	18
3.4	Estratégias de Mapeamento Digital de Solos.....	22
3.4.1	Compatibilização da base de dados e softwares utilizados .....	23
3.4.2	Procedimentos de amostragem .....	24
3.4.3	Modelagem pedométrica .....	28
3.4.4	Validação dos modelos preditores.....	30
4	RESULTADOS E DISCUSSÃO.....	31
4.1	Pedoambientes.....	31
4.2	Análise das Covariáveis Morfométricas e Espectrais .....	35
4.3	Análise dos Modelos .....	39
4.3.1	Modelagem da árvore de decisão .....	39
4.3.2	Modelagem por <i>random forest</i> .....	41
4.3.3	Comparação entre os modelos preditivos.....	43
5	CONCLUSÕES.....	53
6	CONSIDERAÇÕES FINAIS .....	54
7	REFERÊNCIAS BIBLIOGRÁFICAS .....	55
8	ANEXO .....	64

## 1 INTRODUÇÃO

Os solos podem ser definidos como, coleção de corpos naturais, constituídos por partes sólidas, líquidas e gasosas, tridimensionais, dinâmicos, formados por materiais minerais e orgânicos, que ocupam a maior parte da superfície emersa das extensões continentais do nosso planeta, contém matéria viva e podem ser vegetados na natureza, onde ocorrem e ocasionalmente podem ter sido modificados por atividades humanas (EMBRAPA, 2018).

Pode ser resumido na seguinte equação:  $S = f(mo, cl, r, o, t)$  a qual expressa "o solo (S) como função das interações entre os fatores ambientais: material de origem (mo), clima (cl), relevo (r), organismos vivos (o), atuando ao longo do tempo (t)" (Dokuchaev, 1883). Ou seja, são corpos naturais que guardam características próprias desenvolvidas durante seu processo de gênese, o qual é condicionado pelos fatores ambientais. Em 1941 valendo-se dos preceitos de Dokuchaev, H. Jenny escreveu o livro intitulado *Factors of Soil Formation* (Jenny, 1941), onde cunhou um acrônimo para os fatores de formação do solo: CLORPT (CL = clima, O = organismos, R = relevo, P = material parental e T = tempo). Essa representação dos fatores, de fácil associação e memorização, se tornou base para o ensino de pedogênese.

Dentre os recursos naturais em nosso planeta, o solo tem grande relevância pelos serviços ambientais que presta aos sistemas e as populações humanas, sobretudo a produção de alimentos, como filtro de contaminantes, controle de fluxos de gases para a atmosfera, sustentação de biodiversidade e no armazenamento de água, garantindo a reposição das nascentes e mananciais (Lepsch, 2002). Assim, o conhecimento sobre as propriedades dos solos e sua distribuição na paisagem é importante para o seu manejo e para o planejamento territorial.

Uma das formas de se obter informações sobre os solos e sua distribuição, em nível regional, municipal ou mesmo em detalhe de uma propriedade, é através de levantamentos de solos. Entretanto esses estudos, em geral, possuem elevados custos e podem ser demorados em função da escala de trabalho. Este fato contribuiu para a escassez de levantamentos de solos em escalas maiores (mais detalhadas) no Brasil, que são exigidos para atender as várias demandas de planejamento. Entre elas, os zoneamentos agroecológicos e econômico-ecológicos, o planejamento de uso da terra (rural e urbano), bem como o planejamento de uso, manejo e conservação do solo em escalas de bacias hidrográficas.

O uso de técnicas computacionais modernas para manipulação e análise de dados espaciais têm contribuído de maneira relevante para maior rapidez e redução da relação custo/benefício da execução dos levantamentos de solos, além de torná-los mais quantitativos e possibilitar avaliar as incertezas associadas aos produtos gerados (Pinheiro, 2015). Os produtos do levantamento podem conter grande volume de dados quantitativos e qualitativos, sobre os aspectos do meio físico, além de conferir maior facilidade às interpretações e facilitar a apresentação das informações para vários usos (Moore et al., 1993; Lagacheire e Voltz, 2000; Zhu et al, 2001; Smith et al., 2006). O aprimoramento das técnicas de investigação das informações de solos, aliado à introdução de novos conceitos e metodologias, tais como o aprendizado de máquina e a simulação computacional, pode auxiliar os pesquisadores no entendimento dos processos que ocorrem nos solos (Samuel-Rosa et al., 2015).

Os métodos de mineração de dados aumentam o poder de interpretação dos mesmos e podem fornecer soluções para auxiliar na extração automática de informações de um conjunto de dados disponíveis (Behrens et al., 2005). Aliado a estratégias como a extrapolação das relações-solo-paisagem de uma área de referência para regiões fisiograficamente semelhantes,

onde essas relações ainda não são conhecidas, pode-se reduzir o tempo necessário para execução de levantamentos de solos.

A hipótese principal do estudo é que a partir das técnicas de mapeamento digital por áreas de referência é possível prever a distribuição espacial das unidades de solos. Como segunda hipótese, derivada da primeira, tem-se que através da seleção de variáveis preditoras e da avaliação de métodos preditivos quantitativos pode-se aperfeiçoar o mapeamento, reduzindo o caráter subjetivo da interpretação do pedólogo e facilitar a comunicação entre dados, além de conferir um caráter quantitativo ao produto final.

Assim, o estudo teve como objetivo avaliar a eficiência dos métodos preditivos baseados em árvores para a extrapolação de unidades de mapeamento de solo nos municípios de Nova Alvorada do Sul e Rio Brilhante no Estado do Mato Grosso do Sul (MS), com base em informações obtidas no levantamento de solos dos municípios de Sidrolândia e Campo Grande (áreas de referência). Também foi feita a análise de padrões morfométricos e espectrais relativos aos fatores de formação para interpretar a gênese dos solos da área em questão.

## 2 REVISÃO DE LITERATURA

### 2.1 Levantamento de Solos no Brasil

A execução de mapeamentos de solo em todos os Estados da Federação brasileira é uma demanda das instituições federais de pesquisa e planejamento, dos estados e municípios, da iniciativa privada e organismos internacionais. As informações nos levantamentos de solos são relevantes para o planejamento adequado do uso racional das terras e para a gestão ambiental, o que é essencial para conciliar o desenvolvimento econômico e social, com a conservação e proteção dos recursos naturais, contemplando assim os requisitos básicos para o desenvolvimento sustentável, previsto na Agenda 21 (EMBRAPA, 2003).

De maneira geral, os mapeamentos executados no Brasil, possuem caráter exploratório ou de reconhecimento (1:1.000.000 a 1:250.000), pois foram feitos com objetivo estratégico de ordenamento territorial, entretanto, essa escala não se mostra adequada para o planejamento em nível mais local, onde se faz necessário um maior nível de detalhamento (PRONASOLOS, 2016). Regiões do país com grande pressão de ocupação agrícola necessitam informações em nível de detalhamento adequados para subsidiar a ocupação ordenada de seu território, uma vez que o inadequado processo de ocupação das terras vem ocasionando a degradação do solo e ambiental. Ainda assim, como citado por Chagas (2006), os levantamentos de solo têm sido considerados atividade não ligada diretamente à produção e de razão custo benefício pouco valorizada.

Os levantamentos de solos em nível de detalhamento apropriado são vitais para o planejamento agrícola e para estimativa da produção, conforme destacado por Singh e Dwivedi (1986). Portanto, uma adequada organização territorial depende da disponibilidade de levantamentos com bom nível de detalhamento, em que a coerência ecológica seja respeitada para o melhor aproveitamento dos recursos naturais, a otimização do setor produtivo e maximização da produtividade das propriedades rurais.

O Instituto José Augusto Trindade, localizado no Município de Sousa, no Estado da Paraíba, é considerado pioneiro no estudo minucioso de solos. Em 1947 foi elaborado o primeiro mapa de solos, denominado mapa agroecológico da bacia de irrigação do açude São Gonçalo, no Estado da Paraíba, mesmo ano de criação da Comissão de Solos do Centro Nacional de Ensino e Pesquisas Agrônomicas do Ministério da Agricultura, que tinha inicialmente, a função coordenar os estudos de solos a serem realizados em diferentes regiões brasileiras. Em 1953, com a reestruturação a Comissão de Solos foi aprovado o Programa de Levantamento de Reconhecimento dos Solos Brasileiros, que iniciou pelos Estados do Rio de Janeiro e Distrito Federal, em 1954 e 1955, respectivamente. Essa comissão teve por incumbências principais, a execução do plano básico de inventário geral de recursos de solos do território nacional e o aperfeiçoamento da capacitação de levantamentos de solos (EMBRAPA, 2019).

O Projeto RADAM criado em 1971 pelo Departamento Nacional da Produção Mineral era voltado para o levantamento dos recursos naturais da Amazônia e, em 1976, foi ampliado para todo território nacional passando a se chamar Projeto RADAMBRASIL, que produziu 38 volumes de mapas de solos exploratórios com escala de 1:1.000.000. Nessa mesma época foram realizados pelo Serviço Nacional de Levantamento e Conservação de Solos - SNLCS órgão público responsável pelas informações de solos, os levantamentos de Solos de nove estados brasileiros, o Mapa Esquemático dos Solos das Regiões Norte, Meio-Norte e Centro-Oeste do Brasil e, em 1981, o Mapa de Solos do Brasil na escala 1:5.000.000 (BRASIL, 1982). Após 1981 o SNLCS publicou ainda os levantamentos de solos dos Estados do Paraná, Maranhão e Piauí, enquanto o Projeto RADAMBRASIL publicou, até 1987, mais 12 folhas que pertencem aos volumes de números 26 a 34 (BRASIL, 1982).

Os levantamentos fornecem uma base de informação com descrições dos solos, consistindo de dados morfológicos, físicos, químicos e mineralógicos, associados às informações ambientais como clima, vegetação, geologia, relevo, drenagem e uso da terra. Tais informações pedológicas e suas correlações com os dados ambientais são publicadas nos boletins e relatórios técnicos que acompanham os mapas de solos. Essas informações possibilitaram a conceituação e a identificação das classes de solos no território brasileiro e são a base do desenvolvimento do Sistema Brasileiro de Classificação dos Solos - SiBCS (Santos et al., 2018).

Relatório recente feito por técnicos do Tribunal de Contas da União (TCU) apontou vários problemas com as informações sobre solos no País que levaram a proposição de um Programa Nacional de Solos (Pronasolos), que envolve diversos ministérios e órgãos federais liderado pela Embrapa Solos e com participação de instituições de pesquisa, ensino e extensão de todo o país, além de representação da Sociedade Brasileira de Ciência do Solo (PRONASOLOS, 2016). Foi ressaltado no documento que o conhecimento atual sobre os solos em escala adequada não é suficiente, há dificuldade no acesso aos poucos dados disponíveis, inconsistências nas informações oficiais de ocupação do território, sobreposição e lacunas de atuação governamental, bem como complexidade e dispersão da legislação brasileira sobre o assunto (Relatório de Auditoria TC 011.713/2015-1).

O Pronasolos tem como meta executar o mapeamento do solo de todo o país, no período entre 10 e 30 anos, em escalas que permitam tomada de decisão e estabelecimento de políticas públicas nos níveis municipal, estadual e federal – 1:25 mil, 1:50 mil, 1:100 mil, respectivamente, diante desse programa espera-se a retomada das ações de levantamento de solos no Brasil (PRONASOLOS, 2016). A iniciativa traz como perspectiva a aplicação e desenvolvimento de novas tecnologias, entre elas o uso das técnicas de mapeamento digital de solos (MDS) como uma das estratégias de mapeamento para o território brasileiro.

Nesse contexto, o Zoneamento agroecológico do Estado do Mato Grosso do Sul (ZAE-MS), programa que visa mapear o estado do Mato Grosso do Sul em uma escala de mapeamento de 1:100.000, apresenta-se como oportunidade para a aplicação de técnicas de mapeamento digital de solos para o levantamento dos solos de forma pioneira dentro do Pronasolos e possibilita o desenvolvimento das técnicas e teste dessa abordagem perante o desafio de melhorar as bases de solos existentes no país, e contribuir para a retomada do mapeamento de solos no Brasil.

Além disso o programa tem como um dos seus objetivos estabelecer uma base de dados integrada, na qual as informações de solos provenientes de trabalhos anteriores (dados legados) e as que vierem a ser produzidas estejam organizadas e sistematizadas para consulta do público em geral (PRONASOLOS, 2016).

## **2.2 Mapeamento Digital para Predição de Solos**

O mapeamento digital de solos (MDS) tem se desenvolvido principalmente devido à crescente demanda por informações de solos e disponibilidade de métodos mais eficientes de prospecção de dados, além do rápido avanço de inovações tecnológicas e maior facilidade na obtenção de mapas auxiliares, com destaque para as imagens de sensores remotos e os atributos derivados dos modelos digitais de elevação. O estudo das relações entre solo e paisagem pode identificar padrões de ocorrência de solos em função das diferentes combinações de atributos que exercem influência direta na pedogênese.

O fundamento principal do MDS consiste na predição de classes e/ou atributos por meio de equações matemáticas, de classes e propriedades de solos, de forma espacialmente contínua, onde o conjunto de dados permanece para análise e interpretações em qualquer

época além de introduzir vantagens em relação ao método tradicional para o mapeamento de solos, sendo uma alternativa rápida e econômica, e que vem aumentando consideravelmente a confiabilidade dos dados gerados (McBratney et al., 2003). O MDS é conceituado por Lagacherie e McBratney (2007) como a “*criação e a população de sistemas de informação espacial de solos por meio de modelos numéricos para inferir as variações espaciais e temporais de classes de solo e suas propriedades, a partir de observações, conhecimento e de dados de variáveis ambientais relacionados*” (tradução do autor).

O mapeamento digital depende diretamente da existência de uma base de dados, que nem sempre abrange toda a extensão da área de estudo. Segundo Ten Caten (2011), isto tem condicionado a extensão territorial da maioria dos trabalhos, pois a união entre cartas topográficas, mapas temáticos, e outras fontes de informações digitalizadas é necessária e essas bases nem sempre usam um mesmo padrão de projeção, volume de dados, precisão e métodos de obtenção. A pedometria possibilitou a caracterização espacial das classes de solos e unidades de mapeamento com estimativas do erro, intrínsecas a qualquer processo de amostragem. No MDS, amostras de solos e variáveis ambientais são obtidas por metodologias tradicionais ‘*in situ*’ no campo e através de imagens captadas por aviões ou satélites. Esses dados são empregados em modelos preditivos que possibilitam gerar novas informações com estatísticas de qualidade associadas (Minasny et al., 2008).

Estudos da variabilidade do solo têm sido abordados usando métodos quantitativos a partir do final da década de 70. Os estudos usando a geoestatística, com destaque para Matheron (1971) seguido por Journel e Huibregts (1978), Isaaks e Srivastava (1989), Webster e Oliver (1990), Cressie (1991), Deutsch e Journel (1992), foram bastante úteis para o desenvolvimento do conhecimento teórico no estudo da variabilidade do solo.

Já no início dos anos 2000, com os avanços tecnológicos inovadores, vários autores tais como Hengl et al. (2004), Minasny et al. (2003), McBratney et al. (2003), Santos et al. (2003) e Santos et al. (2007), propunham mudanças nas técnicas usadas para o mapeamento de solos. Entre as técnicas apontadas estavam o uso de recursos computacionais, como algoritmos preditivos, avaliação da modelagem global e dinâmica da gênese de solos, integração de sistemas de informações geográficas e ferramentas geoestatísticas. Ainda, o uso de inovações tecnológicas e de imagens de alta resolução, para a produção de mapas funcionais e zoneamentos diversos, podem auxiliar no aumento de escala e na avaliação da acurácia e erro.

O MDS vem quebrando paradigmas e ganhando espaço, principalmente devido à crescente disponibilidade de mapas auxiliares (Minasny e McBratney, 2015), os quais tornam possível a previsão da ocorrência de tipos de solos em áreas ainda não mapeadas. Coelho (2010) destaca como grande vantagem dessa abordagem, a possibilidade da extrapolação dessas relações para áreas adjacente de fisiografia semelhante. A predição de classes de solos utilizando técnicas de mapeamento digital é baseada na escolha de variáveis preditoras que melhor explicam a variabilidade de solos em uma determinada condição de paisagem. Portanto, a elaboração dos algoritmos preditivos, e a extrapolação para outras áreas de ecologia e geomorfologia semelhantes, ainda é um procedimento que deve ser minuciosamente analisado (Pinheiro, 2012).

Duas abordagens contribuem para a modelagem preditiva: a modelagem dos dados e o aprendizado de máquinas (Breiman, 2001). A modelagem dos dados avalia a qualidade de ajuste através do resíduo para fornecer mais informações sobre como se relacionam as variáveis de resposta com as variáveis de entrada. O aprendizado de máquinas avalia a precisão preditiva e é considerado como caixa preta uma vez que não é possível compreender exatamente quais critérios adotados na tomada de decisão. Geralmente possuem maior poder preditivo com menos esforços de modelagem (Chagas et al., 2013). Embora, segundo Santos

et al. (2007), as limitações quanto à qualidade da informação obtida em campo e a precisão dos métodos analíticos utilizados continuem as mesmas do mapeamento tradicional.

No delineamento preliminar de unidades de solo no MDS são aplicadas técnicas que integram atributos do terreno derivados de um modelo digital de elevação (MDE) para prever classes do solo (McBratney et al., 2000; Behrens et al., 2010; Jafari et al., 2014; Vaysse e Lagacherie, 2015) além de dados climáticos e geológicos. Embora as abordagens empregadas sejam baseadas em novas tecnologias, o modelo de classificação continua fundamentado na relação solo-paisagem. De acordo com este conceito, uma vez conhecidas as relações entre os solos e seus respectivos ambientes, é possível inferir os solos que ocorrem em uma determinada área, por meio da avaliação das condições ambientais. Portanto, para a aplicação dessas técnicas de predição, é necessária uma fase de treinamento, onde os modelos aprendem o conjunto de condições em que cada unidade de mapeamento que se deseja mapear ocorre ao longo da área.

Entre os métodos mais utilizados para extração de informações e reconhecimento de padrões, o aprendizado de máquinas apoiado em modelos de árvore vem sendo usado no MDS, com grande potencial na modelagem ambiental e predição espacial de solos. No Brasil tais métodos foram utilizados por Carvalho Júnior (2005), Chagas et al. (2010, 2013) e Pinheiro et al. (2017), com bons resultados para predição de classes de solos. Além de vários estudos encontrados na literatura (Behrens et al., 2010; Hansen et al., 2009; Minasny e McBratney, 2006; Stumpf et al., 20016; Schmidt et al., 2008; Lemercier et al., 2011; Giasson et al. 2011).

As árvores de decisão são muito úteis por identificar as interações, mas geralmente têm menor poder preditivo do que o método *Random Forest* (RF), que utiliza uma coleção de árvores de decisão independentes de modo que cada árvore tenha sido treinada em um conjunto de dados “*bootstrapped*” com uma seleção aleatória de preditores. Possui maior poder de previsão, porém uma menor interpretabilidade das interações (Breiman, 2001).

Os modelos gerados por esses métodos podem contemplar informações categóricas ou contínuas, como por exemplo, mapas de solo convencionais, além de cartas pluviométricas, isotermas, modelos digitais de elevação (MDEs), mapas categóricos de geologia ou litologia, ou qualquer outro dado que esteja relacionado aos fatores de formação. Nesse sentido, em regiões onde levantamentos e dados de solo são escassos, essa técnica torna-se uma alternativa para planejamento e gestão do uso da terra (Mora-Vallejo et al., 2008; Sindayihubura et al., 2017), pois permite gerar informações de solo através de dados já disponíveis que muitas vezes são de fácil aquisição e baixo custo, exigindo também menor tempo de execução comparado aos mapas convencionais de solo (Hengl et al., 2015; Santos et al., 2007).

Os modelos baseados em árvores classificam a partir da hierarquização das covariáveis de maior influência na variável dependente. No entanto, seus algoritmos diferem-se quanto ao número de árvores – apenas uma ou várias, ou ainda profundidade dos ramos (Breiman, 2001). Os métodos baseados em árvores adotados nesse estudo são detalhados nos itens subsequentes.

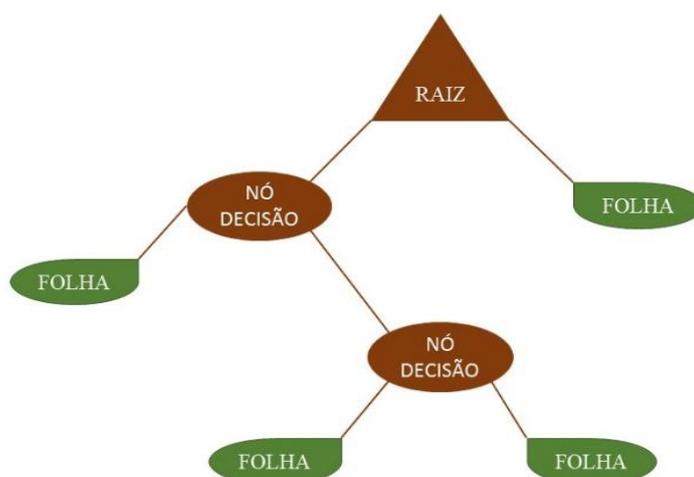
### **2.2.1 Árvore de decisão (AD)**

O procedimento de Árvore de Decisão (AD) constitui-se em uma hierarquia de nós dispostos sequencialmente que separam os dados de entrada de forma binária para um ramo ou para outro, até resultar em um nó terminal onde a resposta final é obtida, no caso a classe de solo predita, conforme pode ser observada na Figura 1. Baseada no *Classification and Regression Tree* (CART) desenvolvido por Breiman et al. (1984) o algoritmo utiliza inteligência artificial e aprendizado de máquinas e apresenta potencial de predição de classes

ou atributos de solo no MDS. É uma ferramenta robusta após análise rigorosa envolvendo a validação cruzada para selecionar a árvore ideal. Diferente das árvores de regressão, que são capazes de estimar dados contínuos, os modelos baseados em árvores de decisão classificam alvos categóricos, através de preditores categóricos ou contínuos (Rokach e Maimon, 2008)

A árvore utiliza treinamento supervisionado para classificação ou predição de amostras desconhecidas, a partir de registros conhecidos, apresenta-se um conjunto de treinamento do qual uma árvore de decisão é montada e determina-se a hierarquia entre os nós conectados por ramos. O nó interno, também conhecido como decisório ou nó intermediário é a unidade de tomada de decisão que avalia, por meio de teste lógico, qual será o próximo nó descendente ou filho. Em contrapartida, um nó externo que não tem nó descendente, também conhecido como folha ou nó terminal, está associado a um rótulo ou valor (Rokach e Maimon, 2008).

Geralmente, esses algoritmos dividem recursivamente conjunto de dados escolhendo interrupções em dados de covariáveis que ajudam a purificar ou aumentar o conteúdo informacional dos nós do modelo (ramos) (Breiman et al., 1984; Pedregosa et al., 2011). Os dados são apresentados ao nó inicial da árvore e, dependendo do resultado do teste lógico usado pelo nó, a árvore ramifica-se para um dos nós filhos. Esse procedimento é repetido até que um nó terminal seja alcançado. A repetição desse procedimento caracteriza a recursividade da árvore de decisão. A Figura 1 mostra a estrutura da árvore de decisão.



**Figura 1:** Arquitetura da Árvore de Decisão onde o nó raiz recebe o dado e repassa para o nó subsequente até chegar à Folha que representa a previsão. (Adaptado de Breiman et al., 1984).

O particionamento recursivo utiliza todos os dados da amostra e considera todos os valores possíveis de todas as variáveis usadas. Iniciando-se pelas variáveis que proporcionam a maior diferenciação dos dados. Seleciona a variável através do valor do nó ( $X=t_1$ ) chamado de "*Split*", se  $X < t_1$  os dados são enviados para a esquerda e se  $X > t_1$  os dados vão para a direita. Em seguida, este mesmo processo é repetido nos dois nós gerados e assim por diante criando, então uma árvore, sendo este particionamento sempre binário (Breiman et al., 1984).

As regras para o particionamento dos dados, no caso da árvore de decisão é determinada pelo índice *Gini* (Gini, 1912) que mede o grau de heterogeneidade dos dados, ou seja mede a impureza de um nó. Assim, para um conjunto de dados  $T$  que contém  $n$  classes, tem-se que:

$$G(T) = 1 - \sum_{j=1}^n (p_j)^2 \quad (\text{Eq. 1})$$

Onde,  $p_j$  é a frequência relativa da classe  $j$  no conjunto  $T$ .

Árvore usa a impureza *Gini* para medir a qualidade das divisões para construção de árvores. A impureza de *Gini* é uma medida da heterogeneidade de classes em um nó e é minimizado no processo de construção de árvores para tentar criar nós-folha com apenas uma classe, ou uma impureza *Gini* valor de zero (Breiman et al., 1984).

É um método preditivo robusto, e tem bom desempenho com grandes quantidades de informação, e tem a vantagem de ser um método não paramétrico (não assume nenhuma distribuição particular para os dados) e não depende da escala das variáveis para a estruturação da árvore. Além disso, fornece automaticamente a importância das variáveis; possibilita a utilização de variáveis discretas e contínuas sem preparação ou normalização; apresenta resultado bastante útil para a interpretação, é um modelo “caixa branca” onde as situações observadas são explicadas facilmente através da lógica booleana; e ao contrário da Regressão Linear, não é sensível a presença de “*outliers*” ou registros faltantes nos dados (Chagas, 2017).

### 2.2.2 Random forest (RF)

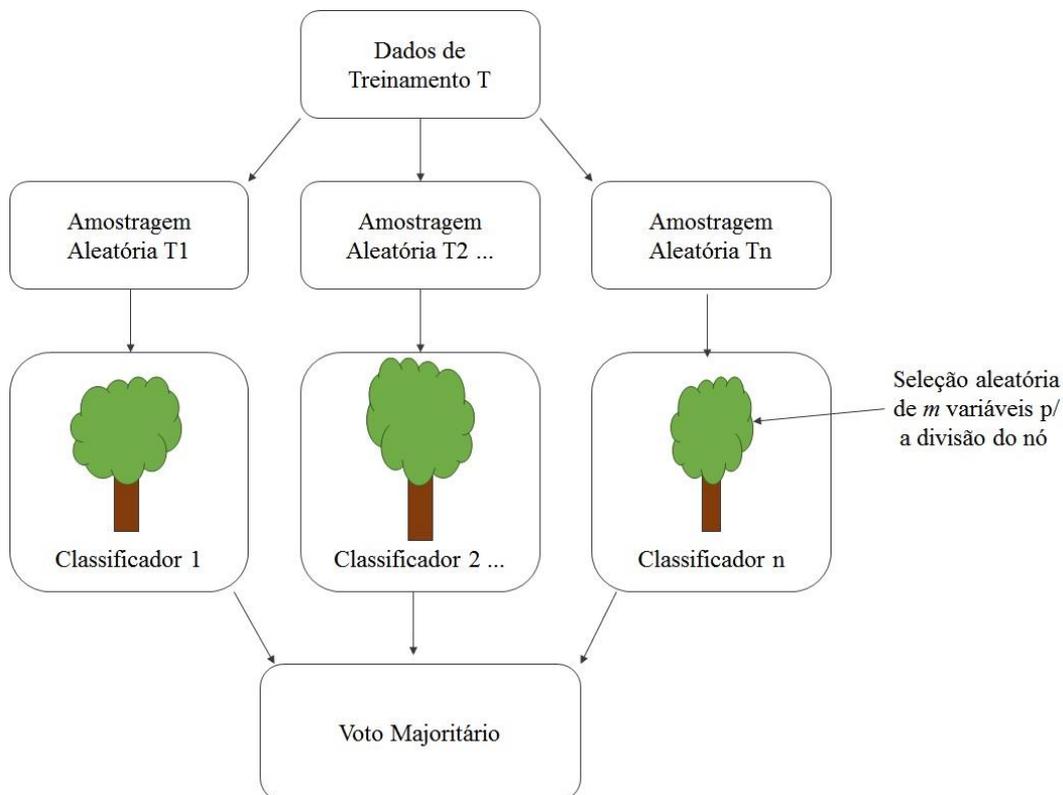
A predição espacial de classes de solos usando métodos baseados em *Random Forest* (RF) vem sendo amplamente utilizada. Foi aplicada por Chagas et al. (2013), Rad et al. (2014), Taghizadeh-mehrjardi et al. (2015), Were et al., (2015), Yang et al., (2016), Jeong et al., (2017), Pinheiro et al. (2017) Meier et al. (2018) e Jeune et al. (2018), sempre com desempenho robusto e satisfatório. O método corresponde a uma evolução do modelo CART proposto por Breiman et al. (1984) e constitui-se, bem como o nome sugere, em uma floresta aleatória composta por um conjunto maior de árvores de decisão construídas a partir de amostras de “*bootstrap*”, onde cada árvore depende dos valores de um vetor aleatório amostrado de forma independente e com a mesma distribuição para todas as árvores na floresta (Breiman, 2001).

Gera vários classificadores que utilizam muitos modelos de Árvores de Decisão e agregam seus resultados (“*ensemble learning*”) o que resulta no aumento das combinações de inúmeras árvores e aumenta o poder preditivo. Funcionam analogamente a uma floresta, têm a capacidade de aprendizado e reconhecimento de padrões, a partir de exemplos que subsidiam o treinamento e conferem capacidade de predição para análises futuras. Cada nó separa um subconjunto aleatório de variáveis, e as Árvores ficam sobrepostas (não podadas), para que o valor previsto no final corresponda a média dos valores das árvores, no caso de regressão e o mais votado no caso de classificação, onde a predição (resultado) agrega as predições de  $n$  árvores. Isso possibilita a hierarquização da importância de cada variável (Strobl et al., 2007).

O modelo pode ser aplicado em regressão ou classificação, ele escolhe um subconjunto aleatório de amostras, a partir do conjunto de treinamento, sempre mantendo a distribuição das classes - treino e validação. Esse processo requer a definição de alguns parâmetros, como o número de covariáveis escolhidas aleatoriamente para cada árvore e o número de árvores na floresta (Liaw e Wiener 2009). O método dispensa o processo de validação cruzada ou separação de conjunto independente de validação para se obter estimativa confiável do erro das predições. Ele faz uma validação interna durante o processo separando aleatoriamente 36% das amostras para calcular o erro (*OOB - Out of Bag error*), onde retira-se uma variável aleatoriamente para a validação. Cada grupo de amostras gera uma árvore que contém um subconjunto aleatório de covariáveis para construir o modelo,

indicando a(s) mais importante(s), evitando assim elevado custo computacional e a possibilidade de *overfitting* (Breiman, 2001).

A Figura 2 mostra um fluxograma do funcionamento do modelo.



**Figura 2** Funcionamento do modelo *Random Forest*

O *Random Forest* também usa o índice *Gini*, emprestado do programa CART, como regra para subdivisão dos nós. O modelo fornece a importância das covariáveis e gera uma matriz de confusão da classificação, onde cada variável tem sua importância associada a uma única classe. Como vantagens podemos destacar: o modelo dispensa a poda das árvores; a acurácia e importância das covariáveis são geradas automaticamente; não tem risco de “*Overfitting*”; também não é sensível a *outliers* no treinamento; é de fácil configuração dos parâmetros (Liaw e Wiener, 2009).

### 2.3 Métodos de Prospecção de Solos: Extração e Aproveitamento de Dados Legados

De forma semelhante a prospecção de dados através de áreas piloto, ou áreas de referência, o aproveitamento de dados legados de mapeamentos prévios, tem como objetivo caracterizar as relações solo-paisagem de uma área, que seja representativa da área que se pretende mapear. Para tal, as principais unidades de solo são identificadas e estabelecidas em termos de pedoambientes, as quais irão conter um finito número de classes de solos formando um padrão reconhecível e repetidamente distribuído na paisagem, e a partir destes a distribuição dos solos de uma área semelhante pode ser predita. O aproveitamento dos dados legados para a expansão da área mapeada baseia-se método de áreas de referência descrito por Favrot (1989) e por Lagacherie et al. (1995).

