Land Cover Change Analysis Pipeline - From PhD Studies on Land Use Change

Appendix C1: Initial Software for Land Use Analysis

UBCO - Interdisciplinary Graduate Studies - Sustainability

PhD Research Proposal - August 2025

Project Title: Mapping the Relationship Between Forest Coverage and Policy Processes in Brazil : A Comparative Study of Indigenous Lands (1980-Present)

Researcher: Leandro Meneguelli Biondo

PhD Committee

- **Supervisor**: Dr. Jon Corbett
- **Committee Member**: Dr. Jonathan Cinnamon
- **Committee Member**: Dr. Robert Friberg
- **Committee Member**: Dr. Tim Paulson

2025 - Leandro Meneguelli Biondo - PhD Student IGS/UBCO - Kelowna/BC-Canada

INMA/MCTI (Santa Teresa/ES-Brazil) - SFB/MMA (Brasilia/DF-Brazil) - github.com/leandromet/

A comprehensive Python-based pipeline for analyzing land cover changes over time using satellite imagery data. This tool processes MapBiomas time series data to extract land cover transitions, create visualizations, and generate statistical summaries for specific geographic regions.

Overview

This pipeline analyzes 40 years of land cover data (1985-2024) from MapBiomas satellite imagery to:

- Extract land cover data for specific polygons/regions
- Calculate pixel-level change frequencies over time
- Generate transition matrices showing land cover class changes
- Create visualizations including maps, Sankey diagrams, and statistical charts
- Produce detailed statistics on land cover persistence and change

Features

Core Functionality

- **Polygon-based Analysis**: Process any GeoJSON polygon to extract land cover data
- **Time Series Analysis**: Track changes across 40 years (1985-2024)
- **Individual Feature Processing**: Handle multiple polygons individually

- **Modular Pipeline**: Skip/include specific processing steps as needed
- **Parallel Processing**: Utilize multi-threading for faster results

Outputs Generated

- **Change Maps**: Visualize frequency of land cover transitions per pixel
- **Land Cover Maps**: Current and historical land cover visualizations
- **Sankey Diagrams**: Interactive flow diagrams showing class transitions
- **Transition Matrices**: Detailed pixel counts for all land cover changes
- **Statistical Reports**: CSV files with comprehensive change statistics
- **GIS-Compatible Files**: PNG files with world files (.pngw) for georeferencing
- **Charts and Graphs**: Visual representations of change statistics

Installation

Requirements

- Python 3.8+
- GDAL/OGR libraries
- Sufficient RAM (8GB+ recommended)
- Storage space for output files

Dependencies

```
pip install -r requirements.txt
```

Key packages:

- `rasterio` - Raster data processing
- `geopandas` - Geospatial vector operations
- `zarr` - Efficient array storage
- `matplotlib` - Visualization
- `plotly` - Interactive charts
- `numpy` - Numerical operations
- `psutil` - Memory monitoring

Usage

Command Line Interface

Basic Usage

```
python main.py --geojson polygons.geojson --vrt mapbiomas_data.vrt --
output-dir ./results
```

Complete Parameter List

```
python main.py \
  --geojson <path_to_geojson_file> \
  --vrt <path_to_vrt_file> \
  --output-dir <output_directory> \
  [--skip-extraction] \
  [--skip-visualization] \
  [--skip-sankey] \
  [--skip-transition-viz] \
  [--verbose]
```

Parameters

Parameter	Required	Description
`--geojson`	Yes	Path to GeoJSON file containing analysis polygons
`--vrt`	Yes	Path to VRT file with MapBiomas time series data
`--output-dir`	Yes	Directory where results will be saved
`--skip-extraction`	No	Skip data extraction, use existing Zarr files
`--skip-visualization`	No	Skip map and chart generation
`--skip-sankey`	No	Skip Sankey diagram creation
`--skip-transition-viz`	No	Skip transition visualization creation
`--verbose`	No	Enable detailed logging output

