

Spatial Analysis Pipeline Documentation

Produced by: Antônio Lucas Barreira Rodrigues

Overview

This R script implements a comprehensive spatial analysis pipeline for the PAN-DELTA project, focusing on species conservation prioritization. The pipeline processes multiple spatial datasets including species occurrences, land use/land cover, conservation units, and various environmental pressures.

Required R Packages

Core Spatial Processing

- `sf` (v1.0+): Simple Features for R - handles vector spatial data (shapefiles, geopackages)
- `terra` (v1.7+): Successor to raster package - processes raster spatial data efficiently
- `dplyr` (v1.1+): Data manipulation and transformation
- `tidyverse` (v2.0+): Collection of data science packages including ggplot2, tidyr, readr, etc.

Additional Utilities

- `knitr`: Dynamic report generation (though not actively used in the script)
- `scales`: Scale functions for data visualization (though not actively used in the script)

Directory Structure Requirements

The script expects a specific directory structure. All paths are relative to a base directory (e.g., `D:/PAN - DELTA`):

```
D:/PAN - DELTA/
├── ANALISES/
│   ├── priorizacao/..... # Working directory and outputs
│   └── dados_analise_r/
│       ├── territorio/..... # PAN territory boundary
│       ├── spp/..... # Species data
│       ├── otto/..... # Ottobacia (watershed) data
│       ├── uc/..... # Conservation units
│       ├── mineracao/..... # Mining areas
│       ├── energia/..... # Energy infrastructure
│       └── app/..... # Permanent Preservation Areas
```

Input Files Required

1. Territory Data (`territorio/`)

- **File:** One shapefile (`.shp`) defining the PAN territory boundary
- **Required fields:** None specific
- **Format:** Any valid shapefile with polygon geometry

2. Species Data (`spp/`)

- **Base shapefile:** `SPP_LISTA_2022_SIRGAS2000_CNCFLORA.shp`
 - Required field: `specie_fb` (species name)
- **CSV files:** Species lists for filtering
 - Format: Single column with species names
 - Creates filtered shapefiles: `spp_alvo.shp` and `spp_beneficiada.shp`

3. Ottobacia Data (`otto/`)

- **File:** One shapefile with watershed boundaries
- **Required field:** `wts_cd_pfa` (watershed code)

4. Conservation Units (`uc/`)

- **File:** One shapefile with conservation units
- **Required field:** `grupo` (must contain "Proteção Integral" values)

5. Mining Areas (`mineracao/`)

- **File:** One shapefile with mining concessions
- **Required field:** `NUMERO` (concession number)

6. Energy Infrastructure (`energia/`)

- **File:** One shapefile with energy infrastructure
- **Required fields:** None specific (presence/absence only)

7. Permanent Preservation Areas (`app/`)

- **Files:** Shapefiles matching pattern `*_APP_USO*.shp`
- **Required:** Fields indicating land use (searched dynamically)
- **Purpose:** Identifies anthropized APP areas for restoration

Functions Documentation

Setup and Initialization Functions

`setup_environment(base_dir)`

- **Purpose:** Creates and validates directory structure

- **Input:** Base directory path
- **Output:** List of directory paths
- **Action:** Stops execution if any directory is missing

`load_pan_territory(pan_dir)`

- **Purpose:** Loads and standardizes PAN territory boundary
- **Input:** Directory containing territory shapefile
- **Output:** sf object in SIRGAS 2000 (EPSG:4674)

`create_standard_raster(pan_sf, resolution)`

- **Purpose:** Creates template raster for standardization
- **Input:** PAN territory sf object, resolution (default: 0.005°)
- **Output:** Terra raster object covering PAN extent

MapBiomass Download Functions

`download_mapbiomas(mapbiomas_url, pan_sf)`

- **Purpose:** Downloads and clips MapBiomass land cover data
- **Input:** MapBiomass URL, PAN territory
- **Output:** Clipped raster and temporary file path
- **MapBiomass layer:** Annual land cover (2023)

`download_mapbiomas_fogo(mapbiomas_fogo_url, pan_sf)`

- **Purpose:** Downloads fire frequency data
- **Input:** Fire data URL, PAN territory
- **Output:** Fire frequency raster
- **MapBiomass layer:** Accumulated burned area (1985-2023)

`download_mapbiomas_pasto(mapbiomas_pasto_url, pan_sf)`

- **Purpose:** Downloads pasture quality data
- **Input:** Pasture URL, PAN territory
- **Output:** Pasture quality raster
- **MapBiomass layer:** Pasture vigor (2023)

`download_mapbiomas_veg_sec(mapbiomas_veg_sec_url, pan_sf)`

- **Purpose:** Downloads secondary vegetation age

- **Input:** Vegetation age URL, PAN territory
- **Output:** Vegetation age raster
- **MapBiomias layer:** Secondary vegetation age (2023)

Species Processing Functions

process_species_from_csv(spp_dir, base_shp_name)

- **Purpose:** Filters species shapefile based on CSV lists
- **Process:**
 1. Reads base shapefile with all species
 2. For each CSV file, filters species in the list
 3. Creates new shapefiles for filtered species
- **Output:** Individual shapefiles for each species list

process_species(spp_file, raster_mapbiomas, output_dir, r_padrao, pan_sf)

- **Purpose:** Processes target species with habitat analysis
- **Process:**
 1. Creates convex hulls for each species
 2. Extracts forest/savanna areas (classes 3,4) from MapBiomias
 3. Generates individual rasters per species
- **Output:**
 - `todas_spp_alvo.gpkg`: All species hulls
 - `spp_alvo/[species_name].tif`: Individual species habitat rasters

process_benefited_species(benefited_spp_file, raster_mapbiomas, output_dir, r_padrao, pan_sf)