Conforme destacado por Silva et al. (2014), os mapas são feitos com as melhores ferramentas e dados disponíveis no momento de sua criação, mas devem ser atualizados

quando mais informações forem adquiridas. Deste ponto de vista, os mapas de solo no Brasil, dos quais a maioria foi criada antes do advento das técnicas de MDS, poderiam ser refinados com as atuais ferramentas de mapeamento e neles inseridos novos dados, coletados ao longo dos últimos anos. Muitas iniciativas relacionadas ao mapeamento digital permitem recuperar esse conhecimento dos mapas legados e correlacioná-los com fatores ambientais, tais como geologia, topografia e vegetação (Taghizadeh-Mehrjardi et al., 2015; Samuel-Rosa et al., 2015).

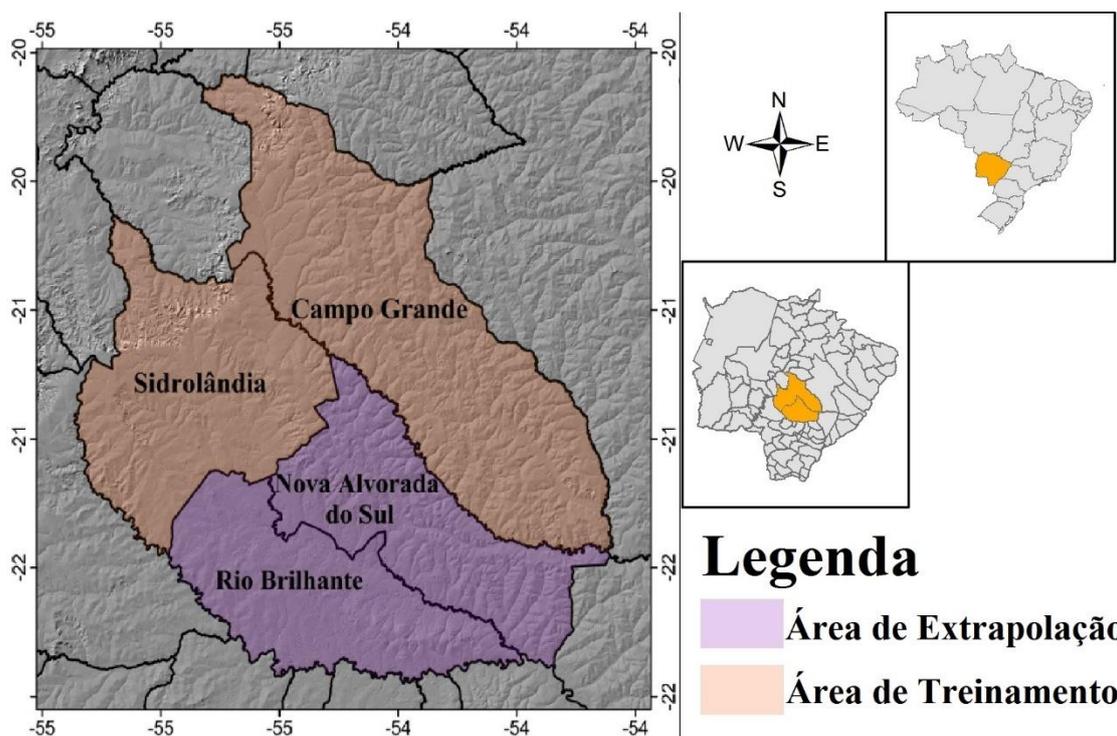
Mapas legados vêm sendo usados como informação básica para o treinamento de modelos preditivos de classes ou propriedades de solos, para extrapolar as relações solo-paisagem de áreas previamente conhecidas para áreas adjacentes fisiograficamente semelhantes (Grinand et al., 2008; Sarmiento et al., 2014; Silva et al., 2013; Höfig et al., 2014). Essa extrapolação, para outra área na região onde essas relações ainda não são conhecidas, pode ajudar a reduzir o tempo e o custo demandado nas prospecções de solo, além de dar subsídios para mapeamentos futuros. Pode-se então, através de técnicas de mapeamento digital, facilitar e acelerar a investigação de solos em outras áreas da mesma região que apresente condições naturais semelhantes (Favrot, 1989; Lagacherie et al., 1995).

O uso de dados legados no mapeamento digital de solos associados a métodos preditivos pode ser observado nos estudos de Bagatini et al. (2016) e Grinand et al. (2008), que aplicaram o método de árvores de decisão; Kempen et al. (2009) através de regressão logística; e Yigini e Panagos (2014) usando regressão e krigagem. O método de prospecção baseado em áreas de referência permite aproveitar dados de levantamentos pré-existentes para predição da distribuição espacial de unidades de solos, em áreas adjacentes e com condições ambientais semelhantes, porém ainda não mapeadas. Desta forma, constitui uma alternativa para otimizar os recursos empregados nos levantamentos pedológicos, disponibilizando além do mapa de solos a avaliação da incerteza e modelos numéricos do terreno, que caracterizam a paisagem; por sua vez, esses dados também servem de subsídio para mapeamentos futuros.

### 3 MATERIAL E MÉTODOS

#### 3.1 Caracterização da Área

A área de estudo compreende os municípios de Sidrolândia, Campo Grande (Áreas de Referência), Nova Alvorada do Sul e Rio Brillhante (Áreas de Extrapolação), no estado do Mato Grosso do Sul (MS), conforme mostra a Figura 3. A área total dos quatro municípios é de aproximadamente 21.389 km<sup>2</sup> e está localizada entre as coordenadas 20°9'57''S e 55°19'50''W; 21°57'58''S e 53°36'2''W; com elevação variando entre 194m e 722m acima do nível médio do mar (WGS 1984).



**Figura 3.** Localização da área de estudo no Estado do Mato Grosso do Sul, com indicação de área de treinamento e de extrapolação.

A caracterização da área, bem como, os perfis de solos utilizados no treinamento dos algoritmos foi obtida a partir dos levantamentos preexistentes na região, que incluíram os dados e relatórios da Folha SF.21 Campo Grande, volume 28 do mapeamento do projeto RADAMBRASIL (BRASIL, 1982) e dos Levantamentos de reconhecimento de baixa intensidade dos solos dos municípios de Campo Grande (EMBRAPA, 2009) e de Sidrolândia (EMBRAPA, 2010), ambos pertencentes a segunda fase do Zoneamento agroecológico do Estado do Mato Grosso do Sul – ZAE-MS. O qual contempla área que compreende os municípios de Sidrolândia, Campo Grande, Nova Alvorada do Sul e Rio Brillhante, situados no estado do Mato Grosso do Sul.

#### 3.2 Fatores de Formação

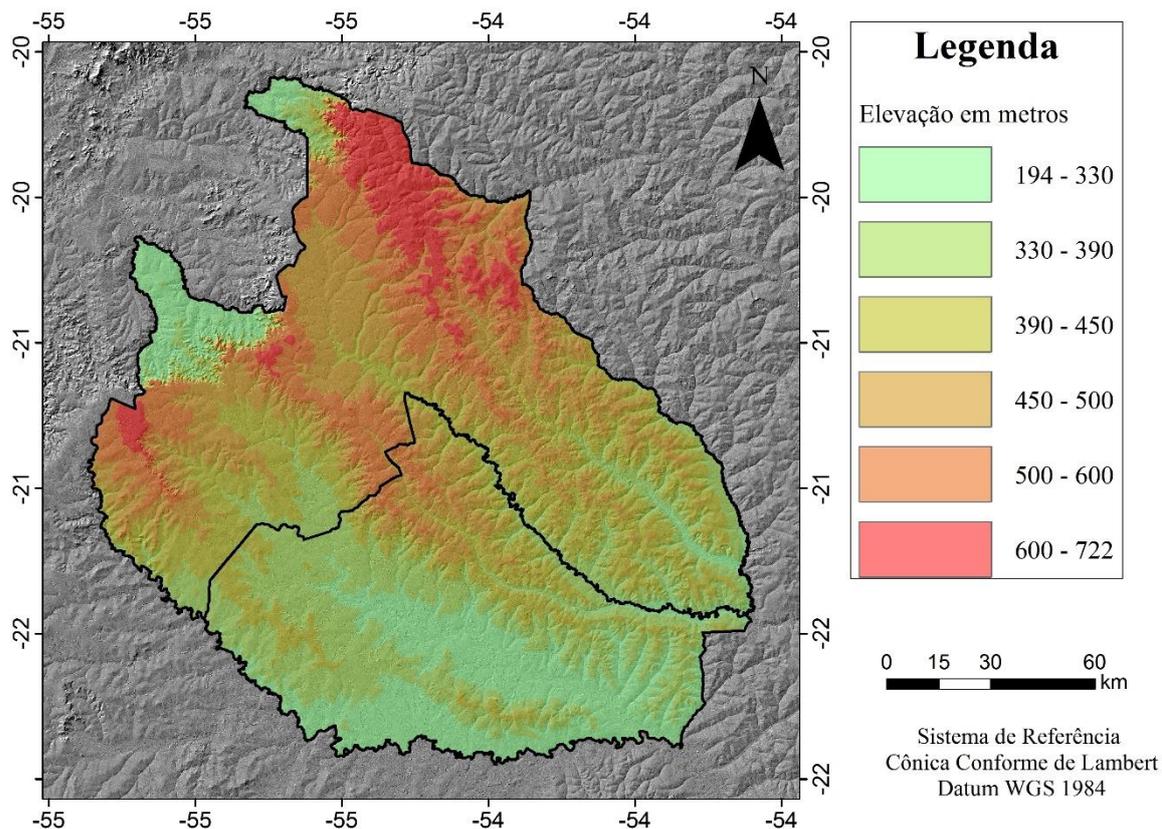
A partir da relação entre solo e paisagem podem ser estabelecidos padrões de correspondência entre unidades de solos com diferentes formas do terreno e covariáveis morfométricas. O reconhecimento destes padrões depende da organização e agrupamento de

variáveis com comportamento semelhante, para sua identificação, delimitação e classificação. Para a aplicação dos modelos preditivos ao mapeamento de unidades de mapeamento, covariáveis morfométricas foram obtidas a partir de MDE, dados de sensoriamento remoto espectral (SRE) e mapas categóricos legados. As variáveis usadas para representar os fatores de formação subsidiaram a identificação dos padrões da paisagem e condições de ocorrência dos tipos de solos, tendo como base o modelo conceitual de gênese de solos, apresentado por Jenny (1941). Os principais fatores de formação serão descritos nos itens subsequentes.

### 3.2.1 Relevo

O relevo da região é diversificado, porém é dominado por grandes planaltos de morfoestruturas complexas, circundados por extensas depressões. A unidade geomorfológica é o Planalto de Maracaju-Campo Grande, onde foram caracterizadas duas unidades estratigráficas: Formação Serra Geral, do Grupo São Bento e Formação Bauru (BRASIL, 1982). O Planalto de Maracaju-Campo Grande constitui extensa superfície suavemente dissecada, com predomínio de formas tabulares muito amplas e trechos conservados. Secundariamente ocorrem formas de dissecção de topo convexo, em geral relacionadas à presença de drenos mais importantes ou às bordas do planalto. Na área de estudo observam-se formas dissecadas de grande amplitude, aliadas a formas conservadas, de grande extensão. Corresponde à terminação sul da borda ocidental da bacia sedimentar do Paraná, separando parte das águas que fluem para o rio Paraguai daquelas que fazem a rede de afluentes do rio Paraná (BRASIL, 1982).

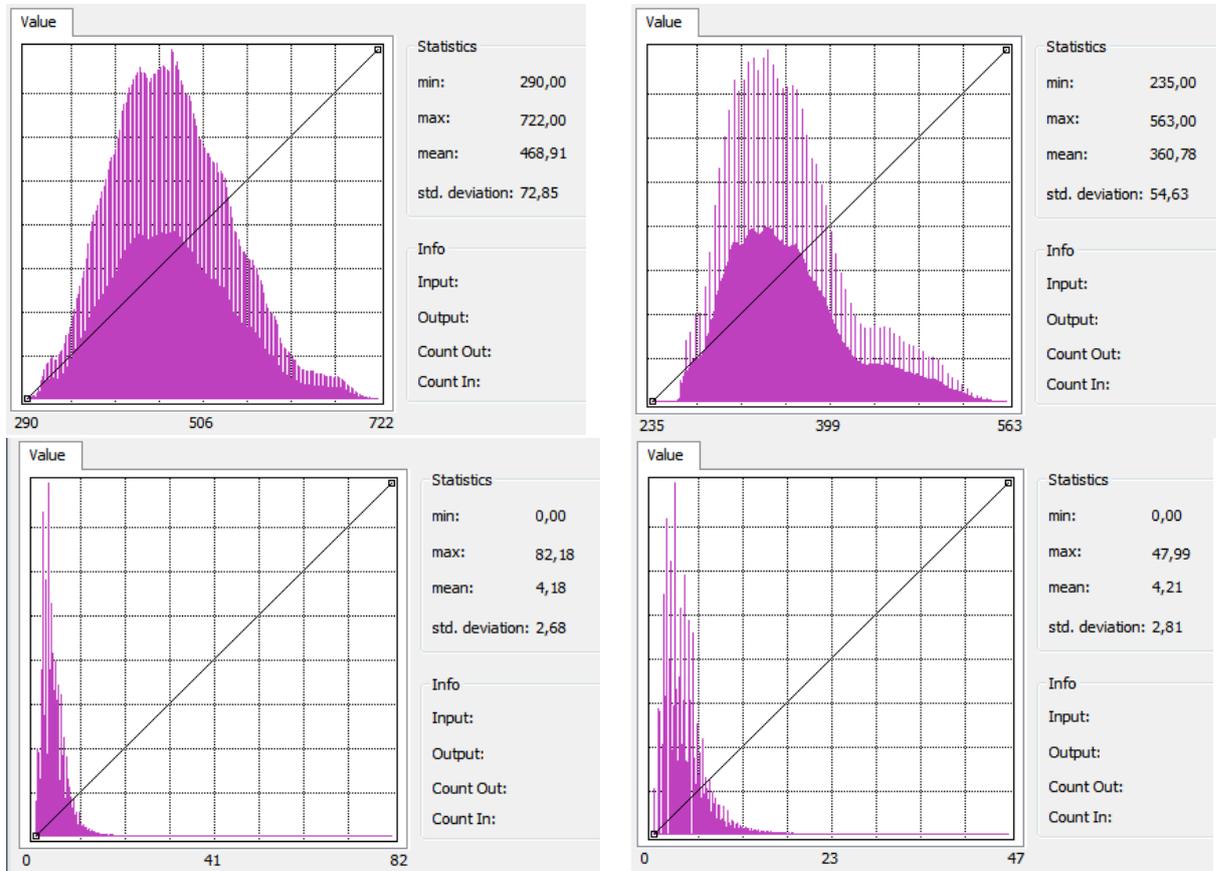
Para representar o relevo foram usadas covariáveis morfométricas derivadas do MDE, tais como a Elevação (Figura 4), que possuem relação direta com o formato e posição na paisagem e, portanto, têm potencial para representar numericamente esse fator.



**Figura 4** Modelo Digital de Elevação para os municípios de Sidrolândia, Campo Grande,

Nova Alvorada do Sul e Rio Brilhante, no Estado do Mato Grosso do Sul.

Para esse estudo aproveitou-se apenas uma parte representativa do relevo, descontando-se as áreas que ocorrem dentro da zona de depressão que diferem da área a ser extrapolada. Ambas as áreas, de treinamento e de extrapolação, apresentaram distribuição, média e desvio padrão semelhantes, no que se refere a elevação e declividade. O que é ilustrado nos histogramas a seguir (Figura 5).

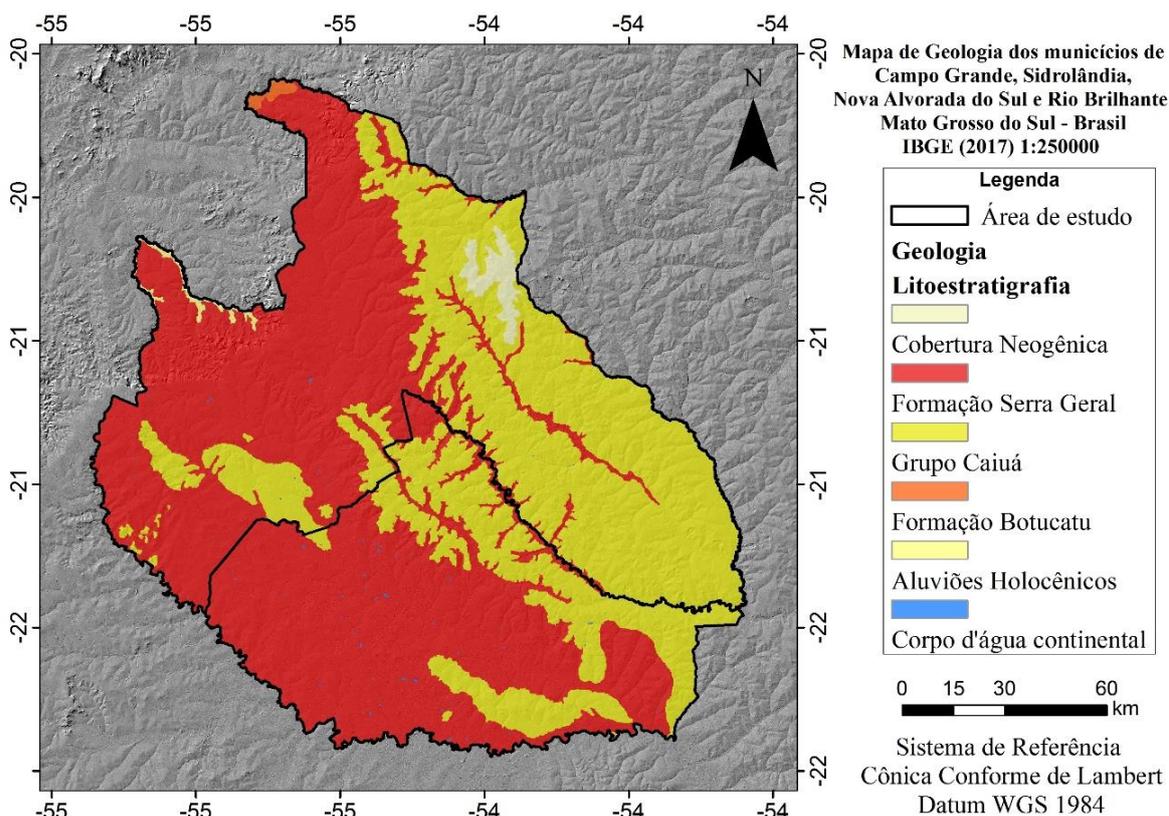


**Figura 5** Histogramas comparativos da Elevação e Declividade das áreas de treinamento e extrapolação.

### 3.2.2 Material de origem e tempo (Litoestratigrafia)

Outra covariável usada no estudo foi o Mapa Temático de Geologia para o Estado do Mato Grosso do Sul, na escala de 1:250.000 (Figura 6) e o respectivo relatório técnico (IBGE, 2017). Segundo essas fontes, o material de origem na área de treinamento e na área de extrapolação apresenta ocorrência e padrão de distribuição semelhante.

Tanto o relatório técnico da geologia do Mato Grosso do Sul, Lacerda Filho et al. (2006), confeccionado pela CPRM, quanto o relatório final do Projeto RADAMBRASIL Folha S.F 21 (1982) afirmam que a estruturação geológica do Mato Grosso do Sul é formada por três compartimentações geotectônicas distintas: Cráton Amazônico, Província Tocantins (Faixa Paraguai) e Bacias Sedimentares Fanerozóicas, para além da sedimentação recente (Cenozoica). Assim, foi escolhida como base a classificação do relatório técnico da CPRM (Lacerda Filho et al., 2006), pois, além de coincidir com o IBGE, é mais atual que o Projeto RADAMBRASIL e não tem foco na geologia evolutiva como a proposta por Milani (1994, 2007).



**Figura 6** Mapa de geologia dos municípios de Sidrolândia, Campo Grande, Nova Alvorada do Sul e Rio Brillhante no Estado do Mato Grosso do Sul (IBGE, 2017)

A análise da geologia possibilita distinção das unidades de mapeamento e pode otimizar o ajuste do modelo preditivo. Tal ocorre, já que as unidades litoestratigráficas com natureza mineral marcável distinta podem subsidiar a distinção dos tipos de solos oriundos dos basaltos (com elevados teores de minerais ferromagnesianos) e arenitos (onde predominam quartzo e feldspatos), formando solos com textura mais argilosa ou arenosa, respectivamente.

A área de estudo situa-se na porção da Bacia do Paraná (Fanerozóica: Paleo-Mesozóica), onde foram caracterizadas unidades geológicas intercaladas do Grupo São Bento (Formações Botucatu e Formação Serra Geral) e do Grupo Caiuá. A Bacia do Paraná é uma ampla região sedimentar-magmática do continente Sul-americano onde um sem-número de unidades foram formalizadas em estudos da bacia para descrever o arcabouço estratigráfico e compreender seu desenvolvimento geológico (Milani, 2007). No Mato Grosso do Sul a bacia ocupa cerca de 250.000 km<sup>2</sup> (Lacerda Filho et. al., 2006).

De forma geral os litotipos da formação Botucatu são arenitos médios a finos avermelhados, ricos em quartzo, dispostos em sets e/ou co-sets de estratificações cruzadas de grande porte, interpretados como de deposição de dunas, numa assinatura faciológica muito característica que possibilita um pronto reconhecimento do “deserto Botucatu” em todos os pontos em que aflora (Milani et al., 2007).

A Formação Serra Geral, é caracterizada por rochas de natureza vulcânica, dominadas por basaltos toleíticos e andesitos basálticos, ocorrendo subordinadas quantidades de riolitos e riolacitos (Peate et. al, 1992). Encontra-se sobreposta aos arenitos eólicos da Formação Botucatu, e sotoposta as rochas sedimentares dos Grupos Bauru e Caiuá ou depósitos

sedimentares inconsolidados do Cenozóico (Machado et al, 2009). Os minerais essenciais do basalto compreendem labradorita e augita, a 10% de magnetita e outros opacos, e traços de apatita (Lastoria, 2002).

O Grupo Caiuá é composto de arenitos arcoseanos vermelhos a roxos, bimodais (muito finos e grossos), com grãos angulosos a subarredondados, na fração fina e, arredondados na fração grossa, com esfericidade variável. A estrutura sedimentar predominante são estratificações cruzadas tangenciais de pequeno a grande porte e, secundariamente, laminação planoparalela. (Lacerda Filho, 2006). O ambiente de deposição é interpretado como fluvial na base e eólico no topo. É o registro sedimentar mesozóico posterior ao vulcanismo Serra Geral considerados como o registro final da Bacia do Paraná.

As observações de campo na Folha SF.21 Campo Grande (BRASIL, 1982) não mostraram relações de contato nítidas devido ao grau de desagregação dos arenitos da Formação Bauru que comumente estão misturados com basaltos alterados da Formação Serra Geral inferior.

Para treinamento dos modelos foram usados dados espectrais obtidos pelo sensor remoto LANDSAT 8 para estimar índices espectrais com relação direta ao material de origem, tais como: *Iron Oxide* e *Clay Minerals* enquanto o *Grain Size Index*, apresenta boa correlação com a textura do solo, que por sua vez é consequência direta do tipo de material de origem (Chagas et al., 2018) e tem suas sintaxes apresentadas abaixo (Equações 2, 3 e 4). Tais índices vêm sendo usados no sensoriamento remoto geológico para distinguir propriedades da superfície, possibilitando distinguir entre propriedades físicas e mineralógicas (Sabins, 1997).

$$\text{Iron Oxide Index} = \text{banda 4}/\text{banda 2} \quad (\text{Eq. 2})$$

$$\text{Clay Minerals Index} = \text{banda 6}/\text{banda 7} \quad (\text{Eq. 3})$$

$$\text{Grain Size Index} = (b4-b2)/(b2+b3+b4) \quad (\text{Eq. 4})$$

### 3.2.3 Clima

Em estudo para o zoneamento climático do café no estado do Mato Grosso do Sul, Alfonsi et al. (2002) identificaram a área do estudo como uma zona de transição climática, com a atuação de diversas massas de ar, o que implica em contrastes térmicos acentuados, tanto espacial quanto temporalmente. A região está em uma zona de encontro de diversas massas que atuam no território brasileiro.

A classificação climática situa-se na faixa de transição entre “Cfa” mesotérmico úmido sem estiagem – em que a temperatura do mês mais quente é superior a 25°C, tendo o mês mais seco mais de 30 mm de precipitação e o sub-tipo “Aw” segundo classificação de Köppen (1948), descrita como: clima tropical, com inverno seco. Apresenta estação chuvosa no verão, de novembro a abril, e nítida estação seca no inverno, de maio a outubro (julho é o mês mais seco). A temperatura média do ar do mês mais frio é superior a 18°C. As médias pluviométricas são superiores a 750 mm anuais, atingindo 1.800 mm, na estação chuvosa (novembro a abril). Apresenta estação seca que varia de 3 a 4 meses e estende-se entre os meses de maio a setembro, onde os totais pluviométricos médios são inferiores a 50 mm (Alfonsi et al., 2002).

Devido à dificuldade de obter informações climáticas detalhadas na área de estudo e o seu tamanho reduzido considerou-se para esse estudo que o clima em toda a área é homogêneo e, portanto, optou-se pela não utilização das covariáveis desse fator. Entretanto, como podem haver variações microclimáticas associadas a feições específicas de relevo e

material de origem, com conseqüentes variações de vegetação (organismos), esses fatores serão observados conjuntamente através da resposta espectral de bandas do satélite LANDSAT 8 e os índices derivados.

### 3.2.4 Organismos (Vegetação)

Na área foram identificadas as unidades fisionômico-florísticas de Savanas (Cerrado) e Floresta Estacional Semidecidual. A análise da vegetação foi baseada no Mapa Temático de Vegetação para o Estado do Mato Grosso do Sul na escala de 1:250 mil (IBGE, 2017). Foram contempladas as duas regiões fitoecológicas, bem como as áreas de tensão ecológica. Também foram estimados através da interpretação das imagens de sensoriamento remoto espectral (dados SRE), os índices NDVI e SAVI através das bandas do vermelho (banda 4) e do infravermelho-próximo (banda 5) do LANDSAT 8. Esses índices quantificam além da vegetação (fator organismos), a disponibilidade de água no solo (fator clima e relevo). Em seus estudos Yang et al. (1997) destacaram relações entre índices derivados de dados de sensor remoto e regimes de temperatura, pluviosidade e características físicas e morfológicas do solo.

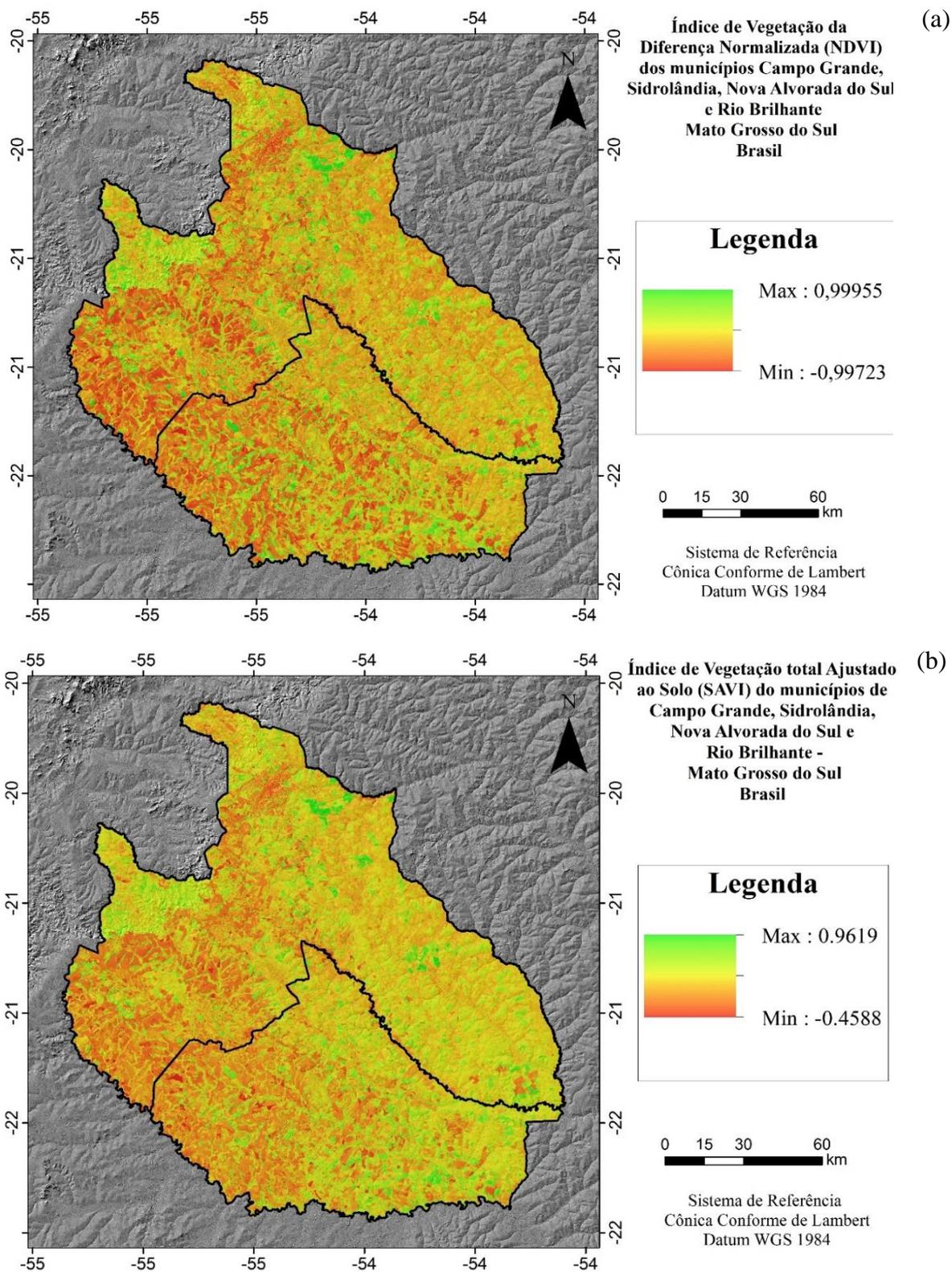
Em estudos ambientais é comum o uso de imagens para a classificação e estimativa da vegetação onde índices espectrais obtidos a partir das bandas espectrais do satélite através das sintaxes apresentadas abaixo, onde a equação (1) representa o índice de vegetação da diferença normalizada - NDVI, (2) representa o índice de vegetação ajustado ao solo – SAVI.

$$NDVI = (banda\ 5 - banda\ 4) / (banda\ 5 + banda\ 4) \quad (Eq.\ 5)$$

$$SAVI = [(1+L) (banda\ 5 - banda\ 4)] / banda\ 5 + banda\ 4 + L \quad (Eq\ 6)$$

Onde o fator L varia entre 0 e 1, de acordo com a densidade da vegetação, e Ponzoni e Shimabukuro (2009), sugerem o valor para o fator L = 0,5, pois engloba maior variação de condições de vegetação. Portanto, esse valor foi escolhido para o cálculo do SAVI, tendo em vista as características da área de estudo.

Através dos índices vegetação NDVI e SAVI (Figura 7) é possível aumentar o contraste dos tipos de cobertura do solo e vegetação, permitindo identificar as áreas de vegetação mais abundantes (mais verdes no mapa) e aquelas onde a vegetação é mais rala e o solo se encontra exposto (mais vermelhas no mapa).



**Figura 7.** (a) Índice de Vegetação da Diferença Normalizada – NDVI (b) Índice de Vegetação Ajustado ao Solo – SAVI para os municípios de Sidrolândia, Campo Grande, Nova Alvorada do Sul e Rio Brillante, no Estado do Mato Grosso do Sul.

