

O SoilGrids, desenvolvido pelo [ISRIC](https://www.isric.org/) World Soil Information, consiste em uma série de mapas raster que retratam as propriedades do solo em diversas profundidades e classes de solo de acordo com o sistema internacional World Reference Base (WRB). Esses mapas oferecem cobertura global e estão disponíveis em resoluções espaciais de 250, 1000 e 5000 metros. Os mapas de 250 metros foram gerados através de interpolação utilizando dados de amostras de solo e covariáveis ambientais. Já os mapas de 1000 e 5000 metros foram obtidos pela agregação (média) dos dados dos produtos de 250 metros [1], [2], [3].

## Opções de Acesso aos Dados do SoilGrids

Os dados de mapas de classes e propriedades do solo do SoilGrids estão acessíveis por meio de várias opções. A plataforma interativa <https://soilgrids.org/> é uma escolha popular para acesso direto aos dados. Além disso, a última versão do SoilGrids pode ser acessada através dos seguintes serviços:

- WMS (Web Map Service): Este serviço oferece acesso para visualização e uma visão geral dos dados. As instruções para utilizar o WMS com softwares GIS comumente utilizados podem ser encontradas em <https://www.isric.org/instruction-wms>.
- WCS (Web Coverage Service): É a melhor opção para obter um subconjunto de um mapa e utilizar o SoilGrids como entrada para outros processos de modelagem. Mais detalhes sobre este serviço e exemplos de como acessar o WCS estão disponíveis em <https://www.isric.org/web-coverage-services-wcs>, incluindo exemplos para Python e R.
- WebDAV: Este serviço permite o download de mapas globais completos em formato VRT ou apenas de telhas (*tiles*) em formato GeoTIFF. Exemplos para acessar os dados através de um navegador de arquivos podem ser encontrados em <https://www.isric.org/explore/soilgrids/soilgrids-access>.

Além dessas opções, as previsões do SoilGrids estão disponíveis no Google Earth Engine como conjuntos de dados contribuídos pela comunidade. Um exemplo de como acessar esses dados através do editor de código do Google Earth Engine

pode ser encontrado em

<https://code.earthengine.google.com/9e11d149a1d28d10619fa5353b895ed2>.

## Como Descarregar Dados usando R e GDAL

Neste tutorial, vamos explorar como descarregar dados do SoilGrids utilizando o serviço WebDAV (Web-based Distributed Authoring and Versioning), a linguagem de programação R e o software GDAL (Geospatial Data Abstraction Library). O WebDAV é uma tecnologia que permite a manipulação remota de arquivos na web e é amplamente utilizado para acesso a dados espaciais.

Os dados do SoilGrids estão disponíveis em dois formatos principais: VRT (Virtual Raster) e COG (Cloud-Optimized GeoTIFF). O formato VRT é um arquivo de mosaico virtual que contém informações sobre a localização, tamanho e formato dos arquivos de imagem subjacentes. Já o formato COG é um tipo especial de GeoTIFF otimizado para acesso e manipulação eficientes em ambientes de computação em nuvem.

Neste tutorial, demonstraremos como acessar e descarregar tanto os dados no formato VRT quanto os dados em formato COG usando R e GDAL. Iremos fornecer instruções passo a passo para configurar a conexão WebDAV, listar os arquivos disponíveis e descarregar os dados desejados. Em seguida, mostraremos como abrir e manusear os arquivos descarregados usando as funcionalidades do GDAL em R.

Ao final deste tutorial, você estará apto a utilizar o WebDAV, R e GDAL para descarregar e manusear dados do SoilGrids, permitindo que você os utilize em suas análises e modelagens de solos de forma eficiente e conveniente.

### **Softwares Requeridos**

Antes de começarmos, é importante garantir que você tenha os seguintes softwares e pacotes instalados em seu sistema:

1. **R e IDE:** Certifique-se de ter o R instalado em seu sistema. Além disso, você precisará de uma IDE (Integrated Development Environment) para escrever e executar código R. Algumas opções populares incluem RStudio, VS Code com extensão para R, ou qualquer outra IDE de sua preferência.
2. **GDAL:** O GDAL (Geospatial Data Abstraction Library) é um conjunto de ferramentas para manipulação de dados geoespaciais e é fundamental para este tutorial. Verifique se o GDAL está instalado em seu sistema. Você pode fazer isso digitando `gdalinfo --version` no terminal. Se o GDAL não estiver instalado, você pode baixá-lo e instalá-lo a partir do site oficial: <https://gdal.org/>.
3. **Pacotes do R:** Certifique-se de ter os seguintes pacotes do R instalados em seu ambiente:
  - a. `geobr`: Este pacote fornece acesso a dados geoespaciais do Brasil. Você pode instalá-lo utilizando o comando `install.packages("geobr")` no console do R.
  - b. `sf`: O pacote `sf` é usado para manipulação de dados espaciais em formato vetorial. Instale-o com o comando `install.packages("sf")`.
  - c. `gdalUtilities`: Este pacote fornece funções úteis para trabalhar com o GDAL em R. Instale-o com o comando `install.packages("gdalUtils")`.