- **Purpose:** Similar to above but for DD/NT species
- **Output:**
 - `todas_spp_beneficiadas.gpkg`
 - `spp_beneficiadas/[species_name]_beneficiada.tif`

process_species_by_otto(spp_file, otto_file, output_dir, r_padrao, pan_sf, tipo_spp)

- **Purpose:** Generates species distribution by watershed
- **Process:** Identifies watersheds containing species occurrences
- **Output:**
 - `spp_[tipo]_ottobacias/[species]_otto.tif`: Rasters by species
 - `ottobacias_[tipo].gpkg`: Summary shapefile

- **Note:** Currently commented out in main function

Environmental Data Processing Functions

`process_otto(otto_dir, r_padrao, output_dir, pan_sf)`

- **Output:** `otto.tif` - Watershed boundaries rasterized

`process_uc(uc_dir, r_padrao, output_dir, pan_sf)`

- **Output:** `ucs_protecao_integral.tif` - Integral protection areas (0=inside, 1=outside)

`process_mineracao(mineracao_dir, r_padrao, output_dir, pan_sf)`

- **Output:** `mineracao.tif` - Mining concession areas

`process_energia(energia_dir, r_padrao, output_dir, pan_sf)`

- **Output:** `energia.tif` - Energy infrastructure presence

`process_app(app_dir, r_padrao, output_dir, pan_sf)`

- **Output:**
 - `app_areas_antropizadas_unificado.shp` - Unified anthropized APPs
 - `app_antropizada.tif` - Rasterized anthropized APPs

MapBiomass Processing Functions

`process_mapbiomas_agricultura(raster_mapbiomas, r_padrao, output_dir, pan_sf)`

- **Classes extracted:** 15, 21, 24, 25, 39, 41, 48 (agricultural uses)
- **Output:** `agro.tif`

`process_mapbiomas_floresta(raster_mapbiomas, r_padrao, output_dir, pan_sf)`

- **Class extracted:** 3 (forest formations)
- **Output:** `floresta.tif`

`process_mapbiomas_savana(raster_mapbiomas, r_padrao, output_dir, pan_sf)`

- **Class extracted:** 4 (savanna formations)
- **Output:** `savana.tif`

`process_fogo(raster_mapbiomas_fogo, temp_file_fogo, r_padrao, output_dir, pan_sf)`

- **Output:** `fogo.tif` - Fire frequency

`process_pasto(raster_mapbiomas_pasto, temp_file_pasto, r_padrao, output_dir, pan_sf)`

- **Classes extracted:** 2, 3 (pasture quality levels)
- **Output:** `pasto_q.tif`

`process_veg_sec(raster_mapbiomas_veg_sec, temp_file_veg_sec, r_padrao, output_dir, pan_sf)`

- **Output:** `veg_secundaria.tif` - Secondary vegetation age

Utility Functions

`create_grid_raster(r_padrao, output_dir, write_vector)`

- **Purpose:** Creates analysis grid
- **Output:**
 - `grid_quadriculas.tif` - Raster grid with unique cell IDs
 - `grid_quadriculas.gpkg` - Vector grid (optional)

Main Pipeline Function

`main(base_dir)`

- **Purpose:** Orchestrates entire processing pipeline
- **Process Flow:**
 1. Setup environment and load territory
 2. Create standard raster template
 3. Download all MapBiomas data
 4. Process species data
 5. Process environmental layers
 6. Generate analysis grids
 7. Clean temporary files

Output Files Summary

All outputs are saved to the `priorizacao/` directory:

Raster Files (.tif)

- Species habitat: `spp_alvo/*.tif`, `spp_beneficiadas/*.tif`
- Environmental: `otto.tif`, `ucs_protecao_integral.tif`, `mineracao.tif`, `energia.tif`
- Land cover: `agro.tif`, `floresta.tif`, `savana.tif`, `fogo.tif`, `pasto_q.tif`, `veg_secundaria.tif`
- Restoration: `app_antropizada.tif`
- Analysis: `grid_quadriculas.tif`

Vector Files

- `todas_spp_alvo.gpkg` - Target species convex hulls
- `todas_spp_beneficiadas.gpkg` - Benefited species convex hulls
- `grid_quadriculas.gpkg` - Analysis grid
- `app_areas_antropizadas_unificado.shp` - Anthropized APPs

Execution Instructions

1. **Prepare directory structure** as shown above
2. **Place all input files** in their respective directories
3. **Update the base directory** path in the script:

```
r  
  
base_dir <- "D:/PAN - DELTA" # Change to your path
```

4. **Run the script:**

```
r  
  
source("script_name.R")
```

Technical Notes

- **Coordinate System:** All data standardized to SIRGAS 2000 (EPSG:4674)
- **Raster Resolution:** 0.005° (approximately 550m at the equator)
- **Memory Management:** Script includes garbage collection after processing large datasets
- **Compression:** Output rasters use LZW compression with predictor
- **GDAL Cache:** Set to 2048MB for MapBiomass processing

Troubleshooting

1. **Memory errors:** Increase GDAL cache or process smaller areas
2. **Missing files:** Check directory structure and file naming patterns
3. **CRS issues:** Ensure all input files can be transformed to EPSG:4674
4. **MapBiomass downloads:** Verify URLs are current and accessible

MapBiomass Classes Reference

- **Forest:** Class 3
- **Savanna:** Class 4
- **Agriculture:** Classes 15, 21, 24, 25, 39, 41, 48

- **Pasture Quality:** Classes 2 (moderate), 3 (low)

This pipeline provides a comprehensive framework for spatial conservation prioritization, integrating species distributions with environmental pressures and opportunities.