### **a) Região da floresta estacional semidecidual**

Define-se diretamente pela condição climática que apresenta duas estações, uma chuvosa e outra de estiagem acentuada, o que ocasiona a decidualidade parcial de suas espécies vegetais arbóreas, principalmente as que ocupam o dossel superior, as quais têm adaptação à deficiência hídrica. No caso da Floresta Estacional Semidecidual, a porcentagem das árvores caducifólias que compõem o conjunto florestal deve situar-se em torno de 30% na época desfavorável. Esse caráter caducifólio da região fitoecológica proporciona que o chão fique completamente recoberto pelas folhas secas, desprendidas das árvores, formando, no fim da estação seca, um verdadeiro tapete. Apoiando-se nos dados legados de BRASIL (1982), foi possível mapear duas fisionomias da Floresta Estacional Semidecidual, correspondendo às Formações Floresta Estacional Semidecidual Aluvial e Floresta Estacional Semidecidual Submontana, além de formações Pioneiras com Influência Fluvial.

### **b) Região da savana (Cerrado)**

A Savana possui vegetação xeromórfica, com fisionomia diversificada, variando de arbórea densa a gramíneo-lenhosa. Geralmente, caracteriza-se por fanerófitas de porte reduzido, isoladas ou agrupadas sobre um revestimento de gramíneas. Apresenta vegetação lenhosa, composta de brotos foliares bem protegidos, apresentando casca grossa e rugosa, muitas vezes com estruturas de reserva subterrânea. As áreas da Savana não detêm uma composição florística uniforme, contudo suas espécies se repetem com bastante frequência, notadamente aquelas dos gêneros *Qualea* e *Callisthene*, formando uma paisagem visivelmente monótona, o que vem se traduzir em fisionomias bastante repetitivas e cansativas. Apoiando-se nos dados legados do projeto RADAMBRASIL (BRASIL,1982), foi possível mapear quatro fisionomias da Savana, correspondendo às Formações Savana Arborizada com Floresta de Galeria, Savana Arborizada sem Floresta de Galeria, Savana Florestada e Savana Parque sem Floresta de Galeria.

### **c) Áreas de tensão ecológica**

Correspondem a áreas que apresentam um uso intensivo das terras onde a vegetação original não se faz presente, como áreas reflorestadas (eucalipto), áreas agrícolas (pecuária e lavouras), áreas de influência urbana e zonas de contato.

## **3.3 Obtenção e Processamento de Dados**

As covariáveis morfométricas foram derivadas a partir de Modelo de Digital Elevação (MDE) com resolução espacial de 30 m, gerado a partir do SRTM, com o uso do *software* livre SAGA v 7.0.0 - Sistema para Análises Geocientíficas Automatizadas (Conrad et al., 2018, disponível em [www.saga-gis.org](http://www.saga-gis.org)) através das ferramentas disponíveis no módulo “*Terrain Analysis*”.

Como covariáveis foram utilizados 26 atributos morfométricos (Tabela 1) derivados do modelo digital de elevação (MDE) do SRTM, com resolução espacial de 30 m, com o uso do *software* livre SAGA v 7.0.0 - Sistema para Análises Geocientíficas Automatizadas (Conrad et al., 2018), através do módulo Terrain Analysis, com exceção da covariável euclidean distance que foi gerada no ArcGIS v.10 (ESRI, 2010). Esse atributos morfométricos expressam modelos numéricos contínuos em formato matricial e apresentam relação direta com o formato e posição na paisagem e, portanto, têm potencial para serem utilizados para representar numericamente o fator relevo.

**Tabela 1** Ferramentas do SAGA utilizadas para a derivação das covariáveis morfométricas.

<b>Ferramenta</b>	<b>Método</b>	<b>Covariáveis Derivadas;</b>
<i>Terrain Analysis=&gt; Channels =&gt; Valley Depth</i>	Conrad, 2012	<i>Valley Depth;</i>
<i>Terrain Analysis=&gt; Channels =&gt; Vertical Distance to Channel Network</i>	Conrad, 2002	<i>Vertical Distance to Channel Network</i>
<i>Terrain Analysis=&gt; Morphometry =&gt; Relative Heights and Slope Position</i>	Boehner e Conrad, 2008	<i>Slope Height; Valley Depth; Normalized Height; Standardized Height; Mid-Slope Position</i>
<i>Terrain Analysis=&gt; Morphometry =&gt; Slope, Aspect, Curvature</i>	Conrad, 2001	<i>Slope; General Curvature; Plan Curvature; Profile Curvature; Tangential Curvature; Longitudinal Curvature; Cross-Sectional Curvature; Minimal Curvature; Maximal Curvature; Total Curvature; Flow Line Curvature</i>
<i>Terrain Analysis=&gt; Morphometry =&gt; Convergence Index</i>	Conrad, 2001	<i>Convergence Index</i>
<i>Terrain Analysis=&gt; Morphometry =&gt; Curvature Classification</i>	Conrad, 2001	<i>Curvature Classification</i>
<i>Terrain Analysis=&gt; Hydrology =&gt; SAGA Wetness Index</i>	Boehner e Conrad, 2001	<i>Catchment Area; Catchment Slope; Modified Catchment Area; Topographic Wetness Index</i>
<i>Terrain Analysis=&gt; Hydrology =&gt; Stream Power Index</i>	Conrad, 2003	<i>Stream Power Index</i>
<i>Terrain Analysis=&gt; Morphometry =&gt; Morphometric Features</i>	Conrad, 2013	<i>Morphometric Features</i>
<i>Terrain Analysis=&gt; Morphometry =&gt; Multi-Scale Topographic Position Index (TPI)</i>	Conrad, 2016	<i>Topographic Position Index</i>
<i>Spatial Analyst Tools =&gt; Distance =&gt; Euclidean Distance*</i>	Default do Arcgis	<i>Euclidean Distance</i>

Os dados espectrais foram obtidos a partir das cenas adquiridas do site do *United States Geological Survey - USGS* (<https://earthexplorer.usgs.gov/>). São imagens do satélite LANDSAT 8 (lançado em novembro de 2013), na altitude de 705km e em órbita circular, polar, heliossíncronal, com inclinação de 98,2° e tempo de duração igual a 99 min, contando com os sensores OLI e TIRS. As características dos sensores é apresentada na Tabela 2.

**Tabela 2.** Especificações técnicas dos sensores LANDSAT 8 OLI TIRS (USGS, 2013).

<b>Banda</b>	<b>Resolução Espectral</b>	<b>Resolução Espacial</b>	<b>Utilidade no Mapeamento</b>
Banda 1 – Ultra Azul	0,435 - 0,451 $\mu\text{m}$	30 m	Estudos de Aerossóis e Costeiros
Banda 2 – Azul	0,452 - 0,512 $\mu\text{m}$	30 m	Mapeamento batimétrico, distinção de solo, vegetação decídua e vegetação de coníferas
Banda 3 – Verde	0,533 - 0,590 $\mu\text{m}$	30 m	Enfatiza a o pico da vegetação, que é útil para avaliar o vigor das plantas
Banda 4 – Vermelho	0,636 - 0,673 $\mu\text{m}$	30 m	Discrimina as encostas da vegetação
Banda 5 – Infravermelho Próximo (NIR)	0,851 - 0,879 $\mu\text{m}$	30 m	Enfatiza o conteúdo de biomassa e linhas costeiras
Banda 6 – Infravermelho Termal (SWIR1)	1,566 - 1,651 $\mu\text{m}$	30 m	Discrimina o teor de umidade do solo e/ou da vegetação; penetra nuvens finas
Banda 7 – Infravermelho Médio (SWIR2)	2,107 - 2,294 $\mu\text{m}$	30 m	Melhora a estimativa da umidade do solo e/ou vegetação e penetra nuvens finas
Banda 8 – Pancromática	0,503 - 0,676 $\mu\text{m}$	15 m	Definição de imagem mais nítida
Banda 9 – Cirrus	1,363 - 1,384 $\mu\text{m}$	30 m	Melhor detecção de contaminação por nuvens cirrus
Banda 10 - TIRS 1	10,60 – 11,19 $\mu\text{m}$	100 m	Mapeamento térmico e estimativa da umidade do solo
Banda 11 - TIRS 2	11,50 - 12,51 $\mu\text{m}$	100 m	Aprimora o mapeamento térmico e estima a umidade do solo

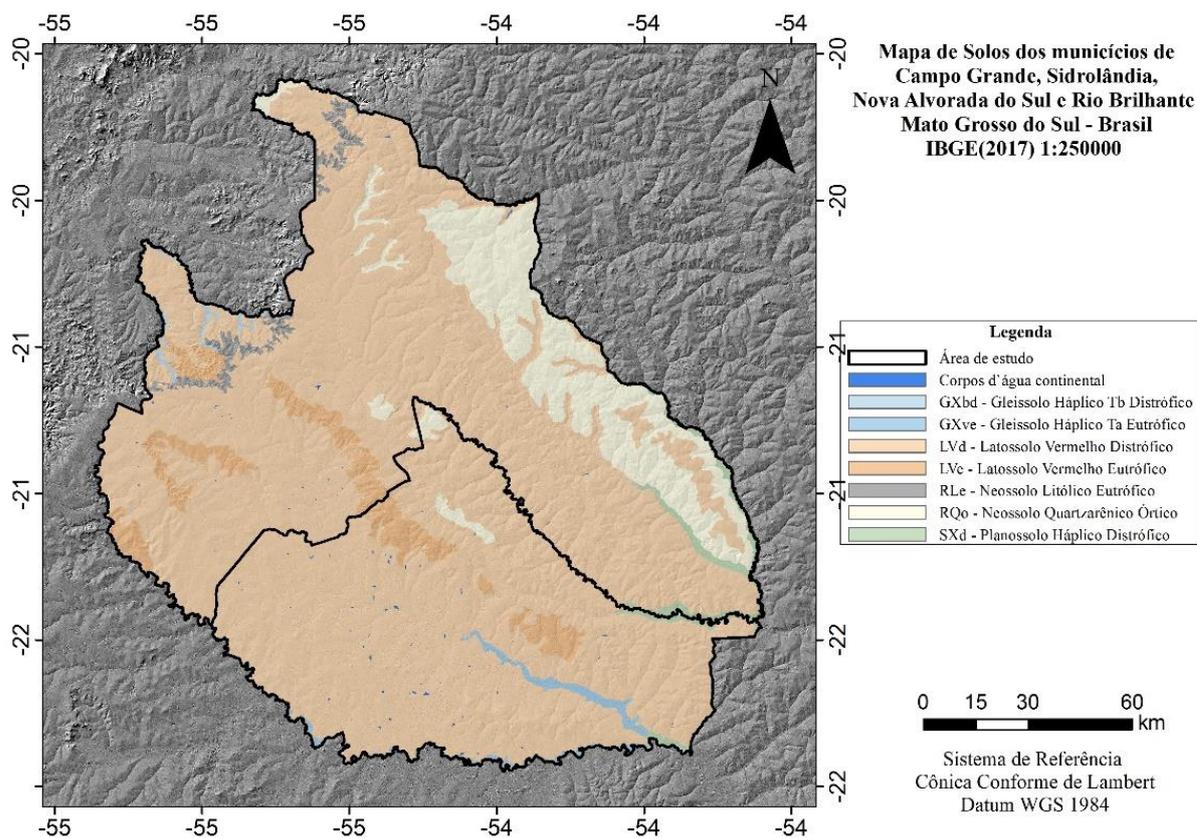
A escolha por imagens obtidas pelo LANDSAT 8 se deu pela disponibilidade gratuita das cenas, além das inovações no satélite, destacando-se a resolução radiométrica de 16 bits, com ganho quali-quantitativo em função da maior precisão radiométrica registrada em comparação a imagens de outros sensores, que costumavam variar entre 8 e 11 bits. Em comparação aos sensores dos LANDSAT anteriores, o sensor OLI também apresenta redução nas larguras espectrais de todas as bandas, com destaque para a banda do infravermelho próximo o que acarreta ganho na interpretação das imagens (USGS, 2013).

Também foram usados dados espectrais obtidos do sensor remoto LANDSAT 8, cenas 2240/74, 2240/75, 2250/74 e 2250/75 todas de agosto de 2017. Primeiramente, as imagens passaram por correções radiométricas para a eliminação de ruídos e correção atmosférica. A correção atmosférica foi feita através do modelo de transferência radioativa na atmosfera usando o 6S (Second Simulation of Satellite Signal in the Solar Spectrum), desenvolvido por

Vermote et al. (1997) e adaptado por Antunes et al., (2012). A partir das imagens corrigidas foram estimados índices espectrais, que possuem alguma relação com os fatores de formação do solo (material de origem, tempo, clima e organismos).

Dessa maneira foram derivados os índices: Clay Minerals (banda 6/banda 7) (Xiao et al., 2006; Chagas et al., 2018), que apresentam relação com o material de origem e tempo; e os índices NDVI  $((\text{banda } 5 - \text{banda } 4)/(\text{banda } 5 + \text{banda } 4))$  e SAVI  $([(1+L) (\text{banda } 5 - \text{banda } 4)]/ \text{banda } 5 + \text{banda } 4 + L)$ . Esses índices quantificam além da vegetação (fator organismos), a disponibilidade de água no solo (fatores clima e relevo). Em seus estudos Yang et al. (1997) destacaram relações entre índices derivados de dados de sensor remoto e regimes de temperatura, pluviosidade e características físicas e morfológicas do solo.

Adicionalmente, utilizou-se como covariável o mapa de Geologia, no qual se observa a ocorrência de basalto da Formação Serra Geral e arenitos da Formação Bauru (Figura 2), e o mapa de solos do Estado do Mato Grosso do Sul, ambos na escala de 1:250.000, e seus respectivos relatórios técnicos (IBGE, 2017).



**Figura 8** Mapa de solos dos municípios de Sidrolândia, Campo Grande, Nova Alvorada do Sul e Rio Brillhante no Estado do Mato Grosso do Sul (IBGE, 2017)

Os atributos derivados tanto do MDE quanto do SRE podem apresentar relações diretas com a pedogênese no que tange aos fatores de formação preconizados por Jenny (1941) e descritos anteriormente (clima, relevo, organismos, material de origem e tempo). Totalizando 41 variáveis (Tabela 3) que expressam modelos numéricos contínuos em formato matricial, possibilitando uma abordagem quantitativa desses fatores.

**Tabela 3.** Fatores de formação e relação com as covariáveis testadas

<b>Fator de Formação</b>	<b>Fonte</b>	<b>Variáveis</b>
Relevo	MDE	<i>Valley Depth; Vertical Distance to Channel Network; Slope Height; Valley Depth; Normalized Height; Standardized Height; Mid-Slope Position; Slope; Aspect; General Curvature; Plan Curvature; Profile Curvature; Tangential Curvature; Longitudinal Curvature; Cross-Sectional Curvature; Minimal Curvature; Maximal Curvature; Total Curvature; Flow Line Curvature; LS factor; Convergence Index; Curvature Classification; Catchment Area; Catchment Slope; Modified Catchment Area; Topographic Wetness Index; Stream Power Index; Morphometric Features; Topographic Position Index e Euclidean Distance</i>
Clima e Organismos		<i>b2; b3; b4; b5; b6; b7; SAVI e NDVI</i>
Material de Origem e Tempo	SRE	<i>b2; b3; b4; b5; b6; b7; Clay Minerals; Iron Oxide e Grain Size Index</i>
	IBGE	Mapa categórico de Geologia

### 3.4 Estratégias de Mapeamento Digital de Solos

O uso de geotecnologias aplicadas à análise e modelagem ambiental permite extrair informações sobre os atributos relevantes na identificação das diferentes classes de solos na área de estudo. Portanto, presume-se que tais métodos de análise podem conferir aos produtos caráter quantitativo ao levantamento de solos, sendo possível estimar o erro associado e desta forma extrapolar a distribuição espacial dos solos através de modelos matemáticos gerados.

As atividades desenvolvidas contemplaram o estudo de técnicas de modelagem espacial preditiva, revisão de literatura e contextualização do estudo, compreensão dos fatores determinantes da gênese dos solos e do modelo solo-paisagem, subsidiado para definir as variáveis discriminantes (ou covariáveis) e algoritmos preditivos adequados para aplicação na área de estudo. A Figura 9 apresenta o fluxograma simplificado dos procedimentos realizados.



**Figura 9.** Fluxograma das atividades de mapeamento digital de solos

### 3.4.1 Compatibilização da base de dados e softwares utilizados

Para que a predição das classes de solo a partir de modelos estatísticos seja assertiva e precisa, é necessário identificar os padrões de ocorrência das classes de solos baseando-se no estudo dos pedoambientes de referência e da avaliação das diferentes feições da superfície. Esses padrões são comparados através de um conjunto de atributos do terreno representados por modelos numéricos matriciais. Como bases para construção dos modelos preditivos, para classes, foram utilizados na análise o conjunto de 26 atributos morfométricos derivados do MDE (primários e secundários), 4 variáveis oriundas de SRE e 2 variáveis categóricas (mapas temáticos), totalizando 32 covariáveis.

A compatibilização dos dados básicos é fundamental pois todos os dados devem estar em mesma projeção geográfica (WGS 1984 Lambert Conformal Conic) e datum (D WGS 1984), mantendo todas em mesma resolução (30m), além dos mapas de geologia, vegetação, e solos, e dados provenientes do MDE e SRE. Todas realizada no programa SAGA - v.7.0.0.

Após a compatibilização dos dados de entrada (*rasters* de *input*) foram analisadas as condições de ocorrência e variabilidade das informações sobre cada camada em cada ponto conhecido (perfis oriundos do levantamento), através da ferramenta “Add Grid Values to Point”, no módulo “Tools; Shapes; Shapes-Grid Tools” do SAGA, o qual pode ser visualizado e trabalhado em planilha, facilitando a interpretação destes dados uma vez que um arquivo “.dbf” é gerado em conjunto com o *shape*.

De posse dessa planilha todas as estatísticas, análises e processamentos dos dados foram realizadas nos programas livres R v.3.4.4 e Rstudio v.1.1.442, (R Core Team, 2017) com o uso dos pacotes: “outliers” (Komsta et al., 2015), “ade4” (Dray e Siberchicot, 2017), “foreign” (R Core Team, 2018), “rpar” (Therneau et al., 2015), “randomForest” (Liaw e

Wiener, 2009), “*rgdal*” (Bivand and Lewin-Koh, 2017), “*sp*” (Pebesma et al., 2018) e “*spsurvey*” (Kincaid et al., 2014).

### 3.4.2 Procedimentos de amostragem

Os pontos coletados e descritos em campo, obtidos a partir de levantamentos de solo prévios na área de estudo, foram selecionados e analisados para excluir pontos não representativos e formar grupamentos de solos com características pedoambientais semelhantes, e com isso determinar as unidades de mapeamento presentes na área. Nessa abordagem foram identificadas um total de 11 classes de solo (considerando até o nível de subgrupo) que foram agrupadas em seis unidades de mapeamento (UM) de acordo com suas características, visando a distinção dos pedoambientes onde se inserem, as quais foram identificadas pela abreviação referente a classe do componente principal de cada unidade. A composição das UMs é apresentada na Tabela 4.

**Tabela 4.** Unidades de Mapeamento das classes de solos identificadas nos municípios de Sidrolândia e Campo Grande - MS

UM	Componentes
GXb	GLEISSOLO HÁPLICO Tb Distrófico típico + GLEISSOLO HÁPLICO Tb Eutrófico típico + (ORGANOSSOLO HÁPLICO Fíbrico + GLEISSOLO MELÂNICO Tb Distrófico organossólico)
LAd	LATOSSOLO AMARELO Distrófico psamítico + LATOSSOLO AMARELO Distrófico plíntico + (GLEISSOLO HÁPLICO Tb Distrófico típico)
LVd	LATOSSOLO VERMELHO Distrófico típico + LATOSSOLO VERMELHO Distrófico psamítico
LVdf	LATOSSOLO VERMELHO Distroférico típico + LATOSSOLO VERMELHO Eutroférico típico + LATOSSOLO VERMELHO Acriférico típico + (NEOSSOLO LITÓLICO Chernossólico)
RQo	NEOSSOLO QUARTZARÊNICO Órtico típico + NEOSSOLO QUARTZARÊNICO Órtico latossólico
SXd	PLANOSSOLO HÁPLICO Distrófico arênico + PLANOSSOLO HÁPLICO Distrófico gleissólico

As seis unidades de mapeamento definidas correspondem as classes de saída (*output*) que foram então utilizadas no mapeamento aplicando as técnicas de MDS. As principais características das unidades de mapeamento são descritas a seguir:

#### **GXb** GLEISSOLO HÁPLICO Tb

**Variações e inclusões:** Unidade composta por GLEISSOLO HÁPLICO Tb Distrófico típico bem como GLEISSOLO HÁPLICO Tb Eutrófico típico, apresentando também inclusão de GLEISSOLO MELÂNICO Tb Distrófico organossólico, geralmente associado a ORGANOSSOLO HÁPLICO Fíbrico.

**Descrição:** Unidade que contempla os solos das áreas de baixada, bem como ao longo dos canais de drenagem, onde a drenagem é muito lenta e o hidromorfismo predomina na

gênese (3 perfis, 2 em Campo Grande 1 em Sidrolândia). Ocorre em relevo plano nas áreas de depressões, principalmente entre as chapadas e próximas aos canais de drenagem onde a drenagem é impedida. São formados predominantemente a partir de sedimentos alóctones e possuem textura variável, de argilosa a arenosa/média dependendo da natureza do sedimento; em áreas onde há aporte de material orgânico podem ocorrer GLEISSOLO MELÂNICO Tb Eutrófico e ORGANOSSOLO HÁPLICO Fíbrico. A cobertura é geralmente de pastagem ou vegetação higrófila de várzea. Tendem a ocorrer em áreas de Sedimentos Quaternários, em relevo plano e cotas baixas no nível de base local.

**Perfil representativo:** PCG-04 (no anexo)

#### **LAd** LATOSSOLO AMARELO Distrófico

**Variações e inclusões:** Unidade composta por LATOSSOLO AMARELO Distrófico plúntico e LATOSSOLO AMARELO Distrófico psamítico, apresentando também inclusões de GLEISSOLO HÁPLICO Tb Distrófico típico em áreas de pequenas depressões presentes ao longo dos planaltos onde a drenagem é mais lenta condicionando uma zona de hidromorfismo mesmo longe dos canais de drenagem.

**Descrição:** Unidade com ocorrência concentrada em áreas de arenito (3 perfis em Campo Grande). Ocorre nas chapadas em relevo plano, ocupam as áreas das chapadas onde o relevo condiciona drenagem moderada ou mais lenta. Desenvolvidos sobre litologia de Arenitos, por vezes com influência de Basaltos, apresentam textura argilosa a média. De uma forma geral, a cobertura original é de cerrado tropical subcaducifólio e estão sob lavouras de cana e soja ou pastagem.

**Perfil representativo:** ECG-08 (no anexo)

#### **LVd** LATOSSOLO VERMELHO Distrófico

**Variações e inclusões:** Unidade composta por LATOSSOLO VERMELHO Distrófico psamítico e LATOSSOLO VERMELHO Distrófico típico (textura média).

**Descrição:** Unidade com ampla ocorrência na área (13 perfis, 6 perfis em Campo Grande e 7 em Sidrolândia), nos topos das chapadas em relevo plano a suave ondulado, bem drenados, profundos, desenvolvidos sobre litologia de Arenito, e apresentam textura média ou mais grossa. De uma forma geral, se encontram sob uso de lavouras anuais, mas também com pastagem plantada.

**Perfil representativo:** PCG-08 (no anexo)

#### **LVdf** LATOSSOLO VERMELHO Distroférico

**Variações e inclusões:** Unidade composta por LATOSSOLO VERMELHO Distroférico típico, LATOSSOLO VERMELHO Eutroférico típico, LATOSSOLO VERMELHO Acriférico típico e LATOSSOLO VERMELHO Distrófico típico (textura argilosa), apresentando também inclusão de NEOSSOLO LITÓLICO Chernossólico típico

**Descrição:** Unidade com ampla ocorrência na área (19 perfis, 7 perfis em Campo Grande e 12 em Sidrolândia), nos topos das chapadas em relevo plano a suave ondulado, bem drenados, profundos, desenvolvidos sobre litologia de Basalto, e apresentam textura argilosa ou mais fina. De uma forma geral, se encontram sob uso de lavouras anuais, mas também com pastagem plantada.

**Perfil representativo:** PS-09 (no anexo)

## RQo NEOSSOLO QUARTZARÊNICO Órtico

**Variações e inclusões:** Unidade composta por NEOSSOLO QUARTZARÊNICO Órtico típico e NEOSSOLO QUARTZARÊNICO Órtico latossólico.

**Descrição:** Unidade com ampla ocorrência (7 perfis em Campo Grande), nos topos das chapadas em relevo plano a suave ondulado, desenvolvidos sobre litologia de Arenitos, apresentando textura arenosa ou mais grossa. De uma forma geral, se encontram em áreas de cerrado tropical subcaducifólio e com uso atual de pastagem. No tangente as características da superfície, estes solos ocorrem em condições muito similares a unidade de mapeamento LVd, porem com teores inferiores de argila.

**Perfil representativo:** PCG-06 (no anexo)

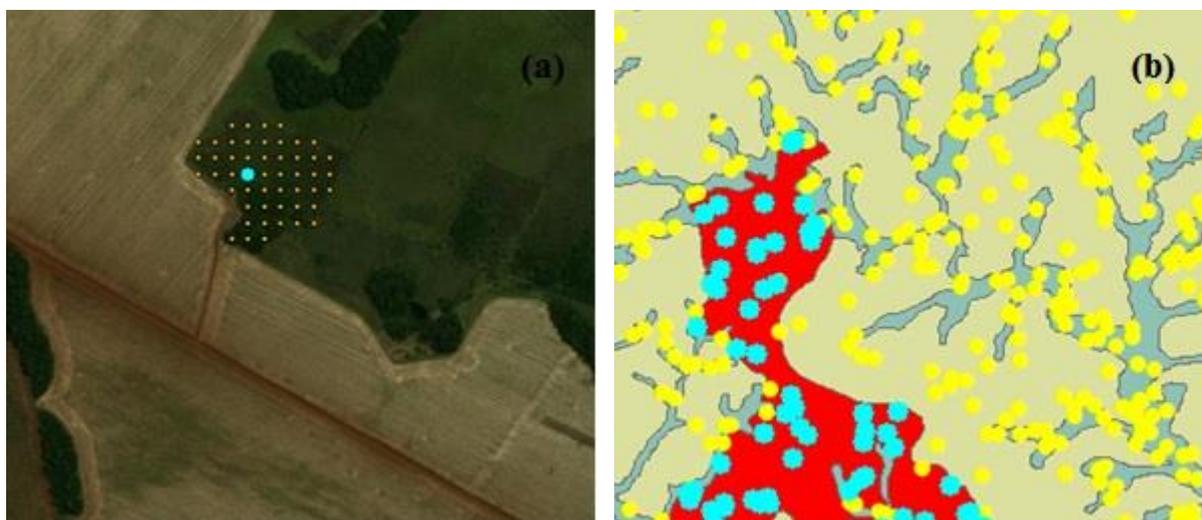
## SXd PLANOSSOLO HÁPLICO Distrófico

**Variações e inclusões:** Unidade composta por PLANOSSOLO HÁPLICO Distrófico arênico e PLANOSSOLO HÁPLICO Distrófico gleissólico.

**Descrição:** Unidade com ocorrência nas áreas próximas as drenagens, geralmente sobre arenito (2 perfis em Campo Grande). Ocorre nas planícies e terço inferior das encostas onde o relevo condiciona drenagem imperfeita. Desenvolvidos sobre litologia de Arenitos, por vezes com influência de Basaltos, apresentam textura arenosa a média. De uma forma geral, a cobertura original é de cerrado tropical subcaducifólio e estão sob lavouras de cana e soja ou pastagem.

**Perfil representativo:** ECG-15 (no anexo)

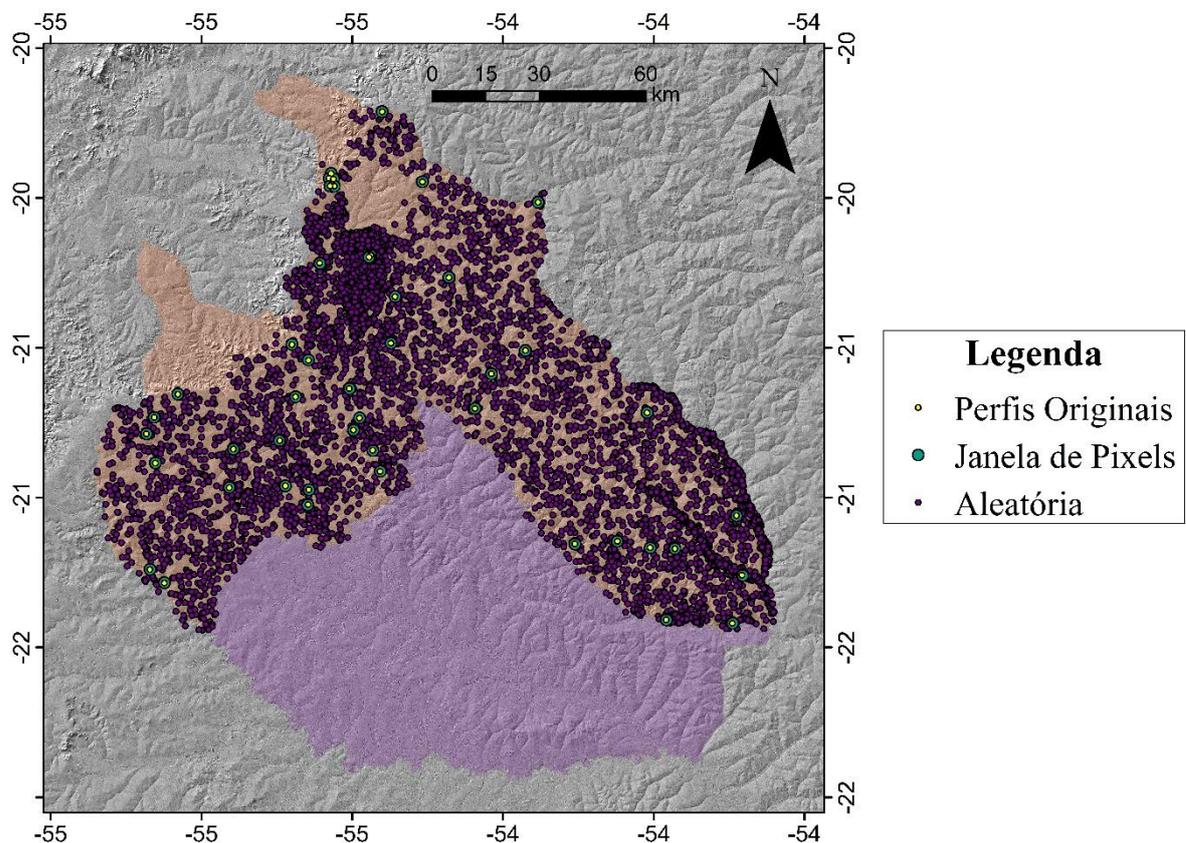
A definição das Unidades de Mapeamento teve como base os dados legados dos Levantamentos de Reconhecimento de Baixa Intensidade dos Solos dos Municípios de Sidrolândia (EMBRAPA, 2010) e Campo Grande (EMBRAPA, 2009), ambos pertencentes à segunda fase do Zoneamento Agroecológico do Estado do Mato Grosso do Sul (ZAE-MS) elaborados na escala 1:100.000, cedidos pela EMBRAPA Solos.



**Figura 10** Obtenção das malhas para o treinamento, (a) por Janela de *pixels* e (b) Aleatória a partir do mapa de referência.

Os mapas desses levantamentos foram compilados e simplificados em um mapa de referência para as unidades de mapeamento encontradas na área de treinamento. A partir desses dados (Pontos de campo e Mapa legado) foram geradas duas malhas amostrais utilizadas para o treinamento dos modelos. Uma malha em janela de *pixels* ao redor dos pontos observados em campo, conforme mostra a Figura 10-a, e outra selecionando *pixels* aleatoriamente dentro das UMs do mapa legado conforme pode ser observado na Figura 10-b.

Definidas as unidades de mapeamento, agrupou-se os perfis representativos de uma mesma UM, onde foi feita contagem das observações de campo para avaliar a necessidade de ampliação do conjunto amostral. Como base, cada UM deveria ter um conjunto com número semelhante de pontos. A partir daí foram criadas janelas de *pixels* ao redor dos perfis observados em campo, onde considerou-se que esses *pixels* próximos aos *pixels* originais e que foram delimitados segundo características semelhantes. Dessa forma, considerou-se que os novos *pixels* também correspondem a unidade de mapeamento do pixel original. Com essa estratégia foi possível aumentar o conjunto amostral para que todas as UMs tivessem número de amostras semelhante.