Certifique-se de que todos os requisitos acima estejam atendidos antes de prosseguir com o tutorial.

## Obtenção dos Limites da Área de Interesse

Antes de prosseguirmos com a obtenção dos dados do SoilGrids, é necessário definir a área de interesse para a qual queremos extrair os dados. Neste tutorial, vamos utilizar os limites do estado do Acre como nossa área de estudo.

Para obter os limites do estado do Acre, podemos utilizar o pacote `geobr` no R. Utilize o seguinte comando para carregar os limites do estado do Acre em um objeto chamado `target`:

```
# Descarregar limite da área de interesse
```

```
target <- geobr::read_state(code_state = "AC")
```

Em seguida, precisamos calcular a caixa delimitadora da área de interesse (*bounding box*). Isso pode ser feito usando o pacote `sf` no R da seguinte forma:

```
# Obter caixa delimitadora  
bbox <- sf::st_bbox(target)
```

É importante notar que o pacote `sf` lida com as caixas delimitadoras de maneira diferente do GDAL. Enquanto o GDAL requer que a caixa delimitadora seja especificada como `<ulx> <uly> <lrx> <lry>`, onde `<ulx>` é o valor X do canto superior esquerdo, `<uly>` é o valor Y do canto superior esquerdo, `<lrx>` é o valor X do canto inferior direito e `<lry>` é o valor Y do canto inferior direito, o pacote `sf` do R utiliza a ordem `c(xmin, ymin, xmax, ymax)` para especificar as coordenadas.

Portanto, para garantir que os limites da área de interesse estejam na ordem correta para serem compatíveis com o GDAL, podemos reorganizar as coordenadas usando o seguinte comando:

```
# Reorganizar caixa delimitadora  
bbox <- bbox[c("xmin", "ymax", "xmax", "ymin")]
```

Com isso, os limites da área de interesse estarão prontos para serem utilizados na obtenção dos dados do SoilGrids.

## Configuração do Acesso ao Serviço WebDAV

A próxima etapa consistirá no descarregamento de dados do serviço WebDAV do SoilGrids. Ele fornece acesso direto aos mapas no formato VRT. No entanto, antes precisamos especificar o URL do serviço WebDAV, que é

<https://files.isric.org/soilgrids/latest/data/>, junto com alguns parâmetros adicionais que são usados para configurar um Sistema de Arquivos Virtual (VFS) para acessar recursos via protocolos web, neste caso, via cURL.

```
# Definir parâmetros de acesso ao WebDAV  
max_retry <- 3  
retry_delay <- 1  
list_dir <- "no"
```

```
url <- "https://files.isric.org/soilgrids/latest/data/"
```

Aqui está uma explicação dos parâmetros:

- `max_retry`: Define o número máximo de tentativas se a conexão falhar.
- `retry_delay`: Define o intervalo de tempo entre as tentativas em segundos.
- `list_dir`: Indica que a listagem de diretórios não é necessária.
- `url`: Este é o URL real do recurso a ser acessado.

Com base nesses parâmetros, podemos construir o URL completo para o serviço WebDAV da seguinte forma:

```
# Construir URL de acesso ao WebDAV
webdav_call <- paste0(
  "/vsicurl?max_retry=", max_retry,
  "&retry_delay=", retry_delay,
  "&list_dir=", list_dir,
  "&url=", url
)
```