Programmatic Usage

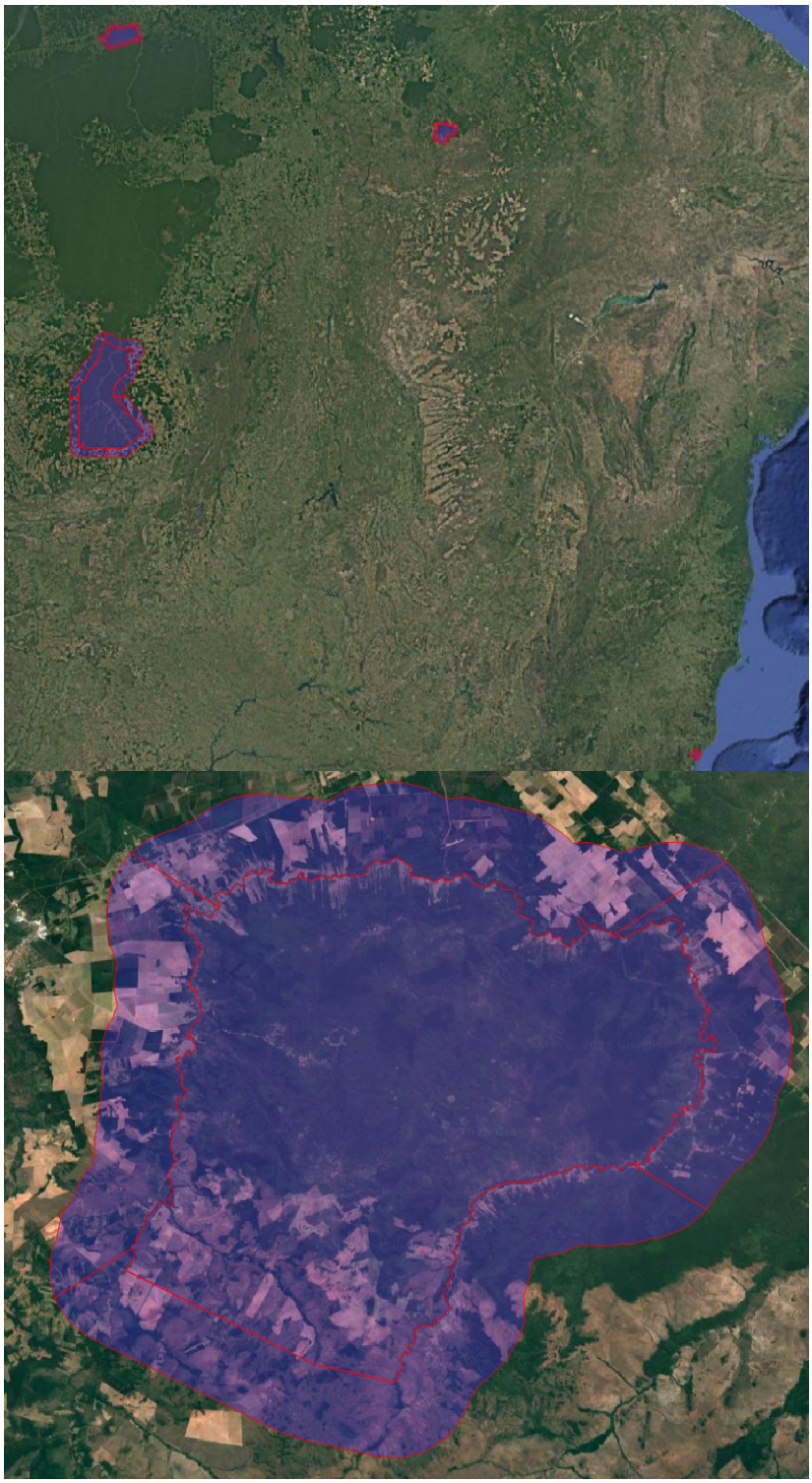
```
from main import run_batch_processing

success = run_batch_processing(
    geojson_path="analysis_areas.geojson",
    vrt_path="/path/to/mapbiomas.vrt",
    output_dir="./results",
    verbose=True
)
```

Input Data Requirements

GeoJSON File

- **Format**: Standard GeoJSON with polygon geometries
- **Coordinate System**: WGS84 (EPSG:4326)
- **Features**: Single or multiple polygons supported
- **Properties**: Optional attributes will be preserved



Example structure:

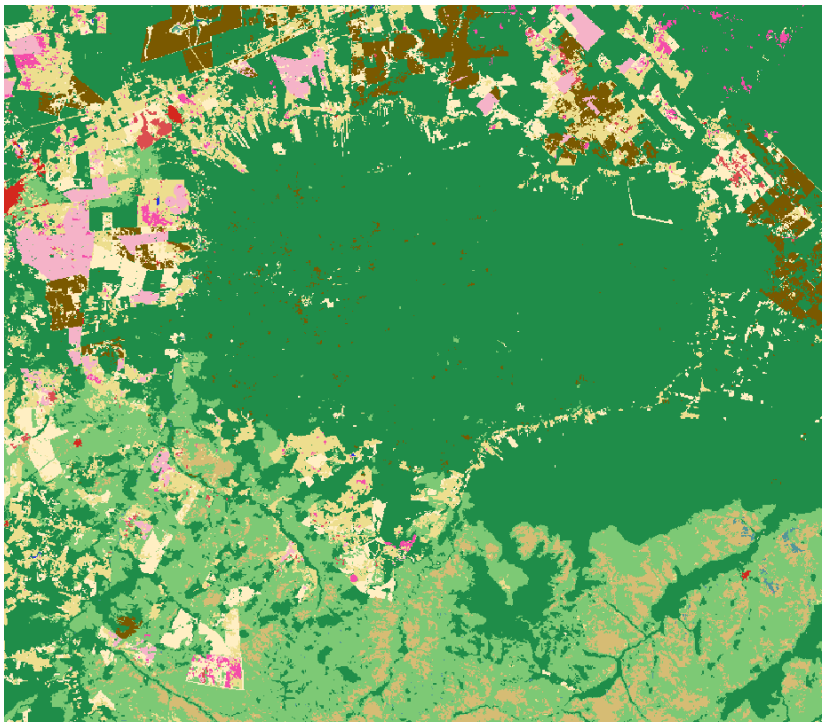
```
{
  "type": "FeatureCollection",
  "features": [
    {
      "type": "Feature",
      "properties": {
        "name": "Study Area 1"
      },
      "geometry": {
        "type": "Polygon",
```

```
          "coordinates": [[[-50.0, -10.0], [-49.0, -10.0], [-49.0, -9.0],
[-50.0, -9.0], [-50.0, -10.0]]]
        }
      }
    ]
}
```