**Figura 11** Malhas amostrais, Perfis Originais de Campo, Janelas de Pixels e Aleatória a partir do mapa legado

Essa forma de seleção por janelas foi utilizada com sucesso por Pinheiro (2015) e favorece o aprendizado do modelo. Com esse procedimento, o modelo é abastecido por um volume maior de dados, além de possibilitar a separação de conjuntos de treinamento e validação, o que pode ser dificultado em situações com número pequeno de amostras ou caso este seja desbalanceado entre as UM, como ocorreu nesse estudo. Foi determinado então o número de 100 *pixels* para cada unidade de forma a igualar a oportunidade de aprendizado de cada UM pelo modelo, totalizando 600 *pixels*.

Os mapas legados também foram aproveitados para extração de dados pelos modelos, onde, os mapas dos levantamentos dos municípios de Sidrolândia e de Campo Grande, na escala de 1:100.000 realizados nas fazes anteriores do ZAE-MS, foram compilados e analisados. Foram excluídas as UM que não seriam representativas para a área de extrapolação, além da simplificação ou junção de algumas unidades de mapeamento, visando a unificação das legendas dos dois mapas legados. A partir desse mapa compilado, composto pelas seis UM acima descritos, foi feita uma amostragem aleatória de *pixels* pela função “*spsample*” no software R (R Development Core Team, 2017). Foram selecionados 1000 *pixels* para o treinamento considerando cada UM observada, totalizando 6000 *pixels* para o treinamento e aplicação dos modelos preditivos, que foram testados contra os dados originais das observações de campo.

Ambas as malhas amostrais criadas (em janela, e dentro das UMs) foram usadas para o treinamento dos modelos preditivos, e os pontos originais dos perfis foram usados para validação dos modelos. Essa forma foi usada para avaliar a qualidade dos modelos gerados, uma vez que ainda não foram coletados dados na área de extrapolação. Essa estratégia visa obter pontos diferentes para validação e treinamento, para evitar a validação tendenciosa. No caso da janela de *pixels*, essa forma de avaliação utiliza um conjunto de dados para o treinamento muito semelhantes aos dados de validação. Os dados de treinamento permitem que os modelos aprendam as relações entre as variáveis ambientais (dados de entrada) usadas na modelagem e as unidades de mapeamento que se pretende mapear (dados de saída). Durante o treinamento são gerados os modelos e, a partir daí, o algoritmo (modelo) gerado é aplicado sobre o conjunto de validação.

Para a validação foi utilizada a malha original de 46 pontos observados em campo (Figura 11) sendo essa malha a mais confiável a nível de campo para avaliação do desempenho de generalização dos algoritmos aprendidos.

### 3.4.3 Modelagem pedométrica

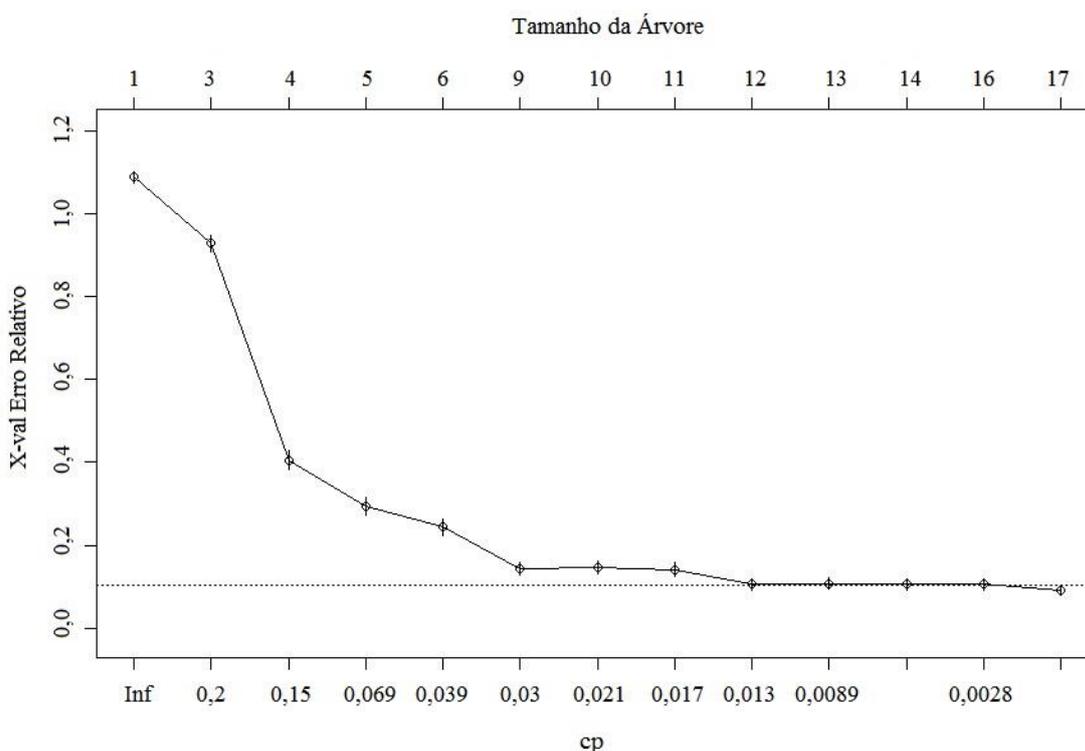
O uso de dados legados baseia-se no princípio de que o conhecimento previamente adquirido sobre uma área semelhante, que possui informações detalhadas e completas, pode ser aproveitado no mapeamento de solos uma vez que facilita a compreensão das relações solo-paisagem e, pode ser útil para o estudo de uma área ainda não mapeada com condições semelhantes quanto às relações solo-paisagem. Para esclarecer as relações solo-paisagem foram utilizadas como referência as áreas dos municípios de Sidrolândia e de Campo Grande, onde foram realizados levantamentos de solos prévios na escala de 1:100.000 oriundos da segunda fase do Zoneamento Agroecológico do estado do Mato Grosso do Sul (ZAE-MS), do qual se aproveitou os dados dos perfis observados para contemplar as variações pedoambientais e incluir todas as unidades do solo. Os dados sobre as variáveis ambientais usadas para alimentar os modelos preditivos provem desses pontos conhecidos das áreas previamente mapeadas.

O conjunto de covariáveis que compoem os dados de entrada nos modelos preditivos foi testado e selecionado a partir de análises de correlação, de forma a selecionar as covariáveis que sejam mais relevantes para o aprendizado da máquina e o ajuste do modelo às condições pedoambientais da área. Esta estratégia permite avaliar o papel de cada variável na predição das unidades de mapeamento do solo.

Todo procedimento de modelagem dos modelos preditivos foi realizado no programa R (R Development Core Team, 2017) através dos pacotes “*rpart*” (AD) (Therneau et al., 2015) e “*randomForest*” (RF) (Liaw e Wiener, 2009).

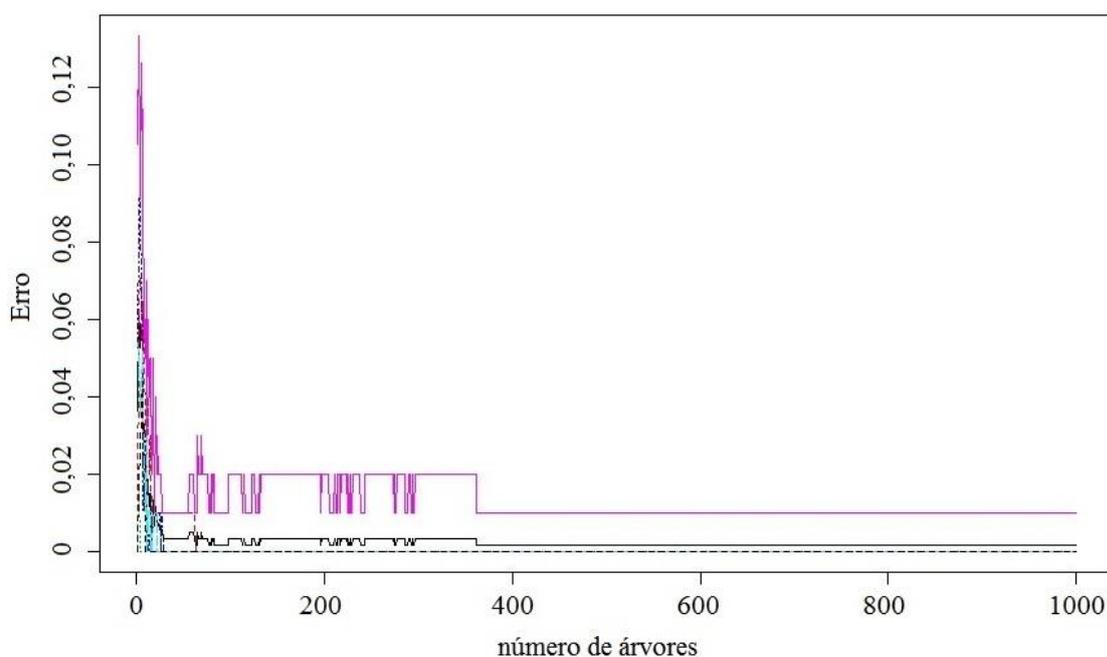
Os modelos preditivos baseados em árvores (AD e RF) foram previamente testados a partir de análises para determinar os melhores parâmetros para a modelagem de forma a torná-la eficiente e assertiva, para ter extrapolações dos dados legados com menor erro possível.

Para o modelo AD o parametro mais importante a ser definido é o tamanho da árvore, que não deve contemplar menos nós do que o necessário, para que sejam contempladas todas as classes de saída, nem tantos nós que torne o modelo complexo demais. As análises no R determinaram que, a partir do valor do parametro de complexidade (cp) 0,001, o tamanho ideal da árvore para esse conjunto de dados foi de 16 nós, uma vez que a partir desse ponto o erro tende a estabilizar, conforme observado na Figura 12.



**Figura 12** Tamanho da árvore (número de nós) em função do parâmetro complexidade (CP)

Para o modelo RF foi necessária a determinação do parametros “*mtry*” e “*ntree*”. O “*mtry*” corresponde ao número de covariáveis usadas em cada árvore. Quanto menor o seu valor, menor a correlação entre as árvores e maior a redução da variância. O “*ntree*” foi determinado a partir de testes que comparam o número de árvores utilizadas e o valor do erro associado. Para esse parâmetro adotou-se então o valor de 400 árvores, uma vez que a partir desse número, o erro fica constante (Figura 13). Já o “*mtry*” foi determinado em 4 conforme o recomendado por Liaw e Wiener (2009) que é a raiz quadrada do número de covariáveis quando a modelagem visa classes. Os resultados, representam a classe com maior probabilidade de ocorrência considerando os resultados de todas as árvores da floresta (Breiman, 2001; Cutler et al., 2009). Além deste, outros parâmetros também podem ser ajustados, como o tamanho de cada árvore (número de nós) “*nsplit*” e a quantidade máxima de nós terminais das árvores “*max nodes*”, entretanto para esses parametros foram utilizados os valores *default* do método.



**Figura 13** Tamanho da floresta em função do erro relativo.

#### 3.4.4 Validação dos modelos preditores

Para validar os modelos preditores, foram usados os pontos originais dos 46 perfis observados em campo e re-classificados segundo o SiBCS (Santos et al., 2018). Um arquivo da base de dados foi gerado a partir dessas informações e um teste "verdadeiro-falso" executado contra o mapa de solos predito. Onde, a informação desses perfis é comparada com a classificação predita.

Durante os eventos de teste, foram feitas as análises estatísticas para cada modelo preditivo abordado, usando os dados de validação. Tais análises geraram o índice Kappa e a exatidão global, obtidas a partir de uma matriz de confusão (Congalton e Green, 1999) e da análise de matriz de significância a partir dos resultados Kappa e variância de cada modelo.

Com base nas análises estatísticas, os melhores resultados obtidos em cada modelo foram comparados entre si para a escolha do mapa de solo. Para generalizar as informações, o arquivo do modelo preditivo escolhido e um arquivo de imagem com os dados de pixel dos arquivos raster foram processados usando o programa R (R Development Core Team, 2017). Nesta fase, o método de predição escolhido foi previamente treinado com os dados legados da fase anterior do levantamento e então usado para processar todos os *pixels* da área extrapolada. O arquivo de imagem final inclui a classificação de cada pixel de acordo com as relações entre as classes de solo e as variáveis ambientais estabelecidas pelo modelo preditivo.

## 4 RESULTADOS E DISCUSSÃO

### 4.1 Pedoambientes

Considerando a variação dos fatores de formação de solos foram destacadas algumas feições que apresentam condições distintas para a pedogênese (Tabela 5). De maneira geral, pode-se observar uma clara diferença entre as condições pedoambientais das chapadas e das depressões (fator relevo).

As chapadas apresentam maior elevação e relevo plano a suave-ondulado, que permite a boa drenagem e promove a alteração do material intemperizado (em função do clima e resistência do material de origem). A vegetação é mais rala com feições típicas do cerrado, onde a fertilidade dos solos é menor, até áreas de transição onde a mata é mais densa (fator organismos). Nas chapadas formam-se solos profundos e, em função do material de origem, se observam LATOSSOLOS VERMELHOS ou NEOSSOLOS QUARTZARÊNICOS.

Nas depressões o relevo é mais movimentado e varia de áreas mais declivosas, onde a remoção mais intensa de partículas favorece o afloramento de rochas e formação de solos rasos como os NEOSSOLOS LITÓLICOS; até áreas planas de baixada, onde o relevo favorece a adição de sedimentos e acúmulo de água, em condições de hidromorfismo formando GLEISSOLOS e ORGANOSSOLOS. Em áreas intermediárias, de relevo mais suave ocorrem solos de melhor drenagem e mais desenvolvidos, mas ainda com algum grau de hidromorfismo, como os PLANOSSOLOS HÁPLICOS Distróficos e mesmo LATOSSOLOS AMARELOS Distróficos plínticos.

O material de origem e o relevo contribuem de forma significativa na pedogênese. O material de origem condiciona a resistência ou vulnerabilidade dos materiais expostos ao intemperismo, que se transformam ao longo do tempo, o que por sua vez também modela a paisagem. O relevo é preponderante na condução da água na superfície e condiciona todo o fluxo de massas e distribuição de partículas ao longo das encostas e vales. Os pedoambientes podem ser separados, em primeira instância, como: sistemas de relevo degradacional, onde são mais intensos os processos de remoção, lixiviação e perda de material; e agradacional, que tende a receber o material removido das partes mais elevadas ou de locais mais distantes.

Os solos em sistemas de relevo degradacional são, em geral, bem drenados e relativamente profundos, ocorrem em feições onde predominam as colinas, morros e morrotes, em distintos materiais de origem, tais como as rochas da Formação Botucatu, Grupo Serra Geral e Grupo Caiuá. Foram identificados nesses sistemas as classes LATOSSOLOS VERMELHOS Distroféricos, relacionados aos corpos intrusivos tabulares (diques de basalto), corroborando a relevância do material de origem na gênese destes solos.

Além desta associação com o material de origem, as condições de relevo e clima favorecem a gênese de solos bem intemperizados e de estrutura bem desenvolvida nas encostas. O clima atua por meio da precipitação pluviométrica, propiciando a transformação dos minerais primários facilmente intemperizados e a translocação de argila. O relevo contribui ao permitir boa drenagem e favorecer perda de frações mais finas em superfície e, em áreas de relevo plano a suave ondulado, intensifica a lixiviação de bases solúveis e o intemperismo. A intemperização do material de origem, com a alteração dos minerais ferromagnesianos e a liberação de íons que depois são precipitados como óxidos de ferro, conduz às cores brunadas a avermelhadas dos solos. Nos arenitos a composição mineral da rocha confere as texturas mais arenosas aos solos e predominam os NEOSSOLOS QUARTZARÊNICOS.

Como sistema de relevo agradacional destacam-se as planícies aluviais próximas aos canais de drenagem, bem como nos vales pouco amplos, onde os solos recebem sedimentos

tanto coluviais como aluviais e têm drenagem imperfeita e mal drenados. Além dos sedimentos o fluxo interno de água influencia as propriedades dos solos dessa feição, em que o relevo é um importante fator, condicionando a influência do lençol freático em distintas épocas do ano. Os solos nesses sistemas foram os: GLEISSOLOS HÁPLICOS e GLEISSOLOS MELÂNICOS, além de PLANOSSOLOS HÁPLICOS, e inclusões de ORGANOSSOLO HÁPLICOS Fíbricos.

As classes de solo observadas na área, classificadas até o segundo nível segundo Santos et al. (2018), foram: LATOSSOLO VERMELHO, LATOSSOLO AMARELO, GLEISSOLO HÁPLICO, GLEISSOLO MELÂNICO, NEOSSOLO QUARTZARÊNICO, NEOSSOLO LITÓLICO, PLANOSSOLO HÁPLICO e ORGANOSSOLO HÁPLICO. As classes foram agrupadas em seis UMs segundo características dos pedoambientes (Tabela 5).

**Tabela 5.** Caracterização de características dos pedoambientes das Unidades de Mapeamento na área de referência (Municípios de Sidrolândia e Campo Grande, MS)

	<b>GXb</b>	<b>LAd</b>	<b>LVd</b>	<b>LVdf</b>	<b>RQo</b>	<b>SXd</b>
<b>Relevo</b>	Plano	Plano a suave ondulado	Plano a suave ondulado	Plano a suave ondulado	Plano a suave ondulado	Plano a suave ondulado
<b>Posição</b>	Baixadas próximas aos drenos	Topo das chapadas	Topo das chapadas	Topo das chapadas	Topo das chapadas	Baixadas próximas aos drenos
<b>Textura</b>	Variável	Argilosa	Média ou mais grossa	Argilosa ou Muito argilosa	Arenosa	Média ou mais grossa
<b>Material de Origem</b>	Sedimentos de Arenitos e Basaltos	Basaltos com influência de Arenitos	Arenitos com influência de Basaltos	Basaltos com influência de Arenitos	Arenitos	Sedimentos de Arenitos e Basaltos
<b>Vegetação original</b>	Campo Higrófilo de várzea	Campo Higrófilo de surgente	Cerradão	Cerradão	Cerradão	Campo Higrófilo de várzea

Um resumo das propriedades dos solos e UMs é apresentado abaixo:

**a) Unidade GXb** - Essa unidade ocorre nas áreas planas de baixada ou próximas dos canais de drenagem. Os solos são, em geral, distróficos, com pouca ou nenhuma agregação nas áreas mais alagadas, com teor de carbono orgânico de moderado a alto e a textura variável com a natureza dos sedimentos; alguns solos têm alta saturação por Al e aumento de argila e silte em profundidade. Geralmente estão associados a vegetação higrófila de várzea. Os fatores que exercem maior influência na formação desses solos são: relevo (posições que favorecem o acúmulo de água), clima (fornecendo umidade por tempo suficiente para condição de hidromorfismo) e material de origem (a textura variável). Os organismos também contribuem quando estão presentes horizontes hísticos, húmicos ou chernozênicos.

**b) Unidade LAd** - Essa unidade tem expressão pontual na área de estudo, aparece geralmente em áreas de relevo plano a suave ondulado, nos topos dos planaltos, onde se formam platôs e concavidades que condicionam uma drenagem é um pouco mais lenta, devido ao próprio formato do relevo e a natureza do material de origem, o que resulta em solos com umidade elevada, caso do Latossolo Amarelo Distrófico plíntico, e do Gleissolo Háplico Distrófico típico presente nas concavidades. Esses solos são, em geral, bastante intemperizados, com baixa atividade de argila e com teor de carbono orgânico de moderado a

baixo. Apresentam textura argilosa. Geralmente são formados de basaltos, alguns solos têm alta saturação por Al, podendo apresentar ligeiro incremento de argila em maior profundidade, mas com ausência ou fraca expressão de cerosidade, estando associados a vegetação de cerradão. Os fatores que exercem maior influência na formação desses solos são: o clima (propiciando a intemperização e remoção das bases), o relevo (condicionando a drenagem moderada), o material de origem (condicionando a textura argilosa) e o tempo (diante da profundidade do *solum* e pelo grau de intemperização dos minerais constituintes).

**c) Unidade LVd** - Essa unidade ocorre em grande extensão ao longo da área de estudo, nos topos planos dos planaltos onde a drenagem é boa e propicia condições para formação de solos profundos e bem estruturados. Esses solos são, em geral, bastante intemperizados, distróficos, com baixa atividade de argila e com teor de carbono orgânico de moderado a baixo. Apresentam textura média a média/arenosa. Geralmente são formados de arenitos, alguns solos têm alta saturação por Al, podendo apresentar ligeiro incremento de argila em maior profundidade, mas com ausência ou fraca expressão de cerosidade, estando associados a vegetação de cerradão. Os fatores que exercem maior influência na formação desses solos são: o clima (propiciando a intemperização e remoção das bases), o relevo (favorecendo a drenagem), o material de origem (condicionando a textura média a média/arenosa) e o tempo (diante da profundidade do *solum* e pelo grau de intemperização dos minerais constituintes).

**d) Unidade LVdf** - Essa unidade ocorre em grande extensão na área de estudo, nos topos planos das chapadas onde a drenagem é boa e propicia a formação de solos profundos e bem estruturados. Esses solos são, em geral bastante intemperizados, podem ser eutróficos ou distróficos, mas sempre com baixa atividade de argila, com teor de carbono orgânico moderado a baixo e apresentam textura argilosa ou muito argilosa. Geralmente são formados de basaltos, podendo apresentar ligeiro incremento de argila em maior profundidade, mas com ausência ou fraca expressão de cerosidade, estando associados a vegetação de cerradão. Os fatores que exercem maior influência na formação desses solos são os mesmos da unidade anterior, diferenciando-se pelo tipo de material de origem que confere aos solos a textura média a argilosa ou mais fina.

**e) Unidade RQo** - Essa unidade ocorre em grande extensão ao longo da área de estudo, nos topos planos das chapadas. Esses solos possuem teores de carbono orgânico baixo a moderado. Apresentam textura arenosa a média/arenosa, com ligeiro aumento de argila e silte em profundidade. Geralmente são formados sobre arenitos e estão associados a vegetação de cerradão. Os fatores que exercem maior influência na formação desses solos são: o material de origem (condicionando a textura arenosa) e o relevo (favorecendo a drenagem).

**f) Unidade SXd** - Essa unidade ocorre nas áreas planas de baixada ou terço inferior das encostas, geralmente próximas dos canais de drenagem. Os solos são, distróficos, com teor de carbono orgânico de moderado a alto e a textura variável com a natureza dos sedimentos; alguns solos têm alta saturação por Al e aumento de argila e silte em profundidade, resultando na formação de um horizonte plânico. Geralmente estão associados a vegetação higrófila de surgente. Os fatores que exercem maior influência na formação desses solos são: relevo (posições que favorecem o acúmulo de água), clima (fornecendo umidade por tempo suficiente para condição de hidromorfismo) e material de origem (a textura variável).

As características morfológicas, físicas e químicas dos perfis representativos das unidades de mapeamento são apresentadas de forma resumida na Tabela 6 (dados completos em anexo).

**Tabela 6.** Características morfológicas, físicas e químicas de perfis representativos das unidades de mapeamento na área de referência (Municípios de Sidrolândia e Campo Grande, MS)

UM	Classificação SiBCS n° do Perfil	Horizonte/ Prof. (cm)	Cor <sup>1</sup>	Classe Textural	Carbono Orgânico g/kg	CTC cmolc/kg	pH	Valor V%
<b>GXb</b>	GLEISSOLO HÁPLICO Tb Distrófico típico PCG-04	Ap / 0-10	10YR 3/2	Areia franca	77	58,3	4,8	11
		AE /10-22	10YR 3/2	Areia franca	37	20	4,9	10
			mq, cm, pq, df 5YR 5/8					
		E1/ 22-36	10YR 5/3 mq cm pq df 5YR 6/8	Areia franca	30	19	4,7	11
		E2/36-46	2,5Y 6/3 mq cm pq df 5YR 6/8	Areia franca	26	14,2	4,7	12
		Btg1/46-58	2,5Y 6/3 mq cm pq df 5YR 6/8	Franco arenosa	28	12,4	4,8	10
		Btgf1/58-130	10 YR 6/3 mq ab gr pr 10 YR 6/8	Franco arenosa	14	7,4	5,1	17
		Btg2/130-175	2,5Y 6/2 mq pc md pr 10YR 6/8	Franco arenosa	9	8,7	5,2	14
			2,5Y 6/2 mq ab md pr 10YR 6/8					
		<b>LAd</b>	LATOSSOLO AMARELO Distrófico plíntico ECG-08	A/0-18	2,5Y 3/2	Argila	47,8	26,6
Bwf/40-60	2,5Y 5/4			Argila	6,2	5,9	5,7	18
	mq ab pq pr 5YR4/6							
Bgcf/80-100	vr 2,5Y 6/2 5YR 5/6			Argila cascalhenta	3,5	5,9	5,3	18
<b>LVd</b>	LATOSSOLO VERMELHO Distrófico típico 06-WN	Ap/0-18	2,5YR 3/3	Areia franca	8,1	29,8	5,4	37
		BA/18-34	2,5YR 3/4	Areia franca	4,1	21,1	4,7	12
				Franco-argilo-arenosa	4,1	17,4	4,7	14
		Bw1/34-59	2,5YR 3/4	Franco-argilo-arenosa	3,1	15,4	4,8	16
		Bw2/59-94	2,5YR 3/4	Franco-argilo-arenosa	2,4	12,4	4,8	16
Bw3/94-135+	2,5YR 3/6	Franco-argilo-arenosa						

UM	Classificação SiBCS n° do Perfil	Horizonte/ Prof.	Cor <sup>1</sup>	Classe Textural	Carbono Orgânico g/kg	CTC cmolc/kg	pH	Valor V%
LVdf	LATOSSOLO VERMELHO Distroférico típico CNPGC-1	A1/0-10	10R 3/3	Argila	20,7	20,9	5,4	27
		A2/10-22	2,5 YR 3/4	Argila	12,4	12,1	5,2	12
		AB/22-48	2,5 YR 3/4	Muito argilosa	9,3	9,8	5,0	5
		BA/48-75	2,5 YR 3/5	Muito argilosa	6,5	8,6	4,9	13
		Bw1/75-110	2,5 YR 3/6	Muito argilosa	5,5	7	4,8	11
		Bw2/110-170	2,5 YR 3/6	Muito argilosa	4,2	5,1	5,2	18
		Bw3/170-310+	2,5 YR 3/6	Muito argilosa	2,8	4	5,4	12
RQo	NEOSSOLO QUARTZARÊNICO Órtico típico	A/0-20	5YR 3/2	Areia	6,2		4,7	
		AC1/20-50	5YR 3/3	Areia	4,8		4,7	
		AC2/50-65	2,5YR 3/4	Areia franca	3		4,7	
		CA/65-85	2,5YR 3/5	Areia franca	3		4,7	
		C1/85-135	2,5YR 3/5	Areia franca	1,9		4,7	
		C2/135-200+	2,5YR 3/5	Areia franca	1,2		4,8	
SXd	PLANOSSOLO HÁPLICO Distrófico arênico ECG-15	A/0-20	2,5Y 2,5/1	Areia franca	13,9	98,4	4,8	7
		E/ 30-80	2,5Y 3,5/1	Areia	2,2	31,6	5	5
		Bt/80-120	2,5Y 4,5/1	Franco arenosa	2,1	25,3	4,5	8

Cor: mq-mosqueado, vr-variegado, pc-pouco, cm-comum, ab-abundante, pq-pequeno, md-médio, gd- grande, df-difuso, pr-proeminente, dt-distinto.

#### 4.2 Análise das Covariáveis Morfométricas e Espectrais

As análises foram realizadas nos dados das três malhas amostrais (Figura 11): as 46 que correspondem aos perfis de campo, (26 perfis localizados no município de Campo Grande e 20 perfis no município de Sidrolândia); as janelas de *pixels* coletadas no entorno das observações originais de campo, totalizando 600 *pixels* (100 para cada UM); além da malha aleatória a partir do mapa legado da área de treinamento, composta por 6000 *pixels* (1000 para cada UM).

As covariáveis que populam os modelos são produtos diretos ou secundários obtidos do MDE, dados do SRE (imagens de satélite) ou dos mapas categóricos, através de ferramentas de processamento digital disponíveis nos SIGs (SAGA e ArcGIS). O MDE (hidrologicamente consistente) utilizado para a geração das covariáveis que, de maneira quantitativa, contribuíram para subsidiar a interpretação do fator Relevo. Os dados obtidos das imagens do sensor espectral (LANDSAT 8), através das bandas e operações entre elas, também forneceram dados quantitativos representando os fatores Clima e Organismos (SAVI) e Material de Origem (*Clay Minerals*), este, que também foi subsidiado por informações dos mapas categóricos (Geologia).

O bom desempenho dos modelos preditivos depende da qualidade dos dados de entrada (Pinheiro, 2015; Arrouays et al., 2014; Hengl et al., 2014). Portanto, a escolha das covariáveis a serem utilizadas depende de análise criteriosa em função da sua importância para a predição, a colinearidade e a redundância.

Uma etapa fundamental, para o processamento correto dos dados e a boa capacidade preditiva das classes de solos pelos modelos, consiste na escolha das variáveis mais adequadas para serem utilizadas na calibração. Outra etapa necessária para a geração de um modelo eficiente, consiste na retirada de variáveis com alta correlação entre si, de forma a simplificar o modelo e evitar redundância de dados. Esse procedimento evita a colinearidade que pode confundir o modelo e prejudicar o ajuste (Ten Caten, 2011).

Nesse sentido, para a seleção das covariáveis foi feita inicialmente uma avaliação visual de todas as variáveis para verificar se o dado estava coerente. Em que foram observadas falhas em algumas variáveis que foram então eliminadas, foram elas: as bandas do espectro visível (2, 3, 4 e 5), os índices *Iron Oxides* e *Grain Size Index*. Isto ocorreu devido a inconsistências observadas nos dados SRE após a junção das imagens, resultando em tons distintos em cada parte da imagem; e também as derivadas do MDE o *Aspect* e *LS Factor*, pois apresentaram valores faltantes para algumas células. Então procedeu-se a análise de correlação entre as 29 variáveis morfométricas restantes bem como para as cinco variáveis espectrais. A correlação entre as variáveis morfométricas é apresentada na Figura 14, onde os círculos maiores significam correlação maior e a cor variando para o azul indica correlação positiva e variando para o vermelho negativa. O “X” destaca que a correlação não foi significativa a 95%.

A presença de covariáveis com elevada correlação entre si pode ocasionar *overfitting* (saturação) e erros no aprendizado do algoritmo devido ao uso de dados redundantes. Segundo Braga et al. (2007), o uso de dados redundantes pode fornecer bons resultados para o treinamento, mas, para os dados de validação, os resultados podem ser totalmente imprecisos. Por isso, para a escolha das covariáveis a serem usadas nos modelos preditivos, são necessárias análises complementares relacionadas a colinearidade e a correlação com as classes de solo.

Na análise dos dados SRE observou-se redundância entre os dados NDVI e SAVI, e portanto, devido as condições da vegetação local o NDVI foi retirado do conjunto. Já na análise do conjunto de dados provenientes do MDE encontrou-se quatro covariáveis apresentando redundância entre si, sendo assim foram eliminadas as covariáveis *Tangencial Crvature* e *Longitudinal Curvature*, pois essas apresentaram a maior soma dos valores de correlação com as demais covariáveis.

Pode-se observar que muitas covariáveis apresentaram valores elevados de correlação, caso por exemplo todos os tipos de curvatura, que marcam uma região inteira de azul na Figura 14. Isto mostra que ainda é possível simplificar o modelo, entretanto, todas as covariáveis que tinham valor de correlação inferior a 95% permaneceram para a modelagem.