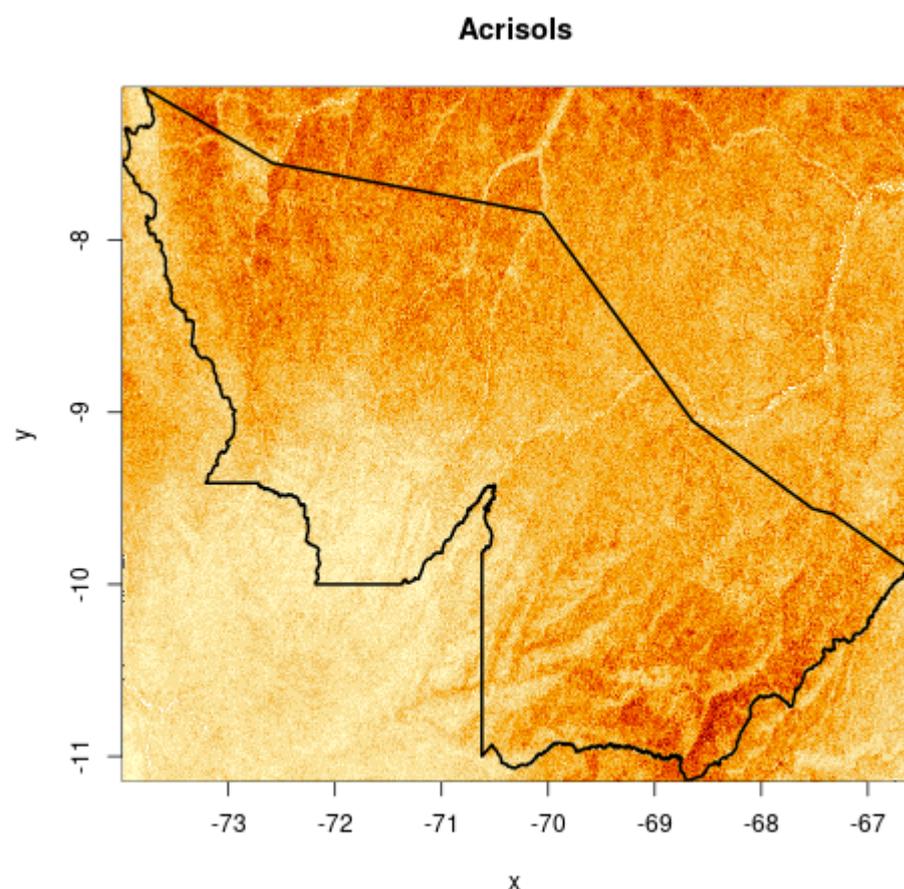
## Descarregando Mapas de Classes de Solo do WRB

O SoilGrids versão 2.0 inclui mapas de classificação de solo. Esses mapas não são o resultado de novas previsões, como acontece com as propriedades do solo. Eles são uma agregação das previsões da versão anterior (2017, Hengl et al.). O SoilGrids versão 2.0 serve apenas para os Grupos de Solos de Referência (RSG) da Base Mundial de Referência para Recursos do Solo (WRB 2006), sem diferenciação por qualificador, em resposta ao feedback recebido dos usuários. Um único mapa é criado para cada RSG individual. Cada mapa mostra a probabilidade estimada de ocorrência do respectivo RSG dentro da célula do mapa. Isso foi obtido somando as probabilidades para cada mapa previsto em 2017 que pode ser associado ao RSG considerado para cada célula do mapa. Mapas de probabilidade de ocorrência para trinta RSGs diferentes estão disponíveis. Antrossolos, Tecnosolos e Retissolos não são considerados, pois não fizeram parte das previsões no SoilGrids versão 2017.

O seguinte comando utiliza a função `gdalUtilities::gdal_translate()` para criar geotiff local na projeção original WGS84 dos mapas WRB.

```
# Descarregar mapas de probabilidade da classe Acrisols (WRB)
map_name <- "wrb/Acrisols.vrt"
output_file <- "data/acrisols.tif"
gdalUtilities::gdal_translate(
  src_dataset = paste0(webdav_call, map_name),
  dst_dataset = output_file,
  projwin = bbox,
  projwin_srs = "EPSG:4326"
)
```

Os nomes usados para os mapas de classe WRB são os mesmos que os nomes das classes WRB, com a primeira letra sempre em maiúsculo. A imagem abaixo mostra os dados descarregados para o estado do Acre.



Nós utilizamos o pacote R `gdalUtilities` para descarregar os dados, mas essa tarefa também poderia ser realizada diretamente usando a função `gdal_translate` do GDAL. O `gdal_translate` é uma ferramenta que permite a conversão entre diferentes formatos de conjuntos de dados raster. Com o `gdal_translate`, é

possível realizar operações como reprojeção, corte, redimensionamento e modificação de formato de arquivos raster. Mais informações sobre a função `gdal_translate` e suas opções estão disponíveis na documentação oficial do GDAL em [https://gdal.org/programs/gdal\\_translate.html](https://gdal.org/programs/gdal_translate.html).

