Latin American Data in Science, v.5, n.1
datainscience.com.br
ISSN: 2763-9290

Soil Aptitude for Livestock (SAFL): A Dataset for Brazilian Municipalities

Recebido:25/01/25|Aceito:02/07/25 | Publicado: 21/07/25
https://doi.org/10.53805/lads.v5i1.72

Felipe M. da Silva [1]

ABSTRACT

This database presents a municipality-level Soil Aptitude for Livestock (SAFL) index, designed to measure each Brazilian locality's suitability for cattle pastures. The index is derived from a combination of georeferenced data sources that capture critical environmental factors, particularly soil type. By processing and standardizing this information in Python, the SAFL index consolidates soil classification details—often overlooked in land-use research—into a single metric suitable for diverse applications. Researchers and practitioners can incorporate the SAFL index into their models to account for the influence of soil-related variables on land-use choices. For instance, it may serve as a control or explanatory factor in econometric panel data, improving the study of phenomena such as agricultural expansion, land-use change, and environmental impacts. Beyond econometric modeling, the dataset offers value for policy planning, risk assessment, and investment decisions by highlighting areas with greater aptitude for livestock activities. The SAFL index and its underlying methodology are intended to be transparent, replicable, and adaptable. Users can modify the index to suit regional or thematic requirements, ensuring broad applicability across disciplines including environmental science, economics, and resource management. By making soil suitability data more accessible and standardized, the SAFL database fills a critical gap in available land-use resources and holds significant potential for advancing research and informing decision-making.

Keywords: Soil Aptitude for Livestock; Georeferenced data; Panel Data; Brazilian Municipalities; Open-source index.

PRIOR PUBLICATIONS

FARIA, W.R. et al. The relationship between rural credit policy and deforestation: evidence from Brazil. Environ Econ Policy Stud (2024). https://doi.org/10.1007/s10018-024-00421-4.

[1] Federal University of Brasília (UnB), Postgraduate Program in Economics (PPGECO), Brasília, Brazil. fms.morelli@gmail.com

DATA IMPORTANCE

- Long-Term Validity: The dataset captures soil traits that remain stable over time, allowing researchers to track phenomena such as deforestation, agricultural expansion, and investment dynamics without frequent updates.
- Standardized Geospatial Index: It consolidates diverse geospatial soil data into a unified metric, facilitating incorporation into econometric or spatial models that require consistent, municipality-level indicators.
- Time-Invariant Variable for Panel Data: Soil suitability serves as a fixed-effect variable, enabling deeper insights into land-use patterns when studying environmental and economic outcomes in panel data research.
- Open-Source Adaptability: Researchers can modify soil scores within the provided CSV file, rerun the index generation, and tailor analyses to specific hypotheses or regional contexts, fostering collaborative refinement.
- Applicability for Policy and Investment: Public agencies and private actors can use this soil-based index to inform land-use policies, plan agricultural ventures, evaluate sustainability initiatives, and guide carbon market investments.

MATERIALS AND METHODS

To obtain a municipality-level Soil Aptitude for Livestock (SAFL) index, we employed a computational approach involving Geographic Information System (GIS) data and Python-based scripts. No field or laboratory experiments were conducted. Below are the main procedures adopted.

- Geospatial Data Acquisition: We used the shapefile of Brazilian municipalities provided by the Brazilian Institute of Geography and Statistics (IBGE, 2019). We obtained a second shapefile from Embrapa (2020), which categorizes soils nationwide according to the World Reference Base (WRB) classification system. Both datasets were stored locally in an input directory for subsequent processing.

- Software and Environment: All procedures were scripted in Python (version 3.8 or higher), using packages such as geopandas, pandas, and shapely to handle geospatial data operations. Three main scripts (step1.py, step2.py, step3.py) automate data processing and index generation. These are provided in supplementary material if needed.

- Geoprocessing and Index Calculation: We confirmed that both shapefiles used the same coordinate reference system (CRS); when discrepancies arose, the soil shapefile was reprojected to match the municipality shapefile. We then used geopandas.overlay() to intersect each municipality's boundary with the soil map, identifying which soil types and how much area (in square kilometers) of each soil fall within each municipality.