VRT File (Virtual Raster)

```
<VRTDataset rasterXSize="154470" rasterYSize="146483">
  <SRS dataAxisToSRSAxisMapping="2,1">GEOGCS["WGS
84",DATUM["WGS_1984",SPHEROID["WGS 84",
6378137,298.257223563,AUTHORITY["EPSG","7030"]],AUTHORITY["EPSG","6326"]],PRIMEM
["Greenwich",0,AUTHORITY["EPSG","8901"]],UNIT["degree",
0.0174532925199433,AUTHORITY["EPSG","9122"]],AXIS["Latitude",NORTH],AXIS["Longi-
tude",EAST],AUTHORITY["EPSG","4326"]]</SRS>
  <GeoTransform> -7.4020999748391759e+01,  2.6949458523585642e-04,
0.0000000000000000e+00,  5.4357057842072249e+00,  0.0000000000000000e+00,
-2.6949458523585642e-04</GeoTransform>
  <VRTRasterBand dataType="Byte" band="1">
    <Metadata>
      <MDI key="STATISTICS_APPROXIMATE">YES</MDI>
      <MDI key="STATISTICS_MAXIMUM">50</MDI>
      <MDI key="STATISTICS_MEAN">7.8170763388796</MDI>
      <MDI key="STATISTICS_MINIMUM">3</MDI>
      <MDI key="STATISTICS_STDDEV">8.5054327907112</MDI>
      <MDI key="STATISTICS_VALID_PERCENT">43.17</MDI>
    </Metadata>
    <NoDataValue>0</NoDataValue>
    <ComplexSource resampling="nearest">
      <SourceFilename relativeToVRT="1">brazil_coverage_1985.tif</
SourceFilename>
      <SourceBand>1</SourceBand>
      <SourceProperties RasterXSize="154470" RasterYSize="146483"
DataType="Byte" BlockXSize="256" BlockYSize="256" />
      <SrcRect xOff="0" yOff="0" xSize="154470" ySize="146483" />
      <DstRect xOff="0" yOff="0" xSize="154470" ySize="146483" />
      <NODATA>0</NODATA>
    </ComplexSource>
  </VRTRasterBand>
  <VRTRasterBand dataType="Byte" band="2">
    <Metadata>
      <MDI key="STATISTICS_APPROXIMATE">YES</MDI>
```

- **Format**: GDAL VRT pointing to MapBiomas time series
- **Bands**: 40 bands (1985-2024, one per year)
- **Projection**: Geographic coordinates (EPSG:4326)
- **No-data values**: Properly defined for water/outside areas



Output Structure

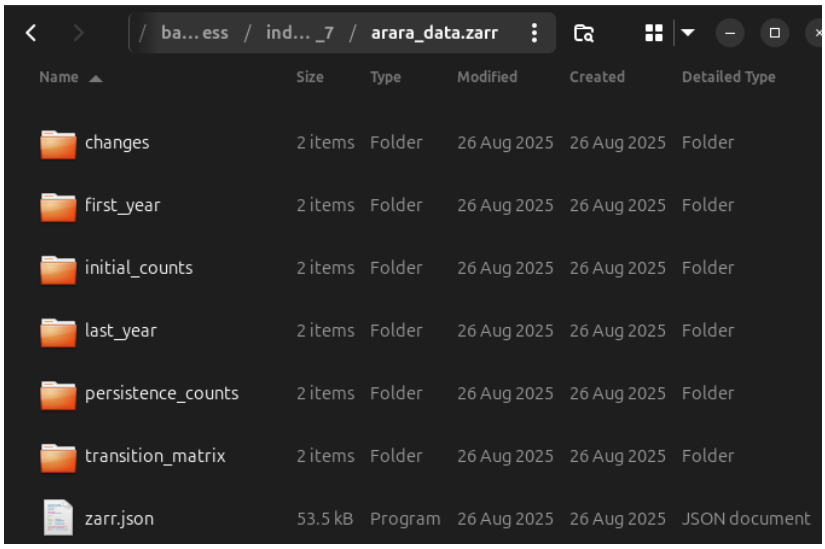
File Organization

```
output_directory/
├── polygon_name_data.zarr/           # Raw extracted data
├── polygon_name_changes_map.png      # Change frequency map
├── polygon_name_landcover_map.png    # Current land cover
├── polygon_name_initial_landcover_map.png  # 1985 land cover
├── polygon_name_sankey_diagram.html  # Interactive flow diagram
├── polygon_name_transition_matrix.csv # Complete transition data
├── polygon_name_change_statistics.csv # Summary statistics
└── polygon_name_*.pngw               # World files for GIS
```

Zarr Data Structure

Efficient compressed arrays containing:

- `changes`: Pixel-level change frequency (0-N transitions)
- `first_year`: Land cover classes in 1985
- `last_year`: Land cover classes in 2024
- `transition_matrix`: Class-to-class transition counts
- `persistence_counts`: Pixels that never changed per class
- `initial_counts`: Starting pixel counts per class

