Project Proposal and EDA

Gunpreet Singh, Gurleen Pelia, and Mariyam Ahmed

Urban Sprawl using Land Cover Data of Canada

Purpose of Project

Land Cover Data : Canada's land cover data is important for several reasons. By offering comprehensive information about ecosystems and forest cover, it plays a crucial role in environmental management and conservation, enabling the preservation of biodiversity and sustainable forest management. Studies on climate change require these data in order to evaluate the effects on ecosystems, vegetation patterns, and carbon sequestration. Land cover data helps manage the risk of disasters, identify locations that are ideal for development while maintaining green spaces, and promote sustainable growth in urban planning and development. All things considered, land cover data is critical to comprehending Canada's landscapes, aiding in conservation efforts, and advancing sustainable development in a number of industries.

Land cover data and urban sprawl are intricately related in thus project that involve studying urban expansion and its impacts on the environment. Understanding this relationship is crucial for analyzing patterns, predicting future changes, and planning sustainable urban development.

Urban Sprawl : Urban sprawl refers to the unplanned and uncontrolled growth of urban areas into neighboring rural areas, which leads to dispersed, low-density development. Longer commutes, a greater reliance on cars, and higher traffic congestion are the results of this phenomenon. It destroys open areas, farms, and natural habitats, upsetting biodiversity and local ecosystems. Infrastructure is also strained by urban sprawl, necessitating large-scale, expensive expenditures in public utilities, roads, and other services. It also adds to environmental problems including pollution and higher emissions of greenhouse gases. Overall, urban sprawl poses serious obstacles to environmental preservation and sustainable growth.

How Land Cover Data Helps Study Urban Sprawl?

1. Identifying Urban Areas:

- Land cover data allows you to identify and delineate urban areas from other land cover types.
- By classifying satellite images, you can determine the extent of urbanization.

2. Temporal Analysis:

- By comparing land cover data from different time periods, you can analyze changes in land use and identify trends in urban expansion.
- This helps in understanding the rate and pattern of urban sprawl.

3. Change Detection:

- Land cover change detection techniques enable you to quantify the conversion of non-urban areas to urban areas.

- This information is crucial for assessing the impact of urban sprawl on natural resources and agricultural land.

4. Modeling and Prediction:

- Using historical land cover data, you can develop models (like CA-Markov) to predict future urban growth.
- This helps in planning and implementing sustainable urban development strategies.

5. Impact Assessment:

- Analyzing land cover changes allows you to assess the environmental impacts of urban sprawl, such as deforestation, loss of biodiversity, and soil degradation.

About the dataset

The CCRS Land Cover dataset is crucial for our project due to its high accuracy, detailed resolution, and comprehensive national coverage. This dataset's 30-meter spatial resolution provides precise and reliable land cover information, essential for a wide range of environmental applications such as climate impact studies, emergency response planning, and wildlife habitat analysis. The dataset spans multiple years (2010, 2015, and 2020), enabling temporal analysis of land cover changes over time. Additionally, its integration into the North American Land Change Monitoring System (NALCMS) underscores its credibility and relevance for both national and regional studies. The solid and accurate data foundation provided by the CCRS Land Cover dataset guarantees that our project is founded on a dependable resource, improving the caliber and significance of our research and assisting with well-informed environmental management and policy decisions.

Description of Data Sources

Primary Data Source: • Source Name: Canada Centre for Remote Sensing (CCRS) Land Cover Map • Description: This dataset provides detailed land cover information for Canada at a 30 m spatial resolution. It is produced using observations from the Operational Land Imager (OLI) Landsat sensor and covers the years 2010, 2015, and 2020. The data is essential for a wide range of environmental applications, including climate impact studies, emergency response, and wildlife habitat analysis. • Citations: Canada Centre for Remote Sensing. "Land Cover Map of Canada (2010, 2015, 2020)." Produced under the North American Land Change Monitoring System (NALCMS).