Next, we assigned a numeric weight (e.g., 0, 0.5, or 1) to each WRB soil group based on its suitability for livestock production; these values, stored in a CSV file, draw on guidelines from the International Soil Reference and Information Center (ISRIC, 2020) and Embrapa. If Ai represents the vector of soil-type proportions in municipality i and Pi the corresponding suitability weights, then the SAFL index Ni is calculated by Ai ’Pi . This step is automated in the second Python script (step2.py), which outputs a shapefile showing the final SAFL scores for all municipalities.

- Output Generation: The third Python script (step3.py) produces a georeferenced map highlighting municipalities according to their

calculated SAFL scores. Final datasets (shapefiles and CSV files with SAFL indices) are saved in the output directory. All scripts and instructions for reproducing the process are shared as supplementary material.

All steps described above are essential for creating the SAFL dataset and do not involve direct manipulation of field-collected or questionnaire-based data. The methodological transparency and open-source coding enable further adaptations and expansions of this index in future research.

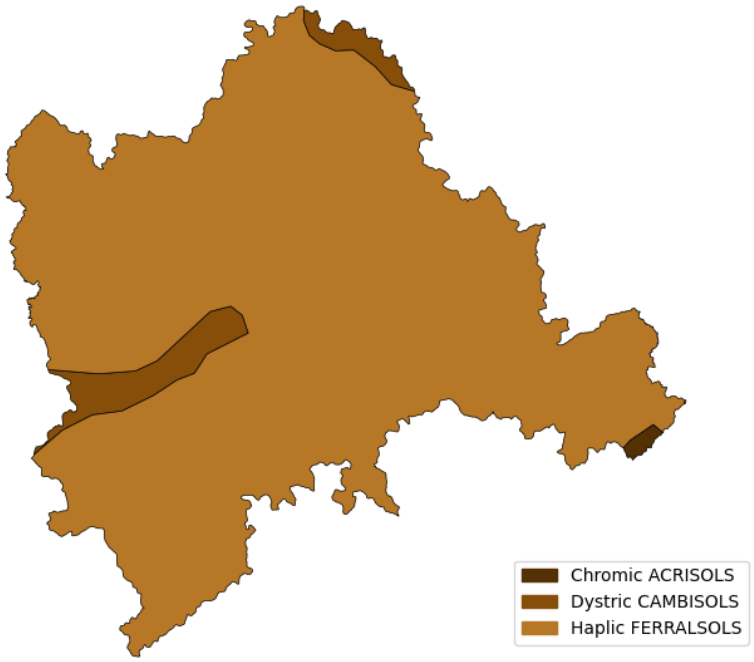
DATA DESCRIPTION

This section presents a brief overview of the Soil Aptitude for Livestock (SAFL) data, followed by a detailed explanation of each file provided. Two main output files—the shapefile and a CSV—are shared to accommodate both GIS experts and researchers less experienced with spatial data. Figures illustrate representative examples and the overall distribution of the SAFL index in Brazil.

In Figure 1, for instance, we show the soil types in the municipality of Juiz de Fora (state of Minas Gerais), highlighting how different soil proportions are used to calculate the index. The municipality's soil composition (e.g., Haplic Ferralsols, Dystric Cambisols, Chromic Acrisols) is multiplied by corresponding suitability weights to yield an overall SAFL score.

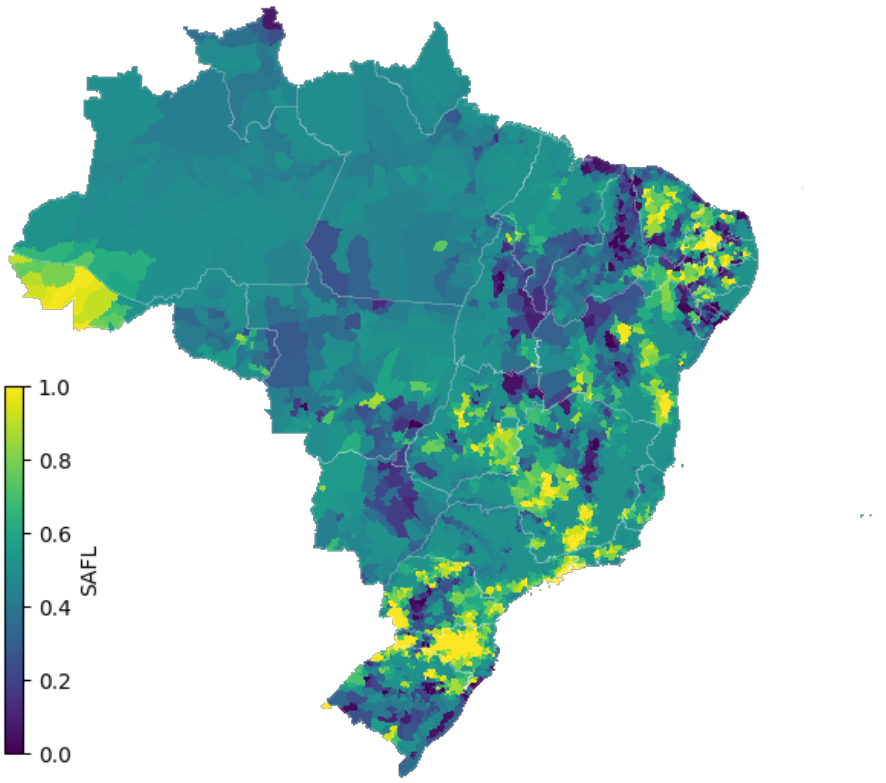
Figure 2 provides a broader view of SAFL distribution across Brazilian municipalities, illustrating how this dataset can be visualized on a national scale.