## Criação de um Mapa Binário dos Acrisols

Os mapas de classes de solo do SoilGrids fornecem informações sobre a probabilidade de ocorrência de cada classe. Essa probabilidade é calculada por um modelo que considera simultaneamente 30 classes do WRB. Portanto, em cada pixel, cada classe tem uma probabilidade de ocorrência entre zero e um, e a soma das probabilidades de todas elas é igual a um. Para gerar um mapa combinado de todas as classes de solo, identificamos, em cada pixel, qual é a classe com a maior probabilidade de ocorrência.

Neste tutorial, estamos interessados apenas na classe Acrisols. Para isso, usamos a probabilidade de ocorrência dos Acrisols para gerar um mapa binário que indica onde os Acrisols ocorrem e onde não ocorrem. Como critério de corte, definimos o valor de probabilidade de 40%. No entanto, este valor pode variar de área para área, e a escolha do valor de corte depende do conhecimento da área e da avaliação dos resultados obtidos utilizando diferentes pontos de corte.

Para gerar o mapa binário, utilizaremos o `gdal_calc.py`. O código fornecido como exemplo é o seguinte:

```
# Criar mapa binário usando gdal_calc.py
# O arquivo de entrada é o arquivo de saída do passo anterior.
input_file <- output_file
prob <- 40
calc <- paste0("where(A>", prob, ", 1, 0)")
output_file <- paste0("data/acrisols_bin", prob, ".tif")
cmd <- paste0(
  "gdal_calc.py -A ", input_file, " --outfile=", output_file, " --calc=\"",
  calc, "\"\"
)
system(cmd)
```

Observe que criamos um novo arquivo de saída com o valor de corte no nome para facilitar a identificação.

O `gdal_calc.py` é uma ferramenta do GDAL que permite a execução de operações aritméticas em conjuntos de dados raster. Com essa ferramenta, é possível realizar cálculos sobre os valores dos pixels de um raster e criar novos conjuntos de dados raster como resultado. Para mais detalhes sobre o uso do `gdal_calc.py`, incluindo uma lista completa de operadores e opções disponíveis, você pode consultar a documentação oficial do GDAL em [https://gdal.org/programs/gdal\\_calc.html](https://gdal.org/programs/gdal_calc.html).

## Filtragem do Mapa Binário

O mapa binário resultante pode conter muitos pixels isolados, uma ocorrência comum em dados raster. Para resolver esse problema, aplicaremos um filtro de peneira (*sieve filter*) ao mapa. Esse filtro exige estabelecer um limiar para remover os pixels isolados. O limiar é um valor definido para determinar se um pixel é considerado isolado ou não. Pixels com menos pixels vizinhos conectados a eles do que o valor do limiar são considerados isolados e serão removidos pelo filtro de peneira.

Podemos determinar o limiar usando dois critérios da cartografia: resolução espacial e escala cartográfica. A resolução espacial refere-se ao tamanho e ao espaçamento dos pixels em um mapa. Se aplicarmos alguns conceitos simples de microscopia e processamento de sinais, podemos entender que a menor capacidade de detalhamento em um mapa raster é aproximadamente o tamanho de dois pixels para objetos lineares e 2 x 2 pixels para objetos retangulares. Chamamos isso de resolução espacial nominal [4], que, nos mapas do SoilGrids, seria de 500 x 500 metros. Assim, podemos considerar que quatro pixels é um valor adequado para remover pixels isolados, especialmente se usarmos os dados filtrados em uma base de dados também no formato raster e com a mesma resolução espacial.

Uma outra aplicação possível dos dados filtrados é a sua integração a uma base de dados espaciais que inclui mapas impressos provenientes de relatórios de

levantamentos que foram digitalizados. Ou mesmo a sua impressão. Para isso, precisamos transformar resolução espacial para escala cartográfica. Podemos assumir que a menor área discernível num mapa digital seja de 1 x 1 mm e, assim, a escala cartográfica seria dada por 1 m/(comprimento lateral de 1000 pixels) [4]. Com isso, um tamanho de pixel e espaçamento de 250x250 m como o do SoilGrids equivaleria a uma escala cartográfica de 1:250.000. Assim a resolução espacial do SoilGrids corresponderia à escala cartográfica dos levantamentos de reconhecimento de baixa ou média intensidade [5].