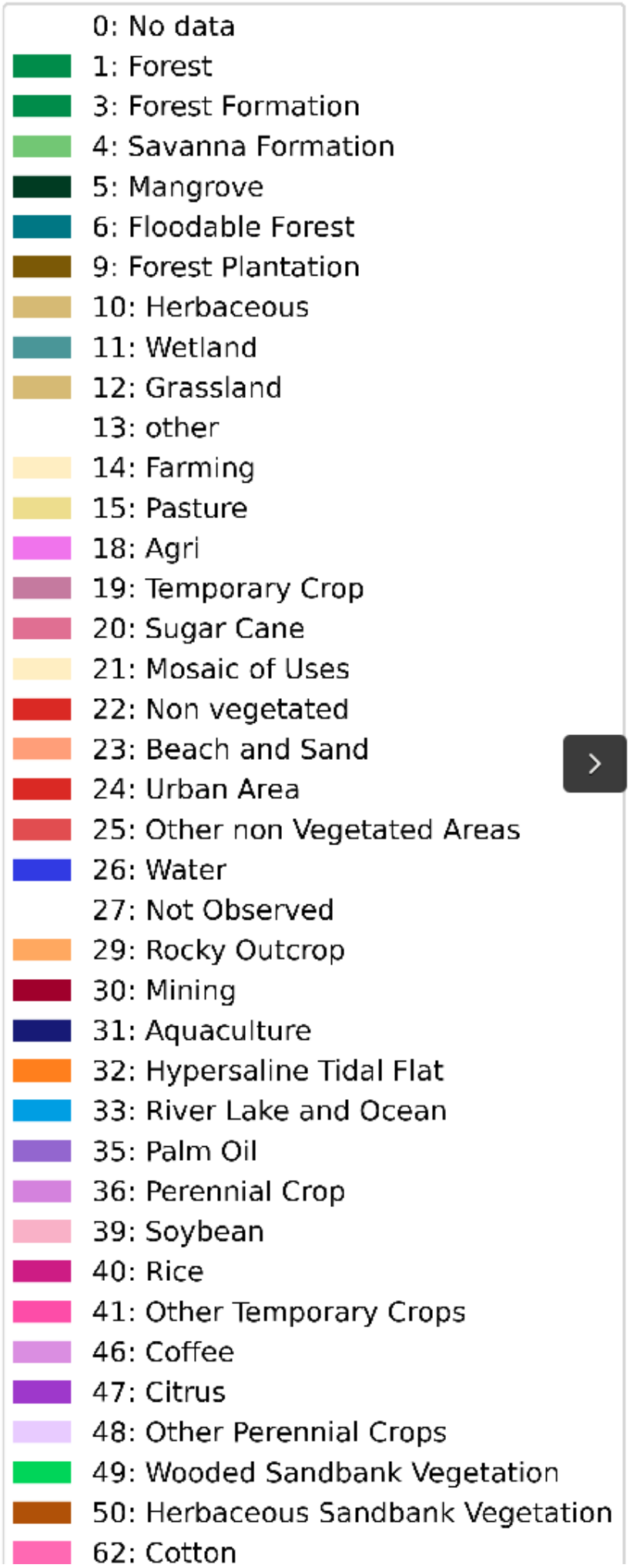


Understanding the Analysis

Land Cover Change Detection

The pipeline tracks changes by:

1. **Reading time series**: All 40 years of MapBiomas data per pixel
2. **Detecting transitions**: Comparing year-to-year land cover classes
3. **Counting changes**: Accumulating total transitions per pixel
4. **Excluding no-data**: Masking water bodies and areas outside Brazil


0: No data
1: Forest
3: Forest Formation
4: Savanna Formation
5: Mangrove
6: Floodable Forest
9: Forest Plantation
10: Herbaceous
11: Wetland
12: Grassland
13: other
14: Farming
15: Pasture
18: Agri
19: Temporary Crop
20: Sugar Cane
21: Mosaic of Uses
22: Non vegetated
23: Beach and Sand
24: Urban Area
25: Other non Vegetated Areas
26: Water
27: Not Observed
29: Rocky Outcrop
30: Mining
31: Aquaculture
32: Hypersaline Tidal Flat
33: River Lake and Ocean
35: Palm Oil
36: Perennial Crop
39: Soybean
40: Rice
41: Other Temporary Crops
46: Coffee
47: Citrus
48: Other Perennial Crops
49: Wooded Sandbank Vegetation
50: Herbaceous Sandbank Vegetation
62: Cotton

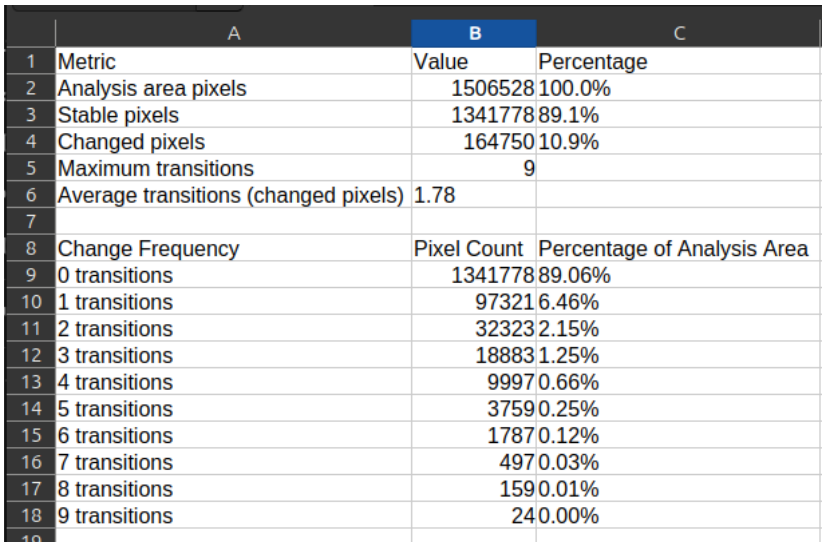
Change Frequency Scale

- **0 transitions**: Stable pixels (same class for 40 years)
- **1-3 transitions**: Low change frequency (typical for protected areas)
- **4-10 transitions**: Moderate change (agricultural rotation, regeneration)
- **11+ transitions**: High change (intense land use dynamics)