Secondary Data Source: • Source Name: National Census Data (2023) • Description: This dataset includes demographic information for residents of Canada, such as age distribution, household income, and population density. It is used to analyze the relationship between land cover and population demographics. • Citations: Government of Canada. "National Census Data, 2023." https://www.census.gc.ca

Collection of Land Cover products for Canada as produced by Natural Resources Canada using Landsat satellite imagery. This collection of cartographic products offers classified Land Cover of Canada at a 30 metre scale, updated on a 5 year basis. * Landcover of Canada 2010 * Landcover of Canada 2015 * Landcover of Canada 2020

Data Pre-Processing

```r
library(raster)
```

```
## Loading required package: sp
```

```r
library(sf)
```

```
## Linking to GEOS 3.11.0, GDAL 3.5.3, PROJ 9.1.0; sf_use_s2() is TRUE
```

```r
library(tmap) # for making the nice plot
```

```
## Breaking News: tmap 3.x is retiring. Please test v4, e.g. with
## remotes::install_github('r-tmap/tmap')
```

```r
library(osmdata)
```

```
## Data (c) OpenStreetMap contributors, ODbL 1.0. https://www.openstreetmap.org/copyright
```

```r
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.0     v stringr   1.5.1
## v ggplot2   3.5.1     v tibble    3.2.1
## v lubridate 1.9.3     v tidyr     1.3.1
## v purrr     1.0.2
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x tidyr::extract() masks raster::extract()
## x dplyr::filter()  masks stats::filter()
## x dplyr::lag()     masks stats::lag()
## x dplyr::select()  masks raster::select()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```r
library(spDataLarge)
library(terra)
```

```
## terra 1.7.79
##
## Attaching package: 'terra'
##
## The following object is masked from 'package:tidyr':
##
##     extract
```

```r
library(spDataLarge)
library(stars)
```

```
## Loading required package: abind
```

```r
library(ggplot2)
library(dplyr)

land_coverage_2010 <- "/Users/gunpreetsingh/Desktop/landcover-2010-classification.tif" |>
  rast()

# Set the CRS to NAD83(CSRS) / Canada Atlas Lambert (EPSG:3979)
crs(land_coverage_2010) <- "EPSG:3979"

# Define the extent of Ontario in lat/lon
ontario_extent_latlon <- ext(-95.1537, -74.3202, 41.6766, 56.8566)

# Create a SpatVector with the extent in lat/lon
ontario_extent_vect <- vect(ontario_extent_latlon, crs = "EPSG:4326")

# Transform the extent to the raster's CRS
ontario_extent_transformed <- project(ontario_extent_vect, "EPSG:3979")

# Extract the transformed extent coordinates
ontario_extent <- ext(ontario_extent_transformed)

# Crop the raster to the transformed extent
land_coverage_2010_cropped <- crop(land_coverage_2010, ontario_extent)
```

```
## |---------|---------|---------|---------|=========================================
```

```r
# Downsample the cropped raster (optional)
land_coverage_2010_downsampled <- aggregate(land_coverage_2010_cropped, fact = 16)
```

```
## |---------|---------|---------|---------|=========================================
```