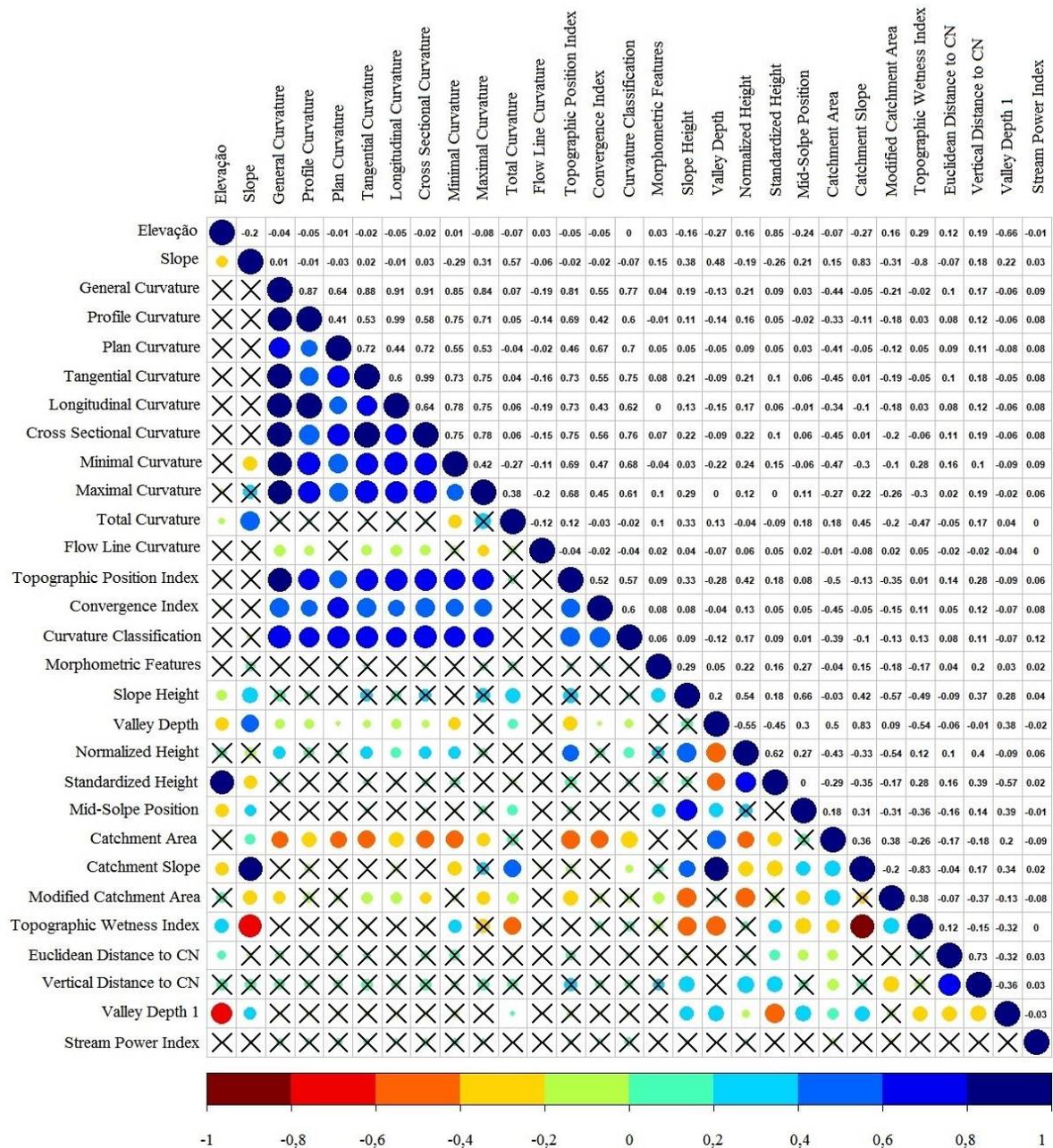
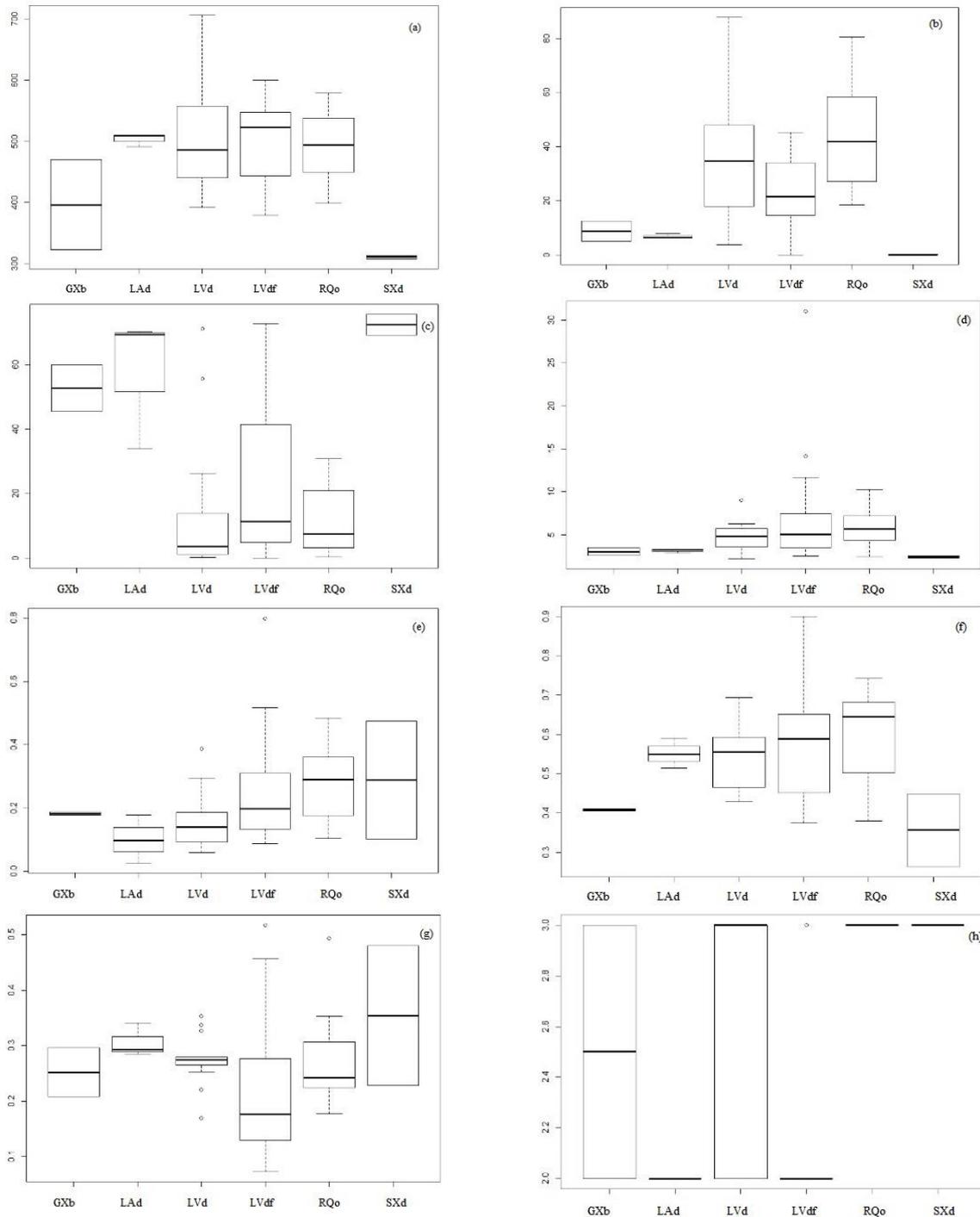


Figura 14. Matriz de correlação das variáveis morfométricas.

A seleção das covariáveis mais relevantes foi baseada na análise exploratória da variabilidade do conjunto de dados com base em valores estatísticos descritivos. A Figura 15 apresenta as variações das covariáveis preditivas em função de cada uma das seis diferentes unidades de mapeamento. Essa análise ajuda a definir o intervalo de variabilidade dos atributos do terreno em cada unidade e facilita a interpretação das relações solo-paisagem.

A variabilidade das covariáveis dentro de cada unidade de mapeamento pode ser observada na Figura 15. Observa-se por exemplo, que as unidades LAd, LVd, LVdf e RQo ocorrem nas posições mais elevadas (Elevação Figura 15-a, e Normalized Height Figura 15-f) sempre em cima dos planaltos, em relevo aplainado, geralmente distantes dos canais de drenagem (Figura 15-b) e nas áreas onde os vales são mais rasos (Figura 15-c); com exceção do LAd que ocorre onde a profundidade do vale é maior. Além disso a Figura 15-h mostra

que todos os LAd bem como os LVdf, são formados a partir de basalto, resultando em solos de textura argilosa, enquanto todos os RQo e a grande maioria dos LVd são formados de arenito.



**Figura 15** Covariáveis derivadas do MDE e SRE relacionadas às UMs, a) Elevação, b) Distância Vertical do Canal de Drenagem, c) Profundidade do Vale, d) *Slope Height*, e) *Mid-Slope Position*, f) *Normalized Height*, g) SAVI, h) Geologia.

GXb - Gleissolo Háptico; LAd - Latossolo Amarelo; LVd - Latossolo Vermelho distrófico; LVdf – Latossolo Vermelho distroférrico; RQo – Neossolo Quartzarênico órtico; SXd – Planossolo Háptico

As áreas de depressão que circundam os planaltos apresentam unidades de solos com hidromorfismo mais acentuado devido à proximidade do canal de drenagem (Figura 15-b), sendo compostas por GXb e SXd. A unidade SXd possui drenagem um pouco melhor, refletindo no desenvolvimento da vegetação dessas áreas, sendo possível diferenciá-la facilmente pelo SAVI (Figura 15-g); além disso os SXd tem uma leve tendência a ter maior declividade média (Figura 15-e) comparada a unidade GXb, com tendência a relevo plano.

### 4.3 Análise dos Modelos

Os métodos *Random Forest* (RF) e *Árvore de Decisão* (AD) foram usados para a extrapolação e criação de mapas de solo, sendo calibrados para ter como saída dados de classes de solos, no caso desse estudo, as unidades de mapeamento.

#### 4.3.1 Modelagem da árvore de decisão

A modelagem da árvore de decisão foi amparada em análises prévias para auxiliar a definição dos parâmetros de modelagem. Foram testados diversos modelos para a escolha de um modelo que conciliasse eficiência e simplicidade. A árvore então foi gerada para apresentar 6 dados de saída (UMs) onde a probabilidade no ultimo nó define a classe de saída. Para definição do número de nós foi usado teste estatístico para comparar o tamanho da árvore e o erro associado, onde foi determinado o parâmetro de complexidade (cp) com valor de 0,001. Por essa análise, é possível observar que é a partir do 16° nó que o erro se torna constante (Figura 12), corroborando a escolha desse valor para o cp adotado no modelo treinado. Assim, com base no baixo erro relativo observado foi definido o valor de 16 nós para a árvore de decisão.

O elevado número de nós se deve ao tamanho do conjunto de dados. Esse aumento no tamanho das árvores em função da densidade amostral também foi observado por Moran e Bui (2002) e por Ten Caten et al. (2013), que destacaram que esses modelos de AD possuem grande capacidade preditiva, porém grande complexidade. Teske et al. (2015) constataram que uma densidade amostral baixa pode resultar na omissão de unidades com pouca expressão na área.

A análise do modelo DT indicou que do total de covariáveis inicialmente consideradas (32 covariáveis), o modelo utilizou efetivamente 12, foram elas: elevação, valley depth (VD), clay minerals (CMI), mid slope position (MSP), catchment slope (CS), slope height (SH), modified catchment area (MCA), geologia, normalized height (NH), SAVI, band 6 (b6) e euclidean distance (ED). A ordem de importância destas covariáveis para a predição do modelo foi: elevação > VD > CMI > b6 > SH > NH > ED > geologia > SAVI > MCA > MSP > CS.

Assim como, verificado por Chagas et al. (2016) e Pinheiro et al. (2017) foi marcante a influência das covariáveis morfométricas, principalmente a elevação e VD. A importância da elevação para predição também foi observada por Adhikari et al. (2014). Dentre as covariáveis espectrais maior destaque foi para CMI e b6, e entre as covariáveis temáticas apenas a geologia foi considerada. Esse resultado difere do observado por Chagas et al. (2016) e Pinheiro (2012), que usaram diferentes abordagens (AD, RF e RNA) para prever as classes de solo na bacia hidrográfica Guapi-Macacu (RJ) de relevo movimentado, e encontraram menor contribuição dos dados derivados das imagens de satélite (Landsat 5) em comparação com o mapa litológico. Essas diferenças podem ser devidas às variações de relevo entre as áreas, pois o presente estudo tem condição de relevo mais aplainado (planaltos e depressões).

A estrutura da AD é apresentada na Figura 16.



A calibração do modelo AD alcançou uma exatidão global de 76,1% e um índice Kappa de 0,66, o que segundo Landis e Koch (1977) corresponde a um desempenho moderado. Esses resultados indicaram melhor desempenho do que os estudos de Giasson et al. (2011), Adhikari et al. (2014), Lorenzetti et al. (2015) e Wolski et al. (2017), enquanto Pinheiro (2015) e Chasmer et al. (2014) obtiveram resultados superiores. Todos utilizando modelos de AD.

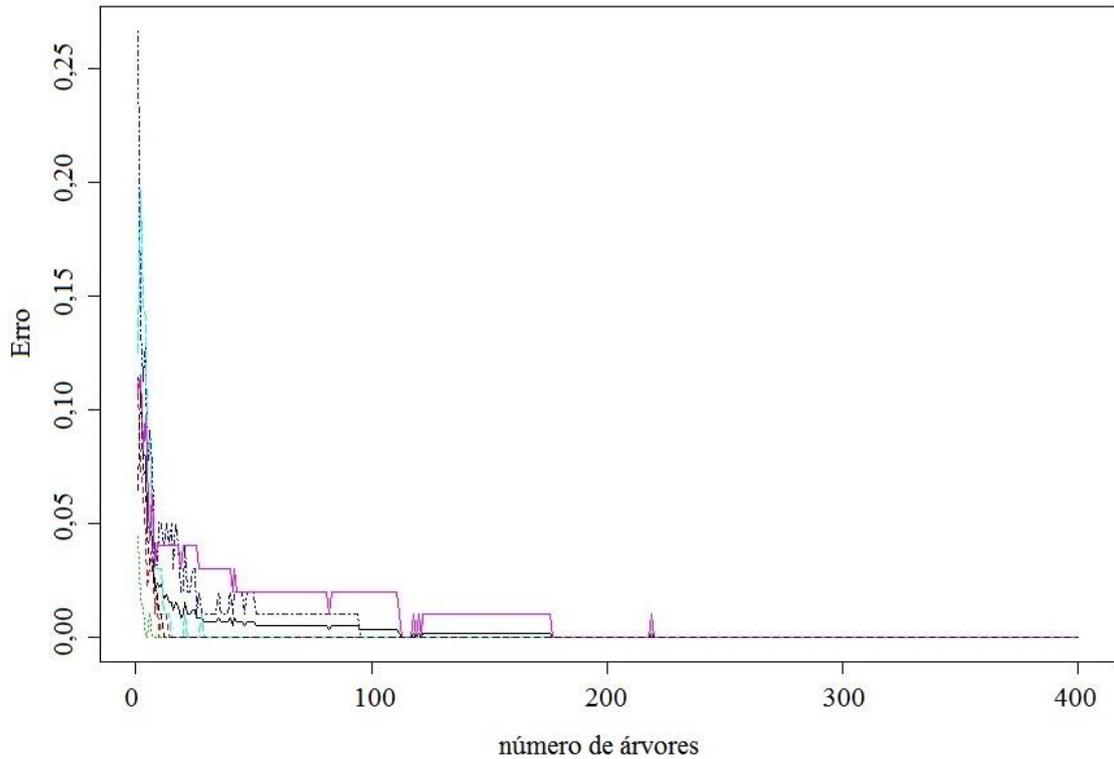
Giasson et al. (2011) testaram várias combinações de algoritmos de árvore de decisão e o número de elementos nos nós terminais das árvores, usando dados legados para diferenciar níveis taxonômicos. O melhor resultado foi obtido pelo algoritmo da árvore de decisão J48, com valores de precisão geral correspondentes a 66,4% e índice Kappa igual a 0,52. Adhikari et al. (2014) utilizaram um modelo de AD para produzir um mapa de solos da Dinamarca, seus resultados indicaram 60% de precisão. Lorenzetti et al. (2015) avaliaram diferentes técnicas de mineração de dados, incluindo árvores de decisão, e obtiveram resultados abaixo do ideal usando esse método. Wolski et al. (2017) obtiveram exatidão global 66,1% e índice Kappa 0,36 em estudo da extrapolação da relação solo paisagem em escala de detalhe em Carajzinho no Estado do Rio Grande do Sul. Pinheiro (2015) observou valores de exatidão global 85% e 0,83 no índice Kappa na predição de classes de solo na bacia Guapi-Macacu (RJ). Chasmer et al. (2014) mostraram uma exatidão global de 85,79 e kappa de 0,91 para o modelo AD.

#### **4.3.2 Modelagem por *random forest***

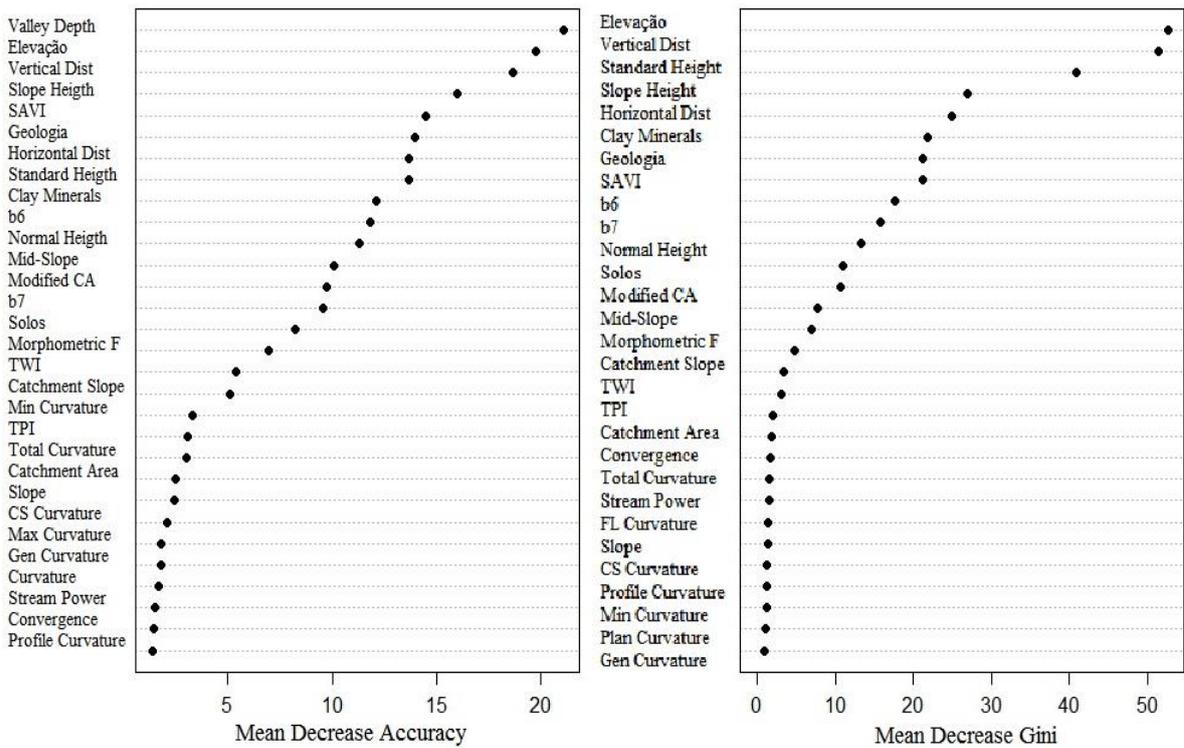
A modelagem pelo *Random Forest* também foi amparada em análises prévias para auxiliar a definição dos parâmetros de modelagem. Assim como a árvore de decisão também apresenta seis classes de dados de saída (UMs), entretanto, as árvores são menores, mas trabalham simultaneamente em conjuntos. Para a finalidade desse estudo foi definido o número de 100 árvores, onde o maior número de respostas para aquele pixel define a classe de saída. Para definição do número de árvores, foi realizado um teste estatístico comparando o tamanho da floresta e o erro associado, testou-se então um conjunto com 1000 árvores, em seguida 400 e a partir do gráfico gerado (Figura 17), foi determinado o número de árvores.

Portanto, baseando no erro relativo apresentado na Figura 17, foi definido o tamanho da floresta com 400 árvores. A figura mostra que o modelo poderia ser um pouco menor, mas outros testes mostraram que este não pode ser muito reduzido e o valor de 400 árvores mostrou-se bastante adequado. Tal definição tem como base o fato de que no atual modelo o resultado é estável e o erro muito baixo e constante, ou seja, logo é suficiente. Considerou-se o princípio de Occam (Young et al., 1996), em que se o erro é constante a partir de um dado valor, o modelo mais simples deve ser adotado.

O modelo RF utilizou todas as covariáveis inicialmente consideradas (32 covariáveis). Esse método de modelagem tem a vantagem de apresentar a ordem de importância das variáveis, através do recurso *out of bag* que retira cada variável do modelo e avalia o erro associado a sua exclusão. A importância de cada variável é então estimada através do aumento do erro pelos métodos da acurácia global e o coeficiente de *Gini*. A partir dessa análise obteve-se então o ranking das covariáveis (Figura 18).



**Figura 17.** Tamanho da floresta (número de árvores) no modelo *Random Forest* em função do erro associado.



**Figura 18** Ranking da importância das covariáveis usadas no modelo RF em função do erro relativo.

Semelhante ao que ocorreu no modelo AD, em que o relevo foi o fator mais importante, o modelo RF também destacou as covariáveis Valley Depth, Elevação, Vertical Distance to Chanel Network e Slope Height, como as mais relevantes. Além destas, também estão entre as mais importantes covariáveis que representam o material de origem e o tempo (Geologia e Clay Minerals), bem como os fatores clima e organismos (SAVI).

Observou-se também que um grande número de covariáveis está contribuindo significativamente para o modelo, e a retirada de qualquer uma delas acarretaria em um aumento do erro dos índices apresentados na Figura 6. A retirada, por exemplo, da variável Valley Depth aumentaria pouco mais de 20% o erro da predição, e em torno 50% do erro no índice Gini. No entanto, observa-se que algumas covariáveis, principalmente as associadas com a curvatura, tiveram baixa contribuição. Ou seja, o modelo poderia ser ainda mais simplificado e otimizado pela retirada dessas covariáveis. Para torná-lo mais simples, foram retiradas as covariáveis de contribuição inferior a 10% (% *increase mean error*) (Figura 6). Sendo assim, o conjunto final foi composto pelas covariáveis: Elevação, Slope Height, Normal Height, Stand Height, Mid Slope Position, Modified Catchment Area, Horizontal Distance to Chanel Network, Vertical Distance to Chanel Network, Valley Depth, Geologia, b6, b7, Clay Minerals e SAVI.

Vale ressaltar que o modelo RF considerou as mesmas covariáveis usadas pelo modelo AD, o que corrobora que os fatores principais envolvidos na formação dos solos da área são de fato o relevo e o material de origem. O material de origem condicionando a natureza do material (composição, resistência, textura, fertilidade, etc) e o relevo influenciando a redistribuição desse material e o fluxo de água ao longo da paisagem.

No modelo RF não é possível ver as arquiteturas individuais de cada árvore, que são bem menores que em uma árvore única (AD). Porém, devido a quantidade de árvores que são processadas simultaneamente, o resultado é bastante confiável, uma vez que resulta da maioria das respostas concordantes apresentadas pelas árvores. A calibração do modelo RF alcançou uma exatidão global de 93,5% e um índice Kappa de 0,91, o que segundo Landis e Koch (1977) corresponde a um desempenho excelente. Esses resultados indicaram melhor desempenho que em Pinheiro (2015), Dias et al. (2015), Vasconcelos (2016) e Ten Caten (2011); enquanto, Chasmer et al. (2014) obtiveram resultados superiores. Todos usando modelos de RF.

Pinheiro (2015) observou valores de exatidão global 67,9% e 0,61 no índice Kappa na predição de classes de solo na bacia Guapi-Macacu (RJ). Dias et al. (2015) utilizaram um modelo de RF para produzir mapa de solos desenvolvidos de arenito em Botucatu (SP), com resultados indicando 56% de exatidão global e índice Kappa 0,39. Vasconcelos (2016) avaliou o mapeamento dos solos do quadrilátero ferrífero (MG) em diferentes escalas e obteve exatidão global de 79% e índice Kappa 0,38. Ten Caten (2011) obteve valores de Kappa que variaram entre 69,7% e 71,1%, em estudo na região central do Estado do Rio Grande do Sul, comparando amostras de campo e mapas inferidos gerados pelo método de random forest. Chasmer et al. (2014) mostraram uma exatidão global de 98,29, no modelo Random Forest.

### 4.3.3 Comparação entre os modelos preditivos

Com os parâmetros definidos e ajustados pelos dados de entrada, os algoritmos preditores foram então avaliados através do desempenho usando uma matriz de confusão com os dados de validação previamente separados e não usados no treinamento. A acurácia dos modelos preditores foi avaliada através da concordância entre as áreas de ocorrência das UMs no mapa gerado pela aplicação desses modelos através do uso de matrizes de erros (Congalton, 1999). A concordância recebeu três denominações: acurácia global (AG); acurácia do usuário (AU) e acurácia do mapeador (AM).

As matrizes de confusão obtidas pelos modelos são apresentadas na Tabela 7.

**Tabela 7.** Matiz de confusão dos modelos *Árvore de Decisão* e *Random Forest* gerados a partir da malha de janela de *pixels* e coleta aleatória a partir do mapa de referência.

<b>Janela de <i>Pixels</i></b>						
<b>Árvore de Decisão (AG = 76,1 e Kappa = 0,658)</b>						
	GXb	LAd	LVd	LVdf	RQo	SXd
GXb	0	0	1	0	0	0
LAd	0	3	0	0	0	0
LVd	0	0	8	2	1	0
LVdf	2	0	3	17	1	0
RQo	0	0	1	0	5	0
SXd	0	0	0	0	0	2
<b>Random Forest (AG = 93,5 e Kappa = 0,909)</b>						
	GXb	LAd	LVd	LVdf	RQo	SXd
GXb	2	0	1	0	0	0
LAd	0	3	0	0	0	0
LVd	0	0	10	0	0	0
LVdf	0	0	1	19	0	0
RQo	0	0	1	0	7	0
SXd	0	0	0	0	0	2
<b>Malha aleatória no Mapa de Referência</b>						
<b>Árvore de Decisão (AG = 67,4 e Kappa = 0,559)</b>						
	GXb	LAd	LVd	LVdf	RQo	SXd
GXb	0	0	1	1	0	0
LAd	1	3	0	2	0	0
LVd	0	0	8	1	3	0
LVdf	0	0	3	14	0	0
RQo	0	0	1	1	4	0
SXd	1	0	0	0	0	2
<b>Random Forest (AG = 65,2 e Kappa = 0,534)</b>						
	GXb	LAd	LVd	LVdf	RQo	SXd
GXb	0	0	2	1	0	0
LAd	1	3	0	1	0	0
LVd	0	0	8	2	3	0
LVdf	0	0	3	13	0	0
RQo	0	0	0	1	4	0
SXd	1	0	0	1	0	2

Os valores de Kappa e exatidão global foram satisfatórios para ambos os modelos, sendo superiores no modelo RF (Kappa de 0,91 e EG de 93,5%) em comparação com o modelo AD (Kappa de 0,66 e EG de 76,1%). De acordo com Landis e Koch (1977), o modelo RF obteve resultado considerado excelente e o modelo AD resultado moderado. Estes resultados foram semelhantes aos obtidos por Pinheiro (2015) que observou índice kappa de 0,61 e exatidão global de 67,89% em mapeamento de solos na bacia do Guapi-Macacu (RJ),

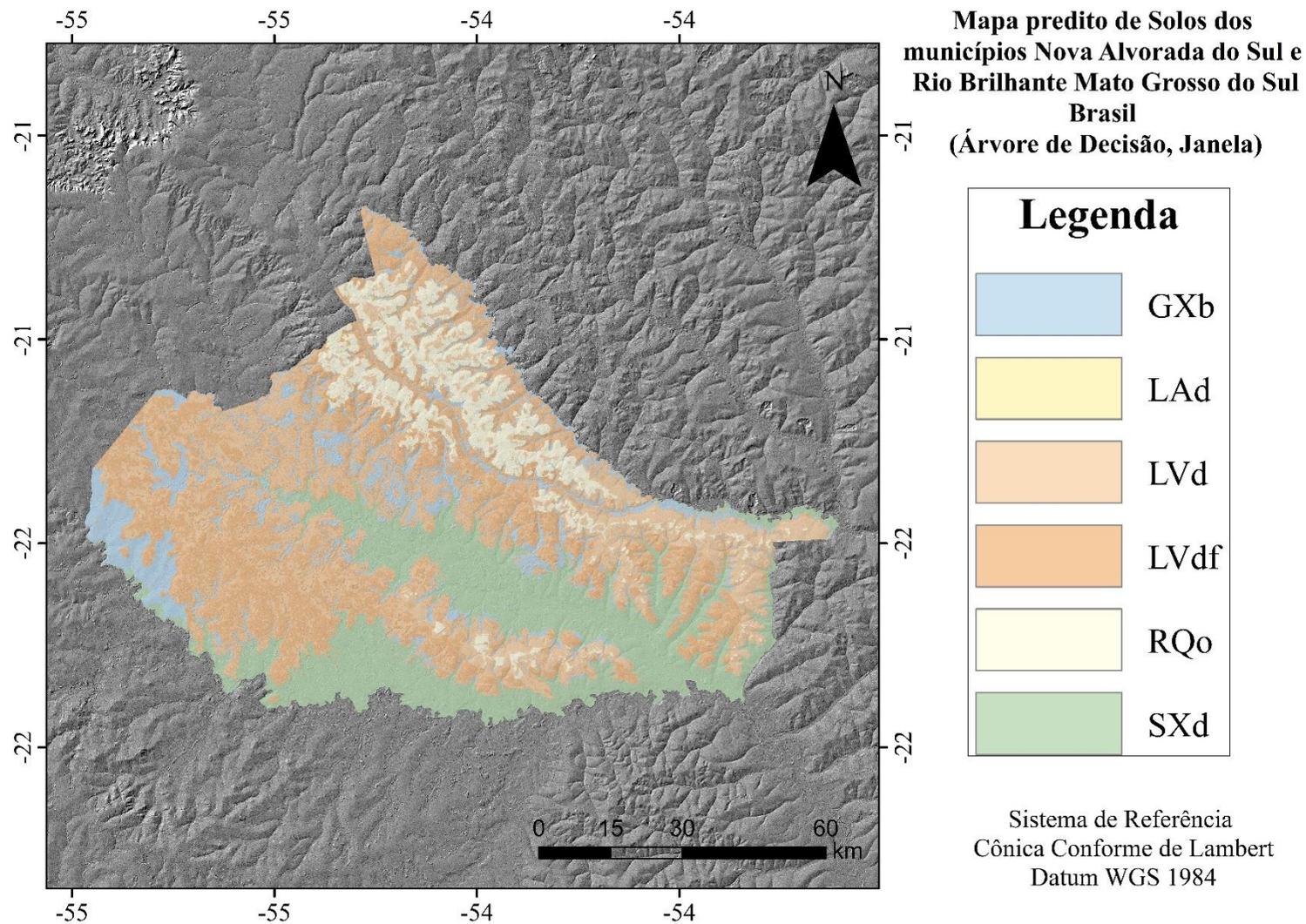
usando o modelo Random Forest. Esses valores indicam uma boa concordância entre o total de pontos amostrados e aqueles na diagonal das matrizes da Tabela 2. No entanto, conforme Congalton (1991), a acurácia global pode superestimar a qualidade do mapeamento e seus valores devem ser tomados com cautela. Conforme Congalton (1991), a acurácia global pode superestimar a qualidade do mapeamento e seus valores devem ser tomados com cautela.