Statistical Outputs

Change Statistics CSV

- **Analysis area pixels**: Total pixels within polygon
- **Stable pixels**: Pixels with 0 transitions (%)
- **Changed pixels**: Pixels with 1+ transitions (%)
- **Maximum transitions**: Highest change frequency observed
- **Change frequency distribution**: Histogram of transition counts



	A	B	C
1	Metric	Value	Percentage
2	Analysis area pixels	1506528	100.0%
3	Stable pixels	1341778	89.1%
4	Changed pixels	164750	10.9%
5	Maximum transitions	9	
6	Average transitions (changed pixels)	1.78	
7			
8	Change Frequency	Pixel Count	Percentage of Analysis Area
9	0 transitions	1341778	89.06%
10	1 transitions	97321	6.46%
11	2 transitions	32323	2.15%
12	3 transitions	18883	1.25%
13	4 transitions	9997	0.66%
14	5 transitions	3759	0.25%
15	6 transitions	1787	0.12%
16	7 transitions	497	0.03%
17	8 transitions	159	0.01%
18	9 transitions	24	0.00%

	A	B	C
1	Metric	Value	Percentage
2	Analysis area pixels	3075833	100.0%
3	Stable pixels	3059083	99.5%
4	Changed pixels	16750	0.5%
5	Maximum transitions	17	
6	Average transitions (changed pixels)	1.97	
7			
8	Change Frequency	Pixel Count	Percentage of Analysis Area
9	0 transitions	3059083	99.46%
10	1 transitions	7289	0.24%
11	2 transitions	6705	0.22%
12	3 transitions	1177	0.04%
13	4 transitions	814	0.03%
14	5 transitions	219	0.01%
15	6 transitions	195	0.01%
16	7 transitions	90	0.00%
17	8 transitions	91	0.00%
18	9 transitions	52	0.00%
19	10 transitions	48	0.00%
20	11 transitions	33	0.00%
21	12 transitions	13	0.00%
22	13 transitions	11	0.00%
23	14 transitions	4	0.00%
24	15 transitions	3	0.00%
25	16 transitions	2	0.00%
26	17 transitions	4	0.00%

Transition Matrix CSV

- **From/To classes**: Land cover transitions between all classes
- **Pixel counts**: Number of pixels for each transition type
- **Temporal aggregation**: Summed across all 40 years

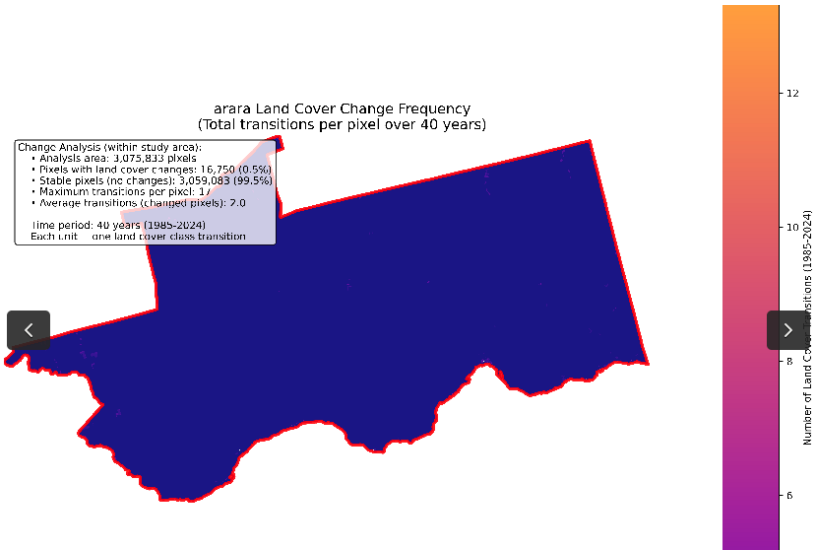
	A	B	C	D
1	Source_Class	Target_Class	Pixels	Percentage_of_Source
2	0: No data	0: No data	1575	55.59%
3	0: No data	3: Forest Formation	325	11.47%
4	0: No data	6: Floodable Forest	115	4.06%
5	0: No data	11: Wetland	17	0.60%
6	0: No data	15: Pasture	623	21.99%
7	0: No data	25: Other non Vegetated Areas	60	2.12%
8	0: No data	33: River Lake and Ocean	107	3.78%
9	0: No data	41: Other Temporary Crops	11	0.39%
10	3: Forest Formation	0: No data	627	0.00%
11	3: Forest Formation	3: Forest Formation	190174943	99.52%
12	3: Forest Formation	6: Floodable Forest	700	0.00%
13	3: Forest Formation	11: Wetland	379	0.00%
14	3: Forest Formation	12: Grassland	6	0.00%
15	3: Forest Formation	15: Pasture	918738	0.48%
16	3: Forest Formation	24: Urban Area	10	0.00%
17	3: Forest Formation	25: Other non Vegetated Areas	3919	0.00%
18	3: Forest Formation	33: River Lake and Ocean	1063	0.00%
19	3: Forest Formation	41: Other Temporary Crops	132	0.00%
20	6: Floodable Forest	0: No data	143	0.00%
21	6: Floodable Forest	3: Forest Formation	668	0.01%
22	6: Floodable Forest	6: Floodable Forest	830791	499.57%
23	6: Floodable Forest	11: Wetland	8200	0.10%
24	6: Floodable Forest	12: Grassland	6	0.00%
25	6: Floodable Forest	15: Pasture	1893	0.02%
26	6: Floodable Forest	25: Other non Vegetated Areas	21	0.00%
27	6: Floodable Forest	33: River Lake and Ocean	25219	0.30%
28	6: Floodable Forest	41: Other Temporary Crops	36	0.00%
29	11: Wetland	0: No data	22	0.01%
30	11: Wetland	3: Forest Formation	274	0.09%