Nos mapas impressos, em vez da resolução nominal, utilizamos o conceito de área mínima mapeável ou área mínima legível. Trata-se da menor área discernível num mapa, que pode variar na literatura entre valores de 0,5 x 0,5 cm e 0,6 x 0,6 cm [6], [7]. Para de reconhecimento de baixa ou média intensidade, a área mínima mapeável seria de 2,25 km<sup>2</sup>:

```
# Cálculo da área mínima mapeável ou legível  
mla <- ((0.6 * 250000) / 100000)^2
```

Essa área mínima mapeável seria aproximadamente equivalente a área formada por aproximadamente 36 pixels de um mapa do SoilGrids:

```
# Compute threshold for sieve filter  
threshold <- round(((sqrt(mla) * 1000) / 250)^2)
```

Note que a estimativa da resolução nominal de um mapa baseada na definição da área mínima legível ou mapeável são muito mais conservadoras do que quando usamos as noções em microscopia e o conceito de processamento de sinais.

Enquanto no primeiro caso o limiar para o filtro era de quatro pixels, aqui são 36.

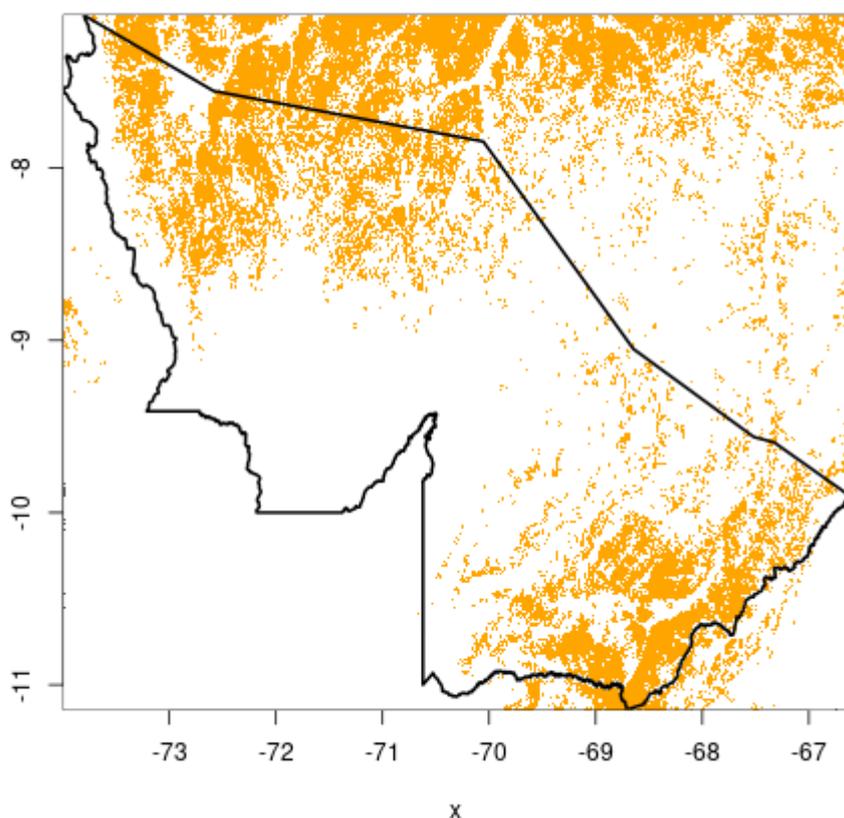
Neste tutorial, estaremos usando o valor da área mínima legível/mapeável como limiar para remover pixels isolados, simulando como os dados do SoilGrids podem ser ajustados para se adequar a uma base de dados poligonais derivada de mapas impressos digitalizados. É importante observar que um limiar mais alto também ajuda a lidar com algumas incertezas no mapa raster, eliminando polígonos de menor extensão.

O código utilizado para aplicar o filtro de peneira usando a função `gdal_sieve.py` do GDAL é o seguinte:

```
# Filtrar mapa binário
input_file <- output_file
threshold <- ceiling(((sqrt(2.5) * 1000) / 250)^2)
connectedness <- 8
output_file <- paste0(
  gsub(".tif", "", input_file), "_sieve", threshold, ".tif"
)
cmd <- paste0(
  "gdal_sieve.py ", input_file, " -st ", threshold, " -", connectedness, "
",
  output_file
)
system(cmd)
```

Observe que o arquivo de saída é nomeado com o sufixo `_sieve` seguido do valor do limiar. A imagem abaixo mostra os dados processados para o estado do Acre.