Os modelos resultaram em distribuições espaciais diferentes das unidades de solos. Os modelos foram extrapolados e os mapas gerados são apresentados nas Figuras 19, 20, 21 e 22. Em ambos os mapas gerados a partir da predição foram estimadas todas as seis UMs de entrada. No entanto, a predição das UMs que têm menor extensão na área é prejudicada pelo método de amostragem aleatória (Giasson et al., 2011; Ten Caten et al., 2012). Nenhuma UM de entrada foi omitida na extrapolação, entretanto a UM GXb só foi classificada corretamente no modelo RF, com o treinamento baseado em janela de *pixels*. Provavelmente devido ao baixíssimo número de perfis dessa unidade, o que acarretou em erros para a validação dessa UM. Entretanto ainda sim o modelo foi aplicado e a distribuição espacial de todas as UMs manteve coerência entre os modelos, inclusive com o mapa de solos da área elaborado pelo IBGE (2017). Porém, como o mapa original do IBGE não contempla as mesmas UMs que os mapas gerados pelos modelos AD e RF não foi possível comparar as áreas e classes de solos.

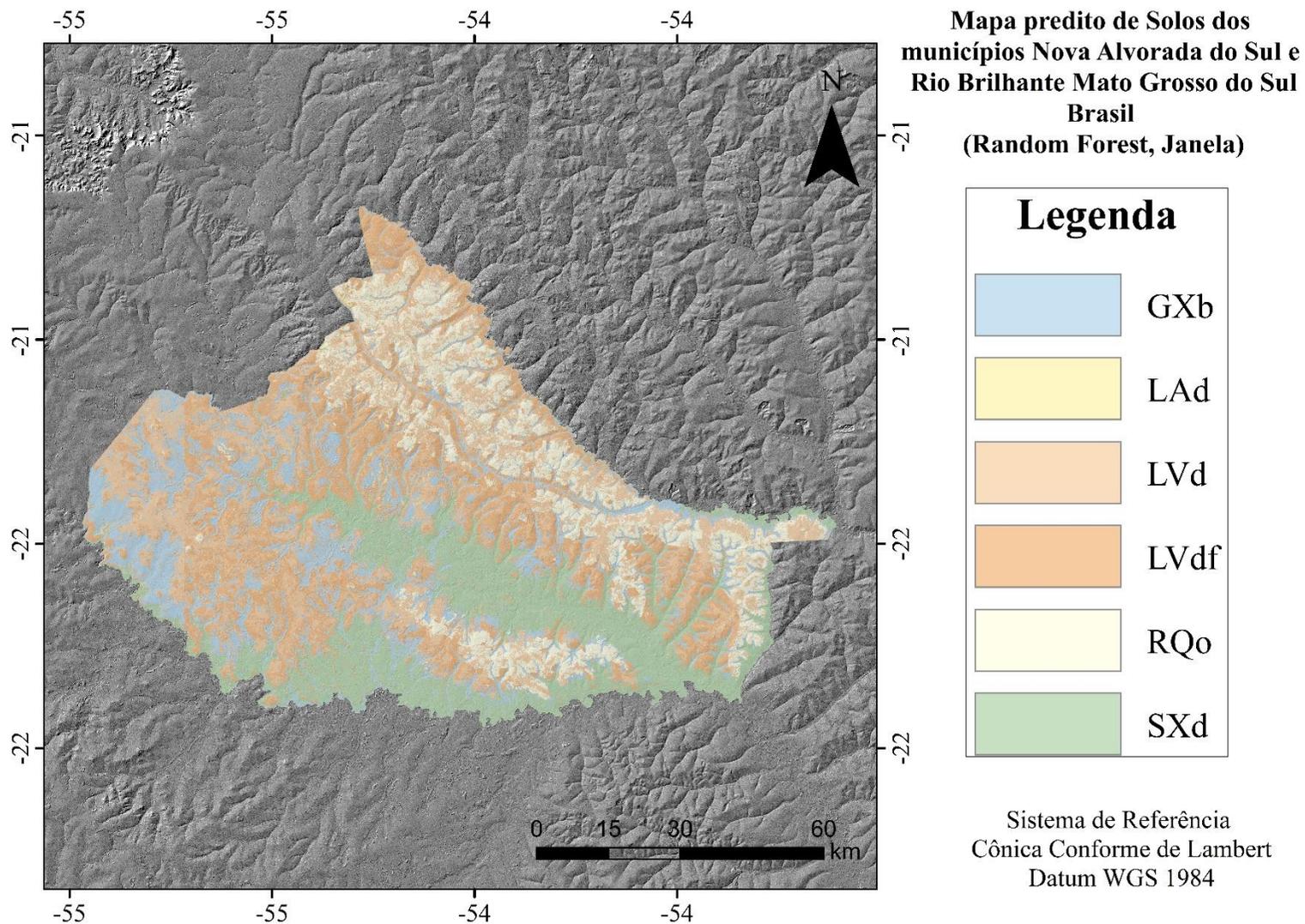
Em ambos os modelos, não houve erros de classificação nas UMs LAd e SXd; o mesmo para GXb, RQo e LVdf no modelo RF (Tabela 7). Este resultado indica que as relações solo-paisagem dessas unidades foram muito bem aprendidas, as duas primeiras para ambos os modelos e as últimas para o modelo RF. O que é atribuído às condições específicas em que estas unidades se encontram, que faz com que sejam dificilmente confundidas pelos modelos. Ou seja, os SXd aparecem sempre em áreas de planícies com baixa elevação, geralmente em áreas mais úmidas, enquanto os LAd apresentam-se sobre planaltos, em relevo plano em áreas de litologia variada sob condições que reduzam a velocidade de infiltração. Já a unidade GXb contempla as áreas mais suscetíveis ao hidromorfismo, logo menor elevação, de relevo plano e com má drenagem, que margeiam os canais de drenagem. Os RQo e os LVdf estão sempre em posição de topo dos planaltos, relevo plano e boa drenagem, diferindo pelo material de origem do qual são formados (Arenito e Basalto respectivamente).

Em todos os modelos, a unidade com menor acerto foi a GXb e a maior confusão se deu com as unidades LAd, SXd, e até mesmo com LVdf. A confusão com as primeiras é aceitável, pois ambas são influenciadas por hidromorfismo. Mas o fator preponderante para o elevado erro pode ser o baixo número de amostras para a aferição dessa UM (apenas duas).

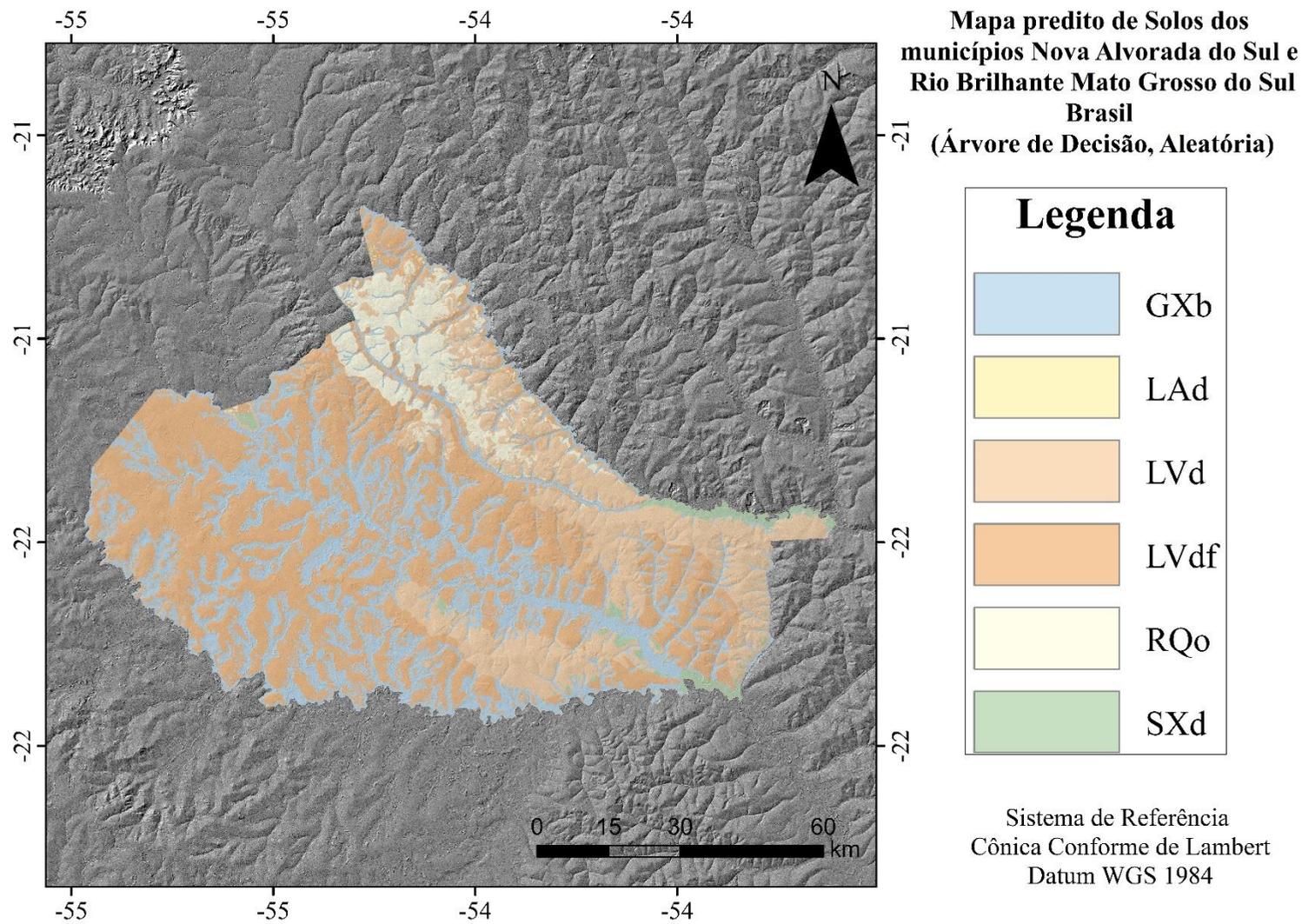
Os solos da unidade LVd apresentaram alguns erros na validação, são solos de textura média ou mais grossa, que se situam em topo de planalto em condições de relevo plano a suave ondulado, têm boa drenagem e estão distantes dos canais de drenagem. Essa unidade erroneamente classificou tanto um LVdf, de textura franco-arenosa, quanto um RQo, de textura mais arenosa. Essas classes podem ser facilmente confundidas entre si até mesmo no campo e por pedólogos experientes quando a distribuição granulométrica se encontra no limite entre grupamentos texturais. Elas podem ocorrer em condições de paisagem similares, diferindo apenas nos teores de argila e com teores muito próximos. O material de origem (arenitos) com maior quantidade de cimentantes ferruginosos irá formar mais argila e terá cor mais vermelha pela presença do ferro, identificando o LVd; enquanto em materiais com maior quantidade de sílica predomina o quartzo e resultam na unidade RQo. O modelo AD também confundiu as unidades LVd e LVdf, provavelmente devido as condições de relevo e drenagem similares. O menor acerto na unidade LVd (textura média), que se mistura com os RQo e também se confunde com os outros LVdf (textura argilosa), corrobora a importância de ter um mapa detalhado de litologia como covariável preditora.



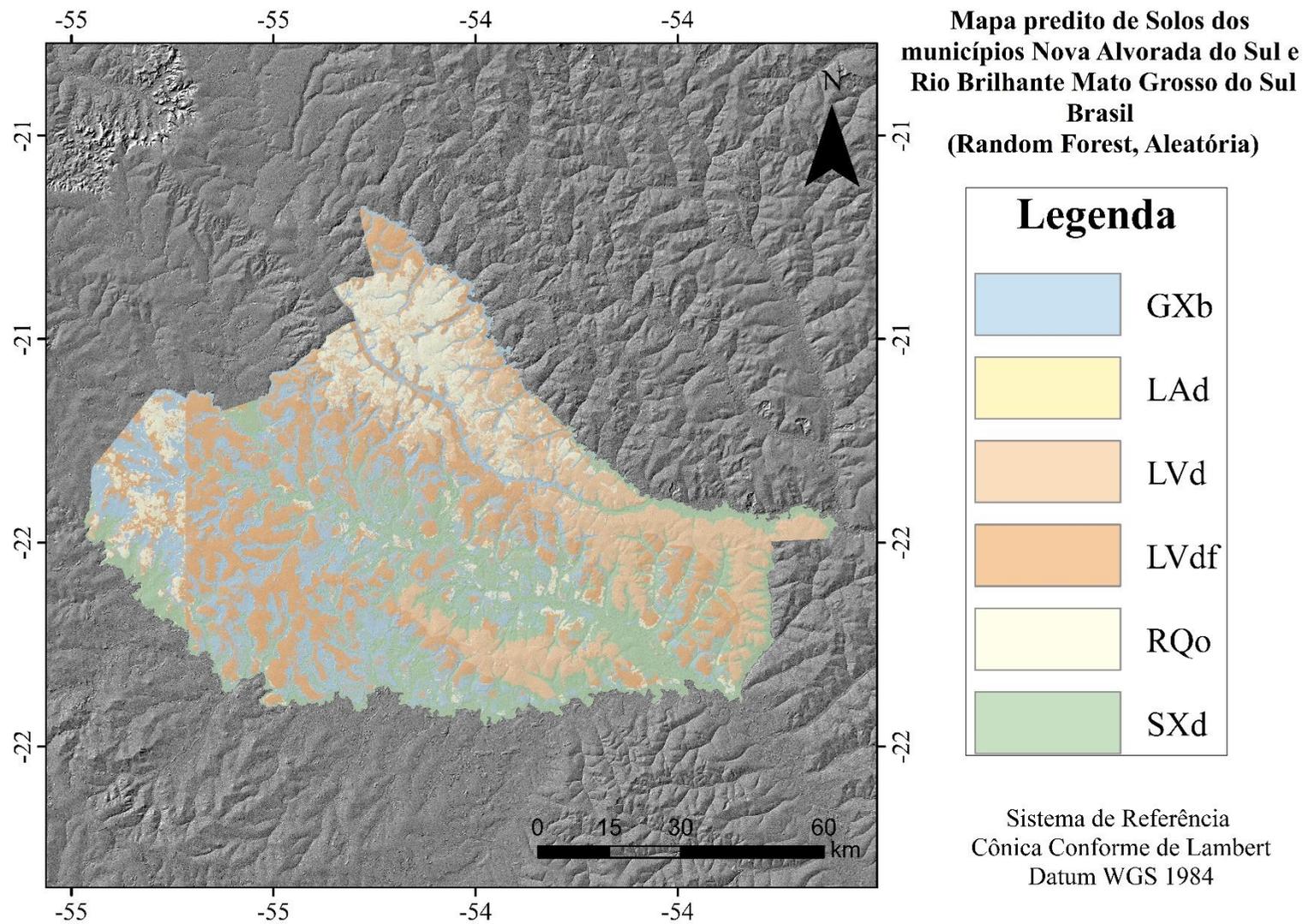
**Figura 19.** Mapa de solos gerado a partir do modelo preditivo Árvore de Decisão e amostragem em Janela de *pixels*, nos municípios de Nova Alvorada do Sul e Rio Brillhante no Estado do Mato Grosso do Sul.



**Figura 20.** Mapa de solos gerado a partir do modelo preditivo *Random Forest* e amostragem em Janela de *pixels*, nos municípios de Nova Alvorada do Sul e Rio Brillante no Estado do Mato Grosso do Sul.



**Figura 21.** Mapa de solos gerado a partir do modelo preditivo Árvore de Decisão e amostragem Aleatória no mapa de referência, nos municípios de Nova Alvorada do Sul e Rio Brillante no Estado do Mato Grosso do Sul.



**Figura 22.** Mapa de solos gerado a partir do modelo preditivo *Random Forest* e amostragem Aleatória no mapa de referência, nos municípios de Nova Alvorada do Sul e Rio Brillante no Estado do Mato Grosso do Sul.

Os modelos gerados a partir de malha aleatória extraída do mapa de referência apresentaram resultados estatisticamente inferiores, com erros em quase todas as UMs testadas, com exceção das SXd e LAd. Entretanto, considerando-se a forma aleatória de amostragem para a criação do modelo, pode-se considerar que os modelos gerados ainda possuem resultado satisfatório para essa técnica de amostragem diretamente do mapa legado. Além disso o modelo *Random Forest* apresentou índices estatísticos superiores ao modelo Árvore de Decisão, na amostragem em janela de pixel, invertendo-se na amostragem aleatória. Provavelmente, devido a maior variação das condições ambientais, além do elevado número de pixels coletados.

Avaliando os mapas gerados a partir da amostragem em janelas de *pixels*, observam-se nas figuras 19 e 20 feições bastante semelhantes. Em quem na porção localizada a norte e nordeste dos mapas observa a feição de planalto com predomínio das unidades RQo, LVd e LVdf e nas demais regiões essas classes foram subestimadas. Por outro lado, nessa feição nota-se predomínio da unidade LVdf ao longo das margens dos canais (mais evidenciado no modelo de árvore de decisão), o que pode ser explicado pela associação com o índice de umidade. Porém, esse resultado não é coerente com o material de origem predominante (arenitos), pois nessa posição deveria predominar a unidade RQo, segundo observações de campo.

A parte central até o oeste da região mostra grande expressão das unidades GXb e SXd, ou seja, parece ter superestimado essas unidades. Isso ocorreu devido aos pontos representativos dessas unidades estarem em posição de cota mais elevadas, o que levou a superestimativa dessas unidades ao longo de toda região sul da área de extrapolação. Como ocorre variação altimétrica, entre a área usada como referência e a área de extrapolação, esta diferença conduziu ao resultado observado. Já a porção leste apresentou grande expressão da unidade SXd, devido à sua baixa cota, associada a natureza do material de origem (composto por sedimentos) e localização no entorno das áreas de drenagem.

Na amostragem por janela de *pixels*, ambos os mapas (modelos AD e RF) mostram padrões semelhantes para as unidades de mapeamento de solo previstas, com exceção das unidades SXd e GXb, que foram superestimadas. Os mapas ilustram confusão potencial na classificação das planícies de inundação, atribuindo área mais extensa para a unidade SXd, provavelmente superestimando a sua distribuição.

Avaliando os mapas gerados a partir da amostragem Aleatória a partir do mapa legado, nas Figuras 21 e 20 notam-se feições semelhantes aos mapas anteriores, principalmente na porção localizada a norte e nordeste, com predomínio das mesmas unidades. Porém, a maior expressão da unidade RQo é coerente com o material de origem (arenitos). Já na parte sul, observam-se diferenças entre os mapas. A parte central até o oeste da região, assim como nos modelos por janelas, mostrou grande expressão das unidades GXb e SXd.

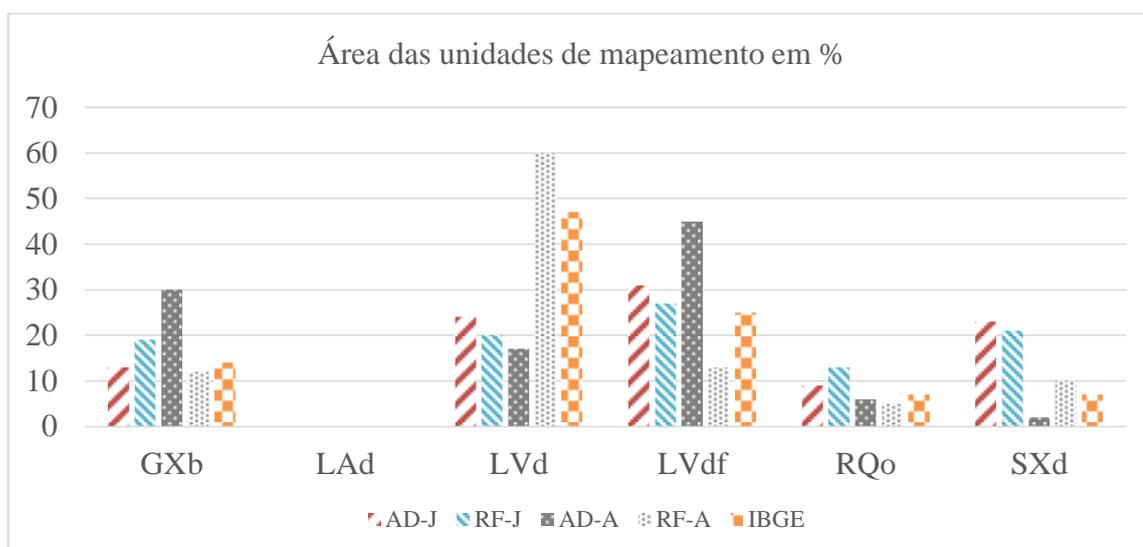
Ambos os modelos delinearão melhor as unidades de solo de melhor drenagem, sendo que o modelo de árvore identificou mais *pixels* como pertencendo a unidade GXd e o RF tendeu a classificar mais *pixels* como SXd.

Os mapas apresentaram maiores diferenças entre os métodos de amostragem do que em função do algoritmo utilizado. Apesar dos resultados com o modelo aleatório mostrarem índices estatísticos inferiores, a visualização dos mapas sugere o contrário. Entretanto, para aferir a veracidade do mapa é necessária uma validação a nível de campo. Na amostragem Aleatória, ambos os mapas (modelos AD e RF) mostram padrões semelhantes para as unidades de mapeamento de solo previstas, e apresentaram também índices estatísticos bastante semelhantes, sendo os valores obtidos no modelo em Árvore de Decisão pouco superiores aos obtidos pelo modelo *Random Forest*.

Os valores dos índices estatísticos obtidos pelos modelos gerados a partir de ambas as malhas (Janela de *pixels*: *Árvore de Decisão* AG = 76,1 e Kappa = 0,658 e *Random Forest* AG = 93,5 e Kappa = 0,909; Amostragem Aleatória: *Árvore de Decisão* AG = 67,4 e Kappa = 0,559; *Random Forest* AG = 65,2 e Kappa = 0,534) se mostraram elevados, entretanto, não foi possível fazer a verificação no campo, sendo está uma etapa ainda necessária para aferição do mapa gerado. Estudos como o de Chasmer et al. (2014) mostraram uma acurácia global foram de 85,79, para o modelo AD, e de 98,29, usando o modelo *Random Forest*. Ten Caten (2011) obteve valores de Kappa que variaram entre 69,7% e 71,1%, na região central do estado do Rio Grande do Sul, comparando amostras de campo e mapas inferidos gerados pelo método de árvores de decisão e com conjuntos de amostras de tamanhos diferentes como dados de entrada.

As áreas estimadas por cada modelo preditivo para cada UM estão na Figura 23, bem como a área disponível adaptada do mapa de solos do IBGE (2017) para a área extrapolada.

A diferença na área estimada das UMs de maior expressão geográfica (GXb, LVd e LVdf) foi mais evidentes entre os algoritmos AD e RF do que nas malhas amostrais. A maior diferença em área estimada foi observada na unidade SXd, onde ambos os modelos oriundos de janela de *pixels* reconheceram esse padrão em 23% e 21% da área, para AD e RF respectivamente, enquanto o mapa o IBGE apontava apenas 7%. A unidade SXd foi superestimada em todos os modelos quando comparadas ao mapa do IBGE, exceto para o modelo AD em amostragem aleatória. A unidade GXb, que também foi superestimada por todos os modelos, foi subestimada pelo modelo RF em amostragem aleatória. Por outro lado, as unidades LVd e LVdf foram subestimadas por todos os modelos, exceto pelo modelo AD, e para o modelo RF, ambos em amostragem aleatória, respectivamente. Para as unidades com menor expressão geográfica, na unidade LAd que não estava presente no mapa do IBGE quase não apareceu em termos de área em nenhum dos modelos testados. De maneira geral, o modelo que mais se aproximou ao mapa apresentado pelo IBGE foi o modelo RF a partir de amostragem aleatória, entretanto esse modelo subestimou consideravelmente a unidade LVdf, quantificando-a a área como sendo quase a metade do indicado pelo IBGE.



**Figura 23.** Área estimada para cada unidade de mapeamento de solos nos municípios de Nova Alvorada do Sul e Rio Brillhante, MS, pelos modelos preditivos *árvore de decisão* (AD) e *Random Forest* (RF) utilizando amostragem em Janelas de *Pixels* (J) e Aleatória (A) e pelo mapa do IBGE (2017)

Em relação ao mapa de solos do IBGE (ano), que contempla a área de estudo, vale ressaltar que a escala (1:250.000) é pouco detalhada e o mapa não contempla as mesmas UMs previstas pelos modelos, sendo necessária uma adaptação das unidades para a comparação.

O uso de dados independentes é indicado para validação dos modelos preditores (Grinand et al., 2008; Brus et al., 2011; Teske et al., 2015). Adicionalmente, ao utilizar a amostragem aleatória, a avaliação do modelo pode ser feita com diferentes métodos sem que se obtenham falsos valores da acurácia (sobre ou subestimados). Isso pode estar associado ao fato de a acurácia obtida pela validação cruzada ser calculada a partir de uma matriz de confusão com base no uso de subconjuntos. Assim, sempre que um subconjunto for usado para a avaliação, este será excluído do treinamento do classificador numa mesma rodada, sendo alternadamente trocados até que todos os subconjuntos sejam utilizados. Por isso, a estimativa da acurácia obtida pela validação cruzada resulta de estimativa da média das classificações e é considerada como indicador confiável para estimar o desempenho da predição de algoritmos supervisionados, quando a amostragem for aleatória (Elkan, 2012).

Vale ressaltar que a validação dos modelos foi feita com um conjunto de dados diferente dos conjuntos de treinamento, o que garante a independência dos dados, permitindo obter valores da acurácia semelhantes à acurácia mensurada. Portanto, os resultados obtidos a partir de ambas as amostragens, tanto em janela quanto aleatória a partir do mapa legado, são consistentes.

Essa abordagem possibilitou o treinamento dos modelos a partir de um pequeno conjunto de dados (perfis e mapa) e otimizou o aprendizado dos padrões de ocorrência dos solos em áreas com poucas informações, ainda assim possibilitou a criação de modelos consistentes para áreas sem dados coletados de solos. Apresentando-se, portanto, como uma alternativa viável de implementação de protocolo para o MDS através da expansão da área mapeada utilizando dados legados. Podendo assim ser de grande valia para mapeamento de grandes áreas, bem como boa aplicabilidade no contexto do PRONASOLOS.

A validação de campo se faz necessária, para obter o erro associado mais próximo da realidade local, uma vez que não se tem nenhuma amostra obtida dentro da área de extrapolação que possa confirmar a eficiência dos modelos preditivos usados no estudo, logo o erro calculado pelo modelo pode ser tendencioso pela forma de amostragem. No entanto, os resultados já obtidos corroboram o potencial do uso de modelos baseados em árvores (AD e RF) no MDS. O que pode permitir uma abordagem quantitativa e conferir maior acurácia ao produto final, conforme quantificado pelo erro dos modelos em relação às amostras de validação de campo.

## 5 CONCLUSÕES

Os modelos baseados em árvores (AD e RF) podem contribuir para o conhecimento sobre os fatores de formação e para qualificar sua contribuição na pedogênese, bem como na compreensão dos pedoambientes e distribuição das classes de solo na paisagem. Ambos foram capazes de identificar e prever com bom nível de acerto as unidades de mapeamento de solos na área de extrapolação, de acordo com a abordagem utilizada.

As malhas amostrais adotadas para o treinamento dos modelos mostraram-se eficientes para a extração dos padrões de paisagem e geraram informações que podem subsidiar o planejamento das etapas de levantamento para a expansão da área mapeada, e a validação dos modelos a partir dos pontos originais garante um resultado mais próximo da condição observada em campo.

O uso de dados legados dos mapeamentos de solos dos municípios vizinhos que apresentaram fisiografia semelhante, mostrou-se promissor para o processo de aprendizado dos padrões morfométricos, bem como para a extrapolação das unidades de mapeamento para toda a área.

Apesar de apresentar valores estatísticos inferiores aos obtidos nos modelos gerados por janelas de *pixels*, a amostragem aleatória apresentou índices estatísticos satisfatórios. Logo é uma técnica com potencial em MDS, já que a forma de amostragem é simples, embora dependa da qualidade do mapa legado.

O modelo *Random Forest* apresentou melhor desempenho preditivo em todos os índices estatísticos avaliados, confirmando a robustez do modelo para predição quando comparado ao modelo de uma única árvore.

## **6 CONSIDERAÇÕES FINAIS**

A qualidade dos modelos apresentados expressa índices estatísticos gerados a partir de validação usando os dados oriundos das áreas de treinamento, e, portanto, ainda não é conclusiva. Para a verificação da qualidade dos mapas gerados, uma nova validação é necessária, a partir de dados coletados na área mapeada por extrapolação dos modelos, sendo esta etapa prevista para a próxima fase do projeto.

## 7 REFERÊNCIAS BIBLIOGRÁFICAS

- ALFONSI, R. R.; PINTO, H. S.; ZULLO JÚNIOR, J.; CORAL, G.; ASSAD, E. D.; EVANGELISTA, B. A.; LOPES, T. S. S.; MARRA, E.; BEZERRA, H. S.; HISSA, H. R.; FIGUEIREDO, A. F.; SILVA, G. G.; SUCHAROV, E. C.; ALVES, J.; MARTORANO, L. G.; BOUHID, A.; ROMÍSIO, G.; BASTOS ANDRADE, W. E. **Zoneamento Climático da Cultura do Café (*Coffea arabica*) no Estado de Mato Grosso do Sul**. Campinas. Campinas: IAC: UNICAMP; Brasília: Embrapa Cerrados; Niterói: Pesagro-Rio; Rio de Janeiro: SIMERJ: Embrapa Solos. 2002.
- ARROUAYS, D., GRUNDY, M.G., HARTEMINK, A.E., HEMPEL, J.W., HEUVELINK, G.B.M., HONG, S.Y., LAGACHERIE, P., LELYK, G., MCBRATNEY, A.B., MCKENZIE, N.J., MENDONCA-SANTOS, M.D.L., MINASNY, B., MONTANARELLA, L., ODEH, I.O.A., SANCHEZ, P.A., THOMPSON, J.A., ZHANG, G.-L., **GlobalSoilMap: toward a fine-resolution global grid of soil properties**. *Advances in Agronomy*, pp. 93–134. 2014.
- ANTUNES, M. A. H., DEBIASI, P., COSTA, A. R., GLERIANI, J. M., **Correção atmosférica de imagens ALOS/AVNIR-2 utilizando o modelo 6S**. *Revista Brasileira de Cartografia*, v. 64, p. 531-539, 2012.
- BAGATINI, T.; GIASSON, E.; TESKE, R. **Expansão de mapas pedológicos para áreas fisiograficamente semelhantes por meio de mapeamento digital de solos**. *Pesquisa Agropecuária Brasileira*. v.51. 2016
- BIVAND, R.; KEITT, T.; ROWLINGSON, B. **rgdal: Bindings for the geospatial data abstraction library**, 2017. Disponível em: <<https://cran.r-project.org/package=rgdal>>
- BEHRENS, T.; FOSTER, H.; SCHOLTEN, T.; STEINRUCKEN, U.; SPIES, E. D.; GOLDSCHMITT, M. **Digital soil mapping using artificial neural networks**. *Journal of Plant Nutrition and Soil Science* 168: 21-33. 2005.
- BEHRENS, T., ZHU, A.-X., SCHMIDT, K., SCHOLTEN, T., **Multi-scale digital terrain analysis and feature selection for digital soil mapping**. *Geoderma* 155, 175–185. 2010.
- BRAGA, A. de P.; CARVALHO L. F. de; LUDEMIR., T. B. **Redes Neurais Artificiais: Teoria e Aplicações** 2.ed. Rio de Janeiro: LTC – Livros Técnicos e Científicos Editora. 226p. 2007
- BRASIL. Ministério de Minas e Energia. Secretaria Geral. Projeto RADAMBRASIL. **Folha SF. 21, Campo Grande**: geologia, geomorfologia, pedologia, vegetação e uso potencial da terra, v. 28. Rio de Janeiro: 416p. il. 1982.
- BREIMAN, L., **Random forests**. *Machine Learning* 45, 5–32. 2001.
- BREIMAN, L.; FRIEDMAN, J.; OLSHEN, R.; STONE, C. **Classification and regression trees**. The Wadsworth statistics/Probability Series. Wadsworth International Group, Belmont, CA. 358p. 1984.
- BRUS, D. J.; KEMPEN, B.; HEUVELINK, G. B. M. **Sampling for validation of digital soil maps**. *Eur J Soil Sci*. 62:394-407. 2011.