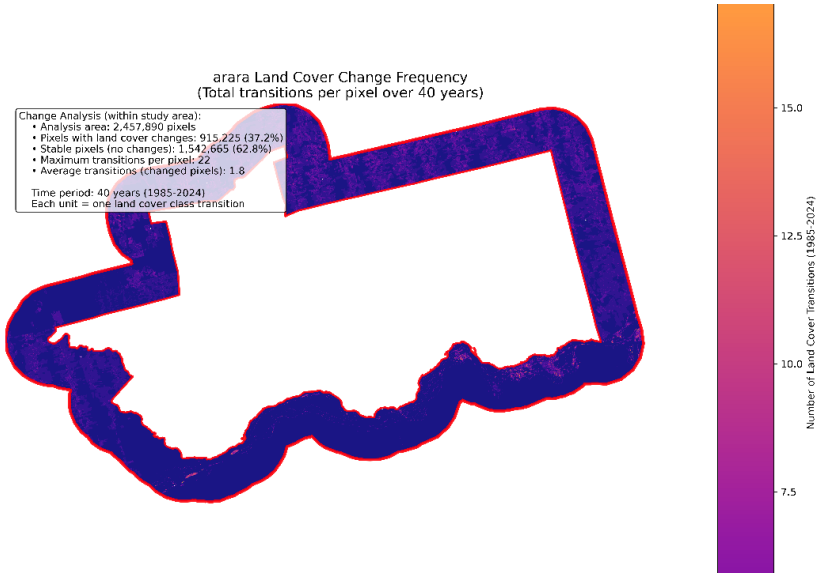
Visualization Guide

Change Maps

- **Color scale**: Purple (low) to yellow (high) change frequency
- **Masked areas**: Gray for outside polygon, white for no-data
- **Red outline**: Polygon boundary overlay

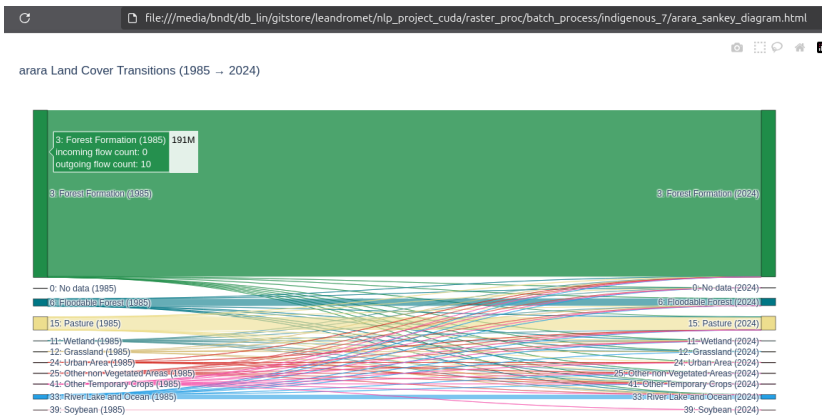
- **Statistics box**: Key metrics displayed on map