Visualizations

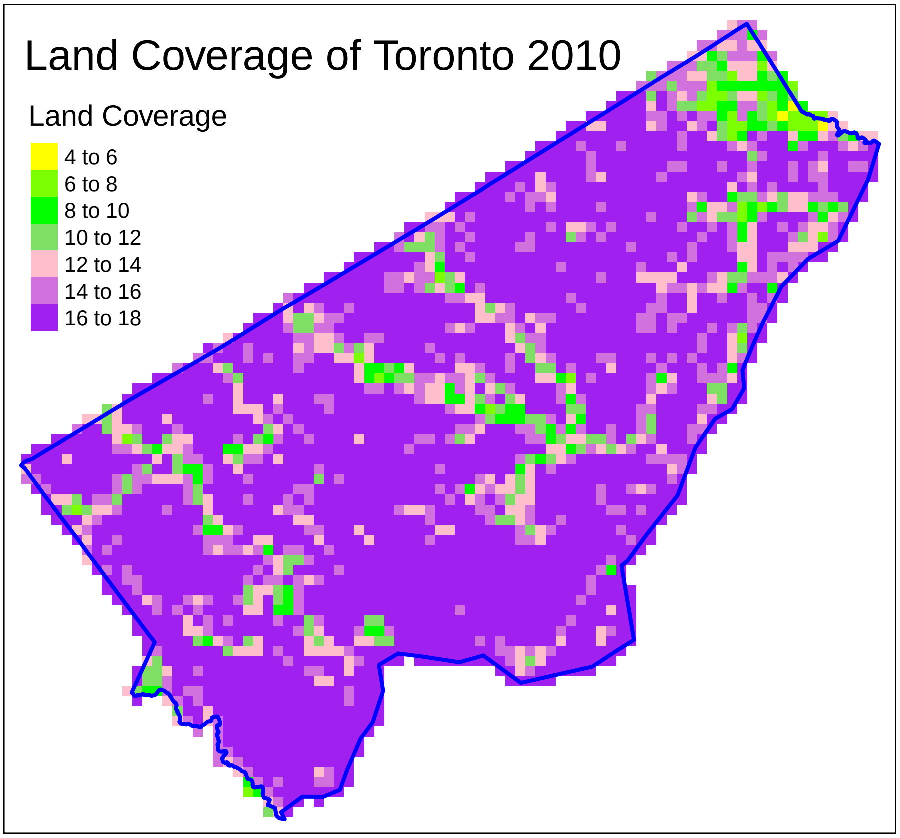
```r
# Define the extent of Toronto in lat/lon
toronto_extent_latlon <- ext(-79.6393, -79.1152, 43.5810, 43.8555)

# Create a SpatVector with the extent in lat/lon
toronto_extent_vect <- vect(toronto_extent_latlon, crs = "EPSG:4326")

# Transform the extent to the raster's CRS
toronto_extent_transformed <- project(toronto_extent_vect, "EPSG:3979")

# Extract the transformed extent coordinates
toronto_extent <- ext(toronto_extent_transformed)

# Crop the raster to the transformed extent
land_coverage_2010_toronto <- crop(land_coverage_2010_cropped, toronto_extent) |>
  aggregate(fact = 16)
```

```r
# Define the bounding box coordinates
bbox <- st_bbox(c(xmin = -79.6393, ymin = 43.5810, xmax = -79.1169, ymax = 43.8555),
                crs = st_crs(3979))

toronto_area = toronto_area <- opq(bbox) |>
  add_osm_feature(key="boundary", value = "administrative") |>
  add_osm_feature(key="name",value = "Toronto") |>
  add_osm_feature(key="admin_level",value = "6") |>
  add_osm_feature(key="capital",value = "4") |>
  osmdata_sf() |>
  (\(x) x$osm_multipolygons)() |>
  select()
```

```r
toronto_area <- st_transform(toronto_area,crs="EPSG:3979")
toronto_area
```

```
## Simple feature collection with 1 feature and 0 fields
## Geometry type: MULTIPOLYGON
## Dimension:     XY
## Bounding box:  xmin: 1254225 ymin: -454850.2 xmax: 1295082 ymax: -416973.6
## Projected CRS: NAD83(CSRS) / Canada Atlas Lambert
##                          geometry
## 1 MULTIPOLYGON (((1266778 -45...
```

```r
toronto_area_2010_masked = crop(land_coverage_2010_toronto,toronto_area) |>
  mask(toronto_area)

color_palette <-  c("yellow", "green","pink","purple")

tm_shape(toronto_area_2010_masked) +
  tm_raster(palette = color_palette, title = "Land Coverage") +
  tm_shape(toronto_area) +
  tm_borders(col = "blue", lwd = 2) +
  tm_layout(title = "Land Coverage of Toronto 2010")
```