- CARVALHO JÚNIOR, W. de. **Classificação supervisionada de pedopaisagens no domínio dos mares de morros utilizando redes neurais artificiais.** Tese de doutorado. UFV, MG. 160p. 2005.
- CRESSIE, N., **Statistics for Spatial Data.** Wiley, New York, 900 pp. 1991.
- CHAGAS, C. da S. **Mapeamento digital de solos por correlação ambiental e redes neurais em uma bacia hidrográfica de domínio de mar de morros.** Tese de doutorado. UFV, MG. 239p. 2006.
- CHAGAS, C. S.; FERNANDES FILHO, E. I.; VIEIRA, C. A. O.; SCHAEFER, C. E. G. R.; CARVALHO JÚNIOR, W. **Topographic attributes and Landsat 7 data in the digital soil mapping using neural networks.** Pesquisa Agropecuária Brasileira 45: 497-507 (in Portuguese, with abstract in English). 2010.
- CHAGAS, C. S.; FERNANDES FILHO, E. I.; VIEIRA, C. A. O. **Comparison between artificial neural networks and maximum likelihood classification in digital soil mapping.** Revista Brasileira de Ciência do Solo 37: 339-351. 2013.
- CHASMER, L.; HOPKINSON, C.; VENESS, T.; QUINTON, W.; BALTZER, J. **A decision-tree classification for low-lying complex land cover types within the zone of discontinuous permafrost.** Remote Sensing of Environment 143: 73–84. 2014.
- COELHO, F. F. **Comparação de métodos de mapeamento digital de solos através de variáveis geomorfométricas e sistemas de informações geográficas.** Dissertação de Mestrado em Sensoriamento Remoto e Meteorologia. Universidade Federal do Rio Grande do Sul. Porto Alegre, 80 f. 2010.
- CONGALTON, R.G., GREEN, K., **Assessing the Accuracy of Remotely Sensed Data: Principles and Practices.** Lewis Publishers, New York (160 pp.). 1999.
- CONRAD, O., BECHTEL, B., BOCK, M., DIETRICH, H., FISCHER, E., GERLITZ, L., WEHBERG, J., WICHMANN, V., and BOEHNER, J. (2018): **System for Automated Geoscientific Analyses (SAGA) v. 7.0.0.** Geosci. Model Dev., 8, 1991-2007. 2018.
- DEUTSCH, C. V., JOURNEL, A. J. **Geostatistical Software Library and User's Guide.** Oxford Univ. Press, 340 pp. 1992.
- DOKUCHAEV, V. V., 1883/1948/1967. **Russian Chernozem,** in, Selected works of V. V. Dokuchaev, Moscow, 1948, 1, 14–419. Jerusalem: Israel Program for Scientific Translations Ltd. (for USDA-NSF), Publ. by S. Monson, 1967. (Transl. into English by N.Kaner).
- DRAY, S., SIBERCHICOT, M. A., **Package 'ADE4'.** <https://cran.r-project.org/web/packages/ade4/ade4.pdf>. 409p. 2018
- ELKAN, C. **Evaluating classifiers.** San Diego: University of California; 2012.
- EMBRAPA. Centro Nacional de Pesquisa de Solos. **Mapeamento digital de classes e atributos de solos: métodos, paradigmas e novas técnicas.** Rio de Janeiro, RJ: Embrapa Solos. 19p. (Embrapa Solos. Documentos; n. 55) 2003.

EMBRAPA. Centro Nacional de Pesquisa de Solos. **Embrapa Solos e suas origens**. Rio de Janeiro, RJ: Embrapa Solos. Disponível no site: <https://www.embrapa.br/solos/historia>. (acesso 03/04/2019).

EMBRAPA. Centro Nacional de Pesquisa de Solos. **Levantamento de reconhecimento de baixa intensidade dos solos do município de Campo Grande** - Zoneamento agroecológico do Estado do Mato Grosso do Sul. Rio de Janeiro: Embrapa Solos, (Embrapa Solos. Relatório Técnico). 2009.

EMBRAPA. Centro Nacional de Pesquisa de Solos. **Levantamento de reconhecimento de baixa intensidade dos solos do município de Sidrolândia** - Zoneamento agroecológico do Estado do Mato Grosso do Sul. Rio de Janeiro: Embrapa Solos, (Embrapa Solos. Relatório Técnico). 2010.

ESRI. Environmental Systems Research Institute Inc. Redlands. CA. Software: **ArcGIS e ArcINFO** – v.10. 2010.

FAVROT, J. C. **A strategy for large scale soil mapping: the reference areas method**. Science du Sol 27: 351-368 (in French, with abstract in English). 1989.

GIASSON, E.; SARMENTO, E. C.; WEBER, E.; FLORES, C. A. e HASENACK, H. **Decision trees for digital soil mapping on subtropical basaltic steplands**. Sci. Agrí., 68:167- 174, 2011.

GINI, C. **Variabilità e Mutabilità**, Bologna: Università di Cagliari. 158p. 1912.

GRINAND, C.; ARROUAYS, D.; LAROCHE, B.; MARTIN, M. P. **Extrapolating regional soil landscapes from an existing soil map: sampling intensity, validation procedures, and integration of spatial context**. Geoderma 143: 180-190, 2008.

HANSEN, M. K., BROWN, D. J., DENNISON, P. E., GRAVES, S. A., BRICKLEMYER, R. S., **Inductively mapping expert-derived soil-landscape units within dambo wetland catenae using multispectral and topographic data**. Geoderma 150 (1–2), 72–84. <http://dx.doi.org/10.1016/j.geoderma.2009.01.013>. 2009.

HENGL, T. E HEUVELINK, G. B. M. **New Challenges for Predictive Soil Mapping**. Anais. Global Workshop on Digital Soil Mapping. September 14-17, Montpellier AGRO-M/INRA Campus. 2004.

HENGL, T., DE JESUS, J. M., MACMILLAN, R. A., BATJES, N. H., HEUVELINK, G. B. M., RIBEIRO, E., SAMUEL-ROSA, A., KEMPEN, B., LEENAARS, J. G. B., WALSH, M. G., GONZALEZ, M. R., **SoilGrids1km** — global soil information based on automated mapping. 2014.

HENGL, T., HEUVELINK, G. B. M., KEMPEN, B., LEENAARS, J. G. B., TAMENE, L., TONDOH, J. E., **Mapping soil properties of Africa at 250 m resolution: random forests significantly improve current predictions**. 2015.

HÖFIG, P.; GIASSON, E.; VENDRAME, P. R. S. **Mapeamento digital de solos com base na extrapolação de mapas entre áreas fisiograficamente semelhantes**. Pesquisa Agropecuária Brasileira, v.49, p.958-966, 2014. DOI: 10.1590/S0100-204X2014001200006.

- IBGE. Instituto Brasileiro de Geografia e Estatística. **Mapas Temáticos do Brasil 1:250.000**. Disponível em <<https://mapas.ibge.gov.br/tematicos.html>>. Consulta em: 10/04/2017.
- INPE, Manuais; **Tutorial de geoprocessamento**; Operações aritméticas, disponível em: <http://www.dpi.inpe.br/spring/portugues/tutorial/operacao.html>. Acesso em 09/09/2018b.
- ISAAKS, E .H., SRIVASTAVA, R. M., **An Introduction to Applied Geostatistics**. Oxford, New York, 561 pp. 1989.
- JAFARI, A., KHADEMI, H., FINKE, P. A., VAN DE WAUW, J., AYOUBI, S. **Spatial prediction of soil great groups by boosted regression trees using a limited point dataset in an arid region, southeastern Iran**. *Geoderma* 232-234, 148–163. doi:10.1016/j.geoderma.2014.04.029. 2014.
- JENNY, H. **Factors of soil formation: a system of quantitative pedology**. New York: McGraw-Hill. 281p. 1941.
- JENSEN, J.R. **Sensoriamento remoto do ambiente: uma perspectiva em recursos terrestres**. São José dos Campos: Parêntese, p 604, 2009.
- JEONG, G.; OEVERDIECK, H.; PARK, S. J.; HUWE, B.; LIESS, M. **Spatial soil nutrients prediction using three supervised learning methods for assessment of land potentials in complex terrain**. *Catena*, v. 154, p. 73–84, 2017.
- JEUNE, W.; FRANCELINO, M. R.; SOUZA, E. DE; FERNANDES FILHO, E. I.; ROCHA, G. C. **Multinomial logistic regression and random forest classifiers in digital mapping of soil classes in western Haiti**. *Revista Brasileira de Ciencia do Solo*, v. 42, p. 1–20, 2018.
- JOURNEL, A.J., HUIJBREGTS, C.H. J. **Mining Geostatistics**. Academic Press. New York. 1978.
- KEMPEN, B; BRUS, D. J.; HEUVELINK, G. B. M.; STOORVOGEL, J. J. **Updating the 1:50,000 Dutch soil map using legacy soil data: a multinomial logistic regression approach**. *Geoderma* 151: 311-326. 2009.
- KINCAID, T., OLSEN, T., STEVENS, D., PLATT, C., WHITE, C., REMINGTON, R. **Spsurvey package: Spatial Survey Design and Analysis**, CRAN r-project. 193p. 2014.
- KOMSTA, L. **Package ‘outliers’**. <https://cran.r-project.org/web/packages/outliers/outliers.pdf> [20 de Fevereiro, 2015] 15p. 2015
- KÖPPEN, W. **Climatologia: con um estudio de los climas de la Tierra**. México: Fondo de Cultura Economica, 478 p. 1948.
- LACERDA FILHO, J. W.; BRITO, R. S.C.; SILVA, M.G.; OLIVEIRA, C.C. DE; MORETON, L.C.; MARTINS, E.G.; LOPES, R.C.; LIMA, T.M.; LARIZZATTI, J.H.; VALENTE, C.R. **Geologia e Recursos Minerais do Estado de Mato Grosso do Sul**. Programa Integração, Atualização e Difusão de Dados de Geologia do Brasil. Convênio CPRM/SICME-MS, MME, 121 p. 2006.
- LAGACHERIE, P.; LEGROS, J. P.; BURROUGH, P. A. **A soil survey procedure using the knowledge on soil pattern of a previously mapped reference area**. *Geoderma* 65: 283-301, 1995.

- LAGACHERIE, P. e MCBRATNEY, A. B. **Spatial soil information systems and spatial soil inference systems: Perspectives for digital soil mapping.** In: LAGACHERIE, P.; MCBRATNEY, A. e VOLTZ, M., eds. Digital soil mapping: An introductory perspective. Amsterdam, Elsevier, p.3-22. 2007.
- LASTORIA, G. **Hidrogeologia da Formação Serra Geral no Estado de Mato Grosso do Sul.** Rio Claro, 2002. 133 p. Tese (Doutoramento em Geociências e Meio Ambiente). Instituto de Geociências e Ciências Exatas – IGCE – UNESP – Rio Claro, 2002.
- LEMERCIER, B.; LACOSTE, M.; LOUM, M; WALTER, C. **Extrapolation at regional scale of local soil knowledge using boosted classification trees: A two-step approach, Geoderma,** In Press, Corrected Proof. 2011.
- LEPSCH, I. F. **Formação e conservação dos solos.** São Paulo, Oficina de Textos, 178p. 2002.
- LIAW, A.; WIENER, M. **Random Forest: Breiman and Cutler’s Random Forests for Classification and Regression.** URL <http://cran.r-project.org/web/packages/randomForest/index.html>. 2009.
- MACHADO, F. B., NARDY, A. J. R., ROCHA-JÚNIOR, E. R. V., MARQUES, L. S., and OLIVEIRA, M. A. F.: **Geologia e Litogeoquímica da Formação Serra Geral nos Estados de Mato Grosso e Mato Grosso do Sul,** Geociências (São Paulo. Impresso), 28, 523–540, 2009.
- MATHERON, G., **The theory of regionalised variables and its applications.** Les Cahiers du centre de morphologie mathématique de Fontainebleau. Ecole Nationale Supérieure des Mines de Paris. 1971.
- MATO GROSSO DO SUL. SECRETARIA DE PLANEJAMENTO E COORDENAÇÃO GERAL - SEPLAN-MS. **Macrozoneamento geoambiental do Estado de Mato Grosso do Sul.** Campo Grande, 242 p. 1989.
- McBRATNEY, A.B.; MENDONÇA-SANTOS, M.L.; MINASNY, B. **On digital soil mapping.** Geoderma, v.117, p. 3-52. 2003.
- MEIER, M.; SOUZA, E.; FRANCELINO, M. R.; FERNANDES FILHO, E. I.; SCHAEFER, C. E. G. **Digital soil mapping using machine learning algorithms in a tropical mountainous area.** Revista Brasileira de Ciência do Solo, v. 42, p. 1–23, 2018.
- MILANI, E. J.; FRANCA, A.; SCHNEIDER, R. **Bacia do Paraná.** Boletim de Geociências da Petrobrás 8 (1), 69-82. 1994.
- MILANI E. D.; MELO J. H. G.; SOUZA P. A.; FERNANDES L. A.; FRANÇA A. B. **Bacia do Paraná.** Boletim de Geociências da Petrobras, 15(2): 265-287. 2007.
- MINASNY, B.; McBRATNEY, A.B.; SANTOS, M. de L.M.; SANTOS H.G. dos. **Revisão sobre Funções de Pedotransferência (PTFs) e Novos Métodos de Predição de Classes de Solos e Atributos do Solo.** Rio de Janeiro: Embrapa Solos, 50 p. (Embrapa Solos. Documentos n. 45). 2003.

- MINASNY, B. e McBRATNEY, A. B. **A conditioned Latin hypercube method for sampling in the presence of ancillary information.** Computers e Geosciences, v.32. p.1378-1388. 2006.
- MINASNY, B.; McBRATNEY, A. B.; LARK, M. **Digital Soil Mapping Technologies for Countries with Sparse Data Infrastructures.** In: Digital Soil Mapping With Limited Data. HARTEMINK, A. E.; McBRATNEY, A. B.; MENDONÇA-SANTOS, M. de L. (Eds.), Springer. p. 15-30. 2008.
- MINASNY, B.; MCBRATNEY, A. B. 2015. **Digital soil mapping: a brief history and some lessons.** Geoderma. In press.
- MORAN C. J.; BUI E. N.; **Spatial data mining for enhanced soil map modelling.** Int J Geogr Inf Sci.;16:533-49. 2002.
- MORA-VALLEJO, A., CLAESSENS, L., STOOORVOGEL, J., HEUVELINK, G. B. M. **Small scaledigital soil mapping in Southeastern Kenya.** Catena, v. 76, p. 44-53. 2008.
- OLAYA, V. **A gentle introduction to SAGA GIS.** 2004. 202p. Disponível em: <http://downloads.sourceforge.net/saga-gis/SagaManual.pdf>. Acesso em 04/07/18.
- PEATE, D. W.; HAWKESWORTH, C. J.; MANTOVANI, M. SM. **Chemical stratigraphy of the Paraná lavas (South América): classification of magma types and their spatial distribution.** Bulletin of Volcanology, Berlin, v.55, n. 1, p. 119-139. 1992.
- PEBESMA, E., BIVAND, R., ROWLINGSON, B., RUBIO, V. G., HIJMANS, R., SUMNER, M., MACQUENN, D., LEMON, J., O'BRIEN, J., O'ROURKE, J., **Package 'SP'.** <https://cran.r-project.org/web/packages/sp/sp.pdf> 118p. 2018
- PEDREGOSA, F.; VAROQUAUX, G.; GRAMFORT, A.; MICHEL, V.; THIRION, B.; GRISEL, O.; BLONDEL, M.; PRETTENHOFER, P.; WEISS, R.; DUBOURG, V.; VANDERPLAS, J.; PASSOS, A.; COURNAPEAU, D.; BRUCHER, M.; PERROT, M.; DUCHESNAY, E. **Scikit-learn: Machine Learning in Python.** Journal of Machine Learning Research, v. 12, p. 2825-2830, 2011.
- PINHEIRO, H.S.K. **Mapeamento digital de solos por redes neurais artificiais da Bacia Hidrográfica do Rio Guapi-Macacu, RJ.** Dissertação de Mestrado. Universidade Federal Rural do Rio de Janeiro. 151p. 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.** Tese de Doutorado. Universidade Federal Rural do Rio de Janeiro, 180f. 2015.
- PINHEIRO, H. S. K.; OWENS, P. R.; ANJOS, L. H. C.; CARVALHO JÚNIOR, W.; CHAGAS, C. S. **Tree-based techniques to predict soil units,** Soil Research 55(8) 788-798. 2017.
- PRONASOLOS - **Programa Nacional de Solos do Brasil** / José Carlos Polidoro ... [et al]. – Dados eletrônicos. – Rio de Janeiro: Embrapa Solos, 53 p. – (Documentos / Embrapa Solos, ISSN 1517-2627; 183). 2016.

R CORE TEAM, **R: A Language and Environment for Statistical Computing**. Vienna, Austria: R Foundation for Statistical Computing. 2017.

R CORE TEAM, **Package 'FOREIGN'**. <https://cran.r-project.org/web/packages/foreign/foreign.pdf>. 25p. 2018

RAD, M. R. P.; TOOMANIAN, N.; KHORMALI, F.; BRUNGARD, C. W.; KOMAKI, C. B.; BOGAERT, P. **Updating soil survey maps using random forest and conditioned Latin hypercube sampling in the loess derived soils of northern Iran**. *Geoderma*, v. 232–234, p. 97–106, 2014.

ROKACH, L.; MAIMON, O. **Data mining with decision trees**. Theory and applications. World Scientific Publishing. 2008.

SAMUEL-ROSA, A.; HEUVELINK, G. B. M.; VASQUES, G. M.; ANJOS, L. H. C. **Do more detailed environmental covariates deliver more accurate soil maps?** *Geoderma*, v. 243–244, p.214–227. 2015

SANTOS, H. G., JACOMINE, P. K. T, ANJOS, L. H. C., OLIVEIRA, V. A., LUMBRERAS, J.F, COELHO, M. Z., ALMEIDA, J. A., ARAUJO FILHO, J. C., OLIVEIRA, J. B., CUNHA, T. J. F. **Sistema brasileiro de classificação de solos**. 5. ed. Brasília: Embrapa, 356 p. 2018.

SANTOS, M. de L. M., SANTOS, H. G. dos, **Mapeamento digital de classes e atributos de solos**. Embrapa solos. Documentos 55. 2003.

SANTOS, M. L. M., SANTOS, H. G., DART, R.de O, PARES, J. G. **Mapeamento Digital de Classes de Solos no Estado do Rio de Janeiro**. 28p. (Boletim de Pesquisa e Desenvolvimento 119) Dezembro, 2007

SARMENTO, E. C.; GIASSON, E.; WEBER, E. J.; FLORES, C. A.; ROSSITER, D. G.; HASENACK, H. **Caracterização de mapas legados de solos: uso de indicadores em mapas com diferentes escalas no Rio Grande do Sul**. *Revista Brasileira de Ciência do Solo*, v.38, p.1672-1680, 2014. DOI: 10.1590/S0100-06832014000600002.

SCHMIDT, K. BEHRENS, T. SCHOLTEN, T. **Instance selection and classification tree analysis for large spatial datasets in digital soil mapping**. *Geoderma*, v.146, p.138-146, 2008.

SILVA, V. A., CURTI, N., MARQUES, J. J. G., CARVALHO, L. M. T. D., SANTOS, W. J. R., **Soil maps, field knowledge, forest inventory and ecological-economic zoning as a basis for agricultural suitability of lands in Minas Gerais elaborated in GIS**. *Ciênc. Agrotecnologia* 37, 538–549. 2013.

SILVA, S. H. G., OWENS, P. R., DUARTE DE MENEZES, M., REIS SANTOS, W. J., CURTI, N., **A technique for low cost soil mapping and validation using expert knowledge on a watershed in Minas Gerais, Brazil**. *Soil Sci. Soc. Am. J.* 78, 1310–1319. 2014.

SINDAYIHEBURA, A.; OTTOY, S.; DONDEYNE, S.; MEIRVENNE, M. V.; ORSHOVEN, J.V. **Comparing digital soil mapping techniques for organic carbon and clay content: Case study in Burundi's central plateaus**. *Catena*, v. 156, p. 161–175. 2017.

SINGH, A.N. e DWIVEDI, R. S. **The utility of LANDSAT imagery as an integral part of the data base for small-scale soil mapping.** International Journal of Remote Sensing, v. 7, n. 9, p.1099-1108, 1986.

STEYERBERG, E. W. **Validation of prediction models.** In: Steyerberg EW, editor. Clinical prediction models. New York: Springer p.299-311. 2009.

STROBL, C.; BOULESTEIX, A. L.; ZEILEIS, A.; HOTHORN, T. **Bias in random forest variable importance measures: Illustrations, sources and a solution.** BMC bioinformatics, v. 8, n. 1, p.25, 2007.

STUMPF, F.; SCHMIDT, K.; BEHRENS, T.; SCHÖNBRODT-STITT, S.; BUZZO, G.; DUMPERTH, C.; WADOUX, A.; XIANG, W.; SCHOLTEN, T. Incorporating limited field operability and legacy soil samples in a hypercube sampling design for digital soil mapping. **Journal of Plant Nutrition and Soil Science**, v. 000, p. 1–11, 2016.

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 253-254, 67–77. 2015

TEN CATEN, A. **Mapeamento digital de solos: metodologias para atender a demanda por informação espacial em solos.** Tese de Doutorado. Universidade Federal de Santa Maria. 108p. 2011.

TEN CATEN, A.; DALMOLIN, R. S. D.; PEDRON, F. de A.; RUIZ, L. F. C.; SILVA, C. A. da. **An appropriate data set size for digital soil mapping in Erechim, Rio Grande do Sul, Brazil.** Revista Brasileira de Ciência do Solo, v.37, p.359-366, 2013. DOI: 10.1590/S0100-06832013000200007.

THERNEAU, T.; ATKINSON, B.; RIPLEY, B. **RPART: Recursive Partitioning and Regression Trees.** (Author of initial R port) 2015-02-24. 2015.

TESKE, R.; GIASSON, E.; BAGATINI, T. **Comparação de esquemas de amostragem para treinamento de modelos preditores no mapeamento digital de classes de solos,** Rio Grande do Sul, Brazil. Revista Brasileira de Ciência do Solo, v.39, p.14-20, 2015.

USGS GLOBAL VISUALIZATION VIEWER (Usa). **Frequently Asked Questions about the Landsat Missions.** 2013. Disponível em: <[http://landsat.usgs.gov/L8\\_band\\_combos.php](http://landsat.usgs.gov/L8_band_combos.php)>. Acesso em: 22 de novembro. 2017.

VERMOTE, E. F., TANRE, D., DEUZE, J. L., HERMAM, M., MORCRETTE, J. J., **Second Simulation of Satellite Signal in the Solar Spectrum, 6S: An overview.**, IEEE Transactionson Geoscience and Remote Sensing v 35, n 3, p 675-686, 1997.

WEBSTER, R., OLIVER, M. A., **Statistical Methods in Soil and Land Resources Survey.** Oxford Univ. Press. 1990.

WERE, K.; BUI, D. T.; DICK, Ø. B.; SINGH, B. R. **A comparative assessment of support vector regression, artificial neural networks, and random forests for predicting and mapping soil organic carbon stocks across an Afromontane landscape.** Ecological Indicators, v. 52, p. 394–403, 2015.

YANG, R. M.; ZHANG, G. L.; LIU, F.; LU, Y. Y.; YANG, F.; YANG, F.; YANG, M.; ZHAO, Y. G.; LI, D. C. **Comparison of boosted regression tree and random forest models for mapping topsoil organic carbon concentration in an alpine ecosystem.** Ecological Indicators, v. 60, p. 870–878, 2016.

YANG, W.; YANG, L.; MERCHANT, J.W. **An assessment of AVHRR/NDVI-ecoclimatological relations in Nebraska.** USA. International Journal of Remote Sensing, v.10. p. 2161-2180. 1997.

YIGINI, Y.; PANAGOS, P., **Reference area method for mapping soil organic carbon content at regional scale.** Procedia Earth and Planetary Science 10: 330-338, 2014.

YOUNG, P.; PARKINSON, S.; LEES, M. Simplicity out of complexity in environmental modelling: Occam's razor revisited. **Journal of Applied Statistics**, 23(2-3), 165-210. 1996.

## 8 ANEXO

### 8.1. Perfis Representativos das Classes de Solo Identificadas (extraído dos relatórios de campo do levantamento dos municípios de Sidrolândia e Campo Grande, MS)

#### 8.1.1. GLEISSOLO HÁPLICO Tb Distrófico típico (UM-GXbd)

**Número de campo** – PCG-04

**Data** – 21.09.2007

**Classificação atual** – GLEISSOLO HÁPLICO Tb Distrófico típico, textura arenosa/média, A moderado, álico, fase campo higrófilo relevo plano.

**Unidade de mapeamento** – GXbd

**Localização, município, estado e coordenadas** – A 7 km do anel viário de Campo Grande (saída sudoeste, junto ao Bairro Morada do Sol), em direção à Faz. Varjão, lado esquerdo 50 metros. Campo Grande, MS. Folha IBGE Sidrolândia (MI 2587). 0747530 e 7717304 (fuso 21).

**Situação, declive e cobertura vegetal sobre o perfil** – Trincheira aberta em área plana ligeiramente deprimida, sob pastagem.

**Litologia** – Sedimentos areno-argilosos.

**Formação geológica** – Depósitos quaternários.

**Cronologia** – Quaternário

**Material originário** – Sedimentos areno-argilosos quaternários.

**Pedregosidade** – Não pedregosa.

**Rochosidade** – Não rochosa.

**Relevo local** – Plano.

**Relevo regional** – Plano.

**Erosão** – Não aparente.

**Drenagem** – Mal drenado.

**Vegetação primária** – Campo higrófilo de várzea.

**Uso atual** – Pastagem.

**Descrito e coletado por** – Paulo E. F. da Motta e Aílton

#### Descrição Morfológica

- Ap - 0 - 10 cm; bruno-acinzentado muito escuro (10YR 3/2); areia franca; fraca, pequena e média granular; ligeiramente dura, friável, não plástica e não pegajosa; transição plana e abrupta.
- AE - 10 - 22 cm; bruno-acinzentado muito escuro (10YR 3/2); mosqueado comum, pequeno e difuso, vermelho-amarelado (5YR 5/8); areia franca; fraca, pequena e média blocos subangulares; ligeiramente dura, friável, não plástica e não pegajosa; transição plana e clara.
- E1 - 22 - 36 cm; bruno (10YR 5/3); mosqueado comum, pequeno e difuso, amarelo-avermelhado (5YR 6/8); areia franca; fraca, pequena e média blocos subangulares; ligeiramente dura, muito friável, não plástica e não pegajosa; transição plana e gradual.
- E2 - 36 - 46 cm; bruno-amarelado-claro (2,5Y 6/3); mosqueado comum, pequeno e difuso, amarelo-avermelhado (5YR 6/8); areia franca; fraca pequena e média blocos subangulares; ligeiramente dura, muito friável, ligeiramente plástica e ligeiramente pegajosa; transição plana e clara.

- Btg1 - 46 – 58 cm, bruno-amarelado-claro (2,5Y 6/3); mosqueado comum, pequeno e difuso, amarelo-avermelhado (5YR 6/8); franco-arenosa; fraca pequena e média blocos subangulares; friável, ligeiramente plástica e ligeiramente pegajosa; transição plana e abrupta.
- Btgf1 - 58 – 130 cm, cinzento-brunado-claro (10YR 6/3); mosqueado muito, grande e proeminente, amarelo-brunado (10YR 6/8); franco-arenosa; muito friável, ligeiramente plástica e ligeiramente pegajosa; transição ondulada e abrupta.
- Btg2 - 130 – 175 cm, cinzento-brunado-claro (2,5Y 6/2); mosqueado pouco, médio e proeminente, amarelo-brunado (10YR 6/8); franco-arenosa; muito friável, ligeiramente plástica e ligeiramente pegajosa; transição ondulada e abrupta.
- Btgf2 - 175 – 200 cm<sup>+</sup>, cinzento-brunado-claro (2,5Y 6/2); mosqueado muito, médio e proeminente, amarelo-brunado (10YR 6/8); franco-arenosa; muito friável, ligeiramente plástica e ligeiramente pegajosa.
- Raízes - Muitas finas no horizonte Ap; comuns finas no AE; poucas finas no E1, E2 e Bg1; raras finas no Bgf1, Bg2 e Bgf2.
- Observações - Os mosqueados nos horizontes AE, E1, E2 e Bg1 ocorrem preferencialmente em torno dos canais das raízes.
- Pontuações mais avermelhadas distribuídas nos horizontes Bgf1, Bg2 e Bgf2 correspondem a plintita, cuja ocorrência, no entanto, corresponde a menos que 5 % em volume da massa solo, quantidade mínima para caracterizar caráter plíntico, conforme o SiBCS.
  - O lençol freático encontrava-se abaixo de 200 cm. Solo úmido do Bg1 para baixo e horizonte Bg2 consistência quase pastosa

**Análises Físicas e Químicas**

Horizonte		Frações da amostra total g/kg			Composição granulométrica da terra fina (g/kg)				Argila dispersa em água g/kg	Grau de flocculação %	Silte/Argila	Densidade g/cm <sup>3</sup>		Poros cm <sup>3</sup> /100cm <sup>3</sup>
Símbolo	Prof. cm	Calhaus > 20 mm	Cascalho 20-2 mm	Terra fina < 2 mm	Areia grossa 2-0,20 mm	Areia fina 0,20-0,05 mm	Silte 0,05-0,002 mm	Argila < 0,002 mm				Solo	Partículas	
Ap	0-10	0	0	1000	641	217	82	60	40	33	1,37			
AE	-22	0	0	1000	706	171	23	100	40	60	0,23	1,42	2,58	45
E1	-36	0	0	1000	681	185	34	100	40	60	0,34	1,39	2,64	47
E2	-46	0	0	1000	602	229	49	120	0	100	0,41			
Btg	-58	0	0	1000	563	223	53	161	0	100	0,33	1,36	2,63	48
Btgf1	-130	0	0	1000	541	231	67	161	0	100	0,42	1,44	2,62	45
Btg2	-175	0	0	1000	577	209	53	161	0	100	0,33			
Btgf2	-200	0	0	1000	543	225	51	181	0	100	0,28			
Horiz.	pH (1:2,5)		Complexo Sortivo cmol/kg								Valor V %	Al sat %	P mg/kg	
	Água	KCl 1N	Ca <sup>2+</sup>	Mg <sup>2+</sup>	K <sup>+</sup>	Na <sup>+</sup>	Valor S	Al <sup>3+</sup>	H <sup>+</sup>	Valor T				
Ap	4,8	4,0	0,4		0,04	0,01	0,4	0,6	2,5	3,5	11	60	4	
AE	4,9	4,2	0,2		0,01	0,01	0,2	0,4	1,4	2,0	10	67	2	
E1	4,7	4,2	0,2		0,01	0,01	0,2	0,4	1,3	1,9	11	67	1	
E2	4,7	4,1	0,2		0,01	0,01	0,2	0,5	1,0	1,7	12	71	1	
Btg	4,8	4,2	0,2		0,01	0,01	0,2	0,5	1,3	2,0	10	71	1	
Btgf1	5,1	4,5	0,2		0,01	0,01	0,2	0,1	0,9	1,2	17	33	1	
Btg2	5,2	4,2	0,2		0,01	0,01	0,2	0,3	0,9	1,4	14	60	1	
Btgf2	5,2	4,2	0,3		0,01	0,01	0,3	0,3	0,9	1,5	20	50	1	
Horiz.	C (org) g/kg	N g/kg	C/N	Ataque sulfúrico g/kg				Relações Moleculares			Fe <sub>2</sub> O <sub>3</sub> livre g/kg	Equivalente de CaCO <sub>3</sub> g/kg		
				SiO <sub>2</sub>	Al <sub>2</sub> O <sub>3</sub>	Fe <sub>2</sub> O <sub>3</sub>	TiO <sub>2</sub>	SiO <sub>2</sub> /Al <sub>2</sub> O <sub>3</sub> (Ki)	SiO <sub>2</sub> /R <sub>2</sub> O <sub>3</sub> (Kr)	Al <sub>2</sub> O <sub>3</sub> /Fe <sub>2</sub> O <sub>3</sub>				
Ap	7,7	0,9	9											
AE	3,7	0,4	9											
E1	3,0	0,4	7											
E2	2,6	0,3	9											
Btg	2,8	0,3	9											
Btgf1	1,4	0,2	7	57	62	18	6,8	1,56	1,32	5,41				
Btg2	0,9	0,2	4	61	68	12	6,5	1,52	1,37	8,90				
Btgf2	0,9	0,2	4											

### 8.1.2. LATOSSOLO AMARELO Distrófico plúntico (UM-LAd)

**Número de campo** – ECG-08

**Data** – 20.11.2006

**Classificação atual** – LATOSSOLO AMARELO Distrófico plúntico, textura argilosa, A moderado, imperfeitamente drenado, fase campo higrófilo de surgente relevo plano.