**Acrisols**



O `gdal_sieve.py` é uma ferramenta do GDAL projetada para filtrar pixels isolados em dados raster. Essa ferramenta remove esses pixels com base em critérios de

conectividade definidos pelo usuário, sendo útil para eliminar artefatos indesejados em mapas. Para mais informações sobre o `gdal_sieve.py` e como usá-lo, consulte a documentação oficial do GDAL em [https://gdal.org/programs/gdal\\_sieve.html](https://gdal.org/programs/gdal_sieve.html).

## Conversão do Raster Binário para Polígonos

A última etapa de nosso tutorial consiste na conversão do dado raster binário para polígonos. O seguinte comando do GDAL será utilizado para converter o mapa binário em um shapefile, utilizando a função `gdal_polygonize.py`.

```
# Transform the raster data to polygons
# The input file is the output file from the previous step.
input_file <- output_file
output_file <- gsub(".tif", ".shp", input_file)
cmd <- paste0("gdal_polygonize.py ", input_file, " ", output_file)
system(cmd)
```

O `gdal_polygonize.py` é uma ferramenta do GDAL que converte dados raster em polígonos vetoriais. Essa ferramenta é útil para transformar informações raster, como mapas binários, em formatos vetoriais, como shapefiles, facilitando análises espaciais mais avançadas. A documentação oficial do GDAL oferece informações detalhadas sobre como usar o `gdal_polygonize.py` e suas opções de uso. Para mais detalhes, você pode acessar a documentação em [https://gdal.org/programs/gdal\\_polygonize.html](https://gdal.org/programs/gdal_polygonize.html).

## Considerações Finais

O SoilGrids, desenvolvido pelo ISRIC World Soil Information, oferece mapas detalhados das propriedades do solo em todo o mundo. Esses mapas, disponíveis em diferentes resoluções espaciais, são acessíveis através de diversas plataformas, como SoilGrids.org, e serviços como WMS e WCS.

Neste tutorial, exploramos como descarregar e manusear os dados do SoilGrids usando o WebDAV com R e GDAL. Aprendemos a configurar a conexão WebDAV, descarregar os dados desejados e realizar operações como filtragem e conversão de formatos.

É importante lembrar que os dados do SoilGrids podem conter erros desconhecidos e incontroláveis, podendo não refletir completamente a realidade de sua área de estudo. Portanto, ao utilizá-los, é recomendado o uso cauteloso, inclusive ao aplicá-los como covariáveis em modelos preditivos. Sempre leve em consideração a relação entre o solo e a paisagem para uma interpretação adequada dos resultados.

O código R para executar o tutorial está disponível na íntegra em <https://github.com/Laboratorio-de-Pedometria/rtutorial/>.

## Literatura Consultada

- [1] T. Hengl *et al.*, “SoilGrids1km—global soil information based on automated mapping”, *PLoS ONE*, vol. 9, nº 8, p. e105992, 2014, doi: 10.1371/journal.pone.0105992
- [2] T. Hengl *et al.*, “SoilGrids250m: Global gridded soil information based on machine learning”, *PLOS One*, vol. 12, nº 2, p. e0169748, fev. 2017, doi: 10.1371/journal.pone.0169748
- [3] L. Poggio *et al.*, “SoilGrids 2.0: producing soil information for the globe with quantified spatial uncertainty”, *SOIL*, vol. 7, nº 1, p. 217–240, jun. 2021, doi: 10.5194/soil-7-217-2021
- [4] A. B. McBratney, M. L. Mendonça-Santos, e B. Minasny, “On digital soil mapping”, *Geoderma*, vol. 117, p. 3–52, 2003, doi: 10.1016/S0016-7061(03)00223-4
- [5] IBGE, *Manual Técnico de Pedologia*, 3º ed. em Manuais Técnicos em Geociências, no. 4. Rio de Janeiro, Rio de Janeiro, Brasil: Instituto Brasileiro de Geografia e Estatística, 2015, p. 430. Disponível em: <https://biblioteca.ibge.gov.br/visualizacao/livros/liv95017.pdf>
- [6] Soil Survey Staff, *Soil Survey Manual*. Washington: USDA-NRCS, 1993. Disponível em:

[http://www.nrcs.usda.gov/wps/portal/nrcs/detail/soils/planners/?cid=nrcs142p2\\_054262](http://www.nrcs.usda.gov/wps/portal/nrcs/detail/soils/planners/?cid=nrcs142p2_054262)

[7] A. P. A. Vink, *Land use in advancing agriculture*. em Advanced series in agricultural sciences, no. v. 1. Berlin ; New York: Springer-Verlag, 1975.