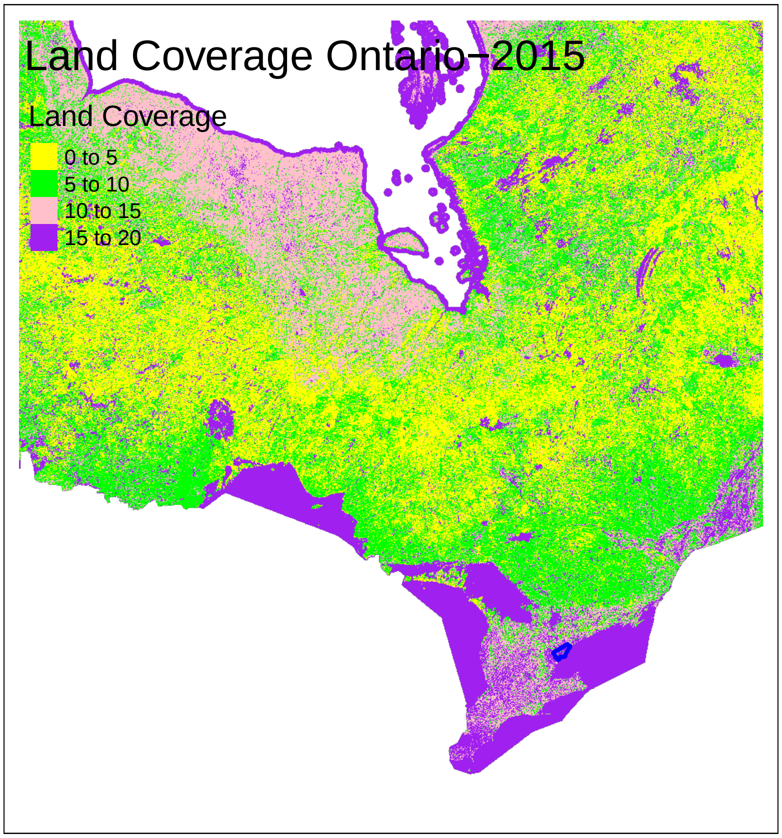



```r
land_coverage_2015 <- "/Users/gunpreetsingh/Desktop/landcover-2015-classification.tif" |>
  rast()

# Set the CRS to NAD83(CSRS) / Canada Atlas Lambert (EPSG:3979)
crs(land_coverage_2015) <- "EPSG:3979"

# Crop the raster to the transformed extent
land_coverage_2015_cropped <- crop(land_coverage_2015, ontario_extent) |>
  aggregate(fact = 16)
```

```
## |---------|---------|---------|---------|=========================================
```

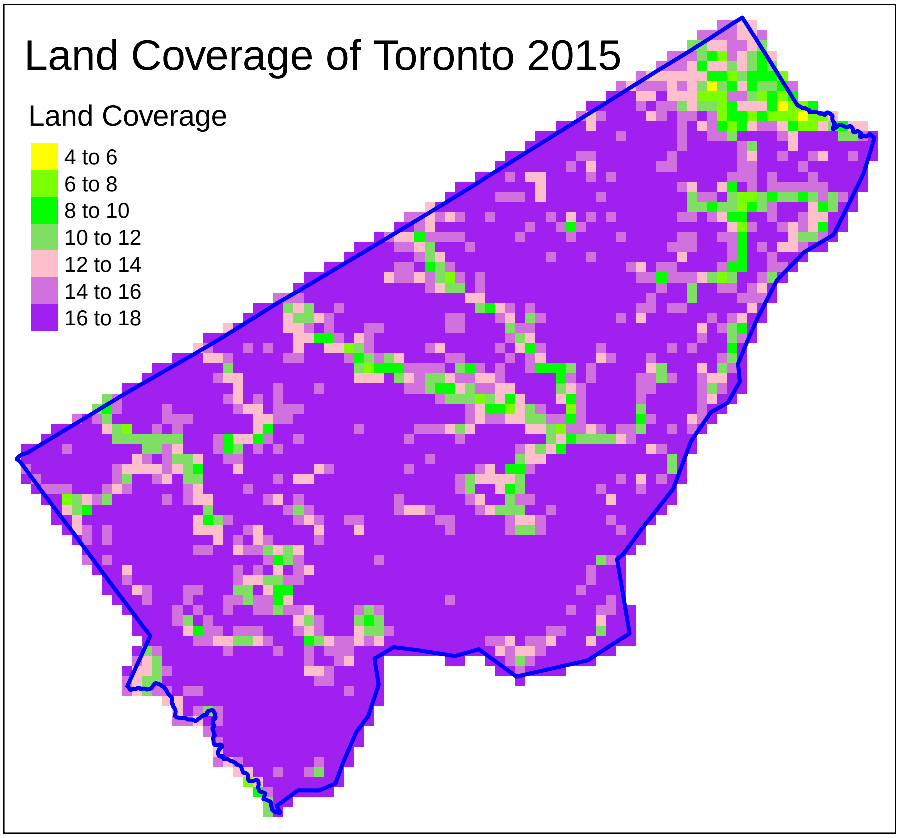
```r
tm_shape(land_coverage_2015_cropped) +
  tm_raster(palette = color_palette, title = "Land Coverage") +
  tm_shape(toronto_area) +
  tm_borders(col = "blue", lwd = 2) +
  tm_layout(title = "Land Coverage Ontario-2015")
```

```
## stars object downsampled to 967 by 1035 cells. See tm_shape manual (argument raster.downsample)
```




```r
# Crop the raster to the transformed extent
land_coverage_2015_toronto <- crop(land_coverage_2015_cropped, toronto_extent)
```

```r
toronto_area_2015_masked = crop(land_coverage_2015_toronto,toronto_area) |>
  mask(toronto_area)

color_palette <-  c("yellow", "green","pink","purple")

tm_shape(toronto_area_2015_masked) +
  tm_raster(palette = color_palette, title = "Land Coverage") +
  tm_shape(toronto_area) +
  tm_borders(col = "blue", lwd = 2) +
  tm_layout(title = "Land Coverage of Toronto 2015")
```