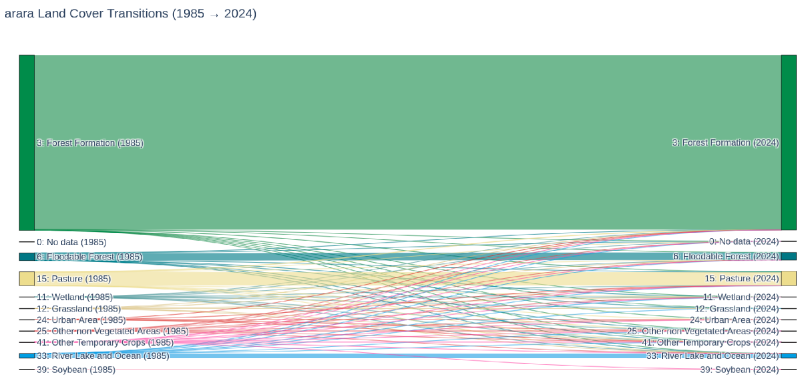




Sankey Diagrams

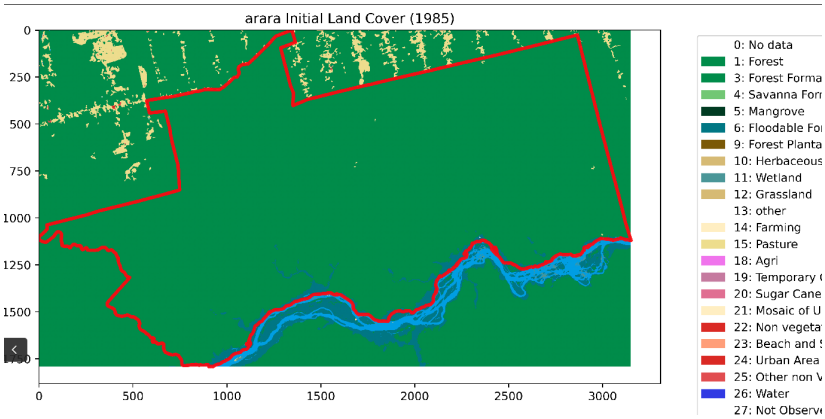
- **Flow width**: Proportional to number of pixels transitioning
- **Color coding**: Follows MapBiomas class colors
- **Interactive**: Hover for details, click to highlight flows
- **Aggregated**: Shows net transitions between classes

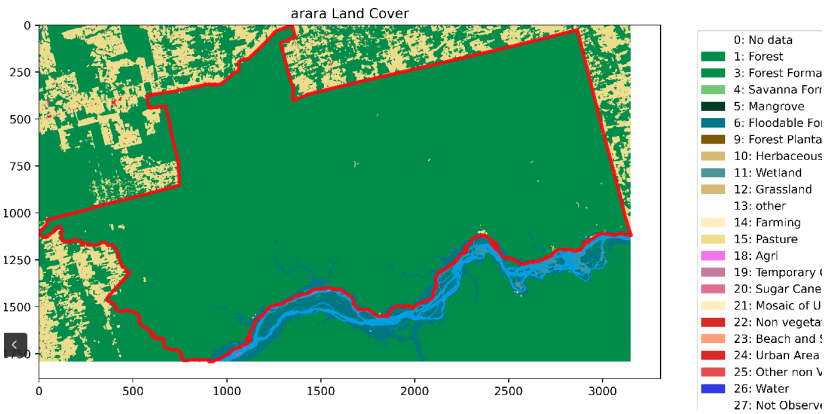




Land Cover Maps

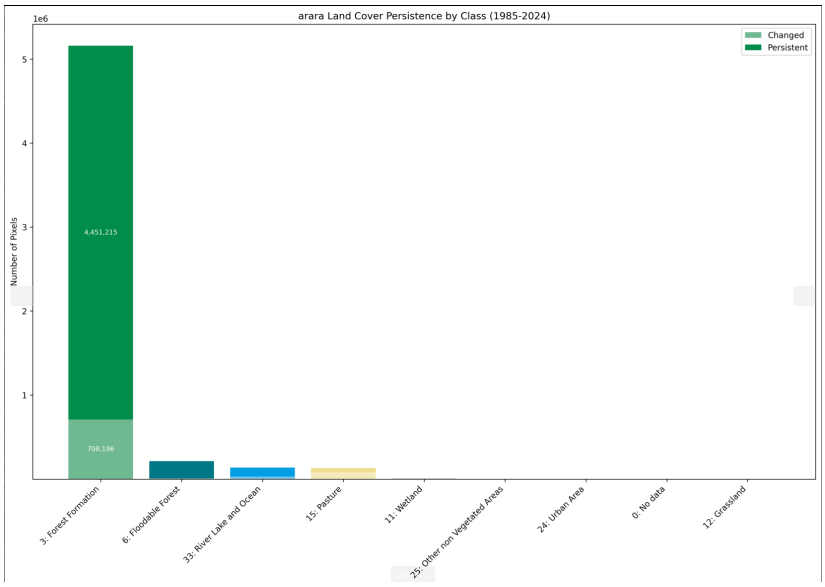
- **MapBiomas colors**: Standard color scheme for land cover classes
- **Legend**: Class codes and descriptions
- **Georeferenced**: Includes .pngw files for GIS import