Figure 1.- Soil Types in Juiz de Fora (MG).



Source: Embrapa soil classification maps. Edited by the authors.

Figure 2: SAFL Distribution Across Brazil



Source: Computed by the authors.

Dataset

1. Shapefile of Municipalities with SAFL (result.shp)

- File Location: output/SAFL/result.shp
- Description: This shapefile stores the geometry of each municipality in Brazil (as polygons), along with relevant attributes.
- Columns:
    - CD_MUN: Numerical code assigned by IBGE (2019 version) to each municipality.
    - NM_MUN: Name of the municipality.
    - SIGLA_UF: Two-letter abbreviation for each Brazilian state.
    - SAFL: Calculated Soil Aptitude for Livestock index, expressed as a continuous value between 0 and 1. A value closer to 1 indicates higher suitability.

2. SAFL Index in CSV Format (result.csv)

- File Location: output/SAFL/result.csv
- Description: Provided as a simpler, non-spatial alternative for those not working with GIS software. Contains two columns matching key data from the shapefile.
- Columns:
    - CD_MUN: IBGE (2019) municipality code.
    - SAFL: Soil Aptitude for Livestock index, matching the score in the shapefile.

3. Supplementary Table of Soil Groups and Suitability Weights

- File Location: input/ISRIC/soil_group.csv

- Description: A reference table (based on the World Reference Base and ISRIC guidelines) that assigns each soil group a suitability weight (e.g., 0, 0.5, or 1). Users can modify these weights to create alternative versions of the SAFL index. This table is essential for interpreting how each soil group contributes to the final index.

All files and figures are provided so researchers can replicate the process, adapt the soil suitability weights, or integrate the SAFL index into further analyses and models.

SUPPLEMENTARY MATERIALS

Repository name: Mendeley Data
DOI of the dataset (when available): http://doi.org/10.17632/rnrd8f7bxb.1
Link to access the data: https://data.mendeley.com/drafts/rnrd8f7bxb

ACKNOWLEDGEMENTS

We thank Coordination for the Improvement of Higher Education Personnel (CAPES) and the University of Brasília (UnB) for their support, which made the development of this data paper possible.

REFERENCES

EMBRAPA - Empresa Brasileira de Pesquisa Agropecuária. Dados geoespaciais em formato de vetor ou raster publicados pela Embrapa e parceiros. 2020. URL: http://geoinfo.cnps.embrapa.br/layers/geonode%5C%3Asolos_br5m_2011_lat_long_wgs84 (visited on 05/06/2024).

EMBRAPA - Empresa Brasileira de Pesquisa Agropecuária. Mapa de aptidão agrícola das terras do Triângulo Mineiro. 2019. URL: http://geoinfo.cnps.embrapa.br/layers/geonode%5C%3Aaptidao_agricola_triangulo_mineiro_lat_long_sirgas2000 (visited on 10/30/2024).

IBGE - Instituto Brasileiro de Geografia e Estatística. Malha Municipal. 2019. URL: https://www.ibge.gov.br/geociencias/organizacao-do-territorio/15774-malhas.html?=&t=o-que-e (visited on 05/06/2024).

ISRIC - International Soil Reference and Information Centre. World Soil Information. 2020. URL: https://www.isric.org/ (visited on 11/30/2024).