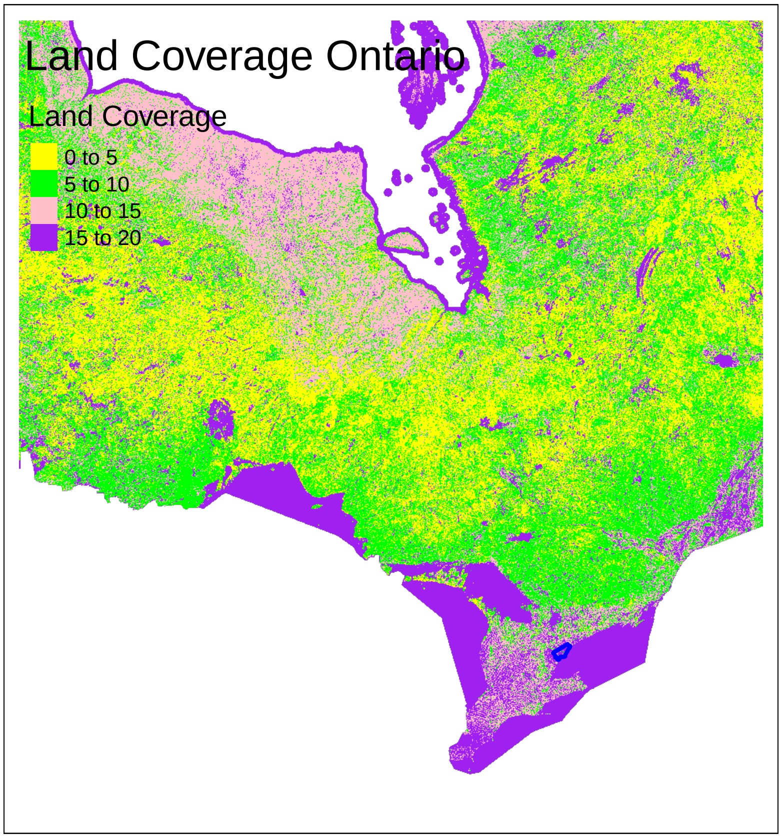



```
land_coverage_2020 <- "/Users/gunpreetsingh/Desktop/landcover-2020-classification.tif" |>
  rast()

# Set the CRS to NAD83(CSRS) / Canada Atlas Lambert (EPSG:3979)
crs(land_coverage_2020) <- "EPSG:3979"

# Crop the raster to the transformed extent
land_coverage_2020_cropped <- crop(land_coverage_2020, ontario_extent) |>
  aggregate(fact = 16)
```

```
## |---------|---------|---------|---------|=========================================
```