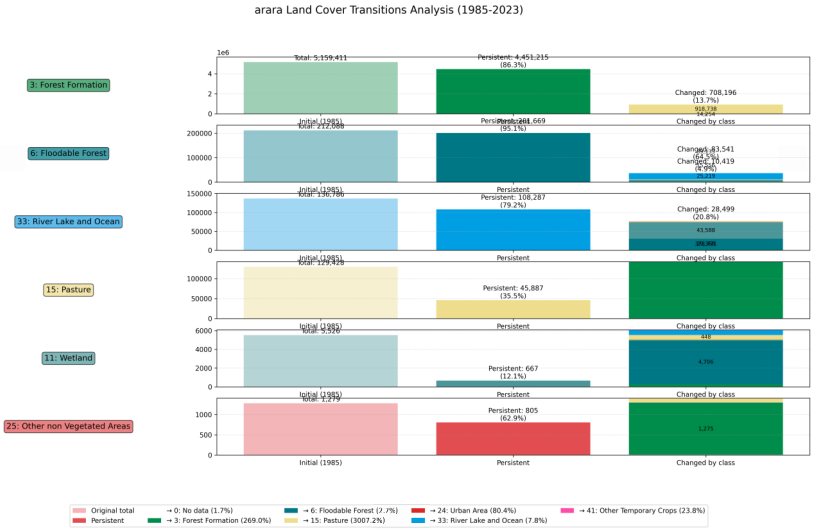


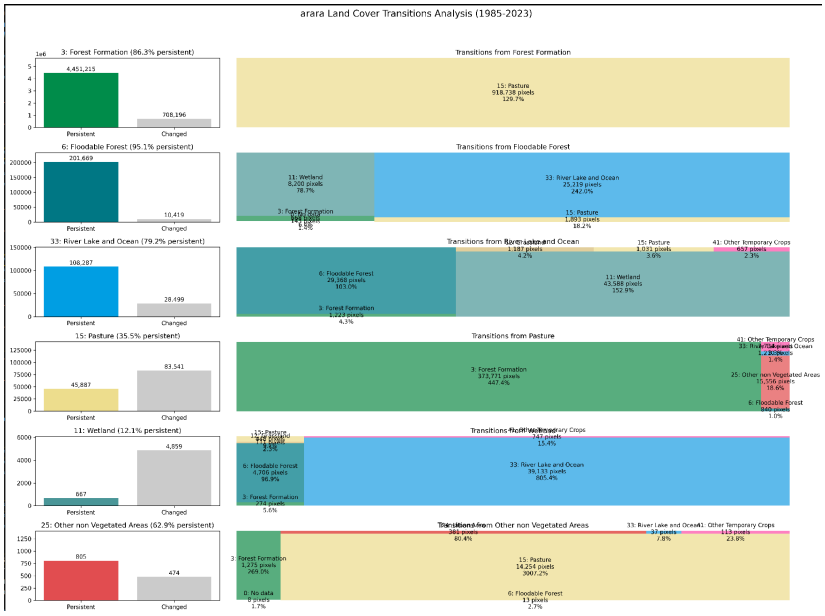


Other Visualization Products - Charts

- **Time series**: Land cover class area evolution over time
- **Bar charts**: Comparative analysis of key land cover changes
- **Pie charts**: Proportional representation of land cover distribution
- **Change frequency**: Distribution of pixel transition counts







Performance Considerations

Memory Requirements

- **Small polygons** (<1000 km²): 4-8 GB RAM

- **Medium polygons** (1000-5000 km²): 8-16 GB RAM
- **Large polygons** (>5000 km²): 16+ GB RAM

Processing Time

- **Data extraction**: 1-10 minutes per polygon (depends on size)
- **Visualizations**: 30 seconds - 2 minutes per polygon
- **Sankey diagrams**: 10-30 seconds per polygon

Optimization Tips

- Use `--skip-extraction` to regenerate only visualizations if ZARR already available
- Process large regions in smaller chunks
- Ensure sufficient disk space (1-5 GB per large polygon)
- Monitor memory usage with `--verbose` flag

Configuration

Color Maps and Labels

Edit `config.py` to customize:

- Land cover class colors
- Class label descriptions
- Processing parameters
- File paths and defaults

Processing Parameters

Key configurable values:

- `PROCESSING_TILE_SIZE`: Memory vs. speed trade-off
- `MAX_WORKERS`: Parallel processing threads
- `FILL_VALUE`: No-data value handling

Troubleshooting

Common Issues

Memory Errors

```
RuntimeError: out of memory
```

Solutions:

- Reduce `PROCESSING_TILE_SIZE` in config.py
- Decrease `MAX_WORKERS`
- Process smaller polygons
- Add more RAM or swap space

Projection Mismatches

```
Warning: CRS mismatch between data sources
```

Solutions:

- Ensure GeoJSON is in WGS84 (EPSG:4326) or SIRGAS 2000 South America - EPSG:4674
- Verify VRT coordinate system
- Check polygon bounds are within Brazil

Missing Dependencies

```
ImportError: No module named 'rasterio'
```

Solutions:

- Install GDAL system libraries first
- Use conda for complex geospatial dependencies
- Check virtual environment activation

Empty Results

```
No valid analysis area found!
```

Solutions:

- Verify polygon intersects with MapBiomas data extent
- Check for very small polygons (<1 pixel)
- Ensure polygon coordinates are valid

Debug Mode

Enable detailed logging:

```
python main.py --verbose [other parameters]
```

Scientific Applications

Use Cases

- **Deforestation monitoring**: Track forest loss in protected areas
- **Agricultural expansion**: Analyze cropland change patterns
- **Urban growth**: Monitor city expansion over time

- **Restoration assessment**: Evaluate reforestation success
- **Policy impact**: Before/after analysis of interventions

Data Quality Considerations

- **MapBiomas accuracy**: ~85% overall classification accuracy
- **Temporal consistency**: Some year-to-year noise expected
- **Spatial resolution**: 30m pixels (0.09 hectares)
- **Class definitions**: Follow MapBiomas collection methodology

Citation

When using this software, please cite:

- MapBiomas Project for the underlying land cover data
- This analysis pipeline (provide repository DOI if applicable)

Support

For questions and issues:

1. Check this documentation
2. Review log files with `--verbose` flag
3. Ensure input data meets requirements
4. Open an issue in the project repository

License

[Specify your license here]

Version: 1.0
Last Updated: August 2025
Compatible with: MapBiomas Collection 8.0+