**Unidade de mapeamento** – LAd

**Localização, município, estado e coordenadas** – Coordenadas UTM 734083 x 7715278. Campo Grande.

**Situação, declive e cobertura vegetal sobre o perfil** – Microtrincheira aberta em superfície aplainada, com 0-2% de declive sob vegetação natural.

**Altitude** – 490

**Litologia** – Basalto

**Formação geológica** – Formação Serra Geral. Grupo São Bento.

**Cronologia** – Juro-cretáceo.

**Material originário** – Produto da decomposição do basalto.

**Pedregosidade** – Não pedregosa.

**Rochosidade** – Não rochosa.

**Relevo local** – Plano

**Relevo regional** – Plano e suave ondulado.

**Erosão** – Laminar ligeira.

**Drenagem** – Imperfeitamente drenado.

**Vegetação primária** – Campo higrófilo de surgente.

**Uso atual** – Pastagem.

**Clima** –

**Descrito e coletado por** – Paulo E. F. da Motta e Afilton

#### Descrição Morfológica:

A	0-18cm; bruno-acinzentado muito escuro (2,5Y 3/2); argila..
Bwf	40-60 cm; bruno-oliváceo-claro (2,5Y 5/4); mosqueado pequeno, abundante e proeminente vermelho-amarelado (5YR 4/6); argila.
Bgcf	80-100 cm; variegado composto de cinzento-brunado-claro (2,5Y 6/2, úmida), cinzento (2,5Y 6/1, úmida) e vermelho-amarelado (5YR 5/6, úmida); bruno-amarelado-claro (10YR 6/4, úmida amassada); argila cascalhenta.
Observações	O cascalho presente no horizonte Bgcf é constituído por petroplintita. O horizonte Bw estende-se de 20 a 70 cm de profundidade.

**Análises Físicas e Químicas**

Horizonte		Frações da amostra total g/kg			Composição granulométrica da terra fina g/kg				Argila dispersa em água g/kg	Grau de floculação %	Relação Silte/Argila	Densidade g/cm <sup>3</sup>		Porosidade cm <sup>3</sup> /100cm <sup>3</sup>
Símbolo	Profundidade cm	Calhaus > 20 mm	Cascalho 20-2 mm	Terra fina < 2 mm	Areia grossa 2-0,20 mm	Areia fina 0,20-0,05 mm	Silte 0,05-0,002 mm	Argila < 0,002 mm				Solo	Partículas	
A	0-20	0	0	1000	118	61	337	484	0	100	0,70			
Bwf	40-60	0	0	1000	182	111	150	557	0	100	0,27			
Bgcf	80-100	0	249	751	245	78	121	556	0	100	0,22			
Horizonte	pH (1:2,5)		Complexo Sortivo cmol <sub>c</sub> /kg								Valor V %	Al Sat %	P mg/kg	
	Água	KCl 1N	Ca <sup>2+</sup>	Mg <sup>2+</sup>	K <sup>+</sup>	Na <sup>+</sup>	Valor S	Al <sup>3+</sup>	H <sup>+</sup>	Valor T				
A	5,0	4,6	0,8	0,4	0,13	0,04	1,4	0,3	11,2	12,9	11	18	2	
Bwf	5,7	5,2	0,6		0,02	0,01	0,6	0	2,7	3,3	18	0	1	
Bgcf	5,3	5,2	0,6		0,01	0,01	0,6	0	2,7	3,3	18	0	1	
Horizonte	C (org) g/kg	N g/kg	C/N	Ataque sulfúrico g/kg				Relações Moleculares			Fe <sub>2</sub> O <sub>3</sub> livre g/kg	Equivalente de CaCO <sub>3</sub> g/kg		
				SiO <sub>2</sub>	Al <sub>2</sub> O <sub>3</sub>	Fe <sub>2</sub> O <sub>3</sub>	TiO <sub>2</sub>	SiO <sub>2</sub> /Al <sub>2</sub> O <sub>3</sub> (Ki)	SiO <sub>2</sub> /R <sub>2</sub> O <sub>3</sub> (Kr)	Al <sub>2</sub> O <sub>3</sub> /Fe <sub>2</sub> O <sub>3</sub>				
A	47,8	5,3	9						1,10	0,77	2,34			
Bwf	6,2	0,8	8	163	252	169	33,0	1,05	0,75	2,55				
Bgcf	3,5	0,4	9	161	261	161	35,0	1,10	0,77	2,34				

### 8.1.3. LATOSSOLO VERMELHO Distrófico típico (UM-LVd)

**Número de campo** – 06WN

**Data** – 31.05.2006

**Classificação atual** – LATOSSOLO VERMELHO Distrófico típico, textura média, A moderado, relevo plano.

**Unidade de mapeamento** – LVd

**Localização, município, estado e coordenadas** – Coordenadas UTM 694.582 x 7.710.537. Sidrolândia (Quebra-cocô).

**Situação, declive e cobertura vegetal sobre o perfil** – Topo plano

**Altitude** – 264 metros

**Litologia** – Sedimentos areno-argilosos.

**Formação geológica** –

**Material originário** – Quartzito e Basalto.

**Pedregosidade** – Não pedregosa.

**Rochosidade** – Não rochosa.

**Relevo local** – Plano.

**Relevo regional** – Plano.

**Erosão** – Não aparente a laminar ligeira.

**Drenagem** – Bem drenado.

**Vegetação primária** – Floresta/Cerradão

**Uso atual** – Pastagem e ao lado cana.

**Descrito e coletado por** – Nilson/ Waldir.

#### Descrição Morfológica

- Ap 0-18cm; bruno-avermelhado-escuro (2,5YR 3/3, seco); areia franca; moderada pequena blocos subangulares que se desfaz em forte pequena granular; ligeiramente plástica; não pegajosa; transição plana e clara.
- BA 18-34cm; bruno-avermelhado-escuro (2,5YR 3/4, seco); areia franca; fraca pequena blocos subangulares que se desfaz em forte pequena granular; ligeiramente plástica; não pegajosa; transição plana e gradual.
- Bw1 34-59cm; bruno-avermelhado-escuro (2,5YR 3/4, seco); franco-argilo-arenosa; fraca pequena blocos subangulares que se desfaz em forte pequena granular; ligeiramente plástica; não pegajosa; transição plana e difusa.
- Bw2 59-94cm; bruno-avermelhado-escuro (2,5YR 3/4, seco); franco-argilo-arenosa; fraca pequena blocos subangulares que se desfaz em forte pequena granular; ligeiramente plástica; não pegajosa; transição plana e difusa.
- Bw3 94-135cm+; vermelho-escuro (2,5YR 3/6, seco); franco-argilo-arenosa; forte pequena granular (pó de café); ligeiramente plástica; não pegajosa; transição plana e difusa.
- Bw4 160-190cm (trado); franco-argilo-arenosa; forte pequena granular (pó de café).

Poros Muitos pequenos e médios em todos os horizontes.

Raízes Fasciculadas. Muitas e finas no Ap, BA, Bw1 e Bw2; comuns e finas no Bw3 e Bw4.

Observações: Coletadas amostras indeformadas e anéis volumétricos de Ap, BA, Bw1 e Bw2.

**Análises Físicas e Químicas**

Horizonte		Frações da amostra total g/kg			Composição granulométrica da terra fina (g/kg)				Argila dispersa em água g/kg	Grau de flocculação %	Silte/Argila	Densidade g/cm <sup>3</sup>		Poros cm <sup>3</sup> /100cm <sup>3</sup>
Símbolo	Profundidade cm	Calhaus > 20 mm	Cascalho 20-2 mm	Terra fina < 2 mm	Areia grossa 2-0,20 mm	Areia fina 0,20-0,05 mm	Silte 0,05-0,002 mm	Argila < 0,002 mm				Solo	Partículas	
Ap	0-18	0	0	1000	310	507	22	161	121	25	0,14			
BA	-34	0	0	1000	301	507	31	161	161	0	0,19			
Bw1	-59	0	0	1000	274	469	56	201	201	0	0,28			
Bw2	-94	0	0	1000	276	479	44	201	60	70	0,22			
Bw3	-135	0	0	1000	270	479	50	201	0	100	0,25			
Bw4	160-190	0	0	1000	246	501	31	222	0	100	0,14			
Horizonte	pH (1:2,5)		Complexo Sortivo cmol <sub>c</sub> /kg								Valor V %	Al Sat %	P mg/kg	
	Água	KCl 1N	Ca <sup>2+</sup>	Mg <sup>2+</sup>	K <sup>+</sup>	Na <sup>+</sup>	Valor S	Al <sup>3+</sup>	H <sup>+</sup>	Valor T				
Ap	5,4	4,4	0,7	1,0	0,07	0,01	1,8	0,2	2,8	4,8	37	10	1	
BA	4,7	3,9	0,4		0,02	0,01	0,4	0,7	2,3	3,4	12	64	1	
Bw1	4,7	3,9	0,5		0,01	0,01	0,5	0,7	2,3	3,5	14	58	1	
Bw2	4,8	4,0	0,5		0,01	0,01	0,5	0,8	1,8	3,1	16	62	1	
Bw3	4,8	4,0	0,4		0,01	0,01	0,4	0,8	1,3	2,5	16	67	1	
Bw4	5,0	4,1	0,4		0,01	0,01	0,4	0,7	1,1	2,2	18	64	1	
Horizonte	C (org) g/kg	N g/kg	C/N	Ataque sulfúrico g/kg				Relações Moleculares			Fe <sub>2</sub> O <sub>3</sub> livre g/kg	Equivalent e de CaCO <sub>3</sub> g/kg		
				SiO <sub>2</sub>	Al <sub>2</sub> O <sub>3</sub>	Fe <sub>2</sub> O <sub>3</sub>	TiO <sub>2</sub>	SiO <sub>2</sub> /Al <sub>2</sub> O <sub>3</sub> (Ki)	SiO <sub>2</sub> /R <sub>2</sub> O <sub>3</sub> (Kr)	Al <sub>2</sub> O <sub>3</sub> /Fe <sub>2</sub> O <sub>3</sub>				
Ap	8,1	0,8	10	40	50	56	16,5	1,36	0,79	1,40				
BA	4,1	0,6	7	50	60	64	17,1	1,42	0,84	1,47				
Bw1	4,1	0,5	8	59	63	62	18,9	1,59	0,98	1,60				
Bw2	3,1	0,4	8	60	63	62	16,9	1,62	0,99	1,60				
Bw3	2,4	0,4	6	62	67	67	16,5	1,57	0,96	1,57				
Bw4	1,6	0,3	5	64	66	68	16,3	1,65	0,99	1,52				

#### 8.1.4. LATOSSOLO VERMELHO Distroférico típico (UM-LVdf)

**Número de campo** – CNPGC-01

**Data** – 13.07.1977 (EMBRAPA, 1979)

**Classificação atual** – LATOSSOLO VERMELHO Distroférico típico, textura muito argilosa, A moderado, álico, fase cerradão tropical subcaducifólio, relevo plano.

**Unidade de mapeamento** – LVdf

**Localização, município, estado e coordenadas** – Jardim de introdução, no prolongamento da estrada central das parcelas, a 60 m da mesma. Campo Grande, MS. Folha IBGE Campo Grande (MI 2551). 738268 e 7736851 (Fuso 21)

**Situação, declive e cobertura vegetal sobre o perfil** – Perfil descrito e coletado em trincheira situada em local praticamente plano, sob cobertura de pastagem (capim-jaraguá), em área desmatada há cerca de 10 anos

**Altitude** – 530 metros

**Litologia** – Rochas eruptivas básicas.

**Formação geológica** – Grupo São Bento

**Cronologia** – Triássico superior

**Material originário** – Desenvolvido a partir da decomposição de rochas eruptivas básicas (basalto), ligeiramente afetado por material de outra (s) fonte (s).

**Pedregosidade** – Não pedregosa.

**Rochosidade** – Não rochosa.

**Relevo local** – Plano.

**Relevo regional** – Plano e suave ondulado.

**Erosão** – Laminar ligeira.

**Drenagem** – Bem drenado.

**Vegetação primária** – Cerradão tropical subcaducifólio

**Uso atual** – Área experimental.

#### Descrição Morfológica

- A1 - 0 – 10 cm, vermelho-escuro-acinzentado (10 R 3/3, úmida e úmida amassada), bruno-avermelhado-escuro (2,5 YR 3/4, seco) e vermelho-amarelado (5 YR 4/6, seca triturada); argila; forte muito pequena a média granular e grãos simples; dura, friável, plástica e pegajosa; transição plana e gradual.
- A2 - 10 – 22 cm, bruno-avermelhado-escuro (2,5 YR 3/4, úmida e úmida amassada), bruno-avermelhado-escuro (2,5 YR 3/5, seca) e vermelho-amarelado (5 YR 4/6, seca triturada); argila; moderada pequena a média blocos subangulares e moderada a forte muito pequena a pequena granular; dura, friável, muito plástica e muito pegajosa; transição plana e clara.
- AB - 22 – 48 cm, bruno-avermelhado-escuro (2,5 YR 3/4, úmida e úmida amassada), bruno-avermelhado-escuro (2,5 YR 3/5, seco) e vermelho-amarelado (5 YR 4/6, seca triturada); muito argilosa; fraca pequena a média blocos subangulares e fraca pequena granular; ligeiramente dura, friável, muito plástica e muito pegajosa; transição plana e difusa.
- BA - 48 – 75 cm, vermelho-escuro (2,5 YR 3/5); muito argilosa; fraca pequena a média blocos subangulares com aspecto de maciça porosa pouco coesa “in situ”; ligeiramente dura, muito friável, muito plástica e muito pegajosa; transição plana e difusa.
- Bw1 - 75 – 110 cm, vermelho-escuro (2,5 YR 3/6); muito argilosa; fraca muito pequena granular com aspecto de maciça porosa muito pouco coesa “in situ”; ligeiramente dura, muito friável, plástica e pegajosa; transição plana e difusa.

- Bw2 - 110 – 170 cm, vermelho-escuro (2,5 YR 3/6); muito argilosa; fraca muito pequena granular com aspecto de maciça porosa muito pouco coesa “in situ”; ligeiramente dura, friável, plástica e pegajosa; transição plana e difusa.
- Bw3 - 170 – 310 cm+, vermelho-escuro (2,5 YR 3/6); muito argilosa fraca muito pequena granular com aspecto de maciça porosa muito pouco coesa “in situ”; ligeiramente dura, muito friável, plástica e pegajosa.

Raízes - Muitas raízes nos horizontes A1 e A2, comuns no AB e BA e raras no Bw1 e Bw2. Até 170 cm de profundidade ocorrem raras raízes com diâmetro em torno de 5 cm.

- Observações - Trincheira com 190 cm de profundidade, a partir dessa profundidade utilizou-se o trado holandês para coleta de amostra (190 + 120 = 310 cm).
- Presença de carvão até 60 cm de profundidade. Atividade biológica no A1, A2 e AB. Muitos poros muito pequenos e alguns médios ao longo do perfil

### **Análise mineralógica**

- A1 - Areia Grossa – 50% de quartzo, grãos subangulosos, subarredondados, arredondados e bem arredondados, superfícies irregulares, brilhantes e foscas, com aderência e incrustação ferruginosas, brancos, avermelhados e incolores; 50% de material areno-argilo-ferruginoso, hematítico e goetítico, magnetita em proporção dominante e detritos

Areia Fina – 50% de quartzo, grãos angulosos, subangulosos, subarredondados, superfícies irregulares, brilhantes e foscas, com aderência e incrustação ferruginosas, brancos, avermelhadas e incolores; 50% de material areno-argiloso-ferruginoso, hematítico, goetítico e argiloso branco, fragmentos silicosos, magnetita em proporção dominante e detritos.

- Bw1 - Areia Grossa – 50% de quartzo, grãos subangulosos, subarredondados, arredondados e bem arredondados, superfícies irregulares, brilhantes e foscas, com aderência e incrustação ferruginosas, brancos e incolores; 50% de material areno-argilo-ferruginoso hematítico, limonítico, magnetita em proporção dominante e detritos.

- Areia Fina – 50% de quartzo, grãos angulosos, subangulosos, subarredondados, arredondados, superfícies irregulares, brilhantes e foscas, com aderência e incrustação ferruginosas, brancos, avermelhados e incolores; 50% de material areno-argilo-ferruginoso, hematítico, limonítico e argiloso branco, turmalina verde-oliva, magnetita em proporção dominante, rutilo avermelhado e detritos

- Bw2 - Areia Grossa - 50% de quartzo, grãos angulosos, subangulosos, subarredondados, arredondados e bem arredondados, superfícies irregulares, brilhantes e foscas, com aderência e incrustação ferruginosas, brancos, avermelhados e incolores; 50% de material areno-argilo-ferruginoso, hematítico e limonítico, magnetita em proporção dominante e detritos.

Areia Fina – 50% de quartzo, grãos angulosos, subangulosos, subarredondados, arredondados, superfícies irregulares, brilhantes e foscas, com aderência e incrustação ferruginosas, brancos, avermelhados e incolores; 50% de material areno-argilo-ferruginoso, hematítico, limonítico e argiloso branco, magnetita em proporção dominante, turmalina verde-oliva, fragmento silicoso, estaurólita (?) e detritos.

### Análises Físicas e Químicas

Horizonte		Frações da amostra total g/kg			Composição granulométrica da terra fina g/kg				Argila dispersa em água g/kg	Grau de flocculação %	Relação Silte/Argila	Poros cm <sup>3</sup> /100cm <sup>3</sup>	
Símbolo	Profundidade cm	Calhaus > 20 mm	Cascalho 20-2 mm	Terra fina < 2 mm	Areia grossa 2-0,20 mm	Areia fina 0,20-0,05 mm	Silte 0,05-0,002 mm	Argila < 0,002 mm					
A1	0 - 10	0	0	1000	170	160	120	550	360	35	0,22		
A2	- 22	0	0	1000	160	170	100	570	270	53	0,18		
AB	- 48	0	0	1000	140	140	100	620	390	37	0,16		
BA	- 75	0	0	1000	130	140	100	630	0	100	0,16		
Bw1	- 110	0	0	1000	140	130	90	640	0	100	0,14		
Bw2	- 170	0	0	1000	130	130	90	650	0	100	0,14		
Bw3	- 310+	0	0	1000	130	140	110	620	0	100	0,18		
Horizonte	pH (1:2,5)		Complexo Sortivo cmol <sub>c</sub> /kg								Valor V %	Al Sat %	P mg/kg
	Água	KCl 1N	Ca <sup>2+</sup>	Mg <sup>2+</sup>	K <sup>+</sup>	Na <sup>+</sup>	Valor S	Al <sup>3+</sup>	H <sup>+</sup>	Valor T			
A1	5,4	4,2	2,3	0,6	0,17	0,03	3,1	1,2	7,2	11,5	27	28	3
A2	5,2	4,0	0,7		0,07	0,02	0,8	2,1	4,0	6,9	12	72	1
AB	5,0	4,0	0,2		0,04	0,02	0,3	1,9	3,9	6,1	5	86	<1
BA	4,9	4,1	0,6		0,03	0,03	0,7	1,6	3,1	5,4	13	70	<1
Bw1	4,8	4,1	0,5		0,02	0,02	0,5	1,3	2,7	4,5	11	72	<1
Bw2	5,2	4,2	0,5		0,02	0,05	0,6	0,8	1,9	3,3	18	57	<1
Bw3	5,4	4,4	0,3		0,01	0,03	0,3	0,3	1,9	2,5	12	50	<1
Horizonte	C (org) g/kg	N g/kg	C/N	Ataque sulfúrico g/kg				Relações Moleculares			Equivalente de CaCO <sub>3</sub> g/kg		
				SiO <sub>2</sub>	Al <sub>2</sub> O <sub>3</sub>	Fe <sub>2</sub> O <sub>3</sub>	TiO <sub>2</sub>	SiO <sub>2</sub> /Al <sub>2</sub> O <sub>3</sub> (Ki)	SiO <sub>2</sub> /R <sub>2</sub> O <sub>3</sub> (Kr)	Al <sub>2</sub> O <sub>3</sub> /Fe <sub>2</sub> O <sub>3</sub>			
A1	20,7	1,9	11	153	177	298	412	1,47	0,71	0,93			
A2	12,4	1,0	12	160	188	303	393	1,45	0,71	0,97			
AB	9,3	0,9	10	163	193	292	372	1,44	0,73	104			
BA	6,5	0,8	8	172	194	303	387	1,51	0,76	100			
Bw1	5,5	0,8	7	172	196	301	376	1,49	0,75	102			
Bw2	4,2	0,5	8	176	203	292	376	1,47	0,77	109			
Bw3	2,8	0,5	6	177	203	305	382	1,48	0,76	104			

### 8.1.5. NEOSSOLO QUARTZARÊNICO Órtico típico (UM-RQo)

**Número de campo** – PCG-06

**Data** – 24.09.2007

**Classificação atual** – NEOSSOLO QUARTZARÊNICO Órtico típico, A moderado, álico, fase cerrado tropical subcaducifólio relevo plano.

**Unidade de mapeamento** – RQo

**Localização, município, estado e coordenadas** – A 17 km de Anhanduí, em direção a Campo Grande (BR 163), entra-se à direita para fazenda Santa Hilda e Fazendão percorrendo-se cerca de 18 km até a rodovia CG 324 (piso de terra). Entrando-se à direita percorre-se 11 km, quando, novamente, vira-se à direita e percorre-se mais 17 km. Neste ponto vira-se novamente à esquerda e percorre-se 3 km. Campo Grande, MS. Folha IBGE Lagoa Rica (MI 2588). 0787849 e 7689821 (fuso 21).

**Situação, declive e cobertura vegetal sobre o perfil** – Trincheira aberta em área aplainada com 0-2% de declive, sob cerrado.

**Litologia** – Arenito

**Formação geológica** – Formação Bauru

**Cronologia** – Cretácio

**Material originário** – Sedimentos arenoargilosos

**Pedregosidade** – Não pedregosa.

**Rochosidade** – Não rochosa.

**Relevo local** – Plano e **Relevo regional** – Plano a suave ondulado.

**Erosão** – Laminar ligeira.

**Drenagem** – Excessivamente drenado.

**Vegetação primária** – Cerrado tropical subcaducifólio

**Uso atual** – Pastagem plantada

**Descrito e coletado por** – Paulo E. F. da Motta e Afilton

#### **Descrição Morfológica:**

- A 0 - 20 cm; bruno-avermelhado-escuro (5YR 3/2, úmida) e 5YR 4/3, seca); areia; fraca pequena e média granular e grãos simples; macia, solta, não plástica e não pegajosa; transição plana e gradual.
- AC1 20 - 50 cm; bruno-avermelhado-escuro (5YR 3/3; areia; fraca pequena e média granular e grãos simples; macia, solta, não plástica e não pegajosa; transição plana e gradual.
- AC2 50 - 65 cm; bruno-avermelhado-escuro (2,5YR 3/4); areia franca; fraca pequena e média granular; macia, solta, não plástica e não pegajosa; transição plana e clara.
- CA 65 - 85 cm; vermelho-escuro (2,5YR 3/5); areia franca; fraca pequena e média blocos subangulares e granular; ligeiramente dura, muito friável, não plástica e não pegajosa; transição plana e clara.
- C1 85 - 135 cm, vermelho-escuro (2,5YR 3/5); areia franca; fraca, média blocos subangulares; ligeiramente dura, muito friável, não plástica e não pegajosa; transição plana e difusa.
- C2 135 - 200 cm+, vermelho-escuro (2,5YR 3/5); areia franca; fraca, média e grande blocos subangulares; ligeiramente dura, muito friável, não plástica e não pegajosa.
- Raízes Muitas finas, comuns médias e poucas grossas nos horizontes A e AC1; muitas finas e comuns médias no AC2; comuns finas no CA; comuns finas e poucas grossas no C1 e comuns finas no C2.

**Análises Físicas e Químicas**

Horizonte		Frações da amostra total g/kg			Composição granulométrica da terra fina g/kg				Argila dispersa em água g/kg	Grau de flocculação %	Relação Silte/Argila	Densidade g/cm <sup>3</sup>		Porosidade cm <sup>3</sup> /100cm <sup>3</sup>
Símbolo	Profundidade cm	Calhaus > 20 mm	Cascalho 20-2 mm	Terra fina < 2 mm	Areia grossa 2-0,20 mm	Areia fina 0,20-0,05 mm	Silte 0,05-0,002 mm	Argila < 0,002 mm				Solo	Partículas	
A	0-20	0	0	1000	510	377	53	60	20	67	0,88	1,29	2,69	52
AC1	-50	0	0	1000	512	373	55	60	40	33	0,92			
AC2	-65	0	0	1000	483	403	34	80	40	50	0,42			
CA	-85	0	0	1000	495	385	40	80	40	50	0,50			
C1	-135	0	0	1000	487	363	50	100	80	20	0,50			
C2	-200	0	0	1000	493	363	44	100	0	100	0,44			
Horizonte		pH (1:2,5)			Complexo Sortivo cmol <sub>c</sub> /kg						Valor V %			
	Água	KCl 1N	Ca <sup>2+</sup>	Mg <sup>2+</sup>	K <sup>+</sup>	Na <sup>+</sup>	Valor S	Al <sup>3+</sup>	H <sup>+</sup>	Valor T				
A	4,7	3,9	0,7		0,03	0,01	0,7	0,6	7,0	8,3	8	46	1	
AC1	4,7	4,0	0,2		0,01	0,01	0,2	0,5	1,6	2,3	9	71	1	
AC2	4,7	4,0	0,2		0,01	0,01	0,2	0,4	1,3	1,9	11	67	1	
CA	4,7	4,0	0,1		0,01	0,01	0,1	0,5	1,2	1,8	6	83	1	
C1	4,7	4,0	0,1		0,01	0,01	0,1	0,5	0,5	1,1	9	83	1	
C2	4,8	4,1	0,2		0,01	0,01	0,2	0,2	0,3	0,7	29	50	1	
Horizonte		C (org) g/kg	N g/kg	C/N	Ataque sulfúrico g/kg				Relações Moleculares			Fe <sub>2</sub> O <sub>3</sub> livre g/kg	Equivalente de CaCO <sub>3</sub> g/kg	
					SiO <sub>2</sub>	Al <sub>2</sub> O <sub>3</sub>	Fe <sub>2</sub> O <sub>3</sub>	TiO <sub>2</sub>	SiO <sub>2</sub> /Al <sub>2</sub> O <sub>3</sub> (Ki)	SiO <sub>2</sub> /R <sub>2</sub> O <sub>3</sub> (Kr)	Al <sub>2</sub> O <sub>3</sub> /Fe <sub>2</sub> O <sub>3</sub>			
A		6,2	0,5	12										
AC1		4,8	0,4	12										
AC2		3,0	0,3	10										
CA		3,0	0,3	10										
C1		1,9	0,2	10	33	34	22	5,0	1,65	1,17	2,43			
C2		1,2	0,2	6										

### 8.1.6. PLANOSSOLO HÁPLICO Distrófico gleissólico (UM-SXd)

**Número de campo** – ECG-17

**Data** – 02.05.2007

**Classificação atual** – PLANOSSOLO HÁPLICO Distrófico gleissólico textura média, álico, fase campo hidrófilo de várzea, relevo plano .

**Unidade de mapeamento** – SXd

**Localização, município, estado e coordenadas** – Pela BR 267 (trecho Nova Alvorada-São Paulo), a 74,5 km de Nova Alvorada (trevo com a BR 167) entra-se à esquerda em direção ao rio Anhanduí, distante 6,0 km. Amostra coletada a 700 metros após a ponte e a 200 metros do lado direito da estrada, dentro da várzea. Campo Grande, MS. Folha IBGE Santa Sofia (MI 2660). 219181 e 7613138 (fuso 22)

**Situação, declive e cobertura vegetal sobre o perfil** – Amostras coletadas com trado em área plana de várzea, com 0-1% de declive sob vegetação natural.

**Altitude** – 308

**Litologia** – Sedimentos areno-argilosos.

**Formação geológica** –

**Cronologia** – Quaternário. Holoceno

**Material originário** – Sedimentos areno-argilosos com grande contribuição de matéria orgânica.

**Pedregosidade** – Não pedregosa.

**Rochosidade** – Não rochosa.

**Relevo local** – Plano.

**Relevo regional** – Plano..

**Erosão** – Não aparente.

**Drenagem** – Mal drenado.

**Vegetação primária** – Campo hidrófilo de várzea.

**Uso atual** – Pastagem natural

#### **Descrição Morfológica:**

A - 0–20 cm, preto (2,5Y 2,5/1); franco-siltosa

E - 20-40 cm, cinzento-escuro (2,5Y 4/1); franco-arenosa.

Btg - 40-100 cm, bruno-acinzentado (2,5Y 5/2) e mosqueado bruno-amarelado (10YR 5/6); franco-argilo-arenosa.

### Análises Físicas e Químicas

Horizonte		Frações da amostra total g/kg			Composição granulométrica da terra fina g/kg				Argila dispersa em água g/kg	Grau de flocculação %	Relação Silte/Argila	Densidade g/cm <sup>3</sup>		Porosidade cm <sup>3</sup> /100cm <sup>3</sup>
Símbolo	Profundidade cm	Calhaus > 20 mm	Cascalho 20-2 mm	Terra fina < 2 mm	Areia grossa 2-0,20 mm	Areia fina 0,20-0,05 mm	Silte 0,05-0,002 mm	Argila < 0,002 mm				Solo	Partículas	
A	0-20	0	0	1000	122	17	775	86	86	0	9,01			
E	40-60	0	0	1000	424	408	7	161	80	50	0,04			
Btg	80-100	0	0	1000	232	312	151	305	0	100	0,50			
Horizonte	pH (1:2,5)		Complexo Sortivo cmol <sub>c</sub> /kg								Valor V %	Al Sat %	P mg/kg	
	Água	KCl 1N	Ca <sup>2+</sup>	Mg <sup>2+</sup>	K <sup>+</sup>	Na <sup>+</sup>	Valor S	Al <sup>3+</sup>	H <sup>+</sup>	Valor T				
A	4,5	3,5	1,2	1,0	0,38	0,09	2,7	6,3	52,1	61,1	4	70	1	
E	4,7	3,8	0,2		0,02	0,01	0,2	0,8	1,0	2,0	10	80	1	
Btg	4,2	3,4	0,4		0,09	0,01	0,5	3,3	2,1	5,9	8	87	1	
Horizonte	C (org) g/kg	N g/kg	C/N	Ataque sulfúrico g/kg				Relações Moleculares			Fe <sub>2</sub> O <sub>3</sub> livre g/kg	Equivalente de CaCO <sub>3</sub> g/kg		
				SiO <sub>2</sub>	Al <sub>2</sub> O <sub>3</sub>	Fe <sub>2</sub> O <sub>3</sub>	TiO <sub>2</sub>	SiO <sub>2</sub> /Al <sub>2</sub> O <sub>3</sub> (Ki)	SiO <sub>2</sub> /R <sub>2</sub> O <sub>3</sub> (Kr)	Al <sub>2</sub> O <sub>3</sub> /Fe <sub>2</sub> O <sub>3</sub>				
A	223,1	14,3	16											
E	3,0	0,4	7											
Btg	3,6	0,4	9											