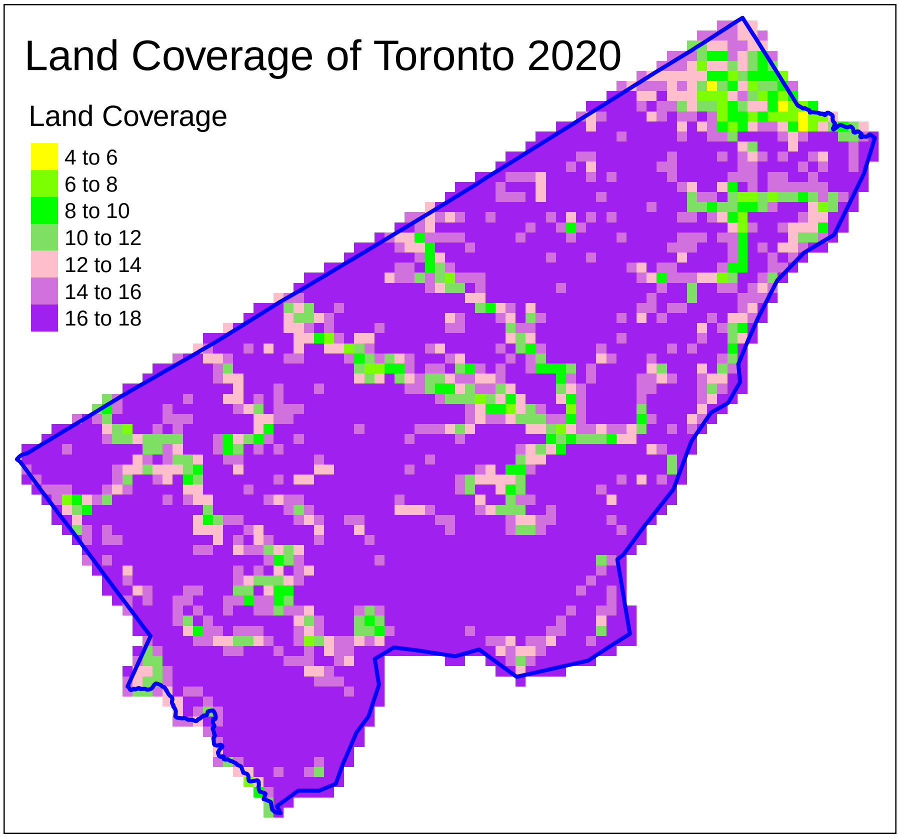
```
tm_shape(land_coverage_2020_cropped) +
  tm_raster(palette = color_palette, title = "Land Coverage") +
  tm_shape(toronto_area) +
  tm_borders(col = "blue", lwd = 2) +
  tm_layout(title = "Land Coverage Ontario")
```

```
## stars object downsampled to 967 by 1035 cells. See tm_shape manual (argument raster.downsample)
```




```r
# Crop the raster to the transformed extent
land_coverage_2020_toronto <- crop(land_coverage_2020_cropped, toronto_extent)
```

```r
toronto_area_2020_masked = crop(land_coverage_2020_toronto,toronto_area) |>
  mask(toronto_area)

color_palette <-  c("yellow", "green","pink","purple")

tm_shape(toronto_area_2020_masked) +
  tm_raster(palette = color_palette, title = "Land Coverage") +
  tm_shape(toronto_area) +
  tm_borders(col = "blue", lwd = 2) +
  tm_layout(title = "Land Coverage of Toronto 2020")
```



Analysis Methods/Modelling

A-Markov modeling is a powerful method for simulating and predicting land use changes over time by combining the strengths of Cellular Automata (CA) and Markov Chains.

Cellular Automata (CA) : Cellular Automata are dynamic models that simulate the evolution of spatial patterns over time. They work on a grid of cells, where each cell can be in one of a finite number of states (e.g., different land cover types). The state of each cell at the next time step is determined by a set of rules that consider the current state of the cell and the states of its neighboring cells. #### Markov Chains: Markov Chains are stochastic models that describe the probability of transitioning from one state to another over time. They are used to model the likelihood of land cover change based on historical data. The transition probabilities are typically derived from observed changes in land cover over a specified period.

CA-Markov Model: The CA-Markov model integrates Cellular Automata and Markov Chains to predict future land cover changes. The Markov Chain component provides the transition probabilities, while the Cellular Automata component spatially allocates these changes based on local neighborhood rules.

Steps to Implement CA-Markov Modeling

1. Data Preparation:

- Collected temporal and spatial land cover data
- Preprocessed the data

2. Land Cover Classification:

- Classified the land cover data into different categories

3. Transition Probability Matrix:

- Calculated the transition probabilities using historical land cover data.

4. Define CA Rules:

- Established rules that determine how cells change state based on their neighbors.

5. Model Calibration and Validation:

- Validated the model using a known time period to ensure accuracy.
- Adjusted parameters to improve the model's predictive capability.

6. Prediction:

- Used the calibrated model to predict future